



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Τεχνικές Μηχανικής Μάθησης για Διαχείριση Γνώσης σε  
Συστήματα Ανάκτησης Εικόνων**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Απόστολος Π. Μαρακάκης

Αθήνα, Ιούνιος 2009





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

## Τεχνικές Μηχανικής Μάθησης για Διαχείριση Γνώσης σε Συστήματα Ανάκτησης Εικόνων

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Απόστολος Π. Μαρακάκης

**Συμβουλευτική Επιτροπή :** Ανδρέας-Γεώργιος Σταφυλοπάτης  
Στέφανος Κόλλιας  
Αριστείδης Λύκας

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 24<sup>η</sup> Ιουνίου 2009.

.....  
Α.-Γ. Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π.

.....  
Στ. Κόλλιας  
Καθηγητής Ε.Μ.Π.

.....  
Α. Λύκας  
Αν. Καθηγητής  
Πανεπιστημίου Ιωαννίνων

.....  
Ν. Γαλατσάνος  
Καθηγητής Πανεπιστημίου  
Πατρών

.....  
Γ. Στάμου  
Λέκτορας Ε.Μ.Π.

.....  
Κ. Νικήτα  
Καθηγήτρια Ε.Μ.Π.

.....  
Π. Τσανάκας  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2009

Η παρούσα διδακτορική διατριβή αποτελεί υποέργο του προγράμματος ΠΕΝΕΔ2003/ΕΠΑΝ με κωδικό 03ΕΔ32 και τίτλο «Τεχνικές Μηχανικής Μάθησης για Διαχείριση Γνώσης σε Πολυμεσικά Δεδομένα». Το πρόγραμμα ΠΕΝΕΔ2003/ΕΠΑΝ συγχρηματοδοτείται από το Ευρωπαϊκό Κοινωνικό Ταμείο (80%) και από Εθνικούς Πόρους (20%) – Γενική Γραμματεία Έρευνας και Τεχνολογίας.

.....  
Απόστολος Π. Μαρακάκης

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Απόστολος Π. Μαρακάκης, 2009.  
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

# Περιεχόμενα

<b>Περιεχόμενα</b>	<b>5</b>
<b>Κατάλογος Σχημάτων</b>	<b>11</b>
<b>Κατάλογος Πινάκων</b>	<b>13</b>
<b>Πρόλογος</b>	<b>15</b>
<b>Περίληψη</b>	<b>17</b>
<b>Abstract</b>	<b>19</b>
<b>1 Εισαγωγή</b>	<b>21</b>
1.1 Εικόνες .....	21
1.1.1 Σχηματισμός Εικόνων .....	21
1.1.2 Χώροι Χρώματος .....	22
1.1.2.1 Γραμμικοί Χώροι Χρώματος .....	23
1.1.2.2 Μη Γραμμικοί Χώροι Χρώματος .....	23
1.1.3 Ψηφιακές Εικόνες .....	25
1.2 Ανάκτηση Εικόνων .....	26
1.2.1 Κατηγορίες Ανάκτησης Εικόνων .....	26
1.2.1.1 Σημασιολογική Ανάκτηση Εικόνων .....	27
1.2.1.2 Ανάκτηση Εικόνων με Βάση το Περιεχόμενο .....	27
1.2.2 Αναπαράσταση Εικόνων .....	28
1.2.3 Διάταξη Εικόνων .....	29

1.3	Συσχετιστική Ανατροφοδότηση .....	31
1.3.1	Περιγραφή της Διαδικασίας .....	31
1.3.2	Υποβολή Παραδειγμάτων .....	32
1.3.3	Αξιοποίηση της Πληροφορίας της Ανατροφοδότησης .....	33
1.3.3.1	Χρήση Ταξινομητών .....	33
1.3.3.2	Τροποποίηση της Αναπαράστασης της Ερώτησης .....	34
1.4	Αξιολόγηση Συστημάτων Ανάκτησης Εικόνων .....	37
1.4.1	Προσομοιώσεις .....	37
1.4.2	Μέτρα Αξιολόγησης .....	38
1.5	Διάρθρωση της Διατριβής .....	39
<b>2</b>	<b>Πιθανοτικά Μοντέλα</b> .....	<b>41</b>
2.1	Βασικές Έννοιες Πιθανοτήτων .....	41
2.2	Ιστογράμματα .....	42
2.3	Γκαουσιανή Κατανομή .....	44
2.3.1	Ορισμός .....	44
2.3.2	Ιδιότητες .....	45
2.3.3	Εκτίμηση Παραμέτρων .....	46
2.4	Γκαουσιανά Μείγματα .....	47
2.4.1	Ορισμός .....	47
2.4.2	Ιδιότητες .....	48
2.4.3	Εκτίμηση Παραμέτρων .....	48
2.4.3.1	Αλγόριθμος EM .....	48
2.4.3.2	Άπληστος Αλγόριθμος EM .....	52
2.4.4	Προσδιορισμός Αριθμού Συνιστωσών .....	56
<b>3</b>	<b>Μέτρα Απόστασης Μεταξύ Γκαουσιανών Μειγμάτων</b> .....	<b>61</b>
3.1	Απόκλιση KL .....	61
3.1.1	Υπολογισμός Μέσω Τυχαίας Δειγματοληψίας .....	62
3.2	Απόσταση EMD .....	63
3.3	Μέτρο Ομοιότητας ALA .....	64
3.3.1	Προσέγγιση της Απόκλισης KL .....	65
3.4	Απόκλιση C2 .....	68
3.4.1	Ιδιότητες .....	70

<b>4</b>	<b>Συσχετιστική Ανατροφοδότηση Μέσω Τροποποίησης της Αναπαράστασης της Ερώτησης</b>	<b>73</b>
4.1	Ανανέωση του Μοντέλου της Ερώτησης .....	73
4.2	Ανανέωση του Μέτρου Απόστασης .....	75
4.3	Ενσωμάτωση των Αρνητικών Παραδειγμάτων .....	76
4.4	Κριτήριο Διάταξης Εικόνων .....	78
4.5	Μέθοδοι Συγκρινόμενες με την Προτεινόμενη .....	78
4.5.1	Παραλλαγή Βασισμένη σε Ιστογράμματα .....	79
4.5.2	Παραλλαγή Βασισμένη στο Μέτρο ALA .....	80
4.5.3	Μέθοδος <i>Kherfi</i> .....	81
4.6	Σύνολα Δεδομένων .....	83
4.6.1	Βάσεις Εικόνων .....	83
4.6.2	Οπτικά Χαρακτηριστικά .....	84
4.7	Θέματα Υλοποίησης .....	88
4.7.1	Προτεινόμενη Μέθοδος .....	88
4.7.1.1	Εκπαίδευση των Γκαουσιανών Μειγμάτων .....	88
4.7.1.2	Ανανέωση του Μοντέλου της Ερώτησης .....	90
4.7.2	Παραλλαγή Βασισμένη σε Ιστογράμματα .....	91
4.7.3	Παραλλαγή Βασισμένη στο Μέτρο ALA .....	92
4.7.4	Μέθοδος <i>Kherfi</i> .....	92
4.8	Πειράματα .....	93
4.8.1	Προσομοιώσεις .....	93
4.8.2	Αποτελέσματα .....	99
4.9	Συμπεράσματα .....	100
<b>5</b>	<b>Ταξινομητές</b>	<b>103</b>
5.1	Προβλήματα Δυαδικής Ταξινόμησης .....	103
5.2	Γραμμική Διαχωρισσιμότητα .....	104
5.3	Μηχανές Διανυσμάτων Υποστήριξης .....	105
5.3.1	Συνάρτηση Πυρήνα .....	106
5.3.2	Πρόβλημα Βελτιστοποίησης .....	109
5.3.3	Επικαλυπτόμενες Κατηγορίες .....	110
5.4	Η Συσχετιστική Ανατροφοδότηση ως Πρόβλημα Δυαδικής Ταξινόμησης .....	112
5.4.1	Συσχετιστική Ανατροφοδότηση και Μηχανές Διανυσμάτων Υποστήριξης .....	113

<b>6</b>	<b>Συσχετιστική Ανατροφοδότηση με Χρήση Ταξινομητών και Γκαουσιανών Μειγμάτων</b>	<b>115</b>
6.1	Συνδυασμός Γκαουσιανών Μειγμάτων και Μηχανών Διανυσμάτων Υποστήριξης .....	115
6.2	Προκαταρκτικά Πειράματα .....	117
6.3	Γκαουσιανά Μείγματα και Μπεϋζιανή Μάθηση .....	121
6.3.1	Μπεϋζιανή Μάθηση .....	122
6.3.2	Εκπαίδευση Γκαουσιανών Μειγμάτων Μέσω Μπεϋζιανής Μάθησης ...	123
6.3.2.1	Γενικό Γκαουσιανό Μείγμα .....	123
6.3.2.2	Αλγόριθμος MAP-EM .....	124
6.3.3	Απόσταση Μεταξύ Γκαουσιανών Μειγμάτων Εκπαιδευμένων Μέσω του Αλγορίθμου MAP-EM .....	127
6.4	Οπτικά Χαρακτηριστικά .....	128
6.4.1	Μέθοδος Υπολογισμού Χαρακτηριστικών SIFT .....	128
6.4.2	Παραλλαγή της Μεθόδου SIFT για Έγχρωμες Εικόνες .....	130
6.4.3	Τελική Περιγραφή των Εικόνων .....	132
6.5	Συγκρινόμενες Μέθοδοι .....	133
6.5.1	Προτεινόμενη Μέθοδος .....	133
6.5.2	Παραλλαγές .....	134
6.5.3	Μέθοδος Βασισμένη σε Οπτικά Λεξικά .....	135
6.6	Πειράματα .....	136
6.6.1	Αποτελέσματα .....	140
6.7	Συμπεράσματα .....	142
<b>7</b>	<b>Συσχετιστική Ανατροφοδότηση και Επιλογή Χαρακτηριστικών</b>	<b>145</b>
7.1	Επιλογή Χαρακτηριστικών .....	145
7.1.1	Εφαρμογή στη Συσχετιστική Ανατροφοδότηση .....	147
7.1.2	Μέθοδοι Επιλογής Χαρακτηριστικών .....	148
7.1.2.1	Αναδρομική Εξάλειψη Χαρακτηριστικών με Χρήση Ταξινομητών SVM .....	148
7.1.2.2	Ελάχιστος Πλεονασμός, Μέγιστη Συνάφεια .....	150
7.2	Οπτικά Χαρακτηριστικά .....	153
7.2.1	Χαρακτηριστικά Χρώματος .....	153
7.2.2	Χαρακτηριστικά Υφής .....	154
7.2.3	Χαρακτηριστικά Σχήματος .....	156
7.2.4	Τελική Αναπαράσταση των Εικόνων .....	157



7.3	Πειράματα .....	158
7.3.1	Αποτελέσματα .....	159
7.4	Συμπεράσματα .....	162
<b>8</b>	<b>Συνολικό Πόρισμα Διατριβής</b>	<b>165</b>
8.1	Γενικά Συμπεράσματα .....	165
8.2	Μελλοντικές Επεκτάσεις .....	167
<b>A</b>	<b>Αποδείξεις</b>	<b>169</b>
<b>B</b>	<b>Βάσεις Εικόνων</b>	<b>179</b>
	<b>Βιβλιογραφία</b>	<b>185</b>
	<b>Δημοσιεύσεις</b>	<b>193</b>
	<b>Βιογραφικό Σημείωμα</b>	<b>195</b>



## Κατάλογος Σχημάτων

1.1	Χώρος χρώματος HSV .....	24
2.1	Μονοδιάστατες Γκαουσιανές κατανομές για διάφορες τιμές του $\sigma$ .....	45
2.2	Ιστόγραμμα με 100 bins για ένα σύνολο παρατηρήσεων .....	55
2.3	Γκαουσιανό Μείγμα 4 συνιστωσών, εκπαιδευμένο μέσω των ίδιων παρατηρήσεων που χρησιμοποιήθηκαν και για το σχήμα 2.2 .....	56
4.1	Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης μέσω τροποποίησης της αναπαράστασης της ερώτησης (1 <sup>ο</sup> πείραμα) – Μέση ακρίβεια για επίπεδο ανάκλισης $RecLev = 0.3$ (βάση εικόνων MSDB) .....	94
4.2	Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης μέσω τροποποίησης αναπαράστασης της ερώτησης (2 <sup>ο</sup> πείραμα) – Μέση ακρίβεια για επίπεδο ανάκλισης $RecLev = 0.3$ (βάση εικόνων MSDB) .....	95
4.3	Σύγκριση μεθόδων συσχετιστικής ανατροφοδότησης που βασίζονται στην τροποποίηση της αναπαράστασης της ερώτησης – Μέση ακρίβεια για εμβέλεια $T = 20$ (βάση εικόνων MSDB) .....	96
4.4	Σύγκριση μέτρων απόστασης μεταξύ Γκαουσιανών Μειγμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης μέσω τροποποίησης της αναπαράστασης της ερώτησης – Μέση ακρίβεια για εμβέλεια $T = 20$ (βάση εικόνων MSDB) .....	97
4.5	Σύγκριση μεθόδων συσχετιστικής ανατροφοδότησης που βασίζονται στην τροποποίηση της αναπαράστασης της ερώτησης – Μέση ακρίβεια για εμβέλεια $T = 20$ (βάση εικόνων Corel9923) .....	97
4.6	Σύγκριση μέτρων απόστασης μεταξύ Γκαουσιανών Μειγμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης μέσω τροποποίησης της αναπαράστασης της ερώτησης – Μέση ακρίβεια για εμβέλεια $T = 20$ (βάση εικόνων Corel9923) .....	98
6.1	Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια για εμβέλεια $T = 10$ .....	119

6.2	Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια για εμβέλεια $T = 20$ .....	120
6.3	Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια για εμβέλεια $T = 30$ .....	120
6.4	Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Οπτικών Λεξικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια $T = 20$ .....	137
6.5	Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Οπτικών Λεξικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια $T = 30$ .....	138
6.6	Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Οπτικών Λεξικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια ανά κατηγορία μετά τον 6 <sup>ο</sup> γύρο της ανατροφοδότησης, σε εμβέλεια $T = 20$ .....	138
6.7	Μελέτη σχετικά με την επίδραση του αριθμού των Γκαουσιανών συνιστωσών των μοντέλων στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια $T = 20$ .....	139
6.8	Σύγκριση μέτρων απόστασης μεταξύ Γκαουσιανών Μειγμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια $T = 20$ .....	139
6.9	Μελέτη σχετικά με την επίδραση του χρησιμοποιούμενου συνόλου εκπαίδευσης του γενικού Γκαουσιανού Μείγματος στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια $T = 20$ .....	140
7.1	Σύγκριση μεθόδων επιλογής χαρακτηριστικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια $T = 20$ , για $K = 25$ επιλεγμένα χαρακτηριστικά .....	160
7.2	Σύγκριση μεθόδων επιλογής χαρακτηριστικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια $T = 20$ , για $K = 50$ επιλεγμένα χαρακτηριστικά .....	161
7.3	Σύγκριση μεθόδων επιλογής χαρακτηριστικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια $T = 20$ , για $K = 75$ επιλεγμένα χαρακτηριστικά .....	161
7.4	Σύγκριση μεταξύ χρήσης και μη χρήσης μεθόδου επιλογής χαρακτηριστικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια $T = 20$ , μετά τον 6 <sup>ο</sup> γύρο της ανατροφοδότησης ...	162
B.1	Ενδεικτικές εικόνες από τις κατηγορίες της βάσης εικόνων MSDB .....	179
B.2	Ενδεικτικές εικόνες από τις κατηγορίες της βάσης εικόνων Corel9923 .....	180
B.3	Ενδεικτικές εικόνες από τις κατηγορίες της βάσης εικόνων Corel4500 .....	182

## Κατάλογος Πινάκων

B.1	Αναλυτικός πίνακας των περιεχομένων ανά κατηγορία για τη βάση εικόνων MSDB .....	179
B.2	Αναλυτικός πίνακας των περιεχομένων ανά κατηγορία για τη βάση εικόνων Core19923 .....	181
B.3	Αναλυτικός πίνακας των περιεχομένων ανά κατηγορία για τη βάση εικόνων Core14500 .....	182



## ΠΡΟΛΟΓΟΣ

Η διατριβή αυτή αποτελεί το αποτέλεσμα της ερευνητικής μου δραστηριότητας πάνω στο πεδίο της μηχανικής μάθησης. Η έρευνα αυτή πραγματοποιήθηκε στο εργαστήριο Ευφυών Υπολογιστικών Συστημάτων υπό την επίβλεψη του καθηγητή κ. Ανδρέα-Γεωργίου Σταφυλοπάτη, τον οποίον ευχαριστώ θερμά για την εμπιστοσύνη που μου έδειξε, καθώς και για την υποστήριξή του τόσο σε επιστημονικό όσο και σε προσωπικό επίπεδο.

Επίσης, θα ήθελα να ευχαριστήσω ιδιαίτερος τον αναπληρωτή καθηγητή κ. Αριστείδη Λύκα, μέλος της συμβουλευτικής επιτροπής, και τον καθηγητή κ. Νικόλαο Γαλατσάνο, μέλος της εξεταστικής επιτροπής, για το ενδιαφέρον, τη στήριξη και τη βοήθειά τους στο πρόσωπο μου, και την εν γένει συνεισφορά τους σε αυτήν την προσπάθεια.

Ακόμα, οφείλω να εκφράσω τις ευχαριστίες μου στον καθηγητή κ. Στ. Κόλλια, μέλος της συμβουλευτικής επιτροπής, για το ενδιαφέρον και τη στήριξή του, καθώς και στα μέλη της εξεταστικής επιτροπής καθηγήτρια κα Κ. Νικήτα, καθηγητή κ. Π. Τσανάκα και λέκτορα κ. Γ. Στάμου για την τιμή που μου έκαναν να συμμετάσχουν στην αξιολόγηση αυτής της διατριβής.

Τέλος, θέλω να ευχαριστήσω και όλους τους υπόλοιπους που συνέβαλαν με τον έναν ή τον άλλο τρόπο σε αυτή μου την προσπάθεια, και κυρίως την οικογένειά μου για τη στήριξη, την εμπιστοσύνη και την υπομονή τους όλα αυτά τα χρόνια.

*Απόστολος Μαρακάκης  
Αθήνα, Ιούνιος 2009*





## ΠΕΡΙΛΗΨΗ

Η κατακόρυφη αύξηση του όγκου των διαθέσιμων ψηφιακών εικόνων τα τελευταία χρόνια, μέσω της εξάπλωσης της χρήσης των ψηφιακών φωτογραφικών μηχανών και της μαζικής διείσδυσης του διαδικτύου στη ζωή του μέσου ανθρώπου, έχει καθιερώσει νομοτελειακά το πρόβλημα της ανάκτησης εικόνων αποθηκευμένων σε ψηφιακή μορφή ως ένα από τα πιο ενδιαφέροντα και δημοφιλή προβλήματα στο χώρο των υπολογιστικών εφαρμογών. Η παρούσα διατριβή πραγματεύεται το πρόβλημα της ανάκτησης εικόνων από τη σκοπιά της χρήσης τεχνικών μηχανικής μάθησης για την αντιμετώπισή του.

Η διαδικασία της ανάκτησης εικόνων προϋποθέτει την πρόθεση από τη μεριά ενός χρήστη για εύρεση εικόνων που ικανοποιούν κάποιο συγκεκριμένο κριτήριο που ο ίδιος καθορίζει. Όμως, τα κριτήρια τα οποία εκφράζουν τις προτιμήσεις των χρηστών είναι σημασιολογικής υφής, και δεν υπάρχει συνήθως προφανής σύνδεση μεταξύ αυτών και των ψηφιακών δεδομένων των εικόνων. Αυτή η αναντιστοιχία μεταξύ των σημασιολογικών κριτηρίων των χρηστών και του περιεχομένου των εικόνων είναι γνωστή ως σημασιολογικό κενό, και καθιστά το πρόβλημα της ανάκτησης εικόνων με βάση το περιεχόμενο ιδιαίτερα δύσκολο.

Στην παρούσα διατριβή εστιάζουμε στην τεχνική της συσχετιστικής ανατροφοδότησης, που έχει προταθεί για την κάλυψη του προαναφερθέντος σημασιολογικού κενού. Μέσω της συσχετιστικής ανατροφοδότησης επιχειρείται η ενσωμάτωση του χρήστη στο σύστημα ανάκτησης, και η αξιοποίηση της αξιολόγησης που αυτός παρέχει πάνω στα επιστρεφόμενα από το σύστημα αποτελέσματα, με σκοπό τη διαδοχική βελτίωση των τελευταίων.

Σε ό,τι αφορά την ανάκτηση εικόνων με βάση το περιεχόμενο είναι απαραίτητη η αναπαράσταση των εικόνων με χρήση οπτικών χαρακτηριστικών που υπολογίζονται από αυτές. Για αυτό το σκοπό βασιστήκαμε σε μεγάλο βαθμό στα Γκαουσιανά Μείγματα, που αποτελούν ένα καταξιωμένο πιθανοτικό μοντέλο, για την περιγραφή της κατανομής των τοπικά υπολογιζόμενων χαρακτηριστικών των εικόνων. Το κύριο θέμα που σχετίζεται με τη χρήση Γκαουσιανών Μειγμάτων σε αυτό το πλαίσιο αφορά τον ορισμό ενός μέτρου απόστασης μεταξύ Γκαουσιανών Μειγμάτων που να μπορεί να υπολογιστεί γρήγορα. Για το λόγο αυτόν προχωρήσαμε σε μια εκτενή μελέτη γύρω από δημοφιλή και νέα μέτρα απόστασης μεταξύ Γκαουσιανών Μειγμάτων.

Μετάπειτα, δύο νέες μέθοδοι συσχετιστικής ανατροφοδότησης αναπτύχθηκαν. Και οι δύο αυτές μέθοδοι χρησιμοποιούν Γκαουσιανά Μείγματα για την αναπαράσταση των εικόνων, και ανάλογα μέτρα απόστασης που μπορούν να υπολογιστούν αποδοτικά. Η πρώτη μέθοδος χρησιμοποιεί μια πιθανοτική τεχνική ανανέωσης του μοντέλου της ερώτησης με βάση τα παραδείγματα του χρήστη, και εκμεταλλεύεται τις ιδιότητες ενός πρόσφατα προταθέντος μέτρου απόστασης, προκειμένου να επιτύχει μια πολύ γρήγορη απόκριση από τη μεριά του συστήματος. Από την άλλη, η δεύτερη μέθοδος που αναπτύχθηκε βασίζεται στη χρήση ταξινομητών, και συγκεκριμένα στις Μηχανές Διανυσμάτων Υποστήριξης, για το διαχωρισμό των σχετικών με την ερώτηση του χρήστη εικόνων από τις μη σχετικές. Επιπροσθέτως, υιοθετεί για την εκτίμηση των παραμέτρων

των Γκαουσιανών Μειγμάτων μια τεχνική που αποδεικνύεται ότι έχει πολλά πλεονεκτήματα σε σχέση με τον συνήθως χρησιμοποιούμενο τρόπο εκπαίδευσής τους.

Τέλος, όσο αφορά την περίπτωση της αναπαράστασης των εικόνων μέσω διανυσμάτων χαρακτηριστικών μεγάλης διάστασης, μελετήθηκε η δυνατότητα χρήσης μεθόδων επιλογής χαρακτηριστικών σε συνδυασμό με την υιοθέτηση ταξινομητών για την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης, με σκοπό τον περιορισμό της διάστασης των αναπαραστάσεων και τη βελτίωση της ικανότητας του ταξινομητή να διακρίνει μεταξύ των σχετικών και των μη σχετικών εικόνων.

Για την τεκμηρίωση των συμπερασμάτων σε κάθε κύριο στάδιο της έρευνας, και για κάθε προτεινόμενη μέθοδο παρατίθενται κατάλληλα σχεδιασμένα πειράματα και συγκρίσεις, που δεν παρέχουν μόνο μια συνολική αξιολόγηση των αποτελεσμάτων των μεθοδολογιών, αλλά παράλληλα φωτίζουν και δικαιολογούν τις επιμέρους επιλογές, και αναδεικνύουν τα πλεονεκτήματά τους.

# ABSTRACT

The rapid increase in the number of available digital images during recent years, due to the widespread use of digital cameras and the popularity of the internet, has definitely established the problem of image retrieval as one of the most interesting problems in the field of computer applications. The current thesis addresses the problem of image retrieval by employing machine learning techniques.

The process of image retrieval is based on the intention of a user to find images relevant to a criterion defined by him. However, the criteria expressing the preferences of the users are semantic in nature, and, thus, there is no obvious correspondence between these criteria and the digital content of the images. This problem is known as the semantic gap, and it makes content-based image retrieval a very difficult task.

In the current thesis, we focus on relevance feedback, which is a technique proposed in order to fill the aforementioned semantic gap. By relevance feedback, the incorporation of the user to the retrieval system and the exploitation of the evaluation that he provides on the retrieval results are attempted, aiming at the sequential improvement of the results.

In what concerns content-based image retrieval, there is a need of an image representation using visual features extracted from the images. For this purpose, we mainly used Gaussian Mixtures, which constitute a well-established probabilistic model, for the description of the distribution of the locally extracted image features. The main issue regarding the use of Gaussian Mixtures in this context is the definition of a distance measure between Gaussian Mixtures which can be computed in a fast way. For this reason, we proceeded to a thorough examination on popular and recently proposed distance measures between Gaussian Mixtures.

Afterwards, two new relevance feedback methods were implemented. Both of them use Gaussian Mixtures to represent the images, and efficiently computed distance measures. The first method uses a probabilistic technique to update the model of the user's query based on the user's examples. Moreover, it takes advantage of the properties of the distance measure used to achieve a very fast system response. On the other hand, the second method is based on classifiers, and particularly Support Vector Machines, to distinguish between the images that are relevant to the user's query and those that are irrelevant. Furthermore, it adopts a technique for the estimation of the parameters of Gaussian Mixtures which is proven to have various advantages compared to the standard parameter estimation methodology.

Finally, in what concerns the case of representing each image using a multi-dimensional vector, the possibility of using feature selection methods along with classifiers for the implementation of relevance feedback task is studied. The aim is the reduction of the feature space dimensionality and the increase of the classifier ability to distinguish between the relevant and irrelevant images.

In order to derive useful conclusions, at every stage of the research and for each proposed method the results of appropriately designed and performed experiments and comparisons are provided, which do not focus only on the overall evaluation of the

methodologies, but, in parallel, intend to justify the particular choices and to prove their merits.

# Κεφάλαιο 1

## Εισαγωγή

Το αντικείμενο της παρούσας διατριβής είναι η μελέτη, υλοποίηση και εφαρμογή τεχνικών μηχανικής μάθησης σε συστήματα ανάκτησης εικόνων. Ουσιαστικά, ο στόχος είναι η δημιουργία πρωτότυπων μεθόδων που θα έχουν τη δυνατότητα να διαχειριστούν και να ενσωματώσουν την πληροφορία που μπορεί να προέλθει από τη μεριά του χρήστη του συστήματος ανάκτησης εικόνων, ώστε μέσω της ενσωμάτωσης αυτής να καταστεί δυνατή η βελτίωση των αποτελεσμάτων του συστήματος. Σε αυτό το κεφάλαιο γίνεται μια γενική εισαγωγή στο ευρύτερο πεδίο στο οποίο κινείται η σχετιζόμενη με αυτήν τη διατριβή έρευνα.

### 1.1 Εικόνες

Αρχικά, παρουσιάζουμε συνοπτικά και ασχολούμαστε με κάποια εισαγωγικά θέματα που αφορούν τις εικόνες, το σχηματισμό τους, τη μαθηματική περιγραφή τους, τη χρωματική πληροφορία που συνδέεται με αυτές και τον τρόπο αναπαράστασής της, τον τρόπο αποθήκευσής τους σε ψηφιακά μέσα κτλ. Για μια πιο αναλυτική και ολοκληρωμένη παρουσίαση πάνω σε όλα αυτά τα σημαντικά θέματα παραπέμπουμε στις εργασίες [34] και [38].

#### 1.1.1 Σχηματισμός Εικόνων

Οι εικόνες προκύπτουν κυρίως από τη συλλογή σε κάποιο αισθητήρα φωτεινής ροής ενέργειας που εκπέμπεται από τα αντικείμενα του χώρου. Έτσι, για παράδειγμα, συλλέγοντας το φως που εκπέμπεται και/ή ανακλάται από τα διάφορα αντικείμενα μέσω του ανθρώπινου συστήματος όρασης, σχηματίζονται στον ανθρώπινο εγκέφαλο εικόνες, που με κατάλληλη επεξεργασία στο εσωτερικό του εγκέφαλου επιτρέπουν στον άνθρωπο να αντιλαμβάνεται οπτικά το χώρο γύρω του. Το ίδιο συμβαίνει βεβαίως και σε πολλά άλλα ζώα που διαθέτουν τα ανάλογα συστήματα βιολογικών αισθητήρων.

Εκτός όμως από τους βιολογικούς οπτικούς αισθητήρες, για τους οποίους η φύση έχει προνοήσει να υπάρχουν ως εφόδιο ζωής για τους διάφορους έμβιους οργανισμούς, ο άνθρωπος έχει δημιουργήσει και τεχνητούς αισθητήρες (κάμερες, τηλεσκόπια κτλ.), που του επιτρέπουν να καταγράφει εικόνες, ώστε, αποθηκευοντάς τες μετέπειτα σε κάποιο μέσο αποθήκευσης, να έχει τη δυνατότητα να τις ανακαλέσει σε κάποιο μελλοντικό χρόνο. Μάλιστα, δεδομένου ότι το ανθρώπινο σύστημα όρασης επιτρέπει την οπτική αντίληψη

μόνο σε ένα στενό φάσμα ηλεκτρομαγνητικής ακτινοβολίας, και συγκεκριμένα μεταξύ 400nm και 700nm, οι τεχνητοί αισθητήρες δίνουν την επιπλέον δυνατότητα να καταγράφονται εικόνες εκτός αυτού του φάσματος και, μετέπειτα, με κάποιο μετασχηματισμό να μετατρέπονται σε ορατές εικόνες που μπορούν να μελετηθούν οπτικά από τον άνθρωπο. Το γεγονός αυτό δίνει μεγάλη ευελιξία και βρίσκει πολλές εφαρμογές όσο αφορά την ιατρική, τις δορυφορικές απεικονίσεις, τη μελέτη του διαστήματος κτλ.

Σε επίπεδο μαθηματικής αναπαράστασης, οι εικόνες περιγράφονται από μαθηματικές συναρτήσεις οι οποίες αναπαριστούν τη φασματική κατανομή της ενέργειας στα διάφορα σημεία της εικόνας και για τα διάφορα μήκη κύματος του ηλεκτρομαγνητικού φάσματος. Οι συνεχείς συναρτήσεις παίζουν σημαντικό ρόλο όσο αφορά τη θεωρητική ανάλυση και εξαγωγή συμπερασμάτων από τις εικόνες. Αντίθετα, οι διακριτές μαθηματικές συναρτήσεις είναι αυτές που χρησιμοποιούνται στην πράξη για την αναπαράσταση των εικόνων, ειδικά στο πεδίο που μας ενδιαφέρει, αυτό των ηλεκτρονικών υπολογιστών, που ως γνωστόν δέχονται και επεξεργάζονται ψηφιακή πληροφορία.

Ο ανθρώπινος οφθαλμός, που είναι το μέσο με το οποίο εμείς δεχόμαστε την οπτική πληροφορία, έχει τέσσερα είδη φωτοαισθητήρων: τα ραβδία και τρία είδη κωνίων. Τα ραβδία λειτουργούν σε αμυδρό φως και είναι υπεύθυνα για τη σκοτοπική όραση, που στερείται χρώματος. Τα κωνία είναι υπεύθυνα για την έγχρωμη όραση και το καθένα έχει μια πολυχρωματική φασματική απόκριση. Οι φασματικές αποκρίσεις αυτών των τριών ειδών κωνίων είναι επικαλυπτόμενες μεταξύ τους αλλά, παρόλα αυτά, διαφορετικές. Η αλληλεπίδραση του φωτός με τα κωνία, τα οποία είναι διάσπαρτα πάνω στον αμφιβληστροειδή χιτώνα (ρέτινα), οδηγεί στην ανάλυση του προσπίπτοντος φωτός σε τρεις συνιστώσες, με κάθε μια να αντιστοιχεί και σε ένα από τα είδη των κωνίων. Αυτός είναι και ο λόγος της έγχρωμης όρασης, η διαφορετική διέγερση, δηλαδή, που ασκεί το φάσμα του φωτός σε καθένα από τους αισθητήρες αυτούς.

Επομένως, γίνεται αντιληπτό ότι παρότι το φάσμα του φωτός αναλύεται σε άπειρα μήκη κύματος, για να έχουμε την ανθρώπινη αντίληψη της έγχρωμης όρασης αρκεί η ύπαρξη τριών αισθητήρων με διαφορετική φασματική απόκριση ο καθένας. Αυτός είναι και ο λόγος που οι έγχρωμες εικόνες αναλύονται σε τρεις συνιστώσες. Δηλαδή, για την αποτύπωση ενός χρώματος σε κάποιο σημείο μιας εικόνας χρησιμοποιούμε πάντα τρεις συντεταγμένες. Το τι ακριβώς αντιπροσωπεύουν αυτές οι τρεις συντεταγμένες ορίζεται ανάλογα με το χρησιμοποιούμενο χώρο χρώματος (color space). Με τους χώρους χρώματος ασχολούμαστε εκτενέστερα στην ενότητα που ακολουθεί.

### 1.1.2 Χώροι Χρώματος

Με τον όρο «χρώμα» εννοούμε μια συγκεκριμένη φασματική κατανομή ηλεκτρομαγνητικής ενέργειας. Κατανομές σε μήκη κύματος μεταξύ 400nm και 700nm αντιστοιχούν στα ορατά από τον άνθρωπο χρώματα. Συγκεκριμένα, τα 400nm αντιστοιχούν στο ιώδες και τα 700nm στο ερυθρό. Μήκη κύματος μικρότερα από τα 400nm αντιστοιχούν στην υπεριώδη και μήκη κύματος μεγαλύτερα των 700nm στην υπέρυθη ακτινοβολία.

Εάν η φασματική κατανομή αποτελείται από μια και μόνο συνάρτηση Dirac σε κάποιο μήκος κύματος, τότε έχουμε την περίπτωση μονοχρωματικής ακτινοβολίας, το χρώμα της οποίας χαρακτηρίζεται ως καθαρό. Αντίθετα, αν το φάσμα περιέχει και άλλες συνιστώσες, τότε το αντίστοιχο χρώμα θεωρείται σύνθετο.

Όπως αναφέρθηκε προηγουμένως, στην πράξη κάθε χρώμα αναπαριστάται ως διάνυσμα σε έναν τρισδιάστατο χώρο. Πολλοί χώροι χρώματος έχουν προταθεί και μελετηθεί για τις ιδιότητές τους. Δύο είναι οι κύριες κατηγορίες στις οποίες κατατάσσονται οι διάφοροι χώροι χρώματος. Η πρώτη κατηγορία περιλαμβάνει τους γραμμικούς χώρους χρώματος και

η δεύτερη τους μη γραμμικούς. Στους γραμμικούς χώρους χρώματος η υπέρθεση – μείξη διαφορετικών χρωμάτων μπορεί να γίνει με γραμμικό τρόπο, σε αντίθεση με τους μη γραμμικούς. Όμως, παρότι οι γραμμικοί χώροι χρώματος είναι από μαθηματικής άποψης λιγότερο πολύπλοκοι και περισσότερο εύχρηστοι από τους μη γραμμικούς, έχουν ένα σημαντικό μειονέκτημα, ιδίως όσο αφορά τις εφαρμογές που θα μας απασχολήσουν σε αυτήν την διατριβή. Συγκεκριμένα, για αυτούς τους χώρους χρώματος, η απόσταση δύο χρωμάτων, όπως αυτή καθορίζεται από τις απεικονίσεις τους στο χώρο χρώματος, δεν εκφράζει την αίσθηση απόστασης μεταξύ τους που αντιλαμβάνεται ο μέσος άνθρωπος. Δηλαδή, δύο χρώματα που ο άνθρωπος θεωρεί κοντινά μπορεί να έχουν τρισδιάστατες απεικονίσεις δυσανάλογα διαφορετικές, και αντίστροφα. Έτσι, σε εφαρμογές ανάκτησης εικόνων προτιμώνται γενικά οι μη γραμμικοί χώροι χρώματος. Εξάλλου, έχει γίνει εκτεταμένη έρευνα για τον ορισμό τέτοιων χώρων που θα ανταποκρίνονται όσο το δυνατόν καλύτερα στην ανθρώπινη αντίληψη περί χρωματικής απόστασης.

### 1.1.2.1 Γραμμικοί Χώροι Χρώματος

Σε αυτό το σημείο θα αναφερθούμε συνοπτικά σε κάποιους από τους πιο διαδεδομένους γραμμικούς χώρους χρώματος. Ας αρχίσουμε από τον πιο γνωστό χώρο χρώματος, που δεν είναι άλλος από το χώρο RGB. Η βάση του συγκεκριμένου χώρου είναι τα μονοχρωματικά χρώματα R-κόκκινο (645.16nm), G-πράσινο (526.32nm) και B-μπλε (444.44nm). Έτσι, σε αυτόν το χώρο κάθε χρώμα αναλύεται σύμφωνα με την περιεκτικότητά του όσο αφορά τις τρεις προαναφερθείσες συνιστώσες. Για διάφορους λόγους, ο συγκεκριμένος χώρος κρίνεται ως ο λιγότερο κατάλληλος για εφαρμογές που στηρίζονται στην αναπαράσταση του χρώματος με τρόπο διαισθητικό για το μέσο άνθρωπο. Εκτός των άλλων, αξίζει να σημειωθεί ότι έχει αποδειχτεί πως για να ταιριάζει οπτικά κάποιο πραγματικό χρώμα με μια υπέρθεση κόκκινου, πράσινου και μπλε ενδέχεται να απαιτηθεί η χρήση αρνητικών συντελεστών στο μείγμα. Αυτό συμβαίνει γιατί τα πρωταρχικά χρώματα του χώρου RGB αντιστοιχούν σε πραγματικά χρώματα. Για να αποφευχθεί ένα τέτοιο ενδεχόμενο είναι απαραίτητο ένας χώρος να στηρίζεται σε μια βάση μη πραγματικών χρωμάτων.

Προς αποφυγή του παραπάνω φαινομένου δημιουργήθηκε ο χρωματικός χώρος XYZ, με γραμμικό μετασχηματισμό του χώρου RGB. Τα πρωταρχικά χρώματα του χώρου αυτού είναι μη πραγματικά, δηλαδή το φάσμα τους εμφανίζει αρνητικές τιμές για κάποια μήκη κύματος. Παρόλα αυτά, έτσι εξασφαλίζεται το γεγονός ότι θα μπορεί πάντα ο μέσος άνθρωπος να αντιστοιχίζει κάποιο πραγματικό χρώμα που παρατηρεί, με ένα γραμμικό συνδυασμό στη βάση του χώρου XYZ, στον οποίο οι συντελεστές θα είναι πάντα θετικοί. Όμως, και αυτός ο χώρος, παρά την ευρεία χρήση και μελέτη του, δείχνει να έχει ξεπεραστεί, μιας και δεν μπορεί να αποφύγει τα συνολικά μειονεκτήματα των γραμμικών χώρων.

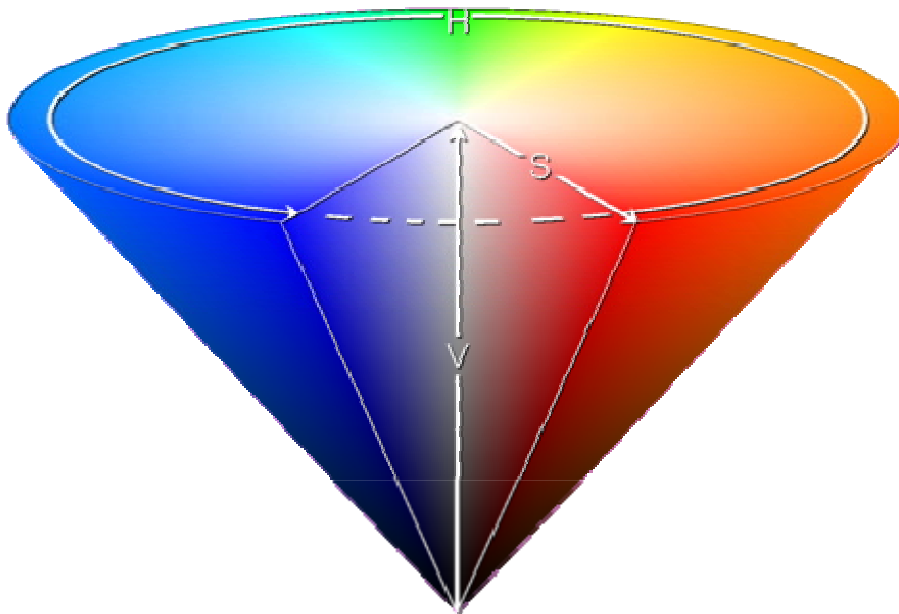
Τέλος, ένας συμπληρωματικός χώρος του RGB είναι ο CMY. Η βάση του περιλαμβάνει το κυανό (C), τη magenta (M) και το κίτρινο (Y). Το κυανό είναι το συμπληρωματικό του κόκκινου, με την έννοια ότι αντιστοιχεί στο χρώμα που θα ανακλαστεί από μια επιφάνεια βαμμένη με μια χρωστική τέτοια ώστε να απορροφά το κόκκινο, όταν αυτή φωτιστεί με λευκό φως. Αντίστοιχα, η magenta είναι το συμπληρωματικό του πράσινου και το κίτρινο το συμπληρωματικό του μπλε.

### 1.1.2.2 Μη Γραμμικοί Χώροι Χρώματος

Το κυριότερο μειονέκτημα σχετικά με τους γραμμικούς χώρους χρώματος έχει να κάνει με την ασυμφωνία που εμφανίζουν όσο αφορά την ανθρώπινη αντίληψη για το χρώμα. Ο

άνθρωπος όσο αφορά το χρώμα αντιλαμβάνεται κυρίως τρεις παραμέτρους: την απόχρωση (hue), τον κορεσμό (saturation) και την φωτεινότητα (brightness). Η απόχρωση σχετίζεται με το κυρίαρχο μήκος κύματος σε μια φασματική κατανομή, και η αίσθηση για αυτή μεταβάλλεται καθώς μετακινούμαστε από κάποιο κύριο χρώμα προς ένα άλλο. Ο κορεσμός, αντίθετα, έχει να κάνει με το βαθμό καθαρότητας του χρώματος. Το μονοχρωματικό φως θεωρείται πλήρως καθαρό, σε αντίθεση με το λευκό φως, που θεωρείται ότι έχει μηδενική καθαρότητα. Τέλος, η φωτεινότητα αυξάνει καθώς πηγαίνουμε από το μαύρο στο λευκό, και σχετίζεται με την ένταση της φωτεινής ακτινοβολίας.

Με βάση αυτές τις διαπιστώσεις έχουν οριστεί οι μη γραμμικοί χώροι HSI (Hue – Saturation – Intensity), HSV (Hue – Saturation – Value) κτλ. Για παράδειγμα, ο χώρος HSV, που χρησιμοποιείται σχετικά συχνά σε εφαρμογές ανάκτησης εικόνων, ορίζεται από έναν κώνο. Ο κατακόρυφος άξονας αντιστοιχεί στη συνιστώσα V (φωτεινότητα), η γωνία γύρω από τον κατακόρυφο άξονα στη συνιστώσα H (απόχρωση), και η κάθετος από το τρέχον σημείο προς τον κατακόρυφο άξονα στη συνιστώσα S (κορεσμός). Ο χώρος HSV παρουσιάζεται εποπτικά στο σχήμα 1.1 [84].



**Σχήμα 1.1:** Χώρος χρώματος HSV

Ο μετασχηματισμός από το χώρο RGB στο χώρο HSV δίνεται από τους παρακάτω τύπους:

$$H = \begin{cases} \left( \frac{G - B}{MAX - MIN} \right) / 6, & \text{αν } R = MAX \\ \left( 2 + \frac{B - R}{MAX - MIN} \right) / 6, & \text{αν } G = MAX \\ \left( 4 + \frac{R - G}{MAX - MIN} \right) / 6, & \text{αν } B = MAX \end{cases} \quad (1.1)$$

$$S = \frac{MAX - MIN}{MAX} \quad (1.2)$$



$$V = MAX \quad (1.3)$$

θεωρώντας ότι το  $MAX$  και το  $MIN$  δηλώνουν τη μέγιστη και ελάχιστη τιμή μεταξύ των  $R$ ,  $G$ ,  $B$ , αντίστοιχα, και ότι οι τελευταίες είναι οι κανονικοποιημένες στο διάστημα  $[0, 1]$  χρωματικές συνιστώσες στο χώρο RGB. Με αυτές τις προϋποθέσεις, οι τιμές που προκύπτουν για τα  $H, S, V$  βρίσκονται και αυτές στο διάστημα  $[0, 1]$ .

Τέλος, πρέπει να αναφέρουμε τους ομοιόμορφους χρωματικούς χώρους, που ως ειδική υποκατηγορία των μη γραμμικών χώρων χρώματος επιδιώκουν να αντιμετωπίσουν το βασικό πρόβλημα της αναντιστοιχίας μεταξύ απόστασης στο χώρο χρώματος και ανθρώπινης αντίληψης χρωματικής διαφορετικότητας. Σε αυτήν την κατηγορία ξεχωρίζουν οι χώροι LUV, UVW, και ο χώρος LAB. Ο τελευταίος είναι σήμερα ο πιο συχνά χρησιμοποιούμενος χώρος αναπαράστασης χρώματος. Η συνιστώσα  $L$  αντιστοιχεί στη φωτεινότητα, η συνιστώσα  $A$  στη διαφορά μεταξύ κόκκινης και πράσινης συνιστώσας, και η συνιστώσα  $B$  στη διαφορά μεταξύ κίτρινης και μπλε συνιστώσας. Ο μετασχηματισμός μεταξύ του χώρου RGB και του χώρου LAB περιγράφεται από τους παρακάτω τύπους:

$$L = 116f(Y/Y_n) - 16 \quad (1.4)$$

$$A = 500[f(X/X_n) - f(Y/Y_n)] \quad (1.5)$$

$$B = 200[f(Y/Y_n) - f(Z/Z_n)] \quad (1.6)$$

όπου

$$f(t) = \begin{cases} t^{1/3}, & \text{αν } t > (6/29)^3 \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 t + \frac{4}{29}, & \text{αλλιώς} \end{cases} \quad (1.7)$$

Με  $X, Y, Z$  συμβολίζονται οι αντίστοιχες συντεταγμένες στο χώρο XYZ, ενώ τα  $X_n, Y_n$  και  $Z_n$  αντιστοιχούν σε τυποποιημένες από την CIE (International Commission on Illumination – Διεθνής Επιτροπή Φωτισμού) XYZ-tristimulus τιμές για το λευκό χρώμα αναφοράς. Όπως αναφέρθηκε και παραπάνω, οι συντεταγμένες  $X, Y, Z$  προκύπτουν από τις  $R, G, B$  μέσω γραμμικού μετασχηματισμού:

$$(X, Y, Z)^T = \mathbf{M}(R, G, B)^T \quad (1.8)$$

όπου ο  $\mathbf{M}$  είναι ο  $3 \times 3$  πίνακας μετασχηματισμού, ο οποίος εξαρτάται από την επιλογή για τις προαναφερθείσες tristimulus τιμές του λευκού χρώματος αναφοράς. Θεωρώντας ότι οι συντεταγμένες  $X, Y, Z$  είναι κανονικοποιημένες στο διάστημα  $[0, 1]$ , οι τιμές των συντεταγμένων  $L, A, B$  κυμαίνονται στο διάστημα  $[0, 100]$ ,  $[-0.86, 0.98]$  και  $[-1.07, 0.94]$ , αντίστοιχα.

### 1.1.3 Ψηφιακές Εικόνες

Στην παρούσα διατριβή θα μας απασχολήσουν οι ψηφιακές εικόνες. Εξάλλου, η ίδια η εντυπωσιακή ανάπτυξη που παρατηρείται τα τελευταία χρόνια στο χώρο της εικόνας είναι σε πολύ μεγάλο βαθμό αποτέλεσμα της ευρείας διάδοσης των ηλεκτρονικών υπολογιστών και των νέων τεχνολογιών γύρω από αυτούς, καθώς και των ψηφιακών καμερών. Οι

ψηφιακές κάμερες επιτρέπουν την πανεύκολη και σχεδόν ανέξοδη δημιουργία φωτογραφιών, οι οποίες, μετέπειτα, μέσω των υπολογιστών και του διαδικτύου, γίνονται κτήμα όχι μόνο του δημιουργού τους αλλά και όλων όσων έχουν πρόσβαση σε αυτές, από το ένα άκρο της γης ως το άλλο.

Τα ηλεκτρονικά μέσα, όμως, που προαναφέρθηκαν είναι ψηφιακά, και ως εκ τούτου χειρίζονται ψηφιακές εικόνες. Μια ψηφιακή εικόνα έχει δύο επίπεδα διακριτοποίησης. Πρώτα από όλα, έχουμε διακριτοποίηση στο επίπεδο της εικόνας. Σχεδόν πάντα μιλάμε για δισδιάστατες εικόνες, οπότε η διακριτοποίηση αυτή συνεπάγεται τον ορισμό στο επίπεδο της εικόνας ενός τετραγωνικού, συνήθως, πλέγματος. Με αυτόν τον τρόπο, αντί να αποθηκεύονται πληροφορίες για κάθε ένα από τα άπειρα σημεία στο επίπεδο της εικόνας (πράγμα αδύνατον), αποθηκεύονται πληροφορίες μόνο για τα πεπερασμένα παράθυρα του πλέγματος. Κάθε τέτοιο παράθυρο του πλέγματος ονομάζεται pixel (εικονοστοιχείο). Τα pixels αντιπροσωπεύουν τη μέγιστη δυνατή ευκρίνεια που έχουμε στη διάθεσή μας, αφού, κατά την απεικόνιση της εικόνας σε κάποιο ηλεκτρονικό μέσο, κάθε pixel απεικονίζεται με ομοιόμορφη πληροφορία σε όλο του το εμβαδόν. Επομένως, αν δύο εικόνες απαθανατίζουν την ίδια σκηνή, αυτή που αποθηκεύεται σε περισσότερα pixels έχει μεγαλύτερη ανάλυση, και, συνεπώς, περιλαμβάνει περισσότερες λεπτομέρειες της σκηνής.

Το δεύτερο επίπεδο διακριτοποίησης έχει να κάνει με την πληροφορία που αποθηκεύεται αυτή καθαυτή. Για ασπρόμαυρες, γκριζες (gray) εικόνες, για τις οποίες δεν αποθηκεύεται πληροφορία χρώματος, μόνο η φωτεινότητα διατηρείται ώστε να χαρακτηρίζει το κάθε pixel. Επομένως, σε κάθε pixel αντιστοιχίζεται ένας αριθμός που εκφράζει τη μέση φωτεινότητα του. Αυτός ο αριθμός για να αποθηκευτεί διακριτοποιείται σε ακρίβεια ενός πλήθους από bits (δυαδικά ψηφία). Ευνόητο είναι ότι όσο περισσότερα είναι τα bits που διατίθενται, τόσο περισσότερες είναι οι διαφορετικές στάθμες κβαντοποίησης του αριθμού, και άρα τόσο περισσότερα τα διαφορετικά επίπεδα φωτεινότητας που μπορούν να αναπαρασταθούν. Συνήθως χρησιμοποιούνται 8 bits, και επομένως  $2^8 = 256$  διαφορετικές στάθμες. Για τις έγχρωμες εικόνες, το μόνο που αλλάζει είναι ότι έχουμε πλέον 3 αριθμούς ανά pixel, όσοι χρειάζονται για να περιγραφεί η χρωματική του πληροφορία. Κάθε μια από τις 3 συνιστώσες χρώματος κβαντοποιείται ξεχωριστά. Επομένως, για 8 bits ανά συνιστώσα, καταλήγουμε τελικά σε 24 bits ανά pixel.

## 1.2 Ανάκτηση Εικόνων

Στα προβλήματα ανάκτησης εικόνων, τα οποία καλείται να επιλύσει κάποιο σύστημα ανάκτησης εικόνων, ο τελικός στόχος είναι ο εντοπισμός από τη μεριά του συστήματος και η επιστροφή στο χρήστη εικόνων που τον ενδιαφέρουν, από ένα μεγάλο σύνολο αναφοράς, που περιέχει εν γένει πάρα πολλές εικόνες, δηλαδή από μια βάση εικόνων. Ένα χαρακτηριστικό αυτής της αναζήτησης είναι ότι το ποσοστό των εικόνων που ενδιαφέρουν το χρήστη, σε σχέση με το συνολικό αριθμό των εικόνων της βάσης, είναι συνήθως πολύ μικρό. Κάτι άλλο που πρέπει να επισημανθεί είναι ότι ο χρήστης, γενικά, δε στοχεύει στην εξαντλητική εύρεση των εικόνων που ανήκουν στην κατηγορία που τον ενδιαφέρει, αλλά περισσότερο επιδιώκει την εύρεση κάποιων αντιπροσωπευτικών εικόνων με βάση τις προτιμήσεις του.

### 1.2.1 Κατηγορίες Ανάκτησης Εικόνων

Οι μέθοδοι ανάκτησης εικόνων, ανάλογα με την πληροφορία που χρησιμοποιούν για την περιγραφή της κάθε εικόνας, χωρίζονται σε δύο κύριες κατηγορίες:

1. Σημασιολογική ανάκτηση εικόνων.
2. Ανάκτηση εικόνων με βάση το περιεχόμενο.

Παρακάτω περιγράφουμε τα κύρια χαρακτηριστικά καθεμιάς από αυτές τις κατηγορίες, εστιάζοντας όμως περισσότερο στη δεύτερη.

### 1.2.1.1 Σημασιολογική Ανάκτηση Εικόνων

Οι προτιμήσεις του χρήστη μπορούν να γίνουν γνωστές στο σύστημα με διάφορους τρόπους. Για παράδειγμα, μέσω κάποιας λέξης ή φράσης κλειδί η οποία θα περιγράφει κάποιο αντικείμενο, τόπο, γεγονός, κατάσταση κτλ. για το οποίο ενδιαφέρεται ο χρήστης. Συνήθως, αυτός ο τρόπος ερώτησης εμπλέκει μόνο έμμεσα τις εικόνες, αφού η αναζήτηση για τις εικόνες που ικανοποιούν τις απαιτήσεις του χρήστη γίνεται με βάση σημασιολογικά δεδομένα που έχουν επισυναφθεί με έμμεσο ή άμεσο τρόπο στις εικόνες της βάσης.

Έχουμε, σε αυτήν την περίπτωση, σημασιολογική ανάκτηση εικόνων (semantic image retrieval). Τα προβλήματα που συνδέονται με αυτόν τον τρόπο ανάκτησης εικόνων έχουν να κάνουν, από τη μια, με την αδυναμία ικανοποιητικής απόδοσης του πλήρους περιεχομένου μιας εικόνας μέσω ενός περιορισμένου αριθμού λέξεων ή φράσεων και, από την άλλη, με το γεγονός ότι, συνήθως, απαιτείται μια μη αυτοματοποιημένη προεργασία για κάθε εικόνα, και συγκεκριμένα η απασχόληση κάποιου ανθρώπου για την περιγραφή του σημασιολογικού της περιεχομένου. Κάτι τέτοιο όμως δεν είναι καθόλου πρακτικό, ειδικά στη σημερινή εποχή, όπου ο τεράστιος όγκος των διαθέσιμων εικόνων και ο κατακλυστικός ρυθμός παραγωγής και διάθεσης καινούργιων καθιστά την ανθρώπινη επεξεργασία τους αδύνατη.

Στην παρούσα διατριβή δε θα ασχοληθούμε από εδώ και στο εξής με τη σημασιολογική ανάκτηση εικόνων. Αντίθετα, θα μας απασχολήσει ιδιαίτερα ο τρόπος ανάκτησης εικόνων που περιγράφεται στην επόμενη ενότητα.

### 1.2.1.2 Ανάκτηση Εικόνων με Βάση το Περιεχόμενο

Μια άλλη μορφή με την οποία μπορεί ο χρήστης να υποβάλλει τις ερωτήσεις του είναι μέσω παραδείγματος. Σε αυτήν την περίπτωση, ο χρήστης επιλέγει μία ή περισσότερες εικόνες, από ένα αντιπροσωπευτικό σύνολο των εικόνων της βάσης για παράδειγμα, που θεωρεί ότι βρίσκονται κοντά σε αυτό που εκείνος ψάχνει, και τις παρουσιάζει στο σύστημα ως παραδείγματα των προτιμήσεών του. Το σύστημα, έπειτα, παράγει μια διάταξη (ranking) των εικόνων της βάσης, συγκρίνοντας κάθε μία από αυτές με τα παραδείγματα του χρήστη μέσω κάποιου εσωτερικού κριτηρίου ομοιότητας, και παρουσιάζει τις πρώτες (αν η διάταξη είναι σε φθίνουσα σειρά όσο αφορά την ομοιότητα με τα παραδείγματα) από αυτές στο χρήστη ως αποτελέσματα της αναζήτησης.

Αυτός ο τρόπος υποβολής ερωτημάτων από το χρήστη ταιριάζει περισσότερο στην περίπτωση κατά την οποία επιχειρείται ανάκτηση εικόνων με βάση το περιεχόμενο (content-based image retrieval), που είναι και το πρόβλημα που θα μας απασχολήσει σε αυτήν τη διατριβή. Κατά την ανάκτηση εικόνων με βάση το περιεχόμενο στηρίζομαστε σε αυτόματα υπολογιζόμενα οπτικά χαρακτηριστικά για την περιγραφή της κάθε εικόνας. Αυτό μας «λύνει τα χέρια», με την έννοια ότι αποφεύγουμε τη δαπανηρή από κάθε άποψη χρησιμοποίηση ανθρώπινου δυναμικού για την περιγραφή του περιεχομένου των εικόνων. Από την άλλη, όμως, έχει αποδειχθεί ότι τα οπτικά χαρακτηριστικά που μπορούμε να υπολογίσουμε από τις εικόνες είναι χαμηλού επιπέδου, σε αντίθεση με την ανθρώπινη

αντίληψη όσο αφορά τις εικόνες, η οποία είναι σε υψηλό, σημασιολογικό επίπεδο. Με απλά λόγια, δεν είναι δυνατόν να περιγραφεί ικανοποιητικά το σημασιολογικό περιεχόμενο μιας εικόνας, για το οποίο ενδιαφέρεται πρωτίστως ο χρήστης, με χρήση αποκλειστικά χαμηλού επιπέδου, οπτικών χαρακτηριστικών. Το πρόβλημα αυτό είναι γνωστό στην επιστημονική κοινότητα ως πρόβλημα σημασιολογικού κενού (semantic gap problem), όπως πολύ καλά αναλύεται στην εργασία [13]. Επιπλέον, για μια συνοπτική όσο και πλήρη επισκόπηση αυτού του πεδίου μπορεί κανείς να ανατρέξει και στις εργασίες [1] και [39].

Από τα παραπάνω, όσο αφορά την ανάκτηση εικόνων με βάση το περιεχόμενο έχουν διαφανεί τα εξής: Πρώτον, χρειάζεται ένας τρόπος αναπαράστασης των εικόνων με χρήση οπτικών χαρακτηριστικών και, δεύτερον, ο ορισμός ενός κριτηρίου διάταξης, που θα αντικατοπτρίζει την κρίση του συστήματος όσο αφορά την ομοιότητα των εικόνων της βάσης με τις εικόνες που δίνει ο χρήστης ως παραδείγματα των προτιμήσεών του. Στις επόμενες ενότητες πραγματευόμαστε τα δύο αυτά σημαντικά θέματα.

## 1.2.2 Αναπαράσταση Εικόνων

Γίνεται εύκολα κατανοητό ότι το μέτρο απόστασης μεταξύ των εικόνων που θα χρησιμοποιηθεί από το σύστημα εξαρτάται άμεσα από τον τρόπο αναπαράστασης των εικόνων. Από την άλλη, οι εικόνες πρέπει οπωσδήποτε να αναπαρασταθούν με κάποιο συμπαγή τρόπο, γιατί ως έχουν είναι δύσκολο να τις χειριστούμε. Για παράδειγμα, αν μιλάμε για ψηφιακές έγχρωμες εικόνες, αυτές αποτελούνται από ένα συνήθως μεγάλο αριθμό pixels, καθένα από τα οποία χαρακτηρίζεται από ένα τρισδιάστατο διάνυσμα που περιλαμβάνει την πληροφορία του χρώματος σε κάποιο χρωματικό χώρο. Αυτός ο όγκος της πληροφορίας, πέραν του ότι είναι ιδιαίτερα μεγάλος για να χρησιμοποιηθεί απευθείας για συγκρίσεις μεταξύ εικόνων, είναι επίσης σε πολύ χαμηλό επίπεδο, με αποτέλεσμα η χρήση του να καθίσταται αναποτελεσματική για τα προβλήματα που μας απασχολούν. Με άλλα λόγια, μέσω μιας προσεγγισμένης συμπαγούς αναπαράστασης των εικόνων, παρότι ο απόλυτος όγκος των δεδομένων που χρησιμοποιούνται για κάθε εικόνα μειώνεται δραματικά, συνήθως αυτό οδηγεί σε κέρδος όσο αφορά την επίδοση των συστημάτων.

Για την αναπαράσταση, με κάποιο τρόπο, των εικόνων απαιτείται πάντα η εξαγωγή οπτικής πληροφορίας με τη μορφή συγκεκριμένων οπτικών χαρακτηριστικών (features). Μέσω του υπολογισμού χαρακτηριστικών από τις εικόνες μας δίνεται η δυνατότητα να μετασχηματίσουμε την πληροφορία, όπως αυτή κρύβεται στον πολύ χαμηλού επιπέδου και αραιό από άποψη πληροφορίας αρχικό όγκο δεδομένων, σε πρότυπα πιο συμπαγή και περιεκτικά πληροφορίας, αλλά και πιο κοντά στην αντίληψη του χρήστη όσο αφορά το περιεχόμενο των εικόνων. Αυτή η διαδικασία υπολογισμού χαρακτηριστικών είναι, λοιπόν, ένα βήμα που κάνουμε για την κάλυψη, σε κάποιο βαθμό, του σημασιολογικού κενού, όπως αυτό ορίστηκε προηγουμένως. Συνήθως, υπολογίζονται χαρακτηριστικά για το χρώμα, την υφή (texture), το σχήμα (shape), την κατανομή όλων αυτών στο επίπεδο της εικόνας κτλ., τα οποία είναι περισσότερο κατάλληλα ως βάση πάνω στην οποία θα οριστεί η όποια απόσταση εικόνων. Πάνω σε αυτή τη λογική έχει αφιερωθεί σημαντική επιστημονική προσπάθεια, ώστε να υπολογισθούν κατάλληλα χαρακτηριστικά για την περιγραφή των εικόνων. Διαφοριστικές εργασίες σχετικά με όλη αυτή την ερευνητική δραστηριότητα είναι, για παράδειγμα, οι [10], [40], [42], [62], [64], [72] και [74].

Όσο αφορά τον τρόπο αναπαράστασης των εικόνων, διακρίνουμε δύο βασικές επιλογές:

1. Διανυσματική αναπαράσταση.
2. Αναπαράσταση με χρήση συνεχούς μοντέλου.

Η πρώτη επιλογή συνίσταται στην περιγραφή κάθε εικόνας μέσω ενός ή σχετικά λίγων διανυσμάτων. Αυτός ο τρόπος περιγραφής ταιριάζει περισσότερο στην περίπτωση κατά την οποία τα τελικά χρησιμοποιούμενα οπτικά χαρακτηριστικά υπολογίζονται καθολικά (globally), δηλαδή από όλη την εικόνα, ή από ένα σύνολο σχετικά λίγων και μεγάλων σε έκταση περιοχών της εικόνας. Με τον ένα ή τον άλλο τρόπο καταλήγουμε εν τέλει στην αναπαράσταση της όποιιας εικόνας από έναν περιορισμένο αριθμό πολυδιάστατων διανυσμάτων χαρακτηριστικών. Η πληροφορία που κωδικοποιείται μέσω αυτών των διανυσμάτων μπορεί να έγκειται στον υπολογισμό κάποιων ροπών (moments) ή ιστογραμμάτων που θα κωδικοποιούν την κατανομή της φωτεινότητας ή τη χρωματική κατανομή της εικόνας, στην περίπτωση που μιλάμε για χρώμα. Ακόμα, μπορεί να βασίζεται σε κάποιο καθολικό μετασχηματισμό, όπως με χρήση wavelets [62], και στον υπολογισμό κάποιων ροπών επί των αποτελεσμάτων για τη μοντελοποίηση της υφής. Τέλος, για τη μοντελοποίηση του σχήματος συχνά χρησιμοποιούνται ιστογράμματα ακμών [64]. Αντιπροσωπευτικές εργασίες που χρησιμοποιούν τέτοιου είδους αναπαράσταση για τις εικόνες είναι, για παράδειγμα, οι [21], [25], [43], [44], [59] και [71].

Η δεύτερη επιλογή που έχουμε όσο αφορά την αναπαράσταση των εικόνων, περιλαμβάνει εργασία σε δύο στάδια. Στο πρώτο στάδιο δουλεύουμε τοπικά. Για κάθε pixel, ομάδα pixels ή μικρή περιοχή της εικόνας εξάγουμε κάποια τοπική περιγραφή που εν γένει ενσωματώνει και αντιπροσωπεύει διαφόρων ειδών πληροφορίες. Για παράδειγμα, τρέχουσα θέση στο επίπεδο της εικόνας, αντιπροσωπευτικό χρώμα, περιγραφή για την υφή κτλ. Έτσι, μετά από την ολοκλήρωση του σταδίου αυτού καταλήγουμε να έχουμε για κάθε εικόνα ένα σημαντικό αριθμό από διανύσματα χαρακτηριστικών, φέροντα την αντίστοιχη τοπική περιγραφή. Τα διανύσματα αυτά έχουν συνήθως αρκετά μικρότερη διάσταση από τα διανύσματα που αντιστοιχούν στον πρώτο τρόπο αναπαράστασης, και κάθε εικόνα χαρακτηρίζεται από ένα σύνολο (bag) τέτοιων ομοειδών διανυσμάτων. Σε μια παρόμοια περιγραφή για τις εικόνες καταλήγουμε και μέσω των πλέον δημοφιλών και καταξιωμένων μεθόδων υπολογισμού χαρακτηριστικών, με αντιπροσωπευτικό παράδειγμα αυτήν που παρουσιάζεται στην εργασία [40], οι οποίες χρησιμοποιούν τεχνικές εντοπισμού ενός συνόλου ιδιαίτερων σημείων στο επίπεδο της εικόνας, και για καθένα από αυτά υπολογίζουν χαρακτηριστικά που κωδικοποιούν την τοπική πληροφορία στην περιοχή της εικόνας γύρω του. Σε αυτήν την περίπτωση, όμως, ο αριθμός των διανυσμάτων χαρακτηριστικών που αντιστοιχούν σε κάθε εικόνα δεν είναι πάρα πολύ μεγάλος, ενώ και η διάσταση του κάθε διανύσματος χαρακτηριστικών είναι αρκετά υψηλή.

Σε δεύτερο στάδιο, τώρα, αυτά τα διανύσματα μπορούν να μοντελοποιηθούν έτσι ώστε να προκύψει τελικά ένα συνεχές μοντέλο που θα περιγράφει όλη την εικόνα. Ένας ενδιαφέρον τρόπος τέτοιας μοντελοποίησης, με τον οποίον θα ασχοληθούμε εκτενώς σε αυτήν τη διατριβή, είναι η χρήση Μοντέλων Γκαουσιανών Μειγμάτων (Gaussian Mixture Models). Τα μοντέλα αυτά έχουν ενδιαφέρουσες μαθηματικές ιδιότητες και πολλά πλεονεκτήματα που απορρέουν από αυτές, και μπορούν να χρησιμοποιηθούν για τη μοντελοποίηση της κατανομής των προαναφερθέντων τοπικών διανυσμάτων στον πολυδιάστατο χώρο στον οποίον αυτά ορίζονται. Ως ενδεικτικές εργασίες πάνω σε αυτόν τον τρόπο περιγραφής και μοντελοποίησης των εικόνων μπορούν να θεωρηθούν οι [5], [17], [18], [45], [49] και [70].

### 1.2.3 Διάταξη Εικόνων

Τα κριτήρια διάταξης εικόνων εξαρτώνται από την αναπαράσταση των εικόνων αλλά και από τη μέθοδο ανάκτησης εικόνων που χρησιμοποιεί το σύστημα. Για το τελευταίο πρέπει να επισημάνουμε ότι πέρα από τη λογική της σύγκρισης των εικόνων της βάσης, μία

προς μία, με τις εικόνες – παραδείγματα που παρέχει ο χρήστης, υπάρχει και η εναλλακτική της εκπαίδευσης κάποιου ταξινομητή (classifier). Όμως, για την εκπαίδευση αυτή απαιτούνται συνήθως εκτός από «θετικά» παραδείγματα, δηλαδή παραδείγματα που περιγράφουν το τι θέλει ο χρήστης, και «αρνητικά» παραδείγματα, δηλαδή παραδείγματα για το τι δε θέλει από το σύστημα να του επιστρέψει. Μετά από την υποβολή των παραδειγμάτων ορίζεται κάποιο σύνορο διαχωρισμού (boundary), και η διάταξη των εικόνων της βάσης επιτυγχάνεται λαμβάνοντας υπόψη τη θέση τους ως προς αυτό. Ως εκ τούτου, αυτή η περίπτωση ταιριάζει περισσότερο στην τεχνική της συσχετιστικής ανατροφοδότησης (relevance feedback) για τη βελτίωση των αποτελεσμάτων του συστήματος ανάκτησης εικόνων, όπου ο χρήστης δίνει από γύρο σε γύρο της ανατροφοδότησης αρκετά θετικά και αρνητικά παραδείγματα. Σε αυτήν την τεχνική θα αναφερθούμε, όμως, αναλυτικότερα σε επόμενη ενότητα.

Αντίθετα, στα κλασικά συστήματα ανάκτησης εικόνων, ή ακόμα και σε αυτά που χρησιμοποιούν συσχετιστική ανατροφοδότηση, κατά την εκκίνηση της διαδικασίας αλληλεπίδρασης με το χρήστη, αυτό που παρέχει ο χρήστης είναι ένα ή το πολύ λίγα θετικά παραδείγματα. Σε αυτήν την περίπτωση, ο τρόπος διάταξης είναι απλός και έγκειται στην ταξινόμηση των εικόνων της βάσης μέσω σύγκρισής τους ως προς την εγγύτητα ή απόστασή τους από τα παραδείγματα του χρήστη.

Όσο αφορά, λοιπόν, αυτό το απλό σενάριο, χρειάζεται ένα μέτρο απόστασης μεταξύ δύο εικόνων. Αυτό, προφανώς, θα εξαρτάται από τον τρόπο αναπαράστασης των εικόνων. Για παράδειγμα, η διανυσματική αναπαράσταση των εικόνων μέσω ενός αριθμού πολυδιάστατων διανυσμάτων χαρακτηριστικών οδηγεί συνήθως στην υιοθέτηση μέτρων που βασίζονται πάνω στις νόρμες  $L_p$ , και ορίζονται στη μορφή

$$d(\mathbf{x}, \mathbf{y}) = L_p(\mathbf{x} - \mathbf{y}) = \left( \sum_{i=1}^k |x_i - y_i|^p \right)^{1/p} \quad (1.9)$$

θεωρώντας ότι τα  $\mathbf{x}$ ,  $\mathbf{y}$  είναι τα διανύσματα προς σύγκριση και ότι καθένα διάνυσμα έχει  $k$  διαστάσεις.

Αντίθετα, στην περίπτωση που κάθε εικόνα αναπαριστάνεται μέσω ενός μοντέλου εκπαιδευμένου πάνω στα υπολογισμένα από αυτήν τοπικά χαρακτηριστικά, όπως τα Γκαουσιανά Μείγματα που αναφέρθηκαν προηγουμένως, απαιτείται κάποιο εξειδικευμένο μέτρο απόστασης, κατάλληλα ορισμένο και προσαρμοσμένο για την αποδοτική σύγκριση μεταξύ των χρησιμοποιούμενων μοντέλων. Γενικά, δεν είναι απαραίτητο αυτό το μέτρο απόστασης να έχει χαρακτηριστικά μετρικής σε κάποιο συγκεκριμένο χώρο, δηλαδή να ικανοποιεί όλα τα ακόλουθα κριτήρια:

$$d(x, y) \geq 0 \quad (1.10)$$

$$d(x, y) = 0 \Leftrightarrow x = y \quad (1.11)$$

$$d(x, y) = d(y, x) \quad (1.12)$$

$$d(x, z) \leq d(x, y) + d(y, z) \quad (1.13)$$

(η τελευταία από τις παραπάνω σχέσεις είναι γνωστή και ως τριγωνική ανισότητα) που πρέπει να ισχύουν για να θεωρηθεί μετρική. Πολλά από τα πιο δημοφιλή κριτήρια διάταξης δεν ικανοποιούν όλες τις παραπάνω απαιτήσεις. Το ζητούμενο στην πράξη είναι το χρησιμοποιούμενο μέτρο να εμφανίζει καλά χαρακτηριστικά όσο αφορά το διαχωρισμό διαφορετικών μοντέλων και, παράλληλα, να μπορεί να υπολογιστεί γρήγορα και, ως εκ τούτου, αποδοτικά.

## 1.3 Συσχετιστική Ανατροφοδότηση

Όπως αναφέρθηκε προηγουμένως, το σημασιολογικό κενό ανάμεσα στα χαμηλού επιπέδου οπτικά χαρακτηριστικά που υπολογίζονται από τις εικόνες και στην ανθρώπινη αντίληψη περί ομοιότητας μεταξύ των εικόνων, που κινείται σε καθαρά σημασιολογικό επίπεδο, κάνει την ανάκτηση εικόνων με βάση το περιεχόμενο να φαντάζει ως ένας πολύ δύσκολος στόχος. Δεν προκαλεί έκπληξη, λοιπόν, το γεγονός ότι τα αποτελέσματα που δίνουν αυτού του είδους τα συστήματα ανάκτησης εικόνων κρίνονται συνήθως από τους χρήστες από καθόλου έως ελάχιστα ικανοποιητικά.

Σε πρώτο στάδιο, αυτή η διαπίστωση οδηγεί στην προσπάθεια βελτίωσης των περιγραφών που χρησιμοποιούνται για τις εικόνες. Όπως έχει διαφανεί ήδη, όμως, αυτή η διαδικασία βελτίωσης των χρησιμοποιούμενων περιγραφών για τις εικόνες έχει ένα όριο ως προς τα αποτελέσματά της, αφού από τη φύση τους τα οπτικά χαρακτηριστικά, όσο προσεγγμένα και αν έχουν εκλεγεί και υπολογισθεί, παρουσιάζουν μια εγγενή αδυναμία να περιγράψουν το σημασιολογικό περιεχόμενο για το οποίο ενδιαφέρεται ο χρήστης.

Οδηγούμαστε λοιπόν νομοτελειακά στην απόφαση για την ενσωμάτωση του χρήστη ως δομικό κομμάτι της λειτουργίας του συστήματος. Εξάλλου, η σημασία μιας τέτοιας απόφασης φαίνεται και από το αν αναλογιστούμε ότι δύο διαφορετικοί χρήστες είναι σχεδόν βέβαιο ότι θα έχουν διαφορετική αντίληψη και άλλα κριτήρια όσο αφορά την ομοιότητα εικόνων. Αλλά ακόμα και ο ίδιος χρήστης μπορεί να μεταβάλλει τα κριτήρια αξιολόγησής του ανάλογα με το τι ψάχνει. Αυτή, επομένως, η ισχυρή εξάρτηση από την κρίση του χρήστη πρέπει κάπως να ενσωματωθεί σε ένα σύστημα, αν αυτό θέλει να παρέχει ικανοποιητικά αποτελέσματα σε αυτόν που στην περίπτωσή μας είναι ο τελικός κριτής των αποτελεσμάτων. Και αυτός δεν είναι άλλος από τον εκάστοτε χρήστη του συστήματος.

### 1.3.1 Περιγραφή της Διαδικασίας

Η συσχετιστική ανατροφοδότηση, λοιπόν, είναι μια τεχνική επιβλεπόμενης μάθησης (supervised learning) που σκοπό έχει να βελτιώσει, με τη βοήθεια του χρήστη, τα αποτελέσματα της ανάκτησης των πληροφοριών (information retrieval) που ζητάει ο τελευταίος. Η διαδικασία που ακολουθείται είναι απλή και βασίζεται στη διαδραστική (interactive) αλληλεπίδραση μεταξύ χρήστη και συστήματος. Τα βήματα της προαναφερθείσης διαδικασίας είναι τα κάτωθι:

1. Ο χρήστης διατυπώνει μια ερώτηση στο σύστημα μέσω παραδειγμάτων αντιπροσωπευτικών των προτιμήσεών του.
2. Το σύστημα χρησιμοποιεί την τρέχουσα ερώτηση ή/και το τρέχον κριτήριο διάταξης για να επιστρέψει ένα σύνολο αποτελεσμάτων στο χρήστη.
3. Εάν ο χρήστης είναι ικανοποιημένος από τα αποτελέσματα, η διαδικασία σταματά, αλλιώς συνεχίζουμε με το βήμα 4.
4. Ο χρήστης αξιολογεί τα αποτελέσματα του συστήματος.
5. Το σύστημα λαμβάνει υπόψιν την αξιολόγηση για να τροποποιήσει την τρέχουσα ερώτηση ή/και το κριτήριο διάταξης που χρησιμοποιεί.
6. Μεταβαίνουμε στο βήμα 2.

Η αρχική χρήση της συσχετιστικής ανατροφοδότησης εστιαζόταν στην αντιμετώπιση των προβλημάτων ανάκτησης κειμένου. Στις συγκεκριμένες εφαρμογές και στην κλασική μορφή της η συσχετιστική ανατροφοδότηση βασίζεται στη θεώρηση του κάθε κειμένου ως

πολυδιάστατου διανύσματος, και η βασική της λειτουργία ορίζεται από τον τύπο του *Rocchio* [46]:

$$\mathbf{Q}' = \alpha\mathbf{Q} + \beta\left(\frac{1}{N_p}\sum_{i=1}^{N_p}\mathbf{V}_i^p\right) - \gamma\left(\frac{1}{N_n}\sum_{i=1}^{N_n}\mathbf{V}_i^n\right) \quad (1.14)$$

όπου  $\mathbf{Q}$  είναι το διάνυσμα που περιγράφει την τρέχουσα ερώτηση και  $\mathbf{Q}'$  το νέο διάνυσμα της ερώτησης μετά την ενσωμάτωση των παραδειγμάτων που έδωσε ο χρήστης σε αυτόν το γύρο (ή εποχή) της ανατροφοδότησης. Επίσης,  $\mathbf{V}_i^p$  είναι τα  $N_p$  το πλήθος διανύσματα που αντιστοιχούν στα θετικά παραδείγματα του χρήστη και  $\mathbf{V}_i^n$  τα  $N_n$  το πλήθος διανύσματα που αντιστοιχούν στα αρνητικά παραδείγματα. Τέλος, με  $\alpha$ ,  $\beta$ ,  $\gamma$  συμβολίζουμε κατάλληλες σταθερές, οι οποίες αντιπροσωπεύουν τη βαρύτητα της τρέχουσας ερώτησης, του διανυσματικού μέσου των θετικών παραδειγμάτων και του διανυσματικού μέσου των αρνητικών παραδειγμάτων, αντίστοιχα, στη διαμόρφωση της νέας ερώτησης. Μέσω του τύπου του *Rocchio*, επομένως, καθορίζεται ο τρόπος με τον οποίον το σύστημα αντιμετωπίζει το πέμπτο βήμα της διαδικασίας της συσχετιστικής ανατροφοδότησης. Στην προκειμένη περίπτωση, μια απλή επισκόπηση του εν λόγω τύπου αρκεί για να αντιληφθούμε ότι αυτό που γίνεται είναι μια τροποποίηση της τρέχουσας ερώτησης, ώστε αυτή να προσεγγίσει τα θετικά παραδείγματα του χρήστη και να απομακρυνθεί από τα αρνητικά.

Λόγω ακριβώς των ιδιαίτερων δυσκολιών του προβλήματος της ανάκτησης εικόνων με βάση το περιεχόμενο, η συσχετιστική ανατροφοδότηση προτάθηκε σχετικά νωρίς ως μια πολλά υποσχόμενη τεχνική για την ενσωμάτωση της ανθρώπινης αντίληψης στα συστήματα που καλούνται να λύσουν το συγκεκριμένο πρόβλημα, και γνώρισε γρήγορα μεγάλη ερευνητική δραστηριότητα.

### 1.3.2 Υποβολή Παραδειγμάτων

Αναλύοντας περαιτέρω το τέταρτο από τα παραπάνω βήματα, πρέπει να τονίσουμε ότι η αξιολόγηση των αποτελεσμάτων γίνεται συνήθως στη βάση δυαδικής λογικής. Δηλαδή, ο χρήστης εξετάζει τις εικόνες (στην περίπτωσή μας) που του επιστρέφει το σύστημα προς αξιολόγηση, και για καθεμιά ή για κάποιες από αυτές προσδιορίζει αν είναι σχετική/σχετικές ή όχι με την αρχική του ερώτηση, όπως αυτή διατυπώθηκε από τα αρχικά παραδείγματα, ή εν πάση περιπτώσει με αυτό που ο ίδιος επιθυμεί να του επιστραφεί από το σύστημα. Αυτές οι εικόνες μαζί με την κρίση του χρήστη που τις συνοδεύει αποτελούν τα θετικά και τα αρνητικά παραδείγματα που ανατροφοδοτούνται στο σύστημα.

Βέβαια, πέρα από αυτήν την περίπτωση, κατά την οποία υπονοείται η ύπαρξη δύο κλασικών και αμοιβαία αποκλειόμενων συνόλων για τις σχετικές και τις μη σχετικές εικόνες, θα μπορούσε να υιοθετηθεί και μια πιο ασαφής λογική, κατά την οποία ο χρήστης θα αποδίδει ένα βαθμό συνάφειας σε κάθε εικόνα προς αξιολόγηση. Για παράδειγμα, θα μπορεί να αξιολογήσει ότι μια εικόνα είναι πολύ σχετική, εν μέρει σχετική, ελάχιστα σχετική ή εντελώς άσχετη κτλ. με την ερώτησή του, και δε θα έχει μόνο δύο επιλογές, δηλαδή να τη θεωρήσει ή εντελώς σχετική ή εντελώς άσχετη.

Ακόμα, πρέπει να τονιστεί ότι ο αριθμός των παραδειγμάτων που δίνει ο χρήστης δεν είναι συνήθως πολύ μεγάλος. Γενικά, γίνεται η ρεαλιστική παραδοχή ότι ο χρήστης είναι ανυπόμονος και δεν έχει τη διάθεση να ξοδέψει πολύ χρόνο ώστε να παράσχει μεγάλο αριθμό παραδειγμάτων στο σύστημα, ή να συμμετάσχει σε πολλούς γύρους ανατροφοδότησης.



### 1.3.3 Αξιοποίηση της Πληροφορίας της Ανατροφοδότησης

Όσο αφορά, τώρα, τον τρόπο με τον οποίον το σύστημα ενσωματώνει την πληροφορία από την αξιολόγηση του χρήστη σε κάθε εποχή της ανατροφοδότησης, ώστε να ανανεώσει τα αποτελέσματα που παρουσιάζει στον τελευταίο, πρέπει να διακρίνουμε δύο βασικές κατηγορίες μεθόδων:

1. Μέθοδοι που χρησιμοποιούν ταξινομητές.
2. Μέθοδοι που βασίζονται στην τροποποίηση της αναπαράστασης της ερώτησης.

Στις δύο επόμενες ενότητες αναλύονται και σχολιάζονται τα βασικά χαρακτηριστικά αυτών των κατηγοριών, και επισημαίνονται τα πλεονεκτήματα και τα μειονεκτήματα που σχετίζονται με αυτές.

#### 1.3.3.1 Χρήση Ταξινομητών

Η πρώτη κατηγορία μεθόδων συσχετιστικής ανατροφοδότησης έχει να κάνει με την εκπαίδευση ενός ταξινομητή, μέσω του οποίου θα γίνει προσπάθεια να προσδιοριστεί το σύνορο μεταξύ των σχετικών και των μη σχετικών με την ερώτηση του χρήστη εικόνων. Το σύνορο αυτό ορίζει δύο περιοχές στο χώρο αναπαράστασης των εικόνων, μια για τις σχετικές και μια για τις μη σχετικές εικόνες. Έπειτα, από την απόσταση, για παράδειγμα, μιας εικόνας της βάσης από το εν λόγω σύνορο, και από το αν αυτή βρίσκεται ή όχι στην περιοχή που θεωρείται από τον ταξινομητή ως περιοχή των σχετικών εικόνων, μπορεί να επιτευχθεί η κατάταξη των εικόνων της βάσης σε μια λίστα που θα αντικατοπτρίζει την κρίση του συστήματος όσο αφορά την συνάφειά τους με την ερώτηση του χρήστη. Εννοείται ότι σε κάθε γύρο της ανατροφοδότησης ο χρήστης προσφέρει νέα παραδείγματα, τα οποία μαζί με τα παραδείγματα των προηγούμενων γύρων χρησιμοποιούνται ώστε να εκπαιδεύσουν ένα νέο ταξινομητή, πράγμα που οδηγεί στην τροποποίηση του συνόρου και, κατ' επέκταση, της διάταξης των εικόνων της βάσης, από την οποία οι θεωρούμενες ως πιο σχετικές θα επιστραφούν στο χρήστη, προκειμένου ένας νέος γύρος ανατροφοδότησης να ακολουθήσει, ενδεχομένως.

Αυτό που πρέπει να σημειωθεί είναι ότι, πέρα από το γεγονός ότι από γύρο σε γύρο της ανατροφοδότησης το σύνολο εκπαίδευσης του ταξινομητή αυξάνει συνεχώς από τα προστιθέμενα νέα παραδείγματα, πράγμα που μπορεί να επιβαρύνει σταδιακά την εκπαίδευση, το σύνολο των θετικών παραδειγμάτων δεν είναι ομοειδές με αυτό των αρνητικών παραδειγμάτων. Με αυτό θέλουμε να πούμε ότι, από τη μια, για μια συγκεκριμένη ερώτηση οι εικόνες της βάσης που δεν ενδιαφέρουν το χρήστη είναι κατά πολύ περισσότερες από αυτές που τον ενδιαφέρουν. Από την άλλη, οι σχετικές εικόνες είναι ως κατηγορία συμπαγής, με την έννοια ότι έχουν κάτι που συνδέει τη μια με την άλλη. Είναι σχετικές με ένα συγκεκριμένο τρόπο, ανήκουν δηλαδή σε μια συγκεκριμένη κατηγορία, έστω σημασιολογική. Οι μη σχετικές εικόνες, αντίθετα, ανήκουν εκ των πραγμάτων σε πολλές σημασιολογικές κατηγορίες, και διαφέρουν η καθεμιά με το δικό της τρόπο από τις σχετικές. Αυτό δυσκολεύει στην πράξη την εκπαίδευση των ταξινομητών, οι οποίοι βασίζονται συνήθως στην υπόθεση ότι οι δύο κατηγορίες που πρέπει να διαχωριστούν είναι ομοειδείς.

Ένα άλλο γεγονός που δυσκολεύει αυτού του είδους την εκπαίδευση είναι ο σχετικά μικρός αριθμός των παραδειγμάτων που παρέχει ο χρήστης, δυσανάλογα μικρός αν συνυπολογίσουμε και τη συνήθως μεγάλη διάσταση των διανυσμάτων χαρακτηριστικών που έχουν υπολογισθεί για τις εικόνες. Εξάλλου, οι κλασικοί ταξινομητές προϋποθέτουν,

ακριβώς, μια διανυσματική αναπαράσταση για κάθε πρότυπο προς ταξινόμηση, αναπαράσταση που, όπως έχει αναφερθεί προηγουμένως, στην περίπτωση των εικόνων οδηγεί σε μεγάλης διάστασης διανύσματα. Στο κεφάλαιο 7 αυτής της διατριβής, προς αντιμετώπιση του παραπάνω προβλήματος θα μελετηθεί η δυνατότητα χρήσης τεχνικών επιλογής χαρακτηριστικών σε κάθε γύρο της ανατροφοδότησης, προκειμένου να μειωθεί η διάσταση των αναπαραστάσεων.

Παρόλα αυτά, για να χρησιμοποιηθεί ένας ταξινομητής σε προβλήματα ανάκτησης εικόνων μέσω συσχετιστικής ανατροφοδότησης δεν είναι αναγκαίο οι εικόνες να αναπαριστούνται ως διανύσματα. Συγκεκριμένα, στο κεφάλαιο 6 προτείνουμε μια μέθοδο συσχετιστικής ανατροφοδότησης που χρησιμοποιεί έναν από τους πιο δημοφιλείς και καθιερωμένους ταξινομητές, τα SVMs (Support Vector Machines – Μηχανές Διανυσμάτων Υποστήριξης), για το διαχωρισμό των θετικών από τα αρνητικά παραδείγματα του χρήστη, θεωρώντας παράλληλα ότι κάθε εικόνα αναπαριστάνεται μέσω ενός Γκαουσιανού Μείγματος.

Αντιπροσωπευτικές εργασίες όσο αφορά τη χρήση ταξινομητών για την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης στο πλαίσιο της ανάκτησης εικόνων μπορούν να θεωρηθούν, για παράδειγμα, οι [7], [9], [19], [20], [47] και [48].

### 1.3.3.2 Τροποποίηση της Αναπαράστασης της Ερώτησης

Μια δεύτερη επιλογή σχετικά με τη διαδικασία της συσχετιστικής ανατροφοδότησης, εκτός από την εκπαίδευση ενός ταξινομητή που περιγράψαμε προηγουμένως, είναι αυτή της τροποποίησης της ερώτησης ή/και του μέτρου απόστασης που χρησιμοποιεί το σύστημα για τη σύγκριση μεταξύ εικόνων, ώστε να συμπεριληφθεί και η πληροφορία που προκύπτει από την ανατροφοδότηση. Με αυτήν την επιλογή συνδέονται δύο υποθέσεις όσο αφορά την κατανομή των παραδειγμάτων του χρήστη στο χώρο αναπαράστασης των εικόνων. Συγκεκριμένα, έχουμε τις εξής εναλλακτικές:

1. Υπόθεση απλής ομάδας.
2. Υπόθεση πολλαπλών ομάδων.

Παρακάτω παρουσιάζουμε μια συνοπτική περιγραφή των βασικότερων αποτελεσμάτων στα οποία οδηγεί καθεμιά από αυτές τις υποθέσεις, καθώς και των κύριων χαρακτηριστικών των αντίστοιχων συστημάτων ανάκτησης εικόνων.

#### Υπόθεση Απλής Ομάδας

Η υπόθεση απλής ομάδας έγκειται στη θεώρηση από τη μεριά του συστήματος ότι τα παραδείγματα που δίνει ο χρήστης σχηματίζουν μια ομάδα (cluster) στο χώρο αναπαράστασης των εικόνων. Μια τέτοια φιλοσοφία εκφράζει, για παράδειγμα, ο τύπος του Rocchio, που παρουσιάστηκε παραπάνω, αφού με απλή επισκόπηση της σχέσης (1.14) βλέπουμε ότι τόσο τα θετικά όσο και τα αρνητικά παραδείγματα λαμβάνονται υπόψιν στον υπολογισμό της νέας αναπαράστασης για την ερώτηση μέσω των αντίστοιχων διανυσματικών μέσων.

Στην περίπτωση που οι εικόνες αναπαριστούνται ως διανύσματα σε έναν πολυδιάστατο χώρο, εκτός από τη δυνατότητα τροποποίησης της τρέχουσας ερώτησης του χρήστη με μια λογική παρόμοια με αυτή του τύπου του Rocchio, ώστε αυτή να προσεγγίσει διανυσματικά τα θετικά παραδείγματα και να απομακρυνθεί από τα αρνητικά, υπάρχει και η δυνατότητα τροποποίησης αυτής καθαυτής της αποστάσεως που χρησιμοποιείται. Σε αυτές τις

περιπτώσεις συνήθως επιλέγεται για να χρησιμοποιηθεί μια απόσταση όπως η απόσταση *Mahalanobis*, που ορίζεται ως

$$d(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{M}(\mathbf{x} - \mathbf{y}) \quad (1.15)$$

όπου τα  $\mathbf{x}$ ,  $\mathbf{y}$  αντιστοιχούν στα διανύσματα των εικόνων, ενώ  $\mathbf{M}$  είναι ένας πίνακας  $k \times k$ , όπου  $k$  η διάσταση των διανυσμάτων.

Αρχικά, ο πίνακας  $\mathbf{M}$  τίθεται ίσος με τον αντίστοιχο μοναδιαίο πίνακα, μια και δεν έχουμε κάποια εκτίμηση για τη σχετική βαρύτητα της κάθε διάστασης στην αντίληψη του χρήστη περί ομοιότητας. Έπειτα, όμως, με τη συσσώρευση των παραδειγμάτων που παρέχει ο χρήστης, μπορούμε να δώσουμε διαφορετική τιμή σε καθένα από τα στοιχεία του  $\mathbf{M}$ . Η λογική γύρω από αυτό είναι απλή. Αν οι συνιστώσες των θετικών παραδειγμάτων για μια συγκεκριμένη διάσταση εμφανίζουν μικρή διασπορά, δηλαδή αν είναι όλες συγκεντρωμένες πολύ κοντά στη μέση τιμή τους, αυτό αποτελεί ένδειξη ότι αυτή η διάσταση κρίνεται ως σημαντική από το χρήστη. Επομένως, το σύστημα τροποποιεί τον πίνακα  $\mathbf{M}$  έτσι ώστε η εν λόγω διάσταση να έχει μεγάλη βαρύτητα στον υπολογισμό της συνολικής απόστασης. Συνήθως, λόγω του μεγάλου μεγέθους του πίνακα αυτού, αφού τα αντίστοιχα διανύσματα των εικόνων περιλαμβάνουν εν γένει πάρα πολλά χαρακτηριστικά, ο πίνακας θεωρείται διαγώνιος, γιατί αν θεωρηθεί πλήρης, με δεδομένο το γεγονός ότι ο αριθμός των παραδειγμάτων που δίνει ο χρήστης είναι τις περισσότερες φορές αρκετά μικρότερος από τη διάσταση των διανυσμάτων, ο υπολογισμός του δεν μπορεί να γίνει με εύρωστο τρόπο.

Η προηγούμενη ανάλυση όσο αφορά την τροποποίηση της απόστασης με βάση τα θετικά παραδείγματα φωτίζει ακόμα περισσότερο την υπόθεση απλής ομάδας. Όμως, παρότι αναφέρθηκε παραπάνω πως, εν γένει, τα θετικά παραδείγματα ανήκουν σε μια σημασιολογική κατηγορία, αυτό δεν προϋποθέτει σε καμία περίπτωση ότι στο χώρο των χαμηλού επιπέδου οπτικών χαρακτηριστικών θα σχηματίζουν μια ομάδα. Πολύ περισσότερο όταν όλη η προηγούμενη προσέγγιση πρακτικά θεωρεί ότι αυτή η ομάδα μπορεί να περιγραφεί μέσω της αντίστοιχης διανυσματικής μέσης τιμής (όπως αποκαλύπτεται από τον τύπο του Rocchio) και ενός πίνακα άμεσα σχετιζόμενου με τη συνδιακύμανση (covariance), αφού στην ουσία ο πίνακας  $\mathbf{M}$ , όπως παρουσιάστηκε παραπάνω, είναι ανάλογος του αντίστροφου του πίνακα συνδιακύμανσης των θετικών παραδειγμάτων. Δηλαδή, ουσιαστικά, η υπόθεση απλής ομάδας ισοδυναμεί με την υπόθεση Γκαουσιανής κατανομής των θετικών παραδειγμάτων. Αν αυτές οι υποθέσεις είναι χονδροειδείς για τα θετικά παραδείγματα, για τα αρνητικά είναι εντελώς απαράδεκτες, γιατί αυτά, εκ των πραγμάτων, δεν είναι ούτε καν σημασιολογικά συμπαγή, όπως έχει αναφερθεί και προηγουμένως. Όλα αυτά, από τη μία δυσκολεύουν την αποτελεσματική ενσωμάτωση των αρνητικών παραδειγμάτων σε τέτοιου είδους μεθόδους και, από την άλλη, οδηγούν τις τελευταίες σε μη ικανοποιητικά αποτελέσματα, λόγω των μη ρεαλιστικών παραδοχών στις οποίες αυτές στηρίζονται αλλά και αριθμητικών προβλημάτων που αναδύονται και εξαναγκάζουν σε ακόμα μεγαλύτερες απλουστεύσεις των υποθέσεων, όπως για παράδειγμα η υιοθέτηση διαγώνιων πινάκων που αναφέρθηκε προηγουμένως.

Αξιόλογες προσπάθειες όσο αφορά αυτήν την κατηγορία μεθόδων αντιπροσωπεύουν οι εργασίες [2], [4], [12], [21], [22] και [23].

## Υπόθεση Πολλαπλών Ομάδων

Λόγω των προβλημάτων που παρουσιάζει η υπόθεση απλής ομάδας, όπως αυτά αναλύθηκαν παραπάνω, κατέστη σύντομα σαφές πως θα έπρεπε να μελετηθούν μοντέλα που, από τη μια, θα είναι μαθηματικά εύρωστα και, από την άλλη, αντί της εξαιρετικά

περιοριστικής υπόθεσης κατανομής των παραδειγμάτων σε μια ομάδα, θα δίνουν την ελευθερία τα τελευταία να μπορούν να κατανέμονται σε πολλές ομάδες στο χώρο των χαρακτηριστικών.

Τα Γκαουσιανά Μείγματα, που παρουσιάζονται εκτενώς σε επόμενο κεφάλαιο της παρούσας διατριβής, θεωρούνται από τα πιο δημοφιλή μοντέλα αυτού του είδους. Οι τρόποι με τους οποίους μπορούν να χρησιμοποιηθούν τα τελευταία στο πλαίσιο των μεθόδων που μελετώνται σε αυτήν την ενότητα είναι δύο, και έχουν άμεση σύνδεση με τον τρόπο αναπαράστασης των εικόνων. Αν, για παράδειγμα, κάθε εικόνα αναπαριστάται με ένα πολυδιάστατο διάνυσμα χαρακτηριστικών, τότε τα (θετικά ή αρνητικά) παραδείγματα που παρέχει στο σύστημα ο χρήστης αποτελούν ένα σύνολο διανυσμάτων σε έναν πολυδιάστατο χώρο. Σε αυτήν την περίπτωση, τα Γκαουσιανά Μείγματα μπορούν να χρησιμοποιηθούν για τη μοντελοποίηση της κατανομής των διανυσμάτων αυτών στον προαναφερθέντα χώρο. Έπειτα, η διάταξη των εικόνων της βάσης με κριτήριο τη συνάφειά τους, όπως αυτή κρίνεται από το σύστημα, με την ερώτηση του χρήστη μπορεί να γίνει με βάση την πιθανοφάνεια (likelihood) της κάθε εικόνας για το διαμορφωθέν μοντέλο. Τα μειονεκτήματα αυτής της προσέγγισης είναι ότι, από τη μία, συνήθως ο αριθμός των παραδειγμάτων είναι σχετικά μικρός και η διάσταση των διανυσμάτων σχετικά μεγάλη, πράγμα που καθιστά δύσκολη τη με εύρωστο τρόπο εκπαίδευση ενός Γκαουσιανού Μείγματος και, από την άλλη, για κάθε εποχή της συσχετιστικής ανατροφοδότησης απαιτείται επανεκπαίδευση του μοντέλου με χρήση και των προηγούμενων παραδειγμάτων αλλά και αυτών που έδωσε ο χρήστης στην τρέχουσα εποχή. Ως αντιπροσωπευτικές αυτού του τρόπου χρήσης των Γκαουσιανών Μειγμάτων, παραθέτουμε τις εργασίες [15], [24] και [25].

Αντίθετα, αν κάθε εικόνα αναπαριστάται μέσω ενός Γκαουσιανού Μείγματος που μοντελοποιεί την κατανομή των τοπικών διανυσμάτων χαρακτηριστικών που υπολογίζονται από αυτήν, και με δεδομένο ότι τα τοπικά διανύσματα χαρακτηριστικών είναι συνήθως πολυάριθμα και η διάσταση τους μικρότερη από αυτήν των διανυσμάτων που χρησιμοποιούνται στην περίπτωση της διανυσματικής αναπαράστασης των εικόνων, η εκπαίδευση του μοντέλου μπορεί να γίνει με πολύ πιο εύρωστο τρόπο. Ακόμα, δεν χρειάζεται επανεκπαίδευση των μοντέλων κατά τις διάφορες εποχές της ανατροφοδότησης, αφού τα μοντέλα αντιπροσωπεύουν τις ίδιες τις εικόνες, και αυτές μένουν αμετάβλητες από εποχή σε εποχή. Επομένως, τα μοντέλα μπορούν να υπολογιστούν εκ των προτέρων, χωρίς ο υπολογισμός αυτός να επιβαρύνει στο ελάχιστο την αλληλεπίδραση του εκάστοτε χρήστη με το σύστημα. Αφού, όμως, με αυτήν την οπτική θεωρούμε την κάθε εικόνα ως ένα Γκαουσιανό Μείγμα αντί για ένα διάνυσμα, πρέπει να βρεθεί ένας τρόπος ανανέωσης της ερώτησης που θα λαμβάνει υπόψιν αυτή τη νέα αναπαράσταση, και να χρησιμοποιηθεί ένα κατάλληλο μέτρο απόστασης για τη διάταξη των εικόνων της βάσης σε κάθε γύρο της συσχετιστικής ανατροφοδότησης, που να μπορεί να υπολογιστεί γρήγορα και, παράλληλα, να διαχωρίζει ικανοποιητικά τα διαφορετικά μοντέλα. Στις εργασίες [26] και [27], η εκ των υστέρων πιθανότητα (posterior probability) του μοντέλου κάθε εικόνας της βάσης δοθέντων των παραδειγμάτων ανατροφοδότησης χρησιμοποιείται ως κριτήριο διάταξης, αλλά αυτό οδηγεί σε μεγάλη υπολογιστική επιβάρυνση, αφού προϋποθέτει τον υπολογισμό της τιμής του μοντέλου για ένα μεγάλο αριθμό τοπικών διανυσμάτων χαρακτηριστικών.

Στο κεφάλαιο 4 της παρούσας διατριβής προτείνουμε μια μέθοδο συσχετιστικής ανατροφοδότησης που εντάσσεται σε αυτήν κατηγορία, και η οποία χρησιμοποιεί Γκαουσιανά Μείγματα για την αναπαράσταση των εικόνων και μια κατάλληλη τεχνική ανανέωσης της ερώτησης, ορισμένη σε πιθανοτικό πλαίσιο. Επίσης, κάνει χρήση των θετικών ιδιοτήτων της παραπάνω τεχνικής και αυτών ενός νέου μέτρου απόστασης μεταξύ των μοντέλων, προκειμένου να ανανεώσει εξαιρετικά γρήγορα τη διάταξη των εικόνων της βάσης μετά την υποβολή των παραδειγμάτων του χρήστη.

## 1.4 Αξιολόγηση Συστημάτων Ανάκτησης Εικόνων

Η τελική αξιολόγηση της αποτελεσματικότητας των συστημάτων ανάκτησης γίνεται από τον τελικό χρήστη, που μένει ή δεν μένει ικανοποιημένος από τα αποτελέσματα που αυτά του παρέχουν μετά από την υποβολή ερωτήσεων του. Σε ερευνητικό επίπεδο, όμως, και για λόγους συγκρισιμότητας μεταξύ των διαφόρων μεθόδων που προτείνονται αλλά και αντικειμενικότητας όσο αφορά την κρίση της απόδοσης των συστημάτων, δεν είναι δυνατόν να χρησιμοποιηθούν άνθρωποι ως χρήστες κατά την αξιολόγηση. Πέρα από όλα τα άλλα, για να είναι τα αποτελέσματα μιας αξιολόγησης στατιστικά σημαντικά χρειάζονται πολλές επαναλήψεις εκτέλεσης του συστήματος με διαφορετικές αρχικές ερωτήσεις και παραδείγματα κάθε φορά. Ακόμα, κατά τη μελέτη μιας μεθόδου απαιτείται η αξιολόγηση να επαναληφθεί πολύ περισσότερες της μιας φορές, κάθε φορά με διαφορετικές ρυθμίσεις και για διάφορες παραλλαγές της μεθόδου. Το πρόβλημα, λοιπόν, και από πρακτικής άποψης καθίσταται αδύνατο στο χειρισμό του με χρήση ανθρώπων ως αξιολογητών.

### 1.4.1 Προσομοιώσεις

Για την αντιμετώπιση όλων των παραπάνω ζητημάτων σχεδιάζονται προσομοιώσεις που προσπαθούν να προσεγγίσουν τη συμπεριφορά του μέσου ανθρώπου κατά την αλληλεπίδρασή του με το σύστημα ανάκτησης.

Για τέτοιου είδους προσομοιώσεις, πρώτα από όλα, απαιτείται η ύπαρξη μιας βάσης εικόνων από την οποία θα επιχειρεί το σύστημα να ανακτήσει εικόνες σχετικές με ερωτήσεις διατυπωμένες προς αυτό. Το ζήτημα που γεννάται αμέσως είναι το πως θα οριστεί αντικειμενικά η έννοια της συνάφειας μεταξύ δύο εικόνων. Αυτή η έννοια είπαμε ότι διαφέρει από άνθρωπο σε άνθρωπο, αλλά και στον ίδιο άνθρωπο ακόμα δε μένει αμετάβλητη. Αυτό όμως δημιουργεί προβλήματα κατά την σε ερευνητικό επίπεδο αξιολόγηση των συστημάτων, και άρα είμαστε αναγκασμένοι να προβούμε σε ένα συμβιβασμό. Ο συμβιβασμός έγκειται στη χρήση εκ των προτέρων κατηγοριοποιημένων βάσεων εικόνων κατά την αξιολόγηση των συστημάτων. Δηλαδή, οι βάσεις που χρησιμοποιούμε χωρίζονται εκ των προτέρων σε έναν αριθμό κατηγοριών, και κάθε εικόνα της βάσης θεωρείται συνήθως ότι ανήκει σε μια και μόνο από τις υπάρχουσες κατηγορίες. Οι κατηγορίες πρέπει να έχουν οριστεί όσο το δυνατόν αντικειμενικά και με σημασιολογικά κριτήρια, μια και τέτοια κριτήρια ακολουθεί η λογική των χρηστών ενός πραγματικού συστήματος. Ευτυχώς, υπάρχουν βάσεις που ικανοποιούν αυτά τα κριτήρια και έχουν προταθεί και καταξιωθεί όσο αφορά την αξιολόγηση τέτοιων συστημάτων. Μια πολύ γνωστή είναι η βάση εικόνων Corel. Ο προαναφερθείς συμβιβασμός είναι αρκετά μεγάλος, βέβαια, ειδικά αν ληφθεί υπόψη ότι η αντιστοίχιση κάθε εικόνας σε μια και μόνο κατηγορία είναι υπερβολικά χονδροειδής. Με τη λογική που περιγράψαμε παραπάνω, πάντως, πετυχαίνουμε τον εξής απλό ορισμό για τη συνάφεια μεταξύ εικόνων: Δύο εικόνες είναι σχετικές μεταξύ τους, αν και μόνο αν ανήκουν στην ίδια κατηγορία της βάσης.

Πλέον τα χέρια μας έχουν λυθεί και οι προσομοιώσεις είναι απλή υπόθεση. Θεωρώντας ότι χρησιμοποιείται και η τεχνική της συσχετιστικής ανατροφοδότησης, για παράδειγμα, επιλέγουμε διάφορες εικόνες της βάσης, που καλύπτουν όλες τις κατηγορίες, και χρησιμοποιούμε την κάθε μια ως αρχική ερώτηση για να εκκινήσουμε ένα κύκλο εποχών συσχετιστικής ανατροφοδότησης. Σε κάθε εποχή της τελευταίας επιλέγουμε αυτόματα ως θετικά παραδείγματα εικόνες της ίδιας κατηγορίας με την τρέχουσα αρχική ερώτηση και ως αρνητικά εικόνες από άλλες κατηγορίες, ενώ, παράλληλα, υπολογίζουμε και κρατάμε την τιμή ενός μέτρου που εκφράζει το πόσο «καλά» είναι τα αποτελέσματα που επέστρεψε το σύστημα στο συγκεκριμένο γύρο. Τελικά, μπορούμε να υπολογίσουμε ένα μέσο όρο αυτού

του μέτρου για όλες τις αρχικές ερωτήσεις και για κάθε εποχή συσχετιστικής ανατροφοδότησης, και να καταλήξουμε σε μια καμπύλη που εκφράζει το πως εξελίσσεται η μέση ποιότητα των αποτελεσμάτων από εποχή σε εποχή.

## 1.4.2 Μέτρα Αξιολόγησης

Συνήθως, ως μέτρο ποιότητας των αποτελεσμάτων χρησιμοποιείται η ακρίβεια (precision) σε μια προκαθορισμένη εμβέλεια (scope)  $T$ . Η ακρίβεια, γενικά, ορίζεται ως

$$\text{precision} = \frac{\text{number of relevant images retrieved}}{\text{total number of images retrieved}} \quad (1.16)$$

Θέτοντας τον παρονομαστή ίσο με  $T$  και υπολογίζοντας πόσες από τις  $T$  πιο σχετικές όπως κρίνονται από το σύστημα εικόνες είναι πράγματι σχετικές σύμφωνα με την κατηγοριοποίηση της βάσης, μπορούμε με μια απλή διαίρεση να αξιολογήσουμε την ποιότητα των  $T$  πρώτων αποτελεσμάτων του συστήματος για μια συγκεκριμένη ερώτηση και εποχή της ανατροφοδότησης. Ως  $T$ , στις περισσότερες των περιπτώσεων επιλέγεται κάποιος σχετικά μικρός αριθμός, λόγω του ότι ο μέσος χρήστης δεν ενδιαφέρεται, όπως είπαμε, παρά για κάποιες αντιπροσωπευτικές εικόνες που να ικανοποιούν τις απαιτήσεις που θέτει, και όχι για μια εξαντλητική ανάκτηση όλων των σχετικών με την ερώτησή του εικόνων της βάσης. Επίσης, καθώς το  $T$  αυξάνεται, η ακρίβεια εμφανίζει την τάση να μειώνεται, λόγω της εισόδου στο σύνολο ανάκτησης όλο και περισσότερων μη σχετικών εικόνων.

Ένα άλλο μέτρο, συμπληρωματικό της ακρίβειας, που ενδέχεται να μας φανεί χρήσιμο είναι η ανάκληση (recall). Αυτή ορίζεται ως

$$\text{recall} = \frac{\text{number of relevant images retrieved}}{\text{total number of relevant images}} \quad (1.17)$$

Εκείνο που διαφοροποιεί αυτό το μέτρο από το προηγούμενο είναι ο παρονομαστής. Στην περίπτωση της ανάκλησης, ο παρονομαστής είναι σταθερός και ίσος με το συνολικό αριθμό εικόνων της υπό εξέταση κατηγορίας (της κατηγορίας της βάσης στην οποία ανήκει η αρχική ερώτηση) που υπάρχουν στη βάση. Είναι φανερό ότι όσο περισσότερες εικόνες ανακτώνται, δηλαδή καθώς μεγαλώνει το  $T$ , η ανάκληση εμφανίζει την τάση να αυξάνεται, αφού ο αριθμητής στη σχέση (1.17) είτε αυξάνεται είτε, στη χειρότερη περίπτωση, μένει σταθερός.

Από την προηγηθείσα ανάλυση παρατηρούμε ότι η ακρίβεια και η ανάκληση εμφανίζουν αντίθετες τάσεις. Δηλαδή, συχνά, η αύξηση της ακρίβειας συνοδεύεται από μείωση της ανάκλησης και, αντίστροφα, η αύξηση της ανάκλησης συνοδεύεται από μείωση της ακρίβειας. Όσο αφορά το λόγο των δύο αυτών μεγεθών, ισχύει η επόμενη ενδιαφέρουσα σχέση:

$$\frac{\text{precision}}{\text{recall}} = \frac{\text{total number of relevant images}}{\text{total number of images retrieved}} \quad (1.18)$$

Όπως αναφέραμε και προηγουμένως, ο αριθμητής του δεξιού μέλους της παραπάνω σχέσης είναι σταθερός. Επομένως, ο λόγος της ακρίβειας προς την ανάκληση εξαρτάται μόνο από τον αριθμό των ανακτώμενων εικόνων, και μάλιστα είναι αντιστρόφως ανάλογος αυτού του

αριθμού. Όσο αφορά την περίπτωση μας, τώρα, ο λόγος της ακρίβειας προς την ανάκληση είναι αντιστρόφως ανάλογος της εμβέλειας  $T$ .

Για μια πιο πλήρη παρουσίαση όσο αφορά την αξιολόγηση των συστημάτων ανάκτησης εικόνων και τα χρησιμοποιούμενα για αυτήν μέτρα, παραπέμπουμε στην εργασία [32].

## 1.5 Διάρθρωση της Διατριβής

Μετά από αυτό το εισαγωγικό κεφάλαιο, στο κεφάλαιο 2 της παρούσας διατριβής γίνεται μια εισαγωγή στο θέμα της χρήσης πιθανοτικών μοντέλων προκειμένου να περιγραφεί ένα σύνολο πολυδιάστατων, εν γένει, παρατηρήσεων. Ιδιαίτερη έμφαση δίνεται στα μοντέλα εκείνα που αποκαλούνται Γκαουσιανά Μείγματα, τα οποία θα μας απασχολήσουν επανειλημμένα στη συνέχεια. Ακολουθώντας, στο κεφάλαιο 3, το θέμα της σύγκρισης και των μέτρων απόστασης μεταξύ Γκαουσιανών Μειγμάτων αναπτύσσεται σε κάποιο βάθος. Ένας αριθμός από δημοφιλή αλλά και νέα μέτρα απόστασης παρουσιάζονται και αναλύονται οι ιδιότητές τους, τα πλεονεκτήματα και τα μειονεκτήματά τους. Στο κεφάλαιο 4 παρουσιάζεται μια νέα μέθοδος συσχετιστικής ανατροφοδότησης που αναπτύχθηκε. Η μέθοδος αυτή χρησιμοποιεί Γκαουσιανά Μείγματα για την αναπαράσταση τόσο των εικόνων της βάσης όσο και των ερωτήσεων του χρήστη, και βασίζεται σε μια συνεπή με το όλο πιθανοτικό πλαίσιο τεχνική για την ανανέωση του μοντέλου της ερώτησης σε κάθε γύρο της ανατροφοδότησης. Στη συνέχεια, στο κεφάλαιο 5 γίνεται μια συνοπτική εισαγωγή όσο αφορά τη χρήση μοντέλων μάθησης για την αντιμετώπιση προβλημάτων ταξινόμησης. Ιδιαίτερη έμφαση δίνεται στο πρόβλημα της δυαδικής ταξινόμησης και στη χρήση ταξινομητών SVM για την αντιμετώπισή του. Επίσης, το πρόβλημα της συσχετιστικής ανατροφοδότησης επανατοποθετείται στο ευρύτερο πλαίσιο των προβλημάτων ταξινόμησης, ώστε να φωτιστεί η δυνατότητα επίλυσής του μέσω της χρήσης ταξινομητών σε κάθε γύρο της ανατροφοδότησης. Στο κεφάλαιο 6, μια ακόμα νέα μέθοδος συσχετιστικής ανατροφοδότησης που βασίζεται στην αναπαράσταση των εικόνων μέσω Γκαουσιανών Μειγμάτων παρουσιάζεται. Αυτή τη φορά, η διαδικασία της συσχετιστικής ανατροφοδότησης υλοποιείται μέσω εκπαίδευσης κατάλληλα ορισμένων ταξινομητών, που μπορούν να χειριστούν τα μοντέλα των εικόνων. Στο κεφάλαιο 7 ακολουθεί μια συνοπτική μελέτη όσο αφορά τη χρήση τεχνικών επιλογής χαρακτηριστικών σε κάθε γύρο της ανατροφοδότησης, παράλληλα με την υιοθέτηση ταξινομητών για το διαχωρισμό των σχετικών από τις μη σχετικές εικόνες. Ως αποτέλεσμα της ενασχόλησής μας με το θέμα, μια πολλά υποσχόμενη μέθοδος επιλογής χαρακτηριστικών προτείνεται. Το κεφάλαιο 8 αποτελεί τον επίλογο της παρούσας διατριβής, συνοψίζοντας τα αποτελέσματα της έρευνας που έχει διεξαχθεί έως τώρα, και αναδεικνύοντας κατευθύνσεις για πιθανή συνέχισή της.





## Κεφάλαιο 2

### Πιθανοτικά Μοντέλα

Στο κεφάλαιο αυτό θα ασχοληθούμε με μοντέλα που χρησιμοποιούνται για να περιγράψουν με ένα πιθανοτικό (probabilistic) τρόπο σύνολα δεδομένων. Όποιος ενδιαφέρεται για περισσότερες πληροφορίες πάνω σε αυτό το θέμα μπορεί να ανατρέξει παράλληλα και στις εργασίες [16], [29], [33], [35], [36] και [37]. Ιδιαίτερο βάρος θα δοθεί σε ό,τι έχει να κάνει με τα Γκαουσιανά Μείγματα, τις ιδιότητές τους και τον τρόπο εκπαίδευσής τους.

#### 2.1 Βασικές Έννοιες Πιθανοτήτων

Θα ξεκινήσουμε με κάποιες βασικές έννοιες της θεωρίας πιθανοτήτων. Στον πυρήνα αυτής της θεωρίας βρίσκεται η έννοια της τυχαίας μεταβλητής. Μια εν γένει πολυδιάστατη τυχαία μεταβλητή, έστω  $\mathbf{X}$ , είναι μια μεταβλητή που μπορεί να πάρει οποιαδήποτε τιμή από ένα συγκεκριμένο σύνολο τιμών. Κάθε μια από τις τιμές αυτού του συνόλου συνδέεται με ένα μέτρο πιθανότητας, που αφορά το πόσο πιθανό είναι η τυχαία μεταβλητή να λάβει την εν λόγω τιμή. Αν το σύνολο τιμών είναι διακριτό, τότε η τυχαία μεταβλητή λέγεται διακριτή και το μέτρο πιθανότητας για κάθε μια από τις δυνατές τιμές της εκφράζει, ακριβώς, την πιθανότητα η τυχαία μεταβλητή να λάβει τη συγκεκριμένη τιμή. Έτσι, ορίζεται μια συνάρτηση

$$P(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in D \quad (2.1)$$

$$\sum_{\mathbf{x} \in D} P(\mathbf{x}) = 1 \quad (2.2)$$

που αποκαλείται συνάρτηση πιθανότητας της τυχαίας μεταβλητής. Το σύνολο  $D$  είναι το σύνολο όλων των δυνατών τιμών που μπορεί να πάρει η τυχαία μεταβλητή, για αυτό και το άθροισμα των τιμών της συνάρτησης για όλα τα στοιχεία αυτού του συνόλου ισούται με τη μονάδα, κάτι που εκφράζει μαθηματικά τη βεβαιότητά μας ότι η τυχαία μεταβλητή μπορεί να πάρει τιμές μόνο μέσα στο  $D$ .

Τώρα, αν το σύνολο  $D$  είναι συνεχές, η τυχαία μεταβλητή λέγεται συνεχής και για αυτήν ορίζεται μια συνάρτηση

$$p(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in D \quad (2.3)$$

$$\int_{\mathbf{x} \in D} p(\mathbf{x}) d\mathbf{x} = 1 \quad (2.4)$$

που αποκαλείται συνάρτηση πυκνότητας πιθανότητας (Probability Density Function – pdf) και η τιμή της στο  $\mathbf{x}$  αντιστοιχεί στην πυκνότητα πιθανότητας της τυχαίας μεταβλητής στο συγκεκριμένο σημείο. Με άλλα λόγια, η  $p(\mathbf{x})$  μας πληροφορεί ότι

$$P(\mathbf{X} \in [\mathbf{x}, \mathbf{x} + d\mathbf{x}]) = p(\mathbf{x})d\mathbf{x} \quad (2.5)$$

όπου το αριστερό μέλος της σχέσης εκφράζει την πιθανότητα κάθε συνιστώσα, έστω  $i$ , της τυχαίας μεταβλητής  $\mathbf{X}$  να λάβει τιμή στο διάστημα  $[x_i, x_i + dx_i]$ , με  $x_i$  να είναι η αντίστοιχη συνιστώσα του  $\mathbf{x}$ . Το  $d\mathbf{x}$  στο δεξιό μέλος εκφράζει έναν απειροστό όγκο στο  $\mathbf{x}$ .

Δύο πολύ σημαντικά χαρακτηριστικά μεγέθη για τυχαίες μεταβλητές είναι η προσδοκώμενη τιμή (expected value) και η διασπορά (variance). Η προσδοκώμενη τιμή για διακριτές τυχαίες μεταβλητές ορίζεται ως

$$E[\mathbf{X}] = \sum_{\mathbf{x} \in D} \mathbf{x}P(\mathbf{x}) \quad (2.6)$$

και για συνεχείς τυχαίες μεταβλητές ως

$$E[\mathbf{X}] = \int_{\mathbf{x} \in D} \mathbf{x}p(\mathbf{x})d\mathbf{x} \quad (2.7)$$

Η διασπορά, από την άλλη, δίνεται από τον τύπο

$$V[\mathbf{X}] = E\{(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T\} \quad (2.8)$$

Προφανώς, ο παραπάνω τύπος λαμβάνει υπόψιν το γεγονός ότι η ποσότητα  $(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T$  εξαρτάται από την τυχαία μεταβλητή  $\mathbf{X}$  και, επομένως, είναι και η ίδια μια τυχαία μεταβλητή.

Εκτός από τους συμβολισμούς που είδαμε παραπάνω, για τα διάφορα μέτρα πιθανότητας μπορούν να χρησιμοποιηθούν και άλλες ισοδύναμες παραλλαγές. Για παράδειγμα, αντί για το συμβολισμό  $P(\mathbf{x})$  μπορεί να χρησιμοποιηθούν οι ισοδύναμοι συμβολισμοί  $P_{\mathbf{x}}(\mathbf{x})$ ,  $P(\mathbf{X} = \mathbf{x})$  κτλ., που καθιστούν όμως πιο ευδιάκριτη την τυχαία μεταβλητή, εδώ τη  $\mathbf{X}$ , στην οποία αναφερόμαστε. Τέλος, υπενθυμίζουμε ότι ένας συμβολισμός όπως ο  $p(\mathbf{x}|a)$  αναφέρεται στην πιθανότητα μιας τιμής  $\mathbf{x}$  με δεδομένο ότι είναι γνωστές κάποιες παράμετροι  $a$ , ή ότι ισχύει κάποια συνθήκη  $a$  κτλ.

## 2.2 Ιστογράμματα

Σε αυτήν την ενότητα θα ασχοληθούμε λίγο με τα Ιστογράμματα, τα οποία θα χρησιμοποιήσουμε αρκετά συχνά στη συνέχεια ως ένα μέσο για να εξάγουμε με απλό τρόπο περιγραφές για τις εικόνες.

Τα Ιστογράμματα αποτελούν πολύ δημοφιλή και παράλληλα απλά διακριτά μοντέλα, μέσω των οποίων μπορούμε να προσεγγίσουμε μια άγνωστη συνεχή κατανομή  $p(\mathbf{x})$ . Με άλλα λόγια, θεωρώντας ότι έχουμε μια συνεχή τυχαία μεταβλητή  $\mathbf{X}$ , που παίρνει τιμές σε ένα σύνολο  $D$ , και ένα σύνολο παρατηρήσεων  $S$  για αυτήν:

$$S = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \quad (2.9)$$

με κάθε  $\mathbf{x}_i \in D$  να αποτελεί ένα στιγμιότυπο της  $\mathbf{X}$ , μπορούμε να χωρίσουμε το σύνολο  $D$  σε ένα σχετικά μεγάλο αριθμό υποσυνόλων  $D_j$ , τα οποία θα καλύπτουν όλο το  $D$  και θα είναι ξένα μεταξύ τους, δηλαδή θα έχουν ανά δύο κενή τομή. Συγκεκριμένα, αν  $Z$  είναι το πλήθος αυτών των υποσυνόλων, θα ισχύει

$$D = \bigcup_{j=1}^Z D_j \quad (2.10)$$

$$D_j \cap D_k = \emptyset, \text{ αν } j \neq k \quad (2.11)$$

Έπειτα, για κάθε ένα από αυτά τα  $D_j$  υπολογίζουμε έναν αριθμό  $c_j$  που εκφράζει το πόσες από τις παρατηρήσεις του  $S$  ανήκουν στο  $D_j$ :

$$c_j = \text{card}(S \cap D_j) \quad (2.12)$$

όπου η συνάρτηση  $\text{card}$  υπολογίζει τον πληθάρημο, δηλαδή το πλήθος στοιχείων ενός συνόλου. Διαιρώντας κάθε  $c_j$  με το συνολικό πλήθος των παρατηρήσεων, παίρνουμε τις ποσότητες

$$c_j^n = \frac{c_j}{N} \quad (2.13)$$

για τις οποίες ισχύει

$$\sum_{j=1}^Z c_j^n = 1 \quad (2.14)$$

και κάθε μια από τις οι οποίες αποτελεί, προφανώς, προσέγγιση της πιθανότητας οι τιμές της μεταβλητής  $\mathbf{X}$  να ανήκουν στο αντίστοιχο υποσύνολο  $D_j$ . Με άλλα λόγια, έχουμε

$$c_j^n \approx P(\mathbf{X} \in D_j) \quad (2.15)$$

Με αυτόν τον τρόπο παίρνουμε μια διακριτή προσέγγιση της κατανομής της συνεχούς μεταβλητής  $\mathbf{X}$ . Είναι προφανές, με βάση το «νόμο των μεγάλων αριθμών» της Στατιστικής, ότι όσο ο αριθμός των παρατηρήσεων και ο αριθμός των υποσυνόλων στα οποία χωρίζεται το πεδίο ορισμού, δηλαδή όσο το  $N$  και το  $Z$ , αυξάνονται, τόσο καλύτερη γίνεται αυτή η προσέγγιση.

Τώρα, στην πράξη, το σύνολο  $D$  είναι κάποιο υποσύνολο του  $\mathcal{R}^d$ ,  $d = 1, 2, \dots$ , δηλαδή του συνόλου των  $d$ -διάστατων πραγματικών διανυσμάτων. Κάθε παρατήρηση, λοιπόν, είναι εν γένει ένα διάνυσμα  $d$  διαστάσεων. Ξέροντας το εύρος των τιμών για κάθε διάσταση ή προσδιορίζοντάς το από τις παρατηρήσεις, καταλήγουμε συνήθως σε  $d$  διαστήματα πραγματικών αριθμών, ένα για κάθε διάσταση. Σε αυτήν την περίπτωση, ο πιο δημοφιλής

τρόπος ορισμού των  $D_j$  που περιγράψαμε παραπάνω είναι να θεωρήσουμε ότι το εύρος τιμών για κάθε διάσταση διαιρείται σε ίσα υποδιαστήματα. Δηλαδή, θεωρώντας για παράδειγμα τη διάσταση  $r$ , με εύρος

$$D^r = [a_r, b_r] \subset \mathfrak{R} \quad (2.16)$$

διαιρούμε το  $D^r$  σε, έστω  $Z_r$ , ίσου μήκους υποδιαστήματα. Αν επαναλάβουμε αυτή τη διαδικασία για όλες τις διαστάσεις, προκύπτουν

$$Z = Z_1 Z_2 \cdots Z_d \quad (2.17)$$

υποσύνολα  $D_j$ , καθένα από τα οποία αποτελεί έναν υπερκύβο, εν γένει, στο  $D$ . Τα «κελιά» αυτά στα οποία υποδιαιρείται ο χώρος  $D$  ονομάζονται και bins.

Από τα παραπάνω αρχίζει να διαφαίνεται και το μεγαλύτερο πρόβλημα των Ιστογραμμάτων, το οποίο περιγράφεται μέσω το όρου «κατάρα της διάστασης» (curse of dimensionality). Με απλά λόγια, το πλήθος των  $D_j$ ,  $Z$ , αυξάνει εκθετικά με τη διάσταση  $d$ . Επομένως, για τυχαίες μεταβλητές πολλών διαστάσεων, προκειμένου να υπολογίσουμε με κάποια αξιοπιστία τις τιμές  $c_j^n$  για κάθε  $D_j$  ένας πολύ μεγάλος αριθμός παρατηρήσεων,  $N$ , είναι αναγκαίος. Συχνά, αυτή η απαίτηση καθίσταται απαγορευτική ακόμα και για προβλήματα σχετικά λίγων διαστάσεων, ακριβώς λόγω αυτής της εκθετικής σχέσης που συνδέει το  $Z$  με το  $d$ .

## 2.3 Γκαουσιανή Κατανομή

Σε αυτό το σημείο θα αναφερθούμε συνοπτικά στο δημοφιλέστερο και περισσότερο μελετημένο και χρησιμοποιημένο πιθανοτικό μοντέλο. Αυτό δεν είναι άλλο από τη Γκαουσιανή (ή Κανονική) κατανομή.

### 2.3.1 Ορισμός

Η Γκαουσιανή κατανομή είναι η πιο ελκυστική επιλογή όταν έχουμε να μοντελοποιήσουμε στατιστικά πρόβλημα, αλλά και γενικότερα για προβλήματα στα οποία κάποια παράμετρος πρέπει να περιγραφεί μέσω κάποιου πιθανοτικού μοντέλου. Το γεγονός αυτό οφείλεται στα πολλά μαθηματικά πλεονεκτήματα που έχει η συγκεκριμένη κατανομή, τα οποία την καθιστούν εύκολη στο χειρισμό της. Συγκεκριμένα, αν θεωρήσουμε ότι έχουμε μια εν γένει πολυδιάστατη συνεχή τυχαία μεταβλητή  $\mathbf{X}$  που ακολουθεί Γκαουσιανή κατανομή, η αντίστοιχη συνάρτηση πυκνότητας πιθανότητας δίνεται από τον παρακάτω τύπο:

$$p(\mathbf{x} | \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right] \quad (2.18)$$

$$\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (2.19)$$

όπου  $\mathbf{x}$  είναι μια πιθανή τιμή της τυχαίας μεταβλητής,  $d$  η διάστασή της,  $\boldsymbol{\mu}$  η μέση τιμή της και  $\boldsymbol{\Sigma}$  ο πίνακας συνδιακύμανσης που της αντιστοιχεί.

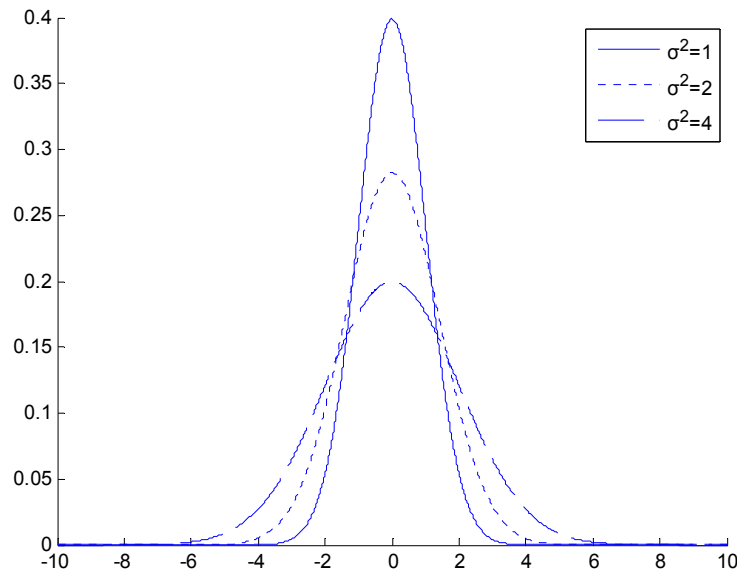
### 2.3.2 Ιδιότητες

Αποδεικνύεται σχετικά εύκολα ότι στην περίπτωση μιας τυχαίας μεταβλητής που ακολουθεί Γκαουσιανή κατανομή ισχύουν οι παρακάτω σχέσεις:

$$E[\mathbf{X}] = \boldsymbol{\mu} \quad (2.20)$$

$$V[\mathbf{X}] = \boldsymbol{\Sigma} \quad (2.21)$$

για την προσδοκώμενη τιμή και τη διασπορά, αντίστοιχα, όπως αυτές ορίστηκαν μέσω των τύπων (2.7) και (2.8).



**Σχήμα 2.1:** Μονοδιάστατες Γκαουσιανές κατανομές για διάφορες τιμές του  $\sigma$

Η συνάρτηση πυκνότητας πιθανότητας  $p(\mathbf{x} | \boldsymbol{\theta})$  της Γκαουσιανής κατανομής εμφανίζει μέγιστη τιμή για  $\mathbf{x} = \boldsymbol{\mu}$ , δηλαδή ακριβώς στη μέση τιμή της κατανομής. Επομένως, αλλάζοντας τη μέση τιμή της κατανομής μπορούμε να μετατοπίσουμε όλη τη συνάρτηση πυκνότητας πιθανότητας όσο και όπου επιθυμούμε στο χώρο τιμών της τυχαίας μεταβλητής, δηλαδή στον  $\mathcal{R}^d$ . Από την άλλη, το πόσο πεπλατυσμένη είναι η Γκαουσιανή κατανομή καθορίζεται από την παράμετρο  $\boldsymbol{\Sigma}$ .

Σε μονοδιάστατες Γκαουσιανές κατανομές η παραπάνω παράμετρος αντιστοιχεί σε έναν θετικό πραγματικό αριθμό. Άρα, θα ισχύει

$$\boldsymbol{\Sigma} = \sigma^2, \quad \sigma \in [0, +\infty) \quad (2.22)$$

με το  $\sigma$  να αποτελεί την τυπική απόκλιση (standard deviation) της κατανομής. Η συνάρτηση πυκνότητας πιθανότητας είναι συμμετρική γύρω από το  $\mu$ , ενώ όσο μεγαλύτερο είναι το  $\sigma$ , τόσο πιο διευρυμένη σε πλάτος και συμπίεσμένη σε ύψος είναι. Αντίστροφα, ένα μικρό  $\sigma$  αντιστοιχεί σε μια κατανομή πολύ συμπίεσμένη γύρω από τη μέση τιμή. Στην ακραία

περίπτωση κατά την οποία ισχύει  $\sigma = 0$ , παίρνουμε την κρουστική συνάρτηση Dirac, με μηδενικό πλάτος και άπειρο ύψος. Αυτή η εξάρτηση από το  $\sigma$  παρουσιάζεται στο σχήμα 2.1, όπου απεικονίζονται 3 μονοδιάστατες Γκαουσιανές κατανομές με  $\mu = 0$  και  $\Sigma = 1, 2$  και  $4$ , αντίστοιχα. Σε αυτό το σχήμα βλέπουμε και τη χαρακτηριστική μορφή «καμπάνας», η οποία αποτελεί σήμα κατατεθέν της Γκαουσιανής κατανομής.

Σε Γκαουσιανές κατανομές  $d > 1$  διαστάσεων, η παράμετρος  $\Sigma$  είναι τετραγωνικός  $d \times d$ , συμμετρικός, θετικά ορισμένος πίνακας. Η μέση τιμή αποτελεί, πάλι, την κορυφή της κατανομής και με κέντρο αυτήν ορίζονται  $d$  άξονες συμμετρίας της κατανομής. Οι άξονες αυτοί προκύπτουν από τα ιδιοδιανύσματα του πίνακα  $\Sigma$ . Ακόμα, για κάθε άξονα συμμετρίας, έστω  $l$ , μπορεί να οριστεί ένα μέγεθος  $s_l$ , αντίστοιχο της τυπικής απόκλισης  $\sigma$  στη μία διάσταση και ίσο με την τετραγωνική ρίζα της αντίστοιχης ιδιοτιμής. Οι ισοϋψείς καμπύλες της πολυδιάστατης Γκαουσιανής συνάρτησης πυκνότητας πιθανότητας έχουν τη μορφή ελλείψεων, με άξονες τους προαναφερθέντες άξονες συμμετρίας και μήκη αξόνων ανάλογα των  $s_l$ .

Η συνάρτηση πυκνότητας πιθανότητας της Γκαουσιανής κατανομής φθίνει ασυμπτωτικά προς το μηδέν σχετικά γρήγορα όσο απομακρυνόμαστε από τη μέση τιμή. Αυτό, μάλιστα, είναι ένα από τα μειονεκτήματα της Γκαουσιανής κατανομής. Το ότι αποδίδει πολύ μικρή πιθανότητα σε τιμές σχετικά απομακρυσμένες από την πιθανότερη τιμή αποτελεί μειονέκτημα σε πρακτικά προβλήματα, στα οποία οι παρατηρήσεις εμπεριέχουν συχνά θόρυβο (noise) ή τιμές που αποκλίνουν αισθητά από το κύριο σώμα των παρατηρήσεων (outliers). Σε τέτοιες περιπτώσεις, η υπόθεση ότι οι παρατηρήσεις ακολουθούν Γκαουσιανή κατανομή μπορεί να οδηγήσει σε ανεπιθύμητα αποτελέσματα, όπως, για παράδειγμα, ο σοβαρός επηρεασμός που ενδέχεται να ασκήσει ο θόρυβος στην εκτίμηση των παραμέτρων της κατανομής. Για την αποφυγή τέτοιων προβλημάτων έχουν προταθεί άλλα πιθανοτικά μοντέλα, αντί του Γκαουσιανού, όπως το Λαπλασιανό (Laplacian) μοντέλο, η κατανομή  $t$  κτλ., που εν γένει εμφανίζουν «βαριές ουρές» (heavy tails), δηλαδή αποδίδουν σχετικά υψηλότερες πιθανότητες σε τιμές απομακρυσμένες από τη μέση τιμή της κατανομής. Για περισσότερες πληροφορίες ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει και στις εργασίες [80], [81] και [82]. Πάντως, σε σχέση με τις παραπάνω κατανομές, η Γκαουσιανή κατανομή έχει ένα πολύ μεγάλο πλεονέκτημα, και αυτό έχει να κάνει με το ότι είναι ασύγκριτα πιο εύχρηστη από μαθηματικής σκοπιάς.

### 2.3.3 Εκτίμηση Παραμέτρων

Ας θεωρήσουμε τώρα ότι έχουμε ένα σύνολο παρατηρήσεων  $S$ , όπως αυτό ορίστηκε από τη σχέση (2.9), για μια τυχαία μεταβλητή  $\mathbf{X}$ . Ας κάνουμε την υπόθεση, επίσης, ότι αυτή η τυχαία μεταβλητή ακολουθεί Γκαουσιανή κατανομή με άγνωστη μέση τιμή  $\boldsymbol{\mu}$  και άγνωστο πίνακα συνδιακύμανσης  $\Sigma$ . Δηλαδή, όπως θα μπορούσαμε να το εκφράσουμε συμβολικά:

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \quad (2.23)$$

Το ερώτημα που γεννάται άμεσα αφορά το πως μπορούμε να εκτιμήσουμε τις βέλτιστες τιμές για τις παραμέτρους  $\boldsymbol{\mu}$ ,  $\Sigma$ , χρησιμοποιώντας τη γνώση που προκύπτει από τις παρατηρήσεις. Για να το πετύχουμε αυτό μπορούμε να βασιστούμε στην αρχή της μέγιστης πιθανοφάνειας (maximum likelihood). Θεωρώντας ότι οι παρατηρήσεις είναι ανεξάρτητες μεταξύ τους, η πιθανοφάνεια του συνόλου παρατηρήσεων  $S$  υπό τη συνάρτηση πυκνότητας πιθανότητας  $p(\mathbf{x} | \boldsymbol{\theta})$  ορίζεται από τη σχέση

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^N p(\mathbf{x}_i | \boldsymbol{\theta}) \quad (2.24)$$

Η  $\mathcal{L}(\boldsymbol{\theta})$  εξαρτάται, προφανώς, από τις παραμέτρους  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  του μοντέλου. Μεγιστοποιώντας το λογάριθμο αυτής της ποσότητας ως προς  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  προκύπτουν οι εκτιμήτριες μέγιστης πιθανοφάνειας

$$\boldsymbol{\mu}_{\text{ML}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (2.25)$$

$$\boldsymbol{\Sigma}_{\text{ML}} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu}_{\text{ML}})(\mathbf{x}_i - \boldsymbol{\mu}_{\text{ML}})^T = \left( \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T \right) - \boldsymbol{\mu}_{\text{ML}} \boldsymbol{\mu}_{\text{ML}}^T \quad (2.26)$$

για τη μέση τιμή και τον πίνακα συνδιακύμανσης της κατανομής, αντίστοιχα. Όμως, η γνώση των τιμών για αυτές τις δύο παραμέτρους αρκεί για τον πλήρη καθορισμό της αντίστοιχης Γκαουσιανής κατανομής. Επομένως, βλέπουμε ότι πολύ εύκολα μπορούμε από ένα σύνολο παρατηρήσεων να καταλήξουμε σε ένα Γκαουσιανό μοντέλο που να τις περιγράφει με βέλτιστο τρόπο.

## 2.4 Γκαουσιανά Μείγματα

Η Γκαουσιανή κατανομή, όπως φαίνεται και από το σχήμα «καμπάνας» που τη χαρακτηρίζει, είναι κατάλληλη για μεταβλητές οι τιμές των οποίων εμφανίζουν μια σχετικά απλή κατανομή γύρω από μια κεντρική τιμή. Για πιο πολύπλοκα κατανεμημένες μεταβλητές χρειάζεται να αναζητηθούν άλλα πιθανοτικά μοντέλα. Συνήθως, ως βάση για τον ορισμό τέτοιων πολύπλοκων μοντέλων χρησιμοποιούνται και πάλι στοιχειώδη μοντέλα. Σε αυτήν την ενότητα θα δούμε πως μπορεί να ορισθεί ένα μοντέλο Γκαουσιανού Μείγματος, χρησιμοποιώντας ως βάση τη Γκαουσιανή κατανομή.

### 2.4.1 Ορισμός

Ας θεωρήσουμε ότι έχουμε μια στοιχειώδη κατανομή με συνάρτηση πυκνότητας πιθανότητας  $\phi(\mathbf{x} | \boldsymbol{\theta})$ , εξαρτώμενη από ένα σύνολο παραμέτρων  $\boldsymbol{\theta}$ . Ένας εύκολος και μαθηματικά εύχρηστος τρόπος να ορίσουμε μια σύνθετη κατανομή με βάση τη  $\phi(\mathbf{x} | \boldsymbol{\theta})$  είναι μέσω του παρακάτω κυρτού συνδυασμού:

$$p(\mathbf{x} | \boldsymbol{\Theta}) = \sum_{j=1}^K \pi_j \phi(\mathbf{x} | \boldsymbol{\theta}_j) \quad (2.27)$$

$$\boldsymbol{\Theta} = (\pi_1, \boldsymbol{\theta}_1, \pi_2, \boldsymbol{\theta}_2, \dots, \pi_K, \boldsymbol{\theta}_K) \quad (2.28)$$

όπου  $\pi_j$  είναι τα βάρη των συνιστωσών κατανομών στο συνδυασμό, για τα οποία ισχύει

$$\pi_j \geq 0, \quad \forall j = 1, \dots, K, \quad \text{και} \quad \sum_j \pi_j = 1 \quad (2.29)$$

Συνδυάζουμε, λοιπόν, με γραμμικό τρόπο και με διαφορετικό βάρος για την καθεμιά, έναν αριθμό  $K$  από διαφορετικές εκδοχές της αρχικής συνάρτησης πυκνότητας πιθανότητας, ώστε να καταλήξουμε σε μια τελική συνάρτηση πυκνότητας πιθανότητας που θα χαρακτηρίζει τη νέα κατανομή.

Μια καλή επιλογή, ώστε να μη χάσουμε τα πολλά πλεονεκτήματα και τις καλές ιδιότητες της Γκαουσιανής κατανομής, είναι να χρησιμοποιήσουμε ως βάση για τη σύνθετη κατανομή τη Γκαουσιανή κατανομή, δηλαδή να θέσουμε

$$\phi(\mathbf{x} | \boldsymbol{\theta}_j) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_j|}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1}(\mathbf{x} - \boldsymbol{\mu}_j)\right] \quad (2.30)$$

$$\boldsymbol{\theta}_j = (\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad (2.31)$$

Έτσι, καταλήγουμε στο πιθανοτικό μοντέλο που αποκαλείται Γκαουσιανό Μείγμα.

## 2.4.2 Ιδιότητες

Ένα Γκαουσιανό Μείγμα, σε αντίθεση με την απλή Γκαουσιανή κατανομή, μπορεί να γίνει όσο πολύπλοκο θέλουμε με κατάλληλη προσαρμογή των παραμέτρων του. Η παράμετρος που ιδίως ελέγχει την πολυπλοκότητά του είναι το  $K$ , ο αριθμός δηλαδή των Γκαουσιανών συνιστωσών ή, όπως αλλιώς θα μπορούσαμε να πούμε, των πυρήνων (kernels) στο μείγμα. Συνεπώς, ένα Γκαουσιανό Μείγμα μας δίνει ως μοντέλο πολύ μεγάλη ευελιξία. Είναι εξάλλου αποδεδειγμένο ότι η οποιαδήποτε συνάρτηση, υπό την προϋπόθεση βέβαια ότι ικανοποιεί κάποιους μαθηματικούς κανόνες ομαλότητας, μπορεί να προσεγγιστεί σε οποιαδήποτε ακρίβεια από ένα Γκαουσιανό Μείγμα με κατάλληλα επιλεγμένο αριθμό πυρήνων.

Πρέπει να σημειωθεί ακόμα ότι ένα Γκαουσιανό Μείγμα έχει εγγενώς ιεραρχική δομή. Για να γίνει αυτό σαφές, ας υποθέσουμε ότι έχουμε μια τυχαία μεταβλητή που ακολουθεί μια κατανομή που περιγράφεται μέσω ενός Γκαουσιανού Μείγματος. Ένα στιγμιότυπο (instance) αυτής της μεταβλητής παράγεται με τον εξής τρόπο:

1. Χρησιμοποιώντας ως πιθανότητα για την επιλογή του  $j$ -οστού πυρήνα το αντίστοιχο βάρος  $\pi_j$  του μείγματος, επιλέγεται ένας από τους  $K$  Γκαουσιανούς πυρήνες.
2. Το στιγμιότυπο παράγεται από την κατανομή του Γκαουσιανού πυρήνα που επιλέχθηκε.

Επομένως, κάθε στιγμιότυπο μιας τυχαίας μεταβλητής που ακολουθεί κατανομή Γκαουσιανού Μείγματος συνδέεται άρρηκτα με ένα συγκεκριμένο πυρήνα του μείγματος. Αυτή η ιδιότητα θα παίζει πρωτεύοντα ρόλο όσο αφορά το πρόβλημα της εκτίμησης των παραμέτρων ενός Γκαουσιανού Μείγματος.

## 2.4.3 Εκτίμηση Παραμέτρων

### 2.4.3.1 Αλγόριθμος EM

Σε αυτό το σημείο θα ασχοληθούμε με το πως μπορούμε να καθορίσουμε τις τιμές των παραμέτρων ενός Γκαουσιανού Μείγματος, ώστε αυτό να περιγράφει κατά τον καλύτερο



δυνατό τρόπο ένα σύνολο παρατηρήσεων. Συγκεκριμένα, έστω ότι έχουμε μια τυχαία μεταβλητή  $\mathbf{X}$  για την οποία υποθέτουμε ότι ακολουθεί μια κατανομή Γκαουσιανού Μείγματος  $K$  συνιστωσών με άγνωστες παραμέτρους  $\Theta$ , όπως αυτές καθορίζονται από τις σχέσεις (2.28) και (2.31). Για την ώρα, θα θεωρήσουμε ότι ο αριθμός,  $K$ , των πυρήνων στο μείγμα είναι γνωστός και σταθερός. Επίσης, θεωρούμε ότι έχουμε ένα σύνολο παρατηρήσεων  $S$  για την τυχαία μεταβλητή  $\mathbf{X}$ , όπως αυτό ορίζεται στη σχέση (2.9). Το πρόβλημα που πρέπει να αντιμετωπίσουμε αφορά την εκτίμηση, με βέλτιστο τρόπο, των τιμών των παραμέτρων  $\Theta$ .

Για μια ακόμα φορά θα στηριχθούμε στην αρχή της μέγιστης πιθανοφάνειας. Όμως εδώ τα πράγματα είναι λίγο πιο περίπλοκα από ό,τι στην περίπτωση της απλής Γκαουσιανής, γιατί, όπως περιγράψαμε παραπάνω, ένα στιγμιότυπο της μεταβλητής παράγεται σε δύο στάδια. Το πρώτο στάδιο, η επιλογή δηλαδή του πυρήνα που αντιστοιχεί στην κάθε παρατήρηση, είναι «κρυφό» στην περίπτωση μας. Με άλλα λόγια, έχουμε μεν τις παρατηρήσεις, αλλά δε μπορούμε να ξέρουμε από ποια Γκαουσιανή του μείγματος προήλθε η κάθε παρατήρηση. Το πρόβλημα αυτό είναι ένα κλασικό πρόβλημα ύπαρξης μη παρατηρήσιμων (unobserved) μεταβλητών. Ο κλασικός αλγόριθμος επίλυσής του είναι ο αλγόριθμος «μεγιστοποίησης της προσδοκίας» (Expectation-Maximization), EM.

Ο αλγόριθμος EM είναι επαναληπτικός και περιλαμβάνει δύο βήματα. Συγκεκριμένα, έχουμε το βήμα E (Expectation) και το βήμα M (Maximization), για τα οποία ισχύουν τα παρακάτω:

1. Βήμα E: Υπολογισμός των προσδοκώμενων τιμών για κάθε κρυφή μεταβλητή, υποθέτοντας ότι οι τρέχουσες εκτιμήσεις των παραμέτρων ισχύουν.
2. Βήμα M: Υπολογισμός μιας νέας εκτίμησης μέγιστης πιθανοφάνειας για τις παραμέτρους, θεωρώντας ότι οι κρυφές μεταβλητές έχουν τιμές ίσες με τις προσδοκώμενες.

Στην προκειμένη περίπτωση θεωρούμε ότι κάθε παρατήρηση  $\mathbf{x}_i$  είναι ένα  $d$ -διάστατο διάνυσμα

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id})^T \quad (2.32)$$

με το οποίο συνδέονται  $K$  κρυφές μεταβλητές,  $z_{i1}, z_{i2}, \dots, z_{iK}$ , οι οποίες καθορίζουν από ποια από τις  $K$  Γκαουσιανές του μείγματος παρήχθη η εν λόγω παρατήρηση. Οι δυνατές τιμές που μπορεί να πάρει κάθε μια από αυτές τις μεταβλητές είναι οι  $\{0,1\}$ . Το 1 (0) αντιστοιχεί στην περίπτωση κατά την οποία η παρατήρηση (δεν) προήλθε από τον αντίστοιχο πυρήνα. Επίσης, για κάθε  $i$  μόνο μια από τις  $K$  αντίστοιχες μεταβλητές φέρει την τιμή 1 και όλες οι άλλες ισούνται με 0, αφού μια συγκεκριμένη παρατήρηση μπορεί να παραχθεί από έναν και μόνο πυρήνα.

Το πρώτο βήμα του αλγορίθμου EM αφορά τον υπολογισμό των προσδοκώμενων τιμών  $E(z_{ij})$  για τις προαναφερθείσες κρυφές μεταβλητές. Όπως έχουμε δει ήδη, η προσδοκώμενη τιμή για μια διακριτή τυχαία μεταβλητή  $Z$  δίνεται από τον παρακάτω τύπο:

$$E[Z] = \sum_{\forall z} zP(Z = z) \quad (2.33)$$

Επομένως, για τις τυχαίες μεταβλητές που εξετάζουμε ισχύει ότι  $E(z_{ij}) = 0P(z_{ij} = 0) + 1P(z_{ij} = 1)$  και άρα

$$E(z_{ij}) = P(z_{ij} = 1) \quad (2.34)$$

Όμως, η  $P(z_{ij} = 1)$ , δηλαδή η πιθανότητα η κρυφή μεταβλητή  $z_{ij}$  να έχει την τιμή 1, προφανώς ισούται με την πιθανότητα ο  $j$ -οστός πυρήνας να έχει παράξει την  $i$ -οστή παρατήρηση. Αυτό μπορούμε να το διατυπώσουμε συμβολικά ως

$$P(z_{ij} = 1) = P(j | \mathbf{x}_i) \quad (2.35)$$

Σύμφωνα με τον τύπο του Bayes, όπως αυτός δίνεται από την παρακάτω σχέση:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (2.36)$$

προκύπτει ότι

$$P(j | \mathbf{x}_i) = \frac{P(j)p(\mathbf{x}_i | j)}{p(\mathbf{x}_i)} \quad (2.37)$$

Ενσωματώνοντας τώρα την υπόθεσή μας ότι οι παρατηρήσεις ακολουθούν κατανομή Γκαουσιανού Μείγματος, παίρνουμε την τελική μορφή:

$$P(j | \mathbf{x}_i) = \frac{\pi_j \phi(\mathbf{x}_i | \boldsymbol{\theta}_j)}{\sum_{k=1}^K \pi_k \phi(\mathbf{x}_i | \boldsymbol{\theta}_k)} \quad (2.38)$$

όπου η  $\phi(\mathbf{x} | \boldsymbol{\theta}_j)$  είναι  $j$ -οστή Γκαουσιανή συνάρτηση πυκνότητας πιθανότητας του μείγματος. Επομένως, από τα παραπάνω, η προσδοκώμενη τιμή για τις κρυφές μεταβλητές  $z_{ij}$  υπολογίζεται ως

$$E[z_{ij}] = \frac{\pi_j \phi(\mathbf{x}_i | \boldsymbol{\theta}_j)}{\sum_{k=1}^K \pi_k \phi(\mathbf{x}_i | \boldsymbol{\theta}_k)} \quad (2.39)$$

Στο δεύτερο βήμα του αλγορίθμου EM χρησιμοποιούμε αυτές τις εκτιμήσεις για τις τιμές των κρυφών μεταβλητών, ώστε να υπολογίσουμε μια νέα εκτίμηση για τις παραμέτρους  $\Theta$  του μοντέλου. Το συγκεκριμένο βήμα ανάγεται απευθείας στην απλή περίπτωση υπολογισμού των παραμέτρων μιας Γκαουσιανής κατανομής από ένα σύνολο παρατηρήσεων, αφού κατά το βήμα αυτό θεωρούμε ότι οι κρυφές μεταβλητές έχουν τιμές ίσες με τις προσδοκώμενες που υπολογίστηκαν στο προηγούμενο βήμα. Όμως, από τη στιγμή που ξέρουμε τις τιμές των κρυφών μεταβλητών, ουσιαστικά ξέρουμε ποιες παρατηρήσεις αντιστοιχούν σε ποιους πυρήνες και, επομένως, από τις παρατηρήσεις που αντιστοιχούν σε κάθε πυρήνα μπορούμε να εκτιμήσουμε τις βέλτιστες τιμές για τις παραμέτρους του. Συνεπώς, το αποτέλεσμα για τις νέες τιμές των παραμέτρων  $\Theta$  προκύπτει με αρκετά διαισθητικό τρόπο, ως

$$\pi_j = \frac{1}{N} \sum_{i=1}^N E[z_{ij}] \quad (2.40)$$

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^N E[z_{ij}] \mathbf{x}_i}{\sum_{i=1}^N E[z_{ij}]} \quad (2.41)$$

$$\boldsymbol{\Sigma}_j = \frac{\sum_{i=1}^N E[z_{ij}] \mathbf{x}_i \mathbf{x}_i^T}{\sum_{i=1}^N E[z_{ij}]} - \boldsymbol{\mu}_j \boldsymbol{\mu}_j^T \quad (2.42)$$

Προφανώς, με βάση την ανωτέρω πιθανοτική ανάλυση και τον υπολογισμό των προσδοκώμενων τιμών για τις κρυφές μεταβλητές  $z_{ij}$ , κάθε παρατήρηση  $\mathbf{x}_i$  δεν αντιστοιχίζεται σε μια μόνο Γκαουσιανή αλλά σε όλες, με διαφορετικό βαθμό στην καθεμιά όμως, και για αυτό το λόγο χρησιμοποιούνται σταθμισμένοι μέσοι όροι για τον υπολογισμό των παραμέτρων.

Πρακτικά μιλώντας πλέον, ο αλγόριθμος EM λειτουργεί ως εξής: Αρχικά θεωρούμε μια πρώτη εκτίμηση των παραμέτρων του μοντέλου. Για παράδειγμα, μπορούμε να θεωρήσουμε τα βάρη  $\pi_j$  των πυρήνων ίσα μεταξύ τους, κάθε μια από τις μέσες τιμές  $\boldsymbol{\mu}_j$  να την εξισώσουμε με κάποια τυχαία επιλεγμένη παρατήρηση  $\mathbf{x}_i$ , και να θεωρήσουμε τους πίνακες συνδιακύμανσης  $\boldsymbol{\Sigma}_j$  όλους ίσους με κάποιο κατάλληλο (θεωρώντας τη διασπορά του συνόλου των παρατηρήσεων) πολλαπλάσιο του αντίστοιχου μοναδιαίου πίνακα. Ένας πιο ενδεδειγμένος από θεωρητικής άποψης τρόπος αρχικοποίησης είναι να στηριχθούμε στον αλγόριθμο ομαδοποίησης (clustering) k-means (k-μέσων), και να χρησιμοποιήσουμε τα αποτελέσματα που μας δίνει αν τον εφαρμόσουμε στο σύνολο των παρατηρήσεων για  $K$  ομάδες (clusters), όσες και οι Γκαουσιανές του μοντέλου. Σε αυτήν την περίπτωση, ο αλγόριθμος k-means θα χωρίσει το σύνολο των παρατηρήσεων σε  $K$  υποσύνολα. Έπειτα, καθένα από αυτά τα υποσύνολα μπορεί να χρησιμοποιηθεί για την αρχικοποίηση μιας από τις Γκαουσιανές του μοντέλου. Συγκεκριμένα, για το  $j$ -οστό υποσύνολο μπορούμε να υπολογίσουμε το ποσοστό του συνόλου των παρατηρήσεων που του αντιστοιχούν, τη μέση τιμή και τη συνδιακύμανσή τους, και με αυτά να αρχικοποιήσουμε τις παραμέτρους  $\pi_j$ ,  $\boldsymbol{\mu}_j$ ,  $\boldsymbol{\Sigma}_j$ , αντίστοιχα, της  $j$ -οστής Γκαουσιανής κατανομής του μείγματος.

Μετά από την αρχικοποίηση, θεωρώντας επομένως ότι έχουμε μια αρχική εκτίμηση για τις παραμέτρους  $\boldsymbol{\Theta}$ , έστω  $\boldsymbol{\Theta}^0$ , εκτελούμε το πρώτο βήμα του αλγορίθμου EM και παίρνουμε μια εκτίμηση για τις κρυφές μεταβλητές  $z_{ij}$ . Έπειτα, εκτελούμε το δεύτερο βήμα του EM και ανανεώνουμε την εκτίμησή μας για τις παραμέτρους  $\boldsymbol{\Theta}$ , από  $\boldsymbol{\Theta}^0$  σε  $\boldsymbol{\Theta}^1$ . Κατόπιν, επιστρέφουμε πάλι στο πρώτο βήμα του EM και ούτω καθεξής. Συνεχίζοντας κατά αυτόν τρόπο και ακολουθώντας επαναληπτικά τα βήματα του αλγορίθμου EM, παίρνουμε διαδοχικές εκτιμήσεις,  $\boldsymbol{\Theta}^0$ ,  $\boldsymbol{\Theta}^1$ , ...,  $\boldsymbol{\Theta}^t$ , ... για τις παραμέτρους του μοντέλου, και σταματάμε όταν ο αλγόριθμος συγκλίνει, δηλαδή όταν οι παράμετροι από επανάληψη σε επανάληψη πάντως να αλλάζουν σημαντικά. Ένα άλλο κριτήριο τερματισμού – σύγκλισης προκύπτει από την παρατήρηση της τρέχουσας πιθανοφάνειας των παρατηρήσεων  $\mathcal{L}(\boldsymbol{\Theta}^t)$  (σχέση (2.24)), που ως μέτρο έχει την τάση να αυξάνεται από επανάληψη σε επανάληψη.

Όταν η πιθανοφάνεια σταθεροποιηθεί μέσα στα όρια κάποιας προκαθορισμένης ακρίβειας, αυτό αποτελεί ένδειξη ότι ο αλγόριθμος έχει συγκλίνει.

Από την προηγηθείσα ανάλυση αποκαλύπτεται, λοιπόν, άλλο ένα πλεονέκτημα των Γκαουσιανών Μειγμάτων, το οποίο είναι η σχετικά εύκολη και μαθηματικά θεμελιωμένη εκπαίδευση του μοντέλου με δεδομένο ένα σύνολο παρατηρήσεων. Όμως, το πρόβλημα που πρέπει να επισημανθεί είναι το εξής: Ο αλγόριθμος EM, όπως αναφέρθηκε και παραπάνω, απαιτεί κάποια αρχικοποίηση, μια αρχική εκτίμηση των παραμέτρων του μοντέλου. Έπειτα, συγκλίνει σε ένα τοπικό μέγιστο της πιθανοφάνειας μέσω της επαναληπτικής εκτέλεσης των βημάτων που αναλύθηκαν προηγουμένως. Ο χώρος των παραμέτρων, όμως, είναι γενικά μεγάλης διάστασης και η συνάρτηση της πιθανοφάνειας πολύπλοκη, με πολλά τοπικά ακρότατα. Αυτό έχει ως αποτέλεσμα η αρχικοποίηση να παίζει μεγάλο ρόλο ως προς το σημείο στο χώρο των παραμέτρων στο οποίο θα συγκλίνει ο αλγόριθμος. Με άλλα λόγια, ο κλασικός αλγόριθμος EM που περιγράψαμε σε αυτήν την ενότητα έχει μεγάλη εξάρτηση από την αρχικοποίηση. Για να αντιμετωπιστεί κάπως αυτό το πρόβλημα έχει προταθεί η εκτέλεση του αλγορίθμου περισσότερες της μίας φορές, με διαφορετική αρχικοποίηση την κάθε φορά, και η επιλογή του αποτελέσματος εκείνου που αντιστοιχεί σε μεγαλύτερη τελική πιθανοφάνεια των παρατηρήσεων. Όμως, αυτό, πέραν του ότι είναι πολύ απλοϊκό ως τρόπος αντιμετώπισης, οδηγεί και σε άσκοπη χρονική επιβάρυνση της εκπαίδευσης, αφού για το ίδιο σύνολο παρατηρήσεων, εκτός από το ένα μοντέλο που θα χρησιμοποιηθεί εν τέλει, εκπαιδεύονται και αρκετά άλλα που μένουν αχρησιμοποίητα.

#### 2.4.3.2 Άπληστος Αλγόριθμος EM

Ένας άλλος περισσότερο ενδιαφέρον τρόπος για να αποφύγουμε τη μεγάλη εξάρτηση του αλγορίθμου EM από την αρχικοποίηση των παραμέτρων είναι να βασιστούμε σε μια παραλλαγή του αλγορίθμου, η οποία έχει προταθεί σχετικά πρόσφατα [31] και ονομάζεται άπληστος (greedy) αλγόριθμος EM. Αυτόν τον αλγόριθμο θα τον χρησιμοποιήσουμε συχνά στα επόμενα κεφάλαια για την εκπαίδευση των μοντέλων Γκαουσιανού Μείγματος που μας χρειάζονται στα διάφορα πειράματα.

Ο συγκεκριμένος αλγόριθμος, όπως αποκαλύπτει και το όνομά του, στηρίζεται σε μια «άπληστη» λογική, με την έννοια ότι ξεκινάει από ένα «μείγμα» μιας Γκαουσιανής κατανομής και προσθέτει κατά βέλτιστο τρόπο, διαδοχικά, νέους Γκαουσιανούς πυρήνες, μέχρι να φτάσουμε σε έναν προκαθορισμένο μέγιστο αριθμό  $K_{\max}$ . Πιο αναλυτικά, ο άπληστος αλγόριθμος EM ακολουθεί τα εξής βήματα:

1. Περιγραφή των παρατηρήσεων προς μοντελοποίηση με μια Γκαουσιανή κατανομή.
2. Αρχικοποίηση με κατάλληλο τρόπο ενός νέου Γκαουσιανού πυρήνα και εισαγωγή του στο ήδη εκπαιδευμένο μοντέλο.
3. Εκπαίδευση με χρήση του κλασικού αλγορίθμου EM του νέου μοντέλου που προκύπτει μετά την εισαγωγή του νέου Γκαουσιανού πυρήνα.
4. Σε περίπτωση που δεν έχει επιτευχθεί ακόμα ο επιθυμητός αριθμός συνιστωσών στο εκπαιδευμένο μοντέλο, μετάβαση στο βήμα 2.

Περιγράφοντας και αναλύοντας με περισσότερο τυπικό τρόπο τα παραπάνω, έστω ότι έχουμε ένα εκπαιδευμένο μοντέλο  $k$  συνιστωσών, το οποίο συμβολίζουμε ως  $p_k(\mathbf{x})$ . Θεωρώντας ότι προσθέτουμε σε αυτό μια ακόμα Γκαουσιανή, έστω την  $\phi(\mathbf{x}|\theta)$ , με ένα βάρος  $a$ , τότε σχηματίζεται ένα νέο μείγμα  $k+1$  συνιστωσών:

$$p_{k+1}(\mathbf{x}) = (1-a)p_k(\mathbf{x}) + a\phi(\mathbf{x} | \boldsymbol{\theta}) \quad (2.43)$$

Ο στόχος μας είναι να βρούμε τις βέλτιστες τιμές με βάση την αρχή της μέγιστης πιθανοφάνειας για τις παραμέτρους  $a$  και  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Όπως αναφέρθηκε στην προηγούμενη ενότητα, ένα τέτοιο πρόβλημα μπορεί να αντιμετωπιστεί με χρήση του κλασικού αλγορίθμου EM. Η διαφορά που πρέπει να επισημανθεί εδώ είναι ότι στον άπληστο αλγόριθμο EM το μοντέλο που πρέπει να εκπαιδευθεί κάθε φορά μετά την ενσωμάτωση ενός νέου Γκαουσιανού πυρήνα έχει λιγότερους βαθμούς ελευθερίας από ό,τι το αντίστοιχο αριθμού πυρήνων μοντέλο που εκπαιδεύεται εξ αρχής μέσω του κλασικού αλγορίθμου EM. Αυτό ισχύει γιατί, όπως βλέπουμε από τη σχέση (2.43), το προηγούμενως εκπαιδευθέν μοντέλο  $k$  συνιστωσών χρησιμοποιείται αυτούσιο στο νέο μείγμα, και μόνο το βάρος μείξης και οι παράμετροι του νέου πυρήνα χρειάζεται να προσδιοριστούν. Στο βήμα 3, λοιπόν, του άπληστου αλγορίθμου EM το πρόβλημα που έχουμε να επιλύσουμε είναι σχετικά απλό, και αυτό αποτελεί πλεονέκτημα και από πλευράς ευρωστίας στον υπολογισμό των παραμέτρων αλλά και από άποψη χρόνου, μια και το εν λόγω πρόβλημα μπορεί να λυθεί σχετικά γρήγορα, πράγμα που επιταχύνει την όλη εκτέλεση του άπληστου αλγορίθμου EM.

Συγκεκριμένα, η χρήση του κλασικού αλγορίθμου EM για τον υπολογισμό των παραμέτρων  $a$ ,  $\boldsymbol{\theta}$  του νέου πυρήνα που εισάγεται στο μοντέλο οδηγεί στους τύπους ανανέωσης

$$E[z_{i,k+1}] = P(k+1 | \mathbf{x}_i) = \frac{a\phi(\mathbf{x}_i | \boldsymbol{\theta})}{(1-a)p_k(\mathbf{x}_i) + a\phi(\mathbf{x}_i | \boldsymbol{\theta})} \quad (2.44)$$

για το βήμα E, και

$$a = \frac{1}{N} \sum_{i=1}^N E[z_{i,k+1}] \quad (2.45)$$

$$\boldsymbol{\mu} = \frac{\sum_{i=1}^N E[z_{i,k+1}] \mathbf{x}_i}{\sum_{i=1}^N E[z_{i,k+1}]} \quad (2.46)$$

$$\boldsymbol{\Sigma} = \frac{\sum_{i=1}^N E[z_{i,k+1}] \mathbf{x}_i \mathbf{x}_i^T}{\sum_{i=1}^N E[z_{i,k+1}]} - \boldsymbol{\mu} \boldsymbol{\mu}^T \quad (2.47)$$

για το βήμα M. Αυτά τα βήματα, όπως αναλύθηκε παραπάνω για τον κλασικό αλγόριθμο EM, εκτελούνται διαδοχικά και επαναληπτικά μέχρι να επέλθει σύγκλιση, ενώ απαιτείται μια αρχική εκτίμηση των παραμέτρων ώστε να αρχίσει η όλη επαναληπτική διαδικασία. Στην προκειμένη περίπτωση, η αρχική εκτίμηση για τις παραμέτρους  $a$ ,  $\boldsymbol{\theta}$  που συνδέονται με τη νέα Γκαουσιανή συνιστώσα που εισάγεται στο μείγμα γίνεται στο βήμα 2 του άπληστου αλγορίθμου EM.

Συγκεκριμένα, για να οδηγηθεί σε μια αρχική εκτίμηση των τιμών των παραπάνω παραμέτρων, που αφορούν τη νεοεισαχθείσα Γκαουσιανή, ο άπληστος αλγόριθμος EM χρησιμοποιεί τη λογαριθμική πιθανοφάνεια των παρατηρήσεων, η οποία για το μοντέλο  $k$  συνιστωσών ορίζεται ως

$$LL_k = \ln \prod_{i=1}^N p_k(\mathbf{x}_i) = \sum_{i=1}^N \ln p_k(\mathbf{x}_i) \quad (2.48)$$

Επομένως, για το μοντέλο  $k+1$  συνιστωσών η λογαριθμική πιθανοφάνεια υπολογίζεται από τον τύπο

$$LL_{k+1} = \sum_{i=1}^N \ln p_{k+1}(\mathbf{x}_i) = \sum_{i=1}^N \ln[(1-a)p_k(\mathbf{x}_i) + a\phi(\mathbf{x}_i | \boldsymbol{\theta})] \quad (2.49)$$

Προφανώς, η  $LL_{k+1}$  είναι συνάρτηση του  $a$  (όπως και του  $\boldsymbol{\theta}$ ). Συνεπώς, μπορούμε να προσεγγίσουμε την παραπάνω συνάρτηση μέσω αναπτύγματος Taylor δευτέρας τάξεως γύρω από το σημείο  $a_0 = 0.5$ , και μετά να μεγιστοποιήσουμε την προσέγγιση αυτή ως προς  $a$ . Αν κάνουμε κάτι τέτοιο, προκύπτει εύκολα ότι ένα τοπικό μέγιστο κόντα στο  $a_0 = 0.5$  υπάρχει για

$$\hat{a} = \frac{1}{2} - \frac{1}{2} \frac{\sum_{i=1}^N \delta(\mathbf{x}_i, \boldsymbol{\theta})}{\sum_{i=1}^N \delta^2(\mathbf{x}_i, \boldsymbol{\theta})} \quad (2.50)$$

όπου

$$\delta(\mathbf{x}, \boldsymbol{\theta}) = \frac{p_k(\mathbf{x}) - \phi(\mathbf{x} | \boldsymbol{\theta})}{p_k(\mathbf{x}) + \phi(\mathbf{x} | \boldsymbol{\theta})} \quad (2.51)$$

και ισούται με

$$m = \sum_{i=1}^N \ln \frac{p_k(\mathbf{x}_i) + \phi(\mathbf{x}_i | \boldsymbol{\theta})}{2} + \frac{1}{2} \frac{\left[ \sum_{i=1}^N \delta(\mathbf{x}_i, \boldsymbol{\theta}) \right]^2}{\sum_{i=1}^N \delta^2(\mathbf{x}_i, \boldsymbol{\theta})} \quad (2.52)$$

Μπορεί, λοιπόν, να χρησιμοποιηθεί ως αρχική εκτίμηση για την παράμετρο  $a$  η τιμή  $\hat{a}$ . Αν η τελευταία βρίσκεται έξω από το διάστημα  $(0,1)$ , στο οποίο αναγκαστικά πρέπει να κινούνται οι τιμές του βάρους  $a$ , γίνεται αρχικοποίηση με  $\hat{a} = 0.5$ , για  $k=1$ , και με  $\hat{a} = 2/(k+1)$ , για  $k \geq 2$ .

Για να χρησιμοποιηθεί η σχέση (2.50) για τον υπολογισμό του  $\hat{a}$  πρέπει πρώτα να έχουν καθοριστεί οι αρχικές τιμές για τις παραμέτρους  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , αφού η παραπάνω σχέση ορίζει το  $\hat{a}$  ως συνάρτηση του  $\boldsymbol{\theta}$ . Για την παράμετρο  $\boldsymbol{\Sigma}$  χρησιμοποιείται μια απλή αρχική εκτίμηση, και συγκεκριμένα:

$$\Sigma = \sigma^2 \mathbf{I} \quad (2.53)$$

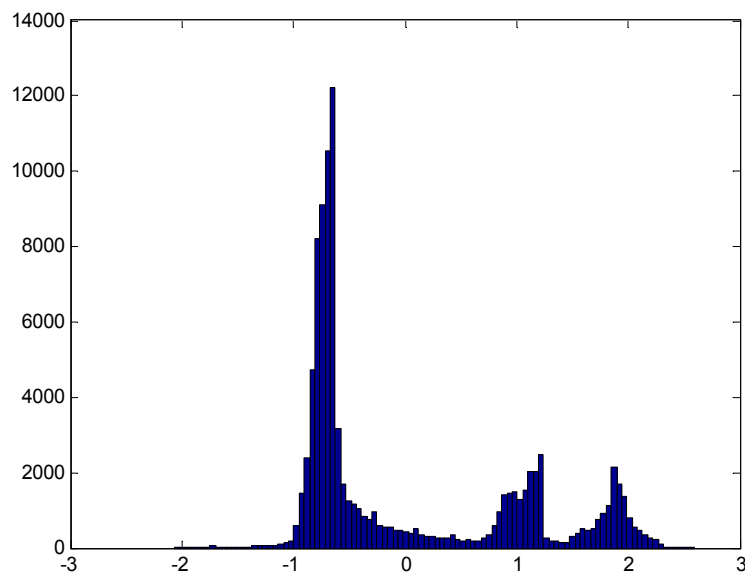
$$\sigma = \beta \left[ \frac{4}{(d+2)N} \right]^{1/(d+4)} \quad (2.54)$$

όπου  $\mathbf{I}$  είναι ο αντίστοιχος μοναδιαίος πίνακας,  $d$  η διάσταση των παρατηρήσεων εκπαίδευσης και  $N$  ο αριθμός τους, ενώ  $\beta$  είναι μια σταθερά. Η τιμή που επιλέχθηκε για την παράμετρο  $\sigma$  είναι ανάλογη με εκείνη που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα στην περίπτωση των μη παραμετρικών (nonparametric) εκτιμητών πυκνότητας πιθανότητας.

Όσο αφορά τον υπολογισμό της αρχικής εκτίμησης του  $\mu$ , παρατηρούμε από τη σχέση (2.52) για το  $m$  ότι για τον καθορισμό της τιμής του τελευταίου μόνο οι παράμετροι  $\theta$  μας χρειάζονται, αφού οι υπόλοιπες ποσότητες είναι ήδη γνωστές. Μάλιστα, με δεδομένη την παραπάνω αρχική τιμή για το  $\Sigma$ , η τιμή του  $m$  εξαρτάται μόνο από το  $\mu$ . Περιορίζοντας λοιπόν την αναζήτησή μας για την καλύτερη τιμή σε σχέση με την παράμετρο  $\mu$  στις παρατηρήσεις  $x_i$ , μπορεί να επιλεγθεί για την αρχικοποίηση της συγκεκριμένης παραμέτρου εκείνη η παρατήρηση που μας δίνει τη μεγαλύτερη τιμή για το  $m$ .

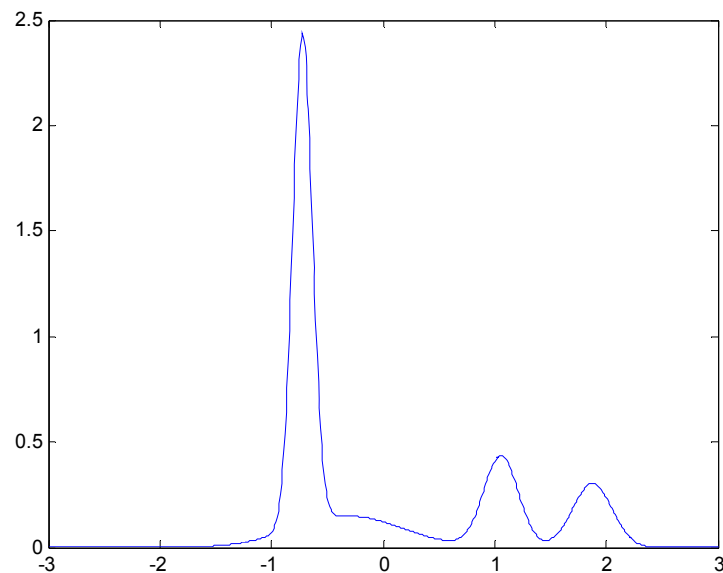
Κλείνοντας, υπενθυμίζουμε ότι άπληστος αλγόριθμος EM ξεκινάει με μια Γκαουσιανή, για την εκτίμηση των παραμέτρων της οποίας δεν απαιτείται προφανώς η χρήση του αλγορίθμου EM. Οι παράμετροί της μπορούν να υπολογιστούν απευθείας από τους τύπους (2.25) και (2.26), που ορίζουν τις παραμέτρους μέγιστης πιθανοφάνειας στην περίπτωση της απλής Γκαουσιανής κατανομής.

### Παράδειγμα Μοντελοποίησης Συνόλου Παρατηρήσεων Μέσω Γκαουσιανού Μείγματος



Σχήμα 2.2: Ιστόγραμμα με 100 bins για ένα σύνολο παρατηρήσεων

Για να παρουσιάσουμε με εποπτικό τρόπο το πως ένα σύνολο παρατηρήσεων μοντελοποιείται μέσω ενός Γκαουσιανού Μείγματος, παραθέτουμε τα σχήματα 2.2 και 2.3. Στο σχήμα 2.2 βλέπουμε το Ιστόγραμμα που προκύπτει από ένα μεγάλο πλήθος μονοδιάστατων παρατηρήσεων που αντιστοιχούν στις τιμές της φωτεινότητας για τα pixels μιας γκριζας εικόνας. Συγκεκριμένα, για τα bins του Ιστογράμματος απεικονίζεται ο αριθμός των παρατηρήσεων και επομένως των pixels που τους αντιστοιχεί, δηλαδή οι τιμές  $c_j$ , όπως αυτές ορίστηκαν στην ενότητα 2.2. Για το Ιστόγραμμα έχει χρησιμοποιηθεί κβαντοποίηση σε 100 bins. Το σχήμα 2.3, από την άλλη, απεικονίζει τη συνάρτηση πυκνότητας πιθανότητας του Γκαουσιανού Μείγματος 4 συνιστωσών που μας δίνει η εφαρμογή του άπληστου αλγορίθμου EM πάνω στο σύνολο των παραπάνω παρατηρήσεων. Από την ομοιότητα των σχημάτων γίνεται και διαισθητικά αντιληπτό αυτό που επιχειρούμε να πράξουμε μέσω των Γκαουσιανών Μειγμάτων, δηλαδή, από ένα πεπερασμένο σύνολο παρατηρήσεων να συνθέσουμε ένα συνεχές μοντέλο που να περιγράφει όσο γίνεται καλύτερα την κατανομή τους.



**Σχήμα 2.3:** Γκαουσιανό Μείγμα 4 συνιστωσών, εκπαιδευμένο μέσω των ίδιων παρατηρήσεων που χρησιμοποιήθηκαν και για το σχήμα 2.2

#### 2.4.4 Προσδιορισμός Αριθμού Συνιστωσών

Σε όλη την προηγούμενη ανάλυση και για τον κλασικό αλγόριθμο EM αλλά και για την παραλλαγή του, τον άπληστο αλγόριθμο EM, μια παράμετρος για την οποία δε μιλήσαμε πολύ είναι αυτή του αριθμού των συνιστωσών του Γκαουσιανού Μείγματος. Η συγκεκριμένη παράμετρος ουσιαστικά καθορίζει την πολυπλοκότητα του μοντέλου, αφού όσο πιο πολλές Γκαουσιανές συνιστώσες έχει ένα Γκαουσιανό Μείγμα, τόσο πιο πολύπλοκο μπορεί να θεωρηθεί. Αξίζει λοιπόν να ασχοληθούμε συνοπτικά με το ρόλο που παίζει η εν λόγω παράμετρος και με το πως θα μπορούσε να καθοριστεί η τιμή της. Για επιπλέον πληροφορίες πάνω σε αυτό το θέμα, ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει παράλληλα στις εργασίες [50], [51] και [52].



Σε κάθε πρόβλημα μοντελοποίησης ενός συνόλου δεδομένων από κάποιο κατάλληλο είδος μοντέλου, ένα από τα ζητήματα που τίθενται είναι το πόσο πολύπλοκο πρέπει να είναι το μοντέλο που θα χρησιμοποιηθεί. Η πολυπλοκότητα συνήθως ορίζεται από τον αριθμό των ανεξάρτητων παραμέτρων που απαιτούνται για τον πλήρη καθορισμό του μοντέλου. Το θέμα αυτό είναι σημαντικό, γιατί αν ένα μοντέλο με δυσανάλογα μεγάλη πολυπλοκότητα σε σχέση με τα δεδομένα επιλεγεί, τότε, εκτός της περιττής επιβάρυνσης που είναι πιθανόν να προκύψει στο χρόνο εκπαίδευσής του, μπορεί η τελευταία να οδηγήσει στη μοντελοποίηση επουσιωδών λεπτομερειών των δεδομένων, και συγκεκριμένα του πάντα υπάρχοντος σε πραγματικά δεδομένα θορύβου, ένα γεγονός που δεν είναι επιθυμητό. Αντίθετα, ένα πολύ απλό μοντέλο μπορεί να αδυνατεί να περιγράψει βασικές δομές των δεδομένων και, επομένως, αν χρησιμοποιηθεί να καταλήξει σε μια πολύ αδρή και μη ικανοποιητική περιγραφή για τα δεδομένα, με σοβαρές επιπτώσεις από άποψη απόδοσης των συστημάτων που θα τη χρησιμοποιήσουν.

Εν γένει, όσο πιο πολύπλοκο είναι το μοντέλο που χρησιμοποιούμε, τόσο πιο καλά θα περιγράψει μετά την εκπαίδευση τα δεδομένα που χρησιμοποιήθηκαν για την τελευταία. Αυτό στην περίπτωση των Γκαουσιανών Μειγμάτων, για παράδειγμα, μεταφράζεται σε μια διαρκώς αυξανόμενη τιμή για την πιθανοφάνεια των παρατηρήσεων, όπως αυτή υπολογίζεται για το εκάστοτε εκπαιδευμένο μοντέλο, καθώς ο αριθμός,  $k$ , των συνιστωσών που περιέχει το μοντέλο αυξάνεται. Όμως, κάτι τέτοιο δε σημαίνει καθόλου ότι όσο πολυπλοκότερο είναι το μοντέλο τόσο καταλληλότερο είναι για την περιγραφή των δεδομένων μας. Και αυτό γιατί τα δεδομένα που χρησιμοποιούμε για την εκπαίδευση δεν αποτελούν παρά ένα μικρό υποσύνολο του συνόλου των δεδομένων που θα μπορούσαν να προκύψουν για το ίδιο πρόβλημα. Και μάλιστα ένα υποσύνολο ευάλωτο στο θόρυβο, στα δεδομένα outliers κτλ., όπως αναφέραμε και παραπάνω. Η υπερβολική πολυπλοκότητα του μοντέλου, λοιπόν, οδηγεί στην υπερβολική προσαρμογή του στο συγκεκριμένο υποσύνολο εκπαίδευσης, με νομοτελειακό αποτέλεσμα την επακόλουθη αδυναμία του να περιγράψει ικανοποιητικά κάποιο άλλο ομοειδές σύνολο δεδομένων, διαφορετικό από αυτό με το οποίο έχει εκπαιδευτεί. Το φαινόμενο αυτό λέγεται υπερεκπαίδευση (overtraining) και πρέπει να αποφεύγεται, εάν θέλουμε τα μοντέλα τα οποία εκπαιδεύονται αναγκαστικά σε ένα περιορισμένο σύνολο δεδομένων, να μπορούν να εμφανίζουν την ιδιότητα της γενίκευσης για δεδομένα για τα οποία δεν έχουν εκπαιδευτεί.

Επομένως, όπως διαφαίνεται από την προηγηθείσα ανάλυση, αυτό που αναζητάμε κάθε φορά είναι ένα μοντέλο αρκετά «πλούσιο», ώστε να μπορεί να περιγράψει τις βασικές δομές των δεδομένων προς μοντελοποίηση, αλλά όχι υπερβολικά πολύπλοκο, ώστε να μην κινδυνεύει να εξειδικευθεί υπερβολικά και να μην ενσωματώνει τις επουσιώδεις λεπτομέρειες που είναι πιθανόν να οφείλονται σε θόρυβο, σε στατιστικές ιδιότητες λόγω του πεπερασμένου συνόλου εκπαίδευσης κτλ. Το πρόβλημα αυτό δεν είναι καθόλου εύκολο, και για την αντιμετώπισή του έχει καταβληθεί μεγάλη ερευνητική προσπάθεια.

Συνήθως, οι μέθοδοι που χρησιμοποιούνται προκειμένου να προσδιοριστεί η κατάλληλη πολυπλοκότητα για ένα μοντέλο που καλείται να μοντελοποιήσει κάποιο συγκεκριμένο σύνολο δεδομένων, στηρίζονται στον ορισμό κάποιου κριτηρίου αξιολόγησης μοντέλων. Το κριτήριο αυτό θα εφαρμοστεί σε ένα σύνολο μοντέλων με διαφορετική πολυπλοκότητα το καθένα, εκπαιδευμένων πάνω στα προαναφερθέντα δεδομένα, ώστε να αναδειχθεί με αυτόν τον τρόπο ένα από αυτά τα μοντέλα ως το καλύτερο. Ένα κριτήριο αυτού του είδους ορίζεται τις περισσότερες φορές ως το άθροισμα δύο όρων. Ο πρώτος από αυτούς τους δύο όρους (TT) εκφράζει το βαθμό της προσαρμογής του μοντέλου στα δεδομένα εκπαίδευσης, καλείται όρος εκπαίδευσης και ορίζεται έτσι ώστε να φθίνει καθώς η πολυπλοκότητα του μοντέλου αυξάνει. Αντίθετα, ο δεύτερος όρος (PT) αντιπροσωπεύει ένα μέτρο της πολυπλοκότητας του μοντέλου, καλείται όρος ποινής

και είναι πάντα αύξουσα συνάρτηση της πολυπλοκότητας. Επομένως, τα κριτήρια για τα οποία μιλάμε έχουν συνήθως τη μορφή

$$CSC = TT + PT \quad (2.55)$$

Ο παραπάνω ορισμός γεννάει την ελπίδα ότι για κάποια συγκεκριμένη πολυπλοκότητα θα υπάρξει ένα ελάχιστο για το κριτήριο αυτό. Για πολυπλοκότητες μικρότερες από αυτήν που δίνει την ελάχιστη τιμή, ο πρώτος όρος θα φθίνει πιο γρήγορα από ό,τι θα αυξάνει ο δεύτερος, πράγμα που θα αναγκάζει τη συνολική τιμή του κριτηρίου να μειώνεται και θα αποτελεί ένδειξη ότι η μοντελοποίηση των κύριων δομών που ενυπάρχουν στα δεδομένα δεν έχει ολοκληρωθεί ακόμα, αφού ο ρυθμός μείωσης του πρώτου όρου είναι σχετικά μεγάλος. Αντίθετα, για πολυπλοκότητες μεγαλύτερες από εκείνη που αντιστοιχεί στην ελάχιστη τιμή του κριτηρίου, ο ρυθμός αύξησης του δεύτερου όρου αναμένεται να υπερνικά το ρυθμό μείωσης του πρώτου όρου και, επομένως, η συνολική τιμή του κριτηρίου θα αυξάνει. Κάτι τέτοιο εκλαμβάνεται ως ένδειξη ότι έχει ξεκινήσει η υπερεκπαίδευση, αφού τα «κέρδη» που έχουμε από την περαιτέρω προσαρμογή στα δεδομένα είναι μικρά σε σχέση με τη «ζημία» από τη ραγδαία αύξηση της πολυπλοκότητας. Επομένως, τελικά, το μοντέλο με εκείνη την πολυπλοκότητα που δίνει την ελάχιστη τιμή για το κριτήριο επιλέγεται ως το καταλληλότερο μοντέλο περιγραφής των δεδομένων.

Παρακάτω αναφερόμαστε συνοπτικά σε δύο από τα πιο δημοφιλή κριτήρια επιλογής μοντέλου, που ακολουθούν τη φιλοσοφία που αναλύσαμε προηγουμένως.

### Κριτήριο MDL

Το πιο δημοφιλές κριτήριο αυτού το είδους είναι το κριτήριο MDL (Minimum Description Length – Ελάχιστο Μήκος Περιγραφής), το οποίο στηρίζεται σε μια προσέγγιση βασισμένη σε θεωρήσεις σχετικές με την κωδικοποίηση της πληροφορίας (information coding). Στην ουσία, μέσω του κριτηρίου MDL, το πρόβλημα της μοντελοποίησης προσεγγίζεται από τη σκοπιά της συμπίεσης δεδομένων (data compression). Σε αυτό το πλαίσιο, ως καλύτερο μοντέλο επιλέγεται εκείνο που περιγράφει τα δεδομένα εκπαίδευσης με τέτοιο τρόπο ώστε να μπορεί να προκληθεί μέγιστη συμπίεση της πληροφορίας που περιέχουν.

Στην περίπτωση των Γκαουσιανών Μειγμάτων που μας απασχολεί, για τον υπολογισμό του κριτηρίου MDL χρησιμοποιείται ο τύπος

$$MDL(k) = -\ln \mathcal{L}(\Theta_k) + \frac{n_k}{2} \ln N \quad (2.56)$$

όπου  $\ln \mathcal{L}(\Theta_k)$  είναι η λογαριθμική πιθανοφάνεια των παρατηρήσεων για το εκπαιδευμένο μοντέλο  $k$  Γκαουσιανών και  $n_k$  το πλήθος των ανεξάρτητων παραμέτρων στο εν λόγω μοντέλο. Με  $N$  συμβολίζουμε και εδώ τον αριθμό των παρατηρήσεων που χρησιμοποιούνται στην εκπαίδευση. Ο πρώτος όρος στο δεξιό μέλος της σχέσης (2.56) αντιστοιχεί στον όρο εκπαίδευσης, ενώ ο δεύτερος όρος στον όρο ποινής, όπως αυτοί αναλύθηκαν θεωρητικά μόλις προηγουμένως. Για ένα Γκαουσιανό Μείγμα  $k$  συνιστωσών, λαμβάνοντας υπόψιν ότι κάθε πίνακας συνδιακύμανσης  $\Sigma_j$  είναι υποχρεωτικά συμμετρικός, και επομένως οι τιμές του κάτω από τη διαγώνιο είναι ίσες με αυτές πάνω από τη διαγώνιο, και ότι τα βάρη  $\pi_j$  αθροίζουν στη μονάδα, με αποτέλεσμα το τελευταίο, για παράδειγμα, να μπορεί να γραφεί ως συνάρτηση των  $k-1$  πρώτων, μπορούμε να

υπολογίσουμε την τιμή του  $n_k$  για το πλήθος των ανεξάρτητων παραμέτρων του μοντέλου βάσει του τύπου

$$n_k = \left[ 1 + d + \frac{d(d+1)}{2} \right] k - 1 \quad (2.57)$$

όπου  $d$  η διάσταση των παρατηρήσεων.

## Κριτήριο AIC

Ένα άλλο δημοφιλές κριτήριο επιλογής μοντέλου είναι το κριτήριο AIC (Akaike's Information Criterion – Κριτήριο Πληροφορίας του Akaike), το οποίο, αντίθετα από το MDL, στηρίζεται πάνω στην έννοια της εντροπίας (entropy), που είναι ένα μέτρο αβεβαιότητας σχετιζόμενο με κάποια τυχαία μεταβλητή, και τη χρησιμοποιεί για την προσέγγιση της απόστασης του εκάστοτε εκπαιδευμένου μοντέλου από το άγνωστο πραγματικό μοντέλο που παρήγαγε τα δεδομένα. Για την εντροπία καθώς και για αποστάσεις μεταξύ πιθανοτικών μοντέλων θα μιλήσουμε αναλυτικότερα στο επόμενο κεφάλαιο. Εδώ, θα αρκεστούμε στο να παραθέσουμε τη σχέση ορισμού του κριτηρίου AIC για την περίπτωση μας:

$$\text{AIC}(k) = -2 \ln \mathcal{L}(\Theta_k) + 2n_k \quad (2.58)$$

με την τιμή του  $n_k$  να υπολογίζεται πάλι από τον τύπο (2.57). Γίνεται αμέσως αντιληπτό με μια απλή επισκόπηση του τύπου ορισμού του ότι και αυτό το κριτήριο ακολουθεί ακριβώς τη λογική που έχει αναλυθεί εκτενώς στις προηγούμενες παραγράφους.

Εκτός από τα προηγούμενα, πολλά άλλα κριτήρια αυτού του είδους έχουν προταθεί. Στα πλαίσια αυτής της διατριβής δεν είναι σκοπός μας η εξαντλητική παρουσίαση όλων των δυνατών τρόπων και προτεινόμενων μεθόδων επιλογής μοντέλου (model selection). Αυτό που πρέπει να αναφέρουμε είναι ότι όλα τα κριτήρια αυτού του είδους για να λειτουργήσουν κατά τον επιθυμητό τρόπο, δηλαδή για να δώσουν μια εκτίμηση σχετικά με το ποια είναι η κατάλληλη πολυπλοκότητα για ένα μοντέλο που περιγράφει ένα συγκεκριμένο σύνολο δεδομένων, χρειάζονται ένα σύνολο εκπαιδευμένων μοντέλων πάνω στα εν λόγω δεδομένα. Τα μοντέλα αυτού του συνόλου πρέπει να έχουν διαφορετική πολυπλοκότητα το καθένα. Τότε, ο υπολογισμός της τιμής του χρησιμοποιούμενου κριτηρίου για όλα τα μοντέλα και η επιλογή εκείνου που την ελαχιστοποιεί μας οδηγούν, άμεσα, σε ένα προτεινόμενο από το κριτήριο μοντέλο και, έμμεσα, σε μια εκτίμηση για τη βέλτιστη πολυπλοκότητα μοντέλου.

Με βάση τα παραπάνω αποκαλύπτεται άλλο ένα πλεονέκτημα του άπληστου αλγορίθμου EM για την εκπαίδευση των Γκαουσιανών Μειγμάτων. Ο συγκεκριμένος αλγόριθμος, ξεκινώντας από μια Γκαουσιανή και μέσω της διαδοχικής προσθήκης νέων πυρήνων, κατά την εκτέλεση του δημιουργεί και εκπαιδεύει όλα τα Γκαουσιανά Μείγματα για τα δεδομένα εισόδου του, από εκείνο που περιλαμβάνει μόνο μια Γκαουσιανή έως και το τελευταίο, που περιλαμβάνει  $K_{\max}$  Γκαουσιανές συνιστώσες. Επομένως, μετά την εκτέλεση του άπληστου αλγορίθμου EM και έχοντας αυτά τα  $K_{\max}$  Γκαουσιανά Μείγματα, που έχουν όλα εκπαιδευτεί έτσι ώστε να περιγράφουν τις παρατηρήσεις, μπορεί αμέσως να χρησιμοποιηθεί το όποιο κριτήριο της παραπάνω μορφής για να επιλεγεί το καλύτερο

μοντέλο για τα συγκεκριμένα δεδομένα, με την προϋπόθεση βέβαια ότι το  $K_{\max}$  είναι αρκετά μεγάλο ώστε να αντιστοιχεί στη μέγιστη πολυπλοκότητα που μας ενδιαφέρει.

Τέλος, πρέπει να επισημάνουμε ότι το πρόβλημα του προσδιορισμού της βέλτιστης πολυπλοκότητας για ένα μοντέλο που περιγράφει κάποια δεδομένα είναι από θεωρητικής άποψης ιδιαίτερα δύσκολο, και καμιά λύση που έχει προταθεί δε μπορεί να θεωρηθεί τελική. Μιλάμε επομένως για ένα ανοικτό πρόβλημα που δεν έχει ακόμα λυθεί ικανοποιητικά. Αυτό έγινε εμφανές και κατά την εκτέλεση των πειραμάτων που παρουσιάζονται στα επόμενα κεφάλαια, οπότε και ελέγχθηκαν διάφορα κριτήρια, συμπεριλαμβανομένων και αυτών που παρουσιάστηκαν παραπάνω. Τα συμπεράσματα που προέκυψαν από τη χρήση τους είναι αρνητικά, με την έννοια ότι η εφαρμογή τους οδηγούσε σταθερά στην επιλογή υπερβολικά πολύπλοκων Γκαουσιανών Μειγμάτων, με μη αποδεκτά μεγάλο αριθμό συνιστωσών.

## Κεφάλαιο 3

# Μέτρα Απόστασης Μεταξύ Γκαουσιανών Μειγμάτων

Όπως έχουμε ήδη αναφέρει στο κεφάλαιο 1, είναι ιδιαίτερα χρήσιμο να μπορούμε, έχοντας δύο συγκεκριμένα Γκαουσιανά Μείγματα, να εκτιμήσουμε το πόσο κοντά ή μακριά βρίσκονται το ένα από το άλλο. Να μπορούμε, με άλλα λόγια, να ορίσουμε και να υπολογίσουμε ένα μέτρο απόστασης μεταξύ τους. Δυστυχώς, κάτι τέτοιο δεν είναι τόσο εύκολο, αφού, πέρα από τις καλές ιδιότητες που απαιτούνται για ένα τέτοιο μέτρο όσο αφορά το διαχωρισμό των διαφορετικών μοντέλων, μια άλλη εξίσου σημαντική απαίτηση είναι το να μπορεί να υπολογιστεί η όποια απόσταση σχετικά γρήγορα. Το τελευταίο μεταφράζεται στην πράξη σε μια απαίτηση για δυνατότητα υπολογισμού της απόστασης σε κλειστή μορφή. Να υπάρχει, δηλαδή, κάποιος κλειστός τύπος, ώστε εφαρμοζόμενος για δύο οποιαδήποτε Γκαουσιανά Μείγματα να μπορεί να μας δώσει την απόσταση που τα χωρίζει. Αυτό στην περίπτωση των Γκαουσιανών Μειγμάτων αποδεικνύεται μια μάλλον δύσκολη απαίτηση, όπως θα δούμε παρακάτω.

Σε αυτό το κεφάλαιο παρουσιάζονται μερικά από τα σημαντικότερα μέτρα απόστασης μεταξύ Γκαουσιανών Μειγμάτων, και αναλύονται οι ενδιαφέρουσες ιδιότητές τους. Για την ανάλυση που ακολουθεί θα υποθέσουμε ότι έχουμε δύο Γκαουσιανά Μείγματα:

$$p_1(\mathbf{x}) = \sum_{j=1}^{K_1} \pi_{1j} \phi(\mathbf{x} | \boldsymbol{\theta}_{1j}) \quad \text{και} \quad p_2(\mathbf{x}) = \sum_{j=1}^{K_2} \pi_{2j} \phi(\mathbf{x} | \boldsymbol{\theta}_{2j}) \quad (3.1)$$

όπου, προφανώς,  $\pi_{ij}$ ,  $\boldsymbol{\theta}_{ij} = (\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij})$  και  $K_i$  είναι το βάρος της  $j$ -οστής συνιστώσας του  $i$ -οστού μείγματος, η μέση τιμή και ο πίνακας συνδιακύμανσης της  $j$ -οστής συνιστώσας του  $i$ -οστού μείγματος και ο αριθμός των συνιστωσών του  $i$ -οστού μείγματος, αντίστοιχα. Επίσης,  $\phi(\mathbf{x} | \boldsymbol{\theta}_{ij}) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij})$ , αφού αναφερόμαστε σε Γκαουσιανά Μείγματα.

### 3.1 Απόκλιση KL

Τα Γκαουσιανά Μείγματα, όπως αναφέραμε εξ αρχής, ανήκουν στην κατηγορία των πιθανοτικών μοντέλων και αποτελούν, στην ουσία, μοντέλα συναρτήσεων πυκνότητας πιθανότητας. Για τέτοιου είδους μοντέλα, το πιο διαδεδομένο μέτρο απόστασης είναι η απόκλιση KL (Kullback-Leibler divergence) [53]. Το μέτρο αυτό χαρακτηρίζεται ως

απόκλιση, γιατί δεν ικανοποιεί όλες τις ιδιότητες της μετρικής, όπως αυτές ορίζονται από τις σχέσεις (1.10)-(1.13).

Η απόκλιση KL για δύο συναρτήσεις πυκνότητας πιθανότητας  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$  δίνεται από τον κάτωθι τύπο:

$$\text{KL}(p_1 \parallel p_2) = \int p_1(\mathbf{x}) \ln \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} d\mathbf{x} \quad (3.2)$$

Στην περίπτωση που τα όρια ολοκλήρωσης παραλείπονται, όπως στην προηγούμενη σχέση, υπονοείται ότι η ολοκλήρωση γίνεται σε όλο το πεδίο ορισμού των αντίστοιχων συναρτήσεων.

Από τον παραπάνω ορισμό βλέπουμε αμέσως ότι η απόκλιση KL δεν είναι συμμετρική, δηλαδή ισχύει ότι

$$\text{KL}(p_1 \parallel p_2) \neq \text{KL}(p_2 \parallel p_1) \quad (3.3)$$

Παρόλα αυτά, μπορούμε εύκολα να ορίσουμε μια συμμετρική εκδοχή της, την SKL, μέσω της παρακάτω σχέσης:

$$\text{SKL}(p_1, p_2) = \frac{1}{2} \text{KL}(p_1 \parallel p_2) + \frac{1}{2} \text{KL}(p_2 \parallel p_1) \quad (3.4)$$

για την οποία ισχύει εμφανώς ότι

$$\text{SKL}(p_1, p_2) = \text{SKL}(p_2, p_1) \quad (3.5)$$

Η απόκλιση KL μπορεί να υπολογιστεί σε κλειστή μορφή στην περίπτωση που οι δύο συγκρινόμενες συναρτήσεις αντιστοιχούν σε Γκαουσιανές κατανομές. Συγκεκριμένα, αν

$$p_1(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \quad \text{και} \quad p_2(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \quad (3.6)$$

τότε η απόκλιση KL μεταξύ τους μπορεί να υπολογιστεί από τον εξής τύπο:

$$\text{KL}(p_1 \parallel p_2) = \frac{1}{2} \left[ \text{tr}(\boldsymbol{\Sigma}_2^{-1} \boldsymbol{\Sigma}_1) - \ln \frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|} - d \right] + \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (3.7)$$

όπου με  $d$  συμβολίζουμε και εδώ τη διάσταση του  $\mathbf{x}$ . Επίσης, με  $\text{tr}(\mathbf{A})$  συμβολίζουμε το άθροισμα των διαγώνιων στοιχείων του πίνακα  $\mathbf{A}$ .

### 3.1.1 Υπολογισμός Μέσω Τυχαίας Δειγματοληψίας

Δυστυχώς, όμως, παρά το ότι, όπως είδαμε αμέσως παραπάνω, η απόκλιση KL μπορεί να υπολογιστεί σε κλειστή μορφή, και επομένως αποδοτικά, για δύο Γκαουσιανές κατανομές, όταν έχουμε δύο Γκαουσιανά Μείγματα τα πράγματα είναι διαφορετικά. Στην περίπτωση των Γκαουσιανών Μειγμάτων η απόκλιση KL δεν μπορεί να υπολογιστεί σε κλειστή μορφή, λόγω του ότι το ολοκλήρωμα που την ορίζει δε μπορεί να υπολογιστεί αναλυτικά. Αυτό το γεγονός γεννάει πολλά προβλήματα όσο αφορά τον υπολογισμό της

αποκλίσεως KL μεταξύ Γκαουσιανών Μειγμάτων. Ένας τρόπος που έχει προταθεί για να καταστεί δυνατός ο υπολογισμός αυτός, είναι η χρήση τεχνικών τυχαίας δειγματοληψίας (random sampling) Monte Carlo.

Εδώ θα παρουσιάσουμε συνοπτικά το πως μπορεί η απόκλιση KL να υπολογιστεί μέσω τυχαίας δειγματοληψίας. Συγκεκριμένα, χρησιμοποιώντας τον ορισμό για την προσδοκώμενη τιμή μιας ποσότητας  $f(\mathbf{X})$ , που εξαρτάται από μια συνεχή τυχαία μεταβλητή  $\mathbf{X}$  κατανεμημένη σύμφωνα με την κατανομή  $p(\mathbf{x})$ , ο οποίος δίνεται από τη σχέση

$$E[f(\mathbf{X})]_p = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \quad (3.8)$$

και εξετάζοντας τη σχέση (3.2), που ορίζει την απόκλιση KL, βλέπουμε ότι αυτή μπορεί να διατυπωθεί σε μια ισοδύναμη μορφή:

$$KL(p_1 \parallel p_2) = E \left[ \ln \frac{p_1(\mathbf{X})}{p_2(\mathbf{X})} \right]_{p_1} \quad (3.9)$$

Η παραπάνω διατύπωση μας πληροφορεί ότι η απόκλιση KL μεταξύ των  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$  είναι ίση με την προσδοκώμενη τιμή της τυχαίας μεταβλητής  $\ln \frac{p_1(\mathbf{X})}{p_2(\mathbf{X})}$ , θεωρώντας ότι η τυχαία μεταβλητή  $\mathbf{X}$  κατανέμεται με βάση την  $p_1(\mathbf{x})$ .

Οι μέθοδοι Monte Carlo, τώρα, στηρίζονται στο γνωστό από τη Στατιστική «νόμο των μεγάλων αριθμών», και υπολογίζουν την απόκλιση KL προσεγγίζοντας την παραπάνω προσδοκώμενη τιμή μέσω της στατιστικής μέσης τιμής. Δηλαδή, η απόκλιση KL υπολογίζεται προσεγγιστικά ως εξής:

$$KL(p_1 \parallel p_2) \approx \frac{1}{N} \sum_{i=1}^N \ln \frac{p_1(\mathbf{x}_i)}{p_2(\mathbf{x}_i)} \quad (3.10)$$

όπου τα  $\mathbf{x}_i$  πρέπει να αποτελούν τυχαία στιγμιότυπα λαμβανόμενα από την κατανομή  $p_1(\mathbf{x})$ . Όσο μεγαλύτερος είναι ο αριθμός,  $N$ , των τυχαίων δειγμάτων  $\mathbf{x}_i$ , τόσο μεγαλύτερη θα είναι η ακρίβεια της προσέγγισης.

Δυστυχώς, αυτός ο τρόπος υπολογισμού της απόκλισης KL είναι υπερβολικά χρονοβόρος, και αυτό γιατί για να προσεγγίσουμε την πραγματική τιμή με καλή ακρίβεια χρειάζεται συνήθως ένας μεγάλος αριθμός τυχαίων δειγμάτων από το ένα από τα δύο Γκαουσιανά Μείγματα.

### 3.2 Απόσταση EMD

Οι σημαντικές δυσκολίες στον υπολογισμό της απόκλισης KL, όπως αυτές παρουσιάστηκαν παραπάνω, μας οδηγούν στο να ψάξουμε για άλλες αποστάσεις μεταξύ Γκαουσιανών Μειγμάτων, με την ελπίδα ότι αυτές θα μπορούν να υπολογιστούν αποδοτικότερα. Μια απόσταση μεταξύ σύνθετων κατανομών που έχει χρησιμοποιηθεί σε διάφορα προβλήματα με ενθαρρυντικά αποτελέσματα είναι η απόσταση EMD (Earth Mover's Distance) [17]. Μεταξύ δύο κατανομών, η συγκεκριμένη απόσταση θεωρεί τη

μάζα της μιας ως σωρούς χρώματος κατανεμημένους στο χώρο ορισμού της και τη μάζα της άλλης ως λακκούβες στον ίδιο χώρο. Θέτοντας έτσι το πρόβλημα, ο υπολογισμός της EMD έγκειται στην εύρεση του ελάχιστου έργου που είναι απαραίτητο για να γεμίσουν οι λακκούβες με το χρώμα από τους σωρούς.

Θεωρώντας τα Γκαουσιανά Μείγματα  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$ , όπως αυτά ορίστηκαν από την (3.1), η απόσταση EMD υπολογίζεται από τον παρακάτω τύπο:

$$\text{EMD}(p_1, p_2) = \frac{\sum_{i=1}^{K_1} \sum_{j=1}^{K_2} f_{ij} d_{\text{ground}}(\phi(\mathbf{x} | \theta_{1i}), \phi(\mathbf{x} | \theta_{2j}))}{\sum_{i=1}^{K_1} \sum_{j=1}^{K_2} f_{ij}} \quad (3.11)$$

με τα βάρη  $f_{ij} \geq 0$  να επιλέγονται έτσι ώστε να ελαχιστοποιείται ο αριθμητής του δεξιού μέλους, υπό τους περιορισμούς

$$\sum_{j=1}^{K_2} f_{ij} \leq \pi_{1i}, \quad \sum_{i=1}^{K_1} f_{ij} \leq \pi_{2j}, \quad \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} f_{ij} = \min\left(\sum_{i=1}^{K_1} \pi_{1i}, \sum_{j=1}^{K_2} \pi_{2j}\right) = 1 \quad (3.12)$$

Είναι γνωστό ότι το έργο που καταναλώνεται κατά τη μεταφορά χρώματος από ένα σημείο σε ένα άλλο είναι ανάλογο της μάζας του χρώματος και της απόστασης των δύο σημείων. Ουσιαστικά, λοιπόν, καθένα από τα  $f_{ij}$  εκφράζει το πόσο «χρώμα» από το «σωρό» που αντιστοιχεί στην  $i$ -οστή συνιστώσα του πρώτου μείγματος θα μεταφερθεί για να καλύψει τη «λακκούβα» που αντιστοιχεί στην  $j$ -οστή συνιστώσα του δεύτερου μείγματος. Αντίστοιχα, η  $d_{\text{ground}}(\phi(\mathbf{x} | \theta_{1i}), \phi(\mathbf{x} | \theta_{2j}))$  είναι η απόσταση ανάμεσα στον προαναφερθέντα «σωρό» και την προαναφερθείσα «λακκούβα». Στην πράξη, βέβαια, η  $d_{\text{ground}}$  είναι μια απόσταση μεταξύ δύο Γκαουσιανών κατανομών. Αλλά, η εύρεση μιας κατάλληλης απόστασης για Γκαουσιανές κατανομές δεν αποτελεί πρόβλημα, αφού μόλις είδαμε παραπάνω ότι η απόκλιση KL, για παράδειγμα, μπορεί να υπολογιστεί σε κλειστή μορφή όταν οι κατανομές που συγκρίνονται είναι Γκαουσιανές. Το ίδιο συμβαίνει και για ένα πλήθος άλλων αποστάσεων, λόγω των καλών μαθηματικών ιδιοτήτων και της απλότητας της Γκαουσιανής κατανομής.

Το πρόβλημα, όμως, με την απόσταση EMD έγκειται αλλού. Για να υπολογιστεί η εν λόγω απόσταση πρέπει να επιλυθεί ένα πρόβλημα ελαχιστοποίησης υπό περιορισμούς, όπως αυτό ορίστηκε προηγουμένως. Το συγκεκριμένο πρόβλημα ελαχιστοποίησης είναι γραμμικό και για την επίλυσή του χρησιμοποιείται γραμμικός προγραμματισμός (linear programming). Δηλαδή, με λίγα λόγια, δεν υπάρχει λύση σε κλειστή μορφή για τον υπολογισμό της απόστασης EMD. Αυτό, μοιραία, επιβαρύνει πολύ τον υπολογισμό της, και καθιστά τη χρήση της πολύ δύσκολη για τα προβλήματα που μας απασχολούν σε αυτήν τη διατριβή, για τα οποία η απαίτηση για γρήγορη απόκριση του συστήματος είναι κεφαλαιώδους σημασίας.

### 3.3 Μέτρο Ομοιότητας ALA

Γίνεται όλο και πιο φανερό από τα παραπάνω ότι αυτό που ζητάμε είναι πάντα μια απόσταση που να μπορεί να υπολογιστεί σε κλειστή μορφή για Γκαουσιανά Μείγματα. Υπό



αυτό το πρίσμα προτάθηκε το μέτρο ομοιότητας ALA (Asymptotic Likelihood Approximation – Ασυμπτωτική Προσέγγιση Πιθανοφάνειας) [18].

Για τα Γκαουσιανά Μείγματα  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$  της (3.1), η ALA ορίζεται μέσω των επόμενων σχέσεων:

$$ALA(p_1 \parallel p_2) = \sum_{i=1}^{K_1} \pi_{1i} \left[ \ln \pi_{2\beta_{12}(i)} + \ln \phi(\boldsymbol{\mu}_{1i} \mid \boldsymbol{\theta}_{2\beta_{12}(i)}) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_{2\beta_{12}(i)}^{-1} \boldsymbol{\Sigma}_{1i}) \right] \quad (3.13)$$

$$\beta_{12}(i) = \arg \min_{j=1, \dots, K_2} \left\{ (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2j})^T \boldsymbol{\Sigma}_{2j}^{-1} (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2j}) - \ln \pi_{2j} \right\} \quad (3.14)$$

Ουσιαστικά, χρησιμοποιείται ένα μέτρο απόστασης μεταξύ Γκαουσιανών συνιστωσών, και συγκεκριμένα το

$$d_{12}(i, j) = (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2j})^T \boldsymbol{\Sigma}_{2j}^{-1} (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2j}) - \ln \pi_{2j} \quad (3.15)$$

το οποίο στηρίζεται στην απόσταση Mahalanobis ανάμεσα στις μέσες τιμές των συνιστωσών των δύο μειγμάτων, προκειμένου να δημιουργηθεί μια αντιστοίχιση μεταξύ των παραπάνω συνιστωσών. Αυτή η αντιστοίχιση συνδέει κάθε συνιστώσα του πρώτου μείγματος με την κοντινότερή της, όπως καθορίζεται από το μέτρο  $d_{12}$ , από το δεύτερο μείγμα. Κατόπιν, για κάθε ζεύγος τέτοιων συνιστωσών υπολογίζεται ένα μέτρο ομοιότητας μεταξύ τους, που ορίζεται μέσω του τύπου

$$s_{12}(i, j) = \ln \pi_{2j} + \ln \phi(\boldsymbol{\mu}_{1i} \mid \boldsymbol{\theta}_{2j}) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_{2j}^{-1} \boldsymbol{\Sigma}_{1i}) \quad (3.16)$$

Τελικά, η τιμή για την ALA προκύπτει ως γραμμικός συνδυασμός, με βάρη αυτά του πρώτου μείγματος, των τιμών του μέτρου  $s_{12}$  για τα ζεύγη των κοντινότερων συνιστωσών των δύο μειγμάτων.

### 3.3.1 Προσέγγιση της Απόκλισης KL

Στην εργασία [18] αποδεικνύεται ότι υπάρχει μια ενδιαφέρουσα σχέση που συνδέει την ALA με την απόκλιση KL. Συγκεκριμένα, αποδεικνύεται ότι η ALA μπορεί να χρησιμοποιηθεί για να υπολογίσουμε μια προσέγγιση της απόκλισης KL.

Ξεκινώντας από τον ορισμό της απόκλισης KL, που δίνεται από τη σχέση (3.2), προκύπτει ότι

$$KL(p_1 \parallel p_2) = EL(p_1 \parallel p_1) - EL(p_1 \parallel p_2) \quad (3.17)$$

όπου

$$EL(p_i \parallel p_j) = E[\ln p_j(\mathbf{X})]_{p_i} = \int p_i(\mathbf{x}) \ln p_j(\mathbf{x}) d\mathbf{x} \quad (3.18)$$

είναι η προσδοκώμενη τιμή του λογαρίθμου του  $p_j(\mathbf{X})$ , θεωρώντας ότι η  $\mathbf{X}$  ακολουθεί την  $p_i(\mathbf{x})$ .

Σε αυτό το σημείο αξίζει να σημειωθεί ότι

$$EL(p \parallel p) = \int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} = -H_p(\mathbf{X}) \quad (3.19)$$

με  $H_p(\mathbf{X})$  να είναι η εντροπία της τυχαίας μεταβλητής  $\mathbf{X}$ , όταν η κατανομή της τελευταίας περιγράφεται από την  $p(\mathbf{x})$ . Η εντροπία αποτελεί ένα μέτρο αβεβαιότητας για μια τυχαία μεταβλητή. Για παράδειγμα, μια μεταβλητή που μπορεί να πάρει μόνο μια συγκεκριμένη τιμή (ουσιαστικά μια μη τυχαία μεταβλητή) έχει εντροπία 0.

Επίσης, για κάθε σύνθετη κατανομή, έστω  $p(\mathbf{x})$ , με  $K$  συνιστώσες, ο χώρος στον οποίον αυτή ορίζεται, δηλαδή ο χώρος στον οποίο κινείται το  $\mathbf{x}$ , μπορεί να χωριστεί σε  $K$  υποσύνολα, λαμβάνοντας υπόψιν τις εκ των υστέρων πιθανότητες  $P(k|\mathbf{x})$  των συνιστώσων, για  $k = 1, \dots, K$ . Συγκεκριμένα, το  $k$ -οστό από αυτά τα υποσύνολα, έστω  $\chi_k$ , περιλαμβάνει τα σημεία  $\mathbf{x}$  για τα οποία ισχύει η σχέση

$$P(k|\mathbf{x}) > P(l|\mathbf{x}), \quad \forall l \neq k \quad (3.20)$$

με την πιθανότητα  $P(k|\mathbf{x})$  να υπολογίζεται κατά τα γνωστά, με βάση τον τύπο του Bayes:

$$P(k|\mathbf{x}) = \frac{P(k)p(\mathbf{x}|k)}{p(\mathbf{x})} = \frac{P(k)p(\mathbf{x}|k)}{\sum_{j=1}^K P(j)p(\mathbf{x}|j)} \quad (3.21)$$

και να αντιστοιχεί στην πιθανότητα το συγκεκριμένο  $\mathbf{x}$  να προέλθει από την  $k$ -οστή συνιστώσα της κατανομής. Επομένως, το σύνολο  $\chi_k$  περιλαμβάνει όλα εκείνα τα σημεία  $\mathbf{x}$  τα οποία είναι πιο πιθανό να προέλθουν από την  $k$ -οστή συνιστώσα παρά από τις υπόλοιπες συνιστώσες της κατανομής.

Ακόμα, για κάθε συνιστώσα της  $p(\mathbf{x})$  ορίζεται και ένα σύνολο υποστήριξης (support). Για παράδειγμα, για την  $j$ -οστή συνιστώσα το σύνολο υποστήριξης, έστω  $S_j$ , περιλαμβάνει τα σημεία εκείνα του χώρου,  $\mathbf{x}$ , για τα οποία η εν λόγω συνιστώσα εμφανίζει μη μηδενική τιμή πυκνότητας πιθανότητας  $p(\mathbf{x}|j)$ .

Χρησιμοποιώντας τους παραπάνω ορισμούς και έννοιες, προκύπτει ότι υπάρχουν δύο συνθήκες που όταν ισχύουν, τότε μέσω της ALA μπορούμε να υπολογίσουμε την απόκλιση KL. Συγκεκριμένα, για τις συναρτήσεις πυκνότητας πιθανότητας  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$ , και θεωρώντας ότι τα σύνολα  $\chi_k$  που αντιστοιχούν στην  $p_2(\mathbf{x})$  συμβολίζονται με  $\chi_k^2$ , ενώ τα σύνολα  $S_j$  που αντιστοιχούν στην  $p_1(\mathbf{x})$  συμβολίζονται με  $S_j^1$ , οι συνθήκες αυτές διατυπώνονται αμέσως παρακάτω:

1<sup>η</sup> συνθήκη: Η πρώτη από τις δύο συνθήκες δίνεται από τη σχέση

$$P_2(k|\mathbf{x}) = 1, \quad \forall \mathbf{x} \in \chi_k^2 \quad (3.22)$$

Αυτό σημαίνει πρακτικά ότι όλα τα σημεία στο  $\chi_k^2$  είναι βέβαιο ότι μπορούν να προκύψουν μόνο από την  $k$ -οστή συνιστώσα της κατανομής  $p_2(\mathbf{x})$ .

2<sup>η</sup> συνθήκη: Η δεύτερη συνθήκη διατυπώνεται ως

$$\forall j, \exists k : S_j^1 \subset \mathcal{X}_k^2 \quad (3.23)$$

και, πρακτικά, σημαίνει ότι κάθε σύνολο υποστήριξης της  $p_1(\mathbf{x})$  πρέπει να περιέχεται πλήρως σε κάποιο από τα  $\mathcal{X}_k^2$  της  $p_2(\mathbf{x})$ .

Όταν αυτές οι δύο συνθήκες ισχύουν, τότε προκύπτει ότι

$$ALA(p_1 \parallel p_1) = EL(p_1 \parallel p_1) \quad \text{και} \quad ALA(p_1 \parallel p_2) = EL(p_1 \parallel p_2) \quad (3.24)$$

και επομένως, λαμβάνοντας υπόψιν και τη σχέση (3.17), καταλήγουμε στο συμπέρασμα ότι

$$KL(p_1 \parallel p_2) = ALA(p_1 \parallel p_1) - ALA(p_1 \parallel p_2) \quad (3.25)$$

Βέβαια, οι παραπάνω συνθήκες είναι πολύ περιοριστικές, και δεν είναι δυνατόν να ισχύουν ακριβώς για Γκαουσιανά Μείγματα, λόγω του ότι οι Γκαουσιανές συνιστώσες που τα αποτελούν έχουν, θεωρητικά, μη μηδενική τιμή για όλα τα  $\mathbf{x}$ . Αν όμως οι μέσες τιμές των Γκαουσιανών συνιστωσών είναι αρκετά απομακρυσμένες σε σχέση με το πλάτος των τελευταίων, που σχετίζεται με τους αντίστοιχους πίνακες συνδιακύμανσης, οι συνθήκες θα ισχύουν προσεγγιστικά.

Σε κάθε περίπτωση, έχουμε μέσω της ALA ένα τρόπο να προσεγγίσουμε την απόκλιση KL με έναν υπολογισμό που μπορεί να γίνει σε κλειστή μορφή. Συγκεκριμένα, λαμβάνοντας υπόψιν την παραπάνω ανάλυση, στηριζόμενοι στην ALA μπορούμε να ορίσουμε ένα μέτρο απόστασης:

$$KL_{\text{mhl-match}}(p_1 \parallel p_2) = ALA(p_1 \parallel p_1) - ALA(p_1 \parallel p_2) \quad (3.26)$$

το οποίο θα αποτελεί προσέγγιση της απόκλισης KL μεταξύ Γκαουσιανών Μειγμάτων. Από τη σχέση (2.30), για τη Γκαουσιανή κατανομή, από τη σχέση (3.13), για τον ορισμό της ALA, αλλά και από τη σχέση (3.7), για τον υπολογισμό της απόκλισης KL στην περίπτωση που οι προς σύγκριση κατανομές είναι Γκαουσιανές, προκύπτει πολύ εύκολα ότι

$$KL_{\text{mhl-match}}(p_1 \parallel p_2) = \sum_{i=1}^{K_1} \pi_{1i} \left[ KL(\phi(\mathbf{x} | \boldsymbol{\theta}_{1i}) \parallel \phi(\mathbf{x} | \boldsymbol{\theta}_{2\beta_{12}(i)})) + \ln \frac{\pi_{1i}}{\pi_{2\beta_{12}(i)}} \right] \quad (3.27)$$

Δηλαδή, όπως γίνεται εύκολα αντιληπτό, αυτό το προσεγγιστικό μέτρο απόστασης ορίζεται με βάση την απόκλιση KL ανάμεσα στα ζεύγη συνιστωσών των δύο μειγμάτων, όπως αυτά καθορίζονται από τη συνάρτηση αντιστοίχισης  $\beta_{12}$ .

Αξίζει να σημειωθεί, επίσης, ότι στην εργασία [54] ένα παρόμοιο μέτρο απόστασης, προσεγγιστικό της απόκλισης KL, ορίζεται από τις σχέσεις

$$KL_{\text{gkl-match}}(p_1 \parallel p_2) = \sum_{i=1}^{K_1} \pi_{1i} \left[ KL(\phi(\mathbf{x} | \boldsymbol{\theta}_{1i}) \parallel \phi(\mathbf{x} | \boldsymbol{\theta}_{2\gamma_{12}(i)})) + \ln \frac{\pi_{1i}}{\pi_{2\gamma_{12}(i)}} \right] \quad (3.28)$$

$$\gamma_{12}(i) = \arg \min_{j=1, \dots, K_2} \{ KL(\phi(\mathbf{x} | \boldsymbol{\theta}_{1i}) \parallel \phi(\mathbf{x} | \boldsymbol{\theta}_{2j})) - \ln \pi_{2j} \} \quad (3.29)$$

Είναι εμφανές ότι αυτό το μέτρο απόστασης είναι εντελώς παρόμοιο με το  $KL_{\text{mhl-match}}$ , που παρουσιάστηκε προηγουμένως. Η μόνη διαφορά έγκειται στον ορισμό της συνάρτησης αντιστοίχισης. Στην περίπτωση της βασιζόμενης στην ALA προσέγγισης, η συνάρτηση αντιστοίχισης,  $\beta_{12}$ , ορίζεται με χρήση της απόστασης Mahalanobis μεταξύ των κέντρων των συνιστωσών των δύο μειγμάτων, όπως φαίνεται από τη σχέση (3.14). Αντίθετα, το μέτρο απόστασης της εργασίας [54] χρησιμοποιεί μια άλλη συνάρτηση αντιστοίχισης,  $\gamma_{12}$ , που έχει ως βάση την απόκλιση KL μεταξύ των συνιστωσών των δύο μειγμάτων.

Οι δύο παραπάνω προσεγγίσεις της απόκλισης KL, όπως και η ίδια η απόκλιση KL, δεν είναι συμμετρικές. Μπορούμε όμως, πάλι, να ορίσουμε με απλό τρόπο μια συμμετρική εκδοχή τους μέσω των τύπων

$$SKL_{\text{mhl-match}}(p_1, p_2) = \frac{1}{2} KL_{\text{mhl-match}}(p_1 \parallel p_2) + \frac{1}{2} KL_{\text{mhl-match}}(p_2 \parallel p_1) \quad (3.30)$$

$$SKL_{\text{gkl-match}}(p_1, p_2) = \frac{1}{2} KL_{\text{gkl-match}}(p_1 \parallel p_2) + \frac{1}{2} KL_{\text{gkl-match}}(p_2 \parallel p_1) \quad (3.31)$$

Από εκεί και πέρα, ένα πρόβλημα των παραπάνω προσεγγίσεων της απόκλισης KL έγκειται στο ότι ο υπολογισμός τους εμφανίζει τετραγωνική πολυπλοκότητα σε σχέση με τον αριθμό των Γκαουσιανών συνιστωσών των μειγμάτων. Πράγματι, για τον υπολογισμό των τιμών των συναρτήσεων αντιστοίχισης  $\beta_{12}$  και  $\gamma_{12}$ , είναι απαραίτητος ο υπολογισμός των αντίστοιχων μέτρων απόστασης μεταξύ όλων των ζευγών συνιστωσών των δύο μειγμάτων. Μετά, για κάθε συνιστώσα του πρώτου μείγματος προσδιορίζεται η εγγύτερη ως προς αυτή συνιστώσα του δεύτερου.

Πάντως, λόγω της δυνατότητας υπολογισμού τους σε κλειστή μορφή, τα δύο παραπάνω μέτρα απόστασης χρησιμοποιήθηκαν σε σημαντικό βαθμό για τις ανάγκες της παρούσας διατριβής. Μάλιστα, στο επόμενο κεφάλαιο προτείνεται ένας τρόπος γρήγορης ανανέωσης της τιμής για το μέτρο ALA, με βάση μια νέα τεχνική ανανέωσης του μοντέλου της ερώτησης στα πλαίσια της συσχετιστικής ανατροφοδότησης. Επίσης, στο κεφάλαιο 6 παρουσιάζεται μια γρήγορη παραλλαγή αυτών των μέτρων απόστασης, που στηρίζεται σε έναν από πολλές απόψεις ενδιαφέροντα τρόπο εκπαίδευσης των Γκαουσιανών Μειγμάτων.

### 3.4 Απόκλιση C2

Ένα άλλο μέτρο απόστασης που έχει προταθεί [28] για σύγκριση μεταξύ πιθανοτικών μοντέλων και εμφανίζει πολύ ευεργετικές ιδιότητες όταν αυτά τα μοντέλα είναι Γκαουσιανά Μείγματα, είναι η απόκλιση C2 (C2 divergence). Οι ιδιότητες του συγκεκριμένου μέτρου απόστασης μελετήθηκαν αναλυτικά, και έγινε εκτενής χρήση του για την αντιμετώπιση του προβλήματος της ανάκτησης εικόνων μέσω συσχετιστικής ανατροφοδότησης, όταν οι εικόνες αναπαριστάνονται ως Γκαουσιανά Μείγματα.

Η απόκλιση C2 μεταξύ δύο συναρτήσεων πυκνότητας πιθανότητας  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$  ορίζεται ως

$$C2(p_1, p_2) = -\ln \frac{2S(p_1, p_2)}{S(p_1, p_1) + S(p_2, p_2)} \quad (3.32)$$

όπου ισχύει

$$S(p_m, p_l) = \int p_m(\mathbf{x})p_l(\mathbf{x})d\mathbf{x} \quad (3.33)$$

για  $m, l = 1, 2$ .

Μπορεί να αποδειχθεί (η απόδειξη δίνεται στο παράρτημα A, και συγκεκριμένα στην ενότητα A1) πως όταν τα  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$  είναι Γκαουσιανά Μείγματα, η απόκλιση C2 είναι δυνατόν να υπολογιστεί σε κλειστή μορφή. Συγκεκριμένα, σε αυτήν την περίπτωση ο υπολογισμός των  $S(p_m, p_l)$  μπορεί να γίνει μέσω των σχέσεων που ακολουθούν:

$$S(p_m, p_l) = \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} \frac{1}{\sqrt{(2\pi)^d |\mathbf{C}_{ml}(i, j)|}} \exp\left[-\frac{1}{2} \mathbf{D}_{ml}(i, j)\right] \quad (3.34)$$

όπου

$$\mathbf{C}_{ml}(i, j) = \mathbf{\Sigma}_{mi} + \mathbf{\Sigma}_{lj} \quad (3.35)$$

$$\mathbf{D}_{ml}(i, j) = (\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj})^T \mathbf{C}_{ml}^{-1}(i, j) (\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj}) \quad (3.36)$$

Στις παραπάνω σχέσεις, το μέτρο  $\mathbf{D}_{ml}(i, j)$  είναι πάλι, όπως βλέπουμε, μια απόσταση Mahalanobis μεταξύ των συνιστωσών των δύο μειγμάτων.

Επίσης, γίνεται εύκολα αντιληπτό ότι το μέτρο S, που ορίζεται από τη σχέση (3.33), δεν είναι άλλο από το εσωτερικό γινόμενο μεταξύ των συναρτήσεων πυκνότητας πιθανότητας  $p_m(\mathbf{x})$  και  $p_l(\mathbf{x})$ .

Υπενθυμίζουμε εδώ, με την ευκαιρία, ότι αν έχουμε δύο συναρτήσεις συνεχούς μεταβλητής  $f(\mathbf{x})$  και  $g(\mathbf{x})$ , τότε το εσωτερικό γινόμενο (inner product) μεταξύ τους ορίζεται ως

$$\langle f, g \rangle = \int f(\mathbf{x})g(\mathbf{x})d\mathbf{x} \quad (3.37)$$

και η νόρμα  $L_2$  ως

$$L_2(f) = \sqrt{\langle f, f \rangle} = \sqrt{\int f^2(\mathbf{x})d\mathbf{x}} \quad (3.38)$$

Από τη σχέση (3.34) προκύπτει επομένως ότι όταν τα  $p_m(\mathbf{x})$  και  $p_l(\mathbf{x})$  αντιστοιχούν σε Γκαουσιανά Μείγματα, το εσωτερικό γινόμενο μεταξύ τους είναι ένα γραμμικός συνδυασμός, και συγκεκριμένα

$$S(p_m, p_l) = \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} s_{ml}(i, j) \quad (3.39)$$

κάποιων ποσοτήτων  $s_{ml}(i, j)$  για τις οποίες ισχύει

$$s_{ml}(i, j) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{C}_{ml}(i, j)|}} \exp\left[-\frac{1}{2} \mathbf{D}_{ml}(i, j)\right] \quad (3.40)$$

Όμως, οι τελευταίες ποσότητες δεν είναι τίποτα άλλο από τα εσωτερικά γινόμενα μεταξύ όλων των δυνατών ζευγών Γκαουσιανών συνιστωσών των δύο μειγμάτων. Βλέπουμε εδώ, πάλι, τις κομψές μαθηματικές ιδιότητες της Γκαουσιανής κατανομής και, κατ' επέκταση, των Γκαουσιανών Μειγμάτων.

Προφανώς, και η απόκλιση C2 χαρακτηρίζεται από τετραγωνική πολυπλοκότητα ως προς τον αριθμό των συνιστωσών των δύο Γκαουσιανών Μειγμάτων που συγκρίνονται, αφού απαιτεί τον υπολογισμό των ποσοτήτων  $s_{ml}(i, j)$  για όλα τα ζεύγη συνιστωσών των μοντέλων. Παρόλα αυτά, στο επόμενο κεφάλαιο θα δούμε μια τεχνική που μπορεί να χρησιμοποιηθεί στα πλαίσια της συσχετιστικής ανατροφοδότησης, προκειμένου να επιτευχθεί η ταχύτερη ανανέωση των τιμών του μέτρου αυτού σε κάθε γύρο της ανατροφοδότησης.

### 3.4.1 Ιδιότητες

Όπως έχουμε δει παραπάνω μέσω των σχέσεων (1.10)-(1.13), 4 είναι οι ιδιότητες που πρέπει να ισχύουν ώστε ένα μέτρο απόστασης να χαρακτηριστεί μετρική. Όσο αφορά τις ιδιότητες αυτές, για την απόκλιση C2 αποδεικνύονται εύκολα τα επόμενα:

1. Η απόκλιση C2 είναι πάντα θετική ποσότητα.

$$C2(p_1, p_2) \geq 0 \quad (3.41)$$

2. Η απόκλιση C2 είναι ίση με το μηδέν, αν και μόνο αν οι συναρτήσεις υπό σύγκριση είναι ίσες για κάθε τιμή του πεδίου ορισμού τους, εκτός ίσως από ένα σύνολο σημείων μηδενικού μέτρου. Στην πράξη, για συνεχείς συναρτήσεις ισχύει

$$C2(p_1, p_2) = 0 \Leftrightarrow p_1(\mathbf{x}) = p_2(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{R}^d \quad (3.42)$$

3. Η απόκλιση C2 είναι συμμετρική.

$$C2(p_1, p_2) = C2(p_2, p_1) \quad (3.43)$$

4. Η τριγωνική ανισότητα δεν ισχύει.

Από τις ανωτέρω διαπιστώσεις καταλήγουμε στο συμπέρασμα ότι το μέτρο απόστασης C2 δε μπορεί να χαρακτηριστεί μετρική, αλλά είναι μια απόκλιση (divergence), όπως εξάλλου και το δημοφιλέστερο μέτρο για σύγκριση μεταξύ πιθανοτικών μοντέλων, δηλαδή η απόκλιση KL.

Επίσης, όσο αφορά τις σχέσεις που συνδέουν την απόκλιση C2 με άλλα ανάλογα μέτρα απόστασης μεταξύ συναρτήσεων πυκνότητας πιθανότητας, αποδείξαμε τα ακόλουθα:

### Σχέση της Απόκλισης C2 με τη Συμμετρική Εκδοχή της Απόκλισης KL

Όσο αφορά τη σχέση της απόκλισης C2 με τη συμμετρική εκδοχή της απόκλισης KL, δηλαδή την SKL, όπως αυτή ορίστηκε από τον τύπο (3.4), ισχύει ότι η διαφορά μεταξύ των δύο αποκλίσεων είναι φραγμένη άνω και κάτω από τις εξής ανισότητες (η απόδειξη δίνεται στην ενότητα A2 του παραρτήματος A):

$$-\Delta_{ij} \leq C2(p_1, p_2) - \text{SKL}(p_1, p_2) \leq \Delta_{ij} \quad (3.44)$$

όπου

$$\Delta_{ii} = \max(\ln E[p_1]_{p_1}, \ln E[p_2]_{p_2}) - \min(E[\ln p_1]_{p_1}, E[\ln p_2]_{p_2}) \geq 0 \quad (3.45)$$

$$\Delta_{ij} = \max(\ln E[p_2]_{p_1}, \ln E[p_1]_{p_2}) - \min(E[\ln p_2]_{p_1}, E[\ln p_1]_{p_2}) \geq 0 \quad (3.46)$$

Οι προσδοκώμενες τιμές στις παραπάνω σχέσεις υπολογίζονται κατά τα γνωστά. Συγκεκριμένα

$$E[\ln p_j]_{p_i} = \int p_i(\mathbf{x}) \ln p_j(\mathbf{x}) d\mathbf{x} \quad (3.47)$$

$$E[p_j]_{p_i} = \int p_i(\mathbf{x}) p_j(\mathbf{x}) d\mathbf{x} \quad (3.48)$$

για  $i, j = 1, 2$ . Από την (3.48) βλέπουμε ότι  $E[p_2]_{p_1} = E[p_1]_{p_2}$ . Άρα, όσο αφορά τον ορισμό της ποσότητας  $\Delta_{ij}$  μέσω της (3.46), ισχύει  $\max(\ln E[p_2]_{p_1}, \ln E[p_1]_{p_2}) = \ln E[p_2]_{p_1} = \ln E[p_1]_{p_2}$ .

Η σχέση (3.44) για τη διαφορά μεταξύ της συμμετρικής απόκλισης KL και της απόκλισης C2 προκύπτει με χρήση της ανισότητας του Jensen (Jensen's inequality), που σε ένα πιθανοτικό πλαίσιο διατυπώνεται ως

$$f(E[X]) \geq E[f(X)] \quad (3.49)$$

για μια μονοδιάστατη τυχαία μεταβλητή  $X$  και για μια κοίλη συνάρτηση  $f$ . Υπενθυμίζουμε ότι μια συνάρτηση είναι κοίλη σε ένα διάστημα του πεδίου ορισμού της, αν για τα σημεία στο διάστημα αυτό η δευτέρα παράγωγος της συνάρτησης είναι αρνητική. Από την άλλη, όταν η συνάρτηση  $f$  είναι γραμμική, δηλαδή όταν είναι της μορφής

$$f(x) = ax + b \quad (3.50)$$

τότε η ανισότητα του Jensen μετατρέπεται σε ισότητα, γιατί, όπως αποδεικνύεται εύκολα από τον ορισμό (3.8), για την  $E(f(X))$  ισχύει

$$E[f(X)] = E[aX + b] = aE[X] + b = f(E[X]) \quad (3.51)$$

Όμως, στην περίπτωση της σχέσης της απόκλισης C2 με την απόκλιση SKL, η συνάρτηση  $f$  είναι αυτή του λογαρίθμου, που είναι πράγματι μια κοίλη συνάρτηση. Παρόλα αυτά, δεδομένου ότι είναι γνωστό πως ο λογάριθμος μπορεί να προσεγγιστεί τοπικά μέσω μιας γραμμικής συνάρτησης, για παράδειγμα

$$\ln x \approx x - 1 \quad \text{για } x \approx 1 \quad (3.52)$$

προκύπτει ότι τα  $\Delta_{ii}$  και  $\Delta_{ij}$  λαμβάνουν μικρότερες τιμές, και ως εκ τούτου ο περιορισμός όσο αφορά τις τιμές της διαφοράς μεταξύ της C2 και της SKL καθίσταται περισσότερο αυστηρός, όταν οι συναρτήσεις πυκνότητας πιθανότητας  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$  είναι σχετικά

ομαλές, έτσι ώστε η υπόθεση περί προσέγγισης του λογαρίθμου τους από μια γραμμική συνάρτηση να είναι ρεαλιστική.

### Σχέση της Απόκλισης C2 με τη Νόρμα $L_2$

Όσο αφορά τη σχέση μεταξύ της απόκλισης C2 και της νόρμας  $L_2$ , όπως η τελευταία ορίστηκε από τους τύπους (3.37)-(3.38), προκύπτει εύκολα ότι

$$C2(p_1, p_2) = -\ln \frac{2 \langle p_1, p_2 \rangle}{L_2^2(p_1) + L_2^2(p_2)} = -\ln \left[ 1 - \frac{L_2^2(p_1 - p_2)}{L_2^2(p_1) + L_2^2(p_2)} \right] \quad (3.53)$$

Χρησιμοποιώντας πάλι τη γραμμική προσέγγιση (3.52) για τη συνάρτηση του λογαρίθμου, παίρνουμε την παρακάτω προσεγγιστική σχέση μεταξύ της απόκλισης C2 και της νόρμας  $L_2$ :

$$C2(p_1, p_2) \approx \frac{L_2^2(p_1 - p_2)}{L_2^2(p_1) + L_2^2(p_2)} \quad \text{για } L_2^2(p_1 - p_2) \ll L_2^2(p_1) + L_2^2(p_2) \quad (3.54)$$



## Κεφάλαιο 4

# Συσχετιστική Ανατροφοδότηση Μέσω Τροποποίησης της Αναπαράστασης της Ερώτησης

Στο κεφάλαιο αυτό γίνεται η παρουσίαση μιας νέας μεθόδου συσχετιστικής ανατροφοδότησης στο πλαίσιο της ανάκτησης εικόνων με βάση το περιεχόμενο. Οι εικόνες της βάσης αναπαριστώνται μέσω Γκαουσιανών Μειγμάτων. Η μέθοδος συσχετιστικής ανατροφοδότησης, από την άλλη, χρησιμοποιεί την απόκλιση C2 ως μέτρο σύγκρισης μεταξύ των Γκαουσιανών Μειγμάτων και στηρίζεται σε ένα μοντέλο για την ερώτηση του χρήστη που ανανεώνεται από εποχή σε εποχή της ανατροφοδότησης ενσωματώνει τα παραδείγματα που δίνει ο τελευταίος. Επιπλέον, επωφελούμενη από τον τρόπο ανανέωσης του μοντέλου της ερώτησης και από τις ιδιότητες της απόκλισης C2, οδηγείται σε μια πολύ αποδοτική ανανέωση της εκτίμησής της για τη συνάφεια μεταξύ της ερώτησης του χρήστη και της κάθε εικόνας της βάσης. Στο τέλος του κεφαλαίου παρατίθενται και σχολιάζονται τα αποτελέσματα μιας σειράς πειραμάτων που έγιναν, και τα οποία αποδεικνύουν τα πλεονεκτήματα της προτεινόμενης μεθόδου. Ως αποτέλεσμα αυτού του σκέλους της έρευνας προέκυψαν οι δημοσιεύσεις [J1], [C1] και [C2].

### 4.1 Ανανέωση του Μοντέλου της Ερώτησης

Όπως αναφέραμε ήδη σε προηγούμενο κεφάλαιο, υπάρχουν δύο τρόποι με τους οποίους μπορούμε να υλοποιήσουμε την τεχνική της συσχετιστικής ανατροφοδότησης. Ο πρώτος έχει να κάνει με την εκπαίδευση ενός ταξινομητή που θα διαχωρίζει τα θετικά από τα αρνητικά παραδείγματα, και μέσω του οποίου θα μπορούμε για οποιαδήποτε εικόνα της βάσης να εκτιμήσουμε πόσο σχετική είναι αυτή με τις προτιμήσεις του χρήστη. Ο δεύτερος τρόπος υλοποίησης, με τον οποίο θα ασχοληθούμε σε αυτό το κεφάλαιο, έχει να κάνει με την τροποποίηση της ερώτησης και του μέτρου απόστασης σε κάθε εποχή της ανατροφοδότησης, έτσι ώστε να λαμβάνονται υπόψιν τα παραδείγματα που παρέχει ο χρήστης.

Θεωρούμε, αρχικά, ότι κάθε εικόνα ή ερώτηση που θα χρησιμοποιηθεί με οποιοδήποτε τρόπο παρακάτω περιγράφεται μέσω ενός πιθανοτικού μοντέλου. Έστω λοιπόν ότι έχουμε μια βάση  $N$  εικόνων, και ότι η  $i$ -οστή εικόνα της βάσης περιγράφεται μέσω ενός μοντέλου  $p_i(\mathbf{x})$ , με  $i = 1, \dots, N$ . Ακόμα, έστω ότι βρισκόμαστε σε μια συγκεκριμένη εποχή της ανατροφοδότησης και ότι στην εποχή αυτή ο χρήστης έχει αποφασίσει ότι ένα σύνολο  $M_T$  εικόνων που περιγράφονται από τα μοντέλα  $p_m^t(\mathbf{x})$ , με  $m = 1, \dots, M_T$ , είναι σχετικές με αυτό

που αναζητά. Αυτό που θέλουμε είναι να ανανεώσουμε το τρέχον μοντέλο της ερώτησης, έστω  $p_q(\mathbf{x})$ , χρησιμοποιώντας τα παραδείγματα του χρήστη.

Ένας εύκολος και διαισθητικός τρόπος για να επιτύχουμε αυτόν τον σκοπό είναι να ορίσουμε το νέο μοντέλο της ερώτησης ως ένα κυρτό συνδυασμό του μοντέλου για την προηγούμενη ερώτηση και των μοντέλων των παραδειγμάτων του χρήστη. Επομένως, αν  $p_{q'}(\mathbf{x})$  είναι το ανανεωμένο μοντέλο της ερώτησης, τότε για αυτό θα ισχύει

$$p_{q'}(\mathbf{x}) = (1 - \Lambda_r)p_q(\mathbf{x}) + \sum_{m=1}^{M_r} \lambda_m^r p_m^r(\mathbf{x}) \quad (4.1)$$

όπου για τα βάρη  $\lambda_m^r$  και  $\Lambda_r$  ισχύει  $0 \leq \lambda_m^r \leq 1$ ,  $\sum_m \lambda_m^r = \Lambda_r$ , και  $0 \leq \Lambda_r \leq 1$ . Το βάρος  $\lambda_m^r$  αντιστοιχεί στο βαθμό συνάφειας της εικόνας με μοντέλο  $p_m^r(\mathbf{x})$ , έστω  $r_m$ , όπως αυτός μπορεί να προσδιοριστεί από το χρήστη. Δηλαδή, το μοντέλο για τη νέα ερώτηση ορίζεται έτσι ώστε να μπορεί εύκολα να ενσωματώσει μια ανατροφοδότηση από τη μεριά του χρήστη η οποία δε θα περιορίζεται στο χαρακτηρισμό μιας εικόνας ως σχετικής ή μη σχετικής, αλλά θα αναθέτει σε κάθε εικόνα που ανατροφοδοτείται και ένα επίπεδο συνάφειας. Επίσης, μέσω της παραμέτρου  $\Lambda_r$  καθορίζεται ο βαθμός συνεισφοράς της τρέχουσας ερώτησης στο μοντέλο για τη νέα ερώτηση μετά την ενσωμάτωση της ανατροφοδότησης, και ο οποίος βαθμός ισούται με  $1 - \Lambda_r$ .

Ο παραπάνω ορισμός για το μοντέλο της νέας ερώτησης έχει, εκτός των άλλων, και ένα ελκυστικό χαρακτηριστικό που έγκειται στη συνέπειά του με το όλο πιθανοτικό πλαίσιο που χρησιμοποιείται για τη μοντελοποίηση των εικόνων. Συγκεκριμένα, είναι φανερό ότι το νέο μοντέλο της ερώτησης είναι σύνθετο, και μπορεί να θεωρηθεί ότι συνίσταται από τον πιθανοτικό συνδυασμό ενός αριθμού από απλούστερα μοντέλα στη μορφή

$$p(\mathbf{x} | q') = P(q)p(\mathbf{x} | q) + \sum_{m=1}^{M_r} P(r_m)p(\mathbf{x} | r_m) \quad (4.2)$$

όπου, προφανώς, τα  $p(\mathbf{x} | q')$ ,  $p(\mathbf{x} | q)$  και  $p(\mathbf{x} | r_m)$  ισούνται με  $p_{q'}(\mathbf{x})$ ,  $p_q(\mathbf{x})$  και  $p_m^r(\mathbf{x})$ , αντίστοιχα, και εκφράζουν, ακριβώς, την πυκνότητα πιθανότητας στο  $\mathbf{x}$  με δεδομένο το αντίστοιχο μοντέλο. Επομένως, για τα βάρη  $\lambda_m^r$  και  $\Lambda_r$  μπορεί να δοθεί μια καθαρά πιθανοτική ερμηνεία. Συγκεκριμένα, για τα βάρη αυτά ισχύει

$$\lambda_m^r = P(r_m) \quad (4.3)$$

και

$$\Lambda_r = 1 - P(q) \quad (4.4)$$

Δηλαδή, το βάρος  $\lambda_m^r$  ισούται με την εκ των προτέρων (όπως αποκαλείται) πιθανότητα (prior probability) του μοντέλου του θετικού παραδείγματος  $r_m$  στο συνδυασμό που ορίζει την κατανομή που αντιστοιχεί στη νέα ερώτηση, ενώ η εκ των προτέρων πιθανότητα  $P(q)$ , που αντιστοιχεί στην τρέχουσα ερώτηση, ισούται με  $1 - \Lambda_r$ .

## 4.2 Ανανέωση του Μέτρου Απόστασης

Θεωρώντας τώρα ότι έχουμε ένα μοντέλο για την ερώτηση του χρήστη, χρειάζεται να χρησιμοποιήσουμε ένα μέτρο απόστασης για να υπολογίσουμε το πόσο κοντά ή μακριά είναι η κάθε εικόνα της βάσης στην ερώτηση αυτή. Αν υποθέσουμε ότι ως μοντέλα για τις εικόνες και την ερώτηση χρησιμοποιούνται τα Γκαουσιανά Μείγματα και ως απόσταση η απόκλιση C2, απαιτείται ο υπολογισμός των ποσοτήτων  $C2(p_q, p_i)$ , για κάθε  $i = 1, \dots, N$ .

Αρχικά, ο χρήστης για να ξεκινήσει τη διαδικασία ανάκτησης εικόνων υποβάλλει στο σύστημα μια εικόνα – παράδειγμα των προτιμήσεών του. Το μοντέλο της εικόνας αυτής χρησιμοποιείται ως αρχικό μοντέλο ερώτησης από το σύστημα. Κατόπιν, σε κάθε εποχή της ανατροφοδότησης το μοντέλο της ερώτησης ανανεώνεται μέσω της σχέσης (4.1). Λαμβάνοντας υπόψιν τόσο τον ορισμό του ανανεωμένου μοντέλου για την ερώτηση όσο και το γεγονός ότι όλες οι εικόνες αναπαριστάνονται ως Γκαουσιανά Μείγματα, καταλήγουμε αναδρομικά στη διαπίστωση ότι σε κάθε εποχή της ανατροφοδότησης το μοντέλο της ερώτησης είναι πράγματι ένα Γκαουσιανό Μείγμα, εν γένει «σύνθετο» όμως, αφού προκύπτει ως κυρτός συνδυασμός άλλων Γκαουσιανών Μειγμάτων.

Επομένως, ο υπολογισμός κάθε ποσότητας  $C2(p_q, p_i)$  είναι πάντα δυνατός σε κλειστή μορφή, χρησιμοποιώντας τις σχέσεις (3.32) και (3.34)-(3.36), που αφορούν τον υπολογισμό της απόκλισης C2 για Γκαουσιανά Μείγματα. Πρέπει όμως να σημειωθεί ότι, προφανώς, καθώς περνούν οι εποχές της ανατροφοδότησης ο συνολικός αριθμός των παραδειγμάτων του χρήστη αυξάνει, με άμεσο αποτέλεσμα την αύξηση της πολυπλοκότητας του μείγματος που μοντελοποιεί την τρέχουσα ερώτηση, αφού αυτό εν τέλει δεν είναι τίποτα άλλο από τον κυρτό συνδυασμό του μοντέλου της αρχικής ερώτησης και των μοντέλων για όλα τα παραδείγματα που έχουν δοθεί ως τώρα. Έτσι, ο απευθείας υπολογισμός της απόκλισης C2 μεταξύ του μοντέλου της τρέχουσας ερώτησης και των μοντέλων των εικόνων της βάσης θα επιβαρύνεται όλο και περισσότερο από εποχή σε εποχή της ανατροφοδότησης. Είναι επιτακτική ανάγκη, λοιπόν, να βρεθεί ένας τρόπος για την αποδοτική ανανέωση της απόστασης μεταξύ της ερώτησης και της κάθε εικόνας της βάσης, από εποχή σε εποχή, λαμβάνοντας υπόψιν τον τύπο ανανέωσης του μοντέλου της ερώτησης.

Πράγματι, τόσο ο ορισμός της απόκλισης C2 όσο και ο τρόπος με τον οποίον ανανεώνουμε την ερώτηση από εποχή σε εποχή της ανατροφοδότησης επιτρέπουν την με αποδοτικό τρόπο ανανέωση της απόκλισης C2. Συγκεκριμένα, η ανανέωση της απόκλισης C2 μπορεί να γίνει μέσω των ακόλουθων σχέσεων:

$$C2(p_{q'}, p_i) = -\ln \frac{2S(p_{q'}, p_i)}{S(p_{q'}, p_{q'}) + S(p_i, p_i)} \quad (4.5)$$

$$S(p_{q'}, p_i) = (1 - \Lambda_r)S(p_q, p_i) + \sum_{m=1}^{M_r} \lambda_m^r S(p_m^r, p_i) \quad (4.6)$$

$$S(p_{q'}, p_{q'}) = (1 - \Lambda_r)^2 S(p_q, p_q) + 2(1 - \Lambda_r) \sum_{m=1}^{M_r} \lambda_m^r S(p_q, p_m^r) + \sum_{m=1}^{M_r} \sum_{m'=1}^{M_r} \lambda_m^r \lambda_{m'}^r S(p_m^r, p_{m'}^r) \quad (4.7)$$

Η απόδειξη αυτών των σχέσεων δίνεται στην ενότητα A3 του παραρτήματος A.

Οι τιμές για τις ποσότητες  $S(p_j, p_i)$  μεταξύ των εικόνων της βάσης μπορούν να υπολογιστούν εκ των προτέρων, λόγω του ότι τα μοντέλα των εικόνων της βάσης δεν εξαρτώνται καθόλου από τη διαδικασία ανάκτησης εικόνων. Μέσα σε αυτές τις ποσότητες συμπεριλαμβάνονται, προφανώς, οι  $S(p_i, p_i)$  των παραπάνω τύπων, καθώς και οι  $S(p_m^r, p_i)$ ,

$S(p_m^r, p_m^r)$ , αφού συνήθως τα παραδείγματα  $r_m$  προέρχονται από τη χρησιμοποιούμενη βάση εικόνων. Επίσης, τόσο η τιμή για την ποσότητα  $S(p_q, p_q)$  μεταξύ της τρέχουσας ερώτησης και του εαυτού της, όσο και οι τιμές των ποσοτήτων  $S(p_q, p_i)$  (ομοίως, συμπεριλαμβάνονται και οι  $S(p_q, p_m^r)$ ) μεταξύ της τρέχουσας ερώτησης και κάθε μιας από τις εικόνες της βάσης μπορούν να αποθηκεύονται σε κάθε εποχή της ανατροφοδότησης. Από τα προηγούμενα γίνεται εύκολα αντιληπτό ότι μπορούμε να έχουμε όλα όσα χρειάζονται ώστε η απόκλιση C2 μεταξύ του νέου μοντέλου για την ερώτηση και των εικόνων της βάσης να είναι δυνατόν να υπολογιστεί ταχύτατα, με την εκτέλεση ως επί το πλείστον γραμμικών συνδυασμών μεταξύ ήδη υπολογισμένων τιμών του μέτρου S (όπως αυτό ορίζεται από τη σχέση (3.33)).

Στην πραγματικότητα, το νέο μοντέλο για την ερώτηση έχει μόνο θεωρητική αξία και δε χρειάζεται να κατασκευαστεί. Αυτό που απαιτείται είναι η χρήση των σχέσεων (4.5)-(4.7), ώστε να κρατάμε ενημερωμένη την τιμή της απόκλισης C2 μεταξύ της τρέχουσας ερώτησης και της κάθε εικόνας της βάσης, και σε αυτούς τους τύπους το ανανεωμένο μοντέλο εμπλέκεται μόνο έμμεσα.

Ακόμα, πρέπει να σημειωθεί εδώ ότι στις περιπτώσεις κατά τις οποίες οι τιμές  $S(p_j, p_i)$  για όλα τα ζεύγη εικόνων της βάσης δε μπορούν να υπολογιστούν ή/και να αποθηκευτούν, λόγω ελλείψεως χρόνου ή/και χώρου όταν η βάση είναι πολύ μεγάλη για παράδειγμα, αυτές οι ποσότητες μπορούν να υπολογιστούν πολύ γρήγορα κατά το χρόνο εκτέλεσης, αφού ορίζονται σε κλειστή μορφή για Γκαουσιανά Μείγματα (σχέσεις (3.34)-(3.36)). Έτσι, κρατώντας ενημερωμένες τις τιμές των  $S(p_q, p_i)$  και  $S(p_q, p_q)$  με χρήση των τύπων (4.6) και (4.7), και υπολογίζοντας κατά το χρόνο εκτέλεσης όποιες τιμές του μέτρου S μεταξύ εικόνων της βάσης μας είναι απαραίτητες, μπορούμε να ανανεώνουμε την απόκλιση C2 πάλι με αποδοτικό τρόπο και χωρίς ούτε αυτή τη φορά να χρειάζεται η πραγματική κατασκευή του νέου μοντέλου ερώτησης.

Μια ακόμα θετική ιδιότητα της παραπάνω μεθόδου συσχετιστικής ανατροφοδότησης είναι ότι η χρήση της τελευταίας δεν περιορίζεται μόνο στην περίπτωση που οι εικόνες της βάσης μοντελοποιούνται ως Γκαουσιανά Μείγματα, αλλά, αντίθετα, μπορεί να χρησιμοποιηθεί αυτούσια για οποιοδήποτε μοντέλο για το οποίο είναι δυνατός ο αποδοτικός υπολογισμός των τιμών του μέτρου S. Εξάλλου, για τον ορισμό των σχέσεων ανανέωσης της απόκλισης C2 πουθενά δεν υποθέσαμε ότι τα χρησιμοποιούμενα μοντέλα είναι Γκαουσιανά Μείγματα, αλλά το μόνο που χρησιμοποιήσαμε είναι ο γενικός ορισμός της απόκλισης C2 και η σχέση με βάση την οποία ανανεώνεται το μοντέλο της ερώτησης. Από αυτό γίνεται εύκολα αντιληπτό ότι η τεχνική ανανέωσης της απόκλισης C2 είναι ανεξάρτητη από το χρησιμοποιούμενο μοντέλο. Αυτήν την ιδιότητα θα την εκμεταλλευτούμε παρακάτω, προκειμένου να ορίσουμε μια παραλλαγή της μεθόδου που θα βασίζεται στη χρήση Ιστογραμμάτων, και η οποία θα χρησιμοποιηθεί για λόγους σύγκρισης.

### 4.3 Ενσωμάτωση των Αρνητικών Παραδειγμάτων

Μέχρι τώρα ασχοληθήκαμε με τα θετικά παραδείγματα που παρέχει στο σύστημα ο χρήστης, δηλαδή με τις εικόνες που αξιολογεί ως σχετικές με την ερώτησή του. Έχει αποδειχτεί, όμως, ότι σημαντική πληροφορία μπορούμε να πάρουμε και από τις εικόνες που ο χρήστης αξιολογεί ως μη σχετικές, δηλαδή από τα αρνητικά παραδείγματα που θα δώσει στο σύστημα.

Στη μέθοδο συσχετιστικής ανατροφοδότησης που προτείνουμε εδώ, τα αρνητικά παραδείγματα μπορούν να ληφθούν υπόψιν με έναν πολύ εύκολο και κομψό τρόπο. Εκτός από την ερώτηση που αντιστοιχεί στα θετικά παραδείγματα που παρέχει ο χρήστης, την οποία ονομάζουμε και «θετική» ερώτηση και της οποίας την ανανέωση μελετήσαμε εκτενώς παραπάνω, μπορούμε να ορίσουμε και μια «αρνητική» ερώτηση, για την ενσωμάτωση των αρνητικών παραδειγμάτων του χρήστη. Σε αντίθεση με τη θετική ερώτηση, η οποία αρχικοποιείται μέσω της αρχικής ερώτησης του χρήστη, δηλαδή μέσω της αρχικής εικόνας που αυτός δίνει ως παράδειγμα των προτιμήσεών του, η αρνητική ερώτηση είναι αρχικά «κενή», δηλαδή δεν περιέχει κάποιο αρνητικό παράδειγμα. Από την πρώτη εποχή της ανατροφοδότησης, όμως, όταν και ο χρήστης αρχίζει να παρέχει στο σύστημα την αξιολόγησή του για τα αποτελέσματα, τα αρνητικά παραδείγματα χρησιμοποιούνται για την ανανέωση της αρνητικής ερώτησης με έναν τρόπο εντελώς ανάλογο με αυτόν που υιοθετείται για την ανανέωση της θετικής ερώτησης.

Συγκεκριμένα, αν σε μια εποχή της ανατροφοδότησης ο χρήστης παρέχει στο σύστημα ένα σύνολο  $M_n$  αρνητικών παραδειγμάτων με μοντέλα  $p_m^n$ ,  $m=1, \dots, M_n$ , τότε, με το μοντέλο για την τρέχουσα αρνητική ερώτηση να συμβολίζεται ως  $p_n(\mathbf{x})$ , το ανανεωμένο μοντέλο της αρνητικής ερώτησης  $p_{n'}(\mathbf{x})$  δίνεται μέσω της σχέσης

$$p_{n'}(\mathbf{x}) = (1 - \Lambda_n) p_n(\mathbf{x}) + \sum_{m=1}^{M_n} \lambda_m^n p_m^n(\mathbf{x}) \quad (4.8)$$

όπου τα βάρη  $\lambda_m^n$  και  $\Lambda_n$  έχουν ανάλογη σημασία και ερμηνεία με αυτήν που δόθηκε νωρίτερα για τα βάρη  $\lambda_m^f$  και  $\Lambda_r$ , αντίστοιχα. Με λίγα λόγια και μιλώντας πρακτικά, το βάρος  $\lambda_m^n$  μπορεί να καθοριστεί από το χρήστη, και αποτελεί ένα μέτρο για το πόσο μακριά από τις προτιμήσεις του είναι το  $m$ -οστό αρνητικό παράδειγμα. Η ποσότητα  $1 - \Lambda_n$ , από την άλλη, είναι το βάρος με το οποίο λαμβάνεται υπόψιν η τρέχουσα αρνητική ερώτηση στο σχηματισμό του μοντέλου της νέας αρνητικής ερώτησης. Όταν δίνονται τα πρώτα αρνητικά παραδείγματα και η τρέχουσα αρνητική ερώτηση είναι «κενή», όπως αναφέραμε και παραπάνω, προφανώς θα ισχύει  $\Lambda_n = 1$ , και το ανανεωμένο μοντέλο θα εξαρτάται μόνο από τα δοθέντα αρνητικά παραδείγματα.

Με παρόμοιους συλλογισμούς με αυτούς που περιγράφηκαν προηγουμένως για την περίπτωση της ανανέωσης της απόκλισης C2 μεταξύ της θετικής ερώτησης και των εικόνων της βάσης, μπορούμε να καταλήξουμε σε μια ακριβώς αντίστοιχη διαδικασία μέσω της οποίας θα ανανεώνονται σε κάθε εποχή της ανατροφοδότησης οι τιμές της απόκλισης C2 μεταξύ της αρνητικής ερώτησης και των εικόνων της βάσης,  $C2(p_n, p_i)$ . Στην πραγματικότητα ισχύουν πάλι οι σχέσεις (4.5)-(4.7), αρκεί να αντικαταστήσουμε τα  $p_q$ ,  $p_q^f$ ,  $\lambda_m^f$ ,  $\Lambda_r$ ,  $M_r$ ,  $p_m^f$  με  $p_n$ ,  $p_{n'}$ ,  $\lambda_m^n$ ,  $\Lambda_n$ ,  $M_n$ ,  $p_m^n$ , αντίστοιχα. Προφανώς, το νέο μοντέλο για την αρνητική ερώτηση σε κάθε εποχή της ανατροφοδότησης δε χρειάζεται να κατασκευάζεται πραγματικά, για τους ίδιους λόγους που κάτι τέτοιο δεν είναι αναγκαίο ούτε για τη θετική ερώτηση.

Πρέπει να επισημανθεί εδώ ότι η υιοθέτηση ενός μοντέλου για την αρνητική ερώτηση που θα συσσωρεύει και θα ενσωματώνει τα αρνητικά παραδείγματα του χρήστη, δεν έχει στην περίπτωσή μας ανάλογο περιοριστικό χαρακτήρα με αυτόν που σχετίζεται με την υπόθεση απλής ομάδας, όπως αυτή αναλύθηκε στο κεφάλαιο 1. Η χρήση Γκαουσιανών Μειγμάτων εξασφαλίζει ότι κανείς ουσιαστικός περιορισμός δεν τίθεται όσο αφορά τον τρόπο κατανομής των χαρακτηριστικών των αρνητικών παραδειγμάτων. Δε γίνεται λοιπόν

καμιά απαράδεκτη υπόθεση ότι τα διανύσματα χαρακτηριστικών των αρνητικών παραδειγμάτων σχηματίζουν μια ομάδα στο χώρο των χαρακτηριστικών. Αντίθετα, μέσω του μοντέλου που επιλέξαμε για τις εικόνες και του τρόπου ανανέωσης της ερώτησης, αποφεύγουμε αυτήν την περιοριστική υπόθεση ακόμα και για τα θετικά παραδείγματα.

#### 4.4 Κριτήριο Διάταξης Εικόνων

Έχοντας τώρα ένα μοντέλο για τη θετική ερώτηση,  $p_q(\mathbf{x})$ , και ένα μοντέλο για την αρνητική ερώτηση,  $p_n(\mathbf{x})$ , καθώς και τις αντίστοιχες αποστάσεις  $C2(p_q, p_i)$  και  $C2(p_n, p_i)$  μεταξύ θετικής και αρνητικής ερώτησης και των εικόνων της βάσης, θέλουμε να ορίσουμε ένα μέτρο μέσω του οποίου θα δημιουργείται μια διάταξη των εικόνων της βάσης τέτοια που να αντικατοπτρίζει την εκτίμηση του συστήματος όσο αφορά το κατά πόσο ικανοποιεί η κάθε εικόνα τις προτιμήσεις του χρήστη.

Αυτό μπορεί να επιτευχθεί πολύ εύκολα μέσω των παρακάτω σχέσεων, που συνδυάζουν με απλό, γραμμικό τρόπο τις αποστάσεις που χωρίζουν την όποια εικόνα της βάσης από τη θετική και την αρνητική ερώτηση:

$$C(i) = a_r d(p_q, p_i) - (1 - a_r) d(p_n, p_i) \quad (4.9)$$

$$d(p_q, p_i) = C2(p_q, p_i) \quad (4.10)$$

$$d(p_n, p_i) = C2(p_n, p_i) \quad (4.11)$$

όπου  $0 \leq a_r \leq 1$  είναι το σχετικό βάρος που αποδίδεται στη θετική ερώτηση κατά τον υπολογισμό του κριτηρίου  $C$  για την  $i$ -οστή εικόνα της βάσης.

Μικρές τιμές για το προαναφερθέν κριτήριο αποτελούν ένδειξη ότι η αντίστοιχη εικόνα της βάσης είναι συμβατή με τις προτιμήσεις του χρήστη, αφού οι μικρές αυτές τιμές σημαίνουν ότι η εικόνα απέχει μικρή απόσταση από τη θετική ερώτηση και μεγάλη απόσταση από την αρνητική. Επομένως, αφού υπολογιστούν οι τιμές του κριτηρίου αυτού για όλες τις εικόνες της βάσης, μπορούν να επιστραφούν στο χρήστη, ως αποτελέσματα της ανάκτησης, οι εικόνες εκείνες με τις μικρότερες τιμές για το εν λόγω κριτήριο.

Για να λάβουμε τα πρώτα αποτελέσματα μετά την υποβολή της αρχικής ερώτησης του χρήστη, αφού, όπως είπαμε, σε αυτήν την αρχική φάση δεν υπάρχουν αρνητικά παραδείγματα και άρα αρνητική ερώτηση, θέτουμε  $a_r = 1$  και, έτσι, στον υπολογισμό των τιμών  $C(i)$  λαμβάνουμε υπόψιν μόνο το μοντέλο της θετικής ερώτησης και την απόσταση των μοντέλων των εικόνων της βάσης από αυτό.

#### 4.5 Μέθοδοι Συγκρινόμενες με την Προτεινόμενη

Προκειμένου να αξιολογήσουμε τη μεθοδολογία συσχειστικής ανατροφοδότησης που παρουσιάστηκε στις προηγούμενες ενότητες, σχεδιάσαμε και εκτελέσαμε ένα σύνολο πειραμάτων. Σε αυτά τα πειράματα προχωρήσαμε σε συγκρίσεις, τόσο από άποψη ποιότητας αποτελεσμάτων όσο και από άποψη ταχύτητας απόκρισης, με άλλες μεθόδους προσαρμοσμένες κατάλληλα, έτσι ώστε τα διάφορα δομικά στοιχεία της προτεινόμενης μεθοδολογίας να μπορούν να αξιολογηθούν επαρκώς. Εδώ παρουσιάζουμε τις διαφορές μεθόδους με τις οποίες συγκρίθηκε η προτεινόμενη μέθοδος.

### 4.5.1 Παραλλαγή Βασισμένη σε Ιστογράμματα

Ένα από τα κύρια στοιχεία της προτεινόμενης μεθοδολογίας είναι τα Γκαουσιανά Μείγματα, τα οποία χρησιμοποιούνται για τη μοντελοποίηση των εικόνων. Όπως, έχουμε ήδη αναφέρει, ένα άλλο δημοφιλές, διακριτό αυτή τη φορά μοντέλο, μέσω του οποίου μπορεί να περιγραφεί πιθανοτικά ένα σύνολο δεδομένων όπως τα τοπικά υπολογιζόμενα χαρακτηριστικά από μια εικόνα, είναι το Ιστόγραμμα. Στην ενότητα 4.2 καταστήσαμε σαφές ότι η προτεινόμενη τεχνική συσχετιστικής ανατροφοδότησης στην πραγματικότητα έχει οριστεί κατά τρόπο ανεξάρτητο από το χρησιμοποιούμενο μοντέλο για την περιγραφή των εικόνων. Μπορούμε, λοιπόν, για λόγους σύγκρισης με τα Γκαουσιανά Μείγματα και αξιολόγησης των αποτελεσμάτων που προκύπτουν από τη χρήση τους, να ορίσουμε μια παραλλαγή της προτεινόμενης μεθοδολογίας, στην οποία οι εικόνες θα μοντελοποιούνται μέσω Ιστογραμμάτων. Σε αυτήν την περίπτωση, το μόνο που πρέπει να ξεκαθαρίσουμε είναι ο τρόπος υπολογισμού της απόκλισης C2 μεταξύ Ιστογραμμάτων.

Στην περίπτωση των Ιστογραμμάτων, όπως αναλύσαμε στην ενότητα 2.2, ο χώρος των δεδομένων διαιρείται σε  $Z$  bins, το  $j$ -οστό από τα οποία συμβολίζεται με  $D_j$ . Για κάθε ένα από αυτά τα bins υπολογίζεται το ποσοστό του συνόλου των δεδομένων που του αντιστοιχεί,  $c_j^n$ . Συμβολίζοντας με  $\mathbf{x}_j^c$  το κέντρο του  $j$ -οστού bin στο χώρο των δεδομένων, το Ιστόγραμμα μπορεί να θεωρηθεί ότι αντιστοιχεί στην εξής συνάρτηση πυκνότητας πιθανότητας:

$$p(\mathbf{x}) = \sum_{j=1}^Z c_j^n \delta(\mathbf{x} - \mathbf{x}_j^c) \quad (4.12)$$

όπου η συνάρτηση  $\delta(\mathbf{x})$  είναι η κρουστική συνάρτηση Dirac. Για τη συγκεκριμένη συνάρτηση είναι γνωστό ότι ισχύουν οι εξής ιδιότητες:

$$\delta(\mathbf{x}) = 0, \quad \forall \mathbf{x} \neq \mathbf{0} \quad (4.13)$$

και

$$\int \delta(\mathbf{x}) d\mathbf{x} = 1 \quad (4.14)$$

Εάν έχουμε, λοιπόν, δύο Ιστογράμματα:

$$p_1(\mathbf{x}) = \sum_{j=1}^Z c_{1j}^n \delta(\mathbf{x} - \mathbf{x}_j^c) \quad \text{και} \quad p_2(\mathbf{x}) = \sum_{j=1}^Z c_{2j}^n \delta(\mathbf{x} - \mathbf{x}_j^c) \quad (4.15)$$

για τα οποία η διαμέριση του χώρου σε bins είναι κοινή, για να υπολογίσουμε την απόκλιση C2 μεταξύ τους αυτό που χρειαζόμαστε είναι ένας τρόπος υπολογισμού των εσωτερικών γινομένων  $S(p_m, p_l) = \int p_m(\mathbf{x}) p_l(\mathbf{x}) d\mathbf{x}$  για  $m, l = 1, 2$ . Αποδεικνύεται εύκολα ότι για την περίπτωση των παραπάνω Ιστογραμμάτων ισχύει

$$S(p_m, p_l) = \sum_{j=1}^Z c_{mj}^n c_{lj}^n \quad (4.16)$$

Επομένως, χρησιμοποιώντας τον παραπάνω τύπο για τον υπολογισμό του μέτρου S μεταξύ δύο Ιστογραμμάτων, έχουμε πλέον ό,τι μας χρειάζεται, και η τεχνική συσχετιστικής ανατροφοδότησης που παρουσιάστηκε στις προηγούμενες ενότητες μπορεί να χρησιμοποιηθεί αυτούσια στην περίπτωση που αντί για Γκαουσιανά Μείγματα οι εικόνες της βάσης αναπαριστάνονται ως Ιστογράμματα.

#### 4.5.2 Παραλλαγή Βασισμένη στο Μέτρο ALA

Ένα δεύτερο δομικό στοιχείο της προτεινόμενης μεθοδολογίας είναι η χρήση της απόκλισης C2 για τη σύγκριση των χρησιμοποιούμενων για την περιγραφή των εικόνων μοντέλων. Όπως αναφέρθηκε στην ενότητα 3.3, ένα άλλο μέτρο σύγκρισης που μπορεί να υπολογιστεί σε κλειστή μορφή μεταξύ Γκαουσιανών Μειγμάτων είναι το μέτρο ALA. Θα ορίσουμε εδώ, λοιπόν, μια παραλλαγή της προτεινόμενης μεθόδου χρησιμοποιώντας την ALA, με σκοπό να αξιολογήσουμε την επιτυχία της επιλογής της απόκλισης C2 για την προτεινόμενη μέθοδο.

Θεωρώντας τον προτεινόμενο τρόπο ανανέωσης του μοντέλου της ερώτησης σε κάθε εποχή της ανατροφοδότησης, όπως αυτός ορίζεται από τη σχέση (4.1), μπορεί να αποδεικτεί ότι για την ALA ισχύει η εξής σχέση (η απόδειξη δίνεται στην ενότητα A4 του παραρτήματος A):

$$ALA(p_q, \| p_i) = (1 - \Lambda_r) ALA(p_q \| p_i) + \sum_{m=1}^{M_r} \lambda_m^r ALA(p_m^r \| p_i) \quad (4.17)$$

Αυτή η σχέση μας αποκαλύπτει έναν τρόπο για να πετύχουμε, όπως και στην περίπτωση της χρήσης της απόκλισης C2, την αυξητική (incremental) ανανέωση του μέτρου σύγκρισης, ώστε αυτό να συμπεριλάβει κάθε φορά τα νέα παραδείγματα που δίνει ο χρήστης. Προφανώς, μια παρόμοια με την παραπάνω σχέση για τη θετική ερώτηση ισχύει και στην περίπτωση της αρνητικής ερώτησης. Συγκεκριμένα, η σχέση ανανέωσης της ALA για την αρνητική ερώτηση προκύπτει από την παραπάνω σχέση με απλή αντικατάσταση των  $p_q, p_{q'}, \lambda_m^r, \Lambda_r, M_r, p_m^r$  με  $p_n, p_{n'}, \lambda_m^n, \Lambda_n, M_n, p_m^n$ , αντίστοιχα.

Επομένως, αν χρησιμοποιήσουμε ως μέτρο σύγκρισης την ALA, έχουμε έναν τρόπο ανανέωσής της ακριβώς ανάλογο με αυτόν που αντιστοιχούσε στην περίπτωση της απόκλισης C2. Βεβαίως, πρέπει να λάβουμε υπόψιν ότι η ALA είναι μέτρο ομοιότητας και όχι μέτρο απόστασης, και για αυτό το λόγο χρειάζεται μια τροποποίηση του ορισμού των  $d(p_q, p_i)$  και  $d(p_n, p_i)$ , όπως αυτός δόθηκε με τις σχέσεις (4.10), (4.11), που χρησιμοποιούνται κατά τον υπολογισμό των τιμών του κριτηρίου διάταξης C για τις εικόνες της βάσης. Συγκεκριμένα, για τον υπολογισμό των παραπάνω ποσοτήτων μπορούμε να χρησιμοποιήσουμε τις σχέσεις

$$d(p_q, p_i) = -ALA(p_q \| p_i) \quad (4.18)$$

$$d(p_n, p_i) = -ALA(p_n \| p_i) \quad (4.19)$$

Με βάση αυτές τις μικρές τροποποιήσεις ορίζεται μια ακόμα παραλλαγή της προτεινόμενης μεθοδολογίας. Η συγκεκριμένη παραλλαγή χρησιμοποιεί το ίδιο μοντέλο για την αναπαράσταση των εικόνων όπως και η προτεινόμενη μεθοδολογία, δηλαδή τα Γκαουσιανά Μείγματα, χρησιμοποιεί την ίδια τεχνική ανανέωσης του μοντέλου της



ερώτησης από εποχή σε εποχή της ανατροφοδότησης, όμως, αντί για την απόκλιση C2, στηρίζεται στην ALA για τη σύγκριση μεταξύ των Γκαουσιανών Μειγμάτων.

### 4.5.3 Μέθοδος *Kherfi*

Εκτός από τις δύο παραλλαγές που παρουσιάστηκαν στις προηγούμενες ενότητες, με τις οποίες συγκρίνουμε την προτεινόμενη μέθοδο αποσκοπώντας στην αξιολόγηση των δύο πυλώνων της, που δεν είναι άλλοι από τα Γκαουσιανά Μείγματα και την απόκλιση C2, επιλέξαμε επίσης να συγκρίνουμε την προτεινόμενη τεχνική συσχετιστικής ανατροφοδότησης αυτή καθαυτή, με μια άλλη πρόσφατα προταθείσα τεχνική.

Η εν λόγω τεχνική συσχετιστικής ανατροφοδότησης παρουσιάζεται στην εργασία [22] και από εδώ και στο εξής θα αποκαλείται «μέθοδος *Kherfi*», από το όνομα του πρώτου συγγραφέα της συγκεκριμένης εργασίας, M.L. Kherfi.

Η μέθοδος *Kherfi*, λοιπόν, βασίζεται, όπως και η προτεινόμενη, στην τροποποίηση της αναπαράστασης της ερώτησης και του χρησιμοποιούμενου μέτρου απόστασης μετά από την υποβολή των παραδειγμάτων του χρήστη, ενώ έχει σχεδιαστεί έτσι ώστε να μπορεί να συμπεριλάβει και θετικά και αρνητικά παραδείγματα ανατροφοδότησης. Λαμβάνοντας μάλιστα υπόψιν ότι υιοθετεί, όπως και οι περισσότερες τεχνικές που έχουν προταθεί στη βιβλιογραφία, αναπαράσταση της κάθε εικόνας μέσω ενός περιορισμένου αριθμού πολυδιάστατων διανυσμάτων χαρακτηριστικών, η τελευταία δυνατότητα είναι αξιοσημείωτη, αφού, όπως έχουμε ήδη αναφέρει, είναι πολύ δύσκολη η ενσωμάτωση των αρνητικών παραδειγμάτων όταν χρησιμοποιείται μια τέτοια αναπαράσταση για τις εικόνες, και δε γίνεται παράλληλα χρήση κάποιου ταξινομητή. Επομένως, συμπεραίνουμε από την παραπάνω ανάλυση ότι τόσο η προτεινόμενη μέθοδος όσο και η μέθοδος *Kherfi* ανήκουν στην ίδια κατηγορία μεθόδων συσχετιστικής ανατροφοδότησης. Επίσης, μπορούν να ενσωματώσουν τα ίδια είδη παραδειγμάτων. Άρα, υπάρχουν όλα εκείνα τα αναγκαία στοιχεία συνάφειας ώστε η σύγκριση μεταξύ τους να είναι δίκαιη.

Η βάση πάνω στην οποία στηρίζει τη διαδικασία της συσχετιστικής ανατροφοδότησης η μέθοδος *Kherfi* είναι η βελτιστοποίηση Lagrange, την οποία χρησιμοποιεί προκειμένου να υπολογιστούν εκείνες οι παράμετροι που ελαχιστοποιούν την «εντός συνόλων» απόσταση (within class distance) των θετικών και των αρνητικών παραδειγμάτων και, παράλληλα, μεγιστοποιούν τη «μεταξύ συνόλων» απόσταση (between class distance) ανάμεσα στα θετικά και τα αρνητικά παραδείγματα.

Συγκεκριμένα, όπως αναφέραμε παραπάνω, κάθε εικόνα περιγράφεται από ένα σύνολο  $I$  πολυδιάστατων διανυσμάτων. Σε κάθε εποχή της ανατροφοδότησης υπολογίζονται οι παρακάτω διανυσματικοί μέσοι:

$$\mathbf{m}_i^1 = \frac{\sum_{n=1}^{N_1} \pi_n^1 \mathbf{x}_{ni}^1}{\sum_{n=1}^{N_1} \pi_n^1} \quad (4.20)$$

$$\mathbf{m}_i^2 = \frac{\sum_{n=1}^{N_2} \pi_n^2 \mathbf{x}_{ni}^2}{\sum_{n=1}^{N_2} \pi_n^2} \quad (4.21)$$

$$\mathbf{q}_i = \frac{\sum_{k=1}^2 \sum_{n=1}^{N_k} \pi_n^k \mathbf{x}_{ni}^k}{\sum_{k=1}^2 \sum_{n=1}^{N_k} \pi_n^k} \quad (4.22)$$

όπου  $i=1, \dots, I$ , και  $\mathbf{m}_i^1, \mathbf{m}_i^2, \mathbf{q}_i$  είναι ο σταθμισμένος μέσος των θετικών παραδειγμάτων, των αρνητικών παραδειγμάτων και του συνόλου των παραδειγμάτων, αντίστοιχα, για το  $i$ -οστό διανυσματικό χαρακτηριστικό. Επίσης,  $N_1$  ( $N_2$ ) είναι ο αριθμός των θετικών (αρνητικών) παραδειγμάτων που έχει δώσει ο χρήστης μέχρι και την τρέχουσα εποχή της ανατροφοδότησης,  $\mathbf{x}_{ni}^1$  ( $\mathbf{x}_{ni}^2$ ) είναι το  $i$ -οστό διανυσματικό χαρακτηριστικό για το  $n$ -οστό θετικό (αρνητικό) παράδειγμα, και  $\pi_n^1$  ( $\pi_n^2$ ) είναι ο βαθμός συνάφειας του  $n$ -οστού θετικού (αρνητικού) παραδείγματος (κάτι ανάλογο, δηλαδή, με τις παραμέτρους  $\lambda_m^r$  και  $\lambda_m^n$  της προτεινόμενης μεθόδου). Τα  $\pi_n^1, \pi_n^2$  είναι κανονικοποιημένα έτσι ώστε να ισχύουν οι παρακάτω σχέσεις:

$$\tilde{\pi}^k = \sum_{n=1}^{N_k} \pi_n^k \quad (4.23)$$

$$\tilde{\pi}^1 + \tilde{\pi}^2 = 1 \quad (4.24)$$

(η παράμετρος  $\tilde{\pi}^1$  αντιστοιχεί στην παράμετρο  $a_r$  της προτεινόμενης μεθόδου, αφού είναι στην ουσία το σχετικό βάρος των θετικών παραδειγμάτων).

Η «εντός συνόλων» απόσταση των θετικών και των αρνητικών παραδειγμάτων ορίζεται από τη σχέση

$$A = \sum_{i=1}^I u_i \sum_{k=1}^2 \sum_{n=1}^{N_k} \pi_n^k (\mathbf{x}_{ni}^k - \mathbf{m}_i^k)^T \mathbf{W}_i (\mathbf{x}_{ni}^k - \mathbf{m}_i^k) \quad (4.25)$$

και η «μεταξύ συνόλων» απόσταση ανάμεσα στα θετικά και τα αρνητικά παραδείγματα, από τη σχέση

$$R = \sum_{i=1}^I u_i \sum_{k=1}^2 \tilde{\pi}^k (\mathbf{m}_i^k - \mathbf{q}_i)^T \mathbf{W}_i (\mathbf{m}_i^k - \mathbf{q}_i) \quad (4.26)$$

με  $u_i$  να είναι το σχετικό βάρος του  $i$ -οστού διανυσματικού χαρακτηριστικού, και  $\mathbf{W}_i$  να είναι ένας συμμετρικός πίνακας βαρών μεταξύ των συνιστωσών του  $i$ -οστού διανυσματικού χαρακτηριστικού.

Η βελτιστοποίηση Lagrange είναι η τεχνική που χρησιμοποιείται προκειμένου να προσδιοριστούν τα  $u_i, \mathbf{W}_i$ , και το αντίστοιχο πρόβλημα προς επίλυση είναι η ελαχιστοποίηση μιας συνάρτησης των  $A, R$ , έστω  $L$ :

$$\min L(A, R) \text{ ως προς } u_1, \dots, u_I, \mathbf{W}_1, \dots, \mathbf{W}_I \quad (4.27)$$

(για παράδειγμα  $L(A, R) = \frac{A}{R}$  ή  $L(A, R) = A - R$ ) υπό τους περιορισμούς

$$\sum_{i=1}^I \frac{1}{u_i} = 1 \quad \text{και} \quad |\mathbf{W}_i| = 1, \quad i = 1, \dots, I \quad (4.28)$$

Μετά από τον προσδιορισμό των  $u_i$  και  $\mathbf{W}_i$ , η διάταξη των εικόνων της βάσης γίνεται μέσω του υπολογισμού ενός μέτρου απόστασης της μορφής

$$D(\mathbf{x}_n) = \sum_{i=1}^I u_i (\mathbf{x}_{ni} - \mathbf{m}_i^1)^T \mathbf{W}_i (\mathbf{x}_{ni} - \mathbf{m}_i^1) - \sum_{i=1}^I u_i (\mathbf{x}_{ni} - \mathbf{m}_i^2)^T \mathbf{W}_i (\mathbf{x}_{ni} - \mathbf{m}_i^2) \quad (4.29)$$

όπου  $\mathbf{x}_n = (\mathbf{x}_{n1}, \dots, \mathbf{x}_{nI})$  είναι τα  $I$  διανύσματα χαρακτηριστικών της  $n$ -οστής εικόνας της βάσης. Όσο μικρότερη είναι η τιμή αυτού του μέτρου για μια εικόνα της βάσης, δηλαδή, στην πράξη, όσο λιγότερο απέχει αυτή από τα θετικά παραδείγματα και όσο περισσότερο απέχει από τα αρνητικά, τόσο πιο σχετική με την ερώτηση του χρήστη θεωρείται από το σύστημα.

Η ανάκτηση εκτελείται σε δύο στάδια. Στο πρώτο στάδιο μόνο τα θετικά παραδείγματα λαμβάνονται υπόψιν, και η διαδικασία βελτιστοποίησης και η διάταξη των εικόνων της βάσης γίνεται ως να μην υπήρχαν καθόλου αρνητικά παραδείγματα. Από αυτή τη διαδικασία, ένας περιορισμένος αριθμός εικόνων που βρίσκονται κοντά στα θετικά παραδείγματα διατηρούνται. Οι εικόνες αυτές στο δεύτερο στάδιο, το οποίο λαμβάνει υπόψιν και τα θετικά και τα αρνητικά παραδείγματα, επαναδιατάσσονται και ένας ακόμα μικρότερος αριθμός από αυτές, εκείνες που βρίσκονται πολύ κοντά στα θετικά παραδείγματα και πολύ μακριά από τα αρνητικά, επιστέφονται στο χρήστη ως αποτελέσματα της ανάκτησης.

## 4.6 Σύνολα Δεδομένων

Για να καταστούν δυνατές οι συγκρίσεις των διαφόρων μεθοδολογιών χρειάζονται προφανώς κατάλληλα σύνολα δεδομένων (datasets) πάνω στα οποία αυτές θα εφαρμοστούν. Έτσι, σε αυτήν την ενότητα περιγράφονται τόσο οι βάσεις εικόνων που χρησιμοποιήθηκαν για την αξιολόγηση της προτεινόμενης μεθόδου, όσο και η διαδικασία υπολογισμού οπτικών χαρακτηριστικών για κάθε μια από τις εικόνες των βάσεων.

### 4.6.1 Βάσεις Εικόνων

Όπως έχουμε αναφέρει στην ενότητα 1.4, η αξιολόγηση των συστημάτων ανάκτησης εικόνων γίνεται συνήθως πάνω σε εκ των προτέρων κατηγοριοποιημένες βάσεις εικόνων, δηλαδή σε βάσεις των οποίων οι εικόνες έχουν καταταξιωθεί «με το χέρι» (manually) σε έναν αριθμό από συγκεκριμένες και γνωστές κατηγορίες, με κριτήριο την όσο το δυνατόν καλύτερη από σημασιολογικής άποψης περιγραφή του περιεχομένου της εκάστοτε εικόνας.

Για τις συγκρίσεις μεταξύ των μεθοδολογιών που παρουσιάστηκαν παραπάνω θα χρησιμοποιήσουμε δύο τέτοιες βάσεις εικόνων:

1. MSDB: Η πρώτη βάση εικόνων περιέχει 3740 εικόνες καθεμιά από τις οποίες έχει ανάλυση  $640 \times 480$  pixels. Αυτές οι εικόνες αποτελούν υποσύνολο της βάσης εικόνων [30] και κατανέμονται σε 17 σημασιολογικές κατηγορίες σύμφωνα με το περιεχόμενό τους. Κάθε μια από αυτές τις κατηγορίες περιγράφεται από μια ετικέτα (label) όπως «αεροπλάνα», «αυτοκίνητα», «πουλιά», «παράθυρα» κτλ. Εν γένει,

υιοθετήσαμε την κατηγοριοποίηση που καθορίζεται από τον πάροχο της βάσεως, εκτός από τις περιπτώσεις λίγων κατηγοριών που ενοποιήθηκαν λόγω του ότι σημασιολογικά βρίσκονταν πολύ κοντά η μία στην άλλη.

2. Corel9923: Η δεύτερη βάση εικόνων που χρησιμοποιήσαμε είναι ένα υποσύνολο της δημοφιλούς βάσης εικόνων Corel, που αποτελεί μια από τις πιο συχνές επιλογές όσο αφορά την αξιολόγηση συστημάτων ανάκτησης εικόνων. Συγκεκριμένα, το εν λόγω υποσύνολο περιέχει 9923 εικόνες, καθεμιά από τις οποίες έχει μέγεθος  $256 \times 384$  pixels. Οι εικόνες αυτές κατανέμονται σε 42 σημασιολογικές κατηγορίες. Παρά το ότι η βάση εικόνων Corel είναι επαγγελματικά επιγεγραμμένη (annotated) και κατηγοριοποιημένη, παρουσιάζεται συχνά το φαινόμενο εικόνες με παρεμφερές σημασιολογικό περιεχόμενο να μοιράζονται σε διαφορετικές κατηγορίες της Corel. Έτσι, προχωρήσαμε στη συγχώνευση διαφόρων κατηγοριών, δημιουργώντας με αυτόν τον τρόπο τη δική μας σημασιολογική κατηγοριοποίηση.

Για περισσότερες πληροφορίες σχετικά με τις παραπάνω βάσεις εικόνων ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει και στο παράρτημα Β.

Η κατηγοριοποίηση που υιοθετείται τελικά για καθεμιά από τις βάσεις αυτές, θα θεωρείται στις προσομοιώσεις που θα ακολουθήσουν ως η ορθή κατηγοριοποίηση (ground truth), με την οποία θα συγκρίνονται τα αποτελέσματα της όποιας μεθοδολογίας.

#### 4.6.2 Οπτικά Χαρακτηριστικά

Όπως έχουμε τονίσει και σε προηγούμενο κεφάλαιο, πριν από τη μοντελοποίηση μιας εικόνας μέσω ενός Γκαουσιανού Μείγματος, για παράδειγμα, απαιτείται ο υπολογισμός ενός αριθμού διανυσμάτων χαρακτηριστικών, καθένα από τα οποία θα περιέχει πληροφορία για το οπτικό περιεχόμενο (visual content) της εικόνας σε τοπικό επίπεδο. Όσο αφορά τα πειράματα που ακολουθούν, επιλέξαμε να περιγράψουμε το τοπικό περιεχόμενο των χρησιμοποιούμενων εικόνων μέσω τριών ειδών χαρακτηριστικών:

1. Χαρακτηριστικά θέσης.
2. Χαρακτηριστικά χρώματος.
3. Χαρακτηριστικά υφής.

Για τον υπολογισμό των χαρακτηριστικών αυτών ακολουθήσαμε σε γενικές γραμμές τη μέθοδο υπολογισμού χαρακτηριστικών της εργασίας [5]. Παρακάτω, θα παρουσιάσουμε συνοπτικά τα χαρακτηριστικά που προκύπτουν με αυτόν τον τρόπο, και τη διαδικασία υπολογισμού τους.

Η διαδικασία υπολογισμού των χαρακτηριστικών θέσης είναι απλή. Συγκεκριμένα, ως χαρακτηριστικά θέσης χρησιμοποιούμε τις συντεταγμένες  $(x, y)$  για κάθε pixel της εικόνας, μετά από κανονικοποίηση μέσω διαίρεσης με το πλάτος της εικόνας για τη διάσταση  $x$ , και με το ύψος της εικόνας για τη διάσταση  $y$ .

#### Χαρακτηριστικά Χρώματος

Ως χαρακτηριστικά χρώματος για κάθε pixel χρησιμοποιούμε τις αντίστοιχες χρωματικές συνιστώσες στο χώρο LAB, ο οποίος, όπως έχουμε αναλύσει στην ενότητα 1.1.2, θεωρείται ότι ανταποκρίνεται καλύτερα στην αντίληψη του μέσου χρήστη περί χρωματικής απόστασης σε σχέση με άλλους χώρους χρώματος. Η ιδιότητα αυτή είναι

ιδιαίτερα ευεργετική όσο αφορά τα προβλήματα ανάκτησης εικόνων, όπου τελικός κριτής είναι πάντα ο χρήστης.

Για να πάρουμε τις τελικές τιμές,  $c_s^L$ ,  $c_s^A$ ,  $c_s^B$ , των χρωματικών χαρακτηριστικών για κάθε pixel (που αντιστοιχούν στις συνιστώσες L, A και B, αντίστοιχα), φιλτράρουμε τοπικά την εικόνα μέσω ενός Γκαουσιανού πυρήνα το πλάτος του οποίου καθορίζεται από την τοπική τιμή της κλίμακας υφής (texture scale). Περί κλίμακας υφής θα μιλήσουμε αναλυτικότερα παρακάτω, όταν θα περιγράψουμε τη διαδικασία υπολογισμού των χαρακτηριστικών υφής. Εδώ πρέπει να σημειώσουμε ότι το προαναφερθέν φιλτράρισμα αποσυμπλέκει την πληροφορία χρώματος από την πληροφορία υφής.

Η διαδικασία του φιλτραρίσματος υλοποιείται μέσω διακριτής συνέλιξης (convolution) με πυρήνα μεταβλητού πλάτους, η οποία στην περίπτωσή μας ορίζεται από την ακόλουθη σχέση:

$$g(x, y) = (\tilde{h} * f)(x, y) = \sum_{(w_1, w_2) \in D(x, y)} \tilde{h}(w_1, w_2 | x, y) f(x - w_1, y - w_2) \quad (4.30)$$

όπου  $\tilde{h}$ ,  $f$  είναι οι συναρτήσεις που συνελίσσονται και  $g(x, y)$  το αποτέλεσμα της πράξης της συνέλιξης (\*) στο σημείο  $(x, y)$ . Με  $\tilde{h}$  συμβολίζουμε μια διακριτή, κανονικοποιημένη προσέγγιση του δισδιάστατου, ισοτροπικού (δηλαδή με  $\Sigma = \sigma^2 \mathbf{I}$ ) και εστιασμένου στην αρχή των αξόνων (δηλαδή με  $\mu = \mathbf{0}$ ) Γκαουσιανού πυρήνα:

$$h(w_1, w_2 | x, y) = \frac{1}{2\pi\sigma^2(x, y)} \exp\left[-\frac{w_1^2 + w_2^2}{2\sigma^2(x, y)}\right] \quad (4.31)$$

η οποία δίνεται από τον τύπο

$$\tilde{h}(w_1, w_2 | x, y) = \frac{h(w_1, w_2 | x, y)}{\sum_{(s_1, s_2) \in D(x, y)} h(s_1, s_2 | x, y)}, \quad \forall (w_1, w_2) \in D(x, y) \quad (4.32)$$

Για το σύνολο σημείων  $D(x, y)$  ισχύει ότι

$$(w_1, w_2) \in D(x, y) \Leftrightarrow w_1, w_2 \in \mathbf{N} \quad \text{και} \quad |w_i| \leq 3\sigma(x, y) \quad (4.33)$$

Δηλαδή, με απλά λόγια, το σύνολο  $D(x, y)$  περιλαμβάνει εκείνες τις δυνάδες ακεραίων για τις οποίες η συνάρτηση  $h(w_1, w_2 | x, y)$  παίρνει μη αμελητέες τιμές. Η συνάρτηση  $f$  είναι μία από τις δισδιάστατες συναρτήσεις  $L(x, y)$ ,  $A(x, y)$  και  $B(x, y)$  των χρωματικών συνιστωσών για τα pixels της εικόνας, ανάλογα με το αν θέλουμε να υπολογίσουμε τις τιμές για τα χαρακτηριστικά  $c_s^L$ ,  $c_s^A$  ή  $c_s^B$ , αντίστοιχα.

## Χαρακτηριστικά Υφής

Η υφή είναι μια ιδιότητα που χαρακτηρίζει περιοχές μιας εικόνας, και έχει να κάνει με τοπικές διακυμάνσεις της φωτεινότητας οι οποίες σχηματίζουν επαναλαμβανόμενα μοτίβα (patterns). Για παράδειγμα, εικόνες με έντονα χαρακτηριστικά υφής είναι αυτές που

απεικονίζουν το σώμα μιας τίγρης ή μιας ζέβρας, που εμφανίζει εκείνες τις χαρακτηριστικές ρίγες, εικόνες που αποτυπώνουν τις ρυτιδώσεις πάνω σε μια ξύλινη επιφάνεια κτλ.

Λαμβάνοντας υπόψιν ότι η υφή συνδέεται πάντα, όπως είπαμε, με περιοχές μιας εικόνας, η μέθοδος που περιγράφεται στην εργασία [5] προσπαθεί να προσδιορίσει μια «γειτονιά» (neighborhood) όσο αφορά την υφή, μέσω της επιλογής μιας κατάλληλης κλίμακας υφής. Συγκεκριμένα, για κάθε θέση pixel  $(x, y)$  υπολογίζεται η κλίση (gradient) της συνιστώσας L του χώρου LAB (που αντιστοιχεί στη φωτεινότητα), έστω  $\nabla L(x, y)$ , η οποία ορίζεται ως

$$\nabla L = \left( \frac{\partial L}{\partial x}, \frac{\partial L}{\partial y} \right)^T \quad (4.34)$$

Στην πράξη ο υπολογισμός γίνεται προσεγγιστικά, υπολογίζοντας τις διαφορές στην τιμή της συνιστώσας L για διπλανά pixels σε κάθε διάσταση.

Επίσης, θεωρώντας έναν αριθμό από κλίμακες:

$$\sigma_k = k/2 \quad \text{με } k = 0, 1, \dots, 7 \quad (4.35)$$

υπολογίζεται σε κάθε  $(x, y)$  ένας ανάλογος αριθμός από πίνακες διάστασης  $2 \times 2$ :

$$\mathbf{M}_{\sigma_k}(x, y) = \tilde{h}_{\sigma_k} * (\nabla L)(\nabla L)^T \quad (4.36)$$

με την  $\tilde{h}_{\sigma_k}$  να αποτελεί διακριτή προσέγγιση Γκαουσιανού πυρήνα, παρόμοια με αυτή που περιγράφηκε παραπάνω για τα χρωματικά χαρακτηριστικά, με τη μόνη διαφορά ότι εδώ το πλάτος του πυρήνα ισούται με  $\sigma_k$  και δεν εξαρτάται από το  $(x, y)$ . Έπειτα, με ανάλυση του πίνακα  $\mathbf{M}_{\sigma_k}(x, y)$  σε ιδιοτιμές και ιδιοδιανύσματα προσδιορίζεται η κύρια κατεύθυνση, δηλαδή η κατεύθυνση του ιδιοδιανύσματος που αντιστοιχεί στη μεγαλύτερη ιδιοτιμή, για τη θέση  $(x, y)$  και στην κλίμακα  $\sigma_k$ .

Χρησιμοποιώντας αυτήν την κατεύθυνση, έστω  $\hat{\mathbf{n}}$ , υπολογίζεται η τιμή της πόλωσης (polarity) για το συγκεκριμένο pixel και τη συγκεκριμένη κλίμακα, μέσω της σχέσης

$$P_{\sigma_k}(x, y) = \frac{|E_+^k(x, y) - E_-^k(x, y)|}{E_+^k(x, y) + E_-^k(x, y)} \quad (4.37)$$

όπου

$$E_+^k(x, y) = (\tilde{h}_{\sigma_k} * \max(\nabla L \cdot \hat{\mathbf{n}}, 0))(x, y) \quad (4.38)$$

$$E_-^k(x, y) = -(\tilde{h}_{\sigma_k} * \min(\nabla L \cdot \hat{\mathbf{n}}, 0))(x, y) \quad (4.39)$$

Ουσιαστικά, το  $E_+^k(x, y)$  ( $E_-^k(x, y)$ ) είναι ένα μέτρο για το πόσα διανύσματα κλίσης στη γειτονιά του  $(x, y)$ , όπως αυτή καθορίζεται από την κλίμακα  $\sigma_k$ , βρίσκονται στην ίδια

(αντίθετη) κατεύθυνση με το διάνυσμα  $\hat{\mathbf{n}}$ , που ορίζει την κύρια κατεύθυνση για αυτό το pixel και για αυτή την κλίμακα.

Μετά από όλα τα παραπάνω, και αφού εξομαλύνουμε για κάθε  $\sigma_k$  τις αντίστοιχες τιμές πόλωσης για όλα τα pixels μέσω φιλτραρίσματος με ένα Γκαουσιανό πυρήνα πλάτους  $2\sigma_k$ ,  $\tilde{h}_{2\sigma_k}$ , καταλήγουμε με τόσες τιμές για την εξομαλυμένη πόλωση σε κάθε pixel, όσες είναι και οι διαφορετικές τιμές  $\sigma_k$  για την κλίμακα. Χρησιμοποιώντας αυτές τις τιμές για την πόλωση, επιλέγουμε ως αντιπροσωπευτική κλίμακα του pixel την πρώτη για την οποία η διαφορά μεταξύ των τιμών πόλωσης για δύο διαδοχικές κλίμακες πέφτει κάτω από το όριο του 2%, δηλαδή εκείνη την κλίμακα για την οποία η πόλωση αρχίζει να σταθεροποιείται. Η κλίμακα αυτή είναι η κλίμακα υφής για το συγκεκριμένο pixel, αυτή που χρησιμοποιείται ως  $\sigma(x, y)$  και για το υπολογισμό των χαρακτηριστικών χρώματος που περιγράψαμε παραπάνω.

Έστω ότι  $P_s$  είναι η ομαλοποιημένη πόλωση για την επιλεγείσα κλίμακα στο pixel με θέση  $(x, y)$ . Αυτή, μαζί με την ανισοτροπία (anisotropy):

$$A_s = 1 - \frac{\lambda_2}{\lambda_1} \quad (4.40)$$

και την αντίθεση (contrast):

$$C_s = 2\sqrt{\lambda_1 + \lambda_2} \quad (4.41)$$

που αντιστοιχούν στο εν λόγω pixel, όπου  $\lambda_1$  και  $\lambda_2$  είναι η μεγάλη και η μικρή ιδιοτιμή για τον πίνακα  $\mathbf{M}_{\sigma_{k^*}}(x, y)$ , με το  $\sigma_{k^*}$  να αντιστοιχεί στην επιλεγείσα κλίμακα, θα χρησιμοποιηθούν για τον υπολογισμό των τελικών χαρακτηριστικών υφής για το pixel στη θέση  $(x, y)$ .

Συγκεκριμένα, ως χαρακτηριστικά υφής επιλέγουμε την αντίθεση,  $t_s^C = C_s$ , το γινόμενο της ανισοτροπίας και της αντίθεσης,  $t_s^{AC} = A_s C_s$ , και το γινόμενο της πόλωσης και της αντίθεσης,  $t_s^{PC} = P_s C_s$ . Η επιλογή αυτή υπαγορεύεται από το γεγονός ότι οι τιμές της ανισοτροπίας και της πόλωσης είναι χωρίς νόημα σε περιοχές με χαμηλή αντίθεση.

Επομένως, τελικά, για κάθε pixel της εικόνας καταλήγουμε να έχουμε ως διάνυσμα χαρακτηριστικών ένα 8-διάστατο διάνυσμα  $\mathbf{x}$ :

$$\mathbf{x} = (f_1, f_2, \dots, f_8)^T = (x^n, y^n, c_s^L, c_s^A, c_s^B, t_s^{AC}, t_s^{PC}, t_s^C)^T \quad (4.42)$$

με

$$x^n = \frac{x}{x_{\max}} \quad (4.43)$$

$$y^n = \frac{y}{y_{\max}} \quad (4.44)$$

όπου  $x_{\max}$  και  $y_{\max}$  είναι το πλάτος και το ύψος, αντίστοιχα, της εικόνας σε pixels.

## Υποδειματοληψία

Πριν από τον υπολογισμό των παραπάνω χαρακτηριστικών για τις εικόνες των βάσεων που παρουσιάστηκαν στην ενότητα 4.6.1, πραγματοποιήθηκε υποδειματοληψία (subsampling) για τη μείωση του συνολικού αριθμού των pixels της κάθε εικόνας. Αυτό κρίθηκε απαραίτητο, ώστε να περιοριστούν οι υπολογιστικές απαιτήσεις τόσο κατά το στάδιο του υπολογισμού των χαρακτηριστικών, όσο και κατά το στάδιο της εκπαίδευσης των Γκαουσιανών Μειγμάτων.

Για την υποδειματοληψία αυτή χρησιμοποιήθηκε ένα χωρικά ομοιόμορφο πλέγμα (grid). Το μέγεθος αυτού του πλέγματος καθορίζεται από το ποσοστό των pixels της εικόνας που θέλουμε να διατηρηθούν μετά την υποδειματοληψία. Για παράδειγμα, αν μια εικόνα έχει ανάλυση  $N_x \times N_y$  pixels και θέλουμε να διατηρήσουμε ένα ποσοστό (περίπου)  $a$  αυτών, με  $0 < a < 1$ , τότε το πλέγμα θα έχει μέγεθος  $\lfloor \sqrt{a}N_x \rfloor \times \lfloor \sqrt{a}N_y \rfloor$  (το  $\lfloor x \rfloor$  σημαίνει στρογγυλοποίηση του  $x$  στον αμέσως μικρότερο ακέραιο). Από τον πίνακα των pixels της εικόνας θα επιλεγούν εκείνες οι γραμμές (στήλες) ξεκινώντας από την πρώτη και προχωρώντας με ομοιόμορφο βήμα, ώστε φτάνοντας στην τελευταία να έχουμε επιλέξει προς διατήρηση  $\lfloor \sqrt{a}N_y \rfloor$  γραμμές ( $\lfloor \sqrt{a}N_x \rfloor$  στήλες).

Πριν από τη διαδικασία της υποδειματοληψίας, φιλτράρουμε την κάθε εικόνα με ένα Γκαουσιανό πυρήνα μικρού πλάτους, ώστε να την ομαλοποιήσουμε κατά τρόπο που να αποφύγουμε το φαινόμενο της αναδίπλωσης φάσματος (aliasing). Αυτό το φαινόμενο παρατηρείται συχνά σε περιπτώσεις δειγματοληψίας σημάτων (όπως είναι εν προκειμένω οι εικόνες), λόγω του ότι το φασματικό περιεχόμενο, στο χώρο της συχνότητας, του σήματος που υφίσταται δειγματοληψία εμφανίζεται σε πολλαπλά αντίγραφα γύρω από τα ακέραια πολλαπλάσια της συχνότητας δειγματοληψίας. Αυτό έχει ως αποτέλεσμα να υπάρχει επικάλυψη των αντιγράφων για τις υψηλές συχνότητες. Μέσω του Γκαουσιανού φιλτραρίσματος που αναφέραμε παραπάνω καταπιέζονται αυτές ακριβώς οι υψηλές συχνότητες του αρχικού φάσματος, με αποτέλεσμα να μειώνεται η επικάλυψη μετά την υποδειματοληψία.

Συγκεκριμένα, όσο αφορά την πρώτη βάση, MSDB, λόγω του ότι οι εικόνες που αυτή περιέχει έχουν σχετικά μεγάλη ανάλυση, μόνο το 15% των pixels κάθε εικόνας διατηρήθηκαν μετά την υποδειματοληψία. Αντίθετα, για τις εικόνες της δεύτερης βάσης, Corel9923, εξαιτίας της μικρότερης ανάλυσής τους αποφασίσαμε να διατηρήσουμε το 50% των pixels της καθεμίας.

## 4.7 Θέματα Υλοποίησης

Στην παρούσα ενότητα αναλύονται ένα σύνολο από θέματα που αφορούν την υλοποίηση τόσο της προτεινόμενης μεθόδου όσο και των άλλων μεθόδων που χρησιμοποιούνται για λόγους σύγκρισης με αυτήν, και παρουσιάζονται οι συγκεκριμένες επιλογές που υιοθετήθηκαν στα πλαίσια των πειραμάτων που διεξήχθησαν.

### 4.7.1 Προτεινόμενη Μέθοδος

#### 4.7.1.1 Εκπαίδευση των Γκαουσιανών Μειγμάτων

Όσο αφορά τα Γκαουσιανά Μείγματα και τη μοντελοποίηση μέσω αυτών των εικόνων των βάσεων, για την εκπαίδευσή τους χρησιμοποιήθηκε ο άπληστος αλγόριθμος EM, που



παρουσιάστηκε στην ενότητα 2.4.3.2. Για κάθε μοντέλο θεωρήσαμε ότι οι πίνακες συνδιακύμανσης των συνιστωσών του είναι πλήρεις (full covariance), δηλαδή δεν κάναμε κάποια υπόθεση απλοποίησης, όπως για παράδειγμα ότι οι παραπάνω πίνακες είναι διαγώνιοι. Σε ό,τι έχει να κάνει με τον αριθμό των συνιστωσών κάθε μοντέλου, έγινε προσπάθεια να χρησιμοποιηθεί κάποιο κριτήριο επιλογής μοντέλου, όπως αυτά που παρουσιάστηκαν στην ενότητα 2.4.4. Όμως, το μοντέλο που επιλέγεται από αυτά τα κριτήρια είναι πάντα υπερβολικά πολύπλοκο, και επομένως η επιλογή αυτή έχει σημαντικό κόστος όσο αφορά το χρόνο εκπαίδευσης του μοντέλου, ενώ, παράλληλα, επιβαρύνει αδικαιολόγητα και την εφαρμογή των μεθόδων που χρησιμοποιούν τα Γκαουσιανά Μείγματα. Αναγκαστήκαμε εκ των πραγμάτων, λοιπόν, να προχωρήσουμε σε μια εμπειρική επιλογή για τον αριθμό των συνιστωσών του κάθε μείγματος. Μετά από έναν αριθμό προκαταρκτικών πειραμάτων με μοντέλα διαφορετικού αριθμού συνιστωσών, καταλήξαμε στο συμπέρασμα ότι η χρησιμοποίηση μειγμάτων 10 Γκαουσιανών συνιστωσών για κάθε εικόνα αντιστοιχεί σε μια ικανοποιητική επιλογή, με την έννοια ότι οδηγεί σε αρκετά ικανοποιητικά αποτελέσματα, ενώ συγχρόνως διατηρεί την πολυπλοκότητα των μοντέλων σε ένα επίπεδο τέτοιο που να μην επιβαρύνεται υπερβολικά η εκπαίδευση και ο χειρισμός τους.

Πριν εφαρμόσουμε τον άπληστο αλγόριθμο EM πάνω στα διανύσματα χαρακτηριστικών που αντιστοιχούν σε κάποια εικόνα για να πάρουμε το εκπαιδευμένο μοντέλο που θα την περιγράφει, κανονικοποιούμε τις τιμές του κάθε χαρακτηριστικού. Η κανονικοποίηση γίνεται κατά τέτοιο τρόπο ώστε οι κανονικοποιημένες τιμές να έχουν μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση. Συγκεκριμένα, αν  $\mathbf{x}_i = (x_{i1}, \dots, x_{id})^T$ , με  $i = 1, \dots, N$ , είναι τα  $N$  διανύσματα χαρακτηριστικών της εικόνας, τότε για τις συντεταγμένες των κανονικοποιημένων διανυσμάτων  $\mathbf{x}_i^n = (x_{i1}^n, \dots, x_{id}^n)^T$  ισχύει

$$x_{ik}^n = \frac{x_{ik} - \tilde{\mu}_k}{\tilde{\sigma}_k} \quad (4.45)$$

με τη στατιστική μέση τιμή για το  $k$ -οστό χαρακτηριστικό να ισούται με

$$\tilde{\mu}_k = \frac{1}{N} \sum_{i=1}^N x_{ik} \quad (4.46)$$

και τη στατιστική τυπική απόκλιση να ισούται με

$$\tilde{\sigma}_k = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_{ik} - \tilde{\mu}_k)^2} \quad (4.47)$$

Αυτή η κανονικοποίηση γίνεται για να εξασφαλίσουμε ότι κανένα χαρακτηριστικό δε θα παίξει λόγω διαφορών όσο αφορά το εύρος τιμών μεγαλύτερο ρόλο από τα άλλα στην εκπαίδευση του μοντέλου. Μετά την εκπαίδευση, όμως, για να μπορεί να γίνει σύγκριση μεταξύ των μοντέλων για τις διάφορες εικόνες πρέπει απαραίτητως τα εκπαιδευμένα μοντέλα να απεικονιστούν στον αρχικό χώρο, δηλαδή αυτόν πριν την κανονικοποίηση. Αυτός ο αντίστροφος μετασχηματισμός είναι εξίσου εύκολος και, θεωρώντας τη μέση τιμή και τον πίνακα συνδιακύμανσης για την όποια Γκαουσιανή συνιστώσα του μείγματος,  $\boldsymbol{\mu}^n = (\mu_1^n, \dots, \mu_d^n)^T$ ,  $\boldsymbol{\Sigma}^n = [s_{kl}^n]$  πριν, και  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)^T$ ,  $\boldsymbol{\Sigma} = [s_{kl}]$  μετά την εφαρμογή του, ορίζεται από τους παρακάτω τύπους:

$$\mu_k = \tilde{\sigma}_k \mu_k^n + \tilde{\mu}_k \quad (4.48)$$

$$s_{kl} = \tilde{\sigma}_k \tilde{\sigma}_l s_{kl}^n \quad (4.49)$$

#### 4.7.1.2 Ανανέωση του Μοντέλου της Ερώτησης

Η ανανέωση των μοντέλων για τις ερωτήσεις γίνεται μέσω των τύπων (4.1) και (4.8). Στον προσδιορισμό των τιμών για τις παραμέτρους  $\lambda_m^r$  και  $\lambda_m^n$ , που χρησιμοποιούνται από την προτεινόμενη μέθοδο για τον ορισμό του ανανεωμένου μοντέλου ερώτησης, πρέπει να ληφθεί υπόψιν ότι κατά την αξιολόγηση των μεθόδων στηριζόμαστε σε εκ των προτέρων κατηγοριοποιημένες βάσεις εικόνων. Οι εικόνες σε αυτήν την περίπτωση κατανέμονται σε συγκεκριμένες κατηγορίες, και επομένως δε μπορεί κατά τις προσομοιώσεις να αποδοθεί διαφορετικός βαθμός συνάφειας σε κάθε παράδειγμα, διότι κάτι τέτοιο δεν είναι συμβατό με την παραπάνω κατηγοριοποίηση. Η εν λόγω κατηγοριοποίηση είναι αυστηρή, με κάθε εικόνα να ανήκει σε μια και μόνο κατηγορία, και είναι αυτή που προσομοιώνει την οπτική του χρήστη. Επομένως, στις προσομοιώσεις για την αξιολόγηση των μεθοδολογιών το μόνο που μπορούμε να υποθέσουμε είναι ότι τα παραδείγματα ανατροφοδοτούνται είτε ως απολύτως θετικά είτε ως απολύτως αρνητικά.

Με βάση τα παραπάνω, σε κάθε εποχή της ανατροφοδότησης οι παράμετροι  $\lambda_m^r$  έχουν ίδια τιμή για κάθε  $m$ , δηλαδή για κάθε θετικό παράδειγμα, και το ίδιο ισχύει όσο αφορά τις παραμέτρους  $\lambda_m^n$  για τα αρνητικά παραδείγματα. Συγκεκριμένα, έστω ότι με  $(M_r)^j$ ,  $(M_n)^j$  συμβολίζουμε τον αριθμό των θετικών και αρνητικών παραδειγμάτων, αντίστοιχα, που δόθηκαν κατά την  $j$ -οστή εποχή. Αν θεωρήσουμε ότι με  $(\lambda_m^r)^e$  και  $(\lambda_m^n)^e$  συμβολίζονται οι τιμές των παραμέτρων  $\lambda_m^r$  και  $\lambda_m^n$ , αντίστοιχα, για την εποχή  $e$  της ανατροφοδότησης, για τον καθορισμό των τιμών αυτών χρησιμοποιούνται οι τύποι

$$(\lambda_m^r)^e = \frac{1}{(t_r)^e}, \quad m = 1, \dots, (M_r)^e \quad (4.50)$$

$$(\lambda_m^n)^e = \frac{1}{(t_n)^e}, \quad m = 1, \dots, (M_n)^e \quad (4.51)$$

όπου

$$(t_r)^e = 1 + \sum_{j=1}^e (M_r)^j \quad (4.52)$$

και

$$(t_n)^e = \sum_{j=1}^e (M_n)^j \quad (4.53)$$

είναι ο συνολικός αριθμός των θετικών και αρνητικών παραδειγμάτων, αντίστοιχα, που έχουν δοθεί έως και την εποχή  $e$ . Στην περίπτωση των θετικών παραδειγμάτων, η μονάδα που προστίθεται επιπλέον στο άθροισμα αντιστοιχεί στην αρχική ερώτηση.

Οι παραπάνω σχέσεις αποσκοπούν στο να αποδίδεται σε κάθε παράδειγμα που έχει δοθεί, ανεξαρτήτως εποχής, ίδιο βάρος. Πράγματι, μέσω των παραπάνω σχέσεων μπορούμε

να προσδιορίσουμε έμμεσα τις τιμές για τις παραμέτρους  $\Lambda_r$  και  $\Lambda_n$ , που χρησιμοποιούνται για τον υπολογισμό των βαρών των προηγούμενων ερωτήσεων κατά το σχηματισμό των νέων μοντέλων. Συγκεκριμένα, όσο αφορά τις παραπάνω παραμέτρους, για την εποχή  $e$  ισχύουν οι σχέσεις

$$(\Lambda_r)^e = \sum_{m=1}^{(M_r)^e} (\lambda_m^r)^e = \frac{(M_r)^e}{(t_r)^e} \quad (4.54)$$

και

$$(\Lambda_n)^e = \sum_{m=1}^{(M_n)^e} (\lambda_m^n)^e = \frac{(M_n)^e}{(t_n)^e} \quad (4.55)$$

Το βάρος με το οποίο λαμβάνεται υπόψιν η προηγούμενη θετική ερώτηση είναι:

$$1 - (\Lambda_r)^e = \frac{(t_r)^e - (M_r)^e}{(t_r)^e} = \frac{(t_r)^{e-1}}{(t_r)^e} \quad (4.56)$$

και, όπως βλέπουμε, είναι ανάλογο του αριθμού των παραδειγμάτων,  $(t_r)^{e-1}$ , που της αντιστοιχούν. Ακριβώς αντίστοιχη σχέση ισχύει και για την περίπτωση της αρνητικής ερώτησης.

## 4.7.2 Παραλλαγή Βασισμένη σε Ιστογράμματα

Σχετικά με τα Ιστογράμματα έχει αναφερθεί στην ενότητα 2.2 ότι το κύριο μειονέκτημά τους σχετίζεται με την «κατάρα της διάστασης». Καθώς η διάσταση του χώρου μεγαλώνει, ο αριθμός των bins για ένα Ιστόγραμμα αυξάνει εκθετικά, με αποτέλεσμα ακόμα και για χώρους σχετικά μικρής διάστασης το Ιστόγραμμα να έχει δυσανάλογα πολλά bins σε σχέση με τα διαθέσιμα δεδομένα.

Στην περίπτωσή μας ο χώρος των χαρακτηριστικών για τις εικόνες έχει 8 διαστάσεις. Ο αριθμός αυτός αποδεικνύεται εξαιρετικά μεγάλος στην πράξη, οδηγώντας σε Ιστογράμματα της τάξης των εκατοντάδων χιλιάδων bins ακόμα και αν η κβαντοποίηση για κάθε διάσταση γίνεται σε σχετικά λίγα διαστήματα. Αυτό μας οδηγεί στην απόφαση αντί για ένα 8-διάστατο Ιστόγραμμα για κάθε εικόνα να χρησιμοποιήσουμε δύο μικρότερης διάστασης, και συγκεκριμένα δύο 5-διάστατα Ιστογράμματα. Το πρώτο από αυτά τα Ιστογράμματα θα αντιστοιχεί στα χαρακτηριστικά θέσης και χρώματος της εικόνας και το δεύτερο στα χαρακτηριστικά θέσης και υφής. Αυτή η επιλογή είναι η μόνη λογική από τη στιγμή που, από τη μια, τα χαρακτηριστικά θέσης δεν έχουν νόημα από μόνα τους, αφού δεν είναι τίποτα άλλο από τις συντεταγμένες των pixels της εικόνας, και, από την άλλη, τόσο η κατανομή του χρώματος όσο και αυτή της υφής εξαρτάται προφανώς από τη θέση στο επίπεδο της εικόνας. Για αυτό, εξάλλου, συμπεριλάβαμε ευθύς εξαρχής τα χαρακτηριστικά θέσης στο σύνολο των χρησιμοποιούμενων χαρακτηριστικών.

Σε αυτό το πλαίσιο, σε κάθε εικόνα και κατ' επέκταση σε κάθε ερώτηση αντιστοιχούν δύο μοντέλα. Ένα Ιστόγραμμα θέσης-χρώματος και ένα θέσης-υφής. Επομένως, προκειμένου να υπολογίσουμε μια συνολική τιμή για την απόσταση μεταξύ της θετικής ή αρνητικής ερώτησης και της  $i$ -οστής εικόνας της βάσης, που μας χρειάζεται για τον υπολογισμό της τιμής του κριτηρίου διάταξης των εικόνων, είναι απαραίτητο να

συνδυαστούν με κάποιο τρόπο οι τιμές της απόκλισης C2, όπως αυτές υπολογίζονται για κάθε ένα από τα παραπάνω Ιστογράμματα. Επιλέξαμε να χρησιμοποιήσουμε για αυτό το σκοπό ένα κυρτό συνδυασμό των αποκλίσεων C2 που αντιστοιχούν στα δύο Ιστογράμματα. Συγκεκριμένα, οι σχέσεις (4.10) και (4.11) τροποποιούνται παίρνοντας την ακόλουθη μορφή:

$$d(p_q, p_i) = l_{pc} C2(p_q^{pc}, p_i^{pc}) + l_{pt} C2(p_q^{pt}, p_i^{pt}) \quad (4.57)$$

$$d(p_n, p_i) = l_{pc} C2(p_n^{pc}, p_i^{pc}) + l_{pt} C2(p_n^{pt}, p_i^{pt}) \quad (4.58)$$

όπου  $l_{pc}, l_{pt} \geq 0$  και  $l_{pc} + l_{pt} = 1$ , ενώ τα  $p_q^{pc}$  και  $p_q^{pt}$  αντιστοιχούν στα Ιστογράμματα θέσης-χρώματος και θέσης-υφής, αντίστοιχα, για τη θετική ερώτηση. Ανάλογα ισχύουν για τα  $p_n^{pc}$  και  $p_n^{pt}$  όσο αφορά την αρνητική ερώτηση, καθώς και για τα  $p_i^{pc}$  και  $p_i^{pt}$  όσο αφορά την ισοστή εικόνα της βάσης.

Όσο αφορά, τώρα, την ανάλυση που χρησιμοποιείται για τα δύο αυτά Ιστογράμματα, επιλέχθηκαν προς σύγκριση δύο διαφορετικές κβαντοποιήσεις. Η πρώτη, έστω Hist1, χρησιμοποιεί λιγότερα bins, και συγκεκριμένα ένα Ιστόγραμμα θέσης-χρώματος  $3 \times 3 \times 4 \times 8 \times 8$  (κάθε αριθμός αντιστοιχεί στο πλήθος των διαστημάτων στα οποία χωρίζεται το εύρος τιμών του αντίστοιχου χαρακτηριστικού) για τα χαρακτηριστικά  $x^n, y^n, c_s^L, c_s^A, c_s^B$ , αντίστοιχα, και ένα Ιστόγραμμα θέσης-υφής  $3 \times 3 \times 4 \times 4 \times 4$  για τα  $x^n, y^n, t_s^{AC}, t_s^{PC}, t_s^C$ , αντίστοιχα. Στη δεύτερη κβαντοποίηση, έστω Hist2, που είναι πιο πυκνή, το Ιστόγραμμα θέσης-χρώματος έχει ανάλυση  $5 \times 5 \times 8 \times 16 \times 16$  και το Ιστόγραμμα θέσης-υφής  $5 \times 5 \times 8 \times 8 \times 8$ .

### 4.7.3 Παραλλαγή Βασισμένη στο Μέτρο ALA

Όσο αφορά την παραλλαγή που στηρίζεται στην ALA αντί της απόκλισης C2 για τη σύγκριση μεταξύ Γκαουσιανών Μειγμάτων, τα χρησιμοποιούμενα μοντέλα είναι τα ίδια που χρησιμοποιούνται και για την προτεινόμενη μέθοδο, η οποία βασίζεται στην απόκλιση C2. Αλλά και σε ό,τι έχει να κάνει με τη διαδικασία της συσχετιστικής ανατροφοδότησης, όπως έχουμε αναφέρει και παραπάνω, οι μόνες διαφοροποιήσεις έγκεινται στην ανανέωση του μέτρου σύγκρισης μειγμάτων, που πραγματοποιείται μέσω του τύπου (4.17) αντί για τους (4.5)-(4.7), και στον ορισμό των  $d(p_q, p_i)$  και  $d(p_n, p_i)$ , που δίνεται μέσω των σχέσεων (4.18) και (4.19) αντί για τις (4.10) και (4.11), αντίστοιχα. Κατά τα άλλα, η μέθοδος υλοποιείται με τον ίδιο ακριβώς τρόπο όπως και η προτεινόμενη. Συνεπώς, για τις παραμέτρους  $\lambda_m^t$  και  $\lambda_m^n$  ισχύει και σε αυτήν την περίπτωση, όπως και στην περίπτωση της παραλλαγής των Ιστογραμμάτων, η ανάλυση που έγινε στα πλαίσια της προτεινόμενης μεθόδου.

### 4.7.4 Μέθοδος *Kherfi*

Η μέθοδος *Kherfi* χρησιμοποιεί, όπως αναφέραμε κατά την παρουσίασή της στην ενότητα 4.5.3, ως περιγραφή για κάθε εικόνα ένα αριθμό διανυσμάτων. Για να υπολογίσουμε κατάλληλα διανύσματα χαρακτηριστικών για αυτή τη μέθοδο, χωρίσαμε κάθε εικόνα σε 9 υποεικόνες (sub-images), διαιρώντας την τόσο κατά πλάτος όσο και κατά ύψος σε 3 ίσα διαστήματα. Για κάθε μια από τις παραπάνω υποεικόνες υπολογίσαμε ένα

Ιστογράμματα χρώματος  $4 \times 8 \times 8$  ( $c_s^L, c_s^A, c_s^B$ ) και ένα Ιστογράμματα υφής  $4 \times 4 \times 4$  ( $t_s^{AC}, t_s^{PC}, t_s^C$ ). Με αυτόν τον τρόπο, για κάθε εικόνα παίρνουμε 18 διανύσματα χαρακτηριστικών, 9 διανύσματα 256 διαστάσεων, που αντιστοιχούν στα Ιστογράμματα χρώματος, και 9 διανύσματα 64 διαστάσεων, που αντιστοιχούν στα Ιστογράμματα υφής.

Από τα παραπάνω γίνεται φανερό ότι τα χαρακτηριστικά που προκύπτουν με αυτόν τον τρόπο είναι σε πλήρη αντιστοιχία με την περιγραφή των εικόνων μέσω της κβαντοποίησης Hist1, η οποία θα χρησιμοποιηθεί για την βασιζόμενη στα Ιστογράμματα παραλλαγή της προτεινόμενης μεθόδου. Τα ανάλογα διανύσματα χαρακτηριστικών που αντιστοιχούν στην κβαντοποίηση Hist2 δε μπορούν να χρησιμοποιηθούν για τη μέθοδο *Kherfi*, λόγω του ότι η χρονική πολυπλοκότητα της συγκεκριμένης μεθόδου αυξάνει ραγδαία καθώς μεγαλώνει η διάσταση των χρησιμοποιούμενων διανυσμάτων χαρακτηριστικών.

Επίσης, για τους ίδιους λόγους που μας οδήγησαν στο να θεωρήσουμε ότι οι τιμές των παραμέτρων  $\lambda_m^r$  και  $\lambda_m^n$  της προτεινόμενης μεθόδου δεν μπορούν να εξαρτώνται κατά τις προσομοιώσεις από τα συγκεκριμένα παραδείγματα που παρέχονται, αποφασίσαμε οι παράμετροι  $\pi_n^1, \pi_n^2$  (βλ. ενότητα 4.5.3) να λαμβάνουν τιμές ανεξάρτητες από το  $n$ , και συγκεκριμένα προχωρήσαμε στην επιλογή

$$\pi_n^k = \frac{\tilde{\pi}^k}{N_k} \text{ για } k=1,2 \text{ και } \forall n=1,\dots,N_k \quad (4.59)$$

Τέλος, όπως έχουμε αναφέρει ήδη, η παράμετρος  $\tilde{\pi}^1$  αντιστοιχεί στην παράμετρο  $a_r$  της προτεινόμενης μεθόδου, και η  $\tilde{\pi}^2$  στο  $1 - a_r$ . Έτσι, για λόγους συγκρισιμότητας, επιλέξαμε να θέσουμε

$$\tilde{\pi}^1 = a_r \quad (4.60)$$

δηλαδή, σε κάθε σύγκριση οι δύο παραπάνω παράμετροι να έχουν την ίδια τιμή.

## 4.8 Πειράματα

### 4.8.1 Προσομοιώσεις

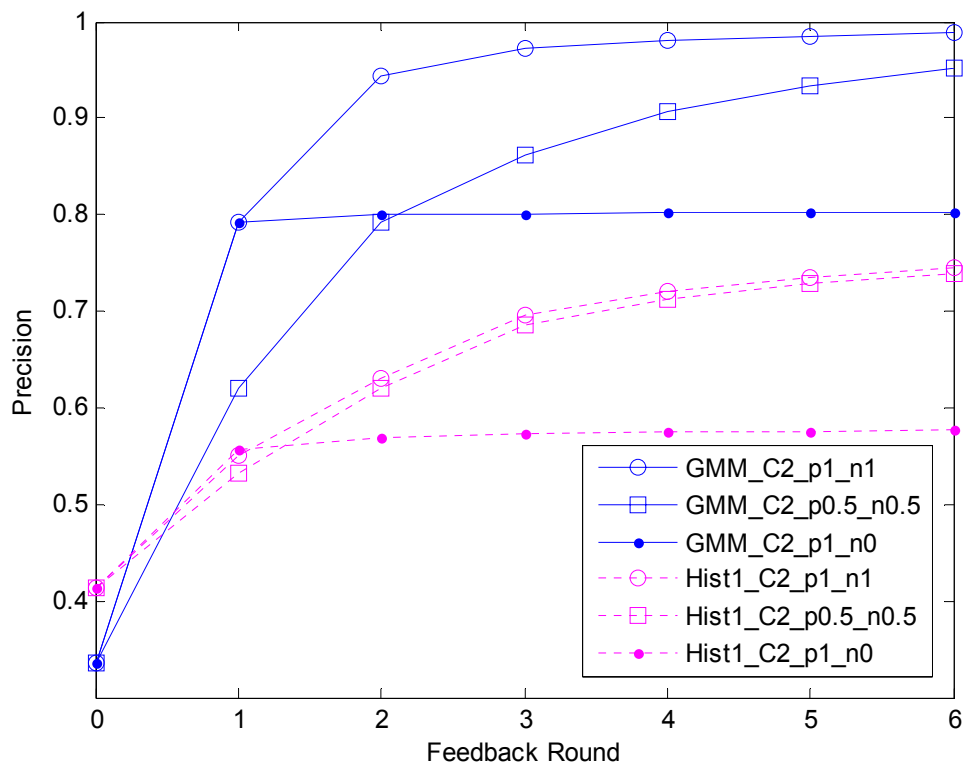
Για την αξιολόγηση της απόδοσης της προτεινόμενης μεθόδου προχωρήσαμε σε μια σειρά συγκρίσεων της με εκείνες τις μεθόδους που παρουσιάστηκαν στην ενότητα 4.5. Για να καταστούν αυτές οι συγκρίσεις εφικτές σχεδιάσαμε κατάλληλες προσομοιώσεις της διαδικασίας της συσχετιστικής ανατροφοδότησης. Οι προσομοιώσεις αυτές λαμβάνουν πάντοτε υπόψιν και στηρίζονται στην εκ των προτέρων γνωστή κατηγοριοποίηση των χρησιμοποιούμενων βάσεων εικόνων. Συγκεκριμένα, το αν μια εικόνα είναι πράγματι σχετική ή όχι με μια ερώτηση καθορίζεται αποκλειστικά και μόνο από το αν ανήκει στην ίδια κατηγορία της βάσης με την ερώτηση αυτή. Αν κάτι τέτοιο ισχύει, τότε η εικόνα θεωρείται ότι είναι πράγματι σχετική με την ερώτηση, ενώ στην αντίθετη περίπτωση θεωρείται άσχετη.

Για τις προσομοιώσεις που ακολουθούνται στα πειράματα, δύο διαφορετικά πλάνα προσομοίωσης σχεδιάστηκαν. Σε κάθε ένα από αυτά, αρχικά σχηματίζεται ένα σύνολο εικόνων που περιλαμβάνει για κάθε κατηγορία της βάσης ένα ποσοστό από τις εικόνες που της αντιστοιχούν. Οι εικόνες αυτού του συνόλου παρουσιάζονται διαδοχικά στο σύστημα

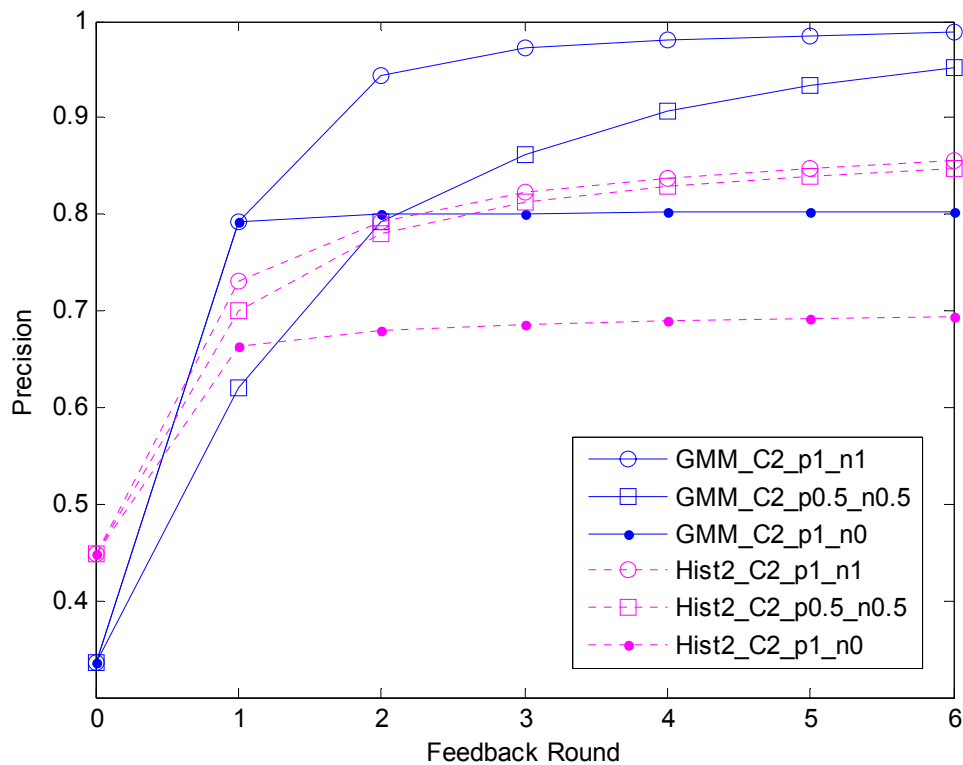
ανάκτησης ως αρχικές ερωτήσεις, προκειμένου για καθεμία από αυτές να ξεκινήσει ένας αντίστοιχος κύκλος γύρων συσχετιστικής ανατροφοδότησης.

### Πρώτο Πλάνο Προσομοίωσης

Όσο αφορά το πρώτο πλάνο προσομοίωσης, σχετικές με βάση την κρίση του συστήματος εικόνες ανακτώνται για κάθε αρχική ερώτηση και εποχή της ανατροφοδότησης, έως ότου ένα συγκεκριμένο επίπεδο ανάκλησης, *RecLev*, επιτευχθεί. Τότε, για το σύνολο των εικόνων που έχουν ανακτηθεί υπολογίζεται η αντίστοιχη τιμή για την ακρίβεια, ενώ παράλληλα προσδιορίζονται και οι πραγματικά σχετικές εικόνες που ανακτήθηκαν καθώς και εκείνες που παρότι ανακτήθηκαν δεν είναι σχετικές με την αρχική ερώτηση (υπενθυμίζουμε εδώ ότι για τα μέτρα της ακρίβειας και της ανάκλησης έχουμε μιλήσει ήδη στην ενότητα 1.4.2). Κατόπιν, ένα ποσοστό,  $P_r$ , των πραγματικά σχετικών εικόνων και ένα άλλο,  $P_n$ , αυτών που ανακτήθηκαν παρότι δεν είναι σχετικές χρησιμοποιούνται ως θετικά και αρνητικά παραδείγματα, αντίστοιχα, στη διαδικασία της ανατροφοδότησης. Αν  $P_r, P_n < 1$ , τότε για τον προσδιορισμό των αντίστοιχων παραδειγμάτων χρησιμοποιείται τυχαία επιλογή.



**Σχήμα 4.1:** Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης μέσω τροποποίησης της αναπαράστασης της ερώτησης (1<sup>ο</sup> πείραμα) – Μέση ακρίβεια για επίπεδο ανάκλησης *RecLev* = 0.3 (βάση εικόνων MSDB)



**Σχήμα 4.2:** Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συστηματικής ανατροφοδότησης μέσω τροποποίησης της αναπαράστασης της ερώτησης (2<sup>ο</sup> πείραμα) – Μέση ακρίβεια για επίπεδο ανάκλησης  $RecLev = 0.3$  (βάση εικόνων MSDB)

Στα σχήματα 4.1 και 4.2 βλέπουμε την πρόοδο όσο αφορά τη μέση τιμή του μέτρου της ακρίβειας για όλες τις αρχικές ερωτήσεις κατά τις διάφορες εποχές της ανατροφοδότησης και για επίπεδο ανάκλησης  $RecLev = 0.3$ . Σε αυτά τα σχήματα, με GMM\_C2 συμβολίζουμε την προτεινόμενη μέθοδο και με Hist1\_C2 (Hist2\_C2) την παραλλαγή που βασίζεται στα Ιστογράμματα και στην κβαντοποίηση Hist1 (Hist2). Επίσης, η ένδειξη “pN” αντιστοιχεί στη σχέση  $P_r = N$  για το ποσοστό των ανατροφοδοτούμενων σχετικών εικόνων και, αντίστοιχα, με “nN” αναφερόμαστε στην επιλογή  $P_n = N$ . Τα αποτελέσματα αφορούν τη βάση MSDB, και το σύνολο των αρχικών ερωτήσεων περιελάμβανε κάθε εικόνα της εν λόγω βάσης.

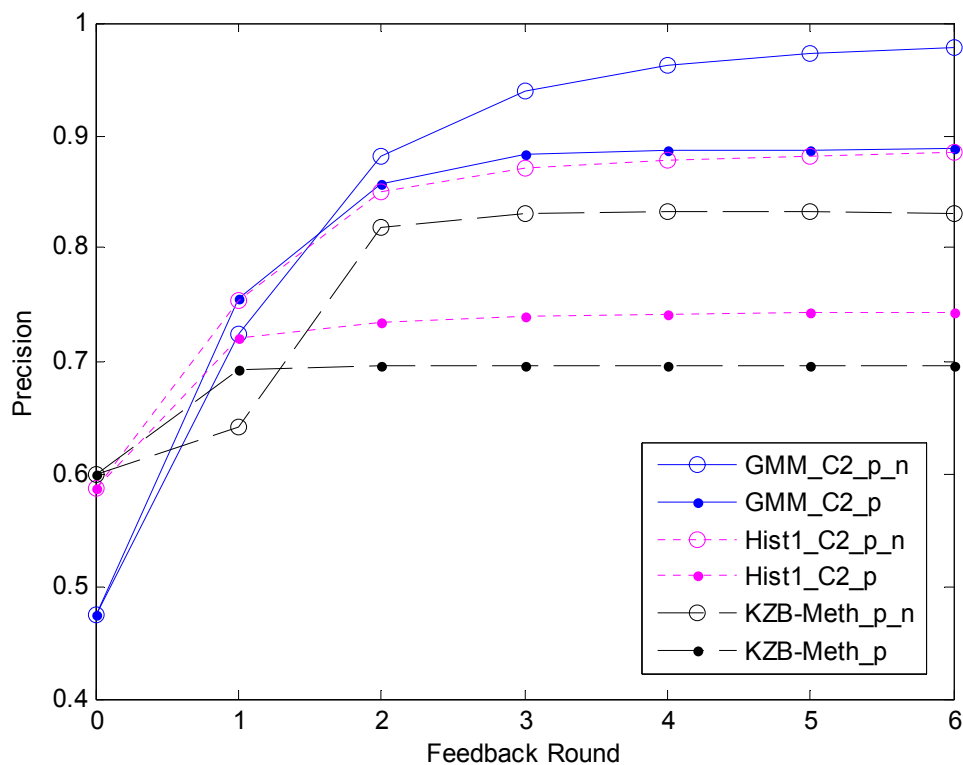
Όταν χρησιμοποιούνται αρνητικά παραδείγματα, για την παράμετρο σχετικού βάρους της θετικής ερώτησης,  $a_r$ , ισχύει  $a_r = 0.55$ . Σε ό,τι έχει να κάνει με τη βασισμένη στα Ιστογράμματα παραλλαγή της προτεινόμενης μεθόδου, για τις παραμέτρους  $l_{pc}$ ,  $l_{pt}$  θέσαμε  $l_{pc} = l_{pt} = 0.5$ .

## Δεύτερο Πλάνο Προσομοίωσης

Το δεύτερο πλάνο προσομοίωσης μοιάζει πολύ με αυτό που ακολουθείται στην εργασία [12]. Συγκεκριμένα, ως μέτρο απόδοσης χρησιμοποιείται η τιμή της ακρίβειας υπολογιζόμενη στις πρώτες  $T$  ανακτημένες εικόνες. Σε κάθε γύρο της ανατροφοδότησης, το πολύ  $K_r$  πραγματικά σχετικές εικόνες επιλέγονται από τις πρώτες  $R$  ανακτημένες εικόνες και ανατροφοδοτούνται στο σύστημα ως θετικά παραδείγματα. Αν ο αριθμός των

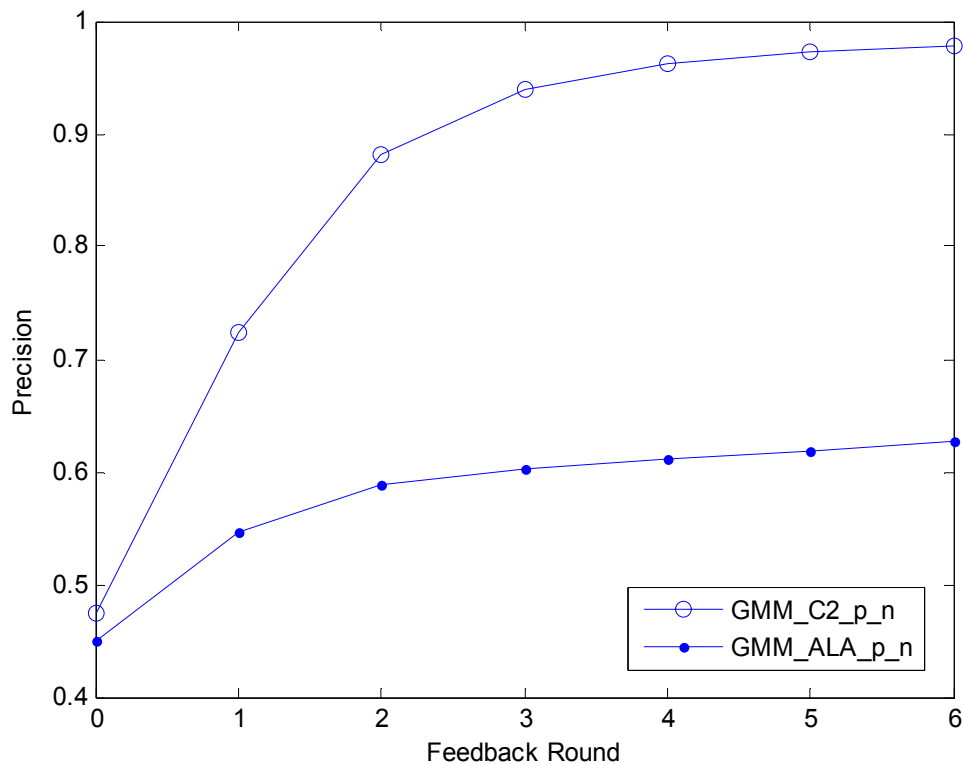
πραγματικά σχετικών εικόνων στο σύνολο των  $R$  πρώτων ανακτημένων εικόνων είναι μεγαλύτερος από  $K_r$ , η επιλογή γίνεται με τυχαίο τρόπο. Αντίθετα, αν υπάρχουν λιγότερες από  $K_r$  σχετικές εικόνες, τότε παρέχονται ως παραδείγματα όλες όσες υπάρχουν. Με ανάλογο τρόπο, επιλέγουμε ως αρνητικά παραδείγματα το πολύ  $K_n$  από τις πραγματικά μη σχετικές εικόνες που υπάρχουν στις πρώτες  $R$  θέσεις της λίστας κατάταξης, όπως αυτή διαμορφώνεται από το αντίστοιχο κριτήριο της εκάστοτε μεθόδου.

Για αυτό το πλάνο προσομοίωσης παρέχουμε πειραματικά αποτελέσματα χρησιμοποιώντας και τις δύο βάσεις της ενότητας 4.6.1. Όσο αφορά την πρώτη βάση, MSDB, το σύνολο των αρχικών ερωτήσεων περιλαμβάνει κάθε εικόνα της, ενώ για τις προσομοιώσεις πάνω στη δεύτερη βάση, Corel9923, στις αρχικές ερωτήσεις περιλαμβάνεται το 40% των εικόνων που αυτή περιέχει, με την επιλογή τους για καθεμιά από τις 42 κατηγορίες να γίνεται με τυχαίο τρόπο. Το παραπάνω ποσοστό για τις αρχικές ερωτήσεις είναι αρκετό, κρινόμενο από στατιστικής απόψεως, ώστε τα αποτελέσματα που θα πάρουμε να μπορούν να θεωρηθούν πράγματι αντιπροσωπευτικά της απόδοσης των μεθόδων. Επίσης, μέσω αυτής της προσομοίωσης θα συγκριθούν μεταξύ τους όλες οι μέθοδοι που παρουσιάστηκαν σε αυτό το κεφάλαιο, ώστε να καταλήξουμε σε ένα ασφαλές συμπέρασμα για την αξία της προτεινόμενης μεθοδολογίας συσχετιστικής ανατροφοδότησης.

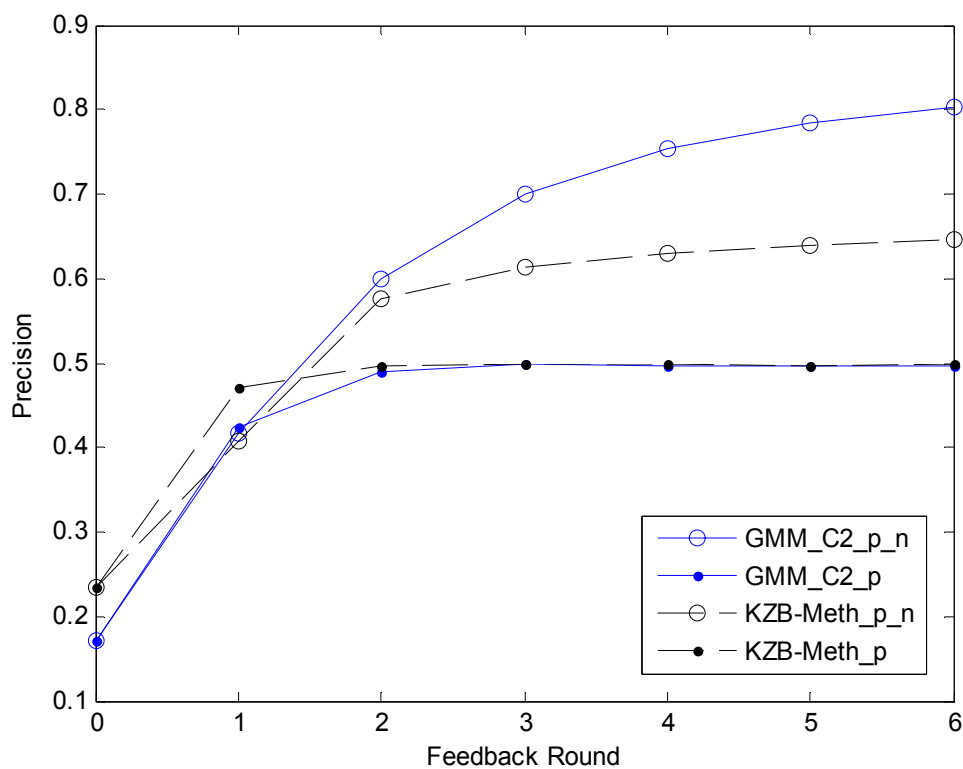


**Σχήμα 4.3:** Σύγκριση μεθόδων συσχετιστικής ανατροφοδότησης που βασίζονται στην τροποποίηση της αναπαράστασης της ερώτησης – Μέση ακρίβεια για εμβέλεια  $T = 20$  (βάση εικόνων MSDB)

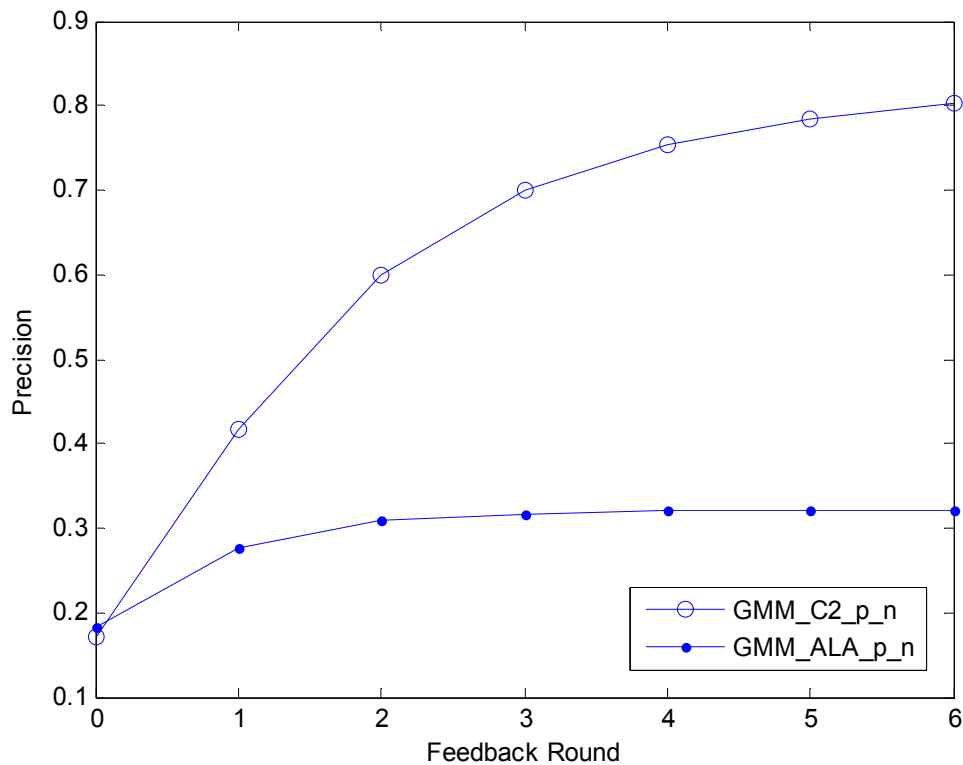




**Σχήμα 4.4:** Σύγκριση μέτρων απόστασης μεταξύ Γκαουσιανών Μειγμάτων στο πλαίσιο της συχετιστικής ανατροφοδότησης μέσω τροποποίησης της αναπαράστασης της ερώτησης – Μέση ακρίβεια για εμβέλεια  $T = 20$  (βάση εικόνων MSDB)



**Σχήμα 4.5:** Σύγκριση μεθόδων συχετιστικής ανατροφοδότησης που βασίζονται στην τροποποίηση της αναπαράστασης της ερώτησης – Μέση ακρίβεια για εμβέλεια  $T = 20$  (βάση εικόνων Corel9923)



**Σχήμα 4.6:** Σύγκριση μέτρων απόστασης μεταξύ Γκαουσιανών Μειγμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης μέσω τροποποίησης της αναπαράστασης της ερώτησης – Μέση ακρίβεια για εμβέλεια  $T = 20$  (βάση εικόνων Core19923)

Για τις συγκεκριμένες τιμές που επιλέχθηκαν όσο αφορά τις παραμέτρους της προσομοίωσης έχουμε τα εξής: Για την εμβέλεια υπολογισμού της ακρίβειας χρησιμοποιείται  $T = 20$ , για το σύνολο ανάκτησης από όπου επιλέγονται τα παραδείγματα  $R = 150$ , για το μέγιστο πλήθος των θετικών παραδειγμάτων ανά εποχή της ανατροφοδότησης  $K_r = 10$ , και για το αντίστοιχο μέγιστο πλήθος των αρνητικών παραδειγμάτων  $K_n = 10$ . Όταν στην ανατροφοδότηση περιλαμβάνονται και αρνητικά παραδείγματα, ισχύει γενικά  $a_r = 0.65$  (εκτός από την περίπτωση της παραλλαγής που βασίζεται στην ALA, για την οποία παίρνουμε τα καλύτερα αποτελέσματα όταν  $a_r = 0.8$ ). Για την παραλλαγή που χρησιμοποιεί Ιστογράμματα θέσαμε και αυτή τη φορά  $l_{pc} = l_{pt} = 0.5$ . Όσο αφορά τη μέθοδο *Kherfi*, όταν χρησιμοποιούνται αρνητικά παραδείγματα, οι  $R (=150)$  πιο κοντινές στα θετικά παραδείγματα εικόνες κατά το πρώτο στάδιο της ανάκτησης, που λαμβάνει υπόψιν μόνο τα θετικά παραδείγματα, διατηρούνται προς αναδιάταξη κατά το δεύτερο στάδιο, όπου λαμβάνονται υπόψιν και τα αρνητικά παραδείγματα.

Στα σχήματα 4.3, 4.4, 4.5 και 4.6 παρουσιάζεται η μέση τιμή για την ακρίβεια, υπολογισμένη για όλες τις αρχικές ερωτήσεις και για κάθε γύρο της ανατροφοδότησης. Σε αυτά τα σχήματα, με GMM\_C2 και Hist1\_C2 συμβολίζουμε πάλι την προτεινόμενη μέθοδο και την παραλλαγή των Ιστογραμμάτων (με κβαντοποίηση Hist1), αντίστοιχα, με GMM\_ALA συμβολίζουμε την παραλλαγή που βασίζεται στο μέτρο ομοιότητας ALA, και με KZB-Meth τη μέθοδο *Kherfi*. Τέλος, η ένδειξη “p” επισημαίνει τη χρήση θετικών παραδειγμάτων και η ένδειξη “n” τη χρήση αρνητικών παραδειγμάτων.

## Απαιτήσεις σε Υπολογιστικό Χρόνο

Προκειμένου να δώσουμε μια εικόνα για την ταχύτητα της προτεινόμενης μεθόδου, αναφέρουμε ενδεικτικά ότι για την πρώτη βάση, MSDB, και για το δεύτερο πλάνο προσομοίωσης, με χρήση τόσο θετικών όσο και αρνητικών παραδειγμάτων και με τις παραμέτρους της προσομοίωσης να έχουν τις τιμές που καθορίστηκαν προηγουμένως, χρειάζονται 166 sec υπολογιστικού χρόνου (σε ένα προσωπικό υπολογιστή με συχνότητα επεξεργαστή 3GHz, που χρησιμοποιεί Matlab) για να εκτελεστούν η αρχική ανάκτηση συν 6 εποχές ανατροφοδότησης για 3740 αρχικές ερωτήσεις. Αυτό αντιστοιχεί σε μέσο χρόνο 0.044 sec για κάθε αρχική ερώτηση ακολουθούμενη από 6 γύρους ανατροφοδότησης.

Όσο αφορά τη βασισμένη στα Ιστογράμματα παραλλαγή, αυτή χρειάζεται κατά προσέγγιση διπλάσιο χρόνο σε σχέση με την προτεινόμενη μέθοδο για την εκτέλεση του ίδιου σεναρίου, λόγω του ότι κάθε εικόνα αναπαριστάνεται μέσω δύο Ιστογραμμάτων αντί για το ένα Γκαουσιανό Μείγμα που προβλέπει η προτεινόμενη μέθοδος. Επομένως, η ανανέωση των τιμών της απόκλισης C2 πρέπει να γίνεται και για τα δύο αυτά Ιστογράμματα, πράγμα που αντιστοιχεί σε διπλάσιο κόπο για κάθε εικόνα της βάσης.

Επίσης, η παραλλαγή που βασίζεται στην ALA, λόγω του ότι χρησιμοποιεί τα ίδια μοντέλα με την προτεινόμενη μέθοδο και παρόμοια αυξητική τεχνική για την ανανέωση της τιμής του μέτρου σύγκρισης, καταλήγει σε μια εκτέλεση με ταχύτητα ανάλογη με αυτήν της προτεινόμενης μεθόδου.

Τέλος, χρησιμοποιώντας τη μέθοδο *Kherfi* για την εκτέλεση του ίδιου σεναρίου προσομοίωσης, χρειάστηκαν περίπου 19000 sec για την ολοκλήρωση της διαδικασίας για όλες τις αρχικές ερωτήσεις, κάτι που σημαίνει ότι απαιτήθηκε χρόνος κατά μέσο όρο ίσος με 5 sec ανά ερώτηση.

### 4.8.2 Αποτελέσματα

Στην ενότητα αυτή θα σχολιάσουμε τα αποτελέσματα των πειραμάτων, όπως αυτά παρουσιάζονται στα σχήματα 4.1 έως 4.6, θα αναφερθούμε γενικότερα σε διάφορα ζητήματα που αφορούν την εμπειρία που κατακτήθηκε από την αξιολόγηση των διαφόρων μεθοδολογιών, θα συνυπολογίσουμε τα δεδομένα που έχουν να κάνουν με την ταχύτητα των μεθόδων, και θα καταλήξουμε σε χρήσιμα συμπεράσματα.

Ξεκινώντας από τα σχήματα 4.1, 4.2 και 4.3, παρατηρούμε ότι όταν χρησιμοποιούνται Γκαουσιανά Μείγματα για την αναπαράσταση των εικόνων, η ακρίβεια είναι σχεδόν πάντα υψηλότερη σε σχέση με αυτήν που επιτυγχάνεται όταν ως μοντέλα για τις εικόνες χρησιμοποιούνται Ιστογράμματα. Εξαιρέση αυτού του κανόνα αποτελούν η αρχική ανάκτηση (πριν από την έναρξη της διαδικασίας της ανατροφοδότησης) και σπάνια οι αρχικοί 1-2 γύροι της ανατροφοδότησης. Η υπεροχή των Γκαουσιανών Μειγμάτων κατά τη σύγκριση με τη βασισμένη στα Ιστογράμματα παραλλαγή παρατηρείται ανεξάρτητα από το ποια από τις δύο διαφορετικές κβαντοποιήσεις, Hist1 ή Hist2, έχει επιλεγεί για τον υπολογισμό των Ιστογραμμάτων. Μετά από λίγους γύρους ανατροφοδότησης, στην περίπτωση των Γκαουσιανών Μειγμάτων καταλήγουμε πάντα σε επίπεδα ακρίβειας σημαντικά υψηλότερα συγκριτικά με τα αντίστοιχα που επιτυγχάνονται με τη χρήση Ιστογραμμάτων. Λαμβάνοντας τώρα υπόψιν και το γεγονός ότι η βασισμένη στα Ιστογράμματα παραλλαγή απαιτεί διπλάσιο κατά μέσο όρο χρόνο για κάθε ερώτηση του χρήστη, λόγω της αρνητικής επίδρασης της «κατάρας της διάστασης», που μας οδήγησε στην ανάγκη αναπαράστασης της κάθε εικόνας μέσω δύο Ιστογραμμάτων, καταλήγουμε στη διαπίστωση ότι μέσω των Γκαουσιανών Μειγμάτων παρέχεται μια συνολικά καλύτερη αναπαράσταση των εικόνων.

Επιπροσθέτως, από τα σχήματα 4.4 και 4.6 γίνεται εύκολα αντιληπτό ότι η χρήση της απόκλισης C2 οδηγεί το σύστημα ανάκτησης πάντα σε σημαντικά καλύτερη απόδοση σε σχέση με την περίπτωση που ως μέτρο σύγκρισης υιοθετείται η ALA, μέσω της αντίστοιχης παραλλαγής της προτεινόμενης μεθόδου. Από χρονικής άποψης, τόσο η προτεινόμενη μέθοδος όσο και η παραλλαγή που χρησιμοποιεί την ALA έχουν πολύ γρήγορη απόκριση, εκμεταλλευόμενες και τον κομψό τρόπο που προτείνεται για την ανανέωση του μοντέλου της ερώτησης από εποχή σε εποχή της ανατροφοδότησης. Όμως, η πολύ υψηλότερη απόδοση της προτεινόμενης μεθόδου όσο αφορά την ποιότητα των αποτελεσμάτων, μας οδηγεί στο συμπέρασμα ότι η επιλογή της απόκλισης C2 για την προτεινόμενη μεθοδολογία είναι εύστοχη.

Επίσης, όπως απεικονίζεται στα σχήματα 4.3 και 4.5, η προτεινόμενη μέθοδος σχεδόν πάντα υπερσχύει της μεθόδου *Kherfi* μετά από τους αρχικούς (1 ή 2) γύρους ανατροφοδότησης. Η διαφορά στην απόδοση είναι πιο θεαματική όταν παρέχονται ως ανατροφοδότηση και θετικά και αρνητικά παραδείγματα. Αυτό αποτελεί σαφή ένδειξη ότι η προτεινόμενη μέθοδος ενσωματώνει πολύ πιο αποτελεσματικά τα αρνητικά παραδείγματα, και εκμεταλλεύεται πολύ καλύτερα την πληροφορία που προέρχεται από αυτά. Ακόμα, η μέθοδος *Kherfi* συγκρινόμενη με τη βασισμένη στα Ιστογράμματα παραλλαγή της προτεινόμενης μεθόδου (όταν χρησιμοποιείται η κβαντοποίηση Hist1), και πάλι εμφανίζει υποδεέστερα αποτελέσματα. Λαμβάνοντας υπόψιν ότι, όπως αναφέρθηκε και προηγουμένως, τα χαρακτηριστικά που χρησιμοποιεί η μέθοδος *Kherfi* ουσιαστικά ταυτίζονται με τα Ιστογράμματα που προκύπτουν από την κβαντοποίηση Hist1, η υψηλότερη απόδοση της παραλλαγής των Ιστογραμμάτων είναι μια σαφής επιπρόσθετη ένδειξη για την ανωτερότητα της προτεινόμενης τεχνικής συσχετιστικής ανατροφοδότησης. Αν συνυπολογίσουμε τώρα και τις ιδιαίτερα αυξημένες χρονικές απαιτήσεις της μεθόδου *Kherfi*, συγκριτικά με τις αντίστοιχες της προτεινόμενης μεθόδου αλλά και των παραλλαγών της, καταλήγουμε στο συμπέρασμα ότι η προτεινόμενη μεθοδολογία συσχετιστικής ανατροφοδότησης εμφανίζει σαφή πλεονεκτήματα σε όλους τομείς.

Τέλος, σχετικά με την προτεινόμενη μέθοδο πρέπει να αναφερθεί ότι στην περίπτωση που χρησιμοποιείται μόνο θετική ανατροφοδότηση, μέσω παραδειγμάτων σχετικών εικόνων, και γενικότερα όταν μια μεγάλη τιμή ανατίθεται στην παράμετρο  $a_r$ , που αντιστοιχεί στο σχετικό βάρος με το οποίο λαμβάνεται υπόψιν η θετική ανατροφοδότηση, παρατηρείται μια απότομη αύξηση της ακρίβειας κατά τον πρώτο γύρο της ανατροφοδότησης, αλλά μετά αυτή σταθεροποιείται γρήγορα σε ένα μέγιστο επίπεδο. Από την άλλη, όταν και θετικά και αρνητικά παραδείγματα παρέχονται και χρησιμοποιούνται κατά την ανατροφοδότηση, και καθώς το  $a_r$  μικραίνει, μένοντας πάντα μεγαλύτερο από την τιμή 0.5, η αύξηση της ακρίβειας γίνεται πιο ομαλή, ένας μεγαλύτερος αριθμός από γύρους ανατροφοδότησης χρειάζονται για να συγκλίνει αυτή στη μέγιστη δυνατή τιμή, παρόλα αυτά όμως, σε πολλές περιπτώσεις το τελικό επίπεδο ακρίβειας που επιτυγχάνεται ενδέχεται να είναι υψηλότερο από ό,τι πριν. Επίσης, όσο αφορά το πρώτο πλάνο προσομοίωσης, η μείωση των ανατροφοδοτούμενων παραδειγμάτων, μέσω ρύθμισης των ποσοστών  $P_r$  και  $P_n$ , οδηγεί συνήθως σε πιο αργή σύγκλιση και σε χαμηλότερο τελικό επίπεδο ακρίβειας.

## 4.9 Συμπεράσματα

Σε αυτό το κεφάλαιο παρουσιάστηκε μια νέα μέθοδος συσχετιστικής ανατροφοδότησης στα πλαίσια της ανάκτησης εικόνων με βάση το περιεχόμενο, που προτάθηκε ως αποτέλεσμα της σχετικής με αυτήν τη διατριβή ερευνητικής δραστηριότητας. Η προτεινόμενη μέθοδος στηρίζεται σε ένα μοντέλο για την ερώτηση που ανανεώνεται με

έναν πιθανοτικό τρόπο σε κάθε γύρο της ανατροφοδότησης. Ως μοντέλα για τις εικόνες της βάσης και τις ερωτήσεις χρησιμοποιούνται Γκαουσιανά Μείγματα. Επίσης, με βάση τις ιδιότητες του χρησιμοποιούμενου μέτρου απόστασης, που είναι η απόκλιση C2, και τον προτεινόμενο τρόπο ανανέωσης του μοντέλου της ερώτησης, μια πολύ γρήγορη ανανέωση της εκτίμησης του συστήματος όσο αφορά τη συνάφεια των εικόνων της βάσης με την ερώτηση του χρήστη επιτυγχάνεται μετά από την υποβολή των παραδειγμάτων. Επιπλέον, η ενσωμάτωση της πληροφορίας των αρνητικών παραδειγμάτων γίνεται με πολύ κομψό και αποτελεσματικό τρόπο, χρησιμοποιώντας και τα πλεονεκτήματα που προσφέρει η μοντελοποίηση μέσω Γκαουσιανών Μειγμάτων.

Μέσω ενός συνόλου πειραμάτων που σχεδιάστηκαν και εκτελέστηκαν, αποδεικνύεται η υπεροχή της χρησιμοποιούμενης αναπαράστασης για τις εικόνες σε σχέση με την κλασική προσέγγιση της αναπαράστασης μέσω Ιστογραμμάτων, καθώς και τα πλεονεκτήματα της απόκλισης C2 σε σύγκριση με τη χρήση του μέτρου ομοιότητας ALA για την εκτίμηση της απόστασης μεταξύ Γκαουσιανών Μειγμάτων στο ίδιο πλαίσιο. Τέλος, τα πειράματα έδειξαν ότι η προτεινόμενη μέθοδος εμφανίζει σημαντικά πλεονεκτήματα εν συγκρίσει με την ανάλογη μέθοδο της εργασίας [22], τόσο σε σχέση με την ακρίβεια των αποτελεσμάτων όσο και σε σχέση με τις υπολογιστικές απαιτήσεις



## Κεφάλαιο 5

### Ταξινομητές

Σε αυτό το κεφάλαιο θα ασχοληθούμε με τη χρήση μοντέλων μάθησης για την αντιμετώπιση του προβλήματος της ταξινόμησης. Κυρίως, θα μας απασχολήσει το πρόβλημα της δυαδικής ταξινόμησης (binary classification), για το οποίο γίνεται η υπόθεση ότι τα δεδομένα χωρίζονται σε δύο κατηγορίες. Επίσης, ως εργαλεία για την επίλυση του συγκεκριμένου προβλήματος θα χρησιμοποιηθούν τα SVMs (Support Vector Machines – Μηχανές Διανυσμάτων Υποστήριξης), τα οποία στηρίζονται σε έναν εξαιρετικά δημοφιλή και επιτυχημένο αλγόριθμο εκπαίδευσης, και η χρήση τους έχει οδηγήσει σε πολύ ελπιδοφόρα αποτελέσματα όσο αφορά ένα εξαιρετικά μεγάλο εύρος εφαρμογών. Όπως είναι αναμενόμενο, η εφαρμογή τους στην οποία θα εστιάσουμε στην παρούσα διατριβή αφορά την ανάκτηση εικόνων, και συγκεκριμένα τη χρήση της τεχνικής της συσχετιστικής ανατροφοδότησης προκειμένου να βελτιωθούν τα αποτελέσματα της ανάκτησης εικόνων. Για περισσότερες πληροφορίες σχετικά με τα θέματα που αναλύονται σε αυτό το κεφάλαιο, ο ενδιαφερόμενος αναγνώστης μπορεί να καταφύγει στις εργασίες [16], [33] και [35].

#### 5.1 Προβλήματα Δυαδικής Ταξινόμησης

Ας θεωρήσουμε ότι έχουμε ένα σύνολο δεδομένων το οποίο χωρίζεται σε δύο διακριτές κατηγορίες, δηλαδή κάθε στοιχείο αυτού του συνόλου μπορεί να ανήκει σε μια από δύο διαφορετικές κατηγορίες. Ας θεωρήσουμε, επίσης, ότι έχουμε ένα σύνολο προτύπων (patterns)  $\mathbf{x}_i$ ,  $i=1, \dots, N$ , αντιπροσωπευτικών του προαναφερθέντος συνόλου δεδομένων, και ότι σε κάθε ένα από αυτά τα πρότυπα αντιστοιχεί μια ετικέτα (label)  $t_i$ , που παίρνει τιμές από το σύνολο  $\{-1, +1\}$ . Κάθε μια από τις δύο τιμές που μπορεί να πάρει η κάθε ετικέτα αντιστοιχεί και σε μια από τις δύο κατηγορίες στις οποίες χωρίζονται τα πρότυπα. Με άλλα λόγια, θεωρούμε ότι για τα συγκεκριμένα πρότυπα είναι γνωστή η κατηγοριοποίησή τους. Τα πρότυπα  $\mathbf{x}_i$  μαζί με τις ετικέτες  $t_i$  που τους αντιστοιχούν αποτελούν παραδείγματα κατηγοριοποίησης του δυνητικά πολύ μεγαλύτερου συνόλου δεδομένων που μας ενδιαφέρει, και μπορούν να χρησιμοποιηθούν ως σύνολο εκπαίδευσης (training set) σε μια διαδικασία επιβλεπόμενης μάθησης, με στόχο τον ορισμό κανόνων μέσω των οποίων θα μπορεί να εκτιμηθεί για ένα οποιοδήποτε μελλοντικό, άγνωστο πρότυπο  $\mathbf{x}$  η κατηγορία στην οποία αυτό ανήκει. Επομένως, το πρόβλημα της δυαδικής ταξινόμησης έγκειται στην όσο το δυνατόν πλησιέστερη στην πραγματική διαμέριση του υπό εξέταση συνόλου δεδομένων, χρησιμοποιώντας μόνο την πληροφορία που προέρχεται από το σύνολο εκπαίδευσης.

## 5.2 Γραμμική Διαχωρισιμότητα

Θεωρώντας, αρχικά, ότι τα πρότυπα  $\mathbf{x}_i$  (αλλά και, γενικά, κάθε πρότυπο  $\mathbf{x}$  του συνόλου δεδομένων) είναι διανύσματα σε έναν εν γένει πολυδιάστατο διανυσματικό χώρο, η πιο απλή περίπτωση όσο αφορά το πρόβλημα της δυαδικής ταξινόμησης είναι αυτή κατά την οποία οι δύο κατηγορίες προτύπων είναι γραμμικά διαχωρίσιμες. Η γραμμική διαχωρισιμότητα θα μπορούσε να αφορά την κατηγοριοποίηση ολόκληρου του συνόλου δεδομένων, αλλά στην πράξη αυτό που γνωρίζουμε είναι η κατηγοριοποίηση του συνόλου εκπαίδευσης, επομένως, το μόνο που μπορούμε να εξετάσουμε είναι η γραμμική διαχωρισιμότητα του συνόλου εκπαίδευσης.

Με τον όρο «γραμμική διαχωρισιμότητα» εννοούμε τη δυνατότητα ορισμού ενός τουλάχιστον υπερεπίπεδου στο χώρο των προτύπων, που θα διαχωρίζει πλήρως τις δύο κατηγορίες. Το υπερεπίπεδο αυτό θα ονομάζεται και όριο (boundary) ανάμεσα στις κατηγορίες. Ως γνωστόν, ένα υπερεπίπεδο σε ένα μονοδιάστατο χώρο αντιστοιχεί σε ένα σημείο, σε ένα δισδιάστατο χώρο σε μια ευθεία γραμμή, σε ένα τρισδιάστατο χώρο σε ένα κλασικό δισδιάστατο επίπεδο, και σε χώρους μεγαλύτερης διάστασης σε αντίστοιχους γεωμετρικούς τύπους σημείων, μη οπτικά αναπαραστάσιμους όμως. Ένα τέτοιο υπερεπίπεδο χωρίζει το χώρο στον οποίον ορίζεται σε δύο υποχώρους. Για γραμμική διαχωρισιμότητα μιλάμε αν όλα τα πρότυπα εκπαίδευσης της μιας κατηγορίας ανήκουν στον έναν από τους δύο αυτούς υποχώρους και, παράλληλα, όλα τα πρότυπα της δεύτερης κατηγορίας ανήκουν στον άλλο υποχώρο. Αυτή η ιδέα εκφράζεται μαθηματικά από τις παρακάτω σχέσεις:

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (5.1)$$

$$t_i y(\mathbf{x}_i) > 0, \quad \forall i = 1, \dots, N \quad (5.2)$$

Το διάνυσμα  $\mathbf{w}$  είναι το κάθετο στο υπερεπίπεδο διάνυσμα, και το βαθμωτό μέγεθος  $b$  αντιστοιχεί σε μια παράμετρο μετατόπισης που επιτρέπει στο υπερεπίπεδο να μη διέρχεται αναγκαστικά από την αρχή των αξόνων.

Αναλύοντας τις προηγούμενες σχέσεις, πρέπει να επισημάνουμε ότι η σχέση (5.1) ορίζει το υπερεπίπεδο και τους δύο υποχώρους που αναφέρθηκαν παραπάνω. Συγκεκριμένα, ο ένας από τους δύο υποχώρους περιλαμβάνει όλα εκείνα τα  $\mathbf{x}$  για τα οποία ισχύει  $y(\mathbf{x}) < 0$ , ο άλλος περιλαμβάνει τα σημεία για τα οποία ισχύει  $y(\mathbf{x}) > 0$ , και το υπερεπίπεδο είναι ο γεωμετρικός τόπος των σημείων που ικανοποιούν τη σχέση  $y(\mathbf{x}) = 0$ . Με βάση αυτή τη λογική, η σχέση (5.2) ερμηνευόμενη μας πληροφορεί ότι για όλα τα πρότυπα εκπαίδευσης  $\mathbf{x}_i$  οι τιμές των αντίστοιχων ετικετών  $t_i$  είναι ομόσημες με τις τιμές  $y(\mathbf{x}_i)$ , και επομένως όλα τα πρότυπα της κατηγορίας «-1» ανήκουν στον υποχώρο  $y(\mathbf{x}) < 0$ , ενώ τα πρότυπα της κατηγορίας «+1» ανήκουν στον υποχώρο  $y(\mathbf{x}) > 0$ . Αν, λοιπόν, ισχύει η σχέση (5.2), γίνεται αντιληπτό ότι μέσω του υπερεπίπεδου που ορίζεται από τις παραμέτρους  $\mathbf{w}$  και  $b$  είναι δυνατός ο απόλυτος διαχωρισμός των προτύπων εκπαίδευσης ανάλογα με την κατηγορία στην οποία αυτά ανήκουν, και έτσι εξηγείται η χρήση του όρου «γραμμική διαχωρισιμότητα».

Στην ουσία, για ένα οποιοδήποτε πρότυπο  $\mathbf{x}$ , μπορούμε να υπολογίσουμε την τιμή της συνάρτησης  $y$ , όπως αυτή ορίζεται από την (5.1), και το πρόσημο της ποσότητας  $y(\mathbf{x})$  θα μας δώσει μια εκτίμηση της κατηγορίας στην οποία αυτό ανήκει. Για αυτό το λόγο η συνάρτηση  $y$  αποκαλείται και συνάρτηση απόφασης (decision function).



Τώρα, μπορεί να γίνει εύκολα αντιληπτό ότι στην περίπτωση που τα πρότυπα εκπαίδευσης μπορούν να διαχωριστούν γραμμικά ανάλογα με την κατηγορίας τους, είναι πολύ πιθανόν να υπάρχουν πολλά υπερεπίπεδα που να επιτυγχάνουν αυτόν το διαχωρισμό. Με άλλα λόγια, ενδέχεται να υπάρχουν πολλές επιλογές για τις παραμέτρους  $w$  και  $b$  της σχέσης (5.1) που να επαληθεύουν τις σχέσεις (5.2). Το πρόβλημα, λοιπόν, μπορεί να επιδέχεται περισσότερων της μιας λύσεων. Πολλοί αλγόριθμοι μάθησης έχουν προταθεί προκειμένου να προσδιοριστεί μια από αυτές τις δυνατές λύσεις. Για παράδειγμα, ο επαναληπτικός αλγόριθμος perceptron εγγυάται την εύρεση μιας λύσης σε πεπερασμένο αριθμό βημάτων. Δυστυχώς, όμως, η λύση που θα βρεθεί εξαρτάται από την αρχική εκτίμηση όσο αφορά τις τιμές των παραμέτρων  $w$  και  $b$ , καθώς και από τη σειρά με την οποία παρουσιάζονται τα πρότυπα εκπαίδευσης στον αλγόριθμο, προκειμένου αυτός να ανανεώσει την εκτίμησή του για τις τιμές των παραμέτρων.

Όπως αναφέραμε προηγουμένως, το σύνολο εκπαίδευσης είναι μόνο ένα υποσύνολο του συνόλου δεδομένων που μας ενδιαφέρει. Ο πραγματικός μας στόχος είναι να είμαστε σε θέση μετά από την εκπαίδευση πάνω στο σύνολο εκπαίδευσης να εκτιμήσουμε την κατηγορία άγνωστων προτύπων, δηλαδή προτύπων που δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση, με τη μικρότερη δυνατή πιθανότητα σφάλματος. Επομένως, αν υπάρχουν πολλές λύσεις για το πρόβλημα της δυαδικής ταξινόμησης, η λύση που είναι προτιμότερη είναι εκείνη που δίνει το μικρότερο σφάλμα γενίκευσης (generalization error). Από αυτή τη σκοπιά, στην αμέσως επόμενη ενότητα θα δούμε πως αντιμετωπίζεται το θέμα αυτό από ένα από τα πιο δημοφιλή μοντέλα μάθησης, και αναφερόμαστε βεβαίως στα SVMs.

Επιπλέον, μπορεί κανείς να καταλάβει εύκολα ότι η περίπτωση κατά την οποία είναι δυνατός ο γραμμικός διαχωρισμός των προτύπων των διαφορετικών κατηγοριών είναι σπάνια, και αφορά μόνο απλά προβλήματα ταξινόμησης. Αντίθετα, τα πιο ενδιαφέροντα προβλήματα περιλαμβάνουν σύνολα δεδομένων στα οποία οι κατηγορίες δεν είναι γραμμικά διαχωρίσιμες. Επομένως, πρέπει με κάποιο τρόπο να αντιμετωπιστεί το πρόβλημα στη γενικότητά του, παρακάμπτοντας την εξαιρετικά περιοριστική υπόθεση της γραμμικής διαχωρισιμότητας και επιτρέποντας τον προσδιορισμό μη γραμμικών ορίων μεταξύ των κατηγοριών.

### 5.3 Μηχανές Διανυσμάτων Υποστήριξης

Σε αυτήν την ενότητα θα περιγράψουμε τα μοντέλα που αποκαλούνται SVMs, και τα οποία έχουν χρησιμοποιηθεί με ιδιαίτερη επιτυχία τόσο σε προβλήματα ταξινόμησης όσο και σε προβλήματα παλινδρόμησης (regression).

Ας συνεχίσουμε, όμως, για λίγο ακόμα πάνω στη λογική της γραμμικής διαχωρισιμότητας των κατηγοριών. Το ζήτημα της ελαχιστοποίησης του σφάλματος γενίκευσης, που όπως είπαμε λίγο παραπάνω μπορεί να χρησιμεύσει ως κριτήριο για την επιλογή μιας από τις εν γένει πολλές δυνατές λύσεις ενός προβλήματος ταξινόμησης, στο πλαίσιο των SVMs αντιμετωπίζεται μέσω της μεγιστοποίησης του περιθωρίου (margin). Το περιθώριο ορίζεται ως η ελάχιστη απόσταση ανάμεσα στο όριο και σε κάθε ένα από τα πρότυπα εκπαίδευσης. Επομένως, ως βέλτιστο υπερεπίπεδο διαχωρισμού θεωρείται εκείνο που, από τη μια, διαχωρίζει τα πρότυπα εκπαίδευσης ανάλογα με την κατηγορία τους και, από την άλλη, βρίσκεται όσο γίνεται περισσότερο μακριά από τα πρότυπα αυτά. Θεωρώντας αυτό το βέλτιστο υπερεπίπεδο, τα πρότυπα εκπαίδευσης που απέχουν από αυτό την ελάχιστη απόσταση σε σχέση με όλα τα υπόλοιπα, δηλαδή τα πρότυπα που απέχουν απόσταση ίση με το περιθώριο, ονομάζονται διανύσματα υποστήριξης (support vectors).

Θεωρώντας το υπερεπίπεδο  $y(\mathbf{x})=0$ , με τη συνάρτηση  $y$  να ορίζεται από τη σχέση (5.1), η (κάθετη) απόσταση ενός διανύσματος  $\mathbf{x}$  από το υπερεπίπεδο ισούται με  $|y(\mathbf{x})|/\|\mathbf{w}\|$ ,

όπου  $\|\mathbf{w}\|$  είναι το μέτρο του διανύσματος  $\mathbf{w}$ . Έτσι, η απαίτηση για μεγιστοποίηση του περιθωρίου, που τίθεται στο πλαίσιο των SVMs, εκφράζεται μαθηματικά ως εξής:

$$\max_{\mathbf{w}, b} \left\{ \min_i \frac{t_i y(\mathbf{x}_i)}{\|\mathbf{w}\|} \right\} = \max_{\mathbf{w}, b} \left\{ \frac{1}{\|\mathbf{w}\|} \min_i [t_i y(\mathbf{x}_i)] \right\} \quad (5.3)$$

όπου, φυσικά, θεωρούμε  $t_i y(\mathbf{x}_i) = |y(\mathbf{x}_i)| > 0$ , αφού μας ενδιαφέρουν, όπως είπαμε, μόνο τα υπερεπίπεδα που διαχωρίζουν γραμμικά το σύνολο εκπαίδευσης. Το δεξιό μέλος της (5.3) προκύπτει από το αριστερό, αν λάβουμε υπόψιν ότι το κάθετο διάνυσμα  $\mathbf{w}$  δεν εξαρτάται από το  $i$ .

Χωρίς να θυσιάσουμε τη γενικότητα της λύσης, μπορούμε να θεωρήσουμε ότι για τα κοντινότερα στο όριο πρότυπα ισχύει

$$t_i y(\mathbf{x}_i) = 1 \quad (5.4)$$

και, επομένως, για κάθε ένα από τα πρότυπα εκπαίδευσης θα έχουμε

$$t_i y(\mathbf{x}_i) \geq 1 \quad (5.5)$$

Με βάση αυτήν την κανονικοποίηση προκύπτει ότι τα διανύσματα υποστήριξης είναι εκείνα ακριβώς τα πρότυπα εκπαίδευσης για τα οποία η απόλυτη τιμή της συνάρτησης απόφασης ισούται με τη μονάδα, και ότι το περιθώριο αντιστοιχεί στην απόσταση των σημείων για τα οποία ισχύει  $|y(\mathbf{x})| = 1$  από το υπερεπίπεδο  $y(\mathbf{x}) = 0$ .

Λαμβάνοντας υπόψιν τα παραπάνω, το πρόβλημα βελτιστοποίησης (5.3) απλοποιείται, και έγκειται πλέον στη μεγιστοποίηση του  $\|\mathbf{w}\|^{-1}$ , πράγμα το οποίο ισοδυναμεί με την ελαχιστοποίηση του  $\|\mathbf{w}\|^2$ . Επομένως, το πρόβλημα της μεγιστοποίησης του περιθωρίου μπορεί να διατυπωθεί στην εξής μορφή:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (5.6)$$

υπό τους περιορισμούς (5.5). Η παράμετρος  $b$  περιλαμβάνεται στους περιορισμούς, οπότε, προφανώς, δε μπορεί να απαλειφθεί από τη βελτιστοποίηση.

### 5.3.1 Συνάρτηση Πυρήνα

Αναφέραμε προηγουμένως ότι η υπόθεση της γραμμικής διαχωρισιμότητας των κατηγοριών είναι εξαιρετικά περιοριστική, και ότι τα περισσότερα προβλήματα ταξινόμησης που καλούμαστε να επιλύσουμε στην πράξη περιλαμβάνουν κατηγορίες που δεν είναι γραμμικά διαχωρίσιμες.

Πρέπει, λοιπόν, με κάποιο τρόπο να καταστεί δυνατή η χρήση κάποιου μη γραμμικού ορίου για το διαχωρισμό των κατηγοριών, στη θέση του γραμμικού υπερεπιπέδου που μας απασχόλησε μέχρι τώρα.

Στα SVMs κάτι τέτοιο επιτυγχάνεται μέσω μιας απεικόνισης των προτύπων σε ένα νέο χώρο, που ονομάζεται χώρος πυρήνα (kernel space), ο οποίος μπορεί να είναι μη γραμμικός

και πολύ μεγαλύτερης διάστασης από τον αρχικό χώρο ορισμού των προτύπων. Η απεικόνιση αυτή:

$$\mathbf{x} \mapsto \phi(\mathbf{x}) \quad (5.7)$$

μας επιτρέπει να χειριστούμε μη γραμμικά διαχωρίσιμα πρόβλημα με έναν πολύ απλό τρόπο. Συγκεκριμένα, υποθέτοντας ότι οι κατηγορίες των προτύπων στο χώρο πυρήνα είναι γραμμικά διαχωρίσιμες, αρκεί να λύσουμε το πρόβλημα στο χώρο αυτόν και μετά να απεικονίσουμε τη λύση στον αρχικό χώρο. Η συνάρτηση απόφασης σε αυτήν την περίπτωση θα πάρει τη μορφή

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (5.8)$$

Προφανώς, στο χώρο πυρήνα τα σημεία  $\phi(\mathbf{x})$  για τα οποία ισχύει  $y(\mathbf{x}) = 0$  σχηματίζουν ένα υπερεπίπεδο, όμως στον αρχικό χώρο η επιφάνεια  $y(\mathbf{x}) = 0$  θα είναι μη γραμμική στην περίπτωση που η συνάρτηση απεικόνισης  $\phi$  είναι μια μη γραμμική συνάρτηση.

Μια έννοια που συνδέεται άμεσα με την απεικόνιση των προτύπων του προβλήματος ταξινόμησης από τον αρχικό χώρο στο χώρο πυρήνα, και η οποία θα μας φανεί ιδιαίτερα χρήσιμη παρακάτω, αφορά τη συνάρτηση πυρήνα (kernel function). Αυτή δεν είναι τίποτα άλλο από το εσωτερικό γινόμενο των απεικονίσεων στο χώρο πυρήνα. Δηλαδή, αν  $k(\mathbf{x}_1, \mathbf{x}_2)$  είναι η τιμή της συνάρτησης αυτής για δύο πρότυπα  $\mathbf{x}_1$  και  $\mathbf{x}_2$ , τότε ισχύει

$$k(\mathbf{x}_1, \mathbf{x}_2) = \phi(\mathbf{x}_1)^T \phi(\mathbf{x}_2) \quad (5.9)$$

Ο προηγούμενος ορισμός έχει νόημα στην περίπτωση που γνωρίζουμε τη συνάρτηση απεικόνισης  $\phi$ . Όμως, το πλεονέκτημα που απορρέει από τη χρήση των SVMs αφορά τις περιπτώσεις που είτε δε γνωρίζουμε καθόλου τη συνάρτηση  $\phi$ , είτε ο νέος χώρος που ορίζεται από τη  $\phi$  είναι πολύ μεγάλης ή και άπειρης διάστασης. Τότε, όπως θα δούμε στην επόμενη ενότητα, αρκεί η γνώση της συνάρτησης πυρήνα για τη λύση του προβλήματος ταξινόμησης στο πλαίσιο της μεγιστοποίησης του περιθωρίου.

Από τα παραπάνω διαφαίνεται ότι στο πλαίσιο των SVMs η μόνη πληροφορία για το χώρο πυρήνα που μας είναι απαραίτητη είναι η συνάρτηση πυρήνα. Στην πράξη, μεγάλο ενδιαφέρον παρουσιάζει και είναι δημοφιλής η απευθείας χρήση συναρτήσεων πυρήνα για την ποσοτικοποίηση της ομοιότητας των απεικονίσεων στο νέο χώρο των προτύπων, αδιαφορώντας πλήρως για τον ίδιο το χώρο πυρήνα. Προφανώς, οι χρησιμοποιούμενες συναρτήσεις πυρήνα οφείλουν να έχουν κάποιες ιδιότητες που θα τις καθιστούν αποδεκτές ως τέτοιες. Για παράδειγμα, βλέπουμε από τον ορισμό (5.9) ότι η συνάρτηση πυρήνα πρέπει να είναι μια συμμετρική συνάρτηση, δηλαδή πρέπει να ισχύει

$$k(\mathbf{x}_1, \mathbf{x}_2) = k(\mathbf{x}_2, \mathbf{x}_1) \quad (5.10)$$

Επίσης, αποδεικνύεται ότι μια ικανή και αναγκαία συνθήκη ώστε μια συνάρτηση πυρήνα να είναι αποδεκτή, είναι ο πίνακας Gram, ο οποίος ορίζεται ως ο συμμετρικός πίνακας με στοιχεία

$$K_{nm} = k(\mathbf{x}_n, \mathbf{x}_m) \quad (5.11)$$

δηλαδή είναι ο πίνακας που περιέχει τις τιμές της συνάρτησης πυρήνα για όλα τα ζεύγη προτύπων ενός συνόλου προτύπων, να είναι θετικά ημιορισμένος (positive semidefinite) για κάθε πιθανό σύνολο προτύπων. Υπενθυμίζουμε εδώ ότι θετικά ημιορισμένος είναι ο πίνακας του οποίου όλες οι ιδιοτιμές είναι μη αρνητικές.

Η πιο δημοφιλής μη γραμμική συνάρτηση πυρήνα, που χρησιμοποιείται για τη λύση πολλών προβλημάτων ταξινόμησης στο πλαίσιο των SVMs, είναι η Γκαουσιανή συνάρτηση ακτινικής βάσης (Gaussian Radial Basis Function – RBF). Η συγκεκριμένη συνάρτηση ορίζεται ως

$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|^2) \quad (5.12)$$

όπου  $\|\mathbf{x}_1 - \mathbf{x}_2\|$  είναι η Ευκλείδεια απόσταση μεταξύ των  $d$ -διάστατων διανυσμάτων  $\mathbf{x}_1$  και  $\mathbf{x}_2$ :

$$\|\mathbf{x}_1 - \mathbf{x}_2\| = \sqrt{(x_{11} - x_{21})^2 + \dots + (x_{1d} - x_{2d})^2} \quad (5.13)$$

και  $\gamma$  μια παράμετρος που σχετίζεται, προφανώς, με το πλάτος του Γκαουσιανού πυρήνα. Η καταλληλότερη τιμή για την παράμετρο αυτή εξαρτάται από το εκάστοτε πρόβλημα προς επίλυση, και ιδίως από τις τιμές που λαμβάνουν τα τετράγωνα των Ευκλείδειων αποστάσεων,  $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ , για τα πρότυπα του συνόλου δεδομένων. Η συγκεκριμένη συνάρτηση πυρήνα έχει αποδειχτεί ότι αντιστοιχεί σε απειροδιάστατο χώρο πυρήνα, δηλαδή το νέο διάνυσμα  $\phi(\mathbf{x})$ , μετά τον προφανώς μη γραμμικό μετασχηματισμό του προτύπου  $\mathbf{x}$ , έχει άπειρες διαστάσεις.

Στην περίπτωση που επιχειρείται η λύση του προβλήματος ταξινόμησης στον αρχικό χώρο ορισμού των προτύπων, τότε για τη συνάρτηση απεικόνισης ισχύει

$$\phi(\mathbf{x}) = \mathbf{x} \quad (5.14)$$

και, επομένως, η αντίστοιχη συνάρτηση πυρήνα είναι η

$$k(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T \mathbf{x}_2 \quad (5.15)$$

Άλλες δημοφιλείς συναρτήσεις πυρήνα, εκτός από τις παραπάνω, είναι η πολυωνυμική, που ορίζεται από τη σχέση

$$k(\mathbf{x}_1, \mathbf{x}_2) = (\gamma \mathbf{x}_1^T \mathbf{x}_2 + \beta)^p \quad (5.16)$$

η Λαπλασιανή συνάρτηση ακτινικής βάσης:

$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|) \quad (5.17)$$

και η σιγμοειδής:

$$k(\mathbf{x}_1, \mathbf{x}_2) = \tanh(\gamma \mathbf{x}_1^T \mathbf{x}_2 + \beta) \quad (5.18)$$

Στις παραπάνω σχέσεις, με  $\tanh$  συμβολίζουμε τη συνάρτηση της υπερβολικής εφαπτομένης, ενώ οι  $\gamma$ ,  $\beta$  και  $\rho$  είναι παράμετροι των συναρτήσεων πυρήνα, οι τιμές των οποίων πρέπει να επιλεγούν με βάση το εκάστοτε πρόβλημα προς επίλυση.

### 5.3.2 Πρόβλημα Βελτιστοποίησης

Στην ενότητα αυτή θα ασχοληθούμε με την επίλυση του προβλήματος βελτιστοποίησης που συνδέεται με τη μεγιστοποίηση του περιθωρίου, που επιδιώκεται από τα SVMs. Θεωρώντας ότι τα πρότυπα έχουν απεικονιστεί ήδη στο χώρο πυρήνα, το ζητούμενο είναι ο γραμμικός διαχωρισμός των κατηγοριών σε αυτόν το νέο χώρο.

Το πρόβλημα βελτιστοποίησης (5.6) υπό τους περιορισμούς (5.5) είναι ένα πρόβλημα τετραγωνικού προγραμματισμού (quadratic programming problem). Για την επίλυσή του μπορεί να χρησιμοποιηθεί η μέθοδος των πολλαπλασιαστών Lagrange. Θεωρώντας τη συνάρτηση απόφασης όπως αυτή ορίζεται από τη σχέση (5.8), προκύπτει η εξής συνάρτηση Lagrange:

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N a_i [t_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) - 1] \quad (5.19)$$

όπου το διάνυσμα  $\mathbf{a} = (a_1, \dots, a_N)$ , με  $a_i \geq 0$ , είναι το διάνυσμα των συντελεστών Lagrange, που κάθε ένας αντιστοιχεί και σε μια από τις γραμμικές ανισότητες που αποτελούν τους περιορισμούς της βελτιστοποίησης. Με το πρόβλημα σε αυτή τη μορφή, το ζητούμενο είναι ο προσδιορισμός των τιμών των παραμέτρων  $\mathbf{w}$ ,  $b$  και  $\mathbf{a}$  που ελαχιστοποιούν τη συνάρτηση  $L$ . Ως γνωστόν, η ελαχιστοποίηση μιας συνάρτησης συνεπάγεται το μηδενισμό των παραγώγων της στα σημεία στα οποία αυτή παίρνει την ελάχιστη τιμή. Επομένως, παραγωγίζοντας ως προς  $\mathbf{w}$  και  $b$ , θέτοντας τις μερικές παραγώγους ίσες με το μηδέν και αντικαθιστώντας τις σχέσεις που προκύπτουν στην (5.19), παίρνουμε τη δυαδική αναπαράσταση του προβλήματος, στην οποία τα  $\mathbf{w}$  και  $b$  έχουν απαλειφθεί. Συγκεκριμένα, στη νέα του μορφή το πρόβλημα διατυπώνεται ως

$$\max \left[ \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j t_i t_j k(\mathbf{x}_i, \mathbf{x}_j) \right] \text{ ως προς } a_i, i = 1, \dots, N \quad (5.20)$$

υπό τους περιορισμούς

$$a_i \geq 0 \text{ και } \sum_{i=1}^N a_i t_i = 0 \quad (5.21)$$

Μετά από τον προσδιορισμό των τιμών των συντελεστών Lagrange που επιλύουν το νέο αυτό πρόβλημα τετραγωνικού προγραμματισμού, η συνάρτηση απόφασης, που μπορεί να χρησιμοποιηθεί για την εκτίμηση της κατηγορίας κάποιου άγνωστου προτύπου  $\mathbf{x}$  το οποίο θα παρουσιαστεί προς κατηγοριοποίηση στον ταξινομητή, δίνεται από τη σχέση

$$y(\mathbf{x}) = \sum_{i=1}^N a_i t_i k(\mathbf{x}, \mathbf{x}_i) + b \quad (5.22)$$

Ακόμα, από τη μορφή του προβλήματος βελτιστοποίησης που αντιμετωπίζουμε προκύπτουν κάποια πολύ ενδιαφέροντα συμπεράσματα. Όπως είπαμε, ο συντελεστής Lagrange  $a_i$  αντιστοιχεί στην  $i$ -οστή ανισότητα των περιορισμών,  $t_i y(\mathbf{x}_i) \geq 1$ , η οποία με τη σειρά της εμπλέκει το  $i$ -οστό πρότυπο εκπαίδευσης,  $\mathbf{x}_i$ . Αποδεικνύεται ότι όσο αφορά τους συντελεστές Lagrange και τα πρότυπα εκπαίδευσης ισχύει είτε  $a_i = 0$  είτε  $t_i y(\mathbf{x}_i) = 1$ . Δηλαδή, με απλά λόγια, μετά τη βελτιστοποίηση, μη μηδενικοί συντελεστές Lagrange θα αντιστοιχούν μόνο στα πρότυπα εκείνα που βρίσκονται πάνω στις επιφάνειες μέγιστου περιθωρίου  $y(\mathbf{x}) = 1$  και  $y(\mathbf{x}) = -1$ , δηλαδή στα διανύσματα υποστήριξης. Σε όλα τα άλλα πρότυπα, που συνήθως είναι πολύ περισσότερα από τα διανύσματα υποστήριξης, θα αντιστοιχούν μηδενικοί συντελεστές Lagrange, και επομένως αυτά δε θα συνεισφέρουν καθόλου στον υπολογισμό της τιμής της συνάρτησης απόφασης για την κατηγοριοποίηση των νέων προτύπων που θα παρουσιάζονται στον ταξινομητή.

Για τον υπολογισμό της τιμής της παραμέτρου  $b$ , που μας είναι απαραίτητη για την πρόβλεψη της κατηγορίας νέων προτύπων όπως βλέπουμε από τη σχέση (5.22), μπορούμε να βασιστούμε στην ιδιότητα  $t_n y(\mathbf{x}_n) = 1$  που χαρακτηρίζει τα διανύσματα υποστήριξης. Για ένα διάνυσμα υποστήριξης, έστω  $\mathbf{x}_n$ , έχουμε επομένως τη σχέση

$$t_n \left( \sum_{i=1}^N a_i t_i k(\mathbf{x}_n, \mathbf{x}_i) + b \right) = 1 \quad (5.23)$$

Επιλύοντας την παραπάνω σχέση ως προς  $b$ , ή, για επίτευξη μεγαλύτερης ευρωστίας, πολλαπλασιάζοντάς την με  $t_n$  (και λαμβάνοντας υπόψιν ότι  $t_n^2 = 1$ ) και παίρνοντας τον μέσο όρο για όλα τα διανύσματα υποστήριξης  $\mathbf{x}_n$ , μπορούμε να υπολογίσουμε την τιμή του  $b$ . Στη δεύτερη περίπτωση για το  $b$  προκύπτει η σχέση

$$b = \frac{1}{N_S} \sum_{n \in S} \left( t_n - \sum_{i=1}^N a_i t_i k(\mathbf{x}_n, \mathbf{x}_i) \right) \quad (5.24)$$

όπου  $S$  είναι το σύνολο των δεικτών που αντιστοιχούν στα διανύσματα υποστήριξης και το  $N_S$  ισούται με τον αριθμό των διανυσμάτων υποστήριξης.

Από όλη την παραπάνω ανάλυση βλέπουμε, επομένως, ότι από τη στιγμή που μετασχηματίσαμε το πρόβλημα στη μορφή που ορίζεται από τις σχέσεις (5.20)-(5.21), καταφέραμε να απαλλαγούμε τελείως από την ανάγκη γνώσης της συνάρτησης απεικόνισης  $\phi$ . Δε μας χρειάζεται δηλαδή, σε τελική ανάλυση, η άμεση γνώση του χώρου πυρήνα για την επίλυση του προβλήματος. Το μόνο που μας χρειάζεται είναι η αντίστοιχη συνάρτηση πυρήνα, δηλαδή το εσωτερικό γινόμενο στο χώρο πυρήνα συναρτήσει της αναπαράστασης των προτύπων στον αρχικό χώρο.

### 5.3.3 Επικαλυπτόμενες Κατηγορίες

Μέχρι τώρα υποθέσαμε ότι τα πρότυπα εκπαίδευσης είναι γραμμικά διαχωρίσιμα με βάση την κατηγορία τους, τουλάχιστον στον εν γένει μη γραμμικό χώρο πυρήνα. Υπό αυτήν την προϋπόθεση, μέσω της εκπαίδευσης ενός SVM όπως αυτό που περιγράφηκε παραπάνω, θα επιτευχθεί ακριβής διαχωρισμός των προτύπων στον αρχικό χώρο, αλλά μέσω ενός μη γραμμικού ορίου μεταξύ των δύο κατηγοριών.

Όμως, σε πραγματικά προβλήματα ενδέχεται οι υπό θεώρηση κατηγορίες να εμφανίζουν μια φυσική επικάλυψη στο χώρο των προτύπων. Δηλαδή, είναι σύνηθες το φαινόμενο οι κατανομές των προτύπων που αντιστοιχούν στις δύο κατηγορίες που αντιμετωπίζουμε να είναι επικαλυπτόμενες. Σε αυτήν την περίπτωση, η επιμονή στη θέση μας για απόλυτα ακριβή διαχωρισμό των προτύπων εκπαίδευσης που αντιστοιχούν στη μια κατηγορία από αυτά που αντιστοιχούν στην άλλη, ενδέχεται να οδηγήσει στην εκπαίδευση ενός ιδιαίτερα πολύπλοκου μοντέλου, υπερβολικά προσαρμοσμένου στα πρότυπα εκπαίδευσης, που θα παρουσιάζει μη ικανοποιητικές δυνατότητες γενίκευσης σε πρότυπα για τα οποία δεν έχει εκπαιδευτεί. Έχουμε εδώ πάλι την εμφάνιση του φαινομένου της υπερεκπαίδευσης, στο οποίο αναφερθήκαμε και όταν ασχολούμασταν με τα Γκαουσιανά Μείγματα.

Επιβάλλεται, λοιπόν, να επανατοποθετήσουμε το πρόβλημα της ταξινόμησης με τέτοιο τρόπο, ώστε να προβλέψουμε και να μπορέσουμε να μοντελοποιήσουμε την ύπαρξη μιας κάποιας επικάλυψης για τις κατηγορίες των προτύπων.

Στο πλαίσιο των SVMs αυτό γίνεται μέσω της εισαγωγής για κάθε πρότυπο εκπαίδευσης  $\mathbf{x}_i$  μιας αντίστοιχης μεταβλητής χαλάρωσης (slack variable)  $\xi_i$ . Στην περίπτωση που το πρότυπο εκπαίδευσης βρίσκεται, σε σχέση με το όριο μεταξύ των κατηγοριών, στην πλευρά που αντιστοιχεί πράγματι στην κατηγορία του και, παράλληλα, χαρακτηρίζεται από μια τιμή για τη συνάρτηση απόφασης τέτοια ώστε να ισχύει  $t_i y(\mathbf{x}_i) \geq 1$ , βρίσκεται δηλαδή έξω από το χώρο που ορίζεται μεταξύ των επιφανειών περιθωρίου  $y(\mathbf{x})=1$  και  $y(\mathbf{x})=-1$ , τότε για τη μεταβλητή χαλάρωσης  $\xi_i$  έχουμε  $\xi_i = 0$ . Σε αντίθετη περίπτωση έχουμε  $\xi_i = |t_i - y(\mathbf{x}_i)|$ , και, επομένως, για πρότυπα μεταξύ των επιφανειών περιθωρίου αλλά ορθά κατηγοριοποιημένα προκύπτει  $0 < \xi_i \leq 1$ , ενώ για πρότυπα λάθος κατηγοριοποιημένα παίρνουμε  $\xi_i > 1$ . Για κάθε πρότυπο εκπαίδευσης, λοιπόν, αντί της σχέσεως (5.5), θα ισχύει ο εξής νέος περιορισμός:

$$t_i y(\mathbf{x}_i) \geq 1 - \xi_i \quad (5.25)$$

Όλη η ιδέα γύρω από αυτήν την ανάλυση έχει να κάνει με τη χαλάρωση του ορισμού της έννοιας του περιθωρίου, και τον ορισμό ενός χαλαρού περιθωρίου, ο οποίος θα επιτρέπει πρότυπα εκπαίδευσης να βρίσκονται σε πιο κοντινή απόσταση από το όριο σε σχέση με την απόσταση περιθωρίου, αλλά ακόμα και πέρα από το όριο στην περιοχή της αντίθετης κατηγορίας, δηλαδή θα επιτρέπει μη ορθά κατηγοριοποιημένα πρότυπα.

Το πρόβλημα βελτιστοποίησης που θα πρέπει να επιλυθεί σε αυτήν την περίπτωση δε θα αφορά μόνο τη μεγιστοποίηση του περιθωρίου, αλλά, προφανώς, και την παράλληλη ελαχιστοποίηση της απόκλισης από τον αυστηρό ορισμό του περιθωρίου. Αυτή η απόκλιση εκφράζεται μέσω των τιμών των μεταβλητών  $\xi_i$  που αντιστοιχούν στα πρότυπα εκπαίδευσης. Για παράδειγμα, αν όλες οι μεταβλητές  $\xi_i$  είχαν μηδενικές τιμές, τότε όλα τα πρότυπα θα ήταν ορθά κατηγοριοποιημένα και είτε πάνω στις επιφάνειες περιθωρίου είτε πιο μακριά ακόμα από το όριο, οπότε θα μιλούσαμε για τον κλασικό ορισμό του περιθωρίου. Επομένως, αντί για την ποσότητα της σχέσεως (5.6), θέλουμε να ελαχιστοποιήσουμε την ποσότητα

$$C \sum_{i=1}^N \xi_i + \frac{1}{2} \|\mathbf{w}\|^2 \quad (5.26)$$

όπου η παράμετρος  $C > 0$  λειτουργεί ως παράγοντας εξισορρόπησης μεταξύ της τάσης ελαχιστοποίησης του σφάλματος εκπαίδευσης (όπως αυτό εκφράζεται από το  $\sum_i \xi_i$ ) και της τάσης ελαχιστοποίησης της πολυπλοκότητας του μοντέλου. Για παράδειγμα, μια τιμή για το  $C$  πολύ μεγάλη (θεωρητικά άπειρη) θα απαγορεύει πρακτικά οποιαδήποτε μη μηδενική τιμή για τα  $\xi_i$ , μέσω της συνεπακόλουθης ραγδαίας αύξησης της ποσότητας προς ελαχιστοποίηση που ορίζεται από τη σχέση (5.26), με αποτέλεσμα να καταλήξουμε πάλι στην ίδια λύση που θα παίρναμε και με τον αυστηρό ορισμό του περιθωρίου.

Χρησιμοποιώντας και αυτή τη φορά την τεχνική των συντελεστών Lagrange, και βασιζόμενοι σε παρόμοιες ιδέες με αυτές που σκιαγραφήθηκαν στην ενότητα 5.3.2, καταλήγουμε σε ένα μετασχηματισμό του προβλήματος βελτιστοποίησης, που εμπλέκει αποκλειστικά τη συνάρτηση πυρήνα και μας απαλλάσσει από την άμεση γνώση του ίδιου του χώρου πυρήνα. Συγκεκριμένα, αποδεικνύεται ότι το τελικό πρόβλημα βελτιστοποίησης ορίζεται πάλι από τη σχέση (5.20), και η μόνη διαφορά έγκειται στους αντίστοιχους περιορισμούς. Αντί, λοιπόν, των περιορισμών (5.21) έχουμε στην περίπτωση του χαλαρού περιθωρίου τους εξής περιορισμούς στο πρόβλημα τετραγωνικού προγραμματισμού:

$$0 \leq a_i \leq C \quad \text{και} \quad \sum_{i=1}^N a_i t_i = 0 \quad (5.27)$$

Ως διανύσματα υποστήριξης ορίζουμε, πάλι, αυτά για τα οποία ισχύει  $a_i > 0$ . Από όλα τα διανύσματα υποστήριξης, εκείνα για τα οποία ισχύει  $a_i < C$  βρίσκονται πάνω στο περιθώριο. Αντίθετα, τα διανύσματα υποστήριξης με  $a_i = C$  μπορεί είτε να είναι ορθά κατηγοριοποιημένα αλλά να βρίσκονται μεταξύ επιφάνειας περιθωρίου και ορίου διαχωρισμού, είτε να είναι μη ορθά κατηγοριοποιημένα.

Επίσης, πρέπει να αναφέρουμε ότι η συνάρτηση απόφασης ορίζεται και σε αυτήν την περίπτωση μέσω του τύπου (5.22) και ότι η τιμή της παραμέτρου  $b$  μπορεί να υπολογιστεί πάλι από τον τύπο (5.24), μόνο που τώρα με  $S$  και  $N_S$  θα συμβολίζουμε το σύνολο των δεικτών και το πλήθος, αντίστοιχα, όχι όλων των διανυσμάτων υποστήριξης, αλλά μόνο εκείνων για τα οποία ισχύει  $a_i < C$ , δηλαδή εκείνων που βρίσκονται πάνω στις επιφάνειες περιθωρίου, γιατί μόνο για αυτά ισχύει η σχέση  $t_n y(\mathbf{x}_n) = 1$ , που είναι αυτή στην οποία στηριζόμαστε για την εξαγωγή του τύπου (5.24).

## 5.4 Η Συσχετιστική Ανατροφοδότηση ως Πρόβλημα Δυναμικής Ταξινόμησης

Όσο αφορά τη συσχετιστική ανατροφοδότηση, αυτή προϋποθέτει, όπως είπαμε αρκετές φορές μέχρι τώρα, την υποβολή από το χρήστη σε κάθε γύρο της διαδικασίας αξιολογήσεων πάνω στα αποτελέσματα της ανάκτησης. Στην περίπτωση, για παράδειγμα, της συσχετιστικής ανατροφοδότησης στα πλαίσια της ανάκτησης εικόνων, οι αξιολογήσεις αυτές συνήθως έχουν να κάνουν με την επιλογή ενός αριθμού εικόνων από το χρήστη, από τις πιο πρόσφατες εικόνες – αποτελέσματα που επέστρεψε το σύστημα, και με την ρητή δήλωση από τον τελευταίο για το αν η καθεμιά από τις εικόνες που επιλέχθηκαν είναι σχετική ή όχι με αυτό που αναζητεί. Αυτές οι εικόνες, μαζί με την κρίση του χρήστη που τις συνοδεύει, ανατροφοδοτούνται στο σύστημα, προκειμένου αυτό να ανανεώσει την εκτίμησή του όσο αφορά τα κριτήρια συνάφειας μεταξύ των εικόνων που χρησιμοποιεί, και να επιστρέψει ένα νέο σύνολο αποτελεσμάτων.



Επομένως, σε κάθε γύρο της διαδικασίας ένας αριθμός από σχετικές κατά την κρίση του χρήστη εικόνες, καθώς και ένας άλλος αριθμός από αντίστοιχες μη σχετικές υποβάλλονται στο σύστημα ως παραδείγματα των προτιμήσεων του τελευταίου. Διακρίνουμε, λοιπόν, αμέσως δύο διακριτές κατηγορίες εικόνων: τις σχετικές εικόνες και τις μη σχετικές εικόνες. Διακρίνουμε επίσης και ένα σύνολο προτύπων εκπαίδευσης, τα οποία δεν είναι άλλα από τις εικόνες που δίνει ο χρήστης ως παραδείγματα. Οι ετικέτες αυτών των παραδειγμάτων προκύπτουν άμεσα από την κρίση του χρήστη. Η ετικέτα «-1», για παράδειγμα, μπορεί να αντιστοιχεί στις μη σχετικές εικόνες, ενώ η ετικέτα «+1» στις σχετικές. Μάλιστα, καθώς περνούν οι γύροι της συσχετιστικής ανατροφοδότησης, ο χρήστης δίνει όλο και περισσότερα θετικά και αρνητικά παραδείγματα, τα οποία συσσωρεύονται, με αποτέλεσμα σε κάθε γύρο της ανατροφοδότησης το σύνολο των προτύπων εκπαίδευσης να περιλαμβάνει, εκτός από τα παραδείγματα που έδωσε ο χρήστης στο συγκεκριμένο γύρο, και όλα αυτά που υπέβαλε στους γύρους της ανατροφοδότησης που προηγήθηκαν.

Ακόμα, ο στόχος ενός συστήματος ανάκτησης εικόνων, σε τελική ανάλυση, είναι να εκτιμήσει για κάθε εικόνα της βάσης εικόνων, αν είναι σχετική ή όχι με αυτό που ζητάει ο χρήστης, αλλά και κατά πόσο ανταποκρίνεται στις προτιμήσεις του τελευταίου.

Από τα παραπάνω καθίσταται προφανές ότι η όλη διαδικασία ανάκτησης εικόνων με χρήση συσχετιστικής ανατροφοδότησης μπορεί να θεωρηθεί ως μια σειρά από προβλήματα δυαδικής ταξινόμησης, ένα για κάθε γύρο της ανατροφοδότησης, για τη λύση καθενός από τα οποία μπορεί να χρησιμοποιηθεί κάποιος αλγόριθμος επιβλεπόμενης μάθησης με πρότυπα εκπαίδευσης τα παραδείγματα του χρήστη μέχρι και τον τρέχοντα γύρο. Έτσι, σε κάθε γύρο της ανατροφοδότησης, συμπεριλαμβάνοντας στην εκπαίδευση του ταξινομητή μαζί με τα προηγούμενα παραδείγματα αυτά που μόλις έδωσε ο χρήστης, καταλήγουμε με ένα νέο εκπαιδευμένο ταξινομητή, και επομένως με ένα ανανεωμένο όριο διαχωρισμού των σχετικών από τις μη σχετικές εικόνες, στο χώρο περιγραφής των τελευταίων. Αυτό το νέο όριο μπορεί να χρησιμοποιηθεί ώστε να ανανεωθεί η διάταξη των εικόνων της βάσης, και άρα η εκτίμηση του συστήματος όσο αφορά τη συνάφεια των εικόνων της βάσης με την ερώτηση του χρήστη, και να επιστραφεί ένα νέο σύνολο αποτελεσμάτων στον τελευταίο.

Θεωρώντας ότι πριν την έναρξη των γύρων της συσχετιστικής ανατροφοδότησης για την υποβολή της αξιολόγησης του χρήστη πάνω στα τρέχοντα αποτελέσματα της ανάκτησης, η όλη διαδικασία της ανάκτησης έχει ξεκινήσει μέσω της υποβολής από το χρήστη στο σύστημα μιας αρχικής ερώτησης – εικόνας αντιπροσωπευτικής των προτιμήσεών του, είναι προφανές ότι τα πρώτα αποτελέσματα της ανάκτησης δεν μπορεί να προέλθουν από την εκπαίδευση κάποιου ταξινομητή, αφού το μόνο που έχουμε αρχικά είναι ένα θετικό παράδειγμα, η αρχική ερώτηση του χρήστη. Η διάταξη των εικόνων της βάσης θα προέλθει σε αυτήν την αρχική φάση μέσω απλού υπολογισμού ενός μέτρου απόστασης μεταξύ της ερώτησης και κάθε εικόνας της βάσης. Επιπλέον, στη διαδικασία της εκπαίδευσης των ταξινομητών, κατά τη διάρκεια των γύρων της ανατροφοδότησης που θα ακολουθήσουν, αυτή η αρχική ερώτηση θα λαμβάνεται υπόψιν ως ένα ακόμα θετικό παράδειγμα, ανάμεσα στα άλλα που θα έχει υποβάλει ο χρήστης μέσω της προβλεπόμενης διαδικασίας αξιολόγησης των αποτελεσμάτων της ανάκτησης.

#### **5.4.1 Συσχετιστική Ανατροφοδότηση και Μηχανές Διανυσμάτων Υποστήριξης**

Όπως έχουμε ήδη αναφέρει, ένα από τα πιο δημοφιλή μοντέλα μάθησης που μπορούν να χρησιμοποιηθούν στην επίλυση προβλημάτων δυαδικής ταξινόμησης είναι τα SVMs. Από τη στιγμή που, όπως περιγράψαμε παραπάνω, η διαδικασία της συσχετιστικής ανατροφοδότησης μπορεί να αντιμετωπιστεί από τη σκοπιά της επίλυσης προβλημάτων

δυναμικής ταξινόμησης, γίνεται προφανές ότι τα SVMs θέτουν ισχυρή υποψηφιότητα όσο αφορά την υιοθέτησή τους ως εργαλεία για την υλοποίηση μιας τεχνικής συσχετιστικής ανατροφοδότησης.

Σε σχέση με μια τέτοια υλοποίηση, λαμβάνοντας υπόψιν ότι αυτό που στην πράξη απαιτείται σε κάθε γύρο της ανατροφοδότησης είναι μια διάταξη των εικόνων της βάσης σε φθίνουσα σειρά (για παράδειγμα) ως προς την εκτιμώμενη συνάφειά τους με την ερώτηση του χρήστη, το βασικό ερώτημα το οποίο ζητάει απάντηση αφορά το πως θα επιτευχθεί αυτή η διάταξη των εικόνων με βάση το εκπαιδευμένο με τα παραδείγματα που έχουν δοθεί μέχρι και τον τρέχοντα γύρο SVM.

Σχετικά με αυτό, έχουμε αναφέρει προηγουμένως ότι η απόσταση ενός προτύπου  $\mathbf{x}$  από το όριο μεταξύ των δύο κατηγοριών ισούται με  $|y(\mathbf{x})|/\|\mathbf{w}\|$ . Το διάνυσμα  $\mathbf{w}$ , και επομένως και το μέτρο του, είναι ανεξάρτητο του προς κατηγοριοποίηση προτύπου  $\mathbf{x}$ . Άρα, η απόλυτη τιμή της συνάρτησης απόφασης ενός προτύπου  $\mathbf{x}$ ,  $|y(\mathbf{x})|$ , μπορεί να χρησιμοποιηθεί ως μέτρο απόστασής του από το όριο διαχωρισμού, ενώ στην περίπτωση που για δύο πρότυπα  $\mathbf{x}_1$  και  $\mathbf{x}_2$  ισχύει  $|y(\mathbf{x}_1)| > |y(\mathbf{x}_2)|$  μπορούμε να είμαστε σίγουροι ότι το πρώτο απέχει περισσότερο από το όριο από ό,τι το δεύτερο. Επιπλέον, είναι προφανές ότι όσο πιο πολύ απέχει ένα πρότυπο από το όριο, τόσο πιο μεγάλη εμπιστοσύνη μπορούμε να έχουμε όσο αφορά την εκτίμησή μας για την κατηγορία στην οποία αυτό ανήκει. Μεγάλες θετικές τιμές αποτελούν ισχυρή ένδειξη ότι το πρότυπο ανήκει στην κατηγορία που αντιστοιχεί στην ετικέτα «+1», ενώ μικρές αρνητικές τιμές αποτελούν αντίστοιχη ισχυρή ένδειξη υπέρ της κατηγορίας με ετικέτα «-1».

Επομένως, έχουμε έναν απλό τρόπο για να διατάξουμε τις εικόνες μιας βάσης εικόνων σε φθίνουσα σειρά σε σχέση με την εκτιμώμενη συνάφειά τους με αυτό που ζητάει ο χρήστης. Συγκεκριμένα, όσο μεγαλύτερη είναι η τιμή της συνάρτησης απόφασης που προκύπτει με βάση τον εκπαιδευμένο ταξινομητή για μια εικόνα, τόσο πιο σχετική πρέπει η εικόνα αυτή να θεωρείται από το σύστημα, με την προϋπόθεση βέβαια ότι η κατηγορία των σχετικών εικόνων αντιστοιχεί στην ετικέτα «+1» και αυτή των μη σχετικών εικόνων στην ετικέτα «-1».

## Κεφάλαιο 6

# Συσχετιστική Ανατροφοδότηση με Χρήση Ταξινομητών και Γκαουσιανών Μειγμάτων

Σε αυτό το κεφάλαιο θα ασχοληθούμε με το ενδιαφέρον θέμα της χρήσης Γκαουσιανών Μειγμάτων για την αναπαράσταση των εικόνων, σε συνδυασμό με την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης μέσω εκπαίδευσης ταξινομητών σε κάθε γύρο της τελευταίας. Κυρίως, θα μας απασχολήσει το πιο δημοφιλές, ίσως, από όλα τα μοντέλα μάθησης, που όπως έχουμε ήδη αναφέρει είναι το SVM. Η χρήση ταξινομητών έχει οδηγήσει σε ιδιαίτερα ενθαρρυντικά αποτελέσματα όσο αφορά την αντιμετώπιση προβλημάτων ανάκτησης εικόνων. Παράλληλα, τα Γκαουσιανά Μείγματα είναι, όπως έχουμε ήδη επισημάνει και διαπιστώσει, ένα πολύ δυνατό εργαλείο για την πιθανοτική μοντελοποίηση συνόλων δεδομένων. Γίνεται, επομένως, εύκολα αντιληπτό ότι ένας συνδυασμός μεταξύ των παραπάνω μοντέλων μάθησης και των πιθανοτικών μοντέλων αναπαράστασης είναι πολλά υποσχόμενος όσο αφορά τη βελτίωση των αποτελεσμάτων της ανάκτησης εικόνων με βάση το περιεχόμενο. Ως αποτέλεσμα αυτού του σκέλους της έρευνας προέκυψε η δημοσίευση [C3] και η υποβληθείσα προς δημοσίευση εργασία [J2].

### 6.1 Συνδυασμός Γκαουσιανών Μειγμάτων και Μηχανών Διανυσμάτων Υποστήριξης

Στο προηγούμενο κεφάλαιο αναλύσαμε συνοπτικά τη φιλοσοφία γύρω από τα SVMs, και είδαμε με ποιο τρόπο αυτό το μοντέλο μάθησης επιχειρεί να λύσει ένα πρόβλημα δυαδικής ταξινόμησης, παρακάμπτοντας το θέμα της μη διαχωρισιμότητας των κατηγοριών, που εμφανίζεται ιδιαίτερα συχνά σε πραγματικά προβλήματα, αλλά και το εξίσου συχνό και σημαντικό ζήτημα της επικάλυψης των κατηγοριών.

Σε όλη την ανάλυσή μας στο προηγούμενο κεφάλαιο αφήσαμε να εννοηθεί ότι τα πρότυπα πάνω στα οποία εφαρμόζεται ο ταξινομητής SVM είναι πολυδιάστατα, εν γένει, διανύσματα. Όμως, θεωρώντας την τελική μορφή στην οποία τίθεται το πρόβλημα μεγιστοποίησης του περιθωρίου, η οποία περιγράφεται μέσω του προβλήματος βελτιστοποίησης που ορίζεται από τις σχέσεις (5.20) και (5.27) (στη γενική περίπτωση, κατά την οποία λαμβάνουμε υπόψιν την πιθανή ύπαρξη επικαλυπτόμενων κατηγοριών), αλλά και τη λύση όσο αφορά την κατηγοριοποίηση νέων προτύπων, που εκφράζεται μέσω της σχέσης (5.22) για τη συνάρτηση απόφασης, βλέπουμε ότι στην πράξη η μόνη πληροφορία που χρησιμοποιείται και αφορά το χώρο ορισμού των προτύπων έχει να κάνει με τη συνάρτηση πυρήνα,  $k(\mathbf{x}_1, \mathbf{x}_2)$ , που εκφράζει, όπως έχουμε αναφέρει ήδη, το

εσωτερικό γινόμενο στο χώρο στο οποίο απεικονίζονται τα πρότυπα μετά την εφαρμογή για το καθένα του μετασχηματισμού  $\phi$ . Έχει καταστεί σαφές, βεβαίως, ότι ο εν λόγω χώρος και οι εικόνες των προτύπων σε αυτόν είναι πρακτικά αδιάφορος, με την έννοια ότι λόγω του ορισμού του προβλήματος και της λύσης, μόνο η γνώση της συνάρτησης πυρήνα απαιτείται. Η συνάρτηση πυρήνα δε, από τον ορισμό της ως εσωτερικό γινόμενο, είναι ένα μέτρο ομοιότητας, υπολογιζόμενο όμως σε ένα γενικά μη γραμμικό χώρο.

Λαμβάνοντας, λοιπόν, υπόψιν την παραπάνω ανάλυση, γίνεται εύκολα αντιληπτό ότι τα πρότυπα μπορούν να έχουν την οποιαδήποτε μορφή, αρκεί να μπορεί να οριστεί για αυτά μια κατάλληλη συνάρτηση πυρήνα, που θα κωδικοποιεί με εν γένει μη γραμμικό τρόπο την ομοιότητα μεταξύ τους. Επομένως, δεν υπάρχει ουσιαστικός περιορισμός ώστε, κατά τη χρήση ταξινομητών SVM σε προβλήματα ανάκτησης εικόνων, οι εικόνες να αναπαριστούνται ως πολυδιάστατα διανύσματα. Θα μπορούσε κάλλιστα οι τελευταίες να αναπαριστούνται μέσω Γκαουσιανών Μειγμάτων, με τη μόνη προϋπόθεση της ύπαρξης μιας κατάλληλης συνάρτησης πυρήνα που θα κωδικοποιεί την ομοιότητα μεταξύ δύο τέτοιων αναπαραστάσεων.

Θεωρητικά, υπάρχουν κάποιες συνθήκες που πρέπει να ισχύουν ώστε μια συγκεκριμένη συνάρτηση να είναι αποδεκτή ως συνάρτηση πυρήνα. Αναφέραμε και στο προηγούμενο κεφάλαιο την απαίτηση, για παράδειγμα, περί θετικά ημιορισμένου πίνακα Gram. Υπενθυμίζουμε εδώ ότι ο τελευταίος είναι ο συμμετρικός πίνακας με στοιχεία τις τιμές της συνάρτησης πυρήνα για όλα τα ζεύγη ενός συνόλου προτύπων. Πάντως, στην πράξη, πολλές συναρτήσεις πυρήνα έχουν προταθεί στη βιβλιογραφία, που η χρήση τους οδηγεί σε πολύ ικανοποιητικά αποτελέσματα χωρίς να παρέχεται αυστηρή μαθηματική απόδειξη για το ότι ικανοποιούν τέτοιου είδους συνθήκες. Ως παραδείγματα τέτοιων εργασιών αναφέρουμε τις [9], [19], [69], [70] και [76].

Το ζητούμενο, επομένως, είναι να οριστεί μια κατάλληλη συνάρτηση πυρήνα που να εκφράζει την ομοιότητα μεταξύ δύο Γκαουσιανών Μειγμάτων. Στην περίπτωση που αντί για Γκαουσιανά Μείγματα έχουμε κλασικά πρότυπα στη μορφή πολυδιάστατων διανυσμάτων, έχουμε ήδη αναφέρει ότι η πιο δημοφιλής και επιτυχημένη μη γραμμική συνάρτηση πυρήνα είναι η Γκαουσιανή συνάρτηση ακτινικής βάσης, που για δύο πρότυπα  $\mathbf{x}_1$  και  $\mathbf{x}_2$  ορίζεται ως  $k(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|^2)$ . Η ποσότητα  $\|\mathbf{x}_1 - \mathbf{x}_2\|^2$  είναι το τετράγωνο της Ευκλείδειας απόστασης μεταξύ των προτύπων και το  $\gamma$  είναι μια παράμετρος που καθορίζεται ανάλογα με το συγκεκριμένο πρόβλημα ταξινόμησης προς επίλυση.

Βασίζόμενοι σε αυτήν τη μορφή συνάρτησης μπορούμε, λοιπόν, με πολύ διαισθητικό τρόπο να ορίσουμε μια ανάλογη συνάρτηση πυρήνα στην περίπτωση που τα πρότυπα αντί μέσω διανυσμάτων αναπαριστούνται με Γκαουσιανά μείγματα. Συγκεκριμένα, με  $p_1$  και  $p_2$  να είναι συναρτήσεις πυκνότητας πιθανότητας που αντιστοιχούν σε δύο Γκαουσιανά Μείγματα, μια συνάρτηση πυρήνα μεταξύ Γκαουσιανών Μειγμάτων μπορεί να οριστεί ως

$$k(p_1, p_2) = \exp[-\gamma d(p_1, p_2)] \quad (6.1)$$

με το  $d(p_1, p_2)$  να είναι κάποιο μέτρο απόστασης μεταξύ Γκαουσιανών Μειγμάτων.

Επομένως, επιστρέφουμε πάλι στο θέμα που μας απασχόλησε και σε προηγούμενα κεφάλαια, και το οποίο έχει να κάνει με την εύρεση κάποιου μέτρου απόστασης μεταξύ Γκαουσιανών Μειγμάτων με καλές ιδιότητες όσο αφορά τη σύγκριση μεταξύ διαφορετικών μοντέλων, που να μπορεί παράλληλα να υπολογιστεί αποδοτικά, ώστε να μην επιβαρύνονται χρονικά οι μέθοδοι που θα βασίζονται πάνω στον υπολογισμό του. Σε ό,τι αφορά το πρόβλημα της ανάκτησης εικόνων μέσω συσχετιστικής ανατροφοδότησης, ο γρήγορος υπολογισμός αυτών των αποστάσεων είναι επιτακτικός, λόγω της απαίτησης για αλληλεπίδραση με το χρήστη σε πραγματικό χρόνο.

Σε αυτό το κεφάλαιο θα ασχοληθούμε κυρίως με τα μέτρα απόστασης/ομοιότητας μεταξύ Γκαουσιανών Μειγμάτων που ορίζονται στις εργασίες [18] και [54], τα οποία και σε διάφορα προκαταρκτικά πειράματα που διεξήχθησαν μας έδωσαν τα καλύτερα αποτελέσματα όσο αφορά την ενσωμάτωση των Γκαουσιανών Μειγμάτων στο πλαίσιο των SVMs. Τα μέτρα αυτά, που παρουσιάστηκαν αναλυτικά στο κεφάλαιο 3, αποτελούν προσεγγίσεις της απόκλισης KL μεταξύ Γκαουσιανών Μειγμάτων, η οποία, όπως έχουμε αναφέρει, δε μπορεί να υπολογιστεί σε κλειστή μορφή, με αποτέλεσμα να καθίσταται μη πρακτική ως προς τη χρήση της για προβλήματα ανάκτησης εικόνων. Τα μέτρα απόστασης που προκύπτουν από τις δύο παραπάνω εργασίες δίνουν μια κάποια λύση σε αυτό το πρακτικό πρόβλημα υπολογισμού της απόκλισης KL, προσεγγίζοντας, όπως βλέπουμε στις σχέσεις (3.27) και (3.14) για το πρώτο και (3.28) και (3.29) για το δεύτερο, την τελευταία μέσω των τιμών της απόκλισης KL μεταξύ γειτονικών Γκαουσιανών συνιστωσών των δύο μειγμάτων. Προκειμένου να προσδιορίσουν αυτά τα ζεύγη των γειτονικών συνιστωσών, χρησιμοποιούν κάποιο μέτρο απόστασης μεταξύ Γκαουσιανών, και εκεί έγκειται και η μόνη διαφορά τους. Γιατί καθένα υιοθετεί δικό του κριτήριο για να καθορίσει ποια συνιστώσα του ενός μείγματος είναι εγγύτερα σε κάποια συνιστώσα του άλλου.

Υπενθυμίζουμε εδώ ότι η απόκλιση KL ανάμεσα σε Γκαουσιανές κατανομές μπορεί να υπολογιστεί σε κλειστή μορφή, πράγμα που εξασφαλίζει τον αποδοτικό και σε κλειστή μορφή υπολογισμό των παραπάνω μέτρων. Βέβαια, έχουμε ήδη αναφέρει ότι ένα πρόβλημα που σχετίζεται με τον υπολογισμό αυτών των μέτρων έχει να κάνει με το ότι ο προσδιορισμός της αντιστοίχισης μεταξύ των συνιστωσών των δύο μειγμάτων εμφανίζει τετραγωνική πολυπλοκότητα ως προς το πλήθος των συνιστωσών, και αυτό γιατί, προφανώς, για κάθε συνιστώσα του ενός μείγματος πρέπει να ελεγχθούν ως προς την εγγύτητά τους όλες οι συνιστώσες του άλλου. Παρακάτω, σε αυτό το κεφάλαιο θα περιγράψουμε έναν τρόπο με τον οποίο μπορούμε να παρακάμψουμε αυτό το πρόβλημα.

Εν κατακλείδι, αυτό που πρέπει να μας μείνει από την παραπάνω ανάλυση είναι ότι χρησιμοποιώντας έναν πυρήνα της μορφής (6.1) και μια κατάλληλη απόσταση μεταξύ Γκαουσιανών Μειγμάτων, μπορούμε να χειριστούμε τα Γκαουσιανά Μείγματα στο πλαίσιο των SVMs ακριβώς με τον ίδιο τρόπο που θα χειριζόμασταν και τα κλασικά διανυσματικά πρότυπα. Επομένως, λαμβάνοντας υπόψιν και τη δυνατότητα υλοποίησης της διαδικασίας της συσχετιστικής ανατροφοδότησης μέσω εκπαίδευσης σε κάθε γύρο της τελευταίας ενός ταξινομητή SVM, μπορούμε να αντιληφθούμε ότι είναι πλέον ανοικτός μπροστά μας ο δρόμος της χρήσης SVMs στα πλαίσια της συσχετιστικής ανατροφοδότησης, για την αντιμετώπιση του προβλήματος της ανάκτησης εικόνων όταν αυτές αναπαριστούνται μέσω Γκαουσιανών Μειγμάτων.

## 6.2 Προκαταρκτικά Πειράματα

Σε αυτήν την ενότητα θα παρουσιαστούν μια σειρά από προκαταρκτικά πειράματα πάνω στην ιδέα που αναπτύχθηκε παραπάνω, και η οποία αφορά την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης χρησιμοποιώντας μοντέλα μάθησης, όταν οι εικόνες αναπαριστούνται μέσω Γκαουσιανών Μειγμάτων.

Για αυτά τα πειράματα θα βασιστούμε στα οπτικά χαρακτηριστικά, και στον τρόπο υπολογισμού τους, που περιγράψαμε και χρησιμοποιήσαμε στα πλαίσια του κεφαλαίου 4. Επίσης, θα χρησιμοποιήσουμε αυτούσια τα Γκαουσιανά Μείγματα που προέκυψαν για τις εικόνες με βάση τα παραπάνω χαρακτηριστικά, και τα Ιστογράμματα στα οποία βασίστηκαν εν μέρει οι συγκρίσεις που διεξήχθησαν για την πειραματική τεκμηρίωση της μεθόδου που παρουσιάστηκε στο κεφάλαιο εκείνο. Τέλος, ως βάση εικόνων για τα

πειράματα αυτής της ενότητας χρησιμοποιείται πάλι η MSDB (βλ. και παράρτημα Β, για περισσότερες πληροφορίες).

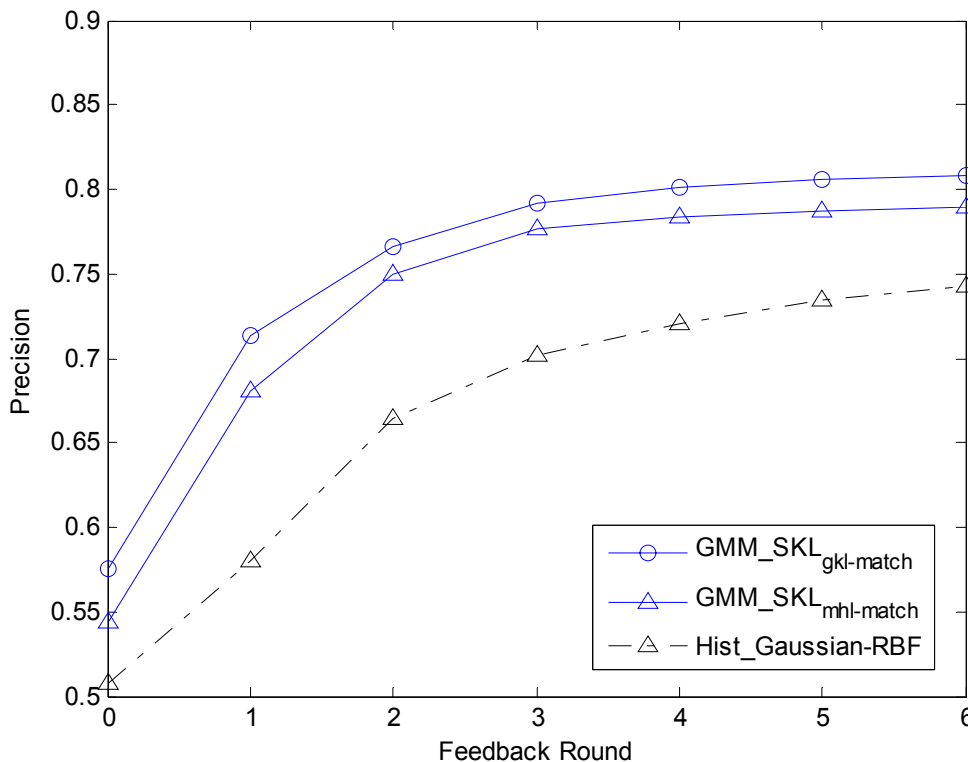
Επομένως, για καθεμιά από τις 3740 εικόνες, χωρισμένες σε 17 σημασιολογικές κατηγορίες, της βάσης MSDB, ένα σύνολο από 8-διάστατα διανύσματα χαρακτηριστικών, που κωδικοποιούν πληροφορία θέσης (συντεταγμένες των pixels), χρώματος (συνιστώσες χρώματος στο χώρο LAB) και υφής (γινόμενο μεταξύ πόλωσης και αντίθεσης, γινόμενο μεταξύ ανισοτροπίας και αντίθεσης, αντίθεση), υπολογίζονται τοπικά και χρησιμοποιούνται κατόπιν σε μια διαδικασία εκτίμησης παραμέτρων ενός Γκαουσιανού Μείγματος, που θα περιγράφει την κατανομή τους και θα αποτελεί την αναπαράσταση της εικόνας. Κάθε Γκαουσιανό Μείγμα έχει 10 συνιστώσες, με καθεμιά από αυτές να χαρακτηρίζεται από έναν πλήρη πίνακα συνδιακύμανσης. Επίσης, για λόγους σύγκρισης, κάθε εικόνα περιγράφεται και από δύο Ιστογράμματα, ένα για το συνδυασμό των χαρακτηριστικών θέσης-χρώματος και ένα δεύτερο για το συνδυασμό των χαρακτηριστικών θέσης-υφής. Όσο αφορά την αναπαράσταση με βάση τα Ιστογράμματα, στα πειράματα που ακολουθούν επιλέχθηκε η κβαντοποίηση Hist1 (όπως αυτή περιγράφεται και στο κεφάλαιο 4). Η συγκεκριμένη κβαντοποίηση οδηγεί σε Ιστογράμματα θέσης-χρώματος με 2304 bins ( $3 \times 3 \times 4 \times 8 \times 8$ ) και σε Ιστογράμματα θέσης-υφής με 576 bins ( $3 \times 3 \times 4 \times 4 \times 4$ ).

Για κάθε εικόνα, οι διανυσματικές αναπαραστάσεις που προκύπτουν για τα παραπάνω Ιστογράμματα συγχωνεύονται, δίνοντας τελικά ένα διάνυσμα 2880 διαστάσεων. Αυτό θα χρησιμοποιηθεί ως τελική διανυσματική αναπαράσταση της εικόνας, κατάλληλη για τη χρήση κλασικών ταξινομητών SVM με συνάρτηση πυρήνα τη Γκαουσιανή συνάρτηση ακτινικής βάσης, που είναι η πιο συχνά χρησιμοποιούμενη στη βιβλιογραφία. Αυτοί οι ταξινομητές, επανεκπαιδευόμενοι σε κάθε εποχή της ανατροφοδότησης με βάση και τα νέα παραδείγματα του χρήστη, θα χρησιμοποιηθούν για λόγους σύγκρισης με την κύρια μεθοδολογία που θα μας απασχολήσει σε αυτό το κεφάλαιο, και η οποία δεν είναι άλλη από τη χρήση για την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης ταξινομητών SVM που μπορούν να χειριστούν αναπαραστάσεις εικόνων που δίνονται μέσω Γκαουσιανών Μειγμάτων.

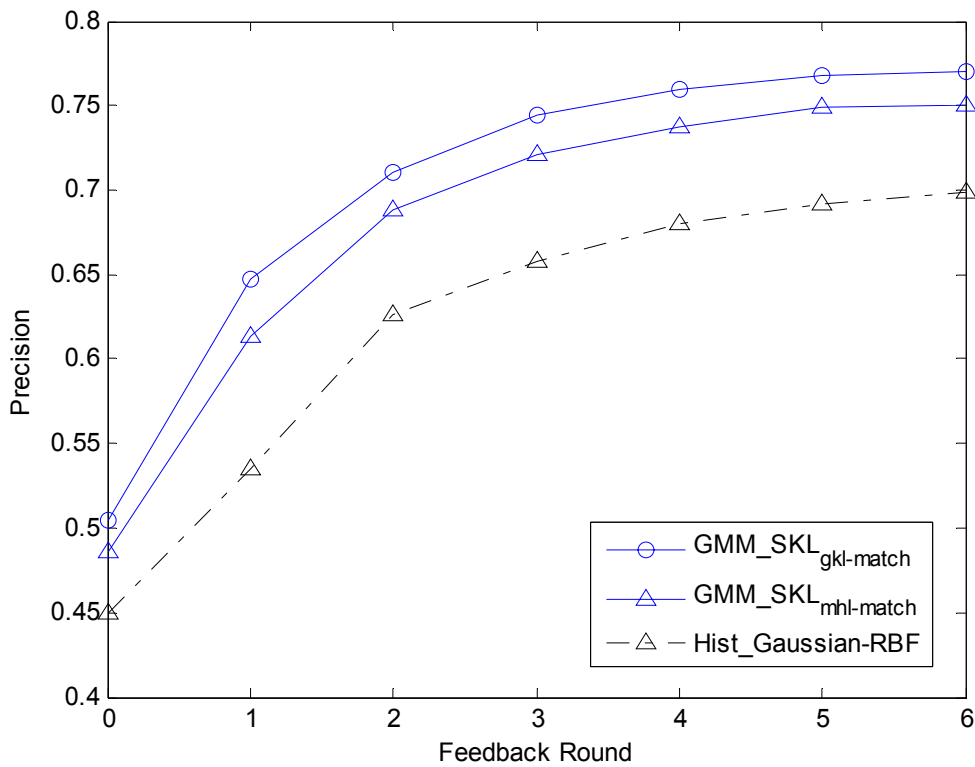
Όπως είπαμε προηγουμένως, για τον ορισμό μιας συνάρτησης πυρήνα μεταξύ Γκαουσιανών Μειγμάτων μέσω της σχέσης (6.1), μια απόσταση μεταξύ Γκαουσιανών Μειγμάτων απαιτείται. Εδώ, επιλέχθηκε να συγκριθούν οι αποστάσεις  $KL_{\text{mhl-match}}$  και  $KL_{\text{gkl-match}}$ , όπως αυτές ορίστηκαν στο κεφάλαιο 3. Υπενθυμίζουμε ότι η πρώτη είναι η προσέγγιση της απόκλισης KL που βασίζεται στο μέτρο ομοιότητας ALA, ενώ η δεύτερη είναι εντελώς παρόμοια με την πρώτη, εκτός του ότι χρησιμοποιεί την απόκλιση KL μεταξύ των Γκαουσιανών συνιστωσών των μειγμάτων για να ορίσει τη συνάρτηση αντιστοίχισης, σε αντίθεση με την πρώτη, που για τον ίδιο σκοπό βασίζεται στην απόσταση Mahalanobis. Υπενθυμίζουμε, επιπλέον, ότι από τον ορισμό της συνάρτησης πυρήνα υπάρχει απαίτηση αυτή να είναι συμμετρική, πράγμα που συνεπάγεται ότι συμμετρική πρέπει να είναι και η χρησιμοποιούμενη στη σχέση (6.1) απόσταση μεταξύ Γκαουσιανών Μειγμάτων. Επομένως, καταλήγοντας, στα πλαίσια αυτής της προκαταρκτικής διερεύνησης των δυνατοτήτων ενσωμάτωσης των Γκαουσιανών Μειγμάτων στη μεθοδολογία των SVMs, ως συναρτήσεις πυρήνα μεταξύ Γκαουσιανών Μειγμάτων θα συγκριθούν παρακάτω αυτές που προκύπτουν από την (6.1) χρησιμοποιώντας τα συμμετρικά μέτρα απόστασης  $SKL_{\text{mhl-match}}$  και  $SKL_{\text{gkl-match}}$ , όπως αυτά ορίζονται από τις σχέσεις (3.30) και (3.31), αντίστοιχα, του κεφαλαίου 3.

Στα πειράματα που ακολουθούν, το σενάριο που χρησιμοποιήθηκε για την προσομοίωση της διαδικασίας της συσχετιστικής ανατροφοδότησης μοιάζει πολύ με το δεύτερο πλάνο προσομοίωσης της ενότητας 4.8.1, και είναι σε γενικές γραμμές το ακόλουθο: Ένα σύνολο αρχικών ερωτήσεων σχηματίζεται μέσω της τυχαίας επιλογής 1000,

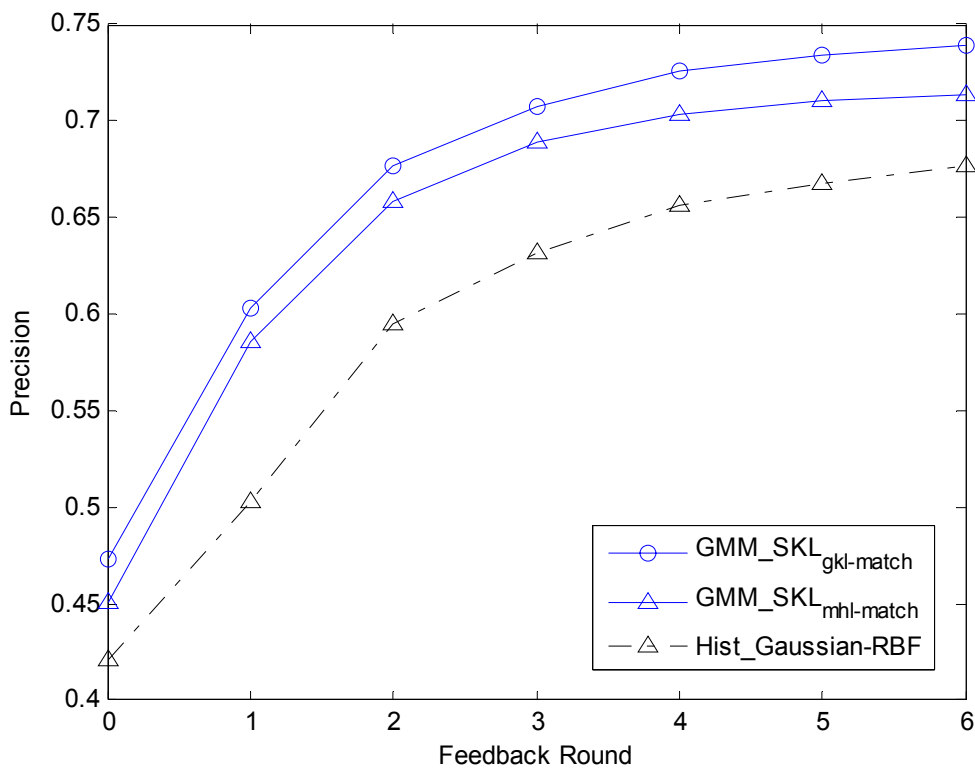
περίπου, εικόνων από τη βάση MSDB. Καθεμιά από αυτές τις αρχικές ερωτήσεις παρουσιάζεται στο σύστημα, ώστε να ξεκινήσει η διαδικασία της ανάκτησης εικόνων. Αρχικά, οι εικόνες της βάσης διατάσσονται ως προς την απόσταση των αναπαραστάσεων τους από αυτήν της αρχικής ερώτησης. Στην περίπτωση της αναπαραστάσεως μέσω Γκαουσιανών Μειγμάτων, μια από τις  $SKL_{mhl-match}$  και  $SKL_{gkl-match}$  χρησιμοποιείται για αυτό το σκοπό. Αντίθετα, στην περίπτωση των διανυσματικών αναπαραστάσεων μέσω Ιστογραμμάτων χρησιμοποιείται το τετράγωνο της Ευκλείδειας απόστασης. Υπενθυμίζουμε ότι σε αυτήν την αρχική φάση, κατά την οποία οι προτιμήσεις του χρήστη προσομοιώνονται μόνο μέσω της αρχικής ερώτησης, δεν είναι δυνατόν να χρησιμοποιηθούν ακόμα ταξινομητές. Από τη διάταξη των εικόνων της βάσης που προκύπτει, τώρα, κατά μέγιστο 3 πραγματικά σχετικές και 3 μη σχετικές εικόνες (όπως αυτό κρίνεται λαμβάνοντας υπόψιν την εκ των προτέρων γνωστή κατηγοριοποίηση της βάσης) επιλέγονται με τυχαίο τρόπο από τις πρώτες 100 θέσεις της διάταξης, και ανατροφοδοτούνται ως θετικά και αρνητικά παραδείγματα, αντίστοιχα. Τα παραδείγματα αυτά μαζί με την αρχική ερώτηση, που λαμβάνεται υπόψιν ως ένα επιπλέον θετικό παράδειγμα, χρησιμοποιούνται ως σύνολο εκπαίδευσης ενός ταξινομητή SVM. Κατόπιν, οι εικόνες της βάσης διατάσσονται σε σχέση με την αντίστοιχη τιμή που υπολογίζεται για τη συνάρτηση απόφασης του ταξινομητή, και ένα καινούργιο σύνολο από 3 θετικά και 3 αρνητικά παραδείγματα, από τις πρώτες 100 θέσεις της νέας διάταξης που προκύπτει, σχηματίζεται. Τα νέα αυτά παραδείγματα χρησιμοποιούνται μαζί με τα προηγούμενα για την εκπαίδευση ενός νέου ταξινομητή SVM και τη συνεπακόλουθη ανανέωση της διάταξης των εικόνων. Αυτή η διαδικασία συνεχίζεται ως ότου συμπληρωθούν 6 γύροι ανατροφοδότησης. Ως μέτρο για την αξιολόγηση των αποτελεσμάτων σε κάθε φάση χρησιμοποιείται η ακρίβεια στις  $T$  πρώτες εικόνες της διάταξης.



**Σχήμα 6.1:** Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συστηματικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια για εμβέλεια  $T = 10$



**Σχήμα 6.2:** Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια για εμβέλεια  $T = 20$



**Σχήμα 6.3:** Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Ιστογραμμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια για εμβέλεια  $T = 30$



Στα σχήματα 6.1, 6.2 και 6.3 περιλαμβάνονται αποτελέσματα για  $T = 10, 20$  και  $30$ , αντίστοιχα. Συγκεκριμένα, για κάθε γύρο της ανάκτησης, ο μέσος όρος των τιμών της ακρίβειας, υπολογισμένος πάνω σε όλες τις αρχικές ερωτήσεις, απεικονίζεται. Με  $GMM\_SKL_{gkl-match}$  συμβολίζουμε τη μέθοδο που βασίζεται στα Γκαουσιανά Μείγματα για την αναπαράσταση των εικόνων, και στην απόσταση  $SKL_{gkl-match}$  για τον ορισμό της συνάρτησης πυρήνα που χρησιμοποιούν τα SVMs σε κάθε γύρο της ανατροφοδότησης. Αντίθετα, με  $GMM\_SKL_{mhl-match}$  αναφερόμαστε στην παραλλαγή που χρησιμοποιεί την απόσταση  $SKL_{mhl-match}$ . Τέλος, με  $Hist\_Gaussian-RBF$  συμβολίζεται η μέθοδος που κάνει χρήση κλασικών διανυσματικών αναπαραστάσεων για τις εικόνες, μέσω των Ιστογραμμάτων που περιγράψαμε παραπάνω, και υιοθετεί τη Γκαουσιανή συνάρτηση ακτινικής βάσης ως συνάρτηση πυρήνα για τα SVMs.

Από τα σχήματα αυτά παρατηρούμε ότι οι μέθοδοι που χρησιμοποιούν τα SVMs σε συνδυασμό με Γκαουσιανά Μείγματα για την αναπαράσταση των εικόνων της βάσης εμφανίζουν σταθερά καλύτερα αποτελέσματα σε σχέση με αυτά που προκύπτουν από τη χρήση κλασικών ταξινομητών SVM. Επιπλέον, η χρήση της απόστασης  $SKL_{gkl-match}$ , που αποτελεί συμμετρική εκδοχή του μέτρου απόστασης  $KL_{gkl-match}$  που προτάθηκε στην εργασία [54], φαίνεται να οδηγεί σε σταθερά υψηλότερα επίπεδα ακρίβειας σε σχέση με την απόσταση  $SKL_{mhl-match}$ , που βασίζεται στο μέτρο ομοιότητας ALA της εργασίας [18].

Αυτά τα αποτελέσματα μας δίνουν μια ένδειξη ότι μέσω της ενσωμάτωσης των Γκαουσιανών Μειγμάτων στο πλαίσιο των SVMs μπορούμε να πετύχουμε σημαντική βελτίωση των αποτελεσμάτων ανάκτησης, σε σχέση με την περίπτωση που γίνεται χρήση κλασικών αναπαραστάσεων για τις εικόνες στο ίδιο πλαίσιο. Μπορούμε πράγματι, λοιπόν, να συνδυάσουμε τα πλεονεκτήματα που προκύπτουν από τη χρήση μοντέλων μάθησης, όπως τα SVMs, για την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης, με αυτά που απορρέουν από τις αναμφισβήτητες περιγραφικές δυνατότητες που διακρίνουν τα πιθανοτικά μοντέλα, όπως τα Γκαουσιανά Μείγματα, για το σχεδιασμό αποτελεσματικών συστημάτων ανάκτησης εικόνων.

Τέλος, πρέπει να αναφέρουμε ότι για τα προηγούμενα πειράματα χρησιμοποιήθηκε η υλοποίηση για τα SVMs που παρέχεται ελεύθερα στο δικτυακό τόπο [67]. Επίσης, ένα άλλο θέμα αφορά τον καθορισμό των τιμών των παραμέτρων που σχετίζονται με τη χρήση των SVMs, και οι οποίες είναι η παράμετρος εκπαίδευσης  $C$  (βλ. κεφάλαιο 5) και η παράμετρος  $\gamma$  της συνάρτησης πυρήνα. Υπενθυμίζουμε, εδώ, ότι η χρησιμοποιούμενη συνάρτηση πυρήνα εξαρτάται από μια παράμετρο  $\gamma$  τόσο κατά την περίπτωση που οι εικόνες αναπαριστούνται μέσω Γκαουσιανών Μειγμάτων (σχέση (6.1)), όσο και κατά την περίπτωση που έχουμε διανυσματική αναπαράσταση μέσω Ιστογραμμάτων (Γκαουσιανή συνάρτηση ακτινικής βάσης, σχέση (5.12)). Σε αυτό το σημείο θα αρκεστούμε να πούμε ότι στα πειράματα που παρουσιάστηκαν σε αυτήν την ενότητα οι τιμές για τις παραπάνω παραμέτρους επιλέχθηκαν εμπειρικά, ώστε να προκύπτουν ανα περίπτωση τα καλύτερα αποτελέσματα. Για αυτές τις παραμέτρους και τον καθορισμό των τιμών τους θα αναφερθούμε αναλυτικότερα παρακάτω στα πλαίσια της παρούσας διατριβής.

### 6.3 Γκαουσιανά Μείγματα και Μπεϋζιανή Μάθηση

Στο κεφάλαιο 2 αναφερθήκαμε αναλυτικά στο πως είναι δυνατόν να εκτιμηθούν οι παράμετροι ενός Γκαουσιανού Μείγματος με βάση ένα σύνολο παρατηρήσεων, χρησιμοποιώντας την αρχή της μέγιστης πιθανοφάνειας. Καταλήξαμε λοιπόν σε έναν επαναληπτικό αλγόριθμο, τον αλγόριθμο EM, ο οποίος μπορεί να χρησιμοποιηθεί για αυτόν

το σκοπό. Σε αυτήν την ενότητα θα αναφερθούμε σε μια ενδιαφέρουσα παραλλαγή του αλγορίθμου EM, που προσεγγίζει το πρόβλημα της εκτίμησης των παραμέτρων από μια άλλη σκοπιά. Η εν λόγω παραλλαγή θα χρησιμοποιηθεί εκτενώς από εδώ και στο εξής για την εκτίμηση των παραμέτρων των μοντέλων των εικόνων, και θα αποδεικτεί, μέσω ενός συνόλου πειραμάτων που θα ακολουθήσουν, ότι συνδέεται με πολλά πλεονεκτήματα τόσο από πλευράς ποιότητας αποτελεσμάτων του συστήματος ανάκτησης όσο και από πλευράς ταχύτητας απόκρισης του τελευταίου.

### 6.3.1 Μπεϋζιανή Μάθηση

Ας θεωρήσουμε αρχικά ότι για μια τυχαία μεταβλητή  $\mathbf{X}$  έχουμε  $N$  παρατηρήσεις,  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ , και ας υποθέσουμε ότι αυτές είναι ανεξάρτητες μεταξύ τους και προέρχονται όλες από ένα πιθανοτικό μοντέλο  $p(\mathbf{x}|\Theta)$ . Με  $\Theta$  συμβολίζουμε τις παραμέτρους του μοντέλου. Για παράδειγμα, όπως έχουμε ήδη δει, στην περίπτωση των Γκαουσιανών Μειγμάτων το  $\Theta$  περιλαμβάνει τα βάρη των συνιστωσών στο μείγμα καθώς και τη μέση τιμή και τον πίνακα συνδιακύμανσης κάθε συνιστώσας.

Η πιθανοφάνεια του παραπάνω συνόλου παρατηρήσεων αντιπροσωπεύει την πιθανότητα  $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | \Theta)$ , και εκφράζει το πόσο πιθανό είναι το γεγονός να προέλθουν όλες αυτές οι παρατηρήσεις από το συγκεκριμένο μοντέλο. Όμως, επειδή έχουμε υποθέσει ότι οι παρατηρήσεις είναι ανεξάρτητες μεταξύ τους, και άρα η τιμή της μιας δεν επηρεάζει την τιμή της άλλης, ισχύει

$$p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | \Theta) = p(\mathbf{x}_1 | \Theta)p(\mathbf{x}_2 | \Theta) \cdots p(\mathbf{x}_N | \Theta) = \prod_{i=1}^N p(\mathbf{x}_i | \Theta) \quad (6.2)$$

Την παραπάνω ποσότητα, και συγκεκριμένα τον λογάριθμό της, επιδιώκει να μεγιστοποιήσει ο αλγόριθμος EM.

Όλη η προηγούμενη ανάλυση υποθέτει ότι δεν έχουμε καμιά απολύτως γνώση για το ποιες μπορεί να είναι οι τιμές των παραμέτρων,  $\Theta$ , του μοντέλου. Προκειμένου να προχωρήσουμε σε μια εκτίμηση για τις τελευταίες, πρέπει να βασιστούμε λοιπόν αποκλειστικά στις παρατηρήσεις  $\mathbf{x}_i$ . Από αυτή τη σκοπιά, λογικά σκεφτόμενοι, επιλέγουμε εκείνες τις τιμές για τις παραμέτρους που ορίζουν το μοντέλο για το οποίο το ενδεχόμενο της παραγωγής του συνόλου των παρατηρήσεων φαντάζει πιο πιθανό. Αυτή είναι η απλή λογική στην οποία βασίζεται η αρχή της μέγιστης πιθανοφάνειας.

Όμως, αν υποθέσουμε ότι με κάποιο τρόπο έχουμε εκ των προτέρων (a priori) μια κάποια γνώση για τις τιμές των παραμέτρων του μοντέλου, τότε τα πράγματα μπορούν να είναι κάπως διαφορετικά. Στο πιθανοτικό πλαίσιο που κινούμαστε, η όποια γνώση σχετικά με τις παραμέτρους του μοντέλου δε μπορεί παρά να είναι πιθανοτική. Έστω λοιπόν ότι γνωρίζουμε εκ των προτέρων, πριν λάβουμε υπόψιν τις παρατηρήσεις  $\mathbf{x}_i$ , πως οι παράμετροι του μοντέλου ακολουθούν μια κατανομή  $p(\Theta)$ . Τότε, οι παρατηρήσεις  $\mathbf{x}_i$  μπορούν να θεωρηθούν ως μια επιπλέον μαρτυρία σχετικά με τις πραγματικές τιμές των παραμέτρων του μοντέλου. Για την εκτίμηση των τιμών των τελευταίων δεν θα ληφθούν, όμως, μόνο αυτές υπόψιν, αλλά θα συνυπολογιστεί και η πρότερη γνώση μας σε σχέση με τις παραμέτρους. Αυτό σε πιθανοτικό πλαίσιο εκφράζεται μέσω της αρχής της μεγιστοποίησης της εκ των υστέρων (maximum a posteriori – MAP) πιθανότητας των παραμέτρων. Με απλά λόγια, αυτή η αρχή μας προτείνει να επιλέξουμε εκείνες τις τιμές για τις παραμέτρους που φαντάζουν πιο πιθανές, με δεδομένες πλέον τις παρατηρήσεις.

Η εκ των υστέρων πιθανότητα των παραμέτρων  $\Theta$ , δοθέντων των παρατηρήσεων  $\mathbf{x}_i$ , συμβολίζεται ως  $p(\Theta | \mathbf{x}_1, \dots, \mathbf{x}_N)$  και μπορεί να υπολογιστεί μέσω του τύπου του Bayes:

$$p(\Theta | \mathbf{x}_1, \dots, \mathbf{x}_N) = \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_N | \Theta)p(\Theta)}{p(\mathbf{x}_1, \dots, \mathbf{x}_N)} \quad (6.3)$$

Η πιθανότητα  $p(\mathbf{x}_1, \dots, \mathbf{x}_N)$  αποτελεί ένα μέτρο του πόσο πιθανές είναι οι παρατηρήσεις  $\mathbf{x}_i$  ανεξαρτήτως μοντέλου, και επομένως είναι ανεξάρτητη του  $\Theta$ . Μπορούμε λοιπόν να την αγνοήσουμε κατά τη διαδικασία της μεγιστοποίησης της εκ των υστέρων πιθανότητας των παραμέτρων. Τελικά, με βάση την παραπάνω ανάλυση, οι τιμές των παραμέτρων που προβλέπονται από την αρχή της μεγιστοποίησης της εκ των υστέρων πιθανότητας, έστω  $\Theta_{\text{MAP}}$ , δίνονται από την παρακάτω σχέση:

$$\Theta_{\text{MAP}} = \arg \max_{\Theta} \left[ p(\Theta) \prod_{i=1}^N p(\mathbf{x}_i | \Theta) \right] \quad (6.4)$$

Όπως φαίνεται από τη σχέση αυτή, έχουμε προχωρήσει ξανά στην υπόθεση περί ανεξαρτησίας των παρατηρήσεων.

Αυτός ο τρόπος εκτίμησης των παραμέτρων ενός μοντέλου με βάση την αρχή της μεγιστοποίησης της εκ των υστέρων πιθανότητας αναφέρεται και ως Μπεϋζιανή μάθηση.

### 6.3.2 Εκπαίδευση Γκαουσιανών Μειγμάτων Μέσω Μπεϋζιανής Μάθησης

Σε αυτήν την ενότητα θα μας απασχολήσει η εφαρμογή της αρχής της μεγιστοποίησης της εκ των υστέρων πιθανότητας, στην περίπτωση που ως μοντέλα των οποίων θέλουμε να εκτιμήσουμε τις παραμέτρους χρησιμοποιούνται τα Γκαουσιανά Μείγματα.

#### 6.3.2.1 Γενικό Γκαουσιανό Μείγμα

Αρχικά, θα εξετάσουμε λίγο σε τι μπορεί να αντιστοιχεί πρακτικά η ύπαρξη εκ των προτέρων γνώσης σχετικά με τις τιμές των παραμέτρων ενός Γκαουσιανού Μείγματος. Ας εστιάσουμε επομένως στο πρόβλημα των εικόνων, που είναι και αυτό που θα μας απασχολήσει στην πράξη.

Έχουμε πει ότι κάθε εικόνα της βάσης περιγράφεται μέσω ενός συνόλου τοπικά υπολογιζόμενων διανυσμάτων χαρακτηριστικών. Αυτά τα διανύσματα χαρακτηριστικών τα χρησιμοποιούμε, μετέπειτα, ώστε να εκτιμήσουμε τις παραμέτρους ενός μοντέλου Γκαουσιανού Μείγματος, που θα περιγράψει την κατανομή τους στο χώρο των χαρακτηριστικών και θα χρησιμεύει ως αναπαράσταση για την εικόνα. Μια βάση εικόνων περιέχει συνήθως πάρα πολλές εικόνες. Ακόμα λοιπόν και αν θεωρήσουμε ότι σε κάθε εικόνα αντιστοιχούν σχετικά λίγα διανύσματα χαρακτηριστικών, πράγμα που όπως θα δούμε παρακάτω δεν είναι σπάνιο, για το σύνολο των εικόνων της βάσης υπάρχει ένας μεγάλος αριθμός διανυσμάτων χαρακτηριστικών.

Στην παραπάνω διαπίστωση κρύβεται και ο κύριος πρακτικός λόγος που συνηγορεί υπέρ της χρήσης μιας λογικής Μπεϋζιανής μάθησης για την εκπαίδευση του μοντέλου της κάθε εικόνας. Συγκεκριμένα, ας θεωρήσουμε ότι έχουμε σχετικά λίγα διανύσματα

χαρακτηριστικών για κάθε εικόνα, συνυπολογίζοντας για παράδειγμα την πολυπλοκότητα του μοντέλου που επιθυμούμε να χρησιμοποιήσουμε, όπως αυτή εκφράζεται από τον αριθμό των Γκαουσιανών συνιστωσών στο μείγμα, και τη διάσταση των διανυσμάτων χαρακτηριστικών. Η τελευταία ενδέχεται να είναι μεγάλη, ειδικά στις περιπτώσεις για τις οποίες εννοείται ο υπολογισμός λίγων διανυσμάτων χαρακτηριστικών ανά εικόνα. Από τα παραπάνω, είναι εύκολο να αντιληφθούμε ότι είναι πολύ δύσκολη η με εύρωστο τρόπο εκτίμηση των παραμέτρων του μοντέλου, αν βασιστούμε μόνο στα λίγα χαρακτηριστικά που υπολογίζονται για την κάθε εικόνα. Αυτό ακριβώς κάνει ο κλασικός αλγόριθμος EM και η λογική της μεγιστοποίησης της πιθανοφάνειας στην οποία στηρίζεται. Χρησιμοποιεί ως μόνη πληροφορία για την εκτίμηση των παραμέτρων του μοντέλου της κάθε εικόνας τα διανύσματα χαρακτηριστικών που της αντιστοιχούν. Συνάγεται λοιπόν από τα παραπάνω ότι η χρήση του κλασικού αλγορίθμου EM σε περιπτώσεις όπως αυτή που αναλύσαμε προηγουμένως ενδέχεται να οδηγήσει σε εκτιμήσεις για τις παραμέτρους η ποιότητα των οποίων δε θα είναι ικανοποιητική, με άμεσο αντίκτυπο στα αποτελέσματα των συστημάτων που θα χρησιμοποιήσουν κατόπιν αυτά τα μοντέλα.

Επομένως, απαιτείται ένας άλλος πιο εύρωστος τρόπος υπολογισμού των τιμών των παραμέτρων των μοντέλων. Και εδώ έρχεται να μας βοηθήσει η διαπίστωση της πληθώρας διανυσμάτων χαρακτηριστικών που υπάρχουν στη βάση. Είναι πράγματι τόσο πολλά που συνήθως δεν μπορούμε αλλά και δε χρειάζεται να τα χειριστούμε όλα. Αρκεί να επιλέξουμε με τυχαίο τρόπο, για παράδειγμα, ένα μικρό σχετικά υποσύνολο όλων αυτών των διανυσμάτων, ώστε να καταλήξουμε σε ένα μεγάλο κατά απόλυτη τιμή αριθμό διανυσμάτων χαρακτηριστικών. Αυτά, όπως καταλαβαίνει κανείς, θα αποτελούν έκφραση της κατανομής των χαρακτηριστικών για όλη τη βάση, και όχι πλέον για μια μεμονωμένη εικόνα. Επιπλέον, ο αριθμός τους επιτρέπει συνήθως έναν εύρωστο υπολογισμό των παραμέτρων, για ένα μοντέλο που αυτή τη φορά θα αφορά τη μοντελοποίηση της κατανομής των διανυσμάτων χαρακτηριστικών για όλη τη βάση εικόνων. Η εκτίμηση των παραμέτρων αυτού του μοντέλου μπορεί να γίνει με βάση την αρχή την μέγιστης πιθανοφάνειας, δηλαδή χρησιμοποιώντας τον κλασικό αλγόριθμο EM.

Έστω λοιπόν ότι μέσω της κλασικής διαδικασίας εκτίμησης παραμέτρων για ένα Γκαουσιανό Μείγμα καταλήγουμε σε ένα γενικό (universal) Γκαουσιανό Μείγμα, με παραμέτρους  $\pi_j^{pr}$ ,  $\mu_j^{pr}$  και  $\Sigma_j^{pr}$  για τα βάρη, τις μέσες τιμές και του πίνακες συνδιακύμανσης των συνιστωσών. Επίσης, έστω ότι  $K$  είναι ο αριθμός των συνιστωσών στο εν λόγω μείγμα. Οι παραπάνω πληροφορίες, σχετικά με το γενικό μοντέλο για τα διανύσματα χαρακτηριστικών της βάσης εικόνων, συνιστούν και την εκ των προτέρων γνώση μας όσο αφορά τις τιμές των παραμέτρων για το μοντέλο της κάθε εικόνας της βάσης.

### 6.3.2.2 Αλγόριθμος MAP-EM

Στην εργασία [78] ορίζεται ένα κατάλληλο πιθανοτικό πλαίσιο σχετικά με την εκτίμηση των παραμέτρων ενός Γκαουσιανού Μείγματος μέσω Μπεϋζιανής μάθησης. Το πλαίσιο αυτό έγκειται κυρίως στην επιλογή κατάλληλου πιθανοτικού μοντέλου,  $p(\Theta)$ , για την εκ των προτέρων κατανομή που αντιστοιχεί στις προς εκτίμηση παραμέτρους. Γίνεται εύκολα αντιληπτό ότι και αυτό το μοντέλο με τη σειρά του πρέπει να είναι παραμετρικό, δηλαδή θα εξαρτάται από κάποιες υπερπαραμέτρους (hyperparameters). Αυτό είναι απαραίτητο για να μπορούν να ληφθούν υπόψιν οι συγκεκριμένες τιμές που έχουν προκύψει για τις παραμέτρους του γενικού Γκαουσιανού Μείγματος. Εν τέλει λοιπόν, οι υπερπαραμέτροι καθορίζονται με βάση τις τιμές που έχουμε για τα  $\pi_j^{pr}$ ,  $\mu_j^{pr}$  και  $\Sigma_j^{pr}$  του γενικού

Γκαουσιανού Μείγματος, και καταλήγουμε σε μια συγκεκριμένη  $p(\Theta)$ , που αντικατοπτρίζει πραγματικά την εκ των προτέρων γνώση μας για τις τιμές των παραμέτρων προς εκτίμηση. Μετά από τον ορισμό της  $p(\Theta)$  επιχειρείται η επίλυση του προβλήματος βελτιστοποίησης της σχέσης (6.4). Για μια αναλυτική περιγραφή των επιμέρους βημάτων που ακολουθούνται, μπορεί κανείς να ανατρέξει στην εργασία [78].

Λαμβάνοντας υπόψιν και τα αποτελέσματα της παραπάνω εργασίας, στην εργασία [77] ορίζεται μια επαναληπτική διαδικασία προσαρμογής των παραμέτρων ενός γενικού Γκαουσιανού Μείγματος με βάση ένα σύνολο παρατηρήσεων. Ο αλγόριθμος που ακολουθείται μοιάζει πολύ με τον κλασικό αλγόριθμο EM, ενώ παράλληλα χρησιμοποιεί μια λογική μεγιστοποίησης της εκ των υστέρων πιθανότητας των παραμέτρων. Για αυτούς τους λόγους μπορεί να ονομαστεί MAP-EM.

Συγκεκριμένα, θεωρώντας τις παραμέτρους  $\pi_j^{\text{pr}}$ ,  $\mu_j^{\text{pr}}$  και  $\Sigma_j^{\text{pr}}$  του γενικού μοντέλου και ένα σύνολο παρατηρήσεων  $\mathbf{x}_i$ , που στην περίπτωσή μας θα είναι τα διανύσματα χαρακτηριστικών που υπολογίζονται από μια συγκεκριμένη εικόνα, θέλουμε να εκτιμήσουμε τις παραμέτρους  $\pi_j$ ,  $\mu_j$  και  $\Sigma_j$  ενός νέου μοντέλου αποκλειστικά για τις  $\mathbf{x}_i$ . Η διαδικασία που θα ακολουθήσουμε περιλαμβάνει και αυτή δύο βήματα, ένα βήμα E και ένα βήμα M. Το βήμα E είναι ακριβώς ίδιο με αυτό του κλασικού αλγορίθμου EM. Δηλαδή, για κάθε παρατήρηση  $\mathbf{x}_i$  και για κάθε συνιστώσα  $j$  υπολογίζεται η προσδοκώμενη τιμή  $E(z_{ij})$  για την αντίστοιχη κρυφή μεταβλητή, μέσω της σχέσεως (2.39). Υπενθυμίζουμε εν τάχει ότι η τιμή της κρυφής μεταβλητής  $z_{ij}$  καθορίζει το αν η  $i$ -οστή παρατήρηση έχει προέλθει από την  $j$ -οστή συνιστώσα του μείγματος, και ότι η αντίστοιχη προσδοκώμενη τιμή ισούται με  $P(j | \mathbf{x}_i)$ , δηλαδή με την πιθανότητα που συγκεντρώνει το ενδεχόμενο η  $j$ -οστή συνιστώσα να παρήγαγε την  $i$ -οστή παρατήρηση.

Το βήμα M του αλγορίθμου MAP-EM, τώρα, είναι λίγο διαφορετικό σε σχέση με το αντίστοιχο του κλασικού αλγορίθμου EM. Συγκεκριμένα, για την ανανέωση των εκτιμήσεων σχετικά με τις παραμέτρους  $\pi_j$ ,  $\mu_j$  και  $\Sigma_j$  του μοντέλου ισχύουν οι παρακάτω σχέσεις:

$$w_j = a_j \frac{\sum_{i=1}^N E[z_{ij}]}{N} + (1 - a_j) \pi_j^{\text{pr}} \quad (6.5)$$

$$\pi_j = \frac{w_j}{\sum_{k=1}^K w_k} \quad (6.6)$$

$$\mu_j = a_j \frac{\sum_{i=1}^N E[z_{ij}] \mathbf{x}_i}{\sum_{i=1}^N E[z_{ij}]} + (1 - a_j) \mu_j^{\text{pr}} \quad (6.7)$$

$$\Sigma_j = a_j \frac{\sum_{i=1}^N E[z_{ij}] \mathbf{x}_i \mathbf{x}_i^T}{\sum_{i=1}^N E[z_{ij}]} + (1 - a_j) [\Sigma_j^{\text{pr}} + (\mu_j^{\text{pr}})(\mu_j^{\text{pr}})^T] - \mu_j \mu_j^T \quad (6.8)$$

όπου

$$a_j = \frac{\sum_{i=1}^N E[z_{ij}]}{\sum_{i=1}^N E[z_{ij}] + \tau} \quad (6.9)$$

Η σχέση (6.6) χρησιμοποιείται για λόγους κανονικοποίησης, ώστε να εξασφαλιστεί ότι τα βάρη όλων των συνιστωσών του μείγματος αθροίζουν στη μονάδα.

Από τις παραπάνω σχέσεις, και αντιπαραβάλλοντάς τες με τις αντίστοιχες του κλασικού αλγορίθμου EM (σχέσεις (2.40)-(2.42)), παρατηρούμε αμέσως για τον υπολογισμό των νέων εκτιμήσεων ως προς τις τιμές των παραμέτρων του μοντέλου, ότι αυτές αντιστοιχούν σε ένα συνδυασμό των εκτιμήσεων που προβλέπονται από τον κλασικό αλγόριθμο EM, οι οποίες βασίζονται αποκλειστικά στις παρατηρήσεις  $\mathbf{x}_i$ , και της πληροφορίας που προέρχεται από τη γνώση των παραμέτρων του γενικού μοντέλου. Το σχετικό βάρος με το οποίο η πληροφορία η προερχόμενη από τις παρατηρήσεις  $\mathbf{x}_i$  λαμβάνεται υπόψη στην ανανέωση των εκτιμήσεων για τις παραμέτρους κάθε συνιστώσας του μοντέλου δίνεται από τον αντίστοιχο συντελεστή  $a_j$ .

Από τον ορισμό του παραπάνω συντελεστή βλέπουμε ότι η τιμή του καθορίζεται από την παράμετρο  $\tau$ , η οποία είναι μια παράμετρος που εκφράζει το βαθμό στον οποίο βασιζόμαστε στις τιμές των παραμέτρων του γενικού μοντέλου για τον καθορισμό των παραμέτρων του μοντέλου των παρατηρήσεων  $\mathbf{x}_i$ , και από την ποσότητα  $\sum_i E[z_{ij}]$ , για την οποίαν ισχύει

$$\sum_{i=1}^N E[z_{ij}] = \sum_{i=1}^N P(j | \mathbf{x}_i) \quad (6.10)$$

λόγω του ότι, όπως αποδείχτηκε στο κεφάλαιο 2,  $E(z_{ij}) = P(j | \mathbf{x}_i)$ . Επομένως, η ποσότητα  $\sum_i E[z_{ij}]$  εκφράζει εν τέλει το πόσες παρατηρήσεις αντιστοιχούν, σε ένα πιθανοτικό πλαίσιο, στην  $j$ -οστή συνιστώσα του μοντέλου. Συνεπώς, παρατηρούμε ότι όσο περισσότερες είναι αυτές οι παρατηρήσεις (δηλαδή όσο μεγαλύτερη είναι η τιμή  $\sum_i E[z_{ij}]$ ) τόσο μεγαλύτερο βάρος δίνεται στη νέα μαρτυρία που προέρχεται από αυτές για τον καθορισμό των νέων τιμών για τις παραμέτρους της αντίστοιχης συνιστώσας. Αντίθετα, εάν στην εν λόγω συνιστώσα αντιστοιχούν πολύ λίγες παρατηρήσεις, σχετικά και με την τιμή της παραμέτρου  $\tau$ , για τον υπολογισμό της νέας εκτίμησης όσο αφορά τις τιμές των παραμέτρων της βασιζόμαστε κυρίως στις ήδη γνωστές τιμές για τις παραμέτρους του γενικού μοντέλου. Αυτή η συμπεριφορά του αλγορίθμου μας εξασφαλίζει την ευρωστία, η οποία είναι εξάλλου και ο λόγος για τον οποίον καταφύγαμε σε ένα νέο τρόπο εκτίμησης για τις παραμέτρους του μοντέλου. Αυτό που θέλαμε να αποφύγουμε εξαρχής είναι ο υπολογισμός των παραμέτρων των συνιστωσών με χρήση ελάχιστων παρατηρήσεων.

Συνοψίζοντας, επομένως, έχουμε να επισημάνουμε ότι ο αλγόριθμος MAP-EM συνίσταται στη διαδοχική επανάληψη των βημάτων E και M που περιγράφηκαν παραπάνω, μέχρι να επέλθει σύγκλιση, η οποία εξακριβώνεται με τον ίδιο ακριβώς τρόπο όπως και στην περίπτωση του κλασικού αλγορίθμου EM. Σε κάθε επανάληψη του αλγορίθμου, και συγκεκριμένα στο βήμα M, η εκτίμηση για τις τιμές των παραμέτρων του μοντέλου ανανεώνεται. Στο βήμα E, η τρέχουσα εκτίμηση ως προς τις τιμές αυτές χρησιμοποιείται

για τον υπολογισμό των προσδοκώμενων τιμών  $E(z_{ij})$ . Ακόμα, το πρόβλημα του κλασικού αλγορίθμου EM, που αφορούσε τον προσδιορισμό μιας αρχικής εκτίμησης για τις τιμές των παραμέτρων, δεν υφίσταται πλέον, γιατί, για την εκκίνηση των επαναλήψεων του αλγορίθμου MAP-EM, ως αρχική εκτίμηση χρησιμοποιούνται πάντα οι τιμές των παραμέτρων του γενικού μοντέλου. Από όλη την παραπάνω ανάλυση συνάγεται, ακόμα, ότι ο αριθμός των συνιστωσών κάθε μοντέλου που προσαρμόζεται από ένα γενικό μοντέλο με βάση ένα σύνολο παρατηρήσεων, ισούται πάντα με τον αριθμό των συνιστωσών του γενικού μοντέλου.

### 6.3.3 Απόσταση Μεταξύ Γκαουσιανών Μειγμάτων Εκπαιδευμένων Μέσω του Αλγορίθμου MAP-EM

Όπως έχουμε αναφέρει πολλές φορές στα πλαίσια της παρούσας διατριβής, το κύριο ζήτημα προς αντιμετώπιση όσο αφορά τη χρήση των Γκαουσιανών Μειγμάτων στα πλαίσια του προβλήματος της ανάκτησης εικόνων έχει να κάνει με τον ορισμό αποδοτικά υπολογιζόμενων μέτρων απόστασης. Ας εστιάσουμε λοιπόν, πάλι, στο πρόβλημα των εικόνων, και ας θεωρήσουμε ότι έχουμε εκπαιδεύσει ένα γενικό Γκαουσιανό Μείγμα χρησιμοποιώντας πληροφορία από όλη τη βάση εικόνων. Επίσης, ας θεωρήσουμε ότι για κάθε εικόνα της βάσης έχουμε εκπαιδεύσει ένα Γκαουσιανό Μείγμα, μέσω προσαρμογής των παραμέτρων του γενικού μείγματος με χρήση του αλγορίθμου MAP-EM, ώστε αυτό να περιγράφει στα πλαίσια της Μπεϋζιανής μάθησης το σύνολο των διανυσμάτων χαρακτηριστικών της εικόνας.

Όπως έχει διαφανεί από τα παραπάνω, κάθε συνιστώσα του μείγματος μιας εικόνας θα έχει προέλθει από την προσαρμογή της αντίστοιχης συνιστώσας του γενικού μείγματος. Αυτό ισχύει για κάθε εικόνα. Δηλαδή, συγκεκριμένα, για κάθε  $i = 1, \dots, K$ , οι  $i$ -οστές συνιστώσες των μοντέλων των εικόνων έχουν προέλθει όλες από την  $i$ -οστή συνιστώσα του γενικού μοντέλου. Επομένως, είναι φυσικό να θεωρήσουμε ότι είναι με κάποιον τρόπο κοντινές μεταξύ τους. Αυτό ορίζει μια φυσική αντιστοιχία μεταξύ των συνιστωσών των μοντέλων των εικόνων.

Από την άλλη, με βάση όσα έχουμε πει στο κεφάλαιο 3 και με μια απλή επισκόπηση των σχέσεων (3.27), (3.14) και (3.28), (3.29), αντιλαμβανόμαστε αμέσως ότι οι προσεγγίσεις της απόκλισης KL μεταξύ Γκαουσιανών Μειγμάτων  $KL_{\text{mhl-match}}$  και  $KL_{\text{gkl-match}}$ , που έχουν προταθεί στις εργασίες [18] και [54], αντίστοιχα, υπολογίζονται μέσω του ίδιου γενικού τύπου:

$$KL_{\text{apprx}}(p_1 \parallel p_2) = \sum_{i=1}^{K_1} \pi_{1i} \left[ KL(\phi(\mathbf{x} \mid \theta_{1i}) \parallel \phi(\mathbf{x} \mid \theta_{2_{m_{12}(i)}})) + \ln \frac{\pi_{1i}}{\pi_{2_{m_{12}(i)}}} \right] \quad (6.11)$$

και διαφέρουν μόνο στον τρόπο με τον οποίον ορίζουν την αντιστοίχιση,  $m_{12}$ , μεταξύ των συνιστωσών των συγκρινόμενων μειγμάτων  $p_1$  και  $p_2$ . Έχουμε αναφέρει επίσης ότι ακριβώς λόγω της ανάγκης υπολογισμού αυτής της συνάρτησης αντιστοίχισης, η πολυπλοκότητα τους είναι τετραγωνική σε σχέση με τον αριθμό των συνιστωσών των δύο μειγμάτων.

Τώρα, αν βασιστούμε στη σχέση (6.11) και εκμεταλλευτούμε τη φυσική αντιστοιχία που δημιουργείται μεταξύ των συνιστωσών στο πλαίσιο της εκπαίδευσης των μοντέλων μέσω του αλγορίθμου MAP-EM, μπορούμε να ορίσουμε μια νέα προσέγγιση της απόκλισης KL μέσω της παρακάτω σχέσης:

$$\text{KL}_{\text{id-match}}(p_1 \parallel p_2) = \sum_{i=1}^K \pi_{1i} \left[ \text{KL}(\phi(\mathbf{x} | \boldsymbol{\theta}_{1i}) \parallel \phi(\mathbf{x} | \boldsymbol{\theta}_{2i})) + \ln \frac{\pi_{1i}}{\pi_{2i}} \right] \quad (6.12)$$

Η συγκεκριμένη απόσταση παρουσιάστηκε και μελετήθηκε με ενθαρρυντικά αποτελέσματα στην εργασία [70].

Η υπόθεση που γίνεται μέσω της σχέσης (6.12), σε αντιπαράβολή και με την (6.11), αφορά προφανώς τη συνάρτηση αντιστοίχισης, για την οποία θεωρούμε ότι ισχύει

$$m_{12}(i) = i, \quad \forall i = 1, \dots, K \quad (6.13)$$

Επίσης, εδώ, ο αριθμός των συνιστωσών και για τα δύο μείγματα είναι κοινός και ίσος με τον αριθμό των συνιστωσών του γενικού μείγματος,  $K$ . Η παραπάνω υπόθεση, με δεδομένη τη χρήση του αλγορίθμου MAP-EM, δεν είναι καθόλου αυθαίρετη, αφού, πραγματικά, στις περισσότερες των περιπτώσεων, ακόμα και αν χρησιμοποιήσουμε τις συναρτήσεις αντιστοίχισης που προτείνονται στις εργασίες [18] και [54], είτε θα ισχύει πράγματι  $m_{12}(i) = i$ , είτε, ακόμα και αν ισχύει  $m_{12}(i) \neq i$ , η υπόθεση  $m_{12}(i) = i$  δε θα οδηγήσει σε πολύ διαφορετική τελική τιμή για το μέτρο απόστασης. Αυτό θα αποδεικτεί και μέσω των πειραμάτων που θα παρουσιαστούν παρακάτω σε αυτό το κεφάλαιο.

Όμως, το πιο ενδιαφέρον σχετικά με την απόσταση της σχέσης (6.12) είναι ότι η πολυπλοκότητα του υπολογισμού της είναι γραμμική όσο αφορά τον αριθμό των συνιστωσών των μοντέλων. Επομένως, μπορεί να υπολογιστεί πολύ ταχύτερα από ό,τι οι αρχικές προσεγγίσεις  $\text{KL}_{\text{mhl-match}}$  και  $\text{KL}_{\text{gkl-match}}$ . Βλέπουμε, λοιπόν, άλλο ένα πλεονέκτημα που έχουμε λόγω της εκτίμησης των παραμέτρων των μοντέλων μέσω του αλγορίθμου MAP-EM και των αρχών της Μπεϋζιανής μάθησης.

## 6.4 Οπτικά Χαρακτηριστικά

Σε προηγούμενη ενότητα αφήσαμε να εννοηθεί ότι τα πλεονεκτήματα από τη χρήση της αρχής της μεγιστοποίησης της εκ των υστέρων πιθανότητας για την εκτίμηση των παραμέτρων των μοντέλων των εικόνων πολλαπλασιάζονται όσο μικραίνει ο αριθμός των ανά εικόνα διαθέσιμων διανυσμάτων χαρακτηριστικών, και αυξάνει παράλληλα η διάσταση τους. Καλώς ή κακώς, πολλές επιτυχημένες μέθοδοι υπολογισμού χαρακτηριστικών καταλήγουν σε περιγραφές για τις εικόνες που περιλαμβάνουν σχετικά λίγα πολυδιάστατα διανύσματα χαρακτηριστικών. Μια τέτοια μέθοδο υπολογισμού χαρακτηριστικών επιλέξαμε να χρησιμοποιήσουμε για την περιγραφή των εικόνων στα πειράματα που ακολουθούν. Στην τρέχουσα ενότητα γίνεται μια συνοπτική παρουσίαση της μεθόδου αυτής και των χαρακτηριστικών που υπολογίζονται τελικά από τις εικόνες.

### 6.4.1 Μέθοδος Υπολογισμού Χαρακτηριστικών SIFT

Η περισσότερο δημοφιλής και καταξιωμένη από άποψη αποτελεσμάτων τεχνική τοπικού υπολογισμού χαρακτηριστικών από εικόνες είναι, όπως καταδεικνύεται για παράδειγμα από την εργασία [74], η SIFT (Scale-Invariant Feature Transform – Μετασχηματισμός Χαρακτηριστικών Αμετάβλητος ως προς την Κλίμακα). Μια λεπτομερής περιγραφή αυτής της τεχνικής δίνεται στην εργασία [40]. Πρέπει να αναφερθεί μάλιστα ότι πολλές παραλλαγές αυτής της μεθόδου έχουν προταθεί, ακριβώς λόγω της επιτυχίας που έχει γνωρίσει και του συνεπαγόμενου ενδιαφέροντος που έχει κερδίσει.



Αναφέρουμε ενδεικτικά τις εργασίες [41], [72] και [73], που κινούνται σε αυτό το πλαίσιο. Η συγκεκριμένη τεχνική χειρίζεται γκριζες εικόνες, δηλαδή εικόνες που περιέχουν μόνο πληροφορία φωτεινότητας και όχι πληροφορία χρώματος. Σε τέτοιες εικόνες, σε κάθε pixel αντιστοιχεί ένας αριθμός που εκφράζει τη φωτεινότητα της εικόνας στο συγκεκριμένο pixel.

Ο υπολογισμός των χαρακτηριστικών γίνεται σε δύο φάσεις. Πρώτα εντοπίζονται σημεία-κλειδιά (keypoints) στο επίπεδο της εικόνας, τα οποία είναι αντιπροσωπευτικά των απεικονιζόμενων αντικειμένων και εμφανίζουν μεγάλη επαναληψιμότητα για μεταβολές όσο αφορά τη θέση και τη γωνία παρατήρησης, το φωτισμό της σκηνής, την εισαγωγή θορύβου κτλ. Για αυτά τα σημεία καθορίζεται η κλίμακα και ο προσανατολισμός που τους αντιστοιχεί, ώστε αυτή η πληροφορία να μπορεί να χρησιμοποιηθεί για τον υπολογισμό χαρακτηριστικών που δεν εξαρτώνται από τις δύο αυτές παραμέτρους, και άρα δε μεταβάλλονται κατά την περιστροφή ενός αντικειμένου ή την απεικόνισή του σε άλλη κλίμακα μεγέθους. Σε δεύτερη φάση, ένα διάνυσμα χαρακτηριστικών υπολογίζεται για κάθε σημείο-κλειδί που εντοπίστηκε. Αυτό το διάνυσμα κωδικοποιεί την τοπική κατανομή των κατευθύνσεων των ακμών γύρω από το εν λόγω σημείο.

Ας θεωρήσουμε μια γκρίζα εικόνα που περιγράφεται μέσω της δισδιάστατης συνάρτησης φωτεινότητας  $I(x, y)$ , με τις δύο διαστάσεις να αντιστοιχούν στο πλάτος και στο ύψος της εικόνας. Η μέθοδος SIFT χρησιμοποιεί ισοτροπικές Γκαουσιανές μεταβλητού πλάτους:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (6.14)$$

Μέσω συνέλιξης με τους παραπάνω Γκαουσιανούς πυρήνες, η εικόνα φιλτράρεται σε πολλαπλές κλίμακες  $\sigma$ , και προκύπτουν οι ομαλοποιημένες εικόνες

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) = \iint I(x - a, y - b) G(a, b, \sigma) da db \quad (6.15)$$

για τις διάφορες τιμές του  $\sigma$ . Η παραπάνω σχέση ορίζει τη συνέλιξη για συνεχείς συναρτήσεις. Στον υπολογιστή, βέβαια, η υλοποίηση της πράξης της συνέλιξης γίνεται με διακριτό τρόπο, όπου πλέον δεν έχουμε ολοκληρώματα αλλά αθροίσματα. Αλλά, για αυτά τα θέματα έχουμε μιλήσει και στο κεφάλαιο 4, και συγκεκριμένα στην ενότητα 4.6.2.

Ο εντοπισμός των σημείων-κλειδιών και ο προσδιορισμός της κλίμακας που τους αντιστοιχεί γίνεται μέσω του προσδιορισμού εκείνων των  $x, y, \sigma$  που αποτελούν τοπικά μέγιστα ή ελάχιστα της συνάρτησης  $D(x, y, \sigma)$ , που ορίζεται από τις διαφορές μεταξύ των ομαλοποιημένων εικόνων σε διαδοχικές κλίμακες. Για τη συνάρτηση αυτή, λοιπόν, ισχύει

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (6.16)$$

όπου το  $k$  παίζει το ρόλο βήματος από μια κλίμακα  $\sigma$  στη διαδοχική της,  $k\sigma$ . Στην πράξη και οι τρεις συντεταγμένες  $x, y, \sigma$  είναι διακριτές, οπότε ο εντοπισμός των μεγίστων και ελαχίστων ανάγεται στην εύρεση εκείνων των τριάδων  $(x, y, \sigma)$  που χαρακτηρίζονται από τιμή για τη  $D(x, y, \sigma)$  μεγαλύτερη ή μικρότερη σε σχέση με αυτήν όλων των γειτονικών τους στο πλέγμα που δημιουργείται. Έτσι, μαζί με τη θέση  $(x, y)$  στο επίπεδο της εικόνας, καθορίζεται και η κλίμακα  $\sigma$  που αντιστοιχεί στο κάθε σημείο-κλειδί.

Έπειτα, για κάθε ένα από τα σημεία-κλειδιά, έστω στη θέση  $(x, y)$  και με τιμή  $\sigma$  για την παράμετρο κλίμακας, ακολουθεί ο προσδιορισμός του κύριου προσανατολισμού που του

αντιστοιχεί. Για αυτό το σκοπό χρησιμοποιείται η αντίστοιχη ομαλοποιημένη εικόνα  $L(x, y, \sigma)$ , και υπολογίζεται η κλίση

$$\nabla L = \left( \frac{\partial L}{\partial x}, \frac{\partial L}{\partial y} \right)^T \quad (6.17)$$

γύρω από τη θέση του σημείου-κλειδιού στο επίπεδο της εικόνας. Ο υπολογισμός γίνεται πάλι προσεγγιστικά, χρησιμοποιώντας, αντί για τις πραγματικές τιμές των παραγώγων, τις διαφορές στις τιμές της  $L$  για γειτονικά pixels. Τελικά, ο κύριος προσανατολισμός στην περιοχή του σημείου-κλειδιού προσδιορίζεται μέσω του σχηματισμού ενός Ιστογράμματος των γωνιών των παραπάνω διανυσμάτων κλίσης για τα pixels της περιοχής, και της εύρεσης της κατεύθυνσης που αντιστοιχεί στη μεγαλύτερη συχνότητα στο Ιστόγραμμα.

Τώρα, όσο αφορά τον προσδιορισμό του διανύσματος χαρακτηριστικών που αντιστοιχεί σε κάθε σημείο-κλειδί, χρησιμοποιείται ένα νέο ιστόγραμμα, τριών διαστάσεων. Οι πρώτες δύο διαστάσεις αντιστοιχούν στο μήκος και το πλάτος γύρω από το σημείο-κλειδί, και το Ιστόγραμμα ως προς καθεμιά από αυτές ορίζεται με ανάλυση 4. Η τρίτη διάσταση του Ιστογράμματος κβαντοποιεί τις τιμές των γωνιών, και έχει ανάλυση 8. Εν τέλει, αυτό που μοντελοποιείται πάλι μέσω του Ιστογράμματος αυτού είναι η τοπική κατανομή των κατευθύνσεων της κλίσης. Μόνο που τώρα από τις γωνίες των διανυσμάτων κλίσης έχει αφαιρεθεί αυτή που αντιστοιχεί στον κύριο προσανατολισμό της περιοχής, ώστε τα χαρακτηριστικά που θα προκύψουν να ορίζονται εν σχέσει με αυτόν, και να είναι αμετάβλητα κατά την περίπτωση περιστροφής ολόκληρης της περιοχής γύρω από το σημείο-κλειδί.

Επομένως, η περιγραφή που προκύπτει από μια γκριζα εικόνα μέσω της μεθόδου SIFT δεν είναι άλλη από ένα σύνολο διανυσμάτων χαρακτηριστικών που το καθένα έχει διάσταση 128, αντιστοιχεί σε ένα σημείο-κλειδί και περιγράφει την κανονικοποιημένη ως προς την κλίμακα και τον προσανατολισμό τοπική κατανομή των γωνιών της κλίσης της εικόνας, μέσω ενός Ιστογράμματος μεγέθους  $4 \times 4 \times 8$ . Αυτά είναι τα αμετάβλητα (invariant) χαρακτηριστικά που παρέχει η μέθοδος. Με κάθε ένα από τα προηγούμενα διανύσματα χαρακτηριστικών συνδέεται και πληροφορία η οποία δεν παραμένει σταθερή σε περιπτώσεις αλλαγής κλίμακας των αντικειμένων, περιστροφής τους, αλλαγής του σημείου παρατήρησης κτλ. Αυτή έχει να κάνει με τη συγκεκριμένη θέση στο επίπεδο της εικόνας, την κλίμακα και τον κύριο προσανατολισμό που υπολογίστηκαν για το αντίστοιχο σημείο-κλειδί. Βλέπουμε, έτσι, ότι ο μετασχηματισμός SIFT καταφέρνει να προσδιορίσει τα πιο χαρακτηριστικά σημεία των αντικειμένων της εικόνας, και να διαχωρίσει την πληροφορία που σχετίζεται με αυτά και είναι επιρρεπής σε μεταβολές, από αυτήν που εμφανίζει μεγαλύτερη σταθερότητα, και αποτελεί επομένως την ταυτότητά τους.

## 6.4.2 Παραλλαγή της Μεθόδου SIFT για Έγχρωμες Εικόνες

Όπως αναφέραμε και παραπάνω, η μέθοδος υπολογισμού χαρακτηριστικών SIFT εφαρμόζεται σε γκριζες εικόνες. Εν γένει, όμως, στα προβλήματα ανάκτησης εικόνων που μας ενδιαφέρουν, οι εικόνες που καλούμαστε να χειριστούμε είναι έγχρωμες. Εξάλλου, η πληροφορία που συνδέεται με το χρώμα είναι πολύ σημαντική, και άρα αξίζει τον κόπο να επιχειρηθεί η ενσωμάτωσή της στις περιγραφές των εικόνων. Έγινε επομένως προσπάθεια για εύρεση μιας τεχνικής που θα ακολουθεί τη φιλοσοφία της μεθόδου SIFT, αλλά παράλληλα θα είναι ικανή να ενσωματώσει χρωματική πληροφορία.

Σε αυτό το πλαίσιο βασιστήκαμε σε μια παραλλαγή της μεθόδου SIFT που προτάθηκε στην εργασία [72]. Για τη συγκεκριμένη παραλλαγή παρέχεται ελεύθερα κώδικας μέσω του δικτυακού τόπου [79]. Αυτή είναι και η υλοποίηση που χρησιμοποιήσαμε για τον υπολογισμό των χαρακτηριστικών των εικόνων στα πειράματα που παρουσιάζονται παρακάτω σε αυτό το κεφάλαιο.

Το θέμα της ενσωμάτωσης της χρωματικής πληροφορίας στο πλαίσιο της τεχνικής υπολογισμού χαρακτηριστικών SIFT αντιμετωπίζεται στην εργασία [72] με ένα πολύ απλό τρόπο. Αντί να υπολογίζουμε και να χρησιμοποιούμε αποκλειστικά την κλίση της συνάρτησης φωτεινότητας της εικόνας, όπως στην τεχνική SIFT, μπορούμε να επεκτείνουμε την ιδέα αυτή, αντιμετωπίζοντας με τον ίδιο τρόπο, εκτός από τις τιμές της φωτεινότητας, και τις τιμές που αντιστοιχούν σε χρωματικές συνιστώσες για κάθε pixel.

Συγκεκριμένα, χρησιμοποιείται το Γκαουσιανό μοντέλο ανταγωνιστικών χρωμάτων της εργασίας [80], που προκύπτει από το χρωματικό χώρο RGB μέσω του γραμμικού μετασχηματισμού

$$\begin{bmatrix} E \\ E_\lambda \\ E_{\lambda\lambda} \end{bmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.60 & 0.17 \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (6.18)$$

Για κάθε pixel, από τις τιμές των συνιστωσών  $R$ ,  $G$  και  $B$  υπολογίζονται οι τιμές  $E$ ,  $E_\lambda$  και  $E_{\lambda\lambda}$ . Η συνιστώσα  $E$  αντιστοιχεί πάλι στη φωτεινότητα, που είναι η μοναδική πληροφορία που λαμβάνεται υπόψιν στην περίπτωση της κλασικής μεθόδου SIFT. Τα άλλα δύο κανάλια,  $E_\lambda$  και  $E_{\lambda\lambda}$ , μεταφέρουν την καθαρά χρωματική πληροφορία. Χρησιμοποιώντας την παραπάνω χρωματική αναπαράσταση, υπολογίζονται για τα pixels της εικόνας τα επόμενα νέα χρωματικά χαρακτηριστικά:

$$C_\lambda = \frac{E_\lambda}{E} \quad (6.19)$$

$$C_{\lambda\lambda} = \frac{E_{\lambda\lambda}}{E} \quad (6.20)$$

Αυτά δεν είναι άλλα από τις κανονικοποιημένες ως προς τη φωτεινότητα χρωματικές συνιστώσες. Αυτή η κανονικοποίηση έχει ως επίπτωση τη σταθερότητα των  $C_\lambda$  και  $C_{\lambda\lambda}$  κατά τις μεταβολές στη σκίαση των αντικειμένων.

Κατόπιν, οι συναρτήσεις που ορίζονται πάνω στο επίπεδο της εικόνας από τις τιμές των  $C_\lambda$  και  $C_{\lambda\lambda}$  για τα διάφορα pixels χρησιμοποιούνται με τον ίδιο τρόπο όπως και η συνάρτηση φωτεινότητας, για να προκύψουν τα τελικά χαρακτηριστικά. Με άλλα λόγια, η διαδικασία υπολογισμού χαρακτηριστικών που αφορούσε τη φωτεινότητα στα πλαίσια της τεχνικής SIFT, εδώ επαναλαμβάνεται, πέρα από τη φωτεινότητα, και για τις χρωματικές συναρτήσεις που αφορούν τα  $C_\lambda$  και  $C_{\lambda\lambda}$ . Έτσι, αντί για ένα διάνυσμα 128 χαρακτηριστικών ανά εντοπιζόμενο σημείο-κλειδί, τώρα παίρνουμε τρία τέτοια διανύσματα. Το ένα από αυτά είναι το ίδιο με αυτό που θα παίρναμε αν εφαρμόζαμε τη μέθοδο SIFT αφού πρώτα μετατρέπαμε την εικόνα μας σε γκρίζα, κρατώντας μόνο την πληροφορία για τη φωτεινότητα των pixels και αγνοώντας τελείως τη χρωματική πληροφορία που συνδέεται με αυτά. Τα άλλα δύο διανύσματα είναι αυτά που περιέχουν τα χαρακτηριστικά χρώματος, υπολογισμένα στο πλαίσιο της προσέγγισης που ακολουθεί η τεχνική SIFT.

Μέσω μιας σειράς πειραμάτων που παρουσιάστηκαν στην εργασία [72], τα παραπάνω χαρακτηριστικά αποδείχτηκαν σταθερά σε αλλαγές στις συνθήκες φωτισμού και στη γεωμετρία των αντικειμένων, σε μεταβολές στις σκιάσεις κτλ. Επίσης, αποδείχτηκαν αρκετά σταθερά σε αλλαγές όσο αφορά το χρησιμοποιούμενο χρώμα για το φωτισμό της σκηνής.

### 6.4.3 Τελική Περιγραφή των Εικόνων

Όπως αναλύσαμε στην παραπάνω ενότητα, μέσω της χρησιμοποιούμενης μεθοδολογίας υπολογισμού χαρακτηριστικών καταλήγουμε σε μια περιγραφή ανά εικόνα που περιλαμβάνει για κάθε σημείο-κλειδί 3 διανύσματα των 128 χαρακτηριστικών το καθένα. Συνενώνοντας σε ένα μεγάλο διάνυσμα τις διανυσματικές περιγραφές που αντιστοιχούν στο ίδιο σημείο-κλειδί, οδηγούμαστε εν τέλει σε ένα σύνολο διανυσμάτων χαρακτηριστικών για κάθε εικόνα, που το καθένα έχει 384 διαστάσεις.

Η απευθείας χρήση διανυσμάτων χαρακτηριστικών τόσο μεγάλης διάστασης, στην περίπτωση που μέσω αυτών θα επιχειρηθεί, όπως θα δούμε παρακάτω, η εκπαίδευση Γκαουσιανών Μειγμάτων που θα περιγράψουν την κατανομή τους, δεν ενδείκνυται, κάτι που επισημάνθηκε και στα πλαίσια της εργασίας [75]. Αντίθετα, η εκ των προτέρων μείωση της διάστασης των διανυσμάτων αυτών ενδέχεται να έχει ευεργετικά αποτελέσματα όσο αφορά την αποσυσχέτιση των χαρακτηριστικών μεταξύ τους, πράγμα που μπορούμε να εκμεταλλευτούμε μετέπειτα, χρησιμοποιώντας διαγώνιους πίνακες συνδιακύμανσης για τις συνιστώσες των μοντέλων που θα εκπαιδευσουμε. Επίσης, η μείωση του αριθμού των χαρακτηριστικών οδηγεί στη μείωση του θορύβου που ενυπάρχει στα δεδομένα και στον πιο γρήγορο και εύρωστο υπολογισμό των παραμέτρων των μοντέλων.

Επομένως, κάτω από αυτό το πρίσμα, αποφασίσαμε να προχωρήσουμε σε μια γενναία μείωση της διάστασης των δεδομένων μας, μέσω της τεχνικής της ανάλυσης κυρίων συνιστωσών (Principal Component Analysis – PCA). Η εν λόγω τεχνική, με δεδομένο ένα σύνολο  $N$   $d$ -διάστατων παρατηρήσεων  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ , χρησιμοποιεί τον στατιστικά υπολογιζόμενο πίνακα συνδιακύμανσης:

$$\tilde{\Sigma} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \tilde{\boldsymbol{\mu}})(\mathbf{x}_i - \tilde{\boldsymbol{\mu}})^T \quad (6.21)$$

όπου

$$\tilde{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (6.22)$$

είναι η αντίστοιχη μέση τιμή των παρατηρήσεων, και υπολογίζει τα ιδιοδιανύσματα και τις ιδιοτιμές που του αντιστοιχούν. Υπενθυμίζουμε, εδώ, ότι αν  $\mathbf{e}_i$  είναι ένα ιδιοδιάνυσμα του  $\tilde{\Sigma}$  και  $\lambda_i$  η αντίστοιχη ιδιοτιμή, τότε θα ισχύει η σχέση

$$\tilde{\Sigma} \mathbf{e}_i = \lambda_i \mathbf{e}_i \quad (6.23)$$

Μετά από αυτή τη διαδικασία, αν  $s$ , με  $s < d$ , είναι η επιθυμητή νέα διάσταση για τα δεδομένα μας, προσδιορίζουμε τις  $s$  μεγαλύτερες ιδιοτιμές και προβάλλουμε τις παρατηρήσεις μας, μετά την αφαίρεση της μέσης τιμής  $\tilde{\boldsymbol{\mu}}$  από καθεμιά από αυτές, στα

αντίστοιχα ιδιοδιανύσματα. Έτσι, αν με  $\mathbf{x}_i^n$  συμβολίζουμε τις νέες παρατηρήσεις μικρότερης διάστασης, για την  $j$ -οστή συνιστώσα καθεμιάς από αυτές θα ισχύει

$$x_{ij}^n = (\mathbf{x}_i - \tilde{\boldsymbol{\mu}})^T \mathbf{e}_j^{\max}, \quad j = 1, \dots, s \quad (6.24)$$

όπου με  $\mathbf{e}_j^{\max}$  συμβολίζονται τα ιδιοδιανύσματα που αντιστοιχούν στις  $s$  μεγαλύτερες ιδιοτιμές.

Την παραπάνω διαδικασία την εφαρμόσαμε για τα διανύσματα χαρακτηριστικών όλων των εικόνων της βάσης, αφού πρώτα, μέσω μιας λογικής παρόμοιας με αυτήν που αναλύθηκε στο κεφάλαιο 4, και συγκεκριμένα στην ενότητα 4.7.1.1, τα κανονικοποιήσαμε ώστε να έχουν μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση σε κάθε διάσταση. Τελικά, αποφασίσαμε τα νέα διανύσματα χαρακτηριστικών, μετά την ανάλυση κυρίων συνιστωσών, να έχουν διάσταση ίση με 50.

Από τον ορισμό της, η τεχνική ανάλυσης κυρίων συνιστωσών απεικονίζει τις παρατηρήσεις σε ένα νέο σύστημα συντεταγμένων, για το οποίο ισχύει ότι η διασπορά των παρατηρήσεων στην  $i$ -οστή διάσταση είναι μεγαλύτερη από τη διασπορά τους στη διάσταση  $i + 1$  και μικρότερη από αυτή που αντιστοιχεί στη διάσταση  $i - 1$ . Με αυτόν τον τρόπο, επιλέγοντας τις πρώτες  $s$  διαστάσεις μέσω της παραπάνω διαδικασίας, διατηρούμε όσο το δυνατόν περισσότερη από τη διασπορά που αντιστοιχούσε στα αρχικά δεδομένα. Από την άλλη, όμως, γίνεται εύκολα αντιληπτό ότι καθεμιά από τις διαστάσεις των νέων δεδομένων μας χαρακτηρίζεται πια από διαφορετική τυπική απόκλιση. Επομένως, απαιτείται μια νέα κανονικοποίηση για το σύνολο των διανυσμάτων χαρακτηριστικών της βάσης εικόνων, ώστε να εξασφαλίσουμε ότι αυτά θα έχουν μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση για καθεμιά από τις 50 διαστάσεις τους.

Μετά και από αυτήν την τελευταία κανονικοποίηση προκύπτουν τα τελικά διανύσματα χαρακτηριστικών, που θα χρησιμοποιηθούν για τα πειράματα τα αποτελέσματα των οποίων παρουσιάζονται και σχολιάζονται σε επόμενη ενότητα αυτού του κεφαλαίου.

## 6.5 Συγκρινόμενες Μέθοδοι

Σε αυτήν την ενότητα παρουσιάζουμε τις μεθόδους που συγκρίνονται στα πειράματα που ακολουθούν, και διευκρινίζουμε διάφορα θέματα σχετικά με την υλοποίηση της καθεμιάς από αυτές. Επισημαίνουμε εξαρχής ότι για όλα τα Γκαουσιανά Μείγματα που χρησιμοποιούνται από τις διάφορες μεθόδους έχουν υιοθετηθεί διαγώνιοι πίνακες συνδιακύμανσης για τις συνιστώσες τους. Αυτή η επιλογή επιταχύνει πολύ την εκπαίδευση και το χειρισμό των μοντέλων, χωρίς να οδηγεί σε αξιόλογη μείωση της απόδοσης των μεθόδων.

### 6.5.1 Προτεινόμενη Μέθοδος

Η κύρια μέθοδος η οποία προτείνεται στα πλαίσια αυτού του κεφαλαίου, περιλαμβάνει τη χρήση Γκαουσιανών Μειγμάτων για τη μοντελοποίηση της κατανομής των διανυσμάτων χαρακτηριστικών που υπολογίζονται τοπικά από τις εικόνες. Συγκεκριμένα, ένα Γκαουσιανό Μείγμα 8 συνιστωσών εκπαιδεύεται για κάθε εικόνα. Για την εκπαίδευση του μοντέλου κάθε εικόνας της βάσης χρησιμοποιείται ο αλγόριθμος MAP-EM, που παρουσιάστηκε στην ενότητα 6.3.2.2.

Αυτός ο αλγόριθμος απαιτεί, όπως είπαμε, ένα εκ των προτέρων εκπαιδευμένο γενικό Γκαουσιανό Μείγμα πάνω σε δεδομένα από όλη τη βάση. Η εκπαίδευση αυτού του

μοντέλου στηρίζεται στην αρχή της μεγιστοποίησης της πιθανοφάνειας, και, επομένως, για αυτό το σκοπό μπορούμε να στηριχθούμε στον ίδιο αλγόριθμο που χρησιμοποιήσαμε και στο κεφάλαιο 4 για την εκτίμηση των παραμέτρων των μειγμάτων. Υπενθυμίζουμε ότι ο αλγόριθμος αυτός είναι ο άπληστος αλγόριθμος EM, και έχει περιγραφεί αναλυτικά στην ενότητα 2.4.3.2 του κεφαλαίου 2. Για την εκπαίδευση του γενικού μοντέλου χρησιμοποιήθηκαν περίπου 65000 διανύσματα χαρακτηριστικών, τυχαία επιλεγμένα από τη βάση εικόνων πάνω στην οποία γίνεται η αξιολόγηση.

Για την εκπαίδευση των μοντέλων των εικόνων μέσω του αλγορίθμου MAP-EM, για την παράμετρο  $\tau$  (βλ. σχέση (6.9)), που έχει να κάνει με το βάρος με το οποίο λαμβάνονται υπόψιν οι παράμετροι του γενικού μοντέλου στον υπολογισμό των μοντέλων των εικόνων, επιλέξαμε να χρησιμοποιήσουμε  $\tau = 15$ . Αυτή η τιμή αποδείχτηκε εμπειρικά ότι αντιστοιχεί σε μια καλή επιλογή, κρινόμενη από τη μεριά των αποτελεσμάτων που προκύπτουν. Εξάλλου, όπως επισημάνθηκε και στην εργασία [77], τα αποτελέσματα από τη μετέπειτα χρήση των Γκαουσιανών Μειγμάτων που εκπαιδεύονται μέσω του αλγορίθμου MAP-EM εμφανίζουν αρκετά μεγάλη σταθερότητα για τιμές της παραμέτρου  $\tau$  μεταξύ του 8 και του 20.

Μετά την εκπαίδευση των μοντέλων για τις εικόνες, για τη σύγκριση μεταξύ τους στηρίζομαστε στην απόσταση που ορίζεται από τη σχέση (6.12) και εκμεταλλεύεται, όπως έχουμε ήδη αναφέρει, τον τρόπο εκτίμησης των παραμέτρων των μοντέλων, με αποτέλεσμα να μπορεί να υπολογιστεί ιδιαίτερα γρήγορα, χωρίς ουσιαστικά την ανάγκη προσδιορισμού μιας αντιστοιχίας μεταξύ των συνιστωσών των υπό σύγκριση μοντέλων. Συγκεκριμένα, χρησιμοποιούμε τη συμμετρική εκδοχή της:

$$SKL_{id-match}(p_1, p_2) = \frac{1}{2}KL_{id-match}(p_1 \parallel p_2) + \frac{1}{2}KL_{id-match}(p_2 \parallel p_1) \quad (6.25)$$

για να ορίσουμε μια κατάλληλη συνάρτηση πυρήνα της μορφής (6.1), που θα μας επιτρέψει να χρησιμοποιήσουμε ταξινομητές SVM για την αντιμετώπιση του προβλήματος της συσχετιστικής ανατροφοδότησης, με δεδομένη την αναπαράσταση των εικόνων μέσω Γκαουσιανών Μειγμάτων.

## 6.5.2 Παραλλαγές

Προκειμένου να αξιολογηθούν οι διάφορες επιμέρους επιλογές που σχετίζονται με την παραπάνω μέθοδο συσχετιστικής ανατροφοδότησης, μια σειρά από παραλλαγές της μεθόδου σχεδιάστηκαν και υλοποιήθηκαν.

Συγκεκριμένα, για τη μελέτη της επίδρασης στην απόδοση του συστήματος του αριθμού των συνιστωσών των μειγμάτων για τις εικόνες, που υπενθυμίζουμε ότι είναι πάντα ίσος με τον αριθμό των συνιστωσών του γενικού μείγματος από το οποίο αυτά προσαρμόστηκαν, ένα σύνολο παραλλαγών της προτεινόμενης μεθόδου υλοποιήθηκαν, με την καθεμιά από αυτές να χρησιμοποιεί μοντέλα διαφορετικής πολυπλοκότητας.

Επίσης, προκειμένου να διαπιστώσουμε το πως επηρεάζονται τα αποτελέσματα από την υιοθέτηση του μέτρου απόστασης  $SKL_{id-match}$ , συγκρίναμε την προτεινόμενη μέθοδο με μια παραλλαγή που χρησιμοποιεί στη θέση του παραπάνω μέτρου το μέτρο  $SKL_{gkl-match}$ , που στα προκαταρκτικά πειράματα της ενότητας 6.2 μας έδωσε καλύτερα αποτελέσματα σε σχέση με το ομοειδές μέτρο  $SKL_{mhl-match}$ . Υπενθυμίζουμε ότι και τα τρία παραπάνω μέτρα απόστασης ορίζονται μέσω του ίδιου γενικού τύπου (σχέση (6.11)), και διαφέρουν μόνο όσο αφορά τη συνάρτηση αντιστοίχισης που υιοθετεί το καθένα.

Επιπλέον, για την εξακρίβωση της σπουδαιότητας των πλεονεκτημάτων που απορρέουν από τη χρήση του αλγορίθμου MAP-EM για την εκτίμηση των παραμέτρων για τα μοντέλα των εικόνων, εν συγκρίσει με τον κλασικό αλγόριθμο EM, μια ακόμα παραλλαγή υλοποιήθηκε. Η συγκεκριμένη παραλλαγή χρησιμοποιεί για κάθε εικόνα ένα Γκαουσιανό Μείγμα εκπαιδευμένο με βάση την αρχή της μέγιστης πιθανοφάνειας και με χρήση του άπληστου αλγορίθμου EM. Σε αυτήν την περίπτωση, η συνάρτηση πυρήνα για τα SVMs που θα χρησιμοποιηθούν σε κάθε γύρο της ανατροφοδότησης ορίζεται μέσω της απόστασης  $SKL_{gkl-match}$ , μια και η απόσταση  $SKL_{id-match}$  δε μπορεί να χρησιμοποιηθεί, αφού δεν ισχύει πλέον η 1-1 φυσική αντιστοιχία μεταξύ των συνιστωσών των μοντέλων για τις εικόνες. Υπενθυμίζουμε ότι η τελευταία είναι άμεσο αποτέλεσμα του τρόπου εκπαίδευσης των μοντέλων μέσω του αλγορίθμου MAP-EM.

Ακόμα, προκειμένου να μελετήσουμε την επίδραση του συνόλου διανυσμάτων χαρακτηριστικών που χρησιμοποιείται για την εκπαίδευση του γενικού μοντέλου, εξετάσαμε μια παραλλαγή για την οποία το γενικό μοντέλο έχει εκπαιδευτεί με διανύσματα χαρακτηριστικών που προέρχονται μόνο από έναν πολύ μικρό αριθμό εικόνων της βάσης, αλλά και μια ακόμη για την οποία το γενικό μοντέλο έχει εκπαιδευτεί πάνω σε διανύσματα χαρακτηριστικών που προέρχονται από διαφορετική βάση εικόνων από αυτήν που χρησιμοποιείται κατά τη διαδικασία της αξιολόγησης.

### 6.5.3 Μέθοδος Βασισμένη σε Οπτικά Λεξικά

Τέλος, η κύρια σύγκριση γίνεται με μια μέθοδο που ως αναπαραστάσεις για τις εικόνες δε χρησιμοποιεί Γκαουσιανά Μείγματα, αλλά διανύσματα. Για να πάρουμε κατάλληλες διανυσματικές αναπαραστάσεις για τις εικόνες καταφεύγουμε στο σχηματισμό ενός Οπτικού Λεξικού (Visual Vocabulary). Η τεχνική αυτή είναι εμπνευσμένη από τη συναφή επιστημονική περιοχή της ανάκτησης κειμένου, και έχει αποδεικτεί ότι οδηγεί σε πολύ αξιολογικά αποτελέσματα.

Η λογική πίσω από αυτήν την προσέγγιση είναι απλή, και αφορά ουσιαστικά μια διαδικασία ομαδοποίησης (clustering) στο χώρο των διανυσμάτων χαρακτηριστικών, μέσω της οποίας επιτυγχάνεται ο σχηματισμός ενός αριθμού ομάδων (clusters). Κάθε ομάδα αντιστοιχεί και σε μια οπτική λέξη (visual word). Για αυτήν την ομαδοποίηση, ένα σύνολο διανυσμάτων χαρακτηριστικών από όλη τη βάση απαιτείται. Μετά το σχηματισμό των ομάδων, για καθένα από τα διανύσματα χαρακτηριστικών κάθε εικόνας διαπιστώνεται σε ποια από τις σχηματισμένες ομάδες ανήκει. Μετρώντας πόσα από τα διανύσματα χαρακτηριστικών της εικόνας ανήκουν σε κάθε ομάδα και κανονικοποιώντας με το συνολικό αριθμό των διανυσμάτων χαρακτηριστικών της εικόνας, σχηματίζεται ένα Ιστογράμμο που μπορεί να χρησιμοποιηθεί για την αναπαράσταση της τελευταίας.

Στα πλαίσια αυτού του κεφαλαίου, για να είμαστε συνεπείς και με την προσέγγιση που ακολουθείται από την κύρια μέθοδο, χρησιμοποιούμε μια παραλλαγή αυτής της ιδέας, που βασίζεται σε Γκαουσιανά Μείγματα. Ένα γενικό Γκαουσιανό Μείγμα εκπαιδεύεται πάνω σε διανύσματα χαρακτηριστικών από όλη τη βάση, ακριβώς όπως στην περίπτωση της κύριας μεθόδου. Έπειτα, για κάθε εικόνα της βάσης υπολογίζονται (κατά τα γνωστά, μέσω του τύπου του Bayes) οι εκ των υστέρων πιθανότητες  $P(j|x_i)$  για κάθε διάνυσμα χαρακτηριστικών που της αντιστοιχεί, έστω  $x_i$ , και για κάθε συνιστώσα  $j$  του γενικού μοντέλου. Τα bins του Ιστογράμματος που θα χρησιμοποιηθεί για την αναπαράσταση της εικόνας είναι ίσα με τον αριθμό των συνιστωσών του γενικού μοντέλου, και κάθε συνιστώσα του τελευταίου αντιστοιχεί σε μια οπτική λέξη. Αν  $N$  είναι ο αριθμός των

διανυσμάτων χαρακτηριστικών της εικόνας, η συχνότητα των τελευταίων για κάθε bin  $j$  υπολογίζεται μέσω της σχέσης

$$c_j^n = \frac{\sum_{i=1}^N P(j | \mathbf{x}_i)}{N} \quad (6.26)$$

όπου, προφανώς, το άθροισμα όλων των  $c_j^n$  ισούται με τη μονάδα, λόγω του ότι ισχύει  $\sum_j P(j | \mathbf{x}_i) = 1$  για κάθε  $\mathbf{x}_i$ , αφού υποθέτουμε ότι το  $\mathbf{x}_i$  προέκυψε από κάποια συνιστώσα του γενικού Γκαουσιανού Μείγματος.

Έτσι, προκύπτει μια διανυσματική αναπαράσταση για κάθε εικόνα, με τόσες διαστάσεις όσες και οι συνιστώσες του γενικού μοντέλου, η οποία χρησιμοποιείται μετέπειτα για την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης με βάση κλασικούς ταξινομητές SVM, και συνάρτηση πυρήνα τη Γκαουσιανή συνάρτηση ακτινικής βάσης.

Για περισσότερες πληροφορίες πάνω στην τεχνική των Οπτικών Λεξικών παραπέμπουμε στις εργασίες [75] και [76].

## 6.6 Πειράματα

Για τη μελέτη της αποτελεσματικότητας και των ιδιοτήτων της προτεινόμενης μεθόδου, όπως αυτή περιγράφηκε στην ενότητα 6.5.1, ένας αριθμός πειραμάτων σχεδιάστηκε και εκτελέστηκε. Εδώ, παραθέτουμε κάποιες λεπτομέρειες σχετικά με αυτά τα πειράματα, και ασφαλώς τα αποτελέσματα που προέκυψαν.

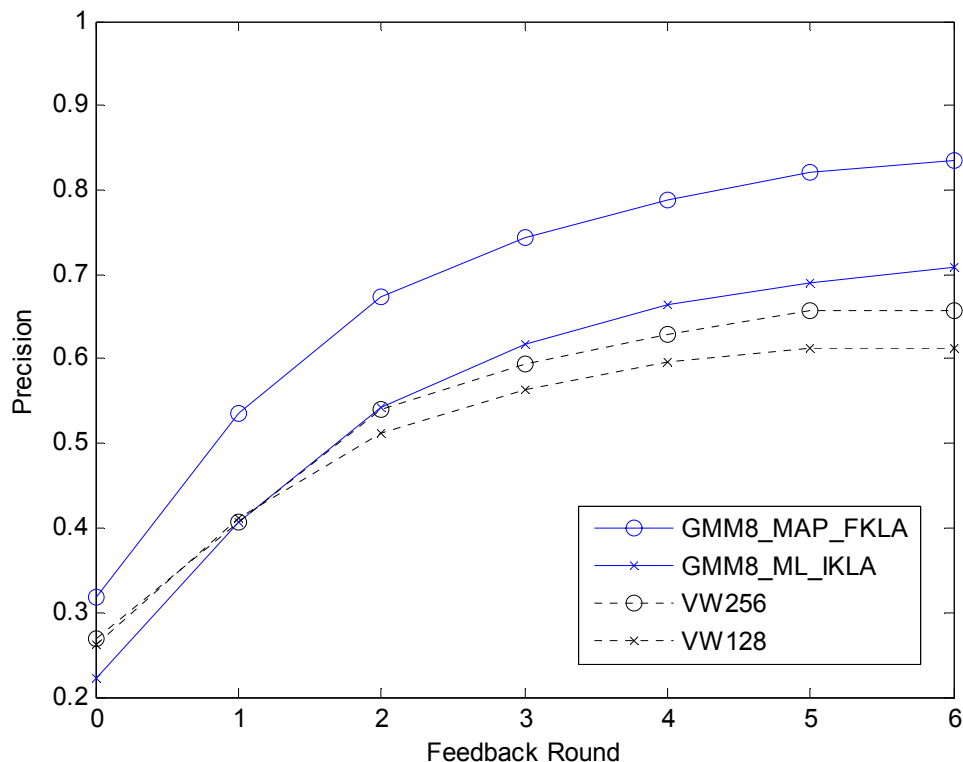
Η βάση εικόνων που χρησιμοποιήθηκε για τα πειράματα είναι ένα υποσύνολο της Corel. Συγκεκριμένα, 4500 εικόνες από την Corel επιλέχθηκαν και χωρίστηκαν σε 25 σημασιολογικές κατηγορίες. Όπως έχουμε επισημάνει ήδη, παρατηρείται πολύ συχνά το φαινόμενο εικόνες της Corel με παρόμοιο σημασιολογικό περιεχόμενο να είναι μοιρασμένες σε διαφορετικές κατηγορίες. Έτσι, αρκετές από τις 25 κατηγορίες της τελικής κατηγοριοποίησης που χρησιμοποιείται παρακάτω προέκυψαν από τη συνένωση δύο ή και περισσότερων κατηγοριών της Corel. Η βάση αυτή από εδώ και πέρα θα αποκαλείται Corel4500. Για περισσότερες πληροφορίες σε σχέση με αυτήν ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στην ενότητα B3 του παραρτήματος Β. Για κάθε εικόνα αυτής της βάσης χρησιμοποιήθηκε η μεθοδολογία που περιγράφηκε στην ενότητα 6.4, προκειμένου να υπολογιστεί ένας αριθμός από διανύσματα χαρακτηριστικών. Κατά μέσο όρο, γύρω στα 700 διανύσματα χαρακτηριστικών υπολογίστηκαν από κάθε εικόνα. Βλέπουμε λοιπόν ότι, πράγματι, σε κάθε εικόνα αντιστοιχούν σχετικά λίγα διανύσματα χαρακτηριστικών.

Το σενάριο προσομοίωσης που χρησιμοποιήθηκε για την ποσοτικοποίηση της απόδοσης των συγκρινόμενων μεθόδων ακολουθεί την ίδια φιλοσοφία με τις υπόλοιπες προσομοιώσεις που παρουσιάζονται στα πλαίσια αυτής της διατριβής. Συγκεκριμένα, 1000 εικόνες επιλέγονται τυχαία από την Corel4500 και παρουσιάζονται στο σύστημα ως αρχικές ερωτήσεις. Από τις 50 πρώτες θέσεις της διάταξης, κατά μέγιστο 5 πραγματικά σχετικές και 5 πραγματικά μη σχετικές εικόνες, με βάση την εκ των προτέρων γνωστή κατηγοριοποίηση των εικόνων, επιλέγονται με τυχαίο τρόπο και υποβάλλονται στο σύστημα ως παραδείγματα ανατροφοδότησης. Μετά από αυτό, ένας ταξινομητής SVM εκπαιδεύεται με βάση αυτά τα παραδείγματα, αλλά και όλα τα άλλα που έχουν δοθεί σε προηγούμενες εποχές, και χρησιμοποιείται για την ανανέωση της διάταξης των εικόνων της βάσης. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να ολοκληρωθούν 6 γύροι

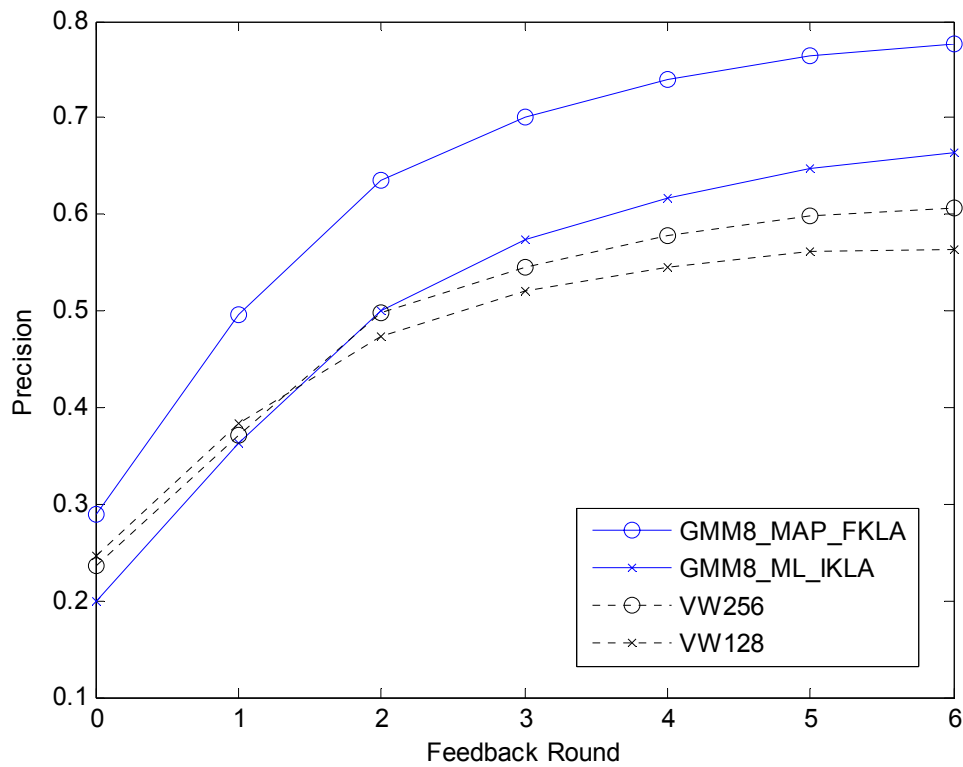


συσχετιστικής ανατροφοδότησης. Ως μέτρο για την αξιολόγηση των αποτελεσμάτων χρησιμοποιείται η ακρίβεια στις πρώτες  $T$  εικόνες της διάταξης.

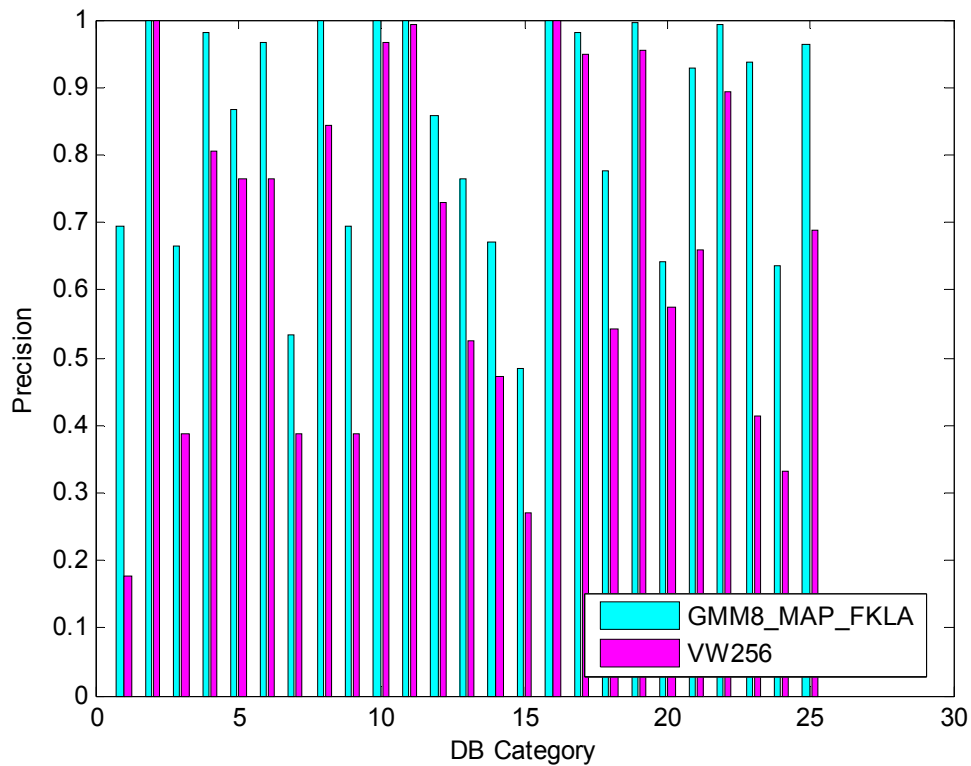
Για τα αποτελέσματα των πειραμάτων που ακολουθούν χρησιμοποιήθηκε πάλι η υλοποίηση της μεθοδολογίας SVM που παρέχεται στο δικτυακό τόπο [67]. Επίσης, οι παράμετροι  $C$  και  $\gamma$  που συνδέονται με τη χρήση των ταξινομητών SVM υπολογίστηκαν εμπειρικά. Συγκεκριμένα, παρατηρήθηκε ότι τα αποτελέσματα δεν είναι ιδιαίτερα ευαίσθητα στις μεταβολές της τιμής της παραμέτρου εκπαίδευσης  $C$  (βλ. κεφάλαιο 5), με την προϋπόθεση ότι για αυτήν έχει επιλεγεί μια σχετικά μεγάλη τιμή. Σε όλα τα πειράματα που ακολουθούν χρησιμοποιείται η τιμή  $C = 100$ . Για τον καθορισμό της τιμής της παραμέτρου  $\gamma$  της συνάρτησης πυρήνα (είτε της Γκαουσιανής συνάρτησης ακτινικής βάσης, είτε αυτής της μορφής (6.1), που χρησιμοποιείται για Γκαουσιανά Μείγματα), απαιτείται περισσότερη προσοχή, γιατί μια λάθος επιλογή μπορεί να επηρεάσει σημαντικά τα αποτελέσματα. Για συναρτήσεις πυρήνα της μορφής  $\exp(-\gamma d(a,b))$ , όπου  $d(a,b)$  είναι το χρησιμοποιούμενο μέτρο απόστασης μεταξύ των προτύπων, στην εργασία [66] έχει προταθεί το  $\gamma$  να τίθεται ίσο με το αντίστροφο της μέσης τιμής του  $d(a,b)$ , όπως αυτή υπολογίζεται σε κάποιο αρκετά μεγάλο σύνολο προτύπων. Παρατηρήσαμε ότι, εν γένει, η καλύτερη τιμή για το  $\gamma$  δεν απέχει πολύ από αυτήν που προκύπτει με τον παραπάνω τρόπο. Επομένως, στην πράξη μπορούμε εύκολα να έχουμε μια αρκετά καλή εκτίμηση για την τιμή και αυτής της παραμέτρου.



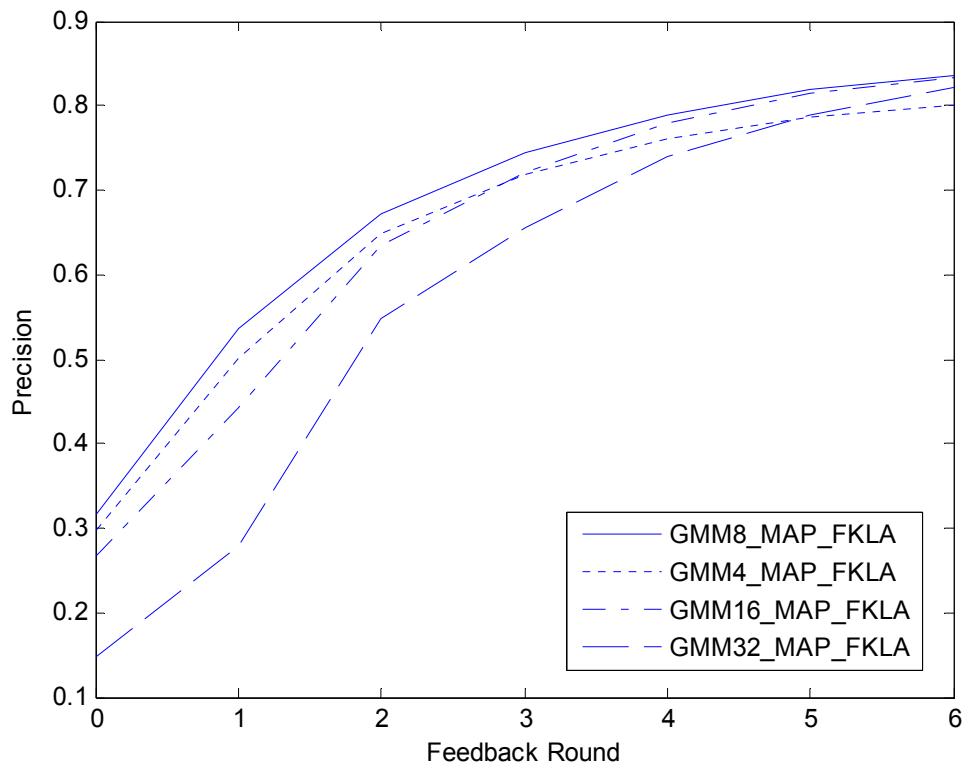
**Σχήμα 6.4:** Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Οπτικών Λεξικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια  $T = 20$



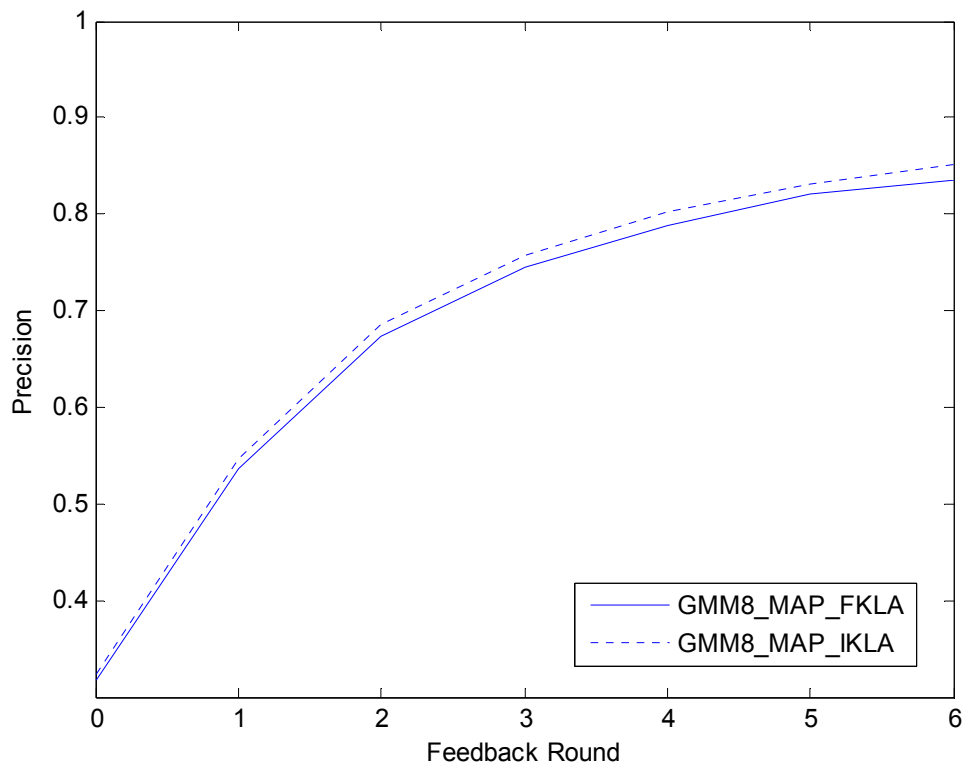
**Σχήμα 6.5:** Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Οπτικών Λεξικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια  $T = 30$



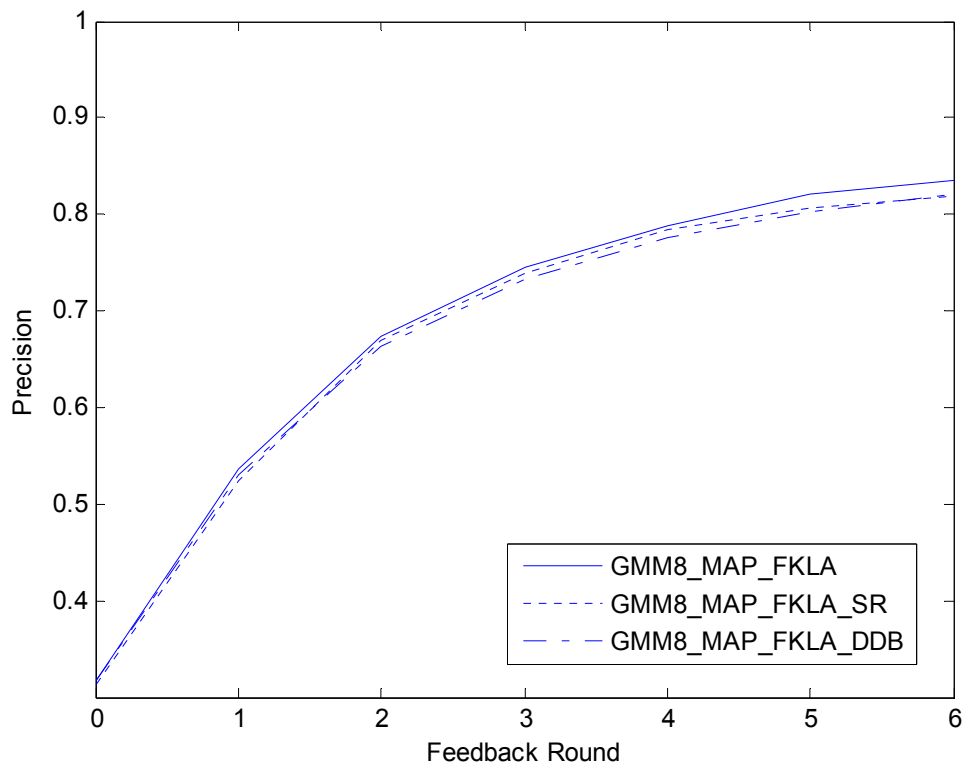
**Σχήμα 6.6:** Σύγκριση μεταξύ Γκαουσιανών Μειγμάτων και Οπτικών Λεξικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια ανά κατηγορία μετά τον 6<sup>ο</sup> γύρο της ανατροφοδότησης, σε εμβέλεια  $T = 20$



**Σχήμα 6.7:** Μελέτη σχετικά με την επίδραση του αριθμού των Γκαουσιανών συνιστωσών των μοντέλων στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια  $T = 20$



**Σχήμα 6.8:** Σύγκριση μέτρων απόστασης μεταξύ Γκαουσιανών Μειγμάτων στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια  $T = 20$



**Σχήμα 6.9:** Μελέτη σχετικά με την επίδραση του χρησιμοποιούμενου συνόλου εκπαίδευσης του γενικού Γκαουσιανού Μείγματος στο πλαίσιο της συστηματικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια  $T = 20$

### 6.6.1 Αποτελέσματα

Τα αποτελέσματα των πειραμάτων που πραγματοποιήθηκαν δίνονται στα σχήματα 6.4 έως 6.9. Συγκεκριμένα, στα σχήματα 6.4, 6.5 και 6.6 η προτεινόμενη μέθοδος, που χρησιμοποιεί για τις εικόνες αναπαράσταση μέσω Γκαουσιανών Μειγμάτων 8 συνιστωσών, εκπαιδευμένων με χρήση του αλγορίθμου MAP-EM, και κατάλληλη συνάρτηση πυρήνα για τα SVMs, της μορφής (6.1), με βάση την απόσταση  $SKL_{id-match}$ , συγκρίνεται με τη μέθοδο που βασίζεται στην αναπαράσταση για τις εικόνες που προκύπτει με χρήση Οπτικών Λεξικών, και χρησιμοποιεί κλασικούς ταξινομητές SVM.

Στα σχήματα 6.4 και 6.5 απεικονίζεται η μέση ακρίβεια σε εμβέλεια  $T = 20$  και 30, αντίστοιχα, όπως υπολογίζεται για όλες τις αρχικές ερωτήσεις. Με GMM8\_MAP\_FKLA συμβολίζεται η προτεινόμενη μέθοδος. Αντίθετα, με VW256 (VW128) αναφερόμαστε στην περίπτωση χρήσης Οπτικών Λεξικών με 256 (128) οπτικές λέξεις. Επίσης, παρουσιάζονται και τα αποτελέσματα μιας παραλλαγής της κυρίας μεθόδου, που συμβολίζεται με GMM8\_ML\_IKLA, η οποία χρησιμοποιεί για κάθε εικόνα ένα μοντέλο 8 συνιστωσών, εκπαιδευμένο με βάση την αρχή της μέγιστης πιθανοφάνειας. Όπως αναφέραμε και παραπάνω, σε αυτήν την περίπτωση δε μπορεί να χρησιμοποιηθεί για τον ορισμό της συνάρτησης πυρήνα το μέτρο  $SKL_{id-match}$ . Αντίθετα, χρησιμοποιείται η αρχική προσέγγιση της απόκλισης KL,  $SKL_{gkl-match}$ . Όπως είναι φανερό από τα σχήματα, η προτεινόμενη μέθοδος οδηγεί σταθερά σε πολύ καλύτερα αποτελέσματα από όλες τις υπόλοιπες με τις οποίες συγκρίνεται. Επιπλέον, ο διπλασιασμός του μεγέθους του Οπτικού Λεξικού, από

128 οπτικές λέξεις σε 256, μας δίνει μια σχετικά μικρή βελτίωση στην επίδοση. Ένα επίσης πολύ ενδιαφέρον συμπέρασμα που προκύπτει από τα σχήματα αυτά είναι ότι ακόμα και τα κλασικά εκπαιδευμένα Γκαουσιανά Μείγματα αναδεικνύονται νικητές στη σύγκριση με τις μεθόδους που χρησιμοποιούν Οπτικά Λεξικά. Και από τα δύο σχήματα βλέπουμε ότι παρότι η μέθοδος GMM8\_ML\_IKLA ξεκινάει από χαμηλότερο επίπεδο ακρίβειας σε σχέση με τις VW256 και VW128, μετά από λίγους γύρους ανατροφοδότησης επιτυγχάνει υψηλότερη ακρίβεια και από τις δύο αυτές μεθόδους. Αυτά τα αποτελέσματα αποτελούν ισχυρή ένδειξη των πλεονεκτημάτων που απορρέουν από τη χρήση των Γκαουσιανών Μειγμάτων για την αναπαράσταση των εικόνων.

Στο σχήμα 6.6, τώρα, βλέπουμε μια ανάλυση για τις 25 κατηγορίες της βάσης Corel4500, της ακρίβειας που επιτυγχάνεται στις πρώτες 20 εικόνες της διάταξης μετά και τον 6<sup>ο</sup> γύρο της ανατροφοδότησης. Η επίδοση της προτεινόμενης μεθόδου, GMM8\_MAP\_FKLA, συγκρίνεται με αυτήν της VW256 για καθεμιά κατηγορία της βάσης. Όπως εύκολα μπορούμε να παρατηρήσουμε, δεν υπάρχει ούτε μία κατηγορία για την οποία η μέθοδος VW256 να επικρατεί της προτεινόμενης. Αντίθετα, σε μεγάλο αριθμό κατηγοριών η προτεινόμενη μέθοδος οδηγεί σε εντυπωσιακά μεγαλύτερη ακρίβεια από αυτήν της VW256.

Όσο αφορά το σχήμα 6.7, αυτό παρουσιάζει μια μελέτη της επίδρασης του χρησιμοποιούμενου αριθμού συνιστωσών για τα μοντέλα των εικόνων. Και πάλι, η μέση ακρίβεια για όλες τις αρχικές ερωτήσεις στις πρώτες 20 εικόνες της διάταξης απεικονίζεται. Ένας αριθμός παραλλαγών της GMM8\_MAP\_FKLA, που χρησιμοποιεί μοντέλα 8 συνιστωσών, αξιολογείται. Συγκεκριμένα, βλέπουμε τα αποτελέσματα για τις GMM4\_MAP\_FKLA, GMM16\_MAP\_FKLA και GMM32\_MAP\_FKLA, που χρησιμοποιούν μοντέλα 4, 16 και 32 συνιστωσών, αντίστοιχα. Όπως είναι εμφανές από το εν λόγω σχήμα, η χρήση 8 Γκαουσιανών συνιστωσών για κάθε εικόνα οδηγεί στα καλύτερα αποτελέσματα. Όταν χρησιμοποιούνται 4 συνιστώσες για την αναπαράσταση των εικόνων, μια κάπως χειρότερη επίδοση επιτυγχάνεται. Από την άλλη μεριά, η αύξηση του αριθμού των συνιστωσών για τα μοντέλα δε δείχνει να βελτιώνει την επίδοση του συστήματος. Αντίθετα, οδηγεί σε υποβάθμιση της τελευταίας, ιδιαίτερα κατά τους πρώτους γύρους της ανάκτησης, όταν ο συνολικός αριθμός των παραδειγμάτων είναι ακόμα μικρός.

Στο σχήμα 6.8 μελετάμε το βαθμό στον οποίον επηρεάζει τα αποτελέσματα η χρήση της συμμετρικής εκδοχής της γρήγορα υπολογιζόμενης προσέγγισης της απόκλισης  $SKL_{id-match}$ . Η προτεινόμενη μέθοδος συγκρίνεται με μια παραλλαγή, που συμβολίζεται ως GMM8\_MAP\_IKLA, η οποία στη συνάρτηση πυρήνα για τα SVMs, αντί της  $SKL_{id-match}$ , χρησιμοποιεί την  $SKL_{gkl-match}$ . Όπως είναι αναμενόμενο, η απόσταση  $SKL_{gkl-match}$ , υπολογίζοντας αναλυτικά την αντιστοιχία μεταξύ των συνιστωσών των συγκρινόμενων μειγμάτων, οδηγεί σε καλύτερα αποτελέσματα. Όμως, όπως είναι εμφανές, η διαφορά στην επίδοση είναι πολύ μικρή. Αν λάβουμε τώρα υπόψιν ότι κατά μέσο όρο ο υπολογισμός της  $SKL_{id-match}$  καταναλώνει χρόνο 0.0008 δευτερόλεπτα σε ένα προσωπικό υπολογιστή με ταχύτητα επεξεργαστή 3GHz, ενώ αυτός της  $SKL_{gkl-match}$  απαιτεί 0.004 δευτερόλεπτα, δηλαδή 5 φορές περισσότερο, καταλαβαίνουμε ότι η μικρή θυσία όσο αφορά τα αποτελέσματα αντισταθμίζεται και με το παραπάνω από τη σημαντική μείωση των απαιτήσεων σε χρόνο.

Τελικά, στο σχήμα 6.9 παρουσιάζονται τα αποτελέσματα μιας μελέτης σχετικά με την ευρωστία των αποτελεσμάτων, σε συνάρτηση με το σύνολο εκπαίδευσης του γενικού μοντέλου που απαιτείται από τον αλγόριθμο MAP-EM. Υπενθυμίζουμε ότι για τα αποτελέσματα των προηγούμενων σχημάτων, το χρησιμοποιούμενο κάθε φορά γενικό μοντέλο είχε εκπαιδευτεί πάνω σε περίπου 65000 διανύσματα χαρακτηριστικών, τυχαία επιλεγμένα από όλες τις εικόνες της Corel4500. Εδώ, συγκρίνονται με την

GMM8\_MAP\_FKLA, που βασίζεται στον παραπάνω τρόπο εκτίμησης των παραμέτρων του γενικού μοντέλου, οι παραλλαγές GMM8\_MAP\_FKLA\_SR και GMM8\_MAP\_FKLA\_DDB. Η πρώτη από τις δύο αυτές παραλλαγές χρησιμοποιεί για την προσαρμογή των μοντέλων των εικόνων μέσω του αλγορίθμου MAP-EM ένα γενικό μοντέλο εκπαιδευμένο πάνω σε περίπου 17500 διανύσματα χαρακτηριστικών που προέρχονται από μόνο 25 τυχαία επιλεγμένες εικόνες της βάσης Corel4500. Η δεύτερη παραλλαγή προχωράει ακόμα παραπέρα, χρησιμοποιώντας για την εκπαίδευση του γενικού μοντέλου περίπου 25000 διανύσματα χαρακτηριστικών, τυχαία επιλεγμένα από τις εικόνες μιας βάσης διαφορετικής από την Corel4500, στην οποία θα γίνει η αξιολόγηση. Συγκεκριμένα, για το σκοπό αυτό χρησιμοποιείται η βάση εικόνων MSDB, που τη συναντάμε και σε πολλά άλλα πειράματα στα πλαίσια αυτής της διατριβής. Όπως γίνεται αμέσως αντιληπτό από τα αντίστοιχα αποτελέσματα που παρουσιάζονται στο σχήμα 6.9, μόνο μια πολύ μικρή μείωση της ακρίβειας παρατηρείται και στις δύο περιπτώσεις, αποδεικνύοντας έτσι την αξιολογική ευρωστία της μεθόδου.

## 6.7 Συμπεράσματα

Σε αυτό το κεφάλαιο μελετήθηκε η δυνατότητα ενσωμάτωσης των Γκαουσιανών Μειγμάτων ως αναπαραστάσεων για τις εικόνες, στο πλαίσιο της χρήσης μοντέλων μάθησης, και ειδικά του μοντέλου SVM, για την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης. Για αυτόν το σκοπό χρησιμοποιήθηκε μια κατάλληλη συνάρτηση πυρήνα μεταξύ Γκαουσιανών Μειγμάτων. Σε μια σειρά από προκαταρκτικά πειράματα καταδείχτηκαν οι δυνατότητες βελτίωσης των αποτελεσμάτων της ανάκτησης που απορρέουν από αυτήν την προσέγγιση, σε σύγκριση με τη χρήση κλασικών ταξινομητών SVM, που βασίζονται σε διανυσματικές αναπαραστάσεις των εικόνων μέσω Ιστογραμμάτων.

Κατόπιν, μελετήθηκε η δυνατότητα περαιτέρω βελτίωσης των αποτελεσμάτων που προκύπτουν, μέσω αλλαγής της όλης φιλοσοφίας γύρω από την εκτίμηση των παραμέτρων των μοντέλων για τις εικόνες. Συγκεκριμένα, αντί της κλασικής εκπαίδευσης των μοντέλων με βάση την αρχή της μεγιστοποίησης της πιθανοφάνειας, μια εκπαίδευσή τους στα πλαίσια της Μπεύζιανής μάθησης, μέσω του αλγορίθμου MAP-EM, υιοθετήθηκε. Τα πλεονεκτήματα αυτής της προσέγγισης μεγιστοποιούνται όταν για κάθε εικόνα υπάρχει διαθέσιμο ένα μικρό σύνολο διανυσμάτων χαρακτηριστικών, όποτε και η αρχή της μέγιστης πιθανοφάνειας δεν καταλήγει σε εύρωστη εκτίμηση των παραμέτρων. Οι πλέον δημοφιλείς μέθοδοι υπολογισμού τοπικών χαρακτηριστικών από εικόνες καταλήγουν, πράγματι, σε περιγραφές με σχετικά λίγα και, παράλληλα, πολυδιάστατα διανύσματα χαρακτηριστικών. Συγκεκριμένα, στα πλαίσια των τελικών μας πειραμάτων χρησιμοποιήσαμε μια παραλλαγή της μεθόδου SIFT, που είναι η πιο καταξιωμένη από τις παραπάνω μεθόδους. Επιπλέον, η υιοθέτηση του αλγορίθμου MAP-EM μας δίνει τη δυνατότητα να επιταχύνουμε κατά πολύ τη διαδικασία σύγκρισης των αναπαραστάσεων των εικόνων, χρησιμοποιώντας μια κατάλληλα ορισμένη προσέγγιση της απόκλισης KL.

Για τις ανάγκες των πειραμάτων που διεξήχθησαν, μια άλλη μέθοδος συσχετιστικής ανατροφοδότησης, που χρησιμοποιεί διανυσματικές αναπαραστάσεις για τις εικόνες, εξετάστηκε. Οι αναπαραστάσεις αυτές προκύπτουν μέσω ενός Οπτικού Λεξικού, το οποίο εκπαιδεύεται και αυτό ως ένα Γκαουσιανό Μείγμα. Μετά, αυτές οι αναπαραστάσεις των εικόνων χρησιμοποιούνται στα πλαίσια της συσχετιστικής ανατροφοδότησης, μέσω κλασικών ταξινομητών SVM. Σε αυτήν την περίπτωση, ως συνάρτηση πυρήνα χρησιμοποιείται η Γκαουσιανή συνάρτηση ακτινικής βάσης.

Τα αποτελέσματα που προέκυψαν από τα πειράματα μας δείχνουν μια αναμφισβήτητη υπεροχή της προτεινόμενης μεθόδου σε σχέση με αυτήν που βασίζεται σε διανυσματικές αναπαραστάσεις που προκύπτουν με χρήση Οπτικού Λεξικού. Επίσης, σε καλύτερα αποτελέσματα σε σχέση με την τελευταία μέθοδο οδηγούμαστε ακόμα και αν χρησιμοποιήσουμε κλασικά εκπαιδευμένα Γκαουσιανά Μείγματα για τις εικόνες, πράγμα που τονίζει με τον πιο σαφή τρόπο τα πλεονεκτήματα της αναπαράστασης των εικόνων μέσω Γκαουσιανών Μειγμάτων. Επιπλέον, τα αποτελέσματα των πειραμάτων καταδεικνύουν ότι ένας μικρός σχετικά αριθμός συνιστωσών για τα μοντέλα των εικόνων είναι αρκετός για να πετύχουμε πολύ καλή απόδοση. Τέλος, πρόσθετα πειράματα κατέδειξαν την ακρίβεια της γρήγορα υπολογιζόμενης προσέγγισης της απόκλισης KL, και την ευρωστία των αποτελεσμάτων της προτεινόμενης μεθόδου, στην περίπτωση που ένα όχι αρκετά αντιπροσωπευτικό σύνολο διανυσμάτων χαρακτηριστικών χρησιμοποιείται για την εκπαίδευση του γενικού μοντέλου, από το οποίο προσαρμόζονται τα επιμέρους μοντέλα για τις εικόνες.





## Κεφάλαιο 7

# Συσχετιστική Ανατροφοδότηση και Επιλογή Χαρακτηριστικών

Σε αυτό το κεφάλαιο θα ασχοληθούμε λίγο με το ζήτημα της επιλογής χαρακτηριστικών, το οποίο αποτελεί ένα ενδιαφέρον θέμα που ανακύπτει σχεδόν πάντα κατά την προσπάθεια επίλυσης προβλημάτων ταξινόμησης. Θα προσπαθήσουμε να θέσουμε το πρόβλημα στο πλαίσιο των συστημάτων ανάκτησης εικόνων με χρήση συσχετιστικής ανατροφοδότησης, και θα ερευνήσουμε τις δυνατότητες ενσωμάτωσης μεθόδων επιλογής χαρακτηριστικών σε τέτοιου είδους συστήματα, καθώς και τα επιμέρους ενδιαφέροντα θέματα που αναδύονται. Ως περιγραφή για τις εικόνες, αντί για τα Γκαουσιανά Μείγματα που μελετήθηκαν μέχρι τώρα, εδώ υιοθετούνται πολυδιάστατα διανύσματα οπτικών χαρακτηριστικών, τα οποία ταιριάζουν περισσότερο στο όλο πλαίσιο της προσέγγισης της επιλογής χαρακτηριστικών. Επίσης, αναφέρουμε προκαταβολικά ότι στα πειράματα που θα ακολουθήσουν, η διαδικασία της συσχετιστικής ανατροφοδότησης θα υλοποιηθεί μέσω της εκπαίδευσης ενός ταξινομητή SVM σε κάθε γύρο. Ως αποτέλεσμα αυτού του σκέλους της έρευνας προέκυψε η δημοσίευση [C4].

### 7.1 Επιλογή Χαρακτηριστικών

Σε αυτήν την ενότητα θα παρουσιάσουμε το θέμα της επιλογής χαρακτηριστικών, ως επιμέρους ζήτημα στο πλαίσιο του γενικότερου προβλήματος της ταξινόμησης ενός συνόλου δεδομένων. Και σε αυτό το κεφάλαιο θα ασχοληθούμε κυρίως με προβλήματα δυαδικής ταξινόμησης, μια και, όπως έχουμε αναλύσει ήδη, παρέχουν το απαιτούμενο πλαίσιο για την ενσωμάτωση και ικανοποίηση των στόχων της συσχετιστικής ανατροφοδότησης.

Έστω λοιπόν ότι έχουμε ένα σύνολο προτύπων σε ένα  $d$ -διάστατο διανυσματικό χώρο, το χώρο των χαρακτηριστικών, τα στοιχεία του οποίου συνόλου μπορούν να ανήκουν σε μια από δύο κατηγορίες. Ο στόχος μας είναι να βρούμε έναν τρόπο, ώστε να μπορούμε να εκτιμήσουμε με όσο το δυνατόν μεγαλύτερη βεβαιότητα την κατηγορία στην οποία ανήκει ένα οποιοδήποτε πρότυπο  $\mathbf{x}$  από το εν λόγω σύνολο, και για αυτό θα βασιστούμε σε ένα σύνολο προτύπων εκπαίδευσης  $\mathbf{x}_i$ ,  $i = 1, \dots, N$ , για κάθε ένα από τα οποία είναι γνωστή η κατηγορία του, και η γνώση αυτή εκφράζεται μέσω της αντίστοιχης ετικέτας  $t_i$ , που παίρνει τιμή  $-1$  για τη μια κατηγορία και τιμή  $+1$  για την άλλη κατηγορία. Προφανώς, κάθε ένα από τα πρότυπα εκπαίδευσης είναι ένα  $d$ -διάστατο διάνυσμα, και κάθε μια από τις  $d$  διαστάσεις αντιστοιχεί και σε ένα χαρακτηριστικό (feature). Θεωρούμε ότι το  $d$  στην

περίπτωσή μας είναι ένας πολύ μεγάλος ακέραιος αριθμός, δηλαδή ότι χρησιμοποιείται ένας πολύ μεγάλος αριθμός χαρακτηριστικών για την περιγραφή των προτύπων. Συνήθως, στα προβλήματα τα οποία μας ενδιαφέρουν ο αριθμός των χρησιμοποιούμενων χαρακτηριστικών είναι της τάξης των αρκετών εκατοντάδων ή και χιλιάδων.

Επομένως, προκύπτει πολλές φορές το φαινόμενο να θέλουμε να λύσουμε ένα πρόβλημα ταξινόμησης, χρησιμοποιώντας ένα μοντέλο μάθησης όπως τα SVMs, για παράδειγμα, που παρουσιάστηκαν σε προηγούμενο κεφάλαιο, στο οποίο τα πρότυπα προς ταξινόμηση είναι διανύσματα πολύ μεγάλης διάστασης, και μάλιστα τόσο μεγάλης ώστε πολλές φορές να υπερβαίνει κατά πολύ σε μέγεθος το πλήθος,  $N$ , των διαθέσιμων προτύπων εκπαίδευσης. Αυτό είναι δυνατόν να οδηγήσει σε μείωση της απόδοσης του ταξινομητή, ο οποίος ενδέχεται, λόγω των λίγων και ιδιαίτερα αραιά κατανεμημένων στον εξαιρετικά πολυδιάστατο χώρο παραδειγμάτων που θα έχει στη διάθεσή του, να αδυνατεί να προσεγγίσει ικανοποιητικά την πραγματική κατανομή των δεδομένων που αντιστοιχεί σε καθεμιά από τις δύο κατηγορίες.

Μια λύση για αυτό το πρόβλημα μπορεί να δοθεί μέσω της με κατάλληλο τρόπο επιλογής ενός σχετικά μικρού αριθμού χαρακτηριστικών, από το σύνολο όλων των  $d$  διαθέσιμων χαρακτηριστικών για τα δεδομένα, και της χρήσης μόνο αυτών τόσο κατά τη διαδικασία εκπαίδευσης του ταξινομητή όσο και κατά τη διαδικασία εκτίμησης της κατηγορίας νέων προτύπων. Αυτό που ουσιαστικά γίνεται είναι ο περιορισμός των προτύπων (είτε αυτά ανήκουν στο σύνολο εκπαίδευσης, είτε παρουσιάζονται στον ταξινομητή προς εκτίμηση της κατηγορίας τους) στο σύνολο των επιλεγέντων χαρακτηριστικών, και η απόρριψη όλων των υπόλοιπων χαρακτηριστικών. Η πρακτική, επομένως, αυτή συνεπάγεται μια συνήθως μεγάλη μείωση όσο αφορά την τελική διάσταση των προτύπων.

Η ακόλουθη εκπαίδευση του ταξινομητή στον προκύπτοντα χώρο περιορισμένης διάστασης μπορεί να αυξήσει την ικανότητά του να προσεγγίσει την κατανομή των προτύπων που αντιστοιχεί σε κάθε κατηγορία, και άρα να βελτιώσει τα αποτελέσματά του όσο αφορά τις εκτιμήσεις του πάνω στην κατηγορία νέων προτύπων που του παρουσιάζονται προς κατηγοριοποίηση. Επίσης, μέσω αυτής της μείωσης της διάστασης των δεδομένων μπορεί να επιτευχθεί μείωση στο χρόνο εκπαίδευσης του ταξινομητή, σε σχέση με τον αντίστοιχο απαιτούμενο χρόνο στον αρχικό χώρο μεγαλύτερης διάστασης, αλλά και στο χρόνο που καταναλώνεται προκειμένου να εκτιμηθεί η κατηγορία του εκάστοτε νέου προτύπου.

Ο τρόπος επιλογής του υποσυνόλου χαρακτηριστικών για το οποίο έγινε λόγος προηγουμένως πρέπει, βεβαίως, να συνάδει με και να υπηρετεί τον τελικό σκοπό, ο οποίος είναι η όσο το δυνατόν αποτελεσματικότερη επίλυση του προβλήματος ταξινόμησης. Για αυτό πρέπει να διατηρηθούν εκείνα τα χαρακτηριστικά τα οποία εμφανίζουν μεγάλη ικανότητα στο να διαχωρίζουν πρότυπα διαφορετικών κατηγοριών και, παράλληλα, να κρατάνε κοντά πρότυπα της ίδιας κατηγορίας. Αντίθετα, χαρακτηριστικά που παρουσιάζουν μικρή διακριτική ικανότητα όσο αφορά τις υπό θεώρηση κατηγορίες θα πρέπει να απορρίπτονται. Τέτοια χαρακτηριστικά είτε περιέχουν υψηλά ποσοστά θορύβου, πράγμα που τα καθιστά ακατάλληλα, αν όχι επιζήμια, όσο αφορά το διαχωρισμό των κατηγοριών των δεδομένων, είτε είναι άσχετα με το πρόβλημα ταξινόμησης που επιδιώκουμε να λύσουμε. Για παράδειγμα, μπορεί να κωδικοποιούν πληροφορία που δεν επηρεάζει και δεν επηρεάζεται από τις κατηγορίες δεδομένων που μας ενδιαφέρουν, με αποτέλεσμα η γνώση της κατηγορίας ενός προτύπου να μην προσφέρει επιπλέον πληροφορία όσο αφορά την τιμή των εν λόγω χαρακτηριστικών και, αντίστροφα, η γνώση της τιμής των εν λόγω χαρακτηριστικών να μη αποτελεί επιπλέον θετική ένδειξη για κάποια από τις δύο κατηγορίες. Μια άλλη κατηγορία χαρακτηριστικών που είναι δυνατόν να απορριφθούν άφοβα, είναι αυτά που κωδικοποιούν την ίδια πάνω-κάτω πληροφορία με

κάποια άλλα χαρακτηριστικά που υπάρχουν στα δεδομένα, και τα οποία έχουμε αποφασίσει ήδη να επιλέξουμε. Σε αυτήν την περίπτωση, η χρήση των εν λόγω χαρακτηριστικών κρίνεται ως περιττή, από τη στιγμή που δεν έχουν να μας προσφέρουν κάποια καινούργια γνώση για τα δεδομένα μας.

Η συχνή ύπαρξη θορυβωδών, άσχετων και περιττών χαρακτηριστικών στα δεδομένα, όπως αυτές οι τρεις κατηγορίες χαρακτηριστικών περιγράφηκαν αμέσως παραπάνω, είναι μια ακόμα αιτία που το ενδεχόμενο της επιλογής χαρακτηριστικών είναι από τα πρώτα που εξετάζονται όταν κανείς ασχολείται με προβλήματα ταξινόμησης.

Για την επιλογή των κατ' εκτίμηση καταλληλότερων χαρακτηριστικών για ένα πρόβλημα ταξινόμησης, ευνόητο είναι ότι θα στηριχθούμε στη γνωστή κατηγοριοποίηση των προτύπων εκπαίδευσης, την οποία μπορούμε να και θα αντιμετωπίσουμε ως περιέχουσα αντιπροσωπευτική πληροφορία σχετικά με τη σύνδεση μεταξύ κατηγοριών και χαρακτηριστικών για το πλήρες σύνολο δεδομένων. Από τα πρότυπα εκπαίδευσης και την κατηγοριοποίησή τους θα εξαχθεί η τελική κρίση για το ποια χαρακτηριστικά εμφανίζουν μεγάλη διακριτική ικανότητα σε σχέση με τις εξεταζόμενες κατηγορίες, και επομένως πρέπει να διατηρηθούν για τις διαδικασίες της εκπαίδευσης και εκτίμησης κατηγορίας, και ποια είναι άσχετα με το πρόβλημα ταξινόμησης ή περιττά, και άρα θα απορριφθούν.

Τέλος, πρέπει να επισημανθεί ότι στην πράξη αυτό που περιμένουμε από μια μέθοδο επιλογής χαρακτηριστικών να επιστρέψει, τουλάχιστον σε πρώτη φάση, είναι μια διάταξη του συνόλου των χαρακτηριστικών σε (για παράδειγμα) φθίνουσα σειρά σπουδαιότητας για το πρόβλημα της ταξινόμησης. Έπειτα, ένα άλλο θέμα είναι το πόσα από τα πρώτα στη λίστα χαρακτηριστικά θα διατηρηθούν. Πάνω σε αυτό το θέμα, στο παρόν κεφάλαιο αυτό που θα εξετάσουμε είναι κυρίως το πως επηρεάζεται η απόδοση ενός συστήματος ανάκτησης εικόνων με χρήση συσχετιστικής ανατροφοδότησης από τον αριθμό των διατηρούμενων χαρακτηριστικών, και δε θα προτείνουμε κάποια συγκεκριμένη μέθοδο προσδιορισμού του βέλτιστου αριθμού χαρακτηριστικών προς διατήρηση.

### **7.1.1 Εφαρμογή στη Συσχετιστική Ανατροφοδότηση**

Ήδη, σε προηγούμενο κεφάλαιο έχουμε περιγράψει πως σε κάθε γύρο της συσχετιστικής ανατροφοδότησης, στο πλαίσιο της ανάκτησης εικόνων, το πρόβλημα της ανανέωσης της διάταξης των εικόνων της βάσης μπορεί να αντιμετωπιστεί ως ένα πρόβλημα δυαδικής ταξινόμησης, με τα πρότυπα εκπαίδευσης να αντιστοιχούν στα παραδείγματα του χρήστη, τα πρότυπα προς κατηγοριοποίηση στις εικόνες της βάσης, και με τις δύο εμπλεκόμενες κατηγορίες να είναι αυτή των σχετικών και αυτή των μη σχετικών με την ερώτηση του χρήστη εικόνων.

Λαμβάνοντας υπόψιν το γεγονός ότι, πάντα, όταν αναφερόμαστε στη διαδικασία της συσχετιστικής ανατροφοδότησης πρέπει να κάνουμε τη ρεαλιστική και στην πράξη τεκμηριωμένη υπόθεση ότι ο χρήστης είναι ανυπόμονος, και για αυτόν το λόγο δεν είναι διατεθειμένος ούτε σε πολλούς γύρους ανατροφοδότησης να συμμετάσχει, ούτε πολλά παραδείγματα να προσφέρει σε κάθε γύρο, μπορούμε με βεβαιότητα να συμπεράνουμε ότι τα παραδείγματα που θα έχουμε διαθέσιμα για την εκπαίδευση του ταξινομητή σε κάθε γύρο της ανατροφοδότησης δε θα είναι πολλά. Αν, τώρα, συνυπολογίσουμε ότι ο κλασικός τρόπος περιγραφής των εικόνων είναι μέσω πολυδιάστατων διανυσμάτων οπτικών χαρακτηριστικών, η διάσταση των οποίων όπως θα δούμε λίγο παρακάτω μπορεί εύκολα να προσεγγίσει το 500, μπορούμε να καταλήξουμε στο ασφαλές συμπέρασμα ότι σχεδόν πάντα ο αριθμός των διαθέσιμων παραδειγμάτων του χρήστη θα υπολείπεται, και μάλιστα κατά πολύ, του αριθμού των χαρακτηριστικών που θα χρησιμοποιούνται για την περιγραφή της κάθε εικόνας.

Επομένως, η χρήση, σε συνδυασμό με κάποιο μοντέλο μάθησης για την αντιμετώπιση του προβλήματος της δυαδικής ταξινόμησης, μεθόδων επιλογής χαρακτηριστικών σε κάθε γύρο της ανατροφοδότησης, για τον περιορισμό της διάστασης των περιγραφών των εικόνων, φαντάζει ως μια ενδιαφέρουσα προοπτική.

## 7.1.2 Μέθοδοι Επιλογής Χαρακτηριστικών

Στη βιβλιογραφία έχουν προταθεί πολλές μέθοδοι επιλογής χαρακτηριστικών στα πλαίσια της αντιμετώπισης προβλημάτων ταξινόμησης προτύπων. Ως αντιπροσωπευτικές εργασίες πάνω σε αυτό το επιστημονικό πεδίο αναφέρουμε τις [47], [55]-[60]. Σε αυτό το κεφάλαιο θα ασχοληθούμε εκτενέστερα με τις μεθόδους που παρουσιάζονται στις εργασίες [55] και [57], που έχουν αποδεικτεί ιδιαίτερα δημοφιλείς για την αντιμετώπιση προβλημάτων επιλογής χαρακτηριστικών.

Η πρώτη από τις δύο παραπάνω μεθόδους είναι και αυτή που προτείνουμε, όσο αφορά τη χρήση της σε συνδυασμό με το μοντέλο μάθησης SVM στα πλαίσια της ανάκτησης εικόνων με συσχετιστική ανατροφοδότηση. Αντίθετα, με τη δεύτερη μέθοδο επιλογής χαρακτηριστικών ασχολούμαστε κυρίως για λόγους σύγκρισης.

### 7.1.2.1 Αναδρομική Εξάλειψη Χαρακτηριστικών με Χρήση Ταξινομητών SVM

Αρχικά, λοιπόν, θα ασχοληθούμε με τη μέθοδο που παρουσιάζεται στην εργασία [55], η οποία, όπως θα δούμε παρακάτω, μας δίνει και την υψηλότερη απόδοση στα πειράματα που πραγματοποιήσαμε. Η μέθοδος αυτή αποκαλείται SVM-RFE (Recursive Feature Elimination using SVM classifiers – Αναδρομική Εξάλειψη Χαρακτηριστικών με χρήση ταξινομητών SVM) και, όπως φανερώνει και το όνομά της, βασίζεται σε μια αναδρομική διαδικασία εξάλειψης των κρινόμενων ως λιγότερο σημαντικών για το αντίστοιχο πρόβλημα ταξινόμησης χαρακτηριστικών, χρησιμοποιώντας για το σκοπό αυτόν τα αποτελέσματα που προκύπτουν από κατάλληλα εκπαιδευμένους ταξινομητές SVM.

Λαμβάνοντας υπόψιν ότι, όπως και στο προηγούμενο κεφάλαιο, θα καταφύγουμε στην εκπαίδευση ενός ταξινομητή SVM σε κάθε γύρο της ανατροφοδότησης, προκειμένου να ανανεώσουμε τα αποτελέσματα της ανάκτησης με βάση και τα νέα παραδείγματα του χρήστη, βλέπουμε αμέσως ότι στην προκειμένη περίπτωση το μοντέλο που θα χρησιμοποιηθεί για το στόχο της επιλογής χαρακτηριστικών είναι ίδιο με αυτό που θα χρησιμοποιηθεί μετά την επιλογή των χαρακτηριστικών, για την υλοποίηση της κυρίως διαδικασίας της συσχετιστικής ανατροφοδότησης. Αυτό το γεγονός οδηγεί στο πλεονέκτημα της επιλογής των πλέον κατάλληλων χαρακτηριστικών για την εκπαίδευση του τελικού ταξινομητή σε κάθε γύρο της ανατροφοδότησης.

Ας πούμε λίγα λόγια, όμως, για τη μέθοδο SVM-RFE. Ας θεωρήσουμε καταρχήν ένα σύνολο προτύπων εκπαίδευσης  $\mathbf{x}_i$ ,  $i = 1, \dots, N$ , και την περίπτωση του ταξινομητή SVM με συνάρτηση απεικόνισης τη  $\phi(\mathbf{x}) = \mathbf{x}$ , δηλαδή το SVM στον αρχικό χώρο των προτύπων. Η συνάρτηση πυρήνα που αντιστοιχεί σε αυτήν τη συνάρτηση απεικόνισης είναι, όπως βλέπουμε και από τη σχέση (5.15), η  $k(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T \mathbf{x}_2$ . Για αυτήν τη συνάρτηση πυρήνα, η σχέση (5.22), που ορίζει τη συνάρτηση απόφασης  $y(\mathbf{x})$  για το πρότυπο προς κατηγοριοποίηση  $\mathbf{x}$ , παίρνει τη μορφή

$$y(\mathbf{x}) = \sum_{i=1}^N a_i t_i \mathbf{x}^T \mathbf{x}_i + b = \left( \sum_{i=1}^N a_i t_i \mathbf{x}_i^T \right) \mathbf{x} + b = \mathbf{w}^T \mathbf{x} + b \quad (7.1)$$

με

$$\mathbf{w} = \sum_{i=1}^N a_i t_i \mathbf{x}_i \quad (7.2)$$

αφού, προφανώς, ισχύει  $\mathbf{x}^T \mathbf{x}_i = \mathbf{x}_i^T \mathbf{x}$ .

Το διάνυσμα  $\mathbf{w}$ , όπως έχουμε αναφέρει ήδη στο κεφάλαιο 5, είναι το κάθετο διάνυσμα στο υπερεπίπεδο  $y(\mathbf{x})=0$ , που χωρίζει τις δύο κατηγορίες προτύπων, και έχει την ίδια διάσταση με τα πρότυπα εκπαίδευσης  $\mathbf{x}_i$ . Αν  $d$  είναι η διάσταση των προτύπων εκπαίδευσης (και του προτύπου προς κατηγοριοποίηση  $\mathbf{x}$ , φυσικά), τότε η σχέση (7.1) αναλυόμενη μας δίνει

$$y(\mathbf{x}) = w_1 x_1 + w_2 x_2 + \dots + w_d x_d + b \quad (7.3)$$

όπου  $\mathbf{w} = (w_1, w_2, \dots, w_d)^T$  και  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$ . Επομένως, όσο μεγαλύτερη είναι η απόλυτη τιμή,  $|w_j|$ , της  $j$ -οστής συνιστώσας του διανύσματος  $\mathbf{w}$  (ή, ισοδύναμα, όσο μεγαλύτερη είναι η τιμή  $w_j^2$ ) τόσο μεγαλύτερη είναι η επίδραση της συγκεκριμένης συνιστώσας στην τιμή της συνάρτησης απόφασης για ένα οποιοδήποτε πρότυπο του οποίου θέλουμε να εκτιμήσουμε την κατηγορία. Αυτή είναι η κεντρική ιδέα πίσω από τη μέθοδο SVM-RFE, αφού, όπως θα δούμε παρακάτω, μπορεί να χρησιμοποιηθεί για να μας δώσει μια διάταξη των χαρακτηριστικών των προτύπων.

Συγκεκριμένα, θεωρούμε ότι έχουμε να λύσουμε ένα πρόβλημα δυαδικής ταξινόμησης, για το οποίο μας δίνονται  $N$  πρότυπα εκπαίδευσης,  $\mathbf{x}_i^c$ ,  $i=1, \dots, N$ . Το σύμβολο  $c$  χρησιμοποιείται ως ένδειξη για το ότι μιλάμε για τα αρχικά πρότυπα, αυτά που περιέχουν το πλήρες σύνολο χαρακτηριστικών. Κάθε πρότυπο  $\mathbf{x}_i^c$  χαρακτηρίζεται από την αντίστοιχη ετικέτα  $t_i \in \{-1, +1\}$ , και έχει διάσταση  $d$ , δηλαδή περιλαμβάνει  $d$  χαρακτηριστικά. Επίσης, έστω ότι  $S_f$  είναι το τρέχον σύνολο χαρακτηριστικών. Κατά την έναρξη της διαδικασίας επιλογής χαρακτηριστικών, και δεδομένου ότι η μέθοδος SVM-RFE είναι αναδρομική, το  $S_f$  αρχικοποιείται έτσι ώστε να ισχύει

$$S_f = \{1, 2, \dots, d\} \quad (7.4)$$

Δηλαδή, το  $S_f$  περιλαμβάνει αρχικά όλα τα χαρακτηριστικά, όπως αυτά περιέχονται στα αρχικά πρότυπα  $\mathbf{x}_i^c$ .

Σε κάθε επανάληψη της διαδικασίας παίρνουμε από τα αρχικά πρότυπα  $\mathbf{x}_i^c$  τα τρέχοντα πρότυπα εκπαίδευσης  $\mathbf{x}_i$ , μέσω της σχέσης

$$\mathbf{x}_i = \mathbf{x}_i^c(S_f), \quad i=1, \dots, N \quad (7.5)$$

όπου με την παραπάνω έκφραση συμβολίζουμε τον περιορισμό των αρχικών προτύπων στο τρέχον σύνολο χαρακτηριστικών. Δηλαδή, απλά, κρατάμε τα χαρακτηριστικά που περιλαμβάνονται στο  $S_f$  και αγνοούμε όλα τα υπόλοιπα. Με αυτά τα πρότυπα εκπαίδευσης που προκύπτουν, και τις αντίστοιχες ετικέτες  $t_i$ , που δεν υφίστανται καμία μεταβολή αλλά

χρησιμοποιούνται ως έχουν, εκπαιδεύουμε ένα SVM με συνάρτηση πυρήνα  $k(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T \mathbf{x}_2$ , δηλαδή στον τρέχοντα χώρο χαρακτηριστικών, και υπολογίζουμε το αντίστοιχο κάθετο διάνυσμα  $\mathbf{w}$ , του οποίου οι συνιστώσες βρίσκονται σε πλήρη αντιστοιχία με τα χαρακτηριστικά που περιέχονται στο τρέχον σύνολο  $S_f$ , από τη σχέση (7.2). Στη συνέχεια εντοπίζουμε τη συνιστώσα του  $\mathbf{w}$  με τη μικρότερη τιμή για την ποσότητα  $w_j^2$ , δηλαδή την

$$z = \arg \min_j w_j^2 \quad (7.6)$$

και αφαιρούμε από το σύνολο  $S_f$  το αντίστοιχο στοιχείο-χαρακτηριστικό, ως το λιγότερο σημαντικό σε σχέση με τα άλλα. Δηλαδή, το καινούργιο σύνολο  $S_f$  θα περιέχει τα χαρακτηριστικά

$$S_f(1, \dots, z-1) \cup S_f(z+1, \dots) \quad (7.7)$$

όπου με την παραπάνω έκφραση συμβολίζουμε, απλά, τη διαγραφή από το σύνολο  $S_f$  του  $z$ -οστού στοιχείου του (θεωρούμε ότι τα στοιχεία έχουν συγκεκριμένη θέση μέσα στο εν λόγω σύνολο, δηλαδή, όπως θα λέγαμε με μαθηματικούς όρους, θεωρούμε ότι το σύνολο είναι διατεταγμένο, ώστε η έκφραση αυτή να έχει νόημα).

Η παραπάνω διαδικασία επαναλαμβάνεται, και οδηγεί σε μια διάταξη των χαρακτηριστικών των προτύπων, αφού το χαρακτηριστικό που αφαιρείται πρώτο είναι αυτό που κρίνεται ως το λιγότερο σημαντικό, αυτό που αφαιρείται αμέσως μετά είναι το δεύτερο λιγότερο σημαντικό κτλ. Αυτές οι διαδοχικές αφαιρέσεις χαρακτηριστικών από το σύνολο  $S_f$  μπορούν να συνεχιστούν, έως ότου ο επιθυμητός αριθμός χαρακτηριστικών μείνει στο  $S_f$ . Τότε, τα χαρακτηριστικά που θα έχουν μείνει στο  $S_f$  θα είναι αυτά που θα έχει επιλέξει η μέθοδος.

### 7.1.2.2 Ελάχιστος Πλεονασμός, Μέγιστη Συνάφεια

Σε αντίθεση με την προηγούμενη ενότητα, όπου παρουσιάσαμε μια μέθοδο επιλογής χαρακτηριστικών που ξεκινάει με το αρχικό σύνολο χαρακτηριστικών και αφαιρεί διαδοχικά από αυτό χαρακτηριστικά μέχρι να καταλήξουμε στον επιθυμητό αριθμό χαρακτηριστικών που θα επιλεγούν, στην παρούσα ενότητα θα δούμε μια άλλη μέθοδο, που ακολουθεί την αντίστροφη λογική. Δηλαδή, ξεκινάει με ένα κενό σύνολο επιλεγέντων χαρακτηριστικών και προσθέτει διαδοχικά σε αυτό χαρακτηριστικά, έως ότου επιτευχθεί ο επιθυμητός αριθμός επιλεγέντων χαρακτηριστικών. Η συγκεκριμένη μέθοδος αποκαλείται mRMR (minimum Redundancy, Maximum Relevance – ελάχιστος Πλεονασμός, Μέγιστη Συνάφεια) και προτείνεται στην εργασία [57].

Η μέθοδος θέτει και εξετάζει το πρόβλημα της επιλογής χαρακτηριστικών από μια πιθανοτική βάση, στηριζόμενη στην έννοια της αμοιβαίας πληροφορίας (mutual information), που είναι ένα μέτρο συνάφειας μεταξύ δύο τυχαίων μεταβλητών. Για δύο συνεχείς μονοδιάστατες τυχαίες μεταβλητές  $Y, Z$ , η αμοιβαία πληροφορία μεταξύ τους ορίζεται από τη σχέση

$$I(Y; Z) = \iint p(y, z) \ln \frac{p(y, z)}{p_Y(y)p_Z(z)} dydz \quad (7.8)$$

όπου  $p_Y(y)$ ,  $p_Z(z)$  είναι οι συναρτήσεις πυκνότητας πιθανότητας των μεταβλητών  $Y$ ,  $Z$ , αντίστοιχα, και  $p(y,z)$  η από κοινού (joint) συνάρτηση πυκνότητας πιθανότητας των  $Y$  και  $Z$ . Στην παραπάνω σχέση οι ολοκληρώσεις εκτείνονται σε όλο το εύρος των τιμών που μπορούν να πάρουν οι αντίστοιχες μεταβλητές. Τώρα, αν οι  $Y$  και  $Z$  είναι διακριτές τυχαίες μεταβλητές, η σχέση (7.8) παίρνει τη μορφή

$$I(Y;Z) = \sum_y \sum_z P(y,z) \ln \frac{P(y,z)}{P_Y(y)P_Z(z)} \quad (7.9)$$

όπου, σε αυτήν την περίπτωση,  $P_Y(y)$ ,  $P_Z(z)$  και  $P(y,z)$  είναι συναρτήσεις πιθανότητας για την τυχαία μεταβλητή  $Y$ , για την τυχαία μεταβλητή  $Z$  και για το συνδυασμό των  $Y$  και  $Z$ , αντίστοιχα. Προφανώς, και εδώ τα αθροίσματα συμπεριλαμβάνουν όλες τις τιμές που μπορούν να πάρουν οι αντίστοιχες μεταβλητές.

Μια μεγάλη τιμή για την αμοιβαία πληροφορία μεταξύ δύο τυχαίων μεταβλητών αποτελεί ένδειξη μεγάλης συνάφειας ανάμεσα στις εν λόγω μεταβλητές. Με απλά λόγια, αυτή η μεγάλη συνάφεια μεταφράζεται στο γεγονός ότι η γνώση της τιμής της μιας μεταβλητής παρέχει σημαντική πληροφορία όσο αφορά τη γνώση της τιμής της άλλης, και αντίστροφα.

Στο πλαίσιο της μεθόδου επιλογής χαρακτηριστικών mRMR, τώρα, κάθε ένα από τα  $d$  χαρακτηριστικά των προτύπων θεωρείται ως μια τυχαία μεταβλητή. Ως τυχαία μεταβλητή μοντελοποιείται, επίσης, και η ετικέτα  $t$  που αντιστοιχεί σε κάθε πρότυπο. Με αυτή τη λογική, και σύμφωνα με αυτά που έχουμε αναφέρει στην αρχή του κεφαλαίου σχετικά με τη φιλοσοφία που πρέπει να ακολουθείται για την επιλογή χαρακτηριστικών, θέλουμε να επιλέξουμε τα χαρακτηριστικά εκείνα που εμφανίζουν μεγάλη συνάφεια με την ετικέτα, που κωδικοποιεί την κατηγορία, και, παράλληλα, εμφανίζουν τη μικρότερη δυνατή συνάφεια μεταξύ τους, ώστε να αποφύγουμε τον πλεονασμό που θα προκύψει από την επιλογή χαρακτηριστικών που περιέχουν παρόμοια πληροφορία.

Αυτή η ιδέα εκφράζεται πολύ εύκολα με μια αυξητική διαδικασία επιλογής χαρακτηριστικών. Θεωρούμε το  $C$  ως το σύνολο όλων των χαρακτηριστικών που περιέχονται στα αρχικά πρότυπα. Επομένως, θα ισχύει

$$C = \{1,2,\dots,d\} \quad (7.10)$$

Επίσης, με  $S_f$  συμβολίζουμε το σύνολο των χαρακτηριστικών που έχουν ήδη επιλεγεί. Αρχικά, βέβαια, το  $S_f$  δεν περιέχει κανένα χαρακτηριστικό, και έτσι έχουμε

$$S_f = \emptyset \quad (7.11)$$

Κατά την πρώτη επανάληψη του αλγορίθμου προσδιορίζουμε εκείνο το χαρακτηριστικό από το  $C$  που εμφανίζει τη μεγαλύτερη συνάφεια με την ετικέτα, με τη συνάφεια να ποσοτικοποιείται μέσω του μέτρου της αμοιβαίας πληροφορίας. Δηλαδή, αν με  $z$  συμβολίσουμε αυτό το χαρακτηριστικό, τότε θα ισχύει

$$z = \arg \max_{f \in C} D(f) \quad (7.12)$$

όπου

$$D(f) = I(f;t) \quad (7.13)$$

Μετά από αυτό το βήμα, προσθέτουμε το χαρακτηριστικό  $z$  στο σύνολο  $S_f$ . Επομένως, θα έχουμε πλέον

$$S_f = \{z\} \quad (7.14)$$

Για τις επόμενες επαναλήψεις το ζητούμενο, όπως είπαμε, είναι να εντοπίσουμε χαρακτηριστικά που δεν έχουν ήδη επιλεγεί, δηλαδή ανήκουν στο σύνολο  $C - S_f$  (αφαίρεση συνόλων), τα οποία έχουν μεγάλη συνάφεια με την ετικέτα  $t$ , και παράλληλα να εξασφαλίσουμε ότι δε θα επιλέξουμε χαρακτηριστικά που θα έχουν μεγάλη συνάφεια με αυτά που έχουν ήδη επιλεγεί, δηλαδή με τα χαρακτηριστικά που υπάρχουν ήδη στο  $S_f$ . Αυτό μαθηματικά εκφράζεται μέσω των παρακάτω σχέσεων:

$$z = \arg \max_{f \in C - S_f} \Phi(f) \quad (7.15)$$

$$\Phi(f) = D(f) - R(f) \quad (7.16)$$

όπου η ποσότητα  $D(f)$  υπολογίζεται μέσω της (7.13), και εκφράζει πάλι τη συνάφεια του τρέχοντος χαρακτηριστικού από το σύνολο  $C - S_f$  με την ετικέτα, ενώ για την  $R(f)$  ισχύει

$$R(f) = \frac{1}{N_f} \sum_{g \in S_f} I(f; g) \quad (7.17)$$

με  $N_f$  να είναι ο αριθμός των χαρακτηριστικών που έχουν ήδη επιλεγεί και περιέχονται στο  $S_f$ . Επομένως, η  $R(f)$  ισούται με τη μέση τιμή της αμοιβαίας πληροφορίας μεταξύ του υποψηφίου προς επιλογή χαρακτηριστικού  $f$  του συνόλου  $C - S_f$  και των ήδη επιλεγμένων χαρακτηριστικών στο  $S_f$ , και χρησιμοποιείται ως ένα μέτρο ενδεικτικό του πλεονασμού που συνδέεται με την επιλογή του χαρακτηριστικού  $f$ . Μετά από την εύρεση του χαρακτηριστικού  $z$  μέσω της σχέσης (7.15), αυτό προστίθεται στο σύνολο  $S_f$ . Έτσι, το ανανεωμένο σύνολο  $S_f$  θα περιέχει τα χαρακτηριστικά

$$S_f \cup \{z\} \quad (7.18)$$

Αυτή η διαδικασία επαναλαμβανόμενη οδηγεί σε μια διάταξη των χαρακτηριστικών των προτύπων, αφού το πρώτο χαρακτηριστικό που προστίθεται στο  $S_f$  είναι αυτό που κρίνεται από τη μέθοδο ως το πλέον σημαντικό χαρακτηριστικό, αυτό που εισάγεται αμέσως μετά μπορεί να θεωρηθεί ως το δεύτερο σημαντικότερο κτλ. Οι επαναλήψεις διαρκούν μέχρι να εισαχθεί στο σύνολο  $S_f$  ο επιθυμητός αριθμός χαρακτηριστικών. Προφανώς, αυτά θα είναι και τα χαρακτηριστικά που επιλέγονται τελικά από τη μέθοδο.

Το μόνο σημείο που χρήζει, εδώ, περαιτέρω ανάλυσης είναι ο τρόπος υπολογισμού των τιμών της αμοιβαίας πληροφορίας που χρησιμοποιούνται από τη μέθοδο, και γενικά ο τρόπος πιθανοτικής μοντελοποίησης τόσο των χαρακτηριστικών όσο και της ετικέτας των προτύπων. Ως προς αυτό, παρατηρούμε αρχικά ότι, όταν αναφερόμαστε σε διακριτές τυχαίες μεταβλητές, η αμοιβαία πληροφορία υπολογίζεται σχετικά εύκολα, αφού, όπως βλέπουμε από τη σχέση (7.9), δεν έχουμε πλέον ολοκληρώματα αλλά αθροίσματα πάνω σε όλες τις πιθανές τιμές των τυχαίων μεταβλητών. Τώρα, για την πιθανοτική μοντελοποίηση



των χαρακτηριστικών και της ετικέτας των προτύπων θα βασιστούμε, όπως ήταν αναμενόμενο, στα πρότυπα εκπαίδευσης.

Συγκεκριμένα, οι τιμές που μπορεί να πάρει η ετικέτα ενός προτύπου είναι από τη φύση τους διακριτές (θα έχει τιμή είτε  $-1$ , είτε  $+1$ ). Από την άλλη, οι τιμές που μπορεί να πάρει το κάθε χαρακτηριστικό μπορούν πολύ εύκολα να διακριτοποιηθούν, αν υπολογίσουμε τη μέση τιμή και την τυπική απόκλιση (μέσω των σχέσεων (4.46) και (4.47)), έστω  $\mu$ ,  $\sigma$ , αντίστοιχα, των τιμών του χαρακτηριστικού για τα πρότυπα του συνόλου εκπαίδευσης. Τότε, για κάθε χαρακτηριστικό ορίζονται τρία μη επικαλυπτόμενα διαστήματα, και συγκεκριμένα τα  $(-\infty, \mu - \sigma]$ ,  $(\mu - \sigma, \mu + \sigma)$  και  $[\mu + \sigma, +\infty)$ . Κβαντοποιώντας τις τιμές του υπό θεώρηση χαρακτηριστικού σε 3 στάθμες (π.χ.  $-1$ ,  $0$  και  $+1$ ), με καθεμία να αντιστοιχεί και σε ένα από αυτά τα διαστήματα, δηλαδή αντιστοιχίζοντας σε οποιαδήποτε τιμή του χαρακτηριστικού εντός καθενός από αυτά τα διαστήματα την αντίστοιχη κοινή στάθμη, καταλήγουμε σε διακριτά χαρακτηριστικά, καθένα από τα οποία μπορεί να πάρει τρεις διαφορετικές τιμές. Από εκεί και πέρα τα πράγματα είναι απλά, και οι τιμές των συναρτήσεων πιθανότητας για τις διάφορες δυνατές τιμές των διακριτοποιημένων χαρακτηριστικών και της ετικέτας των προτύπων, που μας χρειάζονται για τους υπολογισμούς που προβλέπονται από τη μέθοδο, προσδιορίζονται καταμετρώντας τον αριθμό των προτύπων εκπαίδευσης που εμφανίζουν αντίστοιχες τιμές για τα χαρακτηριστικά και την ετικέτα και κανονικοποιώντας με το συνολικό αριθμό των προτύπων εκπαίδευσης.

## 7.2 Οπτικά Χαρακτηριστικά

Στην τρέχουσα ενότητα θα αναφερθούμε στα οπτικά χαρακτηριστικά που χρησιμοποιήθηκαν για τα πειράματα που διεξήχθησαν. Τα πειράματα αυτά αφορούν την ενσωμάτωση των προηγουμένως παρουσιασθέντων μεθόδων επιλογής χαρακτηριστικών στη διαδικασία συσχετιστικής ανατροφοδότησης, στα πλαίσια της ανάκτησης εικόνων, και τα αποτελέσματά τους παρουσιάζονται παρακάτω σε αυτό το κεφάλαιο.

Συγκεκριμένα, για την περιγραφή των εικόνων χρησιμοποιήθηκαν τρεις κατηγορίες οπτικών χαρακτηριστικών:

1. Χαρακτηριστικά χρώματος
2. Χαρακτηριστικά υφής
3. Χαρακτηριστικά σχήματος

Στις ενότητες που ακολουθούν θα δούμε αναλυτικά τα χαρακτηριστικά που περιλαμβάνονται σε κάθε μια από τις παραπάνω κατηγορίες.

### 7.2.1 Χαρακτηριστικά Χρώματος

Ως χαρακτηριστικά χρώματος για τις εικόνες στα πειράματα αυτού του κεφαλαίου χρησιμοποιήθηκαν τα εξής:

1. Ιστόγραμμα στο χώρο χρώματος HSV: Σε διάφορες εργασίες, όπως για παράδειγμα η [61], έχει επισημανθεί και επιβεβαιωθεί πειραματικά ότι όταν θέλουμε να περιγράψουμε εικόνες μέσω Ιστογραμμάτων ο χώρος χρώματος HSV είναι μια από τις καλύτερες επιλογές. Για το χώρο χρώματος HSV έχουμε αναφερθεί στο κεφάλαιο 1, ενώ για τα Ιστογράμματα μιλήσαμε στο κεφάλαιο 2. Το μόνο που

πρέπει να διευκρινίσουμε σε αυτή τη φάση είναι η κβαντοποίηση που υιοθετήσαμε για τα Ιστογράμματα. Συγκεκριμένα, επιλέξαμε να υπολογίσουμε ένα Ιστόγραμμα  $8 \times 8 \times 4$ , δηλαδή να διαιρέσουμε το εύρος τιμών των συνιστωσών H και S σε 8 ίσα διαστήματα, και το εύρος τιμών της συνιστώσας V σε 4. Συνολικά, λοιπόν, το κάθε Ιστόγραμμα περιλαμβάνει 256 bins. Η επιλογή μας αυτή είναι μια από τις τυπικές επιλογές που εμφανίζονται στη βιβλιογραφία. Για παραδείγματα εφαρμογής αυτών των χαρακτηριστικών, ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στις εργασίες [21], [59] και [71].

2. Ροπές χρώματος στο χώρο χρώματος LAB: Συγκεκριμένα, για κάθε μια από τις τρεις συνιστώσες L, A και B, και θεωρώντας τις τιμές που αυτή παίρνει για όλα τα pixels της υπό περιγραφή εικόνας, υπολογίζονται οι τρεις πρώτες ροπές, δηλαδή η μέση τιμή, η τυπική απόκλιση και η ασυμμετρία (skewness). Τη μέση τιμή και την τυπική απόκλιση ενός συνόλου παρατηρήσεων τις έχουμε ορίσει στο κεφάλαιο 4, μέσω των σχέσεων (4.46) και (4.47), αντίστοιχα. Η ασυμμετρία, τώρα, έστω  $s$ , ορίζεται ως

$$s = \frac{\sum_{i=1}^N (x_i - \mu)^3}{(N-1)\sigma^3} \quad (7.19)$$

αν  $x_i$  είναι οι  $N$  το πλήθος (μονοδιάστατες) παρατηρήσεις και  $\mu$ ,  $\sigma$  είναι η υπολογιζόμενη για αυτές μέση τιμή και τυπική απόκλιση, αντίστοιχα. Μετά τον υπολογισμό των τριών αυτών ροπών για καθεμιά από τις τρεις συνιστώσες χρώματος της εικόνας, καταλήγουμε εν τέλει με ένα διάνυσμα ροπών 9 διαστάσεων. Επισημαίνουμε εδώ ότι τέτοιου είδους χαρακτηριστικά είναι πολύ δημοφιλή όσο αφορά την περιγραφή εικόνων στα πλαίσια του προβλήματος της ανάκτησης εικόνων, και συναντώνται πολύ συχνά στη βιβλιογραφία, όπως για παράδειγμα στις εργασίες [21], [59] και [71].

## 7.2.2 Χαρακτηριστικά Υφής

Η υφή, με την οποία ασχοληθήκαμε και στο κεφάλαιο 4, και συγκεκριμένα στην ενότητα 4.6.2, είναι μια πολύ σημαντική ιδιότητα των εικόνων, και για αυτό στις περιγραφές των εικόνων που χρησιμοποιούν τα διάφορα συστήματα ανάκτησης σχεδόν πάντα περιλαμβάνεται και κωδικοποιείται πληροφορία υφής.

Για τον υπολογισμό χαρακτηριστικών υφής από τις εικόνες επιλέξαμε να βασιστούμε στο δενδρικό δομημένο μετασχηματισμό Wavelet (tree-structured Wavelet transform) [62] τριών επιπέδων. Για παραδείγματα εφαρμογής των χαρακτηριστικών που προκύπτουν με αυτόν τον τρόπο, μπορεί κανείς να ανατρέξει στις εργασίες [59], [61] και [63].

Θεωρώντας ότι έχουμε μια γκρίζα εικόνα, η εφαρμογή του μετασχηματισμού Wavelet σε ένα επίπεδο έγκειται στο φιλτράρισμα της εικόνας τόσο κατά γραμμές όσο και κατά στήλες με δύο μονοδιάστατα φίλτρα, ένα βαθυπερατό και ένα υψιπερατό.

Υπενθυμίζουμε επιγραμματικά ότι όταν αναφερόμαστε σε φιλτράρισμα ενός διακριτού σήματος, έστω  $y$ , μέσω ενός φίλτρου με κρουστική απόκριση  $h$ , εννοούμε την πράξη της συνέλιξης μεταξύ των  $y$  και  $h$ :

$$(y * h)(n) = \sum_k h(k)y(n-k) \quad (7.20)$$

Επίσης, ως βαθυπερατό θεωρείται εκείνο το φίλτρο του οποίου η κρουστική απόκριση είναι τέτοια ώστε στο χώρο της συχνότητας, από τη μια, να επιτρέπει να διέλθουν οι χαμηλές φασματικές συνιστώσες του σήματος που φιλτράρεται και, από την άλλη, να καταπιέζει τις υψηλές φασματικές συνιστώσες. Αντίστροφα, υπιπερατό θεωρείται εκείνο το φίλτρο που επιτρέπει να διέλθουν οι υψηλές φασματικές συνιστώσες του σήματος που φιλτράρει, και παράλληλα καταπιέζει τις αντίστοιχες χαμηλές.

Μετά το φιλτράρισμα ενός σήματος με ένα φίλτρο όπως αυτά που περιγράψαμε παραπάνω, λόγω της συρρίκνωσης που προκαλείται στο φάσμα του σήματος, μπορούμε να προχωρήσουμε σε υποδειγματοληψία, μειώνοντας στο μισό τον αριθμό των δειγμάτων του φιλτραρισμένου σήματος.

Επομένως, φιλτράροντας κατά γραμμές και στήλες με τα φίλτρα που αναφέραμε προηγουμένως την εικόνα για την οποία θέλουμε να υπολογίσουμε το μετασχηματισμό Wavelet, και μετά την κατάλληλη υποδειγματοληψία των σημάτων που προκύπτουν, καταλήγουμε με τέσσερα σήματα-συνιστώσες, καθένα από τα οποία έχει τις μισές γραμμές και στήλες σε σχέση με την αρχική εικόνα. Το πρώτο από αυτά τα προκύπτοντα σήματα, συμβολιζόμενο ως LL, αντιστοιχεί σε βαθυπερατό φιλτράρισμα τόσο κατά γραμμές όσο και κατά στήλες και αποτελεί μια προσέγγιση της αρχικής εικόνας, ή, όπως αλλιώς θα μπορούσαμε να πούμε, μια θολωμένη εκδοχή της αρχικής εικόνας, στην οποία έχουν αμβλυνθεί οι απότομες μεταβολές της φωτεινότητας. Τα δύο επόμενα σήματα συμβολίζονται με LH και HL, αντίστοιχα, προκύπτουν από βαθυπερατό φιλτράρισμα κατά τη μία διάσταση και υπιπερατό κατά την άλλη, και κωδικοποιούν τις λεπτομέρειες όσο αφορά τις οριζόντιες και τις κάθετες ακμές της αρχικής εικόνας. Τέλος, το τελευταίο σήμα, HH, είναι αυτό που προκύπτει από υπιπερατό φιλτράρισμα και στις δύο διαστάσεις, και κωδικοποιεί λεπτομέρειες σε σχέση με τις διαγώνιες ακμές της εικόνας.

Ο δενδρικός δομημένος μετασχηματισμός Wavelet, τώρα, προκύπτει από τη διαδοχική, πολυεπίπεδη ανάλυση των συνιστωσών LL, LH και HL. Με άλλα λόγια, στο πρώτο επίπεδο αναλύουμε την αρχική εικόνα στις συνιστώσες LL, LH, HL και HH, όπως περιγράψαμε παραπάνω. Στο δεύτερο επίπεδο κάνουμε ό,τι κάναμε για την αρχική εικόνα για καθεμιά από τις συνιστώσες LL, LH, HL και παίρνουμε, στο σύνολο, 12 νέες συνιστώσες. Τέλος, στο τρίτο επίπεδο, αναλύουμε τις 9 από τις 12 συνιστώσες του προηγούμενου επιπέδου (εκτός των αντίστοιχων HH που προκύπτουν από την ανάλυση των LL, LH και HL) εκ νέου και παίρνουμε  $9 \times 4 = 36$  ακόμα συνιστώσες. Αυτό μπορεί να συνεχιστεί και σε επόμενα επίπεδα, αλλά εδώ, όπως είπαμε, θα χρησιμοποιήσουμε δενδρικό δομημένο μετασχηματισμό Wavelet τριών επιπέδων.

Επομένως, κατά την ανάλυση τριών επιπέδων που παρουσιάστηκε στην προηγούμενη παράγραφο, η αρχική εικόνα αναλύθηκε σε συνολικά  $4 + 12 + 36 = 52$  συνιστώσες. Καθεμιά από τις τελευταίες είναι ένα σήμα με λιγότερες γραμμές και στήλες από ό,τι η αρχική εικόνα. Συγκεκριμένα, οι 4 συνιστώσες του πρώτου επιπέδου έχουν τις μισές γραμμές και στήλες της αρχικής εικόνας, οι 12 συνιστώσες του δεύτερου επιπέδου έχουν το  $1/4$  των γραμμών και στηλών της αρχικής εικόνας, και οι 36 συνιστώσες του τελευταίου επιπέδου το  $1/8$ .

Για να πάρουμε τα τελικά χαρακτηριστικά από αυτήν την ανάλυση της εικόνας που περιγράψαμε, χρησιμοποιούμε όλες τις παραπάνω συνιστώσες, και για καθεμιά από αυτές υπολογίζουμε (κατά τα γνωστά) τη μέση τιμή των απολύτων των συντελεστών που αντιστοιχούν στα pixels της. Αυτή η ποσότητα αποτελεί ένα μέτρο της ενέργειας που περιέχεται στο τμήμα του φάσματος της εικόνας που απομονώνεται μέσω των φιλτραρισμάτων που αντιστοιχούν στη συγκεκριμένη συνιστώσα. Επίσης, υπολογίζουμε, ως ένα ακόμα χαρακτηριστικό, την τυπική απόκλιση των παραπάνω συντελεστών. Επομένως, για κάθε εικόνα υπολογίζουμε 2 χαρακτηριστικά για καθεμιά από τις 52 συνιστώσες του μετασχηματισμού, δηλαδή, συνολικά, 104 χαρακτηριστικά υψής.

Πριν την εφαρμογή του μετασχηματισμού και τον υπολογισμό των χαρακτηριστικών, πρέπει προφανώς να έχουμε ήδη μετατρέψει την εικόνα που αναλύουμε από έγχρωμη σε γκριζα, διατηρώντας μόνο την πληροφορία που σχετίζεται με τη φωτεινότητα για κάθε pixel, και αγνοώντας την αντίστοιχη χρωματική πληροφορία. Ως φίλτρα για την υλοποίηση του μετασχηματισμού Wavelet υιοθετήθηκαν τα ευρέως χρησιμοποιούμενα και παράλληλα ιδιαίτερα απλά φίλτρα Haar. Το βαθυπερατό φίλτρο Haar ουσιαστικά υπολογίζει τη μέση τιμή μεταξύ γειτονικών pixels, και το αντίστοιχο υψιπερατό φίλτρο υπολογίζει τη διαφορά μεταξύ γειτονικών pixels.

### 7.2.3 Χαρακτηριστικά Σχήματος

Το σχήμα ως χαρακτηριστικό των διαφόρων αντικειμένων συνδέεται συνήθως με την έννοια του περιγράμματος των αντικειμένων. Το περίγραμμα ενός αντικειμένου, από την άλλη, ορίζεται μέσω ενός συνόλου ακμών, για τις οποίες η πληροφορία που μας ενδιαφέρει είναι κυρίως η κατεύθυνσή τους. Λαμβάνοντας υπόψιν τα παραπάνω, γίνεται εύκολα αντιληπτό ότι ο υπολογισμός χαρακτηριστικών σχήματος από μια εικόνα θα εμπλέκει κάποιο μηχανισμό περιγραφής της τοπικής κατανομής των εμφανιζόμενων ακμών στο επίπεδο της εικόνας.

Στα πλαίσια των πειραμάτων αυτού του κεφαλαίου, ως χαρακτηριστικά σχήματος για τις εικόνες χρησιμοποιήσαμε εκείνα που προκύπτουν από τον υπολογισμό ενός Ιστογράμματος ακμών (edge histogram) [64]. Για μια εφαρμογή αυτών των χαρακτηριστικών σε προβλήματα ανάκτησης εικόνων, ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στην εργασία [65].

Σε αυτό το σημείο θα αναφερθούμε εν συντομία στον τρόπο υπολογισμού του Ιστογράμματος ακμών για μια εικόνα. Καταρχήν, απαιτείται πάλι η μετατροπή της εικόνας που θέλουμε να αναλύσουμε σε γκριζα. Μετά, η γκριζα εικόνα χωρίζεται σε 16 υποεικόνες, μέσω της διαίρεσης καθεμιάς από τις δύο διαστάσεις της (πλάτος και ύψος) σε 4 ίσα διαστήματα. Έπειτα, καθεμιά από τις παραπάνω υποεικόνες διαιρείται σε έναν αριθμό από μη επικαλυπτόμενα παράθυρα. Στα πειράματα αυτού του κεφαλαίου ο αριθμός αυτός επιλέχθηκε να είναι ίσος με 300, πράγμα που για τη βάση εικόνων που χρησιμοποιούμε (θα αναφερθούμε σε αυτήν αναλυτικότερα παρακάτω) συνεπάγεται ότι το κάθε παράθυρο έχει μέγεθος  $8 \times 8$  pixels. Κατόπιν, το καθένα από αυτά τα παράθυρα χωρίζεται και αυτό με τη σειρά του σε 4 υποπαράθυρα μεγέθους  $4 \times 4$ , για κάθε ένα από τα οποία υπολογίζεται η μέση τιμή της φωτεινότητας για τα pixels που του αντιστοιχούν. Συνεπώς, καταλήγουμε στην αναπαράσταση κάθε παραθύρου μέσω ενός  $2 \times 2$  πίνακα, έστω  $\mathbf{M}$ , με τιμές για τα στοιχεία του τις μέσες τιμές που υπολογίστηκαν πάνω στα αντίστοιχα υποπαράθυρα.

Μετά και από αυτό το βήμα, χρησιμοποιούμε 5 πίνακες – μάσκες, επίσης  $2 \times 2$ , καθέναν από τους οποίους αντιπροσωπεύει και από ένα είδος ακμής. Συγκεκριμένα, ισχύει:

$$\mathbf{h}_v = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad \mathbf{h}_h = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \quad \mathbf{h}_{45} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{bmatrix},$$

$$\mathbf{h}_{135} = \begin{bmatrix} 0 & \sqrt{2} \\ -\sqrt{2} & 0 \end{bmatrix}, \quad \mathbf{h}_{nd} = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} \quad (7.21)$$

όπου  $\mathbf{h}_v$ ,  $\mathbf{h}_h$ ,  $\mathbf{h}_{45}$ ,  $\mathbf{h}_{135}$  και  $\mathbf{h}_{nd}$  είναι οι μάσκες για την κάθετη, οριζόντια, διαγώνια 45 μοιρών, διαγώνια 135 μοιρών και μη κατευθυνόμενη ακμή, αντίστοιχα. Τα στοιχεία του πίνακα  $\mathbf{M}$  που ορίστηκε παραπάνω πολλαπλασιάζονται με τα αντίστοιχα στοιχεία καθενός

από τους πίνακες – μάσκες, τα γινόμενα αθροίζονται και η απόλυτη τιμή του αθροίσματος που προκύπτει, έστω  $e$ , χρησιμοποιείται ως ένδειξη ύπαρξης της αντίστοιχης ακμής στο παράθυρο που αναπαριστάνεται μέσω του πίνακα  $\mathbf{M}$ . Επομένως, για το  $e$  θα έχουμε

$$e = \left| \sum_{i=1}^2 \sum_{j=1}^2 h_{ij} M_{ij} \right| \quad (7.22)$$

όπου  $M_{ij}$  είναι τα στοιχεία του πίνακα  $\mathbf{M}$ , και  $h_{ij}$  είναι τα στοιχεία ενός από τους  $\mathbf{h}_v$ ,  $\mathbf{h}_h$ ,  $\mathbf{h}_{45}$ ,  $\mathbf{h}_{135}$  και  $\mathbf{h}_{nd}$ .

Τέλος, προσδιορίζεται η ακμή που αντιστοιχεί στη μέγιστη τιμή για το μέτρο  $e$ , και αν η αντίστοιχη τιμή για το  $e$  είναι μεγαλύτερη ενός κατωφλίου (που στα πλαίσια των πειραμάτων του παρόντος κεφαλαίου τέθηκε ίσο με 10, θεωρώντας ότι οι τιμές της φωτεινότητας για τα pixels της γκρίζας εικόνας κυμαίνονται μεταξύ του 0 και του 255), τότε θεωρούμε ότι το εξεταζόμενο παράθυρο περιέχει τη συγκεκριμένη ακμή.

Για κάθε, λοιπόν, είδος ακμής, από τα 5 που παρουσιάστηκαν παραπάνω, μπορούμε να προσδιορίσουμε τον αριθμό των παραθύρων κάθε υποεικόνας που ανήκουν σε αυτό. Κανονικοποιώντας μετέπειτα με το συνολικό πλήθος παραθύρων της υποεικόνας, παίρνουμε ένα είδος Ιστογράμματος των ακμών της υποεικόνας, το οποίο περιέχει 5 συχνότητες, με καθεμιά να αντιστοιχεί και σε ένα από τα είδη ακμών που υποστηρίζονται. Αν επαναλάβουμε αυτή τη διαδικασία για καθεμιά από τις 16 υποεικόνες της εικόνας που αναλύουμε, παίρνουμε τελικά 16 Ιστογράμματα ακμών των 5 bins το καθένα. Αυτά αποτελούν και τα χαρακτηριστικά, συνολικά 80, που χρησιμοποιήσαμε στα πειράματα.

## 7.2.4 Τελική Αναπαράσταση των Εικόνων

Για κάθε εικόνα της βάσης, λοιπόν, υπολογίζονται όλα τα παραπάνω χαρακτηριστικά, δηλαδή, ένα Ιστόγραμμα χρώματος στο χώρο HSV, 256 bins, 9 ροπές χρώματος στο χώρο LAB, 104 χαρακτηριστικά υφής μέσω του δενδρικά δομημένου μετασχηματισμού Wavelet τριών επιπέδων, και ένα Ιστόγραμμα ακμών με 80 bins.

Για να προκύψει μια ενιαία αναπαράσταση της εικόνας, τα παραπάνω χαρακτηριστικά συγχωνεύονται σε ένα διάνυσμα χαρακτηριστικών. Το διάνυσμα αυτό έχει  $256 + 9 + 104 + 80 = 449$  συνιστώσες, που είναι ένας αρκετά μεγάλος αριθμός, ώστε να έχει νόημα και ενδιαφέρον η εφαρμογή τεχνικών επιλογής χαρακτηριστικών για τον περιορισμό της διάστασης του αντίστοιχου χώρου χαρακτηριστικών σε κάθε γύρο της συσχετιστικής ανατροφοδότησης. Εξάλλου, είναι προφανές ότι ο συνολικός αριθμός των παραδειγμάτων (για όλους τους γύρους της ανατροφοδότησης) αναμένεται να είναι κατά πολύ μικρότερος από τη διάσταση του συγκεκριμένου χώρου χαρακτηριστικών.

Αυτά τα 449 χαρακτηριστικά που χρησιμοποιούνται έχουν εν γένει διαφορετικό εύρος τιμών το καθένα. Επομένως, για να εξασφαλιστεί η ισορροπημένη συνεισφορά των τιμών του κάθε χαρακτηριστικού στους όποιους υπολογισμούς γίνουν κατόπιν, μια αντίστοιχη διαδικασία με αυτήν που περιγράφηκε στην ενότητα 4.7.1.1, αλλά τώρα όχι για τα τοπικά διανύσματα χαρακτηριστικών μιας εικόνας αλλά για τις διανυσματικές αναπαραστάσεις όλων των εικόνων, απαιτείται για την κανονικοποίηση των τιμών των χαρακτηριστικών, έτσι ώστε για το σύνολο των εικόνων να έχουν μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση. Μέσω αυτού του μετασχηματισμού δε μεταβάλλεται προφανώς η διάσταση του χώρου αναπαράστασης των εικόνων. Μόνο το εύρος τιμών των διάφορων χαρακτηριστικών τροποποιείται, προκειμένου να υπάρχει μεταξύ τους μέτρο σύγκρισης.

## 7.3 Πειράματα

Σε αυτήν την ενότητα περιγράφονται οι λεπτομέρειες και τα αποτελέσματα των πειραμάτων που έγιναν πάνω στην ενσωμάτωση στη διαδικασία της συσχετιστικής ανατροφοδότησης των μεθόδων επιλογής χαρακτηριστικών που παρουσιάσαμε στην ενότητα 7.1.2.

Για τα πειράματα αυτά χρησιμοποιήθηκε η βάση εικόνων MSDB. Τη συγκεκριμένη βάση εικόνων την έχουμε συναντήσει τόσο στα πειράματα του κεφαλαίου 4, όσο και σε αυτά του κεφαλαίου 6. Υπενθυμίζουμε εδώ ότι περιέχει 3740 εικόνες που κατανέμονται σε 17 σημασιολογικές κατηγορίες. Επίσης, στην πρώτη ενότητα του παραρτήματος Β παραθέτουμε πιο αναλυτικές πληροφορίες σχετικά με τα περιεχόμενά της.

Τα αποτελέσματα των πειραμάτων που ακολουθούν, τα οποία μας δίνουν μια ποσοτική εκτίμηση της απόδοσης των συγκρινόμενων μεθόδων, προκύπτουν, όπως και όλα τα αποτελέσματα που παρουσιάζονται στην παρούσα διατριβή, μέσω της εκτέλεσης κατάλληλων σεναρίων προσομοίωσης της διαδικασίας της συσχετιστικής ανατροφοδότησης. Οι προσομοιώσεις που εκτελέστηκαν στα πλαίσια αυτού του κεφαλαίου δε διαφοροποιούνται σημαντικά από αυτές που παρουσιάστηκαν στα προηγούμενα κεφάλαια.

Συγκεκριμένα, περίπου 1000 εικόνες από τη βάση MSDB επιλέχθηκαν με τυχαίο τρόπο. Στη συνέχεια, για κάθε πείραμα, αυτές οι εικόνες παρουσιάζονται διαδοχικά ως αρχικές ερωτήσεις στο σύστημα ανάκτησης εικόνων. Για κάθε μια αρχική ερώτηση ακολουθεί προσομοίωση 6 γύρων συσχετιστικής ανατροφοδότησης. Σε κάθε έναν από αυτούς τους γύρους, το πολύ 3 εικόνες από την ίδια κατηγορία με την αρχική ερώτηση, και το πολύ 3 εικόνες από άλλες κατηγορίες επιλέγονται από τις 50 πρώτες εικόνες (τις θεωρούμενες ως πιο σχετικές) της τρέχουσας διάταξης (τυχαία, στην περίπτωση που υπάρχουν παραπάνω από 3) και παρουσιάζονται στο σύστημα ως θετικά και αρνητικά παραδείγματα, αντίστοιχα. Αυτά τα παραδείγματα, σε συνδυασμό και με όλα τα άλλα που έχουν ανατροφοδοτηθεί σε προηγούμενους γύρους, χρησιμοποιούνται ως παραδείγματα εκπαίδευσης ενός ταξινομητή SVM για το διαχωρισμό των σχετικών από τις μη σχετικές εικόνες. Στην περίπτωση που γίνεται χρήση κάποιας μεθόδου επιλογής χαρακτηριστικών, της εκπαίδευσης του ταξινομητή προηγείται η επιλογή ενός αριθμού,  $K$ , χαρακτηριστικών από το σύνολο των 449 χαρακτηριστικών των περιγραφών των εικόνων. Η διάταξη των εικόνων της βάσης γίνεται σύμφωνα με την τιμή της συνάρτησης απόφασης του SVM, πάνω στη λογική που αναλύσαμε στο κεφάλαιο 5, και συγκεκριμένα στην ενότητα 5.4.1. Μόνο η αρχική διάταξη των εικόνων, πριν την έναρξη των γύρων της ανατροφοδότησης, και επομένως πριν από την υποβολή παραδειγμάτων ανατροφοδότησης που θα επέτρεπαν τη χρήση της όποιας μεθόδου επιλογής χαρακτηριστικών και την εκπαίδευση του ταξινομητή, γίνεται με βάση την Ευκλείδεια απόσταση των εικόνων από την αρχική ερώτηση, όπως αυτή υπολογίζεται στον αρχικό χώρο χαρακτηριστικών.

Ως μέτρο αξιολόγησης των αποτελεσμάτων της ανάκτησης σε κάθε γύρο χρησιμοποιείται η ακρίβεια, δηλαδή το ποσοστό των πραγματικά σχετικών εικόνων, στις πρώτες  $T = 20$  εικόνες της διάταξης. Επίσης, όπως έχουμε αναφέρει πολλάκις, πραγματικά σχετικές θεωρούνται οι εικόνες που ανήκουν στην ίδια κατηγορία με αυτήν της αρχικής ερώτησης. Κατόπιν, για να λάβουμε ένα συνολικό μέτρο της απόδοσης της κάθε μεθόδου, υπολογίζουμε το μέσο όρο της ακρίβειας για όλες τις αρχικές ερωτήσεις και για κάθε γύρο της ανατροφοδότησης.

Για την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης χρησιμοποιήθηκαν σε κάθε γύρο ταξινομητές SVM με συνάρτηση πυρήνα τη Γκαουσιανή συνάρτηση ακτινικής βάσης,  $k(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|^2)$ . Με δεδομένη τη χρήση του

συγκεκριμένου μοντέλου μάθησης και της συγκεκριμένης συνάρτησης πυρήνα, για την εκπαίδευση του πρώτου απαιτείται ο καθορισμός δύο παραμέτρων, της παραμέτρου εκπαίδευσης  $C$  και της παραμέτρου της συνάρτησης πυρήνα  $\gamma$ . Για μια σύντομη αλλά περιεκτική παρουσίαση όσο αφορά αυτές τις παραμέτρους, και γενικά τα SVMs, μπορεί κανείς να ανατρέξει στο κεφάλαιο 5, και συγκεκριμένα στην ενότητα 5.3.

Όπως και στα αντίστοιχα πειράματα του κεφαλαίου 6, παρατηρήσαμε και εδώ ότι η απόδοση του ταξινομητή δεν είναι ιδιαίτερα ευαίσθητη όσο αφορά την τιμή της παραμέτρου  $C$ . Με βάση αυτές μας τις παρατηρήσεις και ένα αριθμό προκαταρκτικών πειραμάτων που πραγματοποιήσαμε, καταλήξαμε πάλι σε μια επιλογή  $C=100$  για τη συγκεκριμένη παράμετρο. Αυτή η τιμή χρησιμοποιήθηκε σε όλα τα πειράματα του παρόντος κεφαλαίου, ακόμα και για την εκπαίδευση των γραμμικών SVMs που χρησιμοποιεί η μέθοδος επιλογής χαρακτηριστικών SVM-RFE. Παρεμπιπτόντως, επισημαίνουμε ότι λόγω της συνάρτησης πυρήνα που αντιστοιχεί στα SVMs της μεθόδου SVM-RFE, εκτός της παραμέτρου  $C$  δεν υπάρχει κάποια άλλη παράμετρος που να τα αφορά και να πρέπει να καθοριστεί.

Όσο αφορά τώρα την παράμετρο  $\gamma$  της Γκαουσιανής συνάρτησης ακτινικής βάσης, τα πράγματα είναι λίγο πιο δύσκολα, όπως επισημάναμε και στο κεφάλαιο 6. Η τιμή αυτής της παραμέτρου μπορεί να επηρεάσει σημαντικά τα αποτελέσματα που θα προκύψουν από τη μετέπειτα χρήση του SVM. Παρόλα αυτά, η τιμή που μας δίνει τα καλύτερα αποτελέσματα δείχνει να εξαρτάται από τις τιμές που παίρνει το τετράγωνο της Ευκλείδειας απόστασης,  $\|\mathbf{x}_1 - \mathbf{x}_2\|^2$ , για τα πρότυπα του προβλήματος προς επίλυση (εδώ, για τις διανυσματικές αναπαραστάσεις των εικόνων), και συγκεκριμένα είναι πάντα κοντά στο αντίστροφο της μέσης τιμής αυτής της ποσότητας. Επομένως, τα συμπεράσματά μας είναι παρεμφερή με αυτά που προκύπτουν μέσω της εργασίας [66]. Βέβαια, ανάλογα με τα χαρακτηριστικά που επιλέγονται για να χρησιμοποιηθούν, και κυρίως ανάλογα με τον αριθμό των τελευταίων, η παραπάνω μέση τιμή μεταβάλλεται, λόγω της μεταβολής του χώρου αναπαράστασης των εικόνων. Πάντως, από την προηγούμενη ανάλυση προκύπτει ότι δεν είναι τόσο δύσκολο να βρούμε μια ικανοποιητική τιμή για την παράμετρο  $\gamma$ . Παρόλα αυτά, και στα πειράματα αυτού του κεφαλαίου η τιμή της προσδιορίστηκε εμπειρικά για κάθε πείραμα, έτσι ώστε να προκύπτουν τα κατά το δυνατόν καλύτερα αποτελέσματα.

Τέλος, πρέπει να αναφέρουμε ότι ως προγραμματιστική υλοποίηση του αλγορίθμου μάθησης των SVMs χρησιμοποιήσαμε και αυτή τη φορά την παρεχόμενη μέσω του δικτυακού τόπου [67]. Επίσης, για μια υλοποίηση της μεθόδου mRMR καταφύγαμε στο δικτυακό τόπο [68].

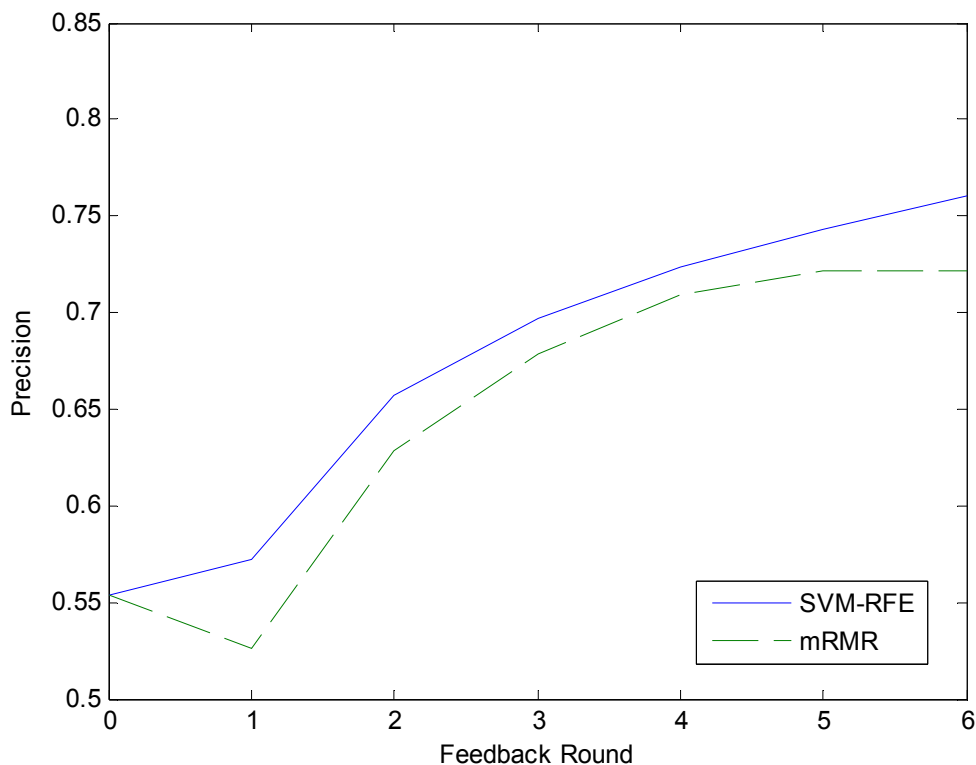
### 7.3.1 Αποτελέσματα

Στην τρέχουσα ενότητα θα παρουσιάσουμε και θα σχολιάσουμε τα αποτελέσματα των πειραμάτων που εκτελέστηκαν στα πλαίσια του παρόντος κεφαλαίου. Συγκεκριμένα, τα αποτελέσματα αυτά δίνονται στα σχήματα 7.1, 7.2, 7.3 και 7.4.

Στα σχήματα 7.1, 7.2 και 7.3 συγκρίνονται οι μέθοδοι επιλογής χαρακτηριστικών που παρουσιάστηκαν στην ενότητα 7.1.2, με βάση τα αποτελέσματα που προκύπτουν κατά την ενσωμάτωσή τους στη διαδικασία της συσχετιστικής ανατροφοδότησης, όπως αυτά ποσοτικοποιούνται μέσω της εξέλιξης της μέσης τιμής του μέτρου της ακρίβειας, για εμβέλεια  $T=20$ . Ο αριθμός των χαρακτηριστικών που επιλέγονται σε κάθε γύρο της ανατροφοδότησης δίνεται ως παράμετρος, και ισούται με 25, 50 και 75 για τα σχήματα 7.1, 7.2 και 7.3, αντίστοιχα. Όπως μπορεί κανείς εύκολα να παρατηρήσει, η χρήση της μεθόδου επιλογής χαρακτηριστικών SVM-RFE οδηγεί σε σταθερά καλύτερα αποτελέσματα σε σχέση με αυτά που προκύπτουν μέσω της μεθόδου mRMR. Επιπροσθέτως, αξίζει εδώ να

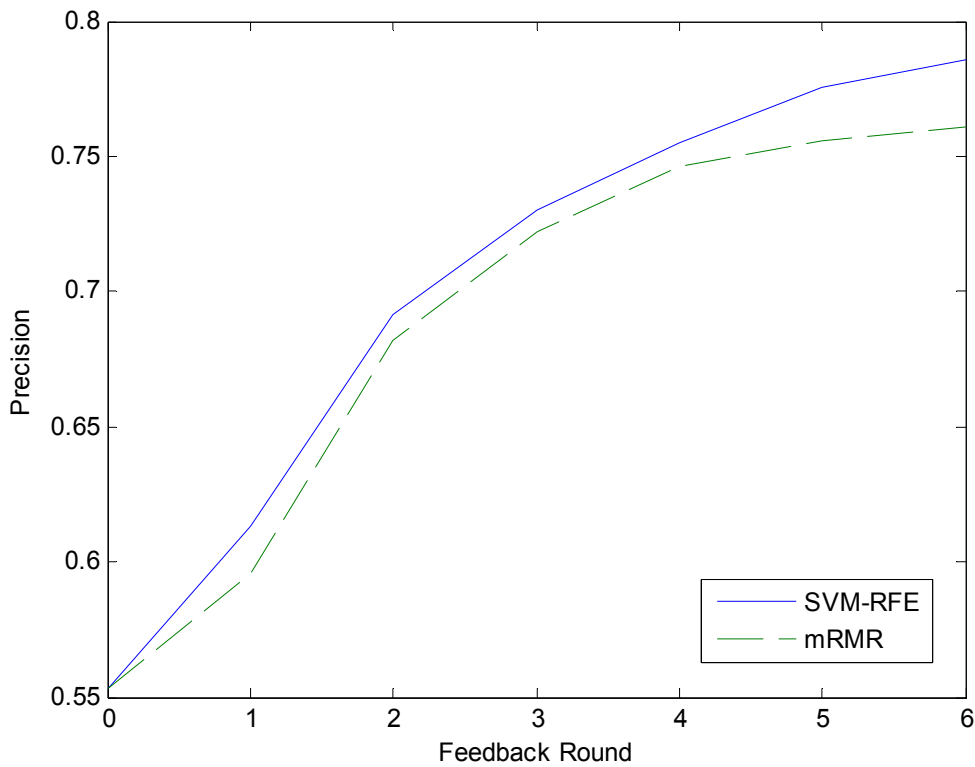
αναφερθεί ότι και από υπολογιστικής απόψεως η ενσωμάτωση στη διαδικασία της συσχετιστικής ανατροφοδότησης της μεθόδου SVM-RFE μας δίνει ένα κατά πολύ ταχύτερο σύστημα ανάκτησης σε σχέση με αυτό που προκύπτει μέσω της μεθόδου mRMR. Συγκεκριμένα, για  $K = 50$ , δηλαδή όταν σε κάθε γύρο της ανατροφοδότησης επιλέγονται 50 χαρακτηριστικά, ο μέσος χρόνος που απαιτείται για να εκτελεστούν οι 6 γύροι της ανατροφοδότησης για μια αρχική ερώτηση είναι περίπου 2.5 δευτερόλεπτα όταν ως μέθοδος επιλογής χαρακτηριστικών χρησιμοποιείται η SVM-RFE, ενώ ανέρχεται στα 15 περίπου δευτερόλεπτα αν χρησιμοποιήσουμε στη θέση της την mRMR. Οι μετρήσεις αυτές έχουν γίνει με χρήση ενός προσωπικού υπολογιστή με συχνότητα επεξεργαστή 3GHz.

Στο σχήμα 4, τώρα, παρουσιάζεται η μέση ακρίβεια (σε εμβέλεια  $T = 20$ , πάλι) που επιτυγχάνεται για διάφορες τιμές της παραμέτρου  $K$  (για παράδειγμα, 10, 25, 50, 75, 100, 200 κτλ.), δηλαδή του αριθμού των χαρακτηριστικών που επιλέγονται, μετά τον έκτο και τελευταίο γύρο της ανατροφοδότησης, όταν ως μέθοδος επιλογής χαρακτηριστικών χρησιμοποιείται η SVM-RFE, και αντιπαραβάλλεται με την αντίστοιχη ακρίβεια που επιτυγχάνεται όταν δε χρησιμοποιείται καμιά μέθοδος επιλογής χαρακτηριστικών, αλλά ο ταξινομητής εκπαιδεύεται πάντα στον αρχικό χώρο χαρακτηριστικών. Και πάλι μπορούμε εύκολα να παρατηρήσουμε ότι με χρήση σχετικά λίγων χαρακτηριστικών από το σύνολο των αρχικών, επιτυγχάνουμε ίση ή και μεγαλύτερη απόδοση σε σχέση με αυτήν που προκύπτει χρησιμοποιώντας όλα τα χαρακτηριστικά. Συγκεκριμένα, για μικρές τιμές του  $K$  η ακρίβεια αυξάνει κατακόρυφα, και πολύ γρήγορα, για  $K \approx 50$ , έχει ήδη εξισωθεί με αυτήν που παίρνουμε όταν και τα 449 χαρακτηριστικά χρησιμοποιούνται για την εκπαίδευση του SVM και τη διάταξη των εικόνων της βάσης.

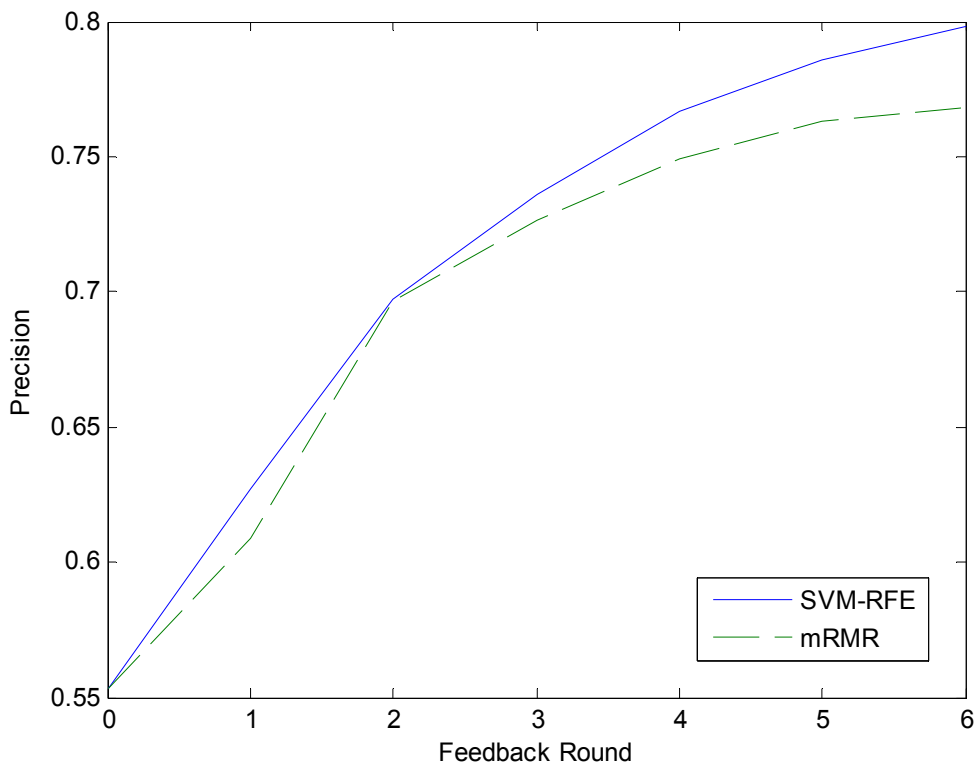


**Σχήμα 7.1:** Σύγκριση μεθόδων επιλογής χαρακτηριστικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια  $T = 20$ , για  $K = 25$  επιλεγμένα χαρακτηριστικά

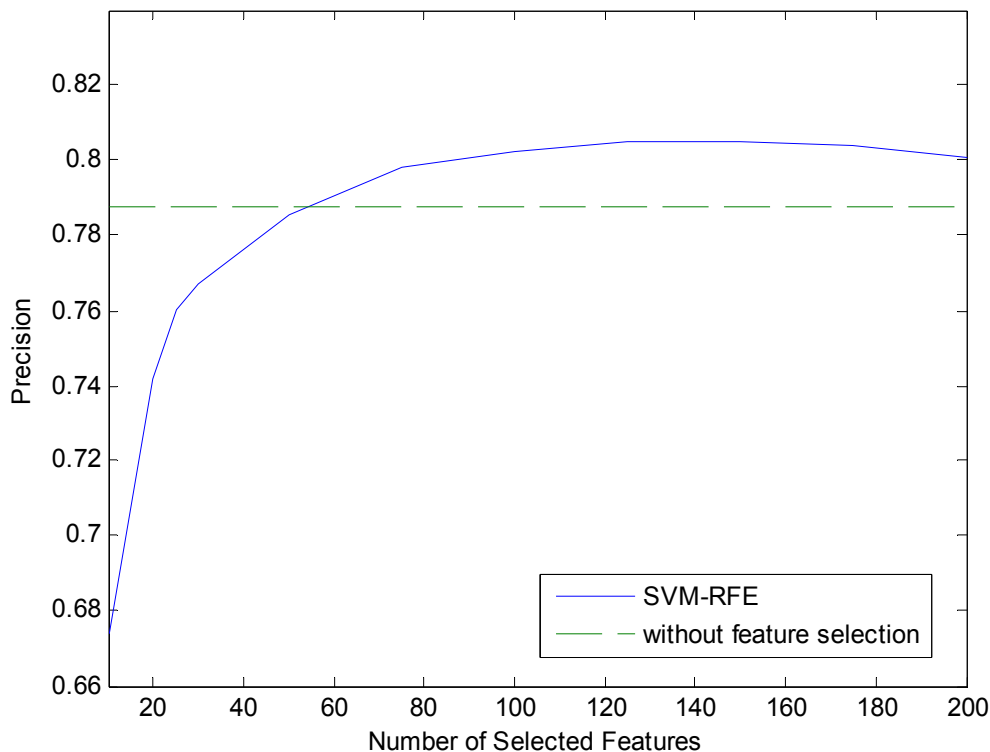




**Σχήμα 7.2:** Σύγκριση μεθόδων επιλογής χαρακτηριστικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια  $T = 20$ , για  $K = 50$  επιλεγμένα χαρακτηριστικά



**Σχήμα 7.3:** Σύγκριση μεθόδων επιλογής χαρακτηριστικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια  $T = 20$ , για  $K = 75$  επιλεγμένα χαρακτηριστικά



**Σχήμα 7.4:** Σύγκριση μεταξύ χρήσης και μη χρήσης μεθόδου επιλογής χαρακτηριστικών στο πλαίσιο της συσχετιστικής ανατροφοδότησης με χρήση ταξινομητών – Μέση ακρίβεια σε εμβέλεια  $T = 20$ , μετά τον 6<sup>ο</sup> γύρο της ανατροφοδότησης

## 7.4 Συμπεράσματα

Σε αυτό το κεφάλαιο μελετήθηκε η χρήση μεθόδων επιλογής χαρακτηριστικών σε κάθε γύρο της συσχετιστικής ανατροφοδότησης, στα πλαίσια του προβλήματος της ανάκτησης εικόνων. Το ζήτημα της συσχετιστικής ανατροφοδότησης αντιμετωπίστηκε θεωρούμενο σε κάθε γύρο ως πρόβλημα δυαδικής ταξινόμησης μεταξύ της κατηγορίας των σχετικών και αυτής των μη σχετικών εικόνων, για τη λύση του οποίου μπορεί να χρησιμοποιηθεί κάποιος ταξινομητής. Αυτή η θεώρηση επιτρέπει και οδηγεί στη με φυσικό τρόπο ενσωμάτωση μεθόδων επιλογής χαρακτηριστικών πριν από την εκπαίδευση του ταξινομητή, με σκοπό τη μείωση της διάστασης των αναπαραστάσεων των εικόνων, την απόρριψη πλεοναζόντων χαρακτηριστικών ή χαρακτηριστικών με μικρή διακριτική ικανότητα μεταξύ των κατηγοριών των σχετικών και μη σχετικών εικόνων, και εν τέλει την αποτελεσματικότερη και αποδοτικότερη εκπαίδευση του ταξινομητή, και την ταχύτερη ανανέωση της εκτίμησης του συστήματος σχετικά με τη συνάφεια μεταξύ εικόνων της βάσης και ερώτησης του χρήστη.

Σε αυτά τα πλαίσια, μια μέθοδος που συνδυάζει το εξαιρετικά δημοφιλές μοντέλο μάθησης των SVMs με μια πρόσφατα προταθείσα τεχνική επιλογής χαρακτηριστικών, που αποκαλείται SVM-RFE και χρησιμοποιεί και η ίδια ταξινομητές SVM, προτάθηκε και αξιολογήθηκε όσο αφορά τα αποτελέσματα που προκύπτουν από τη χρήση της στους διαδοχικούς γύρους της συσχετιστικής ανατροφοδότησης. Για λόγους σύγκρισης, συνδυάστηκε με τα SVMs, στην ίδια λογική, και μια άλλη πολύ δημοφιλής μέθοδος επιλογής χαρακτηριστικών, η mRMR, και αξιολογήθηκε παράλληλα με την προηγούμενη.

Από τα πειράματα που διεξήχθησαν, και τα οποία παρουσιάστηκαν προηγουμένως, προκύπτει ότι η προτεινόμενη μέθοδος πλεονεκτεί και από άποψη αποτελεσμάτων και από άποψη ταχύτητας σε σχέση με την παραλλαγή που βασίζεται στην mRMR. Επιπλέον, από τα παραπάνω πειράματα καταδείχτηκε ότι είναι δυνατόν με μια πολύ μεγάλη μείωση στον αριθμό των χρησιμοποιούμενων χαρακτηριστικών να πάρουμε αποτελέσματα εξίσου καλά ή και ακόμα καλύτερα σε σχέση με αυτά που προκύπτουν χρησιμοποιώντας το αρχικό και εξαιρετικά μεγάλης διάστασης σύνολο χαρακτηριστικών.



## Κεφάλαιο 8

### Συνολικό Πόρισμα Διατριβής

#### 8.1 Γενικά Συμπεράσματα

Το αντικείμενο της έρευνας που μας οδήγησε στην παρούσα διατριβή είναι η χρήση τεχνικών μηχανικής μάθησης για την αντιμετώπιση του προβλήματος της ανάκτησης εικόνων. Συγκεκριμένα, εστιάσαμε πάνω στο ζήτημα της ανάκτησης εικόνων με βάση το περιεχόμενο, και η κύρια τεχνική με την οποία ασχοληθήκαμε είναι αυτή της συσχετιστικής ανατροφοδότησης, που μπορεί να χρησιμοποιηθεί για τη διαδοχική βελτίωση των αποτελεσμάτων που παρέχονται από τα συστήματα ανάκτησης εικόνων, μέσω της ενσωμάτωσης στο σύστημα της υποκειμενικής κρίσης του χρήστη. Με την παρούσα διατριβή επιχειρήθηκε μια όσο το δυνατόν πιο πλήρης και παράλληλα συνοπτική παρουσίαση των αποτελεσμάτων της ερευνητικής δραστηριότητας που διεξήχθη.

Η μελέτη της εκτενούς βιβλιογραφίας πάνω στο θέμα της ανάκτησης εικόνων με βάση το περιεχόμενο αποκαλύπτει αμέσως τα προβλήματα που απορρέουν από την ύπαρξη του σημασιολογικού κενού μεταξύ των περιγραφών που χρησιμοποιούνται για τις εικόνες, οι οποίες βασίζονται στον υπολογισμό χαμηλού επιπέδου οπτικών χαρακτηριστικών, και των υψηλού επιπέδου σημασιολογικών κριτηρίων που ακολουθούν κατά την αναζήτησή τους οι χρήστες. Η τεχνική της συσχετιστικής ανατροφοδότησης, που έχει προταθεί για την κάλυψη αυτού του σημασιολογικού κενού, παρέχει μεγάλες δυνατότητες βελτίωσης των αποτελεσμάτων των συστημάτων ανάκτησης εικόνων, κάτι το οποίο αποδείχτηκε περίτρανα και στα πλαίσια της παρούσας διατριβής για τις διάφορες μεθοδολογίες που υλοποιήθηκαν και μελετήθηκαν. Όπως αναλύθηκε και στο πρώτο κεφάλαιο, δύο είναι οι κύριες προσεγγίσεις όσο αφορά την υλοποίηση της τεχνικής της συσχετιστικής ανατροφοδότησης. Η πρώτη αφορά τη χρήση μιας αναπαράστασης για την ερώτηση του χρήστη, και την τροποποίησή της κάθε φορά σύμφωνα με τα παραδείγματα που παρέχει ο τελευταίος. Αντίθετα, η δεύτερη προσέγγιση έχει να κάνει με την υιοθέτηση και εκπαίδευση κάποιου ταξινομητή σε κάθε γύρο της ανατροφοδότησης, προκειμένου μέσω αυτού να διαχωριστεί η κατηγορία των σχετικών με την ερώτηση του χρήστη εικόνων από αυτήν των μη σχετικών εικόνων.

Επιπλέον, η μελέτη της βιβλιογραφίας αναδεικνύει και δύο δυνατότητες που υπάρχουν όσο αφορά τον τρόπο αναπαράστασης των εικόνων που θα χρησιμοποιηθεί στα πλαίσια ενός συστήματος ανάκτησης. Η πρώτη δυνατότητα που έχουμε είναι να χρησιμοποιήσουμε μια συνήθως ιδιαίτερα μεγάλης διάστασης διανυσματική αναπαράσταση για την κάθε εικόνα. Από την άλλη, στηριζόμενοι στον σε τοπικό επίπεδο υπολογισμό ενός συνόλου διανυσμάτων χαρακτηριστικών από την εικόνα, μπορούμε να καταλήξουμε σε μια ενιαία πιθανοτική περιγραφή της κατανομής των τελευταίων, χρησιμοποιώντας ένα συνεχές

πιθανοτικό μοντέλο για αυτόν το σκοπό. Στην περίπτωση μιας τέτοιας αναπαράστασης για τις εικόνες, το κύριο θέμα προς αντιμετώπιση είναι ο ορισμός κατάλληλων μέτρων σύγκρισης μεταξύ των μοντέλων, που να μπορούν να υπολογιστούν αποδοτικά, με δεδομένη την απαίτηση για γρήγορη αλληλεπίδραση μεταξύ συστήματος ανάκτησης και χρήστη. Στα πλαίσια της έρευνάς μας, όσο αφορά τα πιθανοτικά μοντέλα εστίασαμε στα Γκαουσιανά Μείγματα, οι ιδιότητες των οποίων τα καθιστούν ιδιαίτερα δημοφιλή για τη μοντελοποίηση συνόλων δεδομένων. Υπό αυτό το πρίσμα, στο τρίτο κεφάλαιο αναλύονται διάφορα ενδιαφέροντα μέτρα απόστασης μεταξύ Γκαουσιανών Μειγμάτων. Ιδιαίτερη έμφαση δόθηκε στη σε βάθος ανάλυση των ιδιοτήτων ενός από τα μέτρα αυτά, και συγκεκριμένα της απόκλισης C2.

Κατόπιν, μελετήθηκαν εκτενώς οι δυνατότητες χρήσης των Γκαουσιανών Μειγμάτων ως αναπαράσταση για τις εικόνες, στην περίπτωση που αυτές θα τις χειριστεί ένα σύστημα ανάκτησης που χρησιμοποιεί συσχετιστική ανατροφοδότηση. Σε αυτό το πλαίσιο, δύο νέες μέθοδοι συσχετιστικής ανατροφοδότησης προτάθηκαν. Η πρώτη από αυτές ανήκει στην κατηγορία των μεθόδων που βασίζονται στην τροποποίηση της αναπαράστασης της ερώτησης, και παρουσιάστηκε στο τέταρτο κεφάλαιο. Ο ορισμός του νέου μοντέλου της ερώτησης ακολουθεί ένα συμβατό με τη μοντελοποίηση των εικόνων πιθανοτικό πλαίσιο. Επίσης, με βάση τον τρόπο ανανέωσης της ερώτησης και τις κομψές ιδιότητες της απόκλισης C2, που χρησιμοποιείται ως μέτρο απόστασης μεταξύ Γκαουσιανών Μειγμάτων, επιτυγχάνεται η ταχύτατη ανανέωση της εκτίμησης του συστήματος όσο αφορά τη συνάφεια των εικόνων της βάσης με την ερώτηση του χρήστη, μετά την υποβολή των παραδειγμάτων από τον τελευταίο. Ακόμα, μέσω των ιδιοτήτων των Γκαουσιανών Μειγμάτων καθίσταται δυνατή η αποτελεσματική ενσωμάτωση των αρνητικών παραδειγμάτων και η εκμετάλλευση της πληροφορίας που προέρχεται από αυτά.

Ακολούθως, στο έκτο κεφάλαιο, υποθέτοντας ξανά μοντελοποίηση των εικόνων μέσω Γκαουσιανών Μειγμάτων, προτείνεται μια μέθοδος συσχετιστικής ανατροφοδότησης της δεύτερης κατηγορίας, που χρησιμοποιεί ταξινομητές SVM σε κάθε γύρο της ανατροφοδότησης. Για να καταστεί δυνατή η χρήση Γκαουσιανών Μειγμάτων στο πλαίσιο των SVMs, που αποτελούν ένα από τα πλέον δημοφιλή μοντέλα μάθησης, χρειάστηκε η υιοθέτηση μιας κατάλληλης συνάρτησης πυρήνα που ορίζεται μέσω αποδοτικά υπολογιζόμενων μέτρων απόστασης μεταξύ Γκαουσιανών Μειγμάτων. Επίσης, για την εκπαίδευση των Γκαουσιανών Μειγμάτων για τις εικόνες χρησιμοποιείται ο πολύ ενδιαφέρον αλγόριθμος MAP-EM, που βασίζεται στις αρχές της Μπεϋζιανής μάθησης. Η χρήση του συγκεκριμένου αλγορίθμου, πέρα από τη σημαντική βελτίωση των αποτελεσμάτων σε σχέση με τον κλασικό τρόπο εκπαίδευσης, ειδικά όταν το διαθέσιμο σύνολο εκπαίδευσης είναι μικρό, έχει ως αποτέλεσμα και τη δυνατότητα περαιτέρω επιτάχυνσης της διαδικασίας σύγκρισης μεταξύ Γκαουσιανών Μειγμάτων.

Μέσω κατάλληλα σχεδιασμένων και εκτελεσμένων πειραμάτων, κατέστη δυνατή η αξιολόγηση των παραπάνω μεθόδων και η σύγκρισή τους με άλλες αντιπροσωπευτικές μεθόδους της αντίστοιχης κατηγορίας. Τα αποτελέσματα των συγκρίσεων αποδεικνύουν, εκτός από την επιτυχία και την υπεροχή των προτεινόμενων μεθόδων συσχετιστικής ανατροφοδότησης αυτών καθ'αυτών, τα πλεονεκτήματα της μοντελοποίησης των εικόνων μέσω Γκαουσιανών Μειγμάτων, αφού αυτού του είδους η αναπαράσταση των εικόνων οδηγεί σχεδόν πάντα σε συστήματα ανάκτησης που χαρακτηρίζονται από υψηλότερες επιδόσεις σε σχέση με τα αντίστοιχα συστήματα που χρησιμοποιούν κλασική διανυσματική αναπαράσταση για τις εικόνες.

Τέλος, στο έβδομο κεφάλαιο μελετήθηκε η χρήση μεθόδων επιλογής χαρακτηριστικών σε κάθε γύρο της συσχετιστικής ανατροφοδότησης, με την τελευταία να υλοποιείται πάλι μέσω εκπαίδευσης ταξινομητών. Σε αυτό το πλαίσιο, ως αναπαράσταση για τις εικόνες χρησιμοποιούνται πολυδιάστατα διανύσματα χαρακτηριστικών, που περιλαμβάνουν

διαφόρων ειδών οπτική πληροφορία. Ο τελευταίος τρόπος αναπαράστασης των εικόνων, παρότι κλασικός, εμφανίζει διάφορα μειονεκτήματα. Εξάλλου, αυτός ήταν ο λόγος που ασχοληθήκαμε τόσο πολύ με την αναπαράσταση των εικόνων μέσω συνεχών μοντέλων, με τα θετικά αποτελέσματα που περιγράψαμε παραπάνω. Ένα από τα κύρια μειονεκτήματα της προσέγγισης της διανυσματικής αναπαράστασης αφορά τη συνήθως μεγάλη διάσταση των αναπαραστάσεων των εικόνων, η οποία σε συνδυασμό με το μικρό αριθμό των διαθέσιμων παραδειγμάτων μπορεί να δυσκολέψει πολύ την εκπαίδευση του ταξινομητή σε κάθε γύρο της ανατροφοδότησης. Έτσι, μια διαδικασία επιλογής χαρακτηριστικών μπορεί να χρησιμοποιηθεί για τη μείωση της διάστασης των αναπαραστάσεων, με στόχο τη βελτίωση της κατάστασης. Ως αποτέλεσμα της σχετικής ερευνητικής δραστηριότητάς μας, προτάθηκε η χρήση μιας βασισμένης στα SVMs μεθόδου επιλογής χαρακτηριστικών, που συνδυαζόμενη με ομοειδείς ταξινομητές για την υλοποίηση της συσχετιστικής ανατροφοδότησης φαίνεται να οδηγεί σε σαφή πλεονεκτήματα τόσο σε σχέση με μια άλλη πολύ δημοφιλή ανάλογη τεχνική, όσο και σε σχέση με την περίπτωση κατά την οποία δε χρησιμοποιείται καμιά μέθοδος επιλογής χαρακτηριστικών.

## 8.2 Μελλοντικές Επεκτάσεις

Η έρευνα που έχει διεξαχθεί στα πλαίσια της παρούσας διατριβής με κανέναν τρόπο δε μπορεί να θεωρηθεί ότι εξαντλεί το αντίστοιχο ερευνητικό πεδίο. Αντίθετα, το τελευταίο είναι πάρα πολύ ευρύ, με αποτέλεσμα να παρέχει πολλές κατευθύνσεις για μελλοντική ερευνητική προσπάθεια, χρησιμοποιώντας ως σημείο εκκίνησης τη γνώση που αποκτήθηκε ως τώρα.

Ξεκινώντας από τα Γκαουσιανά Μείγματα, τα οποία χρησιμοποιήθηκαν σε μεγάλη έκταση για την αναπαράσταση των εικόνων κατά τη μέχρι αυτή τη στιγμή διεξαχθείσα έρευνα, ένα σημείο που χρήζει περαιτέρω διερεύνησης έχει να κάνει με τον προσδιορισμό του αριθμού των συνιστωσών που θα περιέχονται στο κάθε μείγμα. Στα πλαίσια της παρούσας διατριβής δεν έγινε εκτενής μελέτη πάνω στις διάφορες μεθόδους που έχουν προταθεί για το σκοπό αυτόν. Πάντως, από μια προκαταρκτική εξέταση διαπιστώσαμε ότι τα συνήθως χρησιμοποιούμενα κριτήρια για την αντιμετώπιση του προβλήματος του καθορισμού της πολυπλοκότητας ενός μοντέλου δεν αποδίδουν ικανοποιητικά αποτελέσματα, τουλάχιστον στην περίπτωση των Γκαουσιανών Μειγμάτων που προκύπτουν κατά τη μοντελοποίηση των οπτικών χαρακτηριστικών των εικόνων.

Επίσης, στη διατριβή αυτή, για τη μοντελοποίηση των εικόνων μέσω συνεχών πιθανοτικών μοντέλων βασιστήκαμε σχεδόν αποκλειστικά στη χρήση Γκαουσιανών Μειγμάτων. Σε αυτή μας την απόφαση καταλυτικό ρόλο έπαιξαν και οι καλές μαθηματικές ιδιότητες των μοντέλων αυτών. Τα ιδιαίτερος θετικά αποτελέσματα που προέκυψαν, πέρα από απόδειξη της αξίας των Γκαουσιανών Μειγμάτων για την αναπαράσταση των εικόνων, μπορούν να θεωρηθούν και ως μια γενική ένδειξη υπέρ των συνεχών μοντέλων. Θα άξιζε επομένως τον κόπο να καταβληθεί προσπάθεια στο ίδιο πνεύμα και όσο αφορά άλλα ανάλογα μοντέλα. Έχουμε, εξάλλου, αναφέρει στο δεύτερο κεφάλαιο κάποια μειονεκτήματα που συνδέονται με τη Γκαουσιανή κατανομή, που αποτελεί τη συνιστώσα κατανομή των Γκαουσιανών Μειγμάτων, και κάποια μοντέλα που έχουν προταθεί για την αντιμετώπιση αυτών των μειονεκτημάτων. Έτσι, για παράδειγμα, μπορεί να διερευνηθεί η δυνατότητα χρήσης μειγμάτων  $t$  κατανομών στη θέση των Γκαουσιανών Μειγμάτων.

Επιπλέον, όσο αφορά το πρόβλημα της επιλογής χαρακτηριστικών, όπως αυτό τέθηκε στο έβδομο κεφάλαιο, η προσέγγισή του μόνο εισαγωγική μπορεί να θεωρηθεί. Απαιτείται περισσότερη ερευνητική προσπάθεια, αρχικά για τη μελέτη και αξιολόγηση και άλλων μεθόδων επιλογής χαρακτηριστικών. Ακόμα, ένα σημαντικό θέμα είναι ο προσδιορισμός

κάθε φορά του βέλτιστου αριθμού όσο αφορά τα χαρακτηριστικά που θα διατηρηθούν, με βάση το τρέχον σύνολο εκπαίδευσης. Επίσης, μια ενδιαφέρουσα προοπτική έχει να κάνει με τη χρήση μεθόδων επιλογής χαρακτηριστικών σε συνδυασμό με τη χρήση Γκαουσιανών Μειγμάτων εκπαιδευμένων πάνω σε σχετικά μεγάλης διάστασης τοπικά υπολογιζόμενα διανύσματα χαρακτηριστικών, όπως αυτά που χρησιμοποιήθηκαν στο έκτο κεφάλαιο. Σε αυτό το πλαίσιο, σε κάθε γύρο της ανατροφοδότησης τα χαρακτηριστικά που θα επιλέγονται, θα χρησιμοποιούνται για την κατάλληλη τροποποίηση των μοντέλων των εικόνων. Μετέπειτα, η διαδικασία της συσχετιστικής ανατροφοδότησης θα υλοποιείται πάλι μέσω κάποιου μοντέλου μάθησης.

Τώρα, σε σχέση με τα μοντέλα μάθησης που μπορούν να χρησιμοποιηθούν για την υλοποίηση της διαδικασίας της συσχετιστικής ανατροφοδότησης, η παρούσα διατριβή δεν προχώρησε σε πλήρη μελέτη όλων των δυνατοτήτων. Στις αντίστοιχες μεθόδους που υλοποιήθηκαν, επιλέχθηκε η χρήση ταξινομητών SVM, λόγω του ότι αυτή η μεθοδολογία είναι εξαιρετικά δημοφιλής και επιτυχημένη. Παρόλα αυτά, πολλοί άλλοι αλγόριθμοι μάθησης θα μπορούσαν να εξεταστούν. Συγκεκριμένα, η μεθοδολογία RVM (Relevance Vector Machine – Μηχανή Διανυσμάτων Συνάφειας) [85] θα μπορούσε να θεωρηθεί μια πρώτη υποψήφια προς διερεύνηση. Το κύριο πλεονέκτημά της συνίσταται στο ότι μέσω αυτής αποφεύγουμε την απαίτηση του λίγο-πολύ εμπειρικού καθορισμού τιμών για τις παραμέτρους που μας απασχόλησαν κατά την εκπαίδευση των SVMs. Επιπλέον, όπως αναφέραμε εξ αρχής, παρότι το πρόβλημα της συσχετιστικής ανατροφοδότησης μπορεί να αντιμετωπιστεί ως πρόβλημα δυαδικής ταξινόμησης, η αντιμετώπιση αυτή δεν είναι ακριβής, με την έννοια ότι τα αρνητικά παραδείγματα δεν είναι απαραίτητα να έχουν μεταξύ τους κάποια συνάφεια. Επομένως, οι κατηγορίες των σχετικών και των μη σχετικών εικόνων δεν είναι ομοειδείς. Έτσι, παρουσιάζει ενδιαφέρον η μελέτη μοντέλων μάθησης που είναι ορισμένα με τέτοιο τρόπο ώστε να μπορούν να αντιμετωπίσουν αυτήν την ανομοιότητα των δύο κατηγοριών του προβλήματος ανάκτησης.

Τέλος, ένα άλλο θέμα που σχετίζεται με τους ταξινομητές και τη χρήση τους σε ό,τι έχει να κάνει με την τεχνική της συσχετιστικής ανατροφοδότησης, αφορά την επιλογή των παραδειγμάτων που θα χρησιμοποιηθούν για την εκπαίδευσή τους. Συγκεκριμένα, έχουν προταθεί μέθοδοι που στηρίζονται σε τεχνικές ενεργητικής μάθησης (active learning) για τη βέλτιστη επιλογή των εικόνων που θα παρουσιαστούν στο χρήστη προς αξιολόγηση, που δείχνουν να βελτιώνουν τα αποτελέσματα της κλασικής προσέγγισης της συσχετιστικής ανατροφοδότησης, για την οποία τα παραδείγματα του χρήστη δίνονται απλά μέσω αξιολόγησης των αποτελεσμάτων της ανάκτησης. Επομένως, φαίνεται ενδιαφέρουσα η έρευνα πάνω σε τέτοιες τεχνικές, που θα αφορούν τον προσδιορισμό κατάλληλων συνόλων εικόνων μέσα από τα οποία θα επιλέγει ο χρήστης τα παραδείγματά του.



# Παράρτημα Α

## Αποδείξεις

### Α1. Απόδειξη Σχέσεων (3.34) – (3.36)

Έστω ότι έχουμε δύο Γκαουσιανά Μείγματα με συναρτήσεις πυκνότητας πιθανότητας  $p_m(\mathbf{x})$  και  $p_l(\mathbf{x})$ , για τις οποίες ισχύει

$$p_m(\mathbf{x}) = \sum_{i=1}^{K_m} \pi_{mi} \phi(\mathbf{x} | \boldsymbol{\theta}_{mi}) \quad \text{και} \quad p_l(\mathbf{x}) = \sum_{j=1}^{K_l} \pi_{lj} \phi(\mathbf{x} | \boldsymbol{\theta}_{lj})$$

όπου

$$\boldsymbol{\theta}_{nk} = (\boldsymbol{\mu}_{nk}, \boldsymbol{\Sigma}_{nk}) \quad \text{και} \quad \phi(\mathbf{x} | \boldsymbol{\theta}_{nk}) = \mathcal{N}(\mathbf{x} | \boldsymbol{\theta}_{nk}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_{nk}|}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{nk})^T \boldsymbol{\Sigma}_{nk}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{nk})\right]$$

Τότε

$$\begin{aligned} \int p_m(\mathbf{x}) p_l(\mathbf{x}) d\mathbf{x} &= \int \left[ \left( \sum_{i=1}^{K_m} \pi_{mi} \phi(\mathbf{x} | \boldsymbol{\theta}_{mi}) \right) \left( \sum_{j=1}^{K_l} \pi_{lj} \phi(\mathbf{x} | \boldsymbol{\theta}_{lj}) \right) \right] d\mathbf{x} \\ &\Rightarrow \int p_m(\mathbf{x}) p_l(\mathbf{x}) d\mathbf{x} = \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} \int \phi(\mathbf{x} | \boldsymbol{\theta}_{mi}) \phi(\mathbf{x} | \boldsymbol{\theta}_{lj}) d\mathbf{x} \end{aligned} \quad (\text{A.1})$$

Όμως

$$\begin{aligned} \int \phi(\mathbf{x} | \boldsymbol{\theta}_{mi}) \phi(\mathbf{x} | \boldsymbol{\theta}_{lj}) d\mathbf{x} &= \int \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_{mi}|}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{mi})^T \boldsymbol{\Sigma}_{mi}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{mi})\right] \times \\ &\quad \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_{lj}|}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{lj})^T \boldsymbol{\Sigma}_{lj}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{lj})\right] d\mathbf{x} \\ &= \frac{1}{(2\pi)^d \sqrt{|\boldsymbol{\Sigma}_{mi}| |\boldsymbol{\Sigma}_{lj}|}} \int \exp\left\{-\frac{1}{2}[(\mathbf{x} - \boldsymbol{\mu}_{mi})^T \boldsymbol{\Sigma}_{mi}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{mi}) + (\mathbf{x} - \boldsymbol{\mu}_{lj})^T \boldsymbol{\Sigma}_{lj}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{lj})]\right\} d\mathbf{x} \end{aligned}$$

$$= \frac{\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_{mi}^T \boldsymbol{\Sigma}_{mi}^{-1} \boldsymbol{\mu}_{mi} + \boldsymbol{\mu}_{lj}^T \boldsymbol{\Sigma}_{lj}^{-1} \boldsymbol{\mu}_{lj})\right]}{(2\pi)^d \sqrt{|\boldsymbol{\Sigma}_{mi}| |\boldsymbol{\Sigma}_{lj}|}} \int \exp\left[-\frac{1}{2}(\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} - \mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{x})\right] d\mathbf{x} \quad (\text{A.2})$$

με

$$\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{mi}^{-1} + \boldsymbol{\Sigma}_{lj}^{-1})^{-1} \quad (\text{A.3})$$

και

$$\mathbf{M} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma}_{mi}^{-1} \boldsymbol{\mu}_{mi} + \boldsymbol{\Sigma}_{lj}^{-1} \boldsymbol{\mu}_{lj}) \quad (\text{A.4})$$

Ακόμα, θεωρώντας ότι  $\boldsymbol{\theta} = (\mathbf{M}, \boldsymbol{\Sigma})$ , ξέρουμε ότι

$$\begin{aligned} \int \phi(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x} &= \int \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{M})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{M})\right] d\mathbf{x} = 1 \\ \Rightarrow \int \exp\left[-\frac{1}{2}(\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} - \mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{x})\right] d\mathbf{x} &= \frac{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}}{\exp\left(-\frac{1}{2} \mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M}\right)} \end{aligned} \quad (\text{A.5})$$

Από (A.2) και (A.5) έχουμε

$$\begin{aligned} \int \phi(\mathbf{x} | \boldsymbol{\theta}_{mi}) \phi(\mathbf{x} | \boldsymbol{\theta}_{lj}) d\mathbf{x} &= \frac{\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_{mi}^T \boldsymbol{\Sigma}_{mi}^{-1} \boldsymbol{\mu}_{mi} + \boldsymbol{\mu}_{lj}^T \boldsymbol{\Sigma}_{lj}^{-1} \boldsymbol{\mu}_{lj})\right]}{(2\pi)^d \sqrt{|\boldsymbol{\Sigma}_{mi}| |\boldsymbol{\Sigma}_{lj}|}} \frac{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}}{\exp\left(-\frac{1}{2} \mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M}\right)} \\ &= \sqrt{\frac{|\boldsymbol{\Sigma}|}{(2\pi)^d |\boldsymbol{\Sigma}_{mi}| |\boldsymbol{\Sigma}_{lj}|}} e^{-\frac{1}{2} R} \end{aligned} \quad (\text{A.6})$$

με το  $R$  να ορίζεται από τη σχέση

$$\begin{aligned} R &= \boldsymbol{\mu}_{mi}^T \boldsymbol{\Sigma}_{mi}^{-1} \boldsymbol{\mu}_{mi} + \boldsymbol{\mu}_{lj}^T \boldsymbol{\Sigma}_{lj}^{-1} \boldsymbol{\mu}_{lj} - \mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} \\ &= \boldsymbol{\mu}_{mi}^T \boldsymbol{\Sigma}_{mi}^{-1} \boldsymbol{\mu}_{mi} + \boldsymbol{\mu}_{lj}^T \boldsymbol{\Sigma}_{lj}^{-1} \boldsymbol{\mu}_{lj} - (\boldsymbol{\mu}_{mi}^T \boldsymbol{\Sigma}_{mi}^{-1} + \boldsymbol{\mu}_{lj}^T \boldsymbol{\Sigma}_{lj}^{-1}) (\boldsymbol{\Sigma}_{mi}^{-1} + \boldsymbol{\Sigma}_{lj}^{-1})^{-1} (\boldsymbol{\Sigma}_{mi}^{-1} \boldsymbol{\mu}_{mi} + \boldsymbol{\Sigma}_{lj}^{-1} \boldsymbol{\mu}_{lj}) \end{aligned} \quad (\text{A.7})$$

Από το λήμμα αντιστροφής πίνακα (matrix inversion lemma) έχουμε

$$(\boldsymbol{\Sigma}_{mi}^{-1} + \boldsymbol{\Sigma}_{lj}^{-1})^{-1} = \boldsymbol{\Sigma}_{mi} - \boldsymbol{\Sigma}_{mi} (\boldsymbol{\Sigma}_{mi} + \boldsymbol{\Sigma}_{lj})^{-1} \boldsymbol{\Sigma}_{mi} \quad (\text{A.8})$$

και

$$\left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} = \Sigma_{lj} - \Sigma_{lj} \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \Sigma_{lj} \quad (\text{A.9})$$

Επιπροσθέτως, ισχύει

$$\begin{aligned} \Sigma_{mi}^{-1} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{lj}^{-1} &= \left[\Sigma_{lj} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right) \Sigma_{mi}\right]^{-1} = \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \\ \Rightarrow \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} &= \Sigma_{mi} \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \Sigma_{lj} \end{aligned} \quad (\text{A.10})$$

Με παρόμοιο τρόπο καταλήγουμε στη σχέση

$$\left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} = \Sigma_{lj} \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \Sigma_{mi} \quad (\text{A.11})$$

Συνδυάζοντας τις σχέσεις (A.7) – (A.11), παίρνουμε

$$\begin{aligned} R &= \boldsymbol{\mu}_{mi}^T \Sigma_{mi}^{-1} \boldsymbol{\mu}_{mi} + \boldsymbol{\mu}_{lj}^T \Sigma_{lj}^{-1} \boldsymbol{\mu}_{lj} \\ &- \boldsymbol{\mu}_{mi}^T \Sigma_{mi}^{-1} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{mi}^{-1} \boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj}^T \Sigma_{lj}^{-1} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{lj}^{-1} \boldsymbol{\mu}_{lj} \\ &- \boldsymbol{\mu}_{mi}^T \Sigma_{mi}^{-1} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{lj}^{-1} \boldsymbol{\mu}_{lj} - \boldsymbol{\mu}_{lj}^T \Sigma_{lj}^{-1} \left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1} \Sigma_{mi}^{-1} \boldsymbol{\mu}_{mi} \\ \Rightarrow R &= \left(\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj}\right)^T \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \left(\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj}\right) \end{aligned} \quad (\text{A.12})$$

Τέλος, από τις (A.1), (A.3), (A.6) και (A.12) παίρνουμε

$$\begin{aligned} \int p_m(\mathbf{x}) p_l(\mathbf{x}) d\mathbf{x} &= \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} \sqrt{\frac{|\Sigma|}{(2\pi)^d |\Sigma_{mi}| |\Sigma_{lj}|}} e^{-\frac{1}{2} R} \\ &= \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} \sqrt{\frac{\left|\left(\Sigma_{mi}^{-1} + \Sigma_{lj}^{-1}\right)^{-1}\right|}{(2\pi)^d |\Sigma_{mi}| |\Sigma_{lj}|}} \exp\left[-\frac{1}{2} \left(\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj}\right)^T \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \left(\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj}\right)\right] \\ &= \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} \frac{1}{\sqrt{(2\pi)^d |\Sigma_{mi} + \Sigma_{lj}|}} \exp\left[-\frac{1}{2} \left(\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj}\right)^T \left(\Sigma_{mi} + \Sigma_{lj}\right)^{-1} \left(\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj}\right)\right] \end{aligned}$$

Επομένως, έχουμε (σχέση (3.33))

$$\begin{aligned} S(p_m, p_l) &= \int p_m(\mathbf{x}) p_l(\mathbf{x}) d\mathbf{x} \\ \Rightarrow S(p_m, p_l) &= \sum_{i=1}^{K_m} \sum_{j=1}^{K_l} \pi_{mi} \pi_{lj} \frac{1}{\sqrt{(2\pi)^d |\mathbf{C}_{ml}(i, j)|}} \exp\left[-\frac{1}{2} D_{ml}(i, j)\right] \end{aligned}$$

με

$$\begin{aligned} \mathbf{C}_{ml}(i, j) &= \boldsymbol{\Sigma}_{mi} + \boldsymbol{\Sigma}_{lj} \\ D_{ml}(i, j) &= (\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj})^T \mathbf{C}_{ml}^{-1}(i, j) (\boldsymbol{\mu}_{mi} - \boldsymbol{\mu}_{lj}) \end{aligned}$$

## A2. Απόδειξη Σχέσεων (3.44) – (3.46)

Θεωρούμε ότι έχουμε δύο συναρτήσεις πυκνότητας πιθανότητας  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$ . Λόγω της ανισότητας του Jensen ισχύει ότι

$$\ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} \geq \int p_1(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x} \quad (\text{A.13})$$

και

$$\ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} \geq \int p_2(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} \quad (\text{A.14})$$

Η απόκλιση KL και η απόκλιση C2 για τις  $p_1(\mathbf{x})$  και  $p_2(\mathbf{x})$  υπολογίζονται από τους παρακάτω τύπους (βλ. σχέσεις (3.2) και (3.32), (3.33), αντίστοιχα):

$$\text{KL}(p_1 \| p_2) = \int p_1(\mathbf{x}) \ln \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} d\mathbf{x} \quad \text{και} \quad C2(p_1, p_2) = -\ln \frac{2 \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x}}{\int p_1^2(\mathbf{x}) d\mathbf{x} + \int p_2^2(\mathbf{x}) d\mathbf{x}}$$

Επομένως

$$\begin{aligned} C2(p_1, p_2) - \text{KL}(p_1 \| p_2) &= \\ &= -\ln 2 - \ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} + \ln \left( \int p_1^2(\mathbf{x}) d\mathbf{x} + \int p_2^2(\mathbf{x}) d\mathbf{x} \right) \\ &\quad - \int p_1(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} + \int p_1(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (\text{A.15})$$

και λόγω της (A.13) προκύπτει

$$C2(p_1, p_2) - \text{KL}(p_1 \| p_2) \leq -\ln 2 + \ln \left( \int p_1^2(\mathbf{x}) d\mathbf{x} + \int p_2^2(\mathbf{x}) d\mathbf{x} \right) - \int p_1(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} \quad (\text{A.16})$$

Ομοίως, κάνοντας χρήση της (A.14), αποδεικνύεται ότι

$$C2(p_1, p_2) - \text{KL}(p_2 \| p_1) \leq -\ln 2 + \ln \left( \int p_1^2(\mathbf{x}) d\mathbf{x} + \int p_2^2(\mathbf{x}) d\mathbf{x} \right) - \int p_2(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x} \quad (\text{A.17})$$

Αν λάβουμε υπόψιν ότι η συμμετρική εκδοχή της απόκλισης KL ορίζεται από τη σχέση

$$\text{SKL}(p_1, p_2) = \frac{1}{2} \text{KL}(p_1 \| p_2) + \frac{1}{2} \text{KL}(p_2 \| p_1)$$

προσθέτοντας κατά μέλη τις ανισότητες (A.16) και (A.17), και διαιρώντας με 1/2 και τα δύο μέλη παίρνουμε

$$\begin{aligned}
C2(p_1, p_2) - \text{SKL}(p_1, p_2) &\leq -\ln 2 + \ln \left( \int p_1^2(\mathbf{x}) d\mathbf{x} + \int p_2^2(\mathbf{x}) d\mathbf{x} \right) \\
&\quad - \frac{1}{2} \int p_1(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} - \frac{1}{2} \int p_2(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x} \\
&\leq -\ln 2 + \ln \left( 2 \max_{i=1,2} \left( \int p_i^2(\mathbf{x}) d\mathbf{x} \right) \right) - \min_{i=1,2} \left( \int p_i(\mathbf{x}) \ln p_i(\mathbf{x}) d\mathbf{x} \right) \\
&= \ln \left( \max_{i=1,2} \left( \int p_i^2(\mathbf{x}) d\mathbf{x} \right) \right) - \min_{i=1,2} \left( \int p_i(\mathbf{x}) \ln p_i(\mathbf{x}) d\mathbf{x} \right) \\
&= \max_{i=1,2} \left( \ln \int p_i^2(\mathbf{x}) d\mathbf{x} \right) - \min_{i=1,2} \left( \int p_i(\mathbf{x}) \ln p_i(\mathbf{x}) d\mathbf{x} \right) \geq 0
\end{aligned} \tag{A.18}$$

αφού, για κάθε  $i = 1, 2$ , λόγω της ανισότητας του Jensen ισχύει

$$\ln \int p_i^2(\mathbf{x}) d\mathbf{x} \geq \int p_i(\mathbf{x}) \ln p_i(\mathbf{x}) d\mathbf{x} \tag{A.19}$$

Επίσης, λόγω του ότι ισχύει

$$\ln \left( \frac{1}{2} \int p_1^2(\mathbf{x}) d\mathbf{x} + \frac{1}{2} \int p_2^2(\mathbf{x}) d\mathbf{x} \right) \geq \frac{1}{2} \ln \int p_1^2(\mathbf{x}) d\mathbf{x} + \frac{1}{2} \ln \int p_2^2(\mathbf{x}) d\mathbf{x}$$

κάνοντας χρήση πάλι της ανισότητας του Jensen, από την (A.15) προκύπτει ότι

$$\begin{aligned}
C2(p_1, p_2) - \text{KL}(p_1 \parallel p_2) &\geq -\ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} + \frac{1}{2} \ln \int p_1^2(\mathbf{x}) d\mathbf{x} + \frac{1}{2} \ln \int p_2^2(\mathbf{x}) d\mathbf{x} \\
&\quad - \int p_1(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} + \int p_1(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x}
\end{aligned} \tag{A.20}$$

Αντίστοιχα, παίρνουμε

$$\begin{aligned}
C2(p_1, p_2) - \text{KL}(p_2 \parallel p_1) &\geq -\ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} + \frac{1}{2} \ln \int p_1^2(\mathbf{x}) d\mathbf{x} + \frac{1}{2} \ln \int p_2^2(\mathbf{x}) d\mathbf{x} \\
&\quad - \int p_2(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x} + \int p_2(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x}
\end{aligned} \tag{A.21}$$

Οπότε, προσθέτοντας κατά μέλη τις (A.20) και (A.21), και διαιρώντας με 1/2 έχουμε

$$\begin{aligned}
C2(p_1, p_2) - \text{SKL}(p_1, p_2) &\geq -\frac{1}{2} \left[ \ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} - \int p_1(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x} \right] \\
&\quad - \frac{1}{2} \left[ \ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} - \int p_2(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} \right] \\
&\quad + \frac{1}{2} \left[ \ln \int p_1^2(\mathbf{x}) d\mathbf{x} - \int p_1(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} \right] \\
&\quad + \frac{1}{2} \left[ \ln \int p_2^2(\mathbf{x}) d\mathbf{x} - \int p_2(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x} \right]
\end{aligned}$$

και, επομένως, χρησιμοποιώντας και την (A.19), οδηγούμαστε στη σχέση

$$\begin{aligned}
C2(p_1, p_2) - \text{SKL}(p_1, p_2) &\geq -\frac{1}{2} \left[ \ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} - \int p_1(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x} \right] \\
&\quad - \frac{1}{2} \left[ \ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} - \int p_2(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} \right] \\
&\geq -\ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} + \min \left( \int p_1(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x}, \int p_2(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} \right) \leq 0 \quad (\text{A.22})
\end{aligned}$$

λόγω των (A.13) και (A.14).

Συνεπώς, από (A.18) και (A.22) προκύπτει

$$-\Delta_{ij} \leq C2(p_1, p_2) - \text{SKL}(p_1, p_2) \leq \Delta_{ii}$$

$$\begin{aligned}
\Delta_{ii} &= \max_{i=1,2} \left( \ln \int p_i^2(\mathbf{x}) d\mathbf{x} \right) - \min_{i=1,2} \left( \int p_i(\mathbf{x}) \ln p_i(\mathbf{x}) d\mathbf{x} \right) \\
&= \max \left( \ln E[p_1]_{p_1}, \ln E[p_2]_{p_2} \right) - \min \left( E[\ln p_1]_{p_1}, E[\ln p_2]_{p_2} \right) \geq 0
\end{aligned}$$

$$\begin{aligned}
\Delta_{ij} &= \ln \int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} - \min \left( \int p_1(\mathbf{x}) \ln p_2(\mathbf{x}) d\mathbf{x}, \int p_2(\mathbf{x}) \ln p_1(\mathbf{x}) d\mathbf{x} \right) \\
&= \max \left( \ln E[p_2]_{p_1}, \ln E[p_1]_{p_2} \right) - \min \left( E[\ln p_2]_{p_1}, E[\ln p_1]_{p_2} \right) \geq 0
\end{aligned}$$

όπου, προφανώς,  $\max(\ln E[p_2]_{p_1}, \ln E[p_1]_{p_2}) = \ln E[p_2]_{p_1} = \ln E[p_1]_{p_2}$ .

### A3. Απόδειξη Σχέσεων (4.5) – (4.7)

Με δεδομένο ότι ο τύπος για την ανανέωση του μοντέλου της ερώτησης είναι ο παρακάτω (σχέση (4.1)):

$$p_{q'}(\mathbf{x}) = (1 - \Lambda_r) p_q(\mathbf{x}) + \sum_{m=1}^{M_r} \lambda_m^r p_m^r(\mathbf{x})$$

και ότι η απόκλιση C2 μεταξύ του μοντέλου της νέας ερώτησης και αυτού της  $i$ -οστής εικόνας της βάσης υπολογίζεται μέσω της σχέσης

$$C2(p_{q'}, p_i) = -\ln \frac{2S(p_{q'}, p_i)}{S(p_{q'}, p_{q'}) + S(p_i, p_i)}$$

(όπως μπορούμε εύκολα να συμπεράνουμε από την (3.32)) και, επίσης, λαμβάνοντας υπόψιν ότι η ποσότητα

$$S(p_i, p_i) = \int p_i^2(\mathbf{x}) d\mathbf{x}$$

χρησιμοποιεί αποκλειστικά και μόνο το μοντέλο της  $i$ -οστής εικόνας της βάσης, και επομένως δε μεταβάλλεται από τη μια εποχή της ανατροφοδότησης στην άλλη, αυτό που μας μένει να υπολογίσουμε είναι οι ποσότητες  $S(p_{q'}, p_i)$  και  $S(p_{q'}, p_{q'})$ . Συγκεκριμένα, ισχύει

$$\begin{aligned}
S(p_{q'}, p_i) &= \int p_{q'}(\mathbf{x}) p_i(\mathbf{x}) d\mathbf{x} = \int \left[ (1 - \Lambda_r) p_q(\mathbf{x}) + \sum_{m=1}^{M_r} \lambda_m^r p_m^r(\mathbf{x}) \right] p_i(\mathbf{x}) d\mathbf{x} \\
&\Rightarrow S(p_{q'}, p_i) = (1 - \Lambda_r) \int p_q(\mathbf{x}) p_i(\mathbf{x}) d\mathbf{x} + \sum_{m=1}^{M_r} \lambda_m^r \int p_m^r(\mathbf{x}) p_i(\mathbf{x}) d\mathbf{x} \\
&\Rightarrow S(p_{q'}, p_i) = (1 - \Lambda_r) S(p_q, p_i) + \sum_{m=1}^{M_r} \lambda_m^r S(p_m^r, p_i)
\end{aligned}$$

και

$$\begin{aligned}
S(p_{q'}, p_{q'}) &= \int p_{q'}(\mathbf{x}) p_{q'}(\mathbf{x}) d\mathbf{x} \\
&= \int \left[ (1 - \Lambda_r) p_q(\mathbf{x}) + \sum_{m=1}^{M_r} \lambda_m^r p_m^r(\mathbf{x}) \right] \left[ (1 - \Lambda_r) p_q(\mathbf{x}) + \sum_{m=1}^{M_r} \lambda_m^r p_m^r(\mathbf{x}) \right] d\mathbf{x} \\
&= (1 - \Lambda_r)^2 \int p_q^2(\mathbf{x}) d\mathbf{x} + 2(1 - \Lambda_r) \sum_{m=1}^{M_r} \lambda_m^r \int p_q(\mathbf{x}) p_m^r(\mathbf{x}) d\mathbf{x} + \sum_{m=1}^{M_r} \sum_{m'=1}^{M_r} \lambda_m^r \lambda_{m'}^r \int p_m^r(\mathbf{x}) p_{m'}^r(\mathbf{x}) d\mathbf{x} \\
&\Rightarrow S(p_{q'}, p_{q'}) = (1 - \Lambda_r)^2 S(p_q, p_q) + 2(1 - \Lambda_r) \sum_{m=1}^{M_r} \lambda_m^r S(p_q, p_m^r) + \sum_{m=1}^{M_r} \sum_{m'=1}^{M_r} \lambda_m^r \lambda_{m'}^r S(p_m^r, p_{m'}^r)
\end{aligned}$$

#### A4. Απόδειξη Σχέσης (4.17)

Ο τύπος για την ανανέωση του μοντέλου της ερώτησης είναι ο παρακάτω (σχέση (4.1)):

$$p_{q'}(\mathbf{x}) = (1 - \Lambda_r) p_q(\mathbf{x}) + \sum_{m=1}^{M_r} \lambda_m^r p_m^r(\mathbf{x})$$

Θεωρώντας ότι το μοντέλο της τρέχουσας ερώτησης, τα μοντέλα που αντιστοιχούν στα παραδείγματα του χρήστη, και τα μοντέλα των εικόνων της βάσης είναι Γκαουσιανά Μείγματα και ορίζονται ως

$$\begin{aligned}
p_q(\mathbf{x}) &= \sum_{j=1}^{K_q} \pi_{qj} \phi(\mathbf{x} | \theta_{qj}) \quad \text{με} \quad \theta_{qj} = (\boldsymbol{\mu}_{qj}, \boldsymbol{\Sigma}_{qj}), \\
p_m^r(\mathbf{x}) &= \sum_{j=1}^{K_m^r} \pi_{mj}^r \phi(\mathbf{x} | \theta_{mj}^r) \quad \text{με} \quad \theta_{mj}^r = (\boldsymbol{\mu}_{mj}^r, \boldsymbol{\Sigma}_{mj}^r), \\
p_i(\mathbf{x}) &= \sum_{j=1}^{K_i} \pi_{ij} \phi(\mathbf{x} | \theta_{ij}) \quad \text{με} \quad \theta_{ij} = (\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}),
\end{aligned}$$

αντίστοιχα, το νέο μοντέλο της ερώτησης είναι Γκαουσιανό Μείγμα που δίνεται από την παρακάτω σχέση:

$$p_{q'}(\mathbf{x}) = \sum_{j=1}^{K_{q'}} \pi_{q'j} \phi(\mathbf{x} | \theta_{q'j})$$

$$= \sum_{j=1}^{K_q} (1 - \Lambda_r) \pi_{qj} \phi(\mathbf{x} | \boldsymbol{\theta}_{qj}) + \sum_{j=1}^{K_1^r} \lambda_1^r \pi_{1j}^r \phi(\mathbf{x} | \boldsymbol{\theta}_{1j}^r) + \dots + \sum_{j=1}^{K_{M_r}^r} \lambda_{M_r}^r \pi_{M_r j}^r \phi(\mathbf{x} | \boldsymbol{\theta}_{M_r j}^r) \quad (\text{A.23})$$

με  $K_{q'} = K_q + K_1^r + \dots + K_{M_r}^r$ .

Επιπροσθέτως, όπως προκύπτει από τον ορισμό της (σχέσεις (3.13) και (3.14)), η ALA μεταξύ του νέου μοντέλου της ερώτησης και αυτού της  $i$ -οστής εικόνας της βάσης δίνεται από τον τύπο

$$\text{ALA}(p_{q'} \| p_i) = \sum_{j=1}^{K_{q'}} \pi_{q'j} \left[ \ln \pi_{i\beta_{q'i}(j)} + \ln \phi(\boldsymbol{\mu}_{q'j} | \boldsymbol{\theta}_{i\beta_{q'i}(j)}) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_{i\beta_{q'i}(j)}^{-1} \boldsymbol{\Sigma}_{q'j}) \right] \quad (\text{A.24})$$

με

$$\beta_{q'i}(j) = \arg \min_{k=1, \dots, K_i} \left\{ (\boldsymbol{\mu}_{q'j} - \boldsymbol{\mu}_{ik})^T \boldsymbol{\Sigma}_{ik}^{-1} (\boldsymbol{\mu}_{q'j} - \boldsymbol{\mu}_{ik}) - \ln \pi_{ik} \right\} \quad (\text{A.25})$$

Λαμβάνοντας υπόψιν τη σχέση (A.23), τα διατεταγμένα σύνολα των παραμέτρων  $\pi_{q'j}$ ,  $\boldsymbol{\mu}_{q'j}$  και  $\boldsymbol{\Sigma}_{q'j}$  για τις συνιστώσες του νέου μοντέλου της ερώτησης είναι τα ακόλουθα:

$$C_\pi = \left\{ (1 - \Lambda_r) \pi_{q1}, \dots, (1 - \Lambda_r) \pi_{qK_q}, \lambda_1^r \pi_{11}^r, \dots, \lambda_1^r \pi_{1K_1^r}^r, \dots, \lambda_{M_r}^r \pi_{M_r 1}^r, \dots, \lambda_{M_r}^r \pi_{M_r K_{M_r}^r}^r \right\} \quad (\text{A.26})$$

$$C_\mu = \left\{ \boldsymbol{\mu}_{q1}, \dots, \boldsymbol{\mu}_{qK_q}, \boldsymbol{\mu}_{11}^r, \dots, \boldsymbol{\mu}_{1K_1^r}^r, \dots, \boldsymbol{\mu}_{M_r 1}^r, \dots, \boldsymbol{\mu}_{M_r K_{M_r}^r}^r \right\} \quad (\text{A.27})$$

$$C_\Sigma = \left\{ \boldsymbol{\Sigma}_{q1}, \dots, \boldsymbol{\Sigma}_{qK_q}, \boldsymbol{\Sigma}_{11}^r, \dots, \boldsymbol{\Sigma}_{1K_1^r}^r, \dots, \boldsymbol{\Sigma}_{M_r 1}^r, \dots, \boldsymbol{\Sigma}_{M_r K_{M_r}^r}^r \right\} \quad (\text{A.28})$$

Επιπλέον, από την (A.25) βλέπουμε ότι κάθε τιμή  $\beta_{q'i}(j)$  εξαρτάται μόνο από τη μέση τιμή  $\boldsymbol{\mu}_{q'j} \in C_\mu$  και από τις παραμέτρους  $\pi_{ik}$ ,  $\boldsymbol{\mu}_{ik}$ ,  $\boldsymbol{\Sigma}_{ik}$ ,  $k=1, \dots, K_i$ , του μοντέλου της  $i$ -οστής εικόνας της βάσης. Επομένως, οι τιμές της συνάρτησης αντιστοίχισης μεταξύ των συνιστωσών που αποτελούν το μοντέλο για τη νέα ερώτηση και των συνιστωσών του μοντέλου της  $i$ -οστής εικόνας της βάσης δε μεταβάλλονται κατά τη συγχώνευση των Γκαουσιανών Μειγμάτων της προηγούμενης ερώτησης και των παραδειγμάτων του χρήστη στο νέο μοντέλο. Συγκεκριμένα, το διατεταγμένο σύνολο των τιμών της συνάρτησης αντιστοίχισης για τις συνιστώσες του νέου μοντέλου είναι

$$C_\beta = \left\{ \beta_{q'i}(1), \dots, \beta_{q'i}(K_q), \beta_{1i}^r(1), \dots, \beta_{1i}^r(K_1^r), \dots, \beta_{M_r i}^r(1), \dots, \beta_{M_r i}^r(K_{M_r}^r) \right\} \quad (\text{A.29})$$

όπου  $\beta_{mi}^r$  είναι η συνάρτηση αντιστοίχισης μεταξύ του  $m$ -οστού παραδείγματος και της  $i$ -οστής εικόνας της βάσης.

Έτσι, από τις (A.24) και (A.26)-(A.29) προκύπτει ότι για την  $\text{ALA}(p_{q'} \| p_i)$  ισχύει

$$\begin{aligned} \text{ALA}(p_{q'} \| p_i) &= \sum_{j=1}^{K_q} (1 - \Lambda_r) \pi_{qj} \left[ \ln \pi_{i\beta_{q'i}(j)} + \ln \phi(\boldsymbol{\mu}_{qj} | \boldsymbol{\theta}_{i\beta_{q'i}(j)}) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_{i\beta_{q'i}(j)}^{-1} \boldsymbol{\Sigma}_{qj}) \right] \\ &+ \sum_{j=1}^{K_1^r} \lambda_1^r \pi_{1j}^r \left[ \ln \pi_{i\beta_{1i}^r(j)} + \ln \phi(\boldsymbol{\mu}_{1j}^r | \boldsymbol{\theta}_{i\beta_{1i}^r(j)}) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_{i\beta_{1i}^r(j)}^{-1} \boldsymbol{\Sigma}_{1j}^r) \right] \end{aligned}$$



$$+ \dots + \sum_{j=1}^{K_{M_r}^r} \lambda_{M_r}^r \pi_{M_r, j}^r \left[ \ln \pi_{i\beta_{M_r, i}^r(j)} + \ln \phi(\boldsymbol{\mu}_{M_r, j}^r | \boldsymbol{\theta}_{i\beta_{M_r, i}^r(j)}) - \frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_{i\beta_{M_r, i}^r(j)}^{-1} \boldsymbol{\Sigma}_{M_r, j}^r) \right]$$

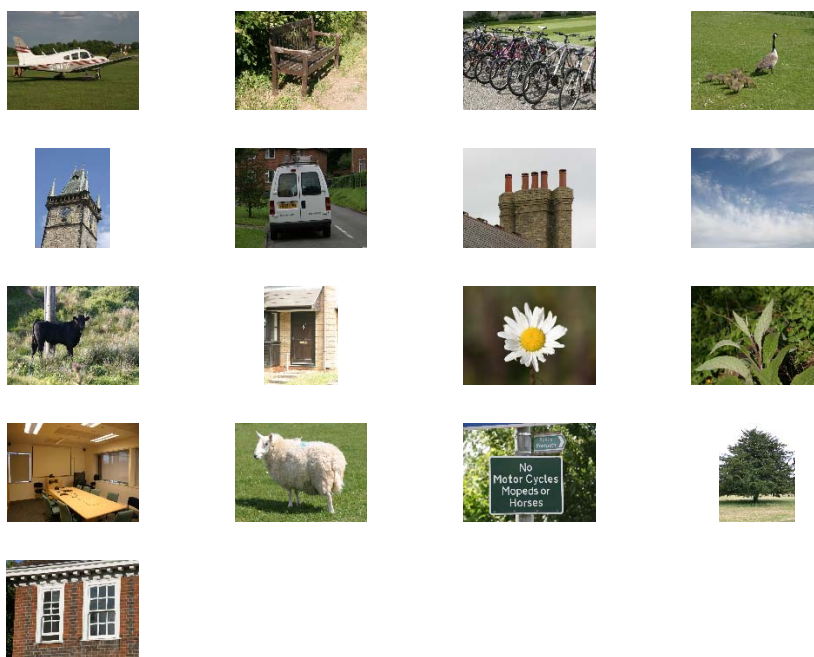
$$\Rightarrow \text{ALA}(p_q \parallel p_i) = (1 - \Lambda_r) \text{ALA}(p_q \parallel p_i) + \sum_{m=1}^{M_r} \lambda_m^r \text{ALA}(p_m^r \parallel p_i)$$



# Παράρτημα Β

## Βάσεις Εικόνων

### Β1. Βάση Εικόνων MSDB



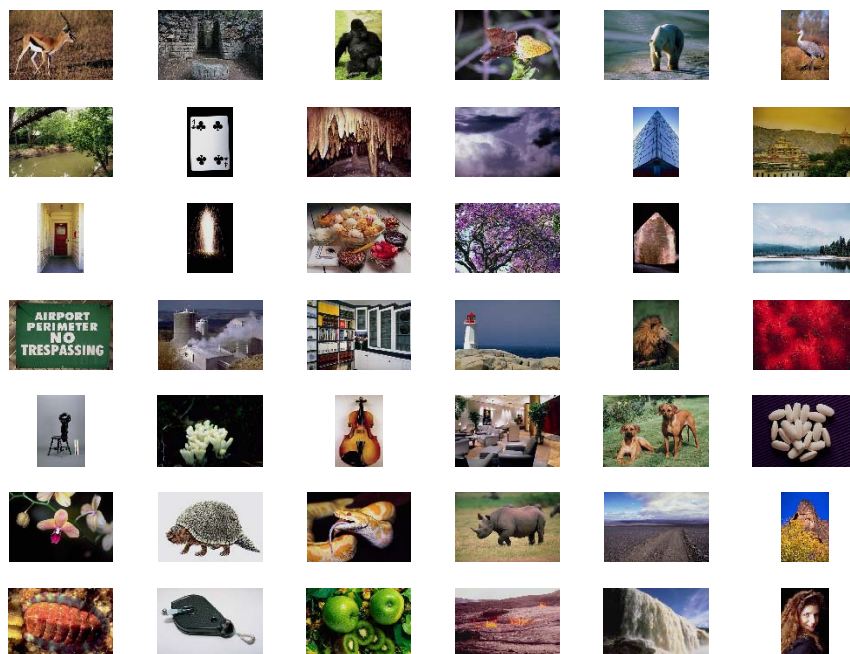
**Σχήμα Β.1:** Ενδεικτικές εικόνες από τις κατηγορίες της βάσης εικόνων MSDB

**Πίνακας Β.1:** Αναλυτικός πίνακας των περιεχομένων ανά κατηγορία για τη βάση εικόνων MSDB

α/α	Κατηγορία	Αριθμός Εικόνων
1	Airplanes	58
2	Benches and Chairs	68

3	Bicycles	272
4	Birds	72
5	Buildings	146
6	Cars	505
7	Chimneys	266
8	Clouds	429
9	Cows	182
10	Doors	166
11	Flowers	166
12	Leaves	118
13	Offices	68
14	Sheep	190
15	Signs	165
16	Trees	217
17	Windows	652
<b>Σύνολο</b>		<b>3740</b>

## B2. Βάση Εικόνων Corel9923

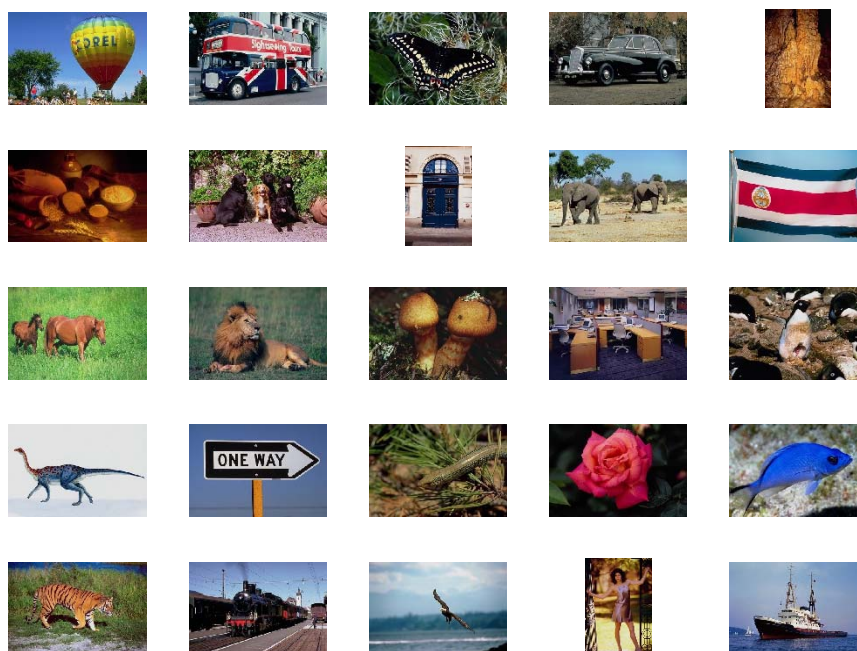


Σχήμα B.2: Ενδεικτικές εικόνες από τις κατηγορίες της βάσης εικόνων Corel9923

**Πίνακας Β.2:** Αναλυτικός πίνακας των περιεχομένων ανά κατηγορία για τη βάση εικόνων Corel9923

<b>α/α</b>	<b>Κατηγορία</b>	<b>Αριθμός Εικόνων</b>
1	African Antelopes	99
2	Ancient Culture	800
3	Apes	100
4	Arthropods and Insects	600
5	Bears	100
6	Birds	793
7	Bridges	157
8	Cards	100
9	Caverns	100
10	Clouds	100
11	Contemporary Buildings	90
12	Dawn and Dusk	284
13	Doors	199
14	Fireworks	197
15	Food	574
16	Forests and Trees	96
17	Gems	100
18	Glaciers and Mountains	198
19	Highway and Street Signs	99
20	Industry	171
21	Kitchens and Bathrooms	96
22	Lighthouses	97
23	Lions	200
24	Marine Life	281
25	Museum Furniture	100
26	Mushrooms	100
27	Musical Instruments	98
28	Office Interiors	100
29	Pets	300
30	Pills	100
31	Plants and Flowers	1255
32	Prehistoric World	100
33	Reptiles and Amphibians	393
34	Rhinos and Hippos	100
35	Roads and Highways	72
36	Rocks	380
37	Shells	100
38	Tools	100
39	Vegetables, Fruits and Nuts	395
40	Volcanic Eruptions	99
41	Waterfalls	95
42	Women	305
<b>Σύνολο</b>		<b>9923</b>

### B3. Βάση Εικόνων Corel4500



Σχήμα Β.3: Ενδεικτικές εικόνες από τις κατηγορίες της βάσης εικόνων Corel4500

Πίνακας Β.3: Αναλυτικός πίνακας των περιεχομένων ανά κατηγορία για τη βάση εικόνων Corel4500

α/α	Κατηγορία	Αριθμός Εικόνων
1	Balloons	200
2	Buses	100
3	Butterflies	200
4	Cars	300
5	Caverns	100
6	Cuisine	300
7	Dogs	300
8	Doors	200
9	Elephants	100
10	Flags	100
11	Horses	100
12	Lions	100
13	Mushrooms	200
14	Office Interiors	100
15	Penguins	100
16	Prehistoric Life	200
17	Road Signs	100
18	Reptiles and Amphibians	200
19	Roses	200

20	Subsea Scenes	300
21	Tigers	100
22	Trains	300
23	Wild Eagles	100
24	Women	400
25	Workships	100
<b>Σύνολο</b>		<b>4500</b>





## Βιβλιογραφία

- [1] DATTA R., LI J., WANG J.Z.: ‘Content-based image retrieval: approaches and trends of the new age’. In Proc. of the 7th ACM SIGMM Int. Workshop on Multimedia Information Retrieval (MIR’05), 2005, pp. 253–262
- [2] ISHIKAWA Y., SUBRAMANYA R., FALOUTSOS C.: ‘MindReader: querying databases through multiple examples’. In Proc. of the 24th Int. Conf. on Very Large Data Bases (VLDB’98), 1998, pp. 218–227
- [3] COX I.J., MILLER M.L., MINKA T.P., PAPATHOMAS T.V., YIANILOS P.N.: ‘The bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments’. IEEE Trans. on Image Processing, 2000, 9, (1), pp. 20–37
- [4] RUI Y., HUANG T.S., ORTEGA M., MEHROTRA S.: ‘Relevance feedback: a power tool for interactive content-based image retrieval’. IEEE Trans. on Circuits and Systems for Video Technology, 1998, 8, (5), pp. 644–655
- [5] CARSON C., BELONGIE S., GREENSPAN H., MALIK J.: ‘Blobworld: image segmentation using expectation-maximization and its application to image querying’. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24, (8), pp. 1026–1038
- [6] CHEN Y., WANG J.Z.: ‘A region-based fuzzy feature matching approach to content-based image retrieval’. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24, (9), pp. 1252–1267
- [7] GUO G.D., JAIN A.K., MA W.Y., ZHANG H.J.: ‘Learning similarity measure for natural image retrieval with relevance feedback’. IEEE Trans. on Neural Networks, 2002, 13, (4), pp. 811–820
- [8] AGGARWAL G., ASHWIN T.V., GHOSAL S.: ‘An image retrieval system with automatic query modification’. IEEE Trans. on Multimedia, 2002, 4, (2), pp. 201–214
- [9] JING F., LI M., ZHANG H.J., ZHANG B.: ‘An efficient and effective region-based image retrieval framework’. IEEE Trans. on Image Processing, 2004, 13, (5), pp. 699–709

- [10] MANJUNATH B.S., OHM J.R., VASUDEVAN V.V., YAMADA A.: ‘Color and texture descriptors’. *IEEE Trans. on Circuits and Systems for Video Technology*, 2001, 11, (6), pp. 703–715
- [11] DO M.N., VETTERLI M.: ‘Wavelet-based texture retrieval using generalized gaussian density and Kullback-Leibler distance’. *IEEE Trans. on Image Processing*, 2002, 11, (2), pp. 146–158
- [12] HSU C.T., LI C.Y.: ‘Relevance feedback using generalized bayesian framework with region-based optimization learning’. *IEEE Trans. on Image Processing*, 2005, 14, (10), pp. 1617–1631
- [13] RUI Y., HUANG T.S., CHANG S.F.: ‘Image retrieval: current techniques, promising directions, and open issues’. *Journal of Visual Communication and Image Representation*, 1999, 10, pp. 39–62
- [14] EL-NAQA I., YANG Y., GALATSANOS N., WERNICK M.: ‘A similarity learning approach to content based image retrieval: application to digital mammography’. *IEEE Trans. on Medical Imaging*, 2004, 23, (10), pp. 1233–1244
- [15] QIAN F., LI M., ZHANG L., ZHANG H.-J., ZHANG B.: ‘Gaussian mixture model for relevance feedback in image retrieval’. In *Proc. of IEEE Int. Conf. on Multimedia and Expo (ICME’02)*, 2002, 1, pp. 229–232
- [16] BISHOP C.M.: ‘Neural networks for pattern recognition’ (Oxford University Press Inc, New York, 1995)
- [17] GREENSPAN H., DVIR G., RUBNER Y.: ‘Context-dependent segmentation and matching in image databases’. *Computer Vision and Image Understanding*, 2004, 93, pp. 86–109
- [18] VASCONCELOS N.: ‘On the efficient evaluation of probabilistic similarity functions for image retrieval’. *IEEE Trans. on Information Theory*, 2004, 50, (7), pp. 1482–1496
- [19] JING F., LI M., ZHANG H.-J., ZHANG B.: ‘Relevance feedback in region-based image retrieval’. *IEEE Trans. on Circuits and Systems for Video Technology*, 2004, 14, (5), pp. 672–681
- [20] TONG S., CHANG E.: ‘Support vector machine active learning for image retrieval’. In *Proc. of ACM Int. Conf. on Multimedia*, 2001, pp. 107–118
- [21] SU Z., ZHANG H., LI S., MA S.: ‘Relevance feedback in content-based image retrieval: bayesian framework, feature subspaces, and progressive learning’. *IEEE Trans. on Image Processing*, 2003, 12, (8), pp. 924–937
- [22] KHERFI M.L., ZIOU D., BERNARDI A.: ‘Combining positive and negative examples in relevance feedback for content-based image retrieval’. *Journal of Visual Communication and Image Representation*, 2003, 14, pp. 428–457

- [23] RUI Y., HUANG T.S.: ‘Optimizing learning in image retrieval’. In Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR’00), 2000, 1, pp. 236–243
- [24] KHERFI M.L., ZIOU D.: ‘Relevance feedback for CBIR: a new approach based on probabilistic feature weighting with positive and negative examples’. IEEE Trans. on Image Processing, 2006, 15, (4), pp. 1017–1030
- [25] MUNEEAWANG P., GUAN L.: ‘Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture’. IEEE Trans. on Neural Networks, 2002, 13, (4), pp. 821–834
- [26] VASCONCELOS N., LIPPMAN A.: ‘Learning over multiple temporal scales in image databases’. In Proc. of the 6th European Conf. on Computer Vision (ECCV’00), 2000, Lecture Notes in Computer Science, 1842, pp. 33–47
- [27] VASCONCELOS N., LIPPMAN A.: ‘Learning from user feedback in image retrieval systems’. In Proc. of Advances in Neural Information Processing Systems 12 (NIPS’99), 1999, pp. 977–986
- [28] SFIKAS G., CONSTANTINOPOULOS C., LIKAS A., GALATSANOS N.P.: ‘An analytic distance metric for gaussian mixture models with application in image retrieval’. In Proc. of the 15th Int. Conf. on Artificial Neural Networks (ICANN’05), 2005, Lecture Notes in Computer Science, 3697, pp. 835–840
- [29] McLACHLAN G.M., PEEL D.: ‘Finite mixture models’ (John Wiley and Sons Inc, New York, 2001)
- [30] Microsoft Research Cambridge Object Recognition Image Database: ‘Version 1.0’. <http://research.microsoft.com/research/downloads/Details/b94de342-60dc-45d0-830b-9f6eff91b301/Details.aspx>
- [31] VLASSIS N., LIKAS A.: ‘A greedy EM algorithm for gaussian mixture learning’. Neural Processing Letters, 2002, 15, (1), pp. 77–87
- [32] BIMBO A.D.: ‘Visual information retrieval’ (Morgan Kaufmann, San Mateo, CA, 1999)
- [33] BISHOP C.M.: ‘Pattern recognition and machine learning’ (Springer, 2006)
- [34] MARAGOS P.: ‘Computer vision’ (NTUA, 2005)
- [35] MITCHELL T.M.: ‘Machine learning’ (McGraw Hill, 1997)
- [36] KAY S.M.: ‘Fundamentals of statistical signal processing: estimation theory’ (Englewood Cliffs, NJ: Prentice-Hall, 1993)
- [37] KAY S.M.: ‘Fundamentals of statistical signal processing: detection theory’ (Englewood Cliffs, NJ: Prentice-Hall, 1998)

- [38] FORSYTH D.A., PONCE J.: ‘Computer vision: a modern approach’ (Prentice Hall, 2003)
- [39] SMEULDERS A.W.M., WORRING M., SANTINI S., GUPTA A., JAIN R.: ‘Content-based image retrieval at the end of the early years’. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000, 22, (12), pp. 1349–1380
- [40] LOWE D.G.: ‘Distinctive image features from scale-invariant keypoints’. *Int. Journal of Computer Vision*, 2004, 60, (2), pp. 91–110
- [41] ABDEL-HAKIM A.E., FARAG A.A.: ‘CSIFT: a SIFT descriptor with color invariant characteristics’. In *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 2006, 2, pp. 1978–1983
- [42] da CUNHA A.L., JIANPING Z., DO M.N.: ‘The nonsubsampling contourlet transform: theory, design, and applications’. *IEEE Trans. on Image Processing*, 2006, 15, (10), pp. 3089–3101
- [43] LEE K.-M., STREET W.N.: ‘Cluster-driven refinement for content-based digital image retrieval’. *IEEE Trans. on Multimedia*, 2004, 6, (6), pp. 817–827
- [44] KSANTINI R., ZIOU D., COLIN B., DUBEAU F.: ‘Weighted pseudometric discriminatory power improvement using a bayesian logistic regression model based on a variational method’. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2008, 30, (2), pp. 253–266
- [45] GOLDBERGER J., GORDON S., GREENSPAN H.: ‘Unsupervised image-set clustering using an information theoretic framework’. *IEEE Trans. on Image Processing*, 2006, 15, (2), pp. 449–458
- [46] ROCCHIO J.J.: ‘Relevance feedback in information retrieval’. In ‘The smart retrieval system: experiments in automatic document processing’ (G. Salton, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1971), pp. 313–323
- [47] JIANG W., ER G., DAI Q., GU J.: ‘Similarity-based online feature selection in content-based image retrieval’. *IEEE Trans. on Image Processing*, 2006, 15, (3), pp. 702–712
- [48] TAO D., TANG X., LI X.: ‘Which components are important for interactive image searching?’. *IEEE Trans. on Circuits and Systems for Video Technology*, 2008, 18, (1), pp. 3–11
- [49] VASCONCELOS N.: ‘Minimum probability of error image retrieval’. *IEEE Trans. on Signal Processing*, 2004, 52, (8), pp. 2322–2336
- [50] SEGHOUANE A.-K., AMARI S.-I.: ‘The AIC criterion and symmetrizing the Kullback-Leibler divergence’. *IEEE Trans. on Neural Networks*, 2007, 18, (1), pp. 97–106

- [51] FIGUEIREDO M., LEITAO J., JAIN A.K.: ‘On fitting mixture models’. In ‘Energy minimization methods in computer vision and pattern recognition’ (E. Hancock and M. Pellilo, eds., Springer Verlag, 1999), pp. 54–69
- [52] GRUNWALD P., MYUNG I.J., PITT M.A.: ‘Advances in minimum description length: theory and applications’ (The MIT Press, 2005)
- [53] FUKUNAGA K.: ‘Introduction to statistical pattern recognition’ (Academic Press, 1990)
- [54] GOLDBERGER J., GORDON S., GREENSPAN H.: ‘An efficient image similarity measure based on approximations of KL-divergence between two gaussian mixtures’. In Proc. of the 9th IEEE Int. Conf. on Computer Vision (ICCV’03), 2003, 1, pp. 487–493
- [55] GUYON I., WESTON J., BARNHILL S., VAPNIK V.: ‘Gene selection for cancer classification using support vector machines’. Machine Learning, 2002, 46, pp. 389–422
- [56] DUAN K., RAJAPAKSE J.C.: ‘SVM-RFE peak selection for cancer classification with mass spectrometry data’. In Proc. of the 3rd Asia-Pacific Bioinformatics Conf., 2005, 1, pp. 191–200
- [57] PENG H., LONG F., DING C.: ‘Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy’. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27, (8), pp. 1226–1238
- [58] DING C., PENG H.: ‘Minimum redundancy feature selection from microarray gene expression data’. Journal of Bioinformatics and Computational Biology, 2005, 3, (2), pp. 185–205
- [59] TAO D., TANG X., LI X., RUI Y.: ‘Direct kernel biased discriminant analysis: a new content-based image retrieval relevance feedback algorithm’. IEEE Trans. on Multimedia, 2006, 8, (4), pp. 716–727
- [60] WANG L.: ‘Feature selection with kernel class separability’. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2008, 30, (9), pp. 1534–1546
- [61] MA W.-Y., ZHANG H.J.: ‘Benchmarking of image features for content-based image retrieval’. In Conf. Record of the Thirty-Second Asilomar Conf. on Signals, Systems and Computers, 1998, 1, pp. 253–257
- [62] CHANG T., KUO C.-C. J.: ‘Texture analysis and classification with tree-structured wavelet transform’. IEEE Trans. on Image Processing, 1993, 2, (4), pp. 429–441
- [63] MANJUNATH B.S., MA W.Y.: ‘Texture features for browsing and retrieval of image data’. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1996, 18, (8), pp. 837–842

- [64] WON C.S., PARK D.K., PARK S.-J.: ‘Efficient use of MPEG-7 edge histogram descriptor’. *Electronics and Telecommunications Research Institute (ETRI) Journal*, 2002, 24, (1), pp. 23–30
- [65] YOON S.J., PARK D.K., PARK S.-J., WON C.S.: ‘Image retrieval using a novel relevance feedback for edge histogram descriptor of MPEG-7’. *Int. Conf. on Consumer Electronics (ICCE’01)*, 2001, pp. 354–355
- [66] ZHANG J., MARSZALEK M., LAZEBNIK S. SCHMID C.: ‘Local features and kernels for classification of texture and object categories: an in-depth study’. *Technical Report RR-5753, INRIA*, 2005
- [67] LIBSVM – A Library for Support Vector Machines,  
<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [68] mRMR Feature Selection Site,  
<http://research.janelia.org/peng/proj/mRMR/>
- [69] VASCONCELOS N., HO P., MORENO P.: ‘The Kullback-Leibler kernel as a framework for discriminant and localized representations for visual recognition’. In *Proc. of the 8th European Conf. on Computer Vision (ECCV’04)*, 2004, *Lecture Notes in Computer Science*, 3023, pp. 430–441
- [70] LIU Y., PERRONNIN F.: ‘A similarity measure between unordered vector sets with application to image categorization’. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR’08)*, 2008, pp. 1–8
- [71] YAP K.-H., WU K.: ‘A soft relevance framework in content-based image retrieval systems’. *IEEE Trans. on Circuits and Systems for Video Technology*, 2005, 15, (12), pp. 1557–1568
- [72] BURGHOOTS G. J., GEUSEBROEK J. M.: ‘Performance evaluation of local colour invariants’. *Computer Vision and Image Understanding*, 2009, 113, pp. 48–62
- [73] KE Y., SUKTHANKAR R.: ‘PCA-SIFT: a more distinctive representation for local image descriptors’. In *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 2004, 2, pp. 506–513
- [74] MIKOLAJCZYK K., SCHMID C.: ‘A performance evaluation of local descriptors’. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2005, 27, (10), pp. 1615–1630
- [75] PERRONNIN F., DANCE C., CSURKA G., BRESSAN M.: ‘Adapted vocabularies for generic visual categorization’. In *Proc. of the 9th European Conference on Computer Vision (ECCV’06)*, 2006, *Lecture Notes in Computer Science*, 3954, pp. 464–475
- [76] FARQUHAR J., SZEDMAK S., MENG H., SHAW-TAYLOR J.: ‘Improving “bag-of-keypoints” image categorisation: generative models and PDF-kernels’. *Technical report, University of Southampton*, 2005

- [77] REYNOLDS D., QUATIERI T., DUNN R.: ‘Speaker verification using adapted gaussian mixture models’. *Digital Signal Processing*, 2000, 10, (1-3), pp. 19–41
- [78] GAUVAIN J.-L., LEE C.-H.: ‘Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains’. *IEEE Trans. on Speech and Audio Processing*, 1994, 2, (2), pp. 291–298
- [79] Homepage of Jan-Mark Geusebroek,  
<http://staff.science.uva.nl/~mark/downloads.html>
- [80] GEUSEBROEK J.M., BOOMGAARD R., SMEULDERS A.W.M., GEERTS H.: ‘Color invariance’. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001, 23, (12), 1338–1350
- [81] LIU C., RUBIN D.B.: ‘ML estimation of the t distribution using EM and its extensions, ECM and ECME’. *Statistica Sinica*, 1995, 5, pp. 19–39
- [82] PEEL D., McLACHLAN G. J.: ‘Robust mixture modelling using the t distribution’. *Statistics and Computing*, 2000, 10, pp. 339–348
- [83] CHEN S., WANG H., LUO B.: ‘Greedy EM algorithm for robust t-mixture modeling’. In *Proc. of the 3rd Int. Conf. on Image and Graphics*, 2004, pp. 548–551
- [84] <http://en.wikipedia.org/wiki/>
- [85] TIPPING M. E.: ‘Sparse bayesian learning and the relevance vector machine’. *Journal of Machine Learning Research*, 2001, 1, pp. 211–244
- [86] GOLDBERGER J., GREENSPAN H.: ‘Context-based segmentation of image sequences’. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006, 28, (3), pp. 463–468
- [87] DOULAMIS A.D., DOULAMIS N.D.: ‘Generalized nonlinear relevance feedback for interactive content-based retrieval and organization’. *IEEE Trans. on Circuits and Systems for Video Technology*, 2004, 14, (5), pp. 656–671
- [88] GOLDBERGER J., ROWEIS S.: ‘Hierarchical clustering of a mixture model’. In *Proc. of Advances in Neural Information Processing Systems 17 (NIPS’04)*, 2004, pp. 505–512
- [89] CARNEIRO G., CHAN A.B., MORENO P.J., VASCONCELOS N.: ‘Supervised learning of semantic classes for image annotation and retrieval’. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007, 29, (3), pp. 394–410
- [90] RASIWASIA N., MORENO P.J., VASCONCELOS N.: ‘Bridging the gap: query by semantic example’. *IEEE Trans. on Multimedia*, 2007, 9, (5), pp. 923–938
- [91] JING F., LI M., ZHANG H.-J., ZHANG B.: ‘A unified framework for image retrieval using keyword and visual features’. *IEEE Trans. on Image Processing*, 2005, 14, (7), pp. 979–989

- [92] HOI S.C.H., LYU M.R., JIN R.: ‘A unified log-based relevance feedback scheme for image retrieval’. *IEEE Trans. on Knowledge and Data Engineering*, 2006, 18, (4), pp. 509–524
- [93] HAN J., NGAN K.N., LI M., ZHANG H.-J.: ‘A memory learning framework for effective image retrieval’. *IEEE Trans. on Image Processing*, 2005, 14, (4), pp. 511–524
- [94] HE X., KING O., MA W.-Y., LI M., ZHANG H.-J.: ‘Learning a semantic space from user’s relevance feedback for image retrieval’. *IEEE Trans. on Circuits and Systems for Video Technology*, 2003, 13, (1), pp. 39–48
- [95] LU Y., ZHANG H., WENYIN L., HU C.: ‘Joint semantics and feature based image retrieval using relevance feedback’. *IEEE Trans. on Multimedia*, 2003, 5, (3), pp. 339–347



## Δημοσιεύσεις

### Περιοδικά

- [J1] MARAKAKIS A., GALATSANOS N., LIKAS A., STAFYLOPATIS A.: ‘Probabilistic relevance feedback approach for content-based image retrieval based on gaussian mixture models’. IET Image Processing, vol. 3, no. 1, pp. 10–25, February 2009
- [J2] MARAKAKIS A., GALATSANOS N., LIKAS A., STALYLOPATIS A.: ‘Using adapted gaussian mixture models and SVM classifiers for relevance feedback in content-based image retrieval’. (submitted)

### Συνέδρια

- [C1] MARAKAKIS A., GALATSANOS N., LIKAS A., STAFYLOPATIS A.: ‘A relevance feedback approach for content based image retrieval using gaussian mixture models’. In Proc. of the 16th International Conference on Artificial Neural Networks (ICANN’06), Lecture Notes in Computer Science, vol. 4132, pp. 84–93, Athens, Greece, 2006
- [C2] MARAKAKIS A., GALATSANOS N., LIKAS A., STAFYLOPATIS A.: ‘Application of relevance feedback in content based image retrieval using gaussian mixture models’. In Proc. of the 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI’08), vol. 1, pp. 141–148, Dayton, Ohio, USA, 2008
- [C3] MARAKAKIS A., GALATSANOS N., LIKAS A., STAFYLOPATIS A.: ‘Combining gaussian mixture models and support vector machines for relevance feedback in content based image retrieval’. In Proc. of the 5th IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI’09), IFIP International Federation for Information Processing, vol. 296, Artificial Intelligence Applications and Innovations III, eds. L. Iliadis, I. Vlahavas, M. Bramer (Springer, Boston), pp. 249–258, Thessaloniki, Greece, 2009
- [C4] MARAKAKIS A., GALATSANOS N., LIKAS A., STAFYLOPATIS A.: ‘Relevance feedback for content-based image retrieval using support vector machines and feature selection’. In the 19th International Conference on Artificial Neural Networks (ICANN’09), Limassol, Cyprus, 2009



## Βιογραφικό Σημείωμα

Ο Απόστολος Μαρακάκης γεννήθηκε στα Χανιά της Κρήτης το Νοέμβριο του 1981. Αποφοίτησε από το 7<sup>ο</sup> Ενιαίο Λύκειο Χανίων με βαθμό «Άριστα» (19.6), ενώ το Νοέμβριο του 2005 απέκτησε το δίπλωμα του Ηλεκτρολόγου Μηχανικού και Μηχανικού Υπολογιστών με βαθμό «Άριστα» (9.36) από τη σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. Κατόπιν, τον ίδιο μήνα έγινε δεκτός ως υποψήφιος διδάκτορας στην παραπάνω σχολή, με ερευνητικό θέμα στην περιοχή της Υπολογιστικής Νοημοσύνης και των Ευφυών Συστημάτων.

Επίσης, προσέφερε επικουρικό εργαστηριακό έργο στο πλαίσιο των μαθημάτων «Νευρωνικά Δίκτυα και Ευφυή Υπολογιστικά Συστήματα» (9<sup>ο</sup> εξάμηνο) και «Προγραμματιστικές Τεχνικές» (2<sup>ο</sup> εξάμηνο), με την παρακολούθηση και εργαστηριακή υποστήριξη σπουδαστών κατά την εκπόνηση εργασιών.

### Επαγγελματικά Σωματεία

2006 -           || Μέλος του Τεχνικού Επιμελητηρίου Ελλάδος.  
                  ||

### Εθελοντικοί Επιστημονικοί Ρόλοι

2009           || Κριτής άρθρων συνεδρίου “International Conference on Artificial Neural  
                  || Networks (ICANN 2009)”

                  ||  
2009           || Κριτής άρθρων περιοδικού “Neural Networks (Elsevier)”  
                  ||

### Συμμετοχή σε Ερευνητικά Προγράμματα

2005-2009    || Πρόγραμμα ΠΕΝΕΔ2003/ΕΠΑΝ, με κωδικό 03ΕΔ32 και τίτλο «Τεχνικές  
                  || Μηχανικής Μάθησης για Διαχείριση Γνώσης σε Πολυμεσικά Δεδομένα»,  
                  || Γενική Γραμματεία Έρευνας και Τεχνολογίας. Χρηματοδοτούμενο από  
                  || την Ευρωπαϊκή Ένωση και από εθνικούς πόρους.  
                  ||

