



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ, ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΣΥΣΤΗΜΑΤΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΤΝΗΣ ΚΑΙ ΜΑΘΗΣΗΣ

Αυτόματη Δημιουργία Λιστών Αναπαραγωγής Μουσικών
Κομματιών με Τεχνικές Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Δαγκλή Σ. Θεοφάνη

Επιβλέπων: Γεώργιος Στάμου

Καθηγητής

Αθήνα, Φεβρουάριος 2022



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ, ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΣΥΣΤΗΜΑΤΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΚΑΙ ΜΑΘΗΣΗΣ

Αυτόματη Δημιουργία Λιστών Αναπαραγωγής Μουσικών Κομματιών με Τεχνικές Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Θεοφάνη Σ. Δαγκλή

Επιβλέπων: Γεώργιος Στάμου
Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 22/03/2022.

.....
Αθανάσιος Βουλόδημος
Επίκουρος Καθηγητής

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής

.....
Γεώργιος Στάμου
Καθηγητής

Αθήνα, Μάρτιος 2022.

.....
Θεοφάνης Σ. Δαγκλής

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών

Copyright©Δαγκλής Θεοφάνης, 2022.

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 τους συγγραφείς.

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.

Περίληψη

Η ιδέα της νοημοσύνης των μηχανών απασχόλησε την επιστημονική και τεχνική σκέψη ήδη από το 1950. Το σημερινό επιστημονικό και τεχνικό υπόβαθρο έχει καταστήσει την Τεχνητή Νοημοσύνη σε ένα σημαντικό επιστημονικό πεδίο με ευρύτατη εφαρμογή. Η ραγδαία αύξηση της υπολογιστικής ισχύος, η βελτίωση των συσκευών αποθήκευσης, τόσο σε όγκο αλλά και σε ταχύτητα, αλλά και η διάδοση της χρήσης αισθητήρων καθιστούν δυνατή την εφαρμογή τεχνικών μηχανικής μάθησης που ήταν απαγορευτικές στο παρελθόν.

Ένα πεδίο της μηχανικής μάθησης με πολύ έντονο ερευνητικό και οικονομικό ενδιαφέρον και ευρεία εφαρμογή αποτελεί η διαμόρφωση εξατομικευμένων προτάσεων. Τα συστήματα προτάσεων (recommendation systems) έχουν ως στόχο την διαλογή αντικειμένων που ενδέχεται να ενδιαφέρουν έναν συγκεκριμένο χρήστη.

Η παρούσα εργασία καταπιάνεται με την διαμόρφωση προτάσεων μουσικών κομματιών βάσει λιστών αναπαραγωγής. Συγκεκριμένα αναλύονται, αναπτύσσονται και συγκρίνονται διαφορετικές τεχνικές μηχανικής μάθησης συνεργατικού φιλτραρίσματος, μιας οικογένειας αλγορίθμων που αποσκοπεί στην εύρεση κατάλληλων προτάσεων για έναν χρήστη, βάσει της αλληλεπίδρασης του με ένα σύνολο αντικειμένων.

Η πρώτη προσέγγιση εφαρμόζει μία ευρέως διαδεδομένη τεχνική βασιζόμενη στην παραγοντοποίηση πινάκων, ενώ η δεύτερη είναι η ανάπτυξη μίας σειράς μοντέλων που βασίζονται στους Αυτοκωδικοποιητές, εκκινώντας από τον απλό Αυτοκωδικοποιητή (Autoencoder), συνεχίζοντας με τον Αυτοκωδικοποιητή Αποθορυβοποίησης (Denoising Autoencoder) και καταλήγοντας στον Αυτοκωδικοποιητή Διακύμανσης (Variational Autoencoders).

Σκοπός της εργασίας είναι η εξέταση της αποτελεσματικότητας των Αυτοκωδικοποιητών Διακύμανσης στην διαμόρφωση προτάσεων μουσικών κομματιών σε λίστες αναπαραγωγής, καθώς και η σύγκριση με τη χρήση της ευρέως διαδεδομένης τεχνικής της Παραγοντοποίησης Πινάκων. Επιπλέον, τα αποτελέσματα οπτικοποιούνται με σκοπό την καλύτερη εποπτεία και τη δυνατότητα σύγκρισης και εξαγωγής ποιοτικών συμπερασμάτων.

Λέξεις κλειδιά: Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Νευρωνικά Δίκτυα, Συστήματα Προτάσεων, Συνεργατικό Φιλτράρισμα, Αυτοκωδικοποιητές Διακύμανσης, Παραγοντοποίηση Πινάκων

Abstract

The idea of intelligent machines has been occupying scientists' and engineers' thoughts even since the 1950s. Modern scientific and technical knowledge base has made artificial intelligence into a significant scientific field with broad scope and application. Rapid improvement in computational power, data storage volume and speed as well as the widespread sensor usage in a multitude of applications enable the application of advanced machine learning techniques infeasible in past years.

Personalized recommendation systems are a rapidly growing field of great scientific and financial importance in machine learning. Recommendation systems aim to select items that might be of interest to specific users.

In this diploma thesis, music recommendations are generated in the form of music track playlists. Specifically, two different collaborative filtering machine learning methods are analyzed, implemented, and compared. Both methods belong to an algorithm family aiming to generate personalized recommendation for a specific user based on their interaction with an item set.

The first approach applies a popular method of matrix factorization, while the second approach is the development of a series of models, starting with the simple Autoencoder, continuing with the Denoising Autoencoder, and finally describing a Variational Autoencoder.

This diploma thesis aims to evaluate the effectiveness of the Variational Autoencoder method in generating recommendations of music playlists, as well as compare this method with the popular matrix factorization method. Furthermore, results are visualized to allow for easier comparison and qualitative conclusion.

Keywords: Artificial Intelligence, Machine learning, Neural Networks, Recommendation Systems, Collaborative Filtering, Variational Autoencoders, Matrix Factorization.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή Γεώργιο Στάμου και το Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης για τη δυνατότητα που μου δόθηκε να εργαστώ πάνω στο συγκεκριμένο θέμα. Ιδιαίτερα θα ήθελα να ευχαριστήσω τον Μιχάλη Γιαζιτζόγλου για την πολύτιμη βοήθεια του.

Θα ήθελα επίσης να ευχαριστήσω όλους αυτούς τους ανθρώπους που με συντρόφευσαν και με βοήθησαν σε αυτό το μεγάλο ταξίδι και που τα ονόματά τους δεν θα μπορούσαν να χωράνε σε αυτή τη μικρή παράγραφο.

Τέλος, θα ήθελα να ευχαριστήσω την μητέρα μου Μαρία και την αδερφή μου Βασιλική.

Περιεχόμενα

Περιεχόμενα	14
1 Βασικές Έννοιες	21
1.1 Τεχνητή Νοημοσύνη	21
1.2 Οι επιστημονικές βάσεις της τεχνητής νοημοσύνης	21
1.3 Συμβολική Τεχνητή Νοημοσύνη (Symbolic Artificial Intelligence)	22
1.4 Υποσυμβολική Τεχνητή Νοημοσύνη (Subsymbolic Artificial Intelligence)	23
1.5 Νευρο-Συμβολική Τεχνητή Νοημοσύνη (Neuro-Symbolic Artificial Intelligence)	23
1.6 Μηχανική Μάθηση	23
1.7 Νευρωνικά Δίκτυα	24
1.8 Βαθιά Νευρωνικά Δίκτυα	25
1.9 Ο απλός Perceptron ή Αντίληπτρο	25
1.10 Multilayer Perceptron - MLP	26
1.10.1 Περιγραφή	26
1.10.2 Η συνάρτηση ενεργοποίησης	27
1.10.3 Η συνάρτηση στόχου ή συνάρτηση κόστους	29
1.11 Μετρικές Απόδοσης (Performance Metrics)	31
2 Συστήματα Προτάσεων/Συστάσεων - Recommendation Systems	35
2.1 Κατηγορίες Συστημάτων Προτάσεων	35
2.1.1 Φιλτράρισμα με βάση το Περιεχόμενο - Content Based Filtering	35
2.1.2 Συνεργατικό Φιλτράρισμα - Collaborative Filtering	35
2.2 Τεχνικές υλοποίησης	36
2.2.1 Συστήματα βασισμένα στη μνήμη	36
2.2.2 Συστήματα βασισμένα σε μοντέλο	37

2.2.3	Συστήματα με χρήση τεχνικών βαθιάς μηχανικής μάθησης	37
2.3	Συnergατικό φίλτράρισμα με Παραγοντοποίηση Πινάκων	37
2.4	Περιγραφή της μεθόδου	38
2.5	Συνάρτηση Κόστους	38
2.5.1	Η μέθοδος των εναλλασσόμενων ελάχιστων τετραγώνων	39
2.6	Ο υπολογισμός των προτάσεων	40
3	Η συμπερασματολογία του Bayes	41
3.1	Στοιχεία Θεωρίας Πληροφορίας	42
3.2	Το θεώρημα Bayes	43
3.3	Εφαρμογή του κανόνα του Bayes	44
3.3.1	Βασικές έννοιες	44
3.4	Εκτίμηση Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimation)	45
3.5	Variational Inference	46
3.5.1	Βασική Προσέγγιση	47
3.5.2	Υπολογισμός της προσεγγιστικής κατανομής με τη χρήση της απόκλισης Kullback Leibler	47
4	Αυτοκωδικοποιητές Διακύμανσης (Variational Autoencoders)	49
4.1	Απλοί Αυτοκωδικοποιητές (Autoencoders)	49
4.1.1	Βασική Αρχιτεκτονική	49
4.1.2	Πεδία εφαρμογής των Αυτοκωδικοποιητών	50
4.1.3	Αυτοκωδικοποιητές Αποθορυβοποίησης (Denoising Autoencoders)	50
4.2	Αυτοκωδικοποιητές Διακύμανσης (Variational Autoencoders)	51
4.2.1	Προσέγγιση του Προβλήματος με χρήση τυχαίων μεταβλητών	51
4.2.2	Προσέγγιση της Μεταγενέστερης Κατανομής	52
4.3	Evidence Lower Bound	53
4.3.1	Το ELBO ως συνάρτηση σφάλματος	53
4.3.2	Μεγιστοποίηση του ELBO	53
4.3.3	Το τρικ επαναπαμετροποίησης (Reparameterization Trick)	54
4.4	Η υλοποίηση ενός Variational Autoencoder	54
4.4.1	Η πλευρά του κωδικοποιητή	55

4.4.2	Η προηγούμενη κατανομή και η δειγματοληψία της τυχαίας μεταβλητής z	55
4.4.3	Η πλευρά του αποκωδικοποιητή	55
4.4.4	Η διαδικασία της εκπαίδευσης	56
5	Σύνολο Δεδομένων	57
5.1	Εύρεση λιστών αναπαραγωγής	57
5.2	Συλλογή και αποθήκευση των δεδομένων	57
5.3	Το σύνολο δεδομένων	58
5.4	Χαρακτηριστικά Μουσικών Κομματιών (Track Features)	63
5.5	Συλλογή ετικετών (tags) μουσικών κομματιών	68
6	Υλοποίηση Μοντέλων	69
6.1	Παραγοντοποίηση Πινάκων με Εναλλασσόμενα Τετράγωνα	69
6.1.1	Πίνακες χαρακτηριστικών/παραγόντων	69
6.1.2	Η διαδικασία εκπαίδευσης	69
6.1.3	Υπολογισμός Προτάσεων	70
6.1.4	Υπολογισμός των Μετρικών στα Σύνολα Δεδομένων Επιθύρωσης και Ελέγχου	70
6.1.5	Το κριτήριο επιλογής βέλτιστου μοντέλου	71
6.1.6	Βελτιστοποίηση υπερπαραμέτρων με τη χρήση Grid Search	71
6.2	Αυτοκωδικοποιητές	74
6.2.1	Απλοί Αυτοκωδικοποιητές και Αυτοκωδικοποιητές Αποθρομβοποίησης	74
6.2.2	Αποκωδικοποιητής	74
6.2.3	Διαδικασία Εκπαίδευσης	74
6.2.4	Αυτοκωδικοποιητές Διακύμανσης	79
6.2.5	Η επίδραση του σταδίου Dropout και του πολλαπλασιαστή β της απόκλισης Kullback Leibler	85
7	Αποτελέσματα και Οπτικοποίηση Δεδομένων	87
7.1	Αποτελέσματα	87
7.2	Οπτικοποίηση δεδομένων με βάση το μουσικό είδος των κομματιών	88
7.3	Ανάλυση χαρακτηριστικών	93
7.3.1	Ανάλυση σε Κύριες Συνιστώσες	93

7.3.2	Συντελεστές Γραμμικής Συσχέτισης	94
7.3.3	Ποιοτική Ανάλυση με Συστατοδοποίηση	95
7.4	Κατεύθυνση Προτάσεων με Ανάλυση Κύριων Συνιστωσών	96
7.5	Σύστημα Προτάσεων	97
8	Επίλογος	99
8.1	Συμπεράσματα	99
8.2	Μελλοντική Έρευνα	100

Κατάλογος Σχημάτων

1.1	Η αρχιτεκτονική του Αντιλήπτρου	25
1.2	Η αρχιτεκτονική του Multilayer Perceptron	27
1.3	Stochastic Gradient Descent vs Gradient Descent	30
1.4	Υπολογισμός της Ακρίβειας (Precision) και της Ανάκλησης (Recall) σε έναν πίνακα σύγχυσης (confusion matrix).	32
3.1	Οι σχέσεις μεταξύ των κατανομών, με βάση την μπεϋζιανή στατιστική	45
4.1	Το στατιστικό μοντέλο Bayes	52
5.1	Παραδείγματα εγγράφων στο Mongoddb Compass	58
5.2	Ο αριθμός των λιστών αναπαραγωγής στο μεγάλο σύνολο δεδομένων βάσει του αριθμού των κομματιών που περιέχουν.	61
5.3	Ο αριθμός των κομματιών στο μεγάλο σύνολο δεδομένων βάσει του αριθμού των λιστών αναπαραγωγής που αυτά περιέχονται.	62
5.4	Ο αριθμός των λιστών αναπαραγωγής στο μικρό σύνολο δεδομένων βάσει το σύνολο των κομματιών που περιέχουν.	62
5.5	Ο αριθμός των κομματιών στο μικρό σύνολο δεδομένων βάσει του αριθμού των λιστών αναπαραγωγής που περιέχονται.	63
5.6	Διαχωρισμός των μουσικών κομματιών ως προς την ταχύτητα	64
5.7	Διαχωρισμός των μουσικών κομματιών ως προς τη χορευτική δυνατότητα	64
5.8	Διαχωρισμός των μουσικών κομματιών ως προς την ενέργεια	65
5.9	Διαχωρισμός των μουσικών κομματιών ως προς την ακουστικότητα	65
5.10	Διαχωρισμός των μουσικών κομματιών ως προς την ορχηστρικήτητα	66
5.11	Διαχωρισμός των μουσικών κομματιών ως προς την ένταση	66
5.12	Διαχωρισμός των μουσικών κομματιών ως προς τη χρήση λέξεων	67
6.1	Διάγραμμα εκπαίδευσης του βέλτιστου μοντέλου παραγαντοποίησης πινάκων για το μεγάλο σύνολο δεδομένων	72

6.2	Διάγραμμα εκπαίδευσης του βέλτιστου μοντέλου παραγοντοποίησης πινάκων για το μικρό σύνολο δεδομένων	73
6.3	Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή για το μεγάλο σύνολο δεδομένων	76
6.4	Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή για το μικρό σύνολο δεδομένων	77
6.5	Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή Αποθροβοποίησης για το μεγάλο σύνολο δεδομένων	78
6.6	Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή Αποθροβοποίησης για το μικρό σύνολο δεδομένων	79
6.7	Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή Διακύμανσης για το μεγάλο σύνολο δεδομένων	83
6.8	Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή Διακύμανσης για το μικρό σύνολο δεδομένων	85
7.1	Αποτελέσματα στα δύο σύνολα δεδομένων ελέγχου	88
7.2	Οπτικοποίηση του μοντέλου παραγωγής πινάκων δύο χαρακτηριστικών	89
7.3	Οπτικοποιήσεις τυχαίων χαρακτηριστικών του βέλτιστου μοντέλου παραγοντοποίησης πινάκων	89
7.4	Οπτικοποίηση του μοντέλου δύο διαστάσεων ενός απλού αυτοκωδικοποιητή	90
7.5	Οπτικοποιήσεις τυχαίων χαρακτηριστικών του βέλτιστου μοντέλου απλού αυτοκωδικοποιητή	91
7.6	Οπτικοποίηση ενός μοντέλου δύο διαστάσεων DAE	91
7.7	Οπτικοποιήσεις τυχαίων χαρακτηριστικών του βέλτιστου μοντέλου DAE	92
7.8	Οπτικοποίηση του μοντέλου VAE	92
7.9	Οπτικοποιήσεις τυχαίων χαρακτηριστικών του βέλτιστου μοντέλου VAE	93
7.10	Διαχωρισμός σε κλάσεις ενός χαρακτηριστικού στις δύο βέλτιστες συνιστώσες	96
7.11	Η επίδραση στις προτάσεις των δύο πρώτων συνιστωσών με τιμή 20	97
7.12	Η επίδραση στις προτάσεις των δύο πρώτων συνιστωσών με τιμή -20	97

Κατάλογος Πινάκων

5.1 Διαχωρισμός των λιστών αναπαραγωγής του μεγάλου συνόλου δεδομένων βάσει του αριθμού των μουσικών κομματιών που περιέχουν.	61
5.2 Διαχωρισμός των μουσικών κομματιών του μεγάλου συνόλου δεδομένων βάσει του αριθμού των λιστών αναπαραγωγής που αυτά περιέχονται	61
5.3 Διαχωρισμός των λιστών αναπαραγωγής του μικρού συνόλου δεδομένων βάσει του αριθμού των μουσικών κομματιών που περιέχουν.	62
5.4 Διαχωρισμός των μουσικών κομματιών του μικρού συνόλου δεδομένων βάσει του αριθμού των λιστών αναπαραγωγής που αυτά περιέχονται	62
5.5 Πίνακας του διαχωρισμού των μουσικών κομματιών ανά κατηγορία μουσικού είδους	68
6.1 Τα αποτελέσματα της μεθόδου παραγοντοποίησης πινάκων στην αναζήτηση Grid για το μεγάλο σύνολο δεδομένων	72
6.2 Τα αποτελέσματα της μεθόδου παραγοντοποίησης πινάκων στην αναζήτηση Grid για το μικρό σύνολο δεδομένων	73
6.3 Τα αποτελέσματα των Απλών Αυτοκωδικοποιητών στην αναζήτηση Grid για το μεγάλο σύνολο δεδομένων	76
6.4 Τα αποτελέσματα των Απλών Αυτοκωδικοποιητών στην αναζήτηση Grid για το μικρό σύνολο δεδομένων	76
6.5 Τα αποτελέσματα των Αυτοκωδικοποιητών Αποθρομβοποίησης στην αναζήτηση Grid για το μεγάλο σύνολο δεδομένων	77
6.6 Τα αποτελέσματα των Αυτοκωδικοποιητών Αποθρομβοποίησης στην αναζήτηση Grid για το μικρό σύνολο δεδομένων	78
6.7 Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μεγάλο σύνολο δεδομένων, με σταδιακή αύξηση του β από το 0 έως το 0.2.	82
6.8 Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μεγάλο σύνολο δεδομένων, με σταδιακή αύξηση του β από το 0 έως το 1.	83
6.9 Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μεγάλο σύνολο δεδομένων, με β σταθερό και ίσο με 20.	83
6.10 Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μικρό σύνολο δεδομένων, με σταδιακή αύξηση του β από το 0 έως το 0.2.	84

6.11 Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μικρό σύνολο δεδομένων, με σταδιακή αύξηση του β από το 0 έως το 1.	84
6.12 Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μικρό σύνολο δεδομένων, με β σταθερό και ίσο με 20.	84
7.1 Αποτελέσματα στα δύο σύνολα δεδομένων ελέγχου	87
7.2 Ποσοστό της διακύμανσης που εξηγείται από την ανάλυση σε 10 κύριες συνιστώσες	94
7.3 Συντελεστές συσχέτισης Pearson για το βέλτιστο μοντέλο ALS	94
7.4 Συντελεστές συσχέτισης Pearson για το βέλτιστο μοντέλο AE	94
7.5 Συντελεστές συσχέτισης Pearson για το βέλτιστο μοντέλο DAE	94
7.6 Συντελεστές συσχέτισης Pearson για το βέλτιστο μοντέλο VAE	95

Εισαγωγή

Η μορφή της κοινωνίας σήμερα και όλες οι πτυχές της οικονομικής, κοινωνικής, πολιτιστικής και πολιτικής ζωής έχουν αδιαμφισβήτητα καθοριστεί από την ανάπτυξη και την ευρεία διάδοση του διαδικτύου. Η ανάπτυξη των κοινωνικών δικτύων έχει επαναπροσδιορίσει την κοινωνικοποίηση και τη διαμόρφωση των ανθρώπινων σχέσεων. Η δυνατότητα πρόσβασης στο συλλογικό γνωσιακό κεκτημένο της ανθρωπότητας και η άμεση διαθεσιμότητα της πληροφορίας συνιστά μια ποιοτική τομή στα πρότερα κοινωνικά δεδομένα. Η οικονομική δραστηριότητα έχει επίσης επανακαθοριστεί σε μεγάλο βαθμό τόσο μέσα από τη ψηφιακή επέκταση παραδοσιακών οικονομικών διεργασιών όπως η διαφήμιση και το εμπόριο. Ταυτόχρονα, έχει αναπτυχθεί ένα μεγάλο πεδίο αμιγώς ψηφιακής κοινωνικής και οικονομικής δραστηριότητας, με χαρακτηριστικότερο ίσως παράδειγμα το Youtube. Ειδική μνεία αξίζει το ολόένα και αυξανόμενο κομμάτι χρηστών που είναι παράλληλα και παραγωγοί περιεχομένου, κάτι που δεν απαιτεί παρά ελάχιστα τεχνικά μέσα και συχνά περιγράφεται με τον όρο prosumer (από τον συνδυασμό των λέξεων παραγωγού και καταναλωτή). Ενδεικτικά παραδείγματα είναι η διαδικτυακή εγκυκλοπαίδεια Wikipedia και η εφαρμογή κοινωνική δικτύωσης Instagram.

Η πραγματικότητα όπως παραπάνω περιγράφεται έχει ως απότοκο, μεταξύ άλλων, την παραγωγή τεράστιου όγκου πληροφορίας και δεδομένων. Ενδεικτικά, τα διαθέσιμα μουσικά κομμάτια στο Spotify είναι πάνω από 50 εκατομμύρια, τα διαθέσιμα video στο Youtube είναι πάνω από 50 δισεκατομμύρια, ενώ μόνο η αγγλική σελίδα της Wikipedia αριθμεί πάνω από 6 εκατομμύρια άρθρα.

Προκύπτει έτσι η ανάγκη διαλογής επιθυμητών, ψηφιακών ή μη, αντικειμένων ως εξατομικευμένες προτάσεις, ώστε ο χρήστης να αλληλεπιδρά με νέο περιεχόμενο και να ανακαλύπτει νέες υπηρεσίες και προϊόντα που ταιριάζουν στις προτιμήσεις του και ενδέχεται να τον ενδιαφέρουν. Η ανάγκη αυτή έχει οδηγήσει στην ανάπτυξη Συστημάτων Προτάσεων (Recommendation Systems), ιδιαίτερα σε πεδία όπου ο αριθμός των επιλογών καθιστά αναποτελεσματικές άλλες παλαιότερες, πιο παραδοσιακές μεθόδους.

Αντικείμενο της εργασίας

Στην παρούσα εργασία επιχειρείται η ανάπτυξη δύο, αρκετά διαφορετικών μεταξύ τους, μοντέλων που στόχο έχουν τη δημιουργία λιστών αναπαραγωγής με βάση τα προϋπάρχοντα μουσικά κομμάτια που αυτή περιέχει. Το κριτήριο με το οποίο γίνονται οι νέες προτάσεις κομματιών είναι το συνεργατικό φίλτράρισμα, δηλαδή η ομοιότητα τους βάσει τις προτιμήσεις άλλων χρηστών (στη συγκεκριμένη περίπτωση, άλλων λιστών αναπαραγωγής).

Για την επίτευξη της εκπαίδευσης χρειάστηκε να δημιουργηθεί το σύνολο δεδομένων εξ' αρχής, αφού κατά την εκκίνηση της εργασίας δεν ήταν δυνατό να βρεθεί αντίστοιχο που να εξυπηρετούσε της ανάγκες της εργασίας.

Στόχος της εργασίας είναι η δυνατότητα εκπαίδευσης των μοντέλων, λαμβάνοντας ως δεδομένο μόνο την ύπαρξη ή μη ενός κομματιού μέσα στη λίστα αναπαραγωγής, χωρίς τη χρήση κανενός άλλου μεταδεδομένου, καθώς και η εφαρμογή ενός μοντέλου βαθιάς μηχανικής μάθησης σε ένα πεδίο εφαρμογής που τα τελευταία χρόνια υπάρχει όλο και πιο έντονο ερευνητικό ενδιαφέρον.

Διάρθρωση της Εργασίας

Η παρούσα εργασία διαρθρώνεται σε 7 κεφάλαια. Το πρώτο κεφάλαιο περιέχει βασικές έννοιες και ορισμούς καθώς και περιγραφή κάποιων γενικών τεχνικών.

Το δεύτερο κεφάλαιο περιγράφει όλα τα βήματα που πραγματοποιήθηκαν για την εξαρχής κατασκευή του συνόλου δεδομένων, από τη συλλογή των λιστών αναπαραγωγής και των μουσικών κομματιών που περιέχονται σε αυτή μέχρι την κατασκευή των συνόλων δεδομένων. Επιπλέον, περιγράφονται οι διαδικασίες που πραγματοποιήθηκαν για την συλλογή και επεξεργασία κάποιων χαρακτηριστικών των μουσικών κομματιών και των ετικετών τους που περιγράφει.

Το τρίτο κεφάλαιο παρουσιάζει τις διαφορετικές προσεγγίσεις στα συστήματα προτάσεων και περιγράφει αναλυτικά το Συνεργατικό Φιλτράρισμα, ως την βασική λογική των εφαρμογών που αναπτύχθηκαν. Επιπλέον αναφέρεται σε μία οικογένεια ευρέως διαδεδομένων τεχνικών για την ανάπτυξη συστημάτων προτάσεων, αυτής της Παραγαντοποίησης Πινάκων. Στη συνέχεια περιγράφεται ένας συγκεκριμένος τρόπος εκπαίδευσης, αυτής των Εναλλασσόμενων Ελαχίστων Τετραγώνων.

Το τέταρτο κεφάλαιο είναι προπαρασκευαστικό, περιγράφοντας βασικές έννοιες της στατιστικής και της συμπερασματολογίας Gauss. Το πέμπτο κεφάλαιο εισάγει την έννοια των Αυτοκωδικοποιητών Διακύμανσης, περιγράφοντας το θεωρητικό τους υπόβαθρο.

Το έκτο κεφάλαιο περιγράφει τις υλοποιήσεις των δύο μοντέλων, τη διαδικασία εκπαίδευσης και τον τρόπο επιλογής του βέλτιστου μοντέλου. Το έβδομο κεφάλαιο παρουσιάζει τα αποτελέσματα των δύο μοντέλων, τη μεταξύ τους σύγκριση και παρουσιάζονται τα συμπεράσματα που προέκυψαν.

Τέλος, το έβδομο κεφάλαιο συνοψίζει τα αποτελέσματα και περιγράφει τα συμπεράσματα που προέκυψαν κατά την εκπόνηση της εργασίας.

Κεφάλαιο 1

Βασικές Έννοιες

Στο παρόν κεφάλαιο γίνεται η προσπάθεια να περιγραφούν βασικές έννοιες των νευρωνικών δικτύων καθώς και να παρουσιαστούν τεχνικές που κρίθηκε αναγκαίο να αναλυθούν και δεν εντάσσονταν σε κάποιο από τα υπόλοιπα κεφάλαια.

1.1 Τεχνητή Νοημοσύνη

Η τεχνητή νοημοσύνη αποτελεί ένα ευρύ επιστημονικό πεδίο που προσπαθεί να προσεγγίσει την έννοια της νοημοσύνης και να αναπτύξει τεχνικές εφαρμογής της. Υπάρχουν αρκετές διαφορετικές προσεγγίσεις για τον ορισμό της, έναν ορισμό που εξελίσσεται με την πάροδο των χρόνων και την αλλαγή των κυρίαρχων παραδειγμάτων. Ένας αρκετά δημοφιλής ορισμός δίνεται από το σύγγραμμα *Artificial Intelligence, Foundations of Computational Agents* των David L. Poole και Alan K. Mackworth[1]:

Η Τεχνητή Νοημοσύνη αποτελεί το επιστημονικό πεδίο εκείνο που συνθέτει και αναλύει υπολογιστικά υποκείμενα/πράκτορες που δρουν με ευφυΐα. Ένα υποκείμενο (ή πράκτορας) χαρακτηρίζεται ως οτιδήποτε μπορεί να αντιληφθεί το περιβάλλον του μέσα από αισθητήρες και να δράσει πάνω σε αυτό μέσω ενεργοποιητών. (Για παράδειγμα ένα υποκείμενο-ρομπότ, αντιλαμβάνεται το περιβάλλον του μέσα από κάμερες και επιδρά στο περιβάλλον με έναν βραχίονα).

Ένα υποκείμενο δρα με ευφυΐα όταν:

- Η δράση του είναι σχετική με τους στόχους του και τις συνθήκες που αυτό βρίσκεται.
- Παρουσιάζει ευελιξία και προσαρμόζεται στις αλλαγές του περιβάλλοντος και των στόχων.
- Μαθαίνει από την εμπειρία του.
- Παίρνει τις αντίστοιχες αποφάσεις, δεδομένων των περιορισμών σε αντίληψη και υπολογιστική ικανότητα.

Ο βασικός επιστημονικός στόχος της τεχνητής νοημοσύνης είναι η κατανόηση των αρχών που επιτρέπουν την δημιουργία ευφυούς συμπεριφοράς συστημάτων ή υποκειμένων.

1.2 Οι επιστημονικές βάσεις της τεχνητής νοημοσύνης

Παρακάτω περιγράφονται οι βασικές επιρροές της τεχνητής νοημοσύνης σύμφωνα με το εγχειρίδιο *Artificial Intelligence*, των Stuart Russell και Peter Norvig[2]:

Η τεχνητή νοημοσύνη έχει επηρεαστεί από διάφορους επιστημονικούς κλάδους. Ενδεικτικά:

Από τα **Μαθηματικά**, αξιοποιούνται η θεωρία της λογικής, η θεωρία πιθανοτήτων και η θεωρία υπολογισμού, με την χρήση σημαντικών θεωρημάτων και κριτηρίων και θεωρητικών μοντέλων, όπως το θεώρημα της μη πληρότητας του Kurt Goedel(1930), η θέση των Turing, Church για τις υπολογίσιμες συναρτήσεις, η μηχανή Turing και το θεώρημα Bayes.

Από την **Οικονομία** μέσα τη χρήση της θεωρίας αποφάσεων, που παρέχει ένα ολοκληρωμένο, τυπικό πλαίσιο για τη λήψη αποφάσεων κάτω από συνθήκες αβεβαιότητας, της θεωρίας παιγνίων και με την ανάπτυξη μεθόδων όπως η διαδικασία αποφάσεων Markov.

Από την **Νευροεπιστήμη**, με σαφή επιρροή στις πρώτες προσπάθειες ανάπτυξης των τεχνητών νευρωνικών δικτύων, μέσα από τις μελέτες γύρω από το νευρικό σύστημα του ανθρώπου, των βιολογικών νευρώνων και την συμβολή τους στις νοητικές διαδικασίες του ανθρώπου.

Από την **Ψυχολογία**, αξιοποιούνται προσεγγίσεις της γνωστικής ψυχολογίας, δηλαδή του κλάδου αυτού της ψυχολογίας που ερμηνεύει την ανθρώπινη συμπεριφορά μέσα από την μελέτη της αντίληψης, της μνήμης, της προσοχής, της νόησης, της δημιουργικότητας και τη χρήση της γλώσσας.

Από την **Επιστήμη των Υπολογιστών**, αξιοποιείται η θεωρία αλγορίθμων για την βελτίωση των μεθόδων υπολογισμού και η πολύ μεγάλη αύξηση της υπολογιστικής και αποθηκευτικής ισχύος, που επιτρέπει την εφαρμογή τεχνικών που δεν ήταν δυνατό στο παρελθόν λόγω του υπολογιστικού κόστους.

Από τον **Αυτόματο Έλεγχο**, με την εισαγωγή της έννοιας του σφάλματος (error) και τη χρήση μίας συνάρτησης στόχου (objective function) που πρέπει να ελαχιστοποιηθεί.

1.3 Συμβολική Τεχνητή Νοημοσύνη (Symbolic Artificial Intelligence)

Η συμβολική τεχνητή νοημοσύνη αποτελείται από ένα σύνολο τεχνικών που βασίζονται σε υψηλού επιπέδου αναπαραστάσεις (δηλαδή κατανοητές από τον άνθρωπο) για την περιγραφή των προβλημάτων, την λογική και την διαδικασία της αναζήτησης. Οι τεχνικές αυτές χρησιμοποιούν εργαλεία όπως προγραμματιστική λογική, κανόνες, σημασιολογικά δίκτυα και βάσεις γνώσης.

Οι τεχνικές της συμβολικής τεχνητής νοημοσύνης αποτέλεσαν το πρώτο κυρίαρχο παράδειγμα στην τεχνητή νοημοσύνη. Από τις αρχές της δεκαετίας του 1950 μέχρι τα τέλη της δεκαετίας του 1960 δημιουργήθηκε μία πληθώρα προγραμμάτων (όπως για παράδειγμα The Logic Theorist, το General Problem Solver, καθώς και προγράμματα που μάθαιναν να παίζουν σκάκι ή ντάμα σε αρκετά καλό επίπεδο), ενώ αναπτύχθηκε η υψηλού επιπέδου γλώσσα προγραμματισμού LISP, που χρησιμοποιήθηκε ευρέως για την συγγραφή προγραμμάτων Τεχνητής Νοημοσύνης.

Την περίοδο αυτή, ακολούθησε η πρώτη ύφεση, με το ερευνητικό ενδιαφέρον και τη χρηματοδότηση να περιορίζεται. Οι αντικειμενικοί παράγοντες που προκάλεσαν αυτή την κρίση, μπορούν να βρεθούν στους τεχνολογικούς περιορισμούς εκείνης της εποχής, όπως για παράδειγμα η μικρή υπολογιστική ισχύς και η περιορισμένη μνήμη και των αντικειμενικών χαρακτηριστικών κάποιων προβλημάτων, όπως ο εκθετικός χρόνος αύξησης των υπολογισμών και η απουσία του αντίστοιχου μαθηματικού πλαισίου (π.χ. η πολυπλοκότητα των αλγορίθμων) που αναπτύχθηκε κάποια χρόνια αργότερα. Στους υποκειμενικούς παράγοντες που συνέβαλαν στην ύφεση αυτή, ήταν η υπεραισιοδοξία των επιστημόνων, που οδήγησε σε αρκετά κακές εκτιμήσεις γύρω από τις δυνατότητες των πρώτων εφαρμογών τεχνητής νοημοσύνης και την πρόοδο της έρευνας και τη δημιουργία παράλογων προσδοκιών στο ευρύ κοινό, λόγω της μεγάλης δημοσιότητας που ακολούθησε τις πρώτες επιτυχίες.

Η δεύτερη εποχή της συμβολικής τεχνητής νοημοσύνης αναδύθηκε με την δημιουργία των συστημάτων γνώσης και των έμπειρων συστημάτων που αξιοποιούν την διαθέσιμη γνώση για την επίλυση πολύπλοκων προγραμμάτων. Ωστόσο, η έλλευση των προσωπικών υπολογιστών καθώς και τεχνικοί περιορισμοί στην ανάπτυξη των έμπειρων συστημάτων οδήγησε σε κρίση τον κλάδο της τεχνητής νοημοσύνης, με βαρύ οικονομικό αντίκτυπο μέσα από

την πτώχευση πολλών εταιριών σχεδιασμού έμπειρων συστημάτων.

1.4 Υποσυμβολική Τεχνητή Νοημοσύνη (Subsymbolic Artificial Intelligence)

Η υποσυμβολική τεχνητή νοημοσύνη προσεγγίζει τη νόηση μέσα από την οργάνωση και διασύνδεση στοιχειωδών υπολογιστικών κόμβων. Στα τέλη της δεκαετίας του 1980, μετά από την δεύτερη κρίση της τεχνητής νοημοσύνης, υπήρχε μεταστροφή του ερευνητικού ενδιαφέροντος σε στοιχειώδη υπολογιστικά μοντέλα και αλγόριθμους που είχαν περιγραφεί στο παρελθόν (όπως για παράδειγμα τα Αντίληπτρα και ο αλγόριθμος οπισθοδρομικής συνάρτησης σφάλματος).

Η δημιουργία, η ανάλυση και η εφαρμογή των νευρωνικών δικτύων αποτελούν ένα από τα βασικά κομμάτια υποσυμβολικής νοημοσύνης. Στη σημερινή εποχή, μαζί με τη μηχανική μάθηση, αποτελούν τα βασικά παραδείγματα ανάπτυξης τεχνητής νοημοσύνης.

1.5 Νευρο-Συμβολική Τεχνητή Νοημοσύνη (Neuro-Symbolic Artificial Intelligence)

Η νευρο-συμβολική τεχνητή νοημοσύνη αποτελεί ένα σχετικά νέο πεδίο έρευνας στη τεχνητή νοημοσύνη και επιχειρεί να συνδυάσει τις μοντέρνες τεχνικές νευρωνικών δικτύων (όπως για παράδειγμα τα βαθιά νευρωνικά δίκτυα) με τεχνικές συμβολικής τεχνητής νοημοσύνης και τη χρήση κανόνων.

Τα νευρωνικά δίκτυα είναι αρκετά αποτελεσματικά στην εξεύρεση χαρακτηριστικών από ένα σύνολο δεδομένων. Παρ' όλα αυτά, αυτά τα χαρακτηριστικά είναι τις περισσότερες φορές δύσκολο να συνδυαστούν με συγκεκριμένες έννοιες κατανοητές από τον άνθρωπο, κάτι που είναι κομβικό σε κλάδους που η ασφάλεια των συμπερασμάτων είναι σημαντική (όπως για παράδειγμα η ιατρική διάγνωση ή η αυτόνομη οδήγηση), καθώς η δυσκολία της ερμηνείας των αποτελεσμάτων μπορεί να αποτρέπει την διάγνωση της υπερεκπαίδευσης ενός μοντέλου.

Ο στόχος των νευρο-συμβολικών μοντέλων είναι η εξεύρεση των χαρακτηριστικών από ένα σύνολο δεδομένων και η μετέπειτα μετατροπή τους με τη χρήση τεχνικών συμβολικής τεχνητής νοημοσύνης. Τα συνδυαστικά αυτά μοντέλα συγκλίνουν ταχύτερα και χρειάζονται σε γενικές γραμμές μικρότερα σύνολα δεδομένων για την εκπαίδευσή τους.[3]

1.6 Μηχανική Μάθηση

Η μηχανική μάθηση αποτελεί έναν κλάδο της τεχνητής νοημοσύνης που ασχολείται με την ανάπτυξη αλγορίθμων, για την πραγματοποίηση της διαδικασίας προσαρμογής των παραμέτρων ενός μοντέλου μέσα από την επίδραση του περιβάλλοντος.

Οι αλγόριθμοι υλοποίησης της μηχανικής μάθησης σε γενικές γραμμές χωρίζονται σε τρεις κατηγορίες.

Επιβλεπόμενη Μάθηση

Στους αλγόριθμους επιβλεπόμενης μάθησης, δίνονται στο δίκτυο ζευγάρια διανυσμάτων εισόδου και επιθυμητής εξόδου. Το δίκτυο υπολογίζει την δική του έξοδο και η διαδικασία εκπαίδευσης ελέγχεται από μία συνάρτηση που επιτηρεί την απόκλιση της εξόδου από την επιθυμητή τιμή. Ένα από τα βασικά μειονεκτήματα των μεθόδων αυτών είναι η απαίτηση της χρήσης κάποιων δεδομένων ως επιθυμητή τιμή. Αυτό συνήθως απαιτεί κάποιου είδους

μη αυτοματοποιημένη ανθρώπινη διαδικασία επεξεργασίας δεδομένων. (για παράδειγμα η κατηγοριοποίηση των στοιχείων ενός συνόλου δεδομένων).

Οι κατηγορίες των προβλημάτων που επιλύονται με αλγόριθμους επιβλεπόμενης μάθησης περιλαμβάνουν τα προβλήματα κατηγοριοποίησης και παλινδρόμησης.

Μη Επιβλεπόμενη Μάθηση

Στη μη επιβλεπόμενη μάθηση, τα νευρωνικά δίκτυα εκπαιδεύονται στην αναγνώριση άγνωστων σχέσεων και μοτίβων σε δεδομένα τα οποία κατά κύριο λόγο δεν κατηγοριοποιούνται.

Σε αντίθεση με την επιβλεπόμενη μάθηση, τα μοντέλα αυτά έχουν τη δυνατότητα της αυτο-οργάνωσης που επιτρέπει την αναγνώριση των σχέσεων που μπορεί να εμφανίζονται στο σύνολο δεδομένων, χωρίς την παρέμβαση του ανθρώπινου παράγοντα με την χρήση ετικετών. Με αυτό τον τρόπο, αίρεται ο περιορισμός που πολλές φορές υπάρχουν στα σύνολα δεδομένων που χρησιμοποιούνται με επιβλεπόμενη μάθηση και αφορά την επίπονη διαδικασία απόδοσης ετικετών με μία σχετικά χειροκίνητη διαδικασία. Οι μέθοδοι μη επιβλεπόμενης μάθησης έχουν, ιστορικά, εφαρμογή σε μοντέλα αναπαραγωγής (generative models).

Παραδείγματα μοντέλων που χρησιμοποιούν μη επιβλεπόμενη μάθηση αποτελούν τα δίκτυα Hopfield, οι μηχανές Boltzman και, τελευταία, οι Αυτοκωδικοποιητές Διακύμανσης (Variational Autoencoders). Μερικοί από τους πιο συνηθισμένους αλγόριθμους που χρησιμοποιούνται είναι οι τεχνικές ομαδοποίησης και η ανίχνευση σφαλμάτων.

Ενισχυτική Μάθηση

Στην ενισχυτική μάθηση, δημιουργείται ένας πράκτορας (agent) στον οποίο δίνεται η ικανότητα να πραγματοποιεί διάφορες ενέργειες. Ως στόχος της μάθησης είναι για κάθε κατάσταση, ο πράκτορας αυτός να αποφασίζει αυτή την ενέργεια που μεγιστοποιεί την ανταμοιβή του. Για να επιτευχθεί αυτό, δοκιμάζει τις διάφορες ενέργειες για να εξεταστεί η ανταμοιβή που προκύπτει από αυτές. Στις πιο ενδιαφέρουσες και σύνθετες περιπτώσεις, οι ενέργειες δεν επηρεάζουν μόνο την απευθείας ανταμοιβή για την τρέχουσα κατάσταση, αλλά συνδέονται με την επόμενη κατάσταση. Αυτά τα δύο χαρακτηριστικά, δηλαδή από τη μία η λογική της ανταμοιβής μέσα από μία διαδικασία δοκιμασίας και λάθους και από την άλλη η λογική της καθυστερημένης ανταμοιβής, αποτελούν τη βασική ραχοκοκαλιά της ενισχυτικής μάθησης.

Η ενισχυτική μάθηση διαφέρει από την επιβλεπόμενη μάθηση γιατί δεν απαιτεί δεδομένα που περιέχουν κάποια ετικέτα. Άλλωστε, σε προβλήματα που απαιτείται αλληλεπίδραση με το περιβάλλον, είναι φυσικό να προκύπτουν καταστάσεις που να μην υπάρχουν επεξεργασμένα δεδομένα. Σε τέτοιες καταστάσεις, πρέπει ο πράκτορας να έχει τη δυνατότητα να μαθαίνει από τα λάθη του και να παίρνει τελικά τις σωστές αποφάσεις.

Η ενισχυτική μάθηση διαφέρει επιπλέον από την μη επιβλεπόμενη μάθηση, γιατί δεν έχει ως βασικό στόχο την ανακάλυψη των λανθάνοντων χαρακτηριστικών ενός συνόλου δεδομένων, αλλά την μεγιστοποίηση της ανταμοιβής, χωρίς να σημαίνει ότι η ανάδειξη αυτών των χαρακτηριστικών δεν μπορεί να αξιοποιηθεί.

Μία από τις προκλήσεις που προκύπτουν κατά τη διάρκεια της μάθησης είναι η ισορροπία μεταξύ της αναζήτησης (exploration) και της εκμετάλλευσης (exploitation). Από τη μία, για να επιτευχθεί η μεγιστοποίηση κέρδους ο πράκτορας οφείλει να πραγματοποιεί αυτές τις δραστηριότητες που του επιφέρουν το μέγιστο κέρδος. Από την άλλη, για να μπορέσει να ανακαλύψει αυτές τις δραστηριότητες, πρέπει να εξερευνά τις δυνατότητες που έχει.

1.7 Νευρωνικά Δίκτυα

Ενώ ιστορικά τα νευρωνικά δίκτυα έχουν την αφετηρία τους στις προσπάθειες μαθηματικής προσέγγισης του τρόπου που ένας ζωντανός οργανισμός επεξεργάζεται την πληροφορία, ο όρος αυτός περιγράφει ένα ευρύ φάσμα

μοντέλων και δικτύων, που η πλειονότητα εξ αυτών δεν έχουν σχέση με τη βιολογία, ούτε από άποψη προσέγγισης του τρόπου λειτουργίας, ούτε από τη σκοπιά των προβλημάτων που αυτά σχεδιάζονται να επιλύουν.[4]

Τα τεχνητά νευρωνικά δίκτυα (artificial neural networks) αποτελούνται από ένα σύνολο υπολογιστικών κόμβων/συναρτήσεων που συνδέονται μεταξύ τους, σχηματίζοντας γράφους. Στην πιο γενική μορφή τους, τα δίκτυα αυτό των πυκνά συνδεδεμένων κόμβων δημιουργούν μοντέλα, που ενώ στην αρχή δεν περιέχουν καμία είδους ευφυΐα, μέσα από μία διαδικασία αλληλεπίδρασης με το περιβάλλον τους και επεξεργασίας των ερεθισμάτων που δέχονται, καταφέρνουν να προσαρμόζονται και να επιλύουν τα προβλήματα. Η επαφή με το περιβάλλον γίνεται μέσα από δεδομένα που παρουσιάζονται στις εισόδους των δικτύων ενώ η προσαρμογή στα ερεθίσματα πραγματοποιείται μέσα από τη μάθηση.[5]

1.8 Βαθιά Νευρωνικά Δίκτυα

Τα βαθιά νευρωνικά δίκτυα αποτελούνται από πολλαπλά επίπεδα νευρώνων μεταξύ της εισόδου και της εξόδου. Μέσα από την διαδικασία εκπαίδευσης τους, καταφέρνουν να προσαρμόσουν τις παραμέτρους του δικτύου τους και εν τέλει να μετατρέπουν την είσοδο στην προβλεπόμενη έξοδο.

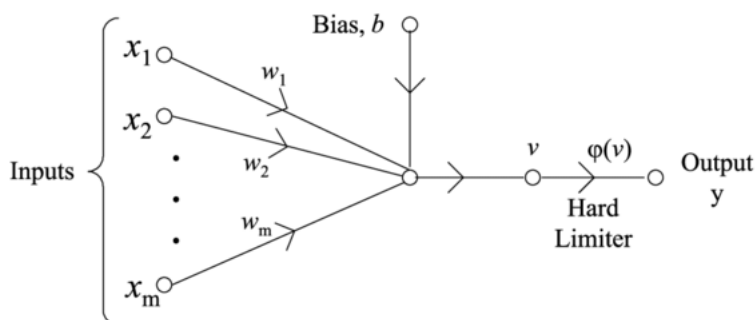
Μία πολύ σημαντική ιδιότητα είναι η δυνατότητα τους να αναγνωρίζουν μη γραμμικές, πολύπλοκες σχέσεις που εμφανίζονται μεταξύ των δεδομένων που χρησιμοποιούνται για την εκπαίδευση. Αυτό τους το πλεονέκτημα καθιστά τις διάφορες κατηγορίες των βαθιών νευρωνικών δικτύων να αναγνωρίζουν σύνθετα στοιχεία (για παράδειγμα, συγκεκριμένα χαρακτηριστικά προσώπου), ή ακόμα και να δημιουργούν σύνθετο περιεχόμενο (όπως για παράδειγμα πρόσωπα). Κάθε επιπλέον επίπεδο μπορεί να αξιοποιηθεί για την αναγνώριση πιο συγκεκριμένων χαρακτηριστικών.

Τα βαθιά νευρωνικά δίκτυα είναι τυπικά δίκτυα εμπρόσθιας τροφοδότησης, με τα δεδομένα να διαδίδονται από την είσοδο προς την έξοδο. Στην αρχή οι παράμετροι του μοντέλου αρχικοποιούνται με τυχαίες αριθμητικές τιμές. Αφού έχει υπολογισθεί, η έξοδος εισέρχεται σε μία συνάρτηση σφάλματος που με τη σειρά της διαδίδει το σφάλμα πίσω στο δίκτυο, αναπροσαρμόζοντας τις παραμέτρους του συστήματος.

Στα διαφορετικά είδη βαθιών νευρωνικών δικτύων κατατάσσονται τα Αντίληπτρα Πολλών Επιπέδων (Multilayer Perceptrons), τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Recurrent Neural Networks), τα οποία αξιοποιούνται σε τεχνικές μοντελοποίησης γλώσσας και τα Συνελκτικα Βαθιά Νευρωνικά Δίκτυα (Convolutional Deep Neural Networks) που χρησιμοποιούνται στην όραση υπολογιστών.

1.9 Ο απλός Perceptron ή Αντίληπτρο

Η θεωρητική σύλληψη του αντιλήπτρου αποτελεί ένας ιστορικός κόμβος στην ανάπτυξη των νευρωνικών δικτύων καθώς αποτελεί το πρώτο νευρωνικό δίκτυο που περιγράφηκε αλγοριθμικά. Αποτελεί το πιο απλό είδος εμπρόσθιου δικτύου (feedforward network). [5]



Σχήμα 1.1: Η αρχιτεκτονική του Αντιλήπτρου

Όπως φαίνεται στο σχήμα, τα βάρη σύναψης του νευρώνα συμβολίζονται με w_1, w_2, \dots, w_m , οι είσοδοι αντίστοιχα με x_1, x_2, \dots, x_m . Ο κόμβος του perceptron αθροίζει τα γινόμενα μεταξύ των βαρών και των εισόδων και στη συνέχεια προστίθεται ο όρος μεροληψίας b .

$$v = \sum_{i=1}^m w_i x_i + b$$

Ο στόχος του perceptron είναι η σωστή ταξινόμηση του διανύσματος εισόδου \mathbf{x} σε μία εκ των δύο δυνατών κλάσεων. Ο κανόνας με τον οποίο αποφασίζεται η ταξινόμηση γίνεται μέσα από την τιμή της εξόδου y . Αν η τιμή είναι 1 τότε επιλέγεται η πρώτη κλάση, ενώ αν η τιμή είναι -1 (ή 0) τότε επιλέγεται η δεύτερη κλάση. Στην πιο απλή μορφή, το perceptron ορίζει ένα υπερεπίπεδο που διαχωρίζει τις δύο περιοχές απόφασης, με εξίσωση:

$$\sum_{i=1}^m w_i x_i + b = 0$$

Ο κανόνας εκπαίδευσης, ο οποίος προκύπτει μέσα από το θεώρημα σύγκλισης του perceptron δίνεται από τη σχέση:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta[d(n) - y(n)]\mathbf{x}(n)$$

Με η ορίζεται το βήμα εκπαίδευσης ενώ η διαφορά $d(n) - y(n)$ συμβολίζει την απόσταση της πρόβλεψης από την επιθυμητή τιμή.

1.10 Multilayer Perceptron - MLP

Τα πολυεπίπεδα αντίληπτρα είναι μία κατηγορία νευρωνικών δικτύων εμπρόσθιας τροφοδότησης. Αποτελούνται από τουλάχιστον τρία επίπεδα. Ένα επίπεδο εισόδου, ένα επίπεδο εξόδου και τουλάχιστον ένα κρυφό επίπεδο. Κάθε νευρώνας, πέρα από αυτούς του επιπέδου εισόδου, χρησιμοποιεί μία μη γραμμική συνάρτηση ενεργοποίησης. Στις περισσότερες περιπτώσεις τα MLP εκπαιδεύονται μέσα από μια τεχνική μάθησης, της οπισθοδρομικής διάδοσης τους σφάλματος (error backpropagation).

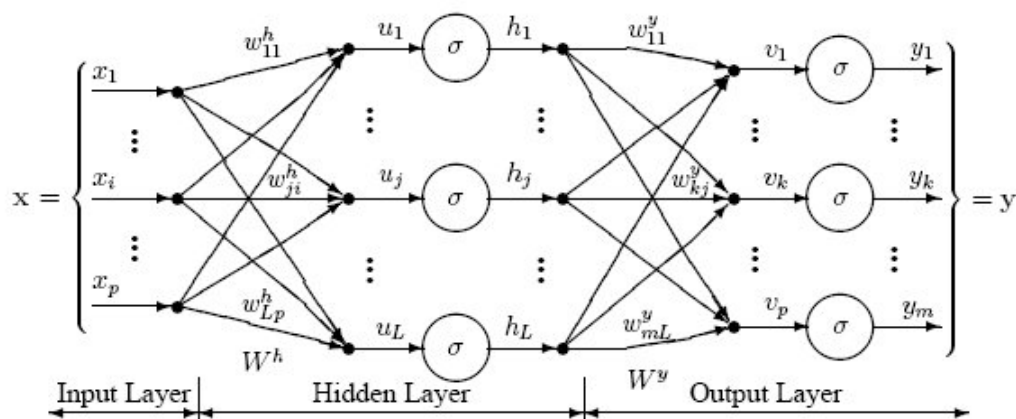
1.10.1 Περιγραφή

Τα βασικά χαρακτηριστικά των MLP είναι τα εξής:[5]

1. Κάθε επιμέρους νευρώνας του δικτύου περιέχει μία μη γραμμική συνάρτηση ενεργοποίησης που είναι διαφορίσιμη.
2. Το δίκτυο περιέχει τουλάχιστον ένα κρυφό επίπεδο.
3. Το δίκτυο πρέπει να είναι σε μεγάλο βαθμό διασυνδεδεμένο.

Κάθε κρυφός νευρώνας ή νευρώνας εξόδου του δικτύου σχεδιάζεται να επιτελεί δύο λειτουργίες:

1. τον υπολογισμό του σήματος που εμφανίζεται στην έξοδο του κάθε κόμβου, που εκφράζεται ως μη γραμμική συνάρτηση του σήματος εισόδου του νευρώνα.
2. τον υπολογισμό της παραγώγου (ή μιας προσέγγισης αυτής) που χρειάζονται για την πίσω διάδοση του σφάλματος μέσα στο δίκτυο.



Σχήμα 1.2: Η αρχιτεκτονική του Multilayer Perceptron

Ο ρόλος των κρυφών επιπέδων του δικτύου

Τα κρυφά επίπεδα του δικτύου ως ανιχνευτές χαρακτηριστικών. Καθώς η διαδικασία εκπαίδευσης προχωράει, τα κρυφά επίπεδα αρχίζουν και να ανακαλύπτουν διάφορα σαφή χαρακτηριστικά του συνόλου δεδομένων. Αυτό επιτυγχάνεται μέσα από τη εφαρμογή μη γραμμικών μετασχηματισμούς πάνω στα δεδομένα εισόδου. Από αυτή τη διαδικασία προκύπτει ένας νέος χώρος που λέγεται χώρος χαρακτηριστικών (feature space).

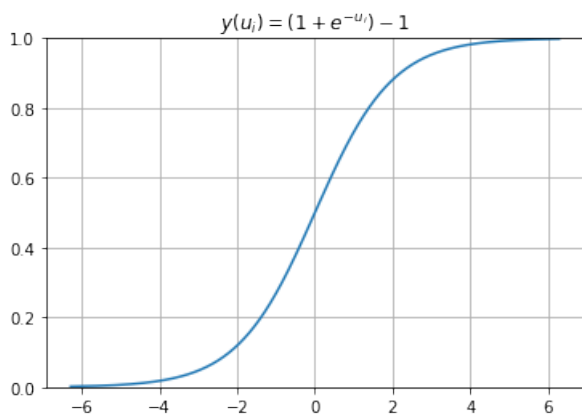
1.10.2 Η συνάρτηση ενεργοποίησης

Η συνάρτηση ενεργοποίησης κάθε κόμβου δέχεται ως όρισμα το άθροισμα των εισόδων (προσαρμοσμένων από τα συνοπτικά βάρη) και καθορίζει την έξοδο του κόμβου. Ιδιαίτερο ενδιαφέρον παρουσιάζουν οι μη γραμμικές συναρτήσεις:

Σιγμοειδής Συναρτήσεις

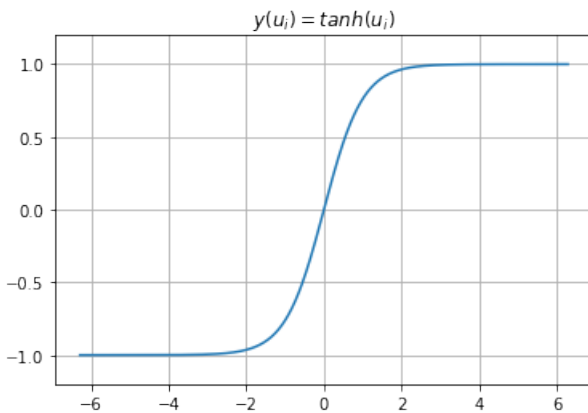
Ιστορικά, στα Multilayer Perceptrons εφαρμόζονται δύο σιγμοειδής συναρτήσεις:
 Η **λογιστική συνάρτηση** με τύπο:

$$y(u_i) = (1 + e^{-u_i})^{-1}$$



Η συνάρτηση υπερβολικής εφαπτομένης με τύπο:

$$y(u_i) = \tanh(u_i)$$



Στις περισσότερες περιπτώσεις που χρειάζεται να εφαρμοσθεί μία σιγμοειδής συνάρτηση, προτιμάται η συνάρτηση υπερβολικής εφαπτομένης λόγω της συμμετρίας της γύρω από το μηδέν και της ομοιότητας της με την $f(u_i) = u_i$ γύρω από αυτό.

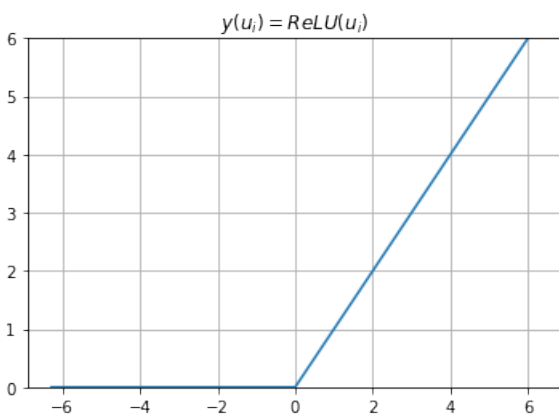
Ένα βασικό μειονέκτημα που εμφανίζεται στις σιγμοειδής συναρτήσεις είναι το πρόβλημα των εξαφανιζόμενων παραγώγων. Κατά τη διάρκεια της πίσω διάδοσης του σφάλματος, αν οι τιμές της συνάρτησης ενεργοποίησης δεν είναι γύρω από το 0, τότε η συνάρτηση μπορεί να φθάνει σε ένα πλάτωμα[6], με αποτέλεσμα οι παράγωγοι σε αυτά τα σημεία να είναι πολύ μικρές. Αν αυτό συμβαίνει σε πολλούς νευρώνες στη σειρά, τότε οι παράγωγοι γίνονται πρακτικά μηδενικές και συνεπώς το δίκτυο δεν εκπαιδεύεται.

Η συνάρτηση Rectifier Linear Unit

Με την ανάπτυξη νέων προσεγγίσεων και τεχνικών στη βαθιά μηχανική μάθηση, η συνάρτηση Rectifier Linear Unit - ReLU έχει γίνει ιδιαίτερα δημοφιλής.

$$f(x) = \max(0, x)$$

Η συνάρτηση αυτή εμφανίζει παράγωγο ίση με 1 για όλες τις τιμές μεγαλύτερες του 0 και έχει γραμμική συμπεριφορά. Ένα από τα βασικά μειονεκτήματα της μεθόδου είναι ότι νευρώνες με συνάρτηση ενεργοποίησης ReLU δεν μπορούν να εκπαιδευθούν όταν η τιμή της συνάρτησης ενεργοποίησης είναι 0. Για την αντιμετώπιση αυτού του ζητήματος έχουν σχεδιασθεί γενικεύσεις που στην γενική περίπτωση λειτουργούν όσο καλά, όσο και η ReLU (Leaky ReLU, parametric ReLU). [6]



1.10.3 Η συνάρτηση στόχου ή συνάρτηση κόστους

Η συνάρτηση στόχου αποτελεί την σχέση αυτή που υπολογίζει την έξοδο του δικτύου, μέσα από την επεξεργασία των δεδομένων και λαμβάνοντας υπόψη τις παραμέτρους του συστήματος. Στα νευρωνικά δίκτυα, ο προσδιορισμός των παραμέτρων προκύπτει μέσα από το πρόβλημα βελτιστοποίηση αυτής της συνάρτησης. Υπάρχουν αρκετές διαφορετικές προσεγγίσεις για την επιλογή της κατάλληλης συνάρτησης κόστους. Σε προβλήματα παλινδρόμησης χρησιμοποιούνται το μέσο τετραγωνικό σφάλμα, το μέσο απόλυτο σφάλμα κ.ο.κ.

Όσον αφορά τα βαθιά νευρωνικά δίκτυα, μια αρκετά δημοφιλής προσέγγιση είναι η χρήση της λογαριθμικής πιθανοφάνειας ως τη συνάρτηση κόστους.[6] Το πρόβλημα ανάγεται στη μεγιστοποίηση της συνάρτησης πιθανοφάνειας που αντιστοιχεί στην ελαχιστοποίηση της διασταυρώμενης εντροπίας (cross entropy).

Η οπισθοδρομική διάδοση του σφάλματος

Η οπισθοδρομική διάδοση είναι ένας ευρέως διαδεδομένος αλγόριθμος εκπαίδευσης νευρωνικών δικτύων εμπρόσθιας διάδοσης. Η γενική ιδέα βασίζεται στον γρήγορο υπολογισμό των μερικών παραγώγων του συστήματος που χρειάζονται για την εκπαίδευση του, μέσα από την εφαρμογή του κανόνα αλυσίδας στην μερική παράγωγο της συνάρτησης σφάλματος, ως προς το συναπτικό βάρος του i νευρώνα του j επιπέδου. Η χρήση του κανόνα αλυσίδας επιτρέπει τον υπολογισμό των παραγώγων του δικτύου, για κάθε επίπεδο, από πίσω προς τα μπρος αποφεύγοντας περιττούς υπολογισμούς όρων. Άρα:

$$\frac{\partial \mathcal{L}}{\partial w_{ij}} = \frac{\partial \mathcal{L}}{\partial o_j} \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}}$$

Ως \mathcal{L} ορίζεται η συνάρτηση σφάλματος, ως o_j ορίζεται η έξοδος ενός κόμβου, ενώ το net_j μας δίνει το σταθμισμένο άθροισμα των εισόδων του κόμβου.

Ο πρώτος όρος $\frac{\partial \mathcal{L}}{\partial o_j}$ υπολογίζεται ως την παράγωγο της συνάρτησης σφάλματος, αν ο κόμβος ανήκει στο επίπεδο εξόδου. Αν όμως ο κόμβος ανήκει σε κάποιο άλλο επίπεδο, τότε για τον δεδομένο νευρώνα, ο όρος εξαρτάται από τις διασυνδέσεις των νευρώνων του επόμενου σταδίου.

Ο δεύτερος όρος $\frac{\partial o_j}{\partial net_j}$ είναι η παράγωγος της συνάρτησης ενεργοποίησης, αφού $o_j = \phi(net_j)$.

Ο τρίτος όρος $\frac{\partial net_j}{\partial w_{ij}}$ ισούται με την έξοδο του i νευρώνα του προηγούμενου επιπέδου.

Πρέπει να τονιστεί ότι ο αλγόριθμος της πίσω διάδοσης περιγράφει έναν τρόπο υπολογισμού των παραγώγων και όχι τον τρόπο που αυτός αξιοποιείται για την μεταβολή των βαρών, η οποία επιτυγχάνεται μέσα από τη χρήση κάποιας επαναληπτικής μεθόδου βελτιστοποίησης (optimizers), όπως η (απλή) Gradient Descent, ο αλγόριθμος Adagrad ή ο αλγόριθμος Adam.

Μια απλή περίπτωση - Σύγκλιση με ελάττωση παραγώγου (Gradient Descent)

Η Gradient Descent είναι ένας επαναληπτικός αλγόριθμος βελτιστοποίησης πρώτης τάξης που οδηγεί στην εύρεση τοπικών ελαχίστων μιας διαφορίσιμης συνάρτησης. Στα νευρωνικά δίκτυα χρησιμοποιείται για να υπολογισθούν τα συνοπτικά βάρη και οι όροι μεροληψίας.

Όπως είναι γνωστό, η παράγωγος μιας συνάρτησης δείχνει την κατεύθυνση και την τιμή της πιο απότομης μεταβολής μιας συνάρτησης. Αυτή η ιδιότητα είναι χρήσιμη κατά τη διάρκεια εκπαίδευσης των νευρωνικών δικτύων. Για ένα σύνολο δεδομένων υπολογίζεται η συνάρτηση κόστους των δεδομένων αυτών. Αξιοποιώντας την οπισθοδρομική διάδοση σφάλματος, υπολογίζονται οι μερικές παράγωγοι όλων των παραμέτρων του δικτύου και υπολογίζονται τα νέα βάρη ως η διαφορά των παλιών βαρών με τις αντίστοιχες μερικές παραγώγους.

Έστω $W(t)$ ένας πίνακας που περιλαμβάνει τις παραμέτρους (τα βάρη και τους συντελεστές μεροληψίας) του δικτύου, τη στιγμή t . Οι νέες παράμετροι μετά την εκτέλεση του αλγορίθμου της Gradient Descent δίνονται

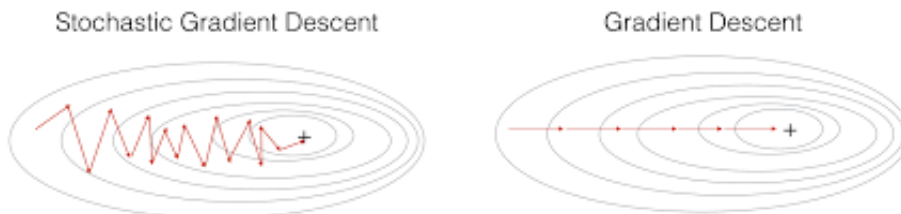
από την εξίσωση:

$$W(t+1) = W(t) - \gamma \nabla F(W(t))$$

Με γ συμβολίζεται το βήμα εκπαίδευσης (training step), ενώ ως $F(W(t))$ ορίζεται ο μέσος όρος ή το άθροισμα των συναρτήσεων σφάλματος όλων των δεδομένων που δίνονται στο σύνολο.

Στην πραγματικότητα, ο αλγόριθμος Gradient Descent παρουσιάζει κάποια μειονεκτήματα. Το πρώτο αφορά το ζήτημα της σύγκλισης. Ενώ είναι δεδομένο ότι ο αλγόριθμος συγκλίνει σε ένα τοπικό ελάχιστο, δεν είναι δεδομένο ότι συγκλίνει σε ένα ολικό ελάχιστο. Το δεύτερο μειονέκτημα αφορά το υπολογιστικό κόστος σε βαθιά νευρωνικά δίκτυα και πολύ μεγάλα Σύνολα Δεδομένων (Dataset), αφού για κάθε αναβάθμιση των παραμέτρων απαιτείται η επεξεργασία όλων των δεδομένων που δίνονται.

Stochastic Gradient Descent και η χρήση Παρτιδών (Batches)



Σχήμα 1.3: Stochastic Gradient Descent vs Gradient Descent

Μία λύση στον περιορισμό των υπολογισμών της Gradient Descent είναι η ενημέρωση των παραμέτρων του συστήματος σε κάθε datapoint ή σε ένα μικρό τμήμα (Minibatch) του συνόλου δεδομένων. Αυτή η προσέγγιση εισάγει στον αλγόριθμο έναν βαθμό στοχαστικότητας, αφού πλέον δεν υπολογίζεται η βέλτιστη κατεύθυνση με βάση το σύνολο δεδομένων, αλλά σε τυχαίο υποσύνολο του. Το πλεονέκτημα είναι πλέον ο αλγόριθμος δεν εξαρτάται από το μέγεθος των δεδομένων και η σύγκλιση μπορεί να προκύψει πριν εισαχθεί το σύνολο του dataset. Ένα ακόμα πλεονέκτημα είναι η πιθανότητα ο αλγόριθμος να ξεφύγει από ένα τοπικό ελάχιστο ή πλάτωμα που εμφανίζονται στις μη κυρτές συναρτήσεις.

Στην πραγματικότητα, και αξιοποιώντας τις παραπάνω προσεγγίσεις, έχουν δημιουργηθεί πολύ αποτελεσματικοί αλγόριθμοι βελτιστοποίησης παραμέτρων, τόσο ως προς την ταχύτητα σύγκλισης, όσο και από τη δυνατότητα να συγκλίνουν και σε μη κυρτές συναρτήσεις. Τέτοιοι αλγόριθμοι είναι οι Adam, AMSGrad κ.ο.κ..

Ο Adam Optimizer

Ο Adam^[7] αποτελεί έναν αλγόριθμο αποτελεσματικής στοχαστικής βελτιστοποίησης, βασισμένο σε προσαρμοσμένες προσεγγίσεις της πρώτης (δηλαδή της μέση τιμή) και της δεύτερης (δηλαδή της μη κεντραρισμένης διακύμανσης) ροπή της κλίσης (gradient), που χρειάζεται μόνο παραγώγους πρώτης τάξης με μικρές απαιτήσεις για μνήμη. Μερικά από τα πλεονεκτήματα του είναι ότι οι τάξεις μεγέθους των ενημερώσεων των παραμέτρων του συστήματος είναι αμετάβλητες των αλλαγών στην κλίμακα της κλίσης, το μέγεθος των βημάτων είναι φραγμένα από την υπερπαραμέτρο που την καθορίζει, δουλεύει καλά με μη στατικές συναρτήσεις στόχου (objective function).

Ο αλγόριθμος υλοποιεί εκθετικούς κινητούς μέσους όρους exponential moving averages m, u , οι οποίοι αποτελούν τις προσεγγίσεις του μέσου όρου (πρώτη ροπή) και της ακεντράριστης διακύμανσης (δεύτερη ροπή) αντίστοιχα. Οι προσεγγίσεις αυτές τείνουν να μεροληπτούν στο 0, ιδιαίτερα στις πρώτες επαναλήψεις, κάτι που οφείλεται στην αρχικοποίησή τους στη τιμή 0. Για αυτό το λόγο χρειάζεται να πραγματοποιηθεί η διόρθωση της μεροληψίας μέσα από τον υπολογισμό των \hat{m}, \hat{u} .

Algorithm 1 Ο αλγόριθμος Adam

Require: α : Stepsize

Require: $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for the moment estimates

Require: $f(\theta)$: Stochastic objective function with parameters θ

Require: θ_0 : Initial parameter vector

$m_0 \leftarrow 0$ (Initialize 1st moment vector)

$u_0 \leftarrow 0$ (Initialize 2nd moment vector)

$t \leftarrow 0$ (Initialize timestep)

while θ_t not converged **do**

$t \leftarrow t + 1$

$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$ (Get gradients with respect to stochastic objective at timestep t)

$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Update biased first moment estimate)

$u_t \leftarrow \beta_2 \cdot u_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Update biased second moment estimate)

$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ (Compute bias-corrected first moment estimate)

$\hat{u}_t \leftarrow u_t / (1 - \beta_2^t)$ (Compute bias-corrected second moment estimate)

$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{u}_t} + \epsilon)$ (Update Parameters)

end while

return θ_t (Resulting parameters)

1.11 Μετρικές Απόδοσης (Performance Metrics)

Στη θεωρία πληροφορίας, οι μετρικές απόδοσης είναι ένας τρόπος εκτίμησης της δυνατότητας ενός μοντέλου ή μίας μεθόδου να αποδίδει τα προσδοκώμενα αποτελέσματα. Για την περιγραφή των διαφορετικών μετρικές, προσδιορίζονται δύο σύνολα. Το πρώτο περιέχει τα εκτιμώμενα αποτελέσματα, ενώ το δεύτερο περιέχει τα πραγματικά αποτελέσματα. Με βάση αυτά, περιγράφονται οι παρακάτω μετρικές:

Ακρίβεια (Precision)

Η ακρίβεια αποτελεί το κλάσμα των αποτελεσμάτων που ανήκουν τόσο στα εκτιμώμενα όσο και στα πραγματικά αποτελέσματα σε σχέση του συνόλου των πραγματικών αποτελεσμάτων. Στη γενική περίπτωση, συμπεριλαμβάνει όλα τα αποτελέσματα του μοντέλου. Ωστόσο, υπάρχει η δυνατότητα να τεθεί ένα άνω όριο στα αποτελέσματα που θα ληφθούν υπ' όψιν. Αυτή η μετρική ονομάζεται ακρίβεια στο n (Precision@ n).

Ανάκληση (Recall)

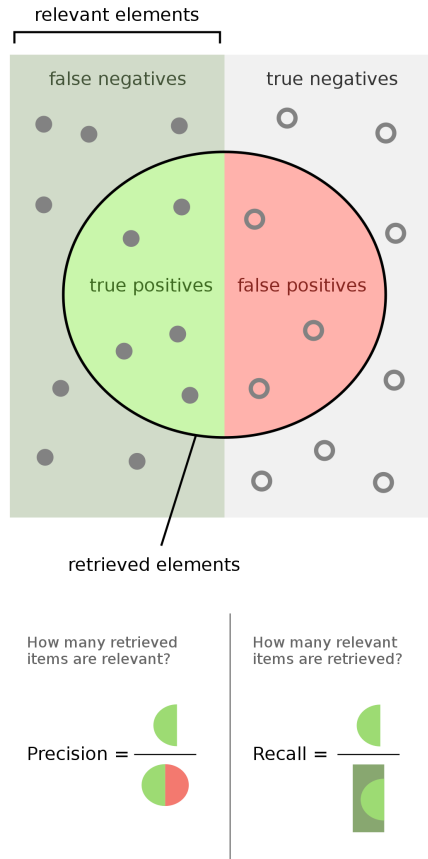
Η ανάκληση αποτελεί το κλάσμα των αποτελεσμάτων που ανήκουν τόσο στα πραγματικά όσο και στα εκτιμώμενα αποτελέσματα ενός μοντέλου σε σχέση με το σύνολο των εκτιμώμενων αποτελεσμάτων. Στη γενική περίπτωση, συμπεριλαμβάνει όλα τα αποτελέσματα του μοντέλου. Ωστόσο, υπάρχει η δυνατότητα να τεθεί ένα άνω όριο στα αποτελέσματα που θα ληφθούν υπ' όψιν. Αυτή η μετρική ονομάζεται ανάκληση στο n (Recall@ n).

Κανονικοποιημένο Εκπτωτικό Άθροισμα Κέρδους (Normalized Discounted Cumulative Gain)

Το Εκπτωτικό Άθροισμα Κέρδους αποτελεί μία μέτρηση της ποιότητας της ταξινόμησης των αποτελεσμάτων ενός μοντέλου. Χρησιμοποιείται με ευρύ τρόπο για την αποτελεσματικότητα των μηχανών αναζήτησης και άλλων συγγενών εφαρμογών.

Ο τρόπος υπολογισμού περιγράφεται αναλυτικά παρακάτω:

- Υπολογισμός αθροίσματος κέρδους (CG): Είναι το άθροισμα των εκτιμώμενων τιμών των αντικειμένων



Σχήμα 1.4: Υπολογισμός της Ακρίβειας (Precision) και της Ανάκλησης (Recall) σε έναν πίνακα σύγκυσης (confusion matrix).

που επιστρέφονται σε μία αναζήτηση.

$$CG_p = \sum_{i=1}^p rel_i \quad (1.1)$$

Με rel_i προσδιορίζεται η βαθμολογία του αποτελέσματος που έχει δοθεί από τον χρήστη, στη θέση i . Η μέτρηση δεν μπορεί να αντιληφθεί τις μεταβολές στη θέση των αποτελεσμάτων.

- Υπολογισμός εκπτώτικου αθροίσματος κέρδους (Discounted CG): Η λογική του Εκπτώτικου Αθροίσματος Κέρδους αποτελεί τον περιορισμό της βαθμολογίας που προσφέρει στο άθροισμα ένα σχετικό αντικείμενο, που όμως βρίσκεται χαμηλά στην κατάταξη των αποτελεσμάτων. Αυτό επιτυγχάνεται μέσα από την διαίρεση με έναν λογαριθμικό παράγοντα, που η τιμή του αυξάνεται ανάλογα με τη θέση κατάταξης.

$$DCG_p = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)} \quad (1.2)$$

ή εναλλακτικά, (1.3)

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2(i+1)} \quad (1.4)$$

- Υπολογισμός του Κανονικοποιημένου Εκπτώτικου Αθροίσματος Κέρδους (nDCG): Ενώ το DCG μπορεί να αναγνωρίσει τις διαφορές στη σειρά των αποτελεσμάτων, δεν υπάρχει κανονικοποίηση. Για την επίλυση

αυτού του προβλήματος, ορίζεται το nDCG:

$$nDCG_p = \frac{DCG_p}{IDCG_p} \quad (1.5)$$

με, (1.6)

$$IDCG_p = \sum_{i=1}^{|REL_p|} \frac{rel_i}{\log_2(i+1)} \quad (1.7)$$

Με το $IDCG_p$, να αποτελεί το αποτέλεσμα του εκπτώτικου αθροίσματος κέρδους, αν τα αποτελέσματα ήταν σε ιδανική σειρά και REL_p να αποτελεί την λίστα με όλα τα σχετικά αντικείμενα, αν αυτά διαταχθούν με τη σειρά σχετικότητά τους.

Παραλλαγή της nDCG αποτελεί η nDCG@n, που λαμβάνει υπ' όψιν μόνο τα n πρώτα αποτελέσματα. Με αυτό τον τρόπο είναι δυνατό να "τιμωρηθεί" το μοντέλο αν δεν επιστρέφει τα σχετικά αντικείμενα νωρίς.

Κεφάλαιο 2

Συστήματα Προτάσεων/Συστάσεων - Recommendation Systems

Σε αυτό το κεφάλαιο θα περιγραφούν τα συστήματα προτάσεων και θα περιγραφούν οι βασικότερες τεχνικές υλοποίησής τους.

2.1 Κατηγορίες Συστημάτων Προτάσεων

Σε γενικές γραμμές, υπάρχουν δύο βασικές προσεγγίσεις:

2.1.1 Φιλτράρισμα με βάση το Περιεχόμενο - Content Based Filtering

Τα συστήματα προτάσεων που βασίζονται στο περιεχόμενο αξιοποιούν περιγραφές στους χρήστες και στα αντικείμενα για να κατασκευάσουν αναπαραστάσεις των αντικειμένων και προφίλ χρηστών. Με αυτό τον τρόπο πραγματοποιούνται προτάσεις αντικειμένων παρόμοιων με αυτά που ένας χρήστης προτίμησε στο παρελθόν. Η βασική διαδικασία αυτής της τεχνικής περιλαμβάνει την ομαδοποίηση των χαρακτηριστικών του εκάστοτε χρήστη με τα χαρακτηριστικά των αντικειμένων. Το αποτέλεσμα είναι μία βαθμολογία σχετικότητας που προβλέπει το επίπεδο ενδιαφέροντος στα αντικείμενα.

Συνήθως τα χαρακτηριστικά που χρησιμοποιούνται για τα αντικείμενα εξάγονται από κάποια χαρακτηριστικά που σχετίζονται με το αντικείμενο (για παράδειγμα, η διάρκεια ενός μουσικού κομματιού, ο ρυθμός κ.ο.κ.), ή κομμάτια κειμένου από την περιγραφή του. Για την δημιουργία του προφίλ ενός χρήστη, τα συστήματα συνήθως επικεντρώνονται σε δύο ειδών δεδομένα. Αυτά που μοντελοποιούν τις προτιμήσεις του χρήστη και αυτά που περιγράφουν το ιστορικό του χρήστη μέσα στο σύστημα.

Ένα βασικό ζήτημα που εμφανίζεται είναι οι περιορισμένες προτάσεις σε αντικείμενα παρόμοια με αυτά που έχει εκδηλώσει ο χρήστης. Εν ολίγοις, η αδυναμία εξερεύνησης αντικειμένων που ξεφεύγουν από τις δηλωμένες προτιμήσεις του χρήστη.

2.1.2 Συνεργατικό Φιλτράρισμα - Collaborative Filtering

Οι τεχνικές Συνεργατικού Φιλτραρίσματος έχουν ως βάση την δημιουργία προτάσεων ενός χρήστη, βασιζόμενες στα μουσικά κομμάτια που άρεσαν σε χρήστες με παρόμοιο γούστο. Η ομοιότητα στο γούστο των χρηστών βασίζεται στην ομοιότητα των βαθμολογιών που έκαναν οι χρήστες στα αντικείμενα. Οι τεχνικές Collaborative

Filtering αποτελούν από τις πιο ευρέως διαδεδομένες τεχνικές στα συστήματα προτάσεων. Στην παρούσα διπλωματική θα υλοποιηθούν δύο διαφορετικά μοντέλα συνεργατικού φιλτραρίσματος.

Βασικό πλεονέκτημα των συστημάτων που αξιοποιούν το συνεργατικό φιλτράρισμα είναι ότι είναι ανεξάρτητες σε περιεχόμενο που έχει προσαρμοστεί να αναλύεται από υπολογιστή και επομένως εκπαιδεύονται στο να πραγματοποιούν σχετικά σύνθετες προτάσεις χωρίς να βασίζονται στην ανάλυση χαρακτηριστικών του ίδιου του αντικειμένου.

Περιορισμοί και προβλήματα

Οι προσεγγίσεις του συνεργατικού φιλτραρίσματος συχνά αντιμετωπίζουν τρία προβλήματα:

- **Κρύα Εκκίνηση (Cold Start):** Για έναν νέο χρήστη ή αντικείμενο, δεν υπάρχουν αρκετά δεδομένα για να γίνουν νέες προτάσεις. Αν δεν υπάρχει κάποια αλληλεπίδραση με ένα καινούριο αντικείμενο, τότε το σύστημα δεν μπορεί να το προτείνει. Επιπλέον, αν υπάρχει μόνο μία μικρή αλληλεπίδραση με το εν λόγω αντικείμενο, τότε ενώ το σύστημα θα μπορεί να το περιλάβει στις προτάσεις του, αυτές θα είναι αρκετά άστοχες. Συνεπώς, επέκταση του προβλήματος κρύας εκκίνησης, είναι η αλληλεπίδραση του συστήματος προτάσεων όχι μόνο με τα νέα μουσικά κομμάτια, αλλά και με αυτά που δεν είναι δημοφιλή. Το φαινόμενο αυτό ονομάζεται μεροληψία δημοτικότητας popularity bias και είναι ιδιαίτερης σημασίας για τη σχεδίαση ενός δικτύου συνεργατικού φιλτραρίσματος, αφού μουσικά κομμάτια με μικρή αλληλεπίδραση που βρίσκονται στο σύστημα για μήνες, είναι πολύ πιθανό να μείνουν σε έναν φαύλο κύκλο. Από τη σκοπιά των χρηστών, το δίκτυο δεν μπορεί να κάνει κάποιου είδους αποτελεσματική και σωστή πρόταση αφού δεν υπάρχει κάποια προϊστορία αλληλεπίδρασης του χρήστη με τα αντικείμενα. Αυτό το πρόβλημα είναι ιδιαίτερης σημασίας γιατί ένας νέος χρήστης που αντιμετωπίζει προτάσεις χαμηλής ποιότητας μπορεί να σταματήσει αρκετά άμεσα τη χρήση του συστήματος.
- **Δυνατότητα Κλιμάκωσης (Scalability):** Σε πολλές εφαρμογές των συστημάτων συνεργατικού φιλτραρίσματος είναι αρκετά πιθανό τόσο οι χρήστες όσο και τα αντικείμενα να ξεπερνούν τις δεκάδες χιλιάδες και να αγγίζουν τα εκατομμύρια. Είναι προφανές, ότι εξαιτίας του μεγάλου μεγέθους της διαθέσιμης πληροφορίας απαιτείται και η αντίστοιχη υπολογιστική ισχύς κάτι που αναδεικνύει την ανάγκη εύρεσης τεχνικών που να μπορούν να εφαρμοστούν σε ολοένα κλιμακούμενα σύνολα δεδομένων, χωρίς μεγάλη επίπτωση στο υπολογιστικό κόστος.
- **Αραιότητα (Sparsity):** Στα περισσότερα προβλήματα που υλοποιούμε συστήματα collaborative filtering ο αριθμός των διαθέσιμων αντικειμένων είναι εξαιρετικά μεγάλος. Στον αντίποδα, οι χρήστες αλληλεπιδρούν με ένα πολύ περιορισμένο αριθμό αυτών, με αποτέλεσμα ακόμη και τα πιο δημοφιλή αντικείμενα να έχουν έναν περιορισμένο αριθμό βαθμολογιών.

2.2 Τεχνικές υλοποίησης

Η υλοποίηση του συνεργατικού φιλτραρίσματος μπορεί να ταξινομηθεί σε δύο κύριες κατηγορίες. Η πρώτη αφορά τα συστήματα βασισμένα στη μνήμη, ενώ η δεύτερη αφορά τα συστήματα βασισμένα στη χρήση μοντέλου. Τέλος, τα τελευταία χρόνια, υπάρχει έντονο ενδιαφέρον στην εξερεύνηση των δυνατοτήτων των μοντέλων βαθιάς μηχανικής μάθησης σε εφαρμογές συνεργατικού φιλτραρίσματος, με σχετικά μεικτά αποτελέσματα.

2.2.1 Συστήματα βασισμένα στη μνήμη

Τα συστήματα με βάση τη μνήμη χρησιμοποιούν στατιστικές τεχνικές πάνω στο σύνολο των δεδομένων για την εξαγωγή των προτάσεων. Η κεντρική ιδέα είναι η χρήση κάποιας τεχνικής για την εύρεση των γειτόνων του χρήστη, που έχουν ένα ιστορικό κοινού γούστου με αυτόν.^[8] Η διαδικασία μπορεί να περιγραφεί με τρία στάδια. Το πρώτο είναι η εύρεση των γειτόνων, το δεύτερο είναι ο υπολογισμός της ομοιότητας ενώ το τελικό στάδιο αποτελεί την εξαγωγή των προτάσεων.

Κάποια από τα πλεονεκτήματα των συστημάτων χρήστη-αντικειμένου που βασίζονται στη μνήμη είναι η ευκολία στην υλοποίηση και η ευκολία στην προσθήκη νέων δεδομένων. Στον αντίποδα, τα βασικά προβλήματα είναι η δυσκολία αντιμετώπισης των αραιών δεδομένων και η κλιμάκωση.[9]

2.2.2 Συστήματα βασισμένα σε μοντέλο

Ο σχεδιασμός και η ανάπτυξη μοντέλων και τεχνικών μηχανικής μάθησης και συλλογής δεδομένων επιτρέπουν πλέον σε μοντέλα συστήματα να ανακαλύπτουν και να αντιλαμβάνονται περίπλοκες σχέσεις των δεδομένων που χρησιμοποιούνται στην εκπαίδευση. Στην προσπάθεια αντιμετώπισης των περιορισμών των συστημάτων με βάση τη μνήμη, αρκετές μέθοδοι έχουν δοκιμασθεί, όπως για παράδειγμα μπειζιανά μοντέλα, μοντέλα με χρήση διάσπασης πίνακα σε ιδιάζουσες τιμές (SVD).[10]

Τα πλεονεκτήματα αυτών των μεθόδων είναι η καλύτερη αντιμετώπιση αραιών δεδομένων, η βελτίωση των προβλέψεων και η δυνατότητα μιας λογικής (για τον άνθρωπο) ερμηνείας των προτάσεων. (για παράδειγμα, μέσα από το εσωτερικό γινόμενο.)

2.2.3 Συστήματα με χρήση τεχνικών βαθιάς μηχανικής μάθησης

Τα τελευταία χρόνια έχει αναπτυχθεί ερευνητικό ενδιαφέρον για την εφαρμογή τεχνικών βαθιάς μηχανικής μάθησης για την ανάπτυξη μοντέλων προτάσεων, είτε μέσα από τη γενίκευση ευρέως εφαρμοσμένων αλγορίθμων (λ.χ. παραγοντοποίηση πινάκων), ή μέσα από την προσαρμογή και δοκιμή νέων τεχνικών (λ.χ. Variational Autoencoders).

Ωστόσο, μία μετά-ανάλυση δημοσιεύσεων σε κορυφαία περιοδικά έδωσε κάποια αποθαρρυντικά αποτελέσματα σχετικά με την αποδοτικότητα τους. Προσπαθώντας να εφαρμόσει τις μεθόδους που οι δημοσιεύσεις αυτές περιέγραφαν, κατάφερε να αναπαράξει μόνο το 40% εξ αυτών, με την πλειονότητα τους να δίνουν χειρότερα αποτελέσματα ακόμα και από πιο απλά και παλιά μοντέλα.[11]

2.3 Συνεργατικό φιλτράρισμα με Παραγοντοποίηση Πινάκων

Η παραγοντοποίηση πινάκων αποτελεί μία από τις πιο διαδεδομένες τεχνικές υλοποίησης συστημάτων συνεργατικού φιλτράρισματος που βασίζονται σε μοντέλο. Αυτές οι τεχνικές διαδόθηκαν ευρέως μέσα από τον διαγωνισμό του Netflix λόγω της σημαντικής βελτίωσης των αποτελεσμάτων σε σχέση με τις τότε προϋπάρχουσες υλοποιήσεις.

Οι τεχνικές αυτές βασίζονται στην ανάλυση ενός πίνακα βαθμολογιών R σε δύο πίνακες. Σε κάθε χρήστη και κάθε αντικείμενο αντιστοιχίζεται ένα διάνυσμα N παραγόντων. Ως αποτέλεσμα ο πρώτος πίνακας αποτελεί το σύνολο των παραγόντων των χρηστών, με διαστάσεις (n_U, N) , ενώ ο δεύτερος αποτελεί το σύνολο των παραγόντων των αντικειμένων, με διάσταση (n_I, N) . Ο αριθμός των παραγόντων για κάθε χρήστη και κάθε αντικείμενο είναι κοινός και αποτελεί μια υπερπαραμέτρο του μοντέλου.

Ο στόχος για την εκπαίδευση του μοντέλου είναι ο πολλαπλασιασμός των δύο αυτών πινάκων να μπορεί να ανακατασκευάζει τον αρχικό πίνακα R . Δηλαδή πρέπει η διαφορά $R - UI^T$ να τείνει στο 0. Επομένως απαιτείται μία επαναληπτική μέθοδος για τον προσδιορισμό των παραγόντων των δύο πινάκων.

Για την διαδικασία εκπαίδευσης θεωρείται η συνάρτηση κόστους (Root Mean Squared Error) με L2 κανονικοποίηση:

$$f(U, M) = \sum_{(u,i) \in S} (r_{ui} - \mathbf{x}_u^T \mathbf{y}_i)^2 + \lambda(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) \quad (2.1)$$

Η επίλυση και η αναπροσαρμογή των τιμών μπορεί να γίνει είτε χρησιμοποιώντας Stochastic Gradient Descent, είτε με τη χρήση των εναλλασσόμενων ελάχιστων τετραγώνων Alternating Least Square. Η μέθοδος SGD είναι πιο αποτελεσματική όταν ο πίνακας των βαθμολογιών είναι σχετικά πυκνός. Στον αντίποδα, η μέθοδος ALS γίνεται πιο αποτελεσματική με την αύξηση του πλήθους των δεδομένων και σε περιπτώσεις που ο πίνακας είναι αραιός, όπως γίνεται στις περισσότερες περιπτώσεις που σχεδιάζονται συστήματα προτάσεων με δεδομένα έμμεσης αναφοράς (implicit).

2.4 Περιγραφή της μεθόδου

Η υλοποίηση της μεθόδου βασίστηκε στη δημοσιευμένη εργασία των Hu, et al. "Collaborative Filtering for Implicit Feedback Datasets". [12] Από το σύνολο δεδομένων προκύπτει ο πίνακας των βαθμολογιών $R = \{r_{ui}\}_{n_u \times n_i}$. Τα στοιχεία του πίνακα έχουν τις τιμές 0 ή 1, επειδή το σύνολο δεδομένων που χρησιμοποιείται έχει ως πληροφορία την ύπαρξη ή την απουσία των κομματιών σε κάθε λίστα αναπαραγωγής. Τα n_u, n_i συμβολίζουν τον αριθμό των χρηστών και των αντικειμένων.

Στη συγκεκριμένη υλοποίηση, ο παραπάνω πίνακας χρησιμοποιείται για τον σχηματισμό δύο νέων. Ο πρώτος πίνακας P αντιπροσωπεύει την προτίμηση ενός χρήστη για ένα αντικείμενο και μοντελοποιεί αν κάποιος χρήστης έχει επιδράσει με κάποιον έμμεσο ή άμεσο τρόπο με το αντικείμενο. Αν έχει επιδράσει, τότε το στοιχείο του πίνακα τίθεται ως 1, ενώ στη διαφορετική περίπτωση ως 0:

$$p_{ui} = \begin{cases} 1, & r_{ui} > 0 \\ 0, & r_{ui} = 0 \end{cases} \quad (2.2)$$

Ο δεύτερος πίνακας αποτυπώνει την εμπιστοσύνη της αρεσκείας του χρήστη για ένα αντικείμενο. Η μοντελοποίηση του πραγματοποιείται μέσα από τη σχέση:

$$c_{ui} = 1 + \alpha r_{ui} \quad (2.3)$$

Όπου α μια υπερπαραμέτρος του συστήματος, η βέλτιστη τιμή της οποίας καθορίζεται κατά τη διάρκεια της εκπαίδευσης, ανάλογα με το πρόβλημα. Ο πίνακας C δεν έχει καμία μηδενική τιμή, αφού ακόμα και στην περίπτωση που η βαθμολογία του r_{ui} είναι 0, η τιμή του στοιχείου του πίνακα είναι άσος.

Θεωρείται επιπλέον η εξής σύμβαση: ως χρήστες θα θεωρηθούν οι λίστες αναπαραγωγής και ως αντικείμενα θα θεωρηθούν τα μουσικά κομμάτια. Για κάθε χρήστη και αντικείμενο δημιουργούνται τα διάνυσμα $x_u, y_u \subseteq R^{n_f}$, όπου το n_f αποτελεί τον αριθμό των παραγόντων. Στη συνέχεια, ορίζονται οι πίνακες $U_f = [x_u], \forall u \in U$, που περιλαμβάνει το σύνολο των διανυσμάτων που περιέχουν τους παράγοντες κάθε χρήστη, και $I_f = [y_i], \forall i \in I$ που περιλαμβάνει το σύνολο των διανυσμάτων που περιέχουν τους παράγοντες κάθε αντικείμενου.

2.5 Συνάρτηση Κόστους

Η συνάρτηση κόστους της υλοποίησης βάσει των αλλαγών που περιγράφηκαν τροποποιείται:

$$f(U, M) = \sum_{(u,i) \in S} c_{ui} (p_{ui} - \mathbf{x}_u^T \mathbf{y}_i)^2 + \lambda (\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) \quad (2.4)$$

Ο παράγοντας $\lambda (\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)$ αποτελεί τον L_2 ρυθμιστικό όρο έτσι ώστε να αποφευχθεί η υπερεκπαίδευση (overfitting) του δικτύου. Ο όρος λ αποτελεί υπερπαραμέτρο του δικτύου και υπολογίζεται πειραματικά.

2.5.1 Η μέθοδος των εναλλασσόμενων ελάχιστων τετραγώνων

Η εκπαίδευση του μοντέλου πραγματοποιείται μέσα από την ελαχιστοποίηση της παραπάνω συνάρτησης κόστους. Λόγω του πολύ μεγάλου μεγέθους των συνόλων δεδομένων, η ελαχιστοποίηση αυτή γίνεται μέσα από τη μέθοδο των εναλλασσόμενων Ελαχίστων Τετραγώνων.

Για να μπορέσουν να προσδιοριστούν οι βέλτιστες τιμές των πινάκων U_f και I_f , εργαζόμαστε με τον εξής τρόπο: Για τον προσδιορισμό του U_f , θεωρείται ο πίνακας των αντικειμένων I_f ως σταθερά. Με αυτό τον τρόπο η συνάρτηση κόστους μπορεί να θεωρηθεί ως μία συνάρτηση δευτέρου βαθμού και το ολικό ελάχιστο μπορεί να υπολογιστεί εύκολα. Μέσα από την παραγωγή της συνάρτησης κόστους, ως προς το διάνυσμα x_u προκύπτει ότι κάθε διάνυσμα παραγόντων του χρήστη:

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u) \quad (2.5)$$

Με Y συμβολίζεται ο πίνακας διαστάσεων $n_i \times n_f$, ενώ το διάνυσμα $p(u)$ περιέχει τις προτιμήσεις του χρήστη u . Είναι δηλαδή η γραμμή u του πίνακα των προτιμήσεων P .

Με C^u ορίζεται ο διαγώνιος πίνακας διαστάσεων $n_i \times n_i$ που η διαγώνιος του περιέχει τα στοιχεία της γραμμής u του πίνακα C , δηλαδή $C_{ii}^u = c_{ui}$. Στις μεγάλες διαστάσεις, ο πίνακας αυτός μπορεί να προκαλέσει προβλήματα διαχείρισης μνήμης (Out of Memory), ενώ ταυτόχρονα γίνονται πάρα πολλοί περιττοί πολλαπλασιασμοί.

Ένας τρόπος για να αντιμετωπιστεί αυτό το πρόβλημα είναι η αξιοποίηση των εξής ιδιοτήτων:

- Του πολλαπλασιασμού μεταξύ ενός διαγώνιου πίνακα με έναν κανονικό πίνακα, καθώς το αποτέλεσμα που προκύπτει είναι ο πολλαπλασιασμός του i στοιχείου της διαγωνίου, με κάθε στοιχείο της i στήλης του κανονικού πίνακα.
- Του γινομένου Hadamard μεταξύ δύο πινάκων ίσων διαστάσεων, που αποτελεί τον πολλαπλασιασμό στοιχείου προς στοιχείο.

Αξιοποιώντας τις παραπάνω ιδιότητες, πραγματοποιούνται οι εξής τροποποιήσεις: Η πρώτη αποτελεί τον υπολογισμό του πολλαπλασιασμού $C_u I$ μέσα από την αντικατάσταση του διαγώνιου πίνακα $n_i \times n_i$, με το διάνυσμα διαστάσεων c_u , το οποίο περιέχει τη διαγώνιο του C^u . Στη συνέχεια, κατασκευάζεται ο πίνακας C'_u διαστάσεων $n_i \times n_f$, που κάθε στήλη του περιέχει το διάνυσμα της διαγωνίου c_u και πολλαπλασιάζεται μέσα από το γινόμενο Hadamard με τον πίνακα I . Δηλαδή:

$$C_u I = I \circ C'_u \quad (2.6)$$

Ο πίνακας C'_u περιέχει, στη γενική περίπτωση, πολύ λιγότερα στοιχεία αφού συνήθως οι διαστάσεις των χαρακτηριστικών n_f , είναι πολύ μικρότερες από το πλήθος των μουσικών κομματιών n_i ή των λιστών αναπαραγωγής n_u .

Η δεύτερη τροποποίηση αφορά το γινόμενο $C^u p(u)$, που αντικαθίσταται με το γινόμενο Hadamard $p(u) \circ c_u$,

Η συνάρτηση κόστους της (2.5) τροποποιείται:

$$x_u = [Y^T (Y \circ C'_u) + \lambda I]^{-1} Y^T p(u) \circ c_u \quad (2.7)$$

Με παρόμοιο τρόπο περιγράφεται η συνάρτηση του διανύσματος που αφορούν τους παράγοντες των αντικειμένων. Συγκεκριμένα αυτά προκύπτουν ως:

$$y_i = [X^T (X \circ C'_i) + \lambda I]^{-1} X^T p(i) \circ c_i \quad (2.8)$$

Με X συμβολίζεται ο πίνακας διαστάσεων $n_i \times n_f$, με c^i ορίζεται το διάνυσμα με διάσταση n_u που περιέχει τα στοιχεία της στήλης i του πίνακα C . Τέλος, το διάνυσμα $p(i)$ περιέχει τις προτιμήσεις των χρηστών που αφορούν το μουσικό κομμάτι i . Είναι δηλαδή η στήλη i του πίνακα των προτιμήσεων P .

Οι παραπάνω σχέσεις αναδεικνύουν ένα από τα μεγάλα πλεονεκτήματα αυτής της μεθόδου αυτής, αφού κάθε διάνυσμα παραγόντων χρήστη και αντικειμένων είναι ανεξάρτητο από τα υπόλοιπα. Είναι δυνατή επομένως η παραλληλοποίηση της διαδικασίας υπολογισμού τους, για παράδειγμα με τη χρήση κάρτας γραφικών (GPU).

2.6 Ο υπολογισμός των προτάσεων

Η πρόταση ενός αντικειμένου σε έναν χρήστη προσδιορίζεται ως το εσωτερικό γινόμενο των δύο διανυσμάτων παραγόντων. Γενικεύοντας, το σύνολο των προτάσεων για έναν δεδομένο χρήστη/λίστα αναπαραγωγής αποτελείται από τα εσωτερικά γινόμενα όλων των διανυσμάτων των παραγόντων των αντικειμένων/κομματιών.

Κεφάλαιο 3

Η συμπερασματολογία του Bayes

Εισαγωγή

Μέσα από την αύξηση της υπολογιστικής ισχύος και λόγω της ευρείας εφαρμογής σε διάφορα επιστημονικά πεδία, η Μπεϋζιανή προσέγγιση της στατιστικής έχει γίνει ιδιαίτερα δημοφιλής. Η ανάπτυξη μεθόδων για τον υπολογισμό των παραμέτρων των μοντέλων, που ήταν αδύνατη στο παρελθόν, και η πρόοδος που αφορά την έρευνα γύρω από τη μορφή των πρότερων κατανομών, έχει οδηγήσει στο καταλάγιασμα των πολύ έντονων διαφωνιών ανάμεσα στους κλασσικούς και τους Μπεϋζιανούς στατιστικολόγους.

Χρησιμοποιώντας τη Μπεϋσιανή ανάλυση, προσπαθούμε να βρούμε τρόπους που θα παράγουν συμπεράσματα από δεδομένα, χρησιμοποιώντας μοντέλα έτσι ώστε να προκύψουν συμπεράσματα για κάποια άγνωστα μεγέθη. Βασικό γνώρισμα των μεθόδων Bayes είναι η αντιμετώπιση όλων των γνωστών και άγνωστων μεγεθών ως τυχαίες μεταβλητές και η χρήση δεσμευμένων ή περιθώριων κατανομών για την προσέγγιση της εκτίμησης που έχουμε για τα διαφορετικά μεγέθη ή τις παραμέτρους των μοντέλων.

Στο παρόν κεφάλαιο θα περιγραφούν τα βασικά σημεία της Μπεϋζιανής προσέγγισης της στατιστικής και στον τρόπο που αυτή χρησιμοποιείται για να εξαχθούν συμπεράσματα. Επιπλέον, θα περιγραφεί μία τεχνική που χρησιμοποιείται για την προσέγγιση δυσεπίλυτων κατανομών που προκύπτουν, ιδιαίτερα στα μεγέθη που περιγράφονται με συνεχείς μεταβλητές πολλών διαστάσεων.

3.1 Στοιχεία Θεωρίας Πληροφορίας

Εντροπία

Στη θεωρία πληροφορίας, η εντροπία μίας τυχαίας μεταβλητής είναι το μέτρο της αβεβαιότητας που συνοδεύει τα αποτελέσματα της τυχαίας μεταβλητής. Για μια διακριτή τυχαία μεταβλητή X και συνάρτηση μάζας πιθανότητας $p(X)$ ορίζεται η εντροπία:

$$H(X) = - \sum_x p(x) \log p(x) \quad (3.1)$$

Απόκλιση Kullback Leibler (Kullback Leibler Divergence) ή Σχετική Εντροπία

Η απόκλιση Kullback Leibler αποτελεί μία μέτρηση του κατά πόσο μια κατανομή πιθανότητας είναι διαφορετική από μία άλλη πιθανότητα αναφοράς. Η απόκλιση Kullback Leibler δεν είναι συμμετρική και δεν ικανοποιεί την τριγωνική ανισότητα και συνεπώς δεν μπορεί να θεωρηθεί μετρική. Στην απλή περίπτωση που το αποτέλεσμα της απόκλισης Kullback Leibler είναι 0, οι δύο κατανομές που εξετάζονται είναι ίδιες.

Για δύο διακριτές κατανομές P και Q , ορισμένες στον ίδιο χώρο πιθανότητας, η απόκλιση Kullback Leibler από την Q μέχρι την P δίνεται από τη σχέση:

$$D_{KL}(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log \left(\frac{P(x)}{Q(x)} \right) \quad (3.2)$$

Εναλλακτικά, μπορεί να ερμηνευτεί ως η προσδοκία της λογαριθμικής διαφοράς των δύο πιθανοτήτων, με την προσδοκία να καθορίζεται βάσει της πιθανότητας P . Υπό αυτό το πρίσμα, η απόκλιση Kullback-Leibler γράφεται:

$$D_{KL}(P||Q) = \mathbb{E}_{P(x)} \left[\log \frac{P(x)}{Q(x)} \right] \quad (3.3)$$

Η απόκλιση KL είναι **πάντα μη αρνητική**. Αυτή η ιδιότητα είναι ταυτόσημη με την **ανισότητα Gibbs**, η οποία δηλώνει ότι η εντροπία $H(P)$ μιας κατανομής P είναι πάντα μεγαλύτερη ή ίση από την διασταυρούμενη εντροπία $H(P, Q)$ αυτής της κατανομής, ως προς μίας οποιαδήποτε άλλης κατανομής Q .

Έστω $P = \{p_1, \dots, p_n\}$ και $Q = \{q_1, \dots, q_n\}$. Η ανισότητα Gibbs δηλώνει ότι:

$$- \sum_{i=1}^n p_i \log p_i \geq - \sum_{i=1}^n p_i \log q_i \quad (3.4)$$

Διασταυρούμενη Εντροπία

Ένα μέγεθος που μοιάζει αρκετά και σχετίζεται με την απόκλιση Kullback Leibler είναι η διασταυρούμενη εντροπία, η οποία ορίζεται ως:

$$H(P, Q) = \mathbb{E}_{P(x)} \log Q(x). \quad (3.5)$$

Με βάση τον ορισμό της απόκλισης KL η παραπάνω σχέση μπορεί να γραφεί ως:

$$H(P, Q) = H(P) + D_{KL}(P||Q) \quad (3.6)$$

Η ελαχιστοποίηση της διασταυρούμενης εντροπίας ως προς Q ισοδυναμεί με την ελαχιστοποίηση της απόκλισης KL. Πρέπει να τονισθεί ότι η διασταυρούμενη εντροπία σχετίζεται με την εκτίμηση μέγιστης πιθανοφάνειας για τον προσδιορισμό των παραμέτρων ενός μοντέλου, καθώς η ελαχιστοποίηση της διασταυρούμενης εντροπίας είναι ταυτόσημη με την μεγιστοποίηση της συνάρτησης πιθανοφάνειας .

3.2 Το θεώρημα Bayes

Η μπεϋζιανή προσέγγιση της στατιστικής βασίζεται στην αξιοποίηση της πεποίθησης που έχουμε για την πραγματοποίηση ενός ενδεχομένου, και τη χρήση του κανόνα ή θεώρημα του Bayes για την εξαγωγή συμπερασμάτων. Το θεώρημα του Bayes περιγράφεται πιο κάτω:

Υποθέτουμε δύο ενδεχόμενα E και H τα οποία δεν είναι ανεξάρτητα μεταξύ τους. Ως E συμβολίζεται η τυχαία μεταβλητή που αντιπροσωπεύει τα **στοιχεία (evidence)**. Ως H συμβολίζεται η τυχαία μεταβλητή που αντιπροσωπεύει την **υπόθεση (hypothesis)**.

Από τα βασικά αξιώματα των πιθανοτήτων γνωρίζουμε ότι η δεσμευμένη πιθανότητα του E , δεδομένου του H δίνεται από τον τύπο:

$$P(E|H) = \frac{P(E, H)}{P(H)} \quad (3.7)$$

$$P(E, H) = P(E|H)P(H) \quad (3.8)$$

με την $P(E|H)$ να ερμηνεύεται ως "η πιθανότητα του E , δεδομένου ότι το H έχει ήδη πραγματοποιηθεί", $P(E, H)$ ως "η πιθανότητα να έχουν πραγματοποιηθεί τόσο το E όσο και το H , ενώ $p(H)$ είναι η πιθανότητα να έχει πραγματοποιηθεί το H . Με αντίστοιχο τρόπο ορίζεται και η δεσμευμένη πιθανότητα στην οποία το ενδεχόμενο E να έχει πραγματοποιηθεί πρώτο:

$$P(H|E) = \frac{P(E, H)}{P(E)} \quad (3.9)$$

$$P(H, E) = P(H|E)P(E) \quad (3.10)$$

Εξισώνοντας την (3.7) και την (3.9) προκύπτει:

$$P(H|E) = \frac{P(H)}{P(E)} P(E|H) \quad (3.11)$$

Ενώ ο κανόνας του Bayes ισχύει για οποιαδήποτε δύο ενδεχόμενα, η προσέγγιση της κλασσικής στατιστικής με αυτή της μπεϋζιανής διαφέρει. Για να γίνει πιο αντιληπτός ο τρόπος χρήσης των πιθανοτήτων από τη σκοπιά της μπεϋζιανής στατιστικής, θα ερμηνεύσουμε τα ενδεχόμενα ως εξής:

- Η μεταβλητή H χαρακτηρίζεται ως **υπόθεση** της οποίας η πιθανότητα μπορεί να επηρεάζεται από κάποια δεδομένα (ή αλλιώς στοιχεία)
- Το στοιχείο E είναι ένα **νέο στοιχείο** που δεν είχε χρησιμοποιηθεί για τον υπολογισμό της πρότερης πιθανότητας
- $P(H)$ ορίζεται ως η **πρωταρχική ή πρότερη (prior)** πιθανότητα, που είναι η πιθανότητα του H πριν την παρατήρηση του E .
- $P(H|E)$ ορίζεται ως η **μεταγενέστερη (posterior)** πιθανότητα, που είναι η πιθανότητα του H , αφού έχει γίνει η παρατήρηση του E .
- $P(E|H)$ ονομάζεται η πιθανότητα της παρατηρητής των στοιχείων E δεδομένης της υπόθεσης H και ονομάζεται **πιθανοφάνεια**
- $P(E)$ η οποία αποτελεί τα **στοιχεία μοντέλου**. Η πιθανότητα αυτή είναι ίδια για τις διαφορετικές υποθέσεις που κάνουμε και επομένως δεν επηρεάζει τις σχετικές πιθανότητες των διαφορετικών υποθέσεων.

3.3 Εφαρμογή του κανόνα του Bayes

3.3.1 Βασικές έννοιες

Η στατιστική συμπερασματολογία ασχολείται με εξαγωγή συμπερασμάτων, μέσα από τη χρήση δεδομένων, για κάποια μεγέθη που δεν έχουν παρατηρηθεί. Διαφοροποιούμε δύο κατηγορίες των μεγεθών αυτών. Η πρώτη περιλαμβάνει τα μεγέθη που είναι πιθανώς παρατηρήσιμα, όπως για παράδειγμα οι μελλοντικές παρατηρήσεις μιας διαδικασίας, ενώ η δεύτερη αφορά τα μεγέθη που δεν είναι άμεσα παρατηρήσιμα, όπως για παράδειγμα οι παράμετροι που προσδιορίζουν μια διαδικασία. Η διάκριση μεταξύ των δύο αυτών μεγεθών δεν είναι πάντα ακριβής, αλλά αποτελεί έναν χρήσιμο τρόπο για να γίνει αντιληπτό πως ένα στατιστικό μοντέλο προσεγγίζει την πραγματικότητα.

Θα θεωρήσουμε ως θ τα μη παρατηρήσιμα μεγέθη/παραμέτρους (που μπορεί να έχουν τη μορφή διανύσματος) ενός μοντέλου, ως x τα παρατηρήσιμα δεδομένα ενώ ως \tilde{x} τα άγνωστα αλλά ενδεχομένως παρατηρήσιμα μεγέθη. Στην Μπεϋζιανή προσέγγιση, το συμπέρασμα για μια παράμετρο θ ή για δεδομένα \tilde{x} που δεν έχουν παρατηρηθεί ακόμα, γίνεται μέσα από μία δήλωση πιθανότητας. Αυτή η πιθανότητα είναι δεσμευμένη ως προς την παρατηρήσιμη τιμή της μεταβλητής x και ορίζεται ως $p(\theta|x)$ ή $p(\tilde{x}|x)$.

Για χάρη ευκολίας, θα ακολουθηθεί ο ακόλουθος συμβολισμός: Με $p(*|*)$ θα συμβολίζεται η δεσμευμένη πυκνότητα πιθανότητας των τυχαίων μεταβλητών που εισάγονται ως ορίσματα της σχέσης, ενώ αντίστοιχα ορίζεται ως $p(*)$ την περιθώρια κατανομή. Οι όροι κατανομή και πυκνότητα θα χρησιμοποιούνται ισοδύναμα.

Καθορισμός του μοντέλου

Για δύο τυχαίες μεταβλητές θ και x , ορίζεται η από κοινού κατανομή του θ και x . Στόχος είναι να περιγραφεί ένα μοντέλο από το οποίο θα προκύπτουν τα συμπεράσματα για τη δεσμευμένη κατανομή του θ με δεδομένο το x . Η από κοινού κατανομή είναι το γινόμενο δύο κατανομών, της πρότερης κατανομής (prior) $p(\theta)$ και της δεσμευμένης κατανομής $p(x|\theta)$ και ονομάζεται δειγματική κατανομή.

$$p(\theta, x) = p(\theta)p(x|\theta) \quad (3.12)$$

Η πρότερη κατανομή (Prior Distribution)

Η πρότερη κατανομή $p(\theta)$ αποτελεί την αρχική πεποίθηση για μία άγνωστη τυχαία μεταβλητή πριν ληφθούν υπόψη δεδομένα που αφορούν αυτή τη μεταβλητή. Μία πρότερη κατανομή μπορεί να κατασκευαστεί με διαφορετικούς τρόπους. Για παράδειγμα μπορεί να είναι το αποτέλεσμα μιας πληροφορίας που αποκτήθηκε στο παρελθόν ή μπορεί να δημιουργηθεί από μία εντελώς υποκειμενική υπόθεση.

Η δειγματική κατανομή (Sample Distribution)

Η δειγματική κατανομή $p(x|\theta)$ ή συνάρτηση πιθανοφάνειας $\mathcal{L}(\theta|x)$ αποτελεί την κατανομή της τυχαίας μεταβλητής των δεδομένων σε σχέση με τις παραμέτρους θ .

Η μεταγενέστερη κατανομή (Posterior Distribution)

Με εφαρμογή του θεωρήματος του Bayes προκύπτει:

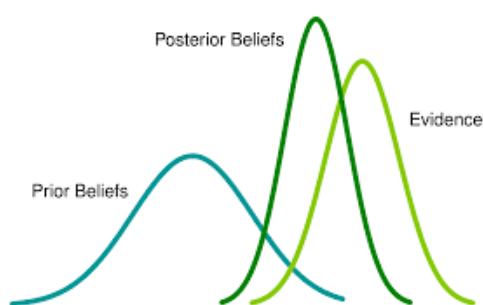
$$p(\theta, x) = p(x, \theta) \quad (3.13)$$

$$p(x|\theta)p(\theta) = p(\theta|x)p(x) \quad (3.14)$$

$$p(\theta|x) = \frac{p(\theta)}{p(x)}p(x|\theta) \quad (3.15)$$

Με $p(x)$ να ονομάζεται η **περιθώρια κατανομή των δεδομένων** (marginal likelihood και $p(x) = \sum_{\theta} p(\theta)p(y|\theta)$ για διακριτές μεταβλητές ή $p(x) = \int_{\theta} p(\theta)p(x|\theta)d\theta$ για συνεχείς μεταβλητές.

Η κατανομή $p(\theta|x)$ ονομάζεται **μεταγενέστερη κατανομή**. Εκφράζει πλέον τη γνώση μας γύρω από την υπόθεση μας, αφού έχουμε λάβει υπόψη τα στοιχεία x



Σχήμα 3.1: Οι σχέσεις μεταξύ των κατανομών, με βάση την μπεϋζιανή στατιστική

Το πρόβλημα της δυσεπίλυτης περιθωριας κατανομής

Η περιθώρια κατανομή γράφεται:

$$p(x) = \int_{\theta} p(x|\theta)p(\theta)d\theta \quad (3.16)$$

Αν η τυχαία μεταβλητή θ είναι πολυδιάστατη, με $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ τότε το παραπάνω ολοκλήρωμα είναι πολύ δύσκολο να υπολογισθεί. Αφού η απειροστή ποσότητα δίνεται από τη σχέση $d\theta = d\theta_1 d\theta_2 \dots$, ο υπολογισμός της $p(x)$ ισοδυναμεί με τον υπολογισμό N ολοκληρωμάτων, που μπορεί, όσο αυξάνεται το μέγεθος των διαστάσεων, να μην μπορεί να υπολογιστεί. Αυτό το πρόβλημα είναι αρκετά πιθανό να εμφανιστεί σε εφαρμογές βαθιάς μηχανικής μάθησης και σε μεγάλα Dataset. Ο τρόπος με τον οποίο αντιμετωπιζόταν μέχρι και αρκετά πρόσφατα ήταν μέσα από αλγόριθμους δειγματοληψίας, που παρ' όλα αυτά έδιναν αποτελέσματα με μεγάλη διακύμανση και με αργή σύγκλιση. Αυτές οι προσεγγίσεις βασίζονται στη χρήση Markov Chain Monte Carlo (Neal, 1993 [13]). Τα πλεονεκτήματα αυτών των μεθόδων είναι η αρκετά απλή υλοποίησή τους και η εξασφάλιση της σύγκλισης τους. Ωστόσο τα μειονεκτήματά τους είναι η αργή σύγκλιση και το ασταφές κριτήριο σύγκλισης.

3.4 Εκτίμηση Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimation)

Η εκτίμηση μέγιστης πιθανοφάνειας αποτελεί ένα κριτήριο για την εύρεση των καλύτερων παραμέτρων ενός στατιστικού μοντέλου. Η επιλογή τους πραγματοποιείται μέσα από την μεγιστοποίηση μίας συνάρτησης πιθανοφάνειας, που περιγράφεται μέσα από τη συνάρτηση πυκνότητας/μάζας πιθανότητας μίας στατιστικής κατανομής ενός συνόλου δεδομένων.[5]

Για την περιγραφή της μεθόδου, θεωρείται αρχικά ένα σύνολο δεδομένων $\mathbf{X} = \{x^{(i)}\}_{i=1}^N$ με N ανεξάρτητα και πανομοιότυπα διανεμημένα (Independent and Identically Distributed - i.i.d.), μίας άγνωστης κατανομής $p_{data}(\mathbf{x})$ που παράγει δεδομένα.

Ορίζεται επιπλέον μία παραμετρική οικογένεια κατανομών πιθανότητας $p_{model}(\mathbf{x}; \theta)$. Ο εκτιμητής μέγιστης πιθανοφάνειας για το θ ορίζεται ως:

$$\theta_{ML} = \arg \max_{\theta} p_{model}(\mathbf{X}; \theta) \quad (3.17)$$

$$= \arg \max_{\theta} \prod_i^N p_{model}(\mathbf{x}_i; \theta) \quad (3.18)$$

Για διάφορους λόγους, όπως λ.χ. η αριθμητική ευστάθεια, η παραπάνω εξίσωση συνήθως εκφράζεται σε λογαριθμική μορφή:

$$\theta_{ML} = \arg \max_{\theta} \sum_i^N \log p_{model}(\mathbf{x}_i; \theta) \quad (3.19)$$

Διαιρώντας με το πλήθος των στοιχείων N και επειδή τα δεδομένα είναι διανεμημένα με ίδιο τρόπο (άρα έχουν ίδια πιθανότητα ίση με $\frac{1}{N}$), η (3.19) μπορεί να εκφραστεί ως η προσδοκώμενη τιμή ως προς την κατανομή πιθανότητας \hat{p}_{model} . Η διαίρεση με έναν αριθμό δεν αλλοιώνει την κατάταξη των παραμέτρων θ :

$$\theta_{ML} = \arg \max_{\theta} \mathbb{E}_{x \sim \hat{p}_{model}} \log p_{model}(\mathbf{x}; \theta) \quad (3.20)$$

Ένας τρόπος να ερμηνευθεί ο εκτιμητής μέγιστης πιθανοφάνειας είναι ως την μείωση της απόστασης μεταξύ των κατανομών \hat{p}_{data} και της κατανομής του μοντέλου. Δηλαδή η ελαχιστοποίηση της απόκλισης Kullback Leibler:

$$D_{KL}(\hat{p}_{data} || p_{model}) = \mathbb{E}_{x \sim \hat{p}_{model}} [\log \hat{p}_{data}(\mathbf{x}) - \log p_{model}(\mathbf{x})] \quad (3.21)$$

Επειδή η συνάρτηση πυκνότητας πιθανότητας \hat{p}_{data} αφορά την διαδικασία παραγωγής των δεδομένων και όχι τις παραμέτρους του μοντέλου, η ελαχιστοποίηση της απόκλισης εξαρτάται μόνο από την ελαχιστοποίηση του δεύτερου όρου, που είναι η διασταυρούμενη εντροπία. Με αυτό τον τρόπο, η εκτίμηση μέγιστης πιθανοφάνειας μπορεί να εκφραστεί ως η ελαχιστοποίηση της αρνητικής λογαριθμικής πιθανότητας, μέσα από τη χρήση της διασταυρούμενης εντροπίας.[6]

3.5 Variational Inference

Οι τεχνικές Variational Inference αποτελούν μία πρόταση για την επίλυση του προβλήματος των δυσεπίλυτων περιθώριων κατανομών στα μπεϋζιανά μοντέλα. Η κεντρική ιδέα των τεχνικών αυτών είναι η αναζήτηση και η εύρεση κάποιων κατανομών που, σε αντίθεση με την περιθώρια κατανομή των στοιχείων, είναι επιλύσιμες και την προσεγγίζουν σε ικανοποιητικό βαθμό. Οι τεχνικές αυτές έχουν δύο βασικούς στόχους:

1. Να δοθεί μία αναλυτική προσέγγιση της μεταγενέστερης κατανομής των μη παρατηρήσιμων μεγεθών έτσι ώστε να εξαχθούν συμπεράσματα για αυτές.
2. Να προκύψει ένα κατώτερο όριο της περιθώριας κατανομής των παρατηρήσιμων δεδομένων. Αυτό δίνει τη δυνατότητα της επιλογής του πιο κατάλληλου μοντέλου, ιδιαίτερα όταν χρειάζεται να γίνει επιλογή παραμέτρων.

3.5.1 Βασική Προσέγγιση

Έστω η μεταγενέστερη κατανομή $p(\boldsymbol{\theta}|\mathbf{y})$ των μη παρατηρήσιμων τυχαίων μεταβλητών $\boldsymbol{\theta} = \{\theta_1 \dots \theta_2\}$ με δεδομένα κάποια \mathbf{y} . Έστω τώρα μια κατανομή $q(\boldsymbol{\theta})$ που τη προσεγγίζει, η οποία ανήκει σε μια παραμετρική οικογένεια κατανομών με πιο απλή μορφή από αυτή της μεταγενέστερης κατανομής. Η ομοιότητα των δύο κατανομών μετράται μέσα από μια συνάρτηση διαφοράς. Η κατάλληλη $q(\boldsymbol{\theta})$ επιλέγεται μέσα από την ελαχιστοποίηση αυτής της συνάρτησης. Από τις πιο ευρέως εφαρμοσμένες συναρτήσεις για την εύρεση της κατάλληλης προσέγγισης είναι η απόκλιση Kullback Leibler.

3.5.2 Υπολογισμός της προσεγγιστικής κατανομής με τη χρήση της απόκλισης Kullback Leibler

Αφού η απόκλιση KL αποτελεί έναν τρόπο σύγκρισης δύο κατανομών, τότε διαισθητικά η απόκλιση KL της μεταγενέστερης κατανομής προς την προσέγγιση της αποτελεί ένα κριτήριο ομοιότητας τους. Τότε:

$$D_{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{x})) = \sum_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) \log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{x})} \quad (3.22)$$

$$= \sum_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) [\log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}, \mathbf{x})}] + \sum_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) \log p(\mathbf{x}) \quad (3.23)$$

$$(3.24)$$

Ο όρος $p(\mathbf{x})$ είναι ανεξάρτητος της μεταβλητής του αθροίσματος και επιπλέον το άθροισμα $\sum_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) = 1$, αφού η $q(\boldsymbol{\theta})$ είναι κατανομή, άρα:

$$D_{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{x})) = \sum_{\boldsymbol{\theta}} q(\boldsymbol{\theta}) [\log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}, \mathbf{x})}] + \log p(\mathbf{x}) \quad (3.25)$$

$$(3.26)$$

Αναδιατάσσοντας τους όρους και λαμβάνοντας υπόψη τον ορισμό της εκτιμώμενης τιμής προκύπτει ότι:

$$\log p(\mathbf{x}) = D_{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{x})) - \mathbb{E}_{\boldsymbol{\theta}}[\log q(\boldsymbol{\theta}) - \log p(\boldsymbol{\theta}, \mathbf{x})] \quad (3.27)$$

$$= D_{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{x})) - \mathcal{L}(q) \quad (3.28)$$

Με $\mathcal{L}(q)$ ορίζεται το κάτω όριο της περιθώριας κατανομής (Evidence Lower Bound - ELBO).

Επειδή η περιθώρια κατανομή $p(\boldsymbol{\theta})$ είναι σταθερή ως προς την κατανομή q , η μεγιστοποίηση του \mathcal{L} οδηγεί στην ελαχιστοποίηση της απόκλισης KL της μεταγενέστερης κατανομής προς την προσέγγιση της. Με κατάλληλη επιλογή της προσέγγισης q , το κάτω όριο μπορεί να υπολογισθεί αναλυτικά.

Κεφάλαιο 4

Αυτοκωδικοποιητές Διακύμανσης (Variational Autoencoders)

Εισαγωγή

Οι Variational Autoencoders αποτελούν μοντέλα βαθιών νευρωνικών δικτύων μη επιβλεπόμενης μάθησης εμπρόσθιας τροφοδότησης. Όπως και οι απλοί Αυτοκωδικοποιητές (Αυτοενζόδες) αποτελούνται από δύο επιμέρους δίκτυα. Έναν **κωδικοποιητή** (Encoder), ο οποίος μετατρέπει το διάνυσμα εισόδου σε ένα διάνυσμα μιας λανθάνουσας μεταβλητής συνήθως μικρότερης διάστασης, και έναν **αποκωδικοποιητή** (Decoder) ο οποίος ανασυνθέτει την είσοδο από την κρυφή μεταβλητή, συμβάλλοντας στην κατασκευή της συνάρτησης κόστους και επομένως στην εκπαίδευση του δικτύου.

Σε αντίθεση όμως με τους απλούς Autoencoders, οι VAE σχεδιάζονται με τέτοιο τρόπο έτσι ώστε η είσοδος του αποκωδικοποιητή να αποτελεί **τυχαία μεταβλητή** μίας στατιστικής κατανομής, ενώ η συνάρτηση σφάλματος και κατά συνέπεια το κριτήριο βελτιστοποίησης διαφέρει.

Ενώ το βασικό πεδίο υλοποίησης τους είναι η λειτουργία ως μοντέλα παραγωγής/αναπαραγωγής (Generative Models), στην παρούσα εργασία εξετάζεται η εφαρμογή τους στο Συνεργατικό Φιλτράρισμα.

4.1 Απλοί Αυτοκωδικοποιητές (Autoencoders)

Οι απλοί Autoencoders αποτελούν νευρωνικά δίκτυα μη επιβλεπόμενης μάθησης. Η βασική λειτουργία τους είναι η αναπαραγωγή της εισόδου, στην έξοδο τους με τη χρήση ενός ενδιάμεσου, λανθάνοντος (latent) επιπέδου. Ο κωδικοποιητής, αρχικά, αναλαμβάνει να απεικονίσει την είσοδο σε ένα διάνυσμα μιας μεταβλητής μικρότερων διαστάσεων, ενώ ύστερα ο αποκωδικοποιητής προσπαθεί να ανακατασκευάσει την έξοδο, ακολουθώντας την αντίστροφη διαδικασία.

Κατά τη διάρκεια της κωδικοποίησης, χρησιμοποιείται ένας ρυθμιστικός παράγοντας L1 ή L2 έτσι ώστε η έξοδος να είναι μια **προσέγγιση** της εισόδου και όχι η ακριβής αναπαραγωγή της.

4.1.1 Βασική Αρχιτεκτονική

Η πιο απλή υλοποίηση έχει ένα επίπεδο εισόδου, ένα επίπεδο εξόδου ίδιων διαστάσεων και τουλάχιστον ένα λανθάνον επίπεδο που τα συνδέει. Όπως έχει ήδη περιγραφεί, το πρώτο κομμάτι του δικτύου, ο κωδικοποιητής έχει ως ρόλο την κωδικοποίηση της πληροφορίας στις διαστάσεις του λανθάνοντος επιπέδου, ενώ το δεύτερο

κομμάτι, ο αποκωδικοποιητής, έχει ως στόχο την αναπαραγωγή της πληροφορίας από αυτό. Έστερα, υπολογίζεται το σφάλμα σφάλμα μεταξύ εισόδου και εξόδου, το οποίο και χρησιμοποιείται για την διόρθωση των βαρών και την εκπαίδευση του δικτύου. Επειδή το σφάλμα δεν περιέχει κάποιου άλλου είδους πληροφορία (π.χ κλάση), οι Autoencoders εκπαιδεύονται με μη επιβλεπόμενο τρόπο.

Συνεπώς μπορούμε να ορίσουμε τις τρεις συναρτήσεις:

$$\begin{aligned}\phi &: \mathcal{X} \rightarrow \mathcal{F} \\ \psi &: \mathcal{F} \rightarrow \mathcal{X} \\ \phi, \psi &= \arg \min_{\phi, \psi} \|X - (\psi \circ \phi)X\|^2\end{aligned}$$

Στην πιο απλή περίπτωση, με μόνο ένα κρυφό επίπεδο, η είσοδος $\mathbf{x} \in \mathbb{R}^d = \mathcal{X}$ μετατρέπεται μέσα από τον κωδικοποιητή στο διάνυσμα $\mathbf{h} \in \mathbb{R}^p = \mathcal{F}$, με:

$$\mathbf{h} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b})$$

Η μεταβλητή σ αποτελεί τη συνάρτηση ενεργοποίησης του κάθε νευρώνα, η μεταβλητή \mathbf{W} συμβολίζει τον πίνακα βαρών και η μεταβλητή \mathbf{b} το διάνυσμα των ρυθμιστικών παραγόντων. Ο πίνακας βαρών και το διάνυσμα με τους ρυθμιστικούς παράγοντες αρχικοποιούνται με τυχαίες τιμές στην αρχή και οι τελικές τιμές υπολογίζονται κατά τη διαδικασία εκπαίδευσης.

Αφού υπολογιστεί η τιμή του διανύσματος των λανθάνουσων μεταβλητών, ξεκινάει η διαδικασία της αποκωδικοποίησης, υπολογίζοντας το διάνυσμα εξόδου ως:

$$\mathbf{x}' = \sigma'(\mathbf{W}'\mathbf{h} + \mathbf{b}')$$

με τα σ' , \mathbf{W}' , \mathbf{b}' μεταβλητές που είναι ασυσχέτιστες με τις αντίστοιχες του κωδικοποιητή.

Η συνάρτηση κόστους δίνεται από την συνάρτηση:

$$\mathcal{L}(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|^2 = \|\mathbf{x} - \sigma'(\mathbf{W}'(\sigma(\mathbf{W}\mathbf{x} + \mathbf{b})) + \mathbf{b}')\|^2$$

Οι Autoencoders εκπαιδεύονται για να ελαχιστοποιούν το σφάλμα ανακατασκευής εισόδου - εξόδου, ενώ η διαδικασία της εκπαίδευσης γίνεται με την προς τα πίσω μετάδοση του σφάλματος (Backpropagation).

4.1.2 Πεδία εφαρμογής των Αυτοκωδικοποιητών

Μία από τις βασικές εφαρμογές των αυτοκωδικοποιητών είναι η χρήση τους για την μείωση της διαστατικότητας (Dimensionality Reduction) ενός συνόλου δεδομένων, αξιοποιώντας τη μη γραμμική συμπεριφορά ως δίκτυα βαθιάς μάθησης.

Οι κλασικοί Autoencoders έχουν αρκετούς περιορισμούς. Ένας από αυτούς έγκειται στο ότι είναι ευάλωτοι σε υπερεκπαίδευση, ειδικά αν εισαχθούν αρκετά επίπεδα. Ένας ακόμη σοβαρός περιορισμός είναι ότι ο χώρος που σχηματίζει το διάνυσμα της κρυφής μεταβλητής δεν είναι ενιαίος, και ως αποτέλεσμα δεν έχουν καμία αναγεννητική δυνατότητα. Τέλος, συγκριτικά με την Ανάλυση σε Κύριες Συνιστώσες (Principal Component Analysis) η απεικόνιση των δεδομένων στο κρυφό επίπεδο δεν αποτελεί ορθογώνιο χώρο.

4.1.3 Αυτοκωδικοποιητές Αποθρομβοποίησης (Denoising Autoencoders)

Οι Denoising Autoencoders αποτελούν μία υποκατηγορία αυτοκωδικοποιητών που έχουν ως στόχο να περιορίσουν την δυνατότητα των απλών αυτοκωδικοποιητών που τείνουν να αναπαράγουν την είσοδο στην έξοδο και άρα να περιορίζεται η υπερεκπαίδευση. Ο τρόπος με τον οποίο επιτυγχάνεται αυτό είναι μέσα από την αλλοίωση της εισόδου, με την εισαγωγή θορύβου. Με αυτό τον τρόπο, το κρυφό στρώμα, μέσα από τη διαδικασία μάθησης, μπορεί να ανακαλύψει πιο ισχυρά χαρακτηριστικά, σε σχέση με τον αντίστοιχο απλό κωδικοποιητή.

Οι DAE έχουν την ίδια αρχιτεκτονική με τους απλούς αυτοκωδικοποιητές με μία απλή διαφοροποίηση, αυτή της εισαγωγής του σταδίου αλλοίωσης εισόδου. Για κάθε στοιχείο x δημιουργείται μία αλλοιωμένη έκδοση $\hat{x} = \text{corrupt}(x)$. Η συνάρτηση μπορεί να είναι οποιαδήποτε συνάρτηση μπορεί να αλλοιώσει την είσοδο και εξαρτάται από τον τύπο εισόδου. Παραδείγματα τέτοιων συναρτήσεων είναι ο θόρυβος Gauss.

Μία αρκετά διαδεδομένη τεχνική για την αλλοίωση της εισόδου είναι η χρήση ενός επιπέδου Dropout. Η μέθοδος λειτουργεί με την στοχαστική απενεργοποίηση των νευρώνων του επιπέδου που εφαρμόζεται, μέσα από τον μηδενισμό της εξόδου τους και την κατάλληλη αναπροσαρμογή των τιμών στους νευρώνες που δεν απενεργοποιούνται, έτσι ώστε το συνολικό άθροισμα των τιμών του στοιχείων να παραμένει σταθερό. Χρησιμοποιείται ευρέως για την αποτροπή της υπερεκπαίδευσης των βαθιών νευρωνικών δικτύων.

4.2 Αυτοκωδικοποιητές Διακύμανσης (Variational Autoencoders)

Ακολουθώντας τα βήματα των Kingma et al. [14], θα αναπτυχθεί ένα μπεύζιανό μοντέλο που έχει ως στόχο την εύρεση των παραμέτρων του, με αποτελεσματικό τρόπο, αντιμετωπίζοντας και επιλύοντας βασικά προβλήματα της μπεύζιανής ανάλυσης. Για αυτό το λόγο θα αξιοποιηθούν νευρωνικά δίκτυα. Η μέθοδος αυτή βασίζεται στην κατασκευή ενός απλού, αμερόληπτου και διαφορίσιμου εκτιμητή Stochastic Gradient Variational Bayes, ο οποίος θα προκύψει από την χρήση ενός απλού τριών επαναπαραμετροποίησης του κατώτατου ορίου ELBO. Το όριο αυτό θα εξαχθεί με τη χρήση Variational Bayes Inference της προσεγγιστικής μεταγενέστερης κατανομής.

Οι παράγωγοι του αμερόληπτου και διαφορίσιμου εκτιμητή υπολογίζονται με αποτελεσματικό τρόπο και μέσα από την εφαρμογή αλγορίθμων βελτιστοποίησης (π.χ. Adam), προσδιορίζονται οι άγνωστες παράμετροι των νευρωνικών δικτύων.

Στη συνέχεια θα περιγραφεί ο Variational Autoencoder, τεχνητού νευρωνικού δικτύου που έχει αρκετά κοινά με έναν αυτοκωδικοποιητή ως συγκεκριμένη εφαρμογή του μπεύζιανού μοντέλου που περιγράφηκε, με συγκεκριμένες υποθέσεις γύρω από την προηγούμενη κατανομή, την μεταγενέστερη κατανομή και την συνάρτηση πιθανοφάνειας. Με αυτό τον τρόπο, κατασκευάζεται ένα δίκτυο που έχει ευρείες εφαρμογές, όπως για παράδειγμα στην αναγνώριση και παραγωγή εικόνων, ή, όπως στην περίπτωση μας, στο συνεργατικό φιλτράρισμα.

4.2.1 Προσέγγιση του Προβλήματος με χρήση τυχαίων μεταβλητών

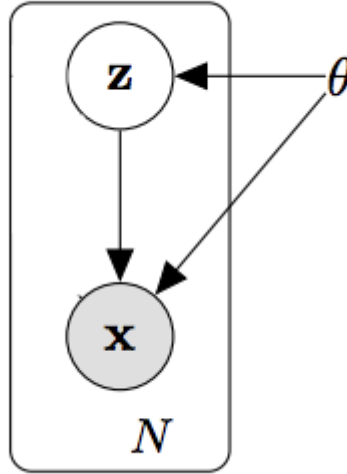
Ορίζεται ένα σύνολο δεδομένων $\mathbf{X} = \{x^{(i)}\}_{i=1}^N$ με N ανεξάρτητα και πανομοιότυπα διανεμημένα (Independent and Identically Distributed - i.i.d.) δείγματα μιας συνεχούς ή διακριτής μεταβλητής. Θεωρούμε ότι τα δεδομένα αυτά παράγονται από μια τυχαία διαδικασία, στην οποία εμπλέκεται μια λανθάνουσα συνεχής τυχαία μεταβλητή z .

1. Από την πρότερη (prior) κατανομή $p_{\theta^*}(z)$ προκύπτει τα δείγματα $z^{(i)}$.
2. Από την κατανομή δεσμευμένης πιθανότητας $p_{\theta^*}(x|z)$, προκύπτουν τα δείγματα $x^{(i)}$.

Έστω ότι, επιπλέον ότι, η προηγούμενη και η δειγματική κατανομή προέρχονται από τις παραμετρικές οικογένειες κατανομών $p_{\theta}(z)$ και $p_{\theta}(x|z)$ και ότι οι συναρτήσεις πυκνότητας πιθανότητας είναι παραγωγίσιμες και συνεχείς για κάθε θ, z . Ωστόσο τόσο οι πραγματικές τιμές των παραμέτρων θ^* όσο και του διανύσματος της λανθάνουσας μεταβλητής \mathbf{z} δεν θεωρούνται γνωστές.

Οι βασικοί στόχοι της μεθόδου που θα περιγραφεί είναι οι εξής:

Η μέθοδος πρέπει να λειτουργεί και στις περιπτώσεις που η περιθώρια κατανομή $p(x)$ (ή η μεταγενέστερη κατανομή $p_{\theta}(z|x)$) είναι δύσκολο να υπολογισθεί, λόγω της πολυπλοκότητας του ολοκληρώματος.



Σχήμα 4.1: Το στατιστικό μοντέλο Bayes

Επιπλέον, πρέπει να μπορεί να αντιμετωπίζει με αποτελεσματικότητα τη συνεχή αύξηση του μεγέθους συνόλων δεδομένων, που αποτελεί και τα μεγαλύτερα προβλήματα εναλλακτικών αλγορίθμων που χρησιμοποιούν δειγματοληψία, λόγω του πολύ μεγάλου υπολογιστικού κόστους (όπως για παράδειγμα ο Monte Carlo EM αλγόριθμος)

Με αυτό τον τρόπο, η μέθοδος που περιγράφεται επιλύει τρία συγγενή μεταξύ τους προβλήματα:

1. Την αποτελεσματική προσέγγιση της μέγιστης συνάρτησης πιθανοφάνειας για τις παραμέτρους θ του δικτύου.
2. Την αποτελεσματική προσέγγιση του συμπεράσματος για την μεταγενέστερη κατανομή της τυχαίας μεταβλητής z , με πεδία εφαρμογής τα προβλήματα αναπαράστασης δεδομένων.
3. Την αποτελεσματική προσέγγιση για τα συμπεράσματα της περιθώριας κατανομής της τυχαίας μεταβλητής x .

4.2.2 Προσέγγιση της Μεταγενέστερης Κατανομής

Θεωρούμε μία νέα κατανομή $q_\phi(z|x)$ που αποτελεί μια προσέγγιση της δυσεπίλυτης μεταγενέστερης κατανομής $p_\theta(z|x)$. Θα προσεγγίσουμε το πρόβλημα με τον εξής τρόπο: Θεωρείται η **απόκλιση Kullback-Leibler** (Kullback–Leibler divergence) της $p_\theta(z|x)$ ως προς $q_\phi(z|x)$. Για να προσεγγίζει αυτή με ικανοποιητική προσέγγιση της μεταγενέστερης πρέπει η απόκλιση KL να ελαχιστοποιηθεί, μέσα από την επιλογή των κατάλληλων παραμέτρων ϕ και θ .

$$\begin{aligned}
 D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z}|\mathbf{x})) &= \mathbb{E}_q[\log q_\phi(\mathbf{z}|\mathbf{x})] - \mathbb{E}_q[\log p(\mathbf{z}|\mathbf{x})] \\
 &= \mathbb{E}_q[\log q_\phi(\mathbf{z}|\mathbf{x})] - \mathbb{E}_q[\log p(\mathbf{x}|\mathbf{z})] - \mathbb{E}_q[\log p(\mathbf{z})] + \mathbb{E}_q[\log p(\mathbf{x})] \\
 &= \mathbb{E}_q[\log q_\phi(\mathbf{z}|\mathbf{x}) - \log p(\mathbf{x}, \mathbf{z})] + \log p(\mathbf{x}) \\
 &= D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})) - \mathbb{E}_q[\log p(\mathbf{x}|\mathbf{z})] + \log p(\mathbf{x})
 \end{aligned}$$

Ο υπολογισμός της απόκλισης KL είναι επίσης δύσκολος, επειδή εμφανίζεται ο όρος της περιθώριας κατανομής $p(\mathbf{x})$. Παρόλα αυτά τώρα είμαστε σε θέση να προσπεράσουμε τη δυσκολία αυτή, μέσα από το Evidence Lower Bound.

4.3 Evidence Lower Bound

Επιλύοντας την εξίσωση του προηγούμενου κεφαλαίου ως προς $\log p(x)$:

$$\log p(\mathbf{x}) = D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(z|x)) + \mathbb{E}_q[\log p(x|z)] - D_{KL}(q_\phi(z|x)||p_\theta(z)) \quad (4.1)$$

Ενώ ο όρος $D_{KL}(q_\phi(z|x)||p_\theta(z|x))$ δεν είναι εύκολα υπολογίσιμος, γνωρίζουμε ότι είναι μη αρνητικός, αφού η απόκλιση KL είναι πάντα μεγαλύτερη ή ίση του μηδενός. Ορίζουμε το Evidence Lower Bound ως $\mathcal{L} = \mathbb{E}_q[\log p(x|z)] - D_{KL}(q_\phi(z|x)||p_\theta(z))$ και η περιθώρια κατανομή μπορεί να γραφεί:

$$\log p(x) = D_{KL}(q_\phi(z|x)||p_\theta(z|x)) + \underbrace{\mathbb{E}_q[\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x)||p_\theta(z))}_{ELBO} \quad (4.2)$$

Με την λογαριθμική πιθανότητα της περιθώριας κατανομής να είναι:

$$\log p(x) \geq \mathcal{L} \Leftrightarrow \quad (4.3)$$

$$\log p(x) \geq \mathbb{E}_q[\log p(x|z)] - D_{KL}(q_\phi(z|x)||p_\theta(z)) \quad (4.4)$$

Η συνάρτηση \mathcal{L} είναι εν τέλει η συνάρτηση που πρέπει να μεγιστοποιήσουμε, έτσι ώστε να μπορέσουν να προσεγγιστούν οι παράμετροι ϕ και θ .

4.3.1 Το ELBO ως συνάρτηση σφάλματος

Το Evidence Lower Bound έχει ορισθεί ως:

$$ELBO = \mathbb{E}_q[\log p(x|z)] - D_{KL}(q_\phi(z|x)||p_\theta(z)) \quad (4.5)$$

Όπως φαίνεται, η μεγιστοποίηση της παραπάνω σχέσης αντιστοιχεί στην εκτίμηση της μέγιστης πιθανοφάνειας και στην ελαχιστοποίηση της απόκλισης Kullback Leibler μεταξύ της προηγούμενης κατανομής και της προσέγγισης της μεταγενέστερης κατανομής. Ο πρώτος όρος μπορεί να ερμηνευθεί ως ο όρος ανακατασκευής (reconstruction term) ενώ ο δεύτερος όρος μπορεί να ερμηνευθεί ως ένας ρυθμιστικός όρος (bias term) που εμποδίζει το δίκτυο από υπερεκπαίδευση:

$$ELBO = \underbrace{\mathbb{E}_q p(x|z)}_{\text{Reconstruction term}} - \underbrace{D_{KL}(q_\phi(z|x)||p_\theta(z))}_{\text{bias term}} \quad (4.6)$$

Με βάση αυτή την οπτική, το κάτω όριο αποτελεί ένα πρόβλημα μεγιστοποίησης της συνάρτησης πιθανοφάνειας με έναν επιπλέον όρο μεροληψίας.

4.3.2 Μεγιστοποίηση του ELBO

Ο υπολογισμός όλων των i.i.d. στοιχείων του συνόλου δεδομένων δίνεται ως το άθροισμα:

$$\mathcal{L}_{\theta,\phi}(\mathcal{D}) = \sum_{\mathbf{x} \in \mathcal{D}} \mathcal{L}_{\theta,\phi}(\mathbf{x}) \quad (4.7)$$

Στη γενική περίπτωση ο υπολογισμός του ELBO είναι δυσεπίλυτος. Είναι όμως δυνατό να κατασκευαστούν αμερόληπτοι εκτιμητές που να προσεγγίζουν τόσο το όριο, που να είναι αμερόληπτοι και παραγωγίσιμοι.

Παραγωγίζοντας τη συνάρτηση ELBO με βάση τις παραμέτρους θ :

$$\nabla_{\theta} \mathcal{L}_{\theta, \phi}(\mathbf{x}) = \nabla_{\theta} \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}|\mathbf{z}) + \log p_{\theta}(\mathbf{z}) - \log q_{\phi}(\mathbf{z}|\mathbf{x})] \quad (4.8)$$

$$= \nabla_{\theta} \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}, \mathbf{z}) - \log q_{\phi}(\mathbf{z}|\mathbf{x})] \quad (4.9)$$

$$= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\nabla_{\theta} (\log p_{\theta}(\mathbf{x}, \mathbf{z}) - \log q_{\phi}(\mathbf{z}|\mathbf{x}))] \quad (4.10)$$

Από την τελευταία εξίσωση μπορεί να προκύψει ο απλός εκτιμητής:

$$\nabla_{\theta} \log p_{\theta}(\mathbf{x}, \mathbf{z}) \quad (4.11)$$

Η εύρεση ενός αμερόληπτου εκτιμητή ως προς τις παραμέτρους ϕ είναι πιο δύσκολο να κατασκευαστούν. Πιο συγκεκριμένα:

$$\nabla_{\phi} \mathcal{L}_{\theta, \phi}(\mathbf{x}) = \nabla_{\phi} \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}|\mathbf{z}) + \log p_{\theta}(\mathbf{z}) - \log q_{\phi}(\mathbf{z}|\mathbf{x})] \quad (4.12)$$

$$\neq \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\nabla_{\phi} (\log p_{\theta}(\mathbf{x}|\mathbf{z}) + \log p_{\theta}(\mathbf{z}) - \log q_{\phi}(\mathbf{z}|\mathbf{x}))] \quad (4.13)$$

4.3.3 Το τρικ επαναπαραμετροποίησης (Reparameterization Trick)

Για να παρακάμψουμε το πρόβλημα με τον εκτιμητή της παραγώγου ως προς ϕ χρησιμοποιούμε μια σχετικά απλή ιδέα. Αν η μεταβλητή $\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})$ εκφρασθεί ως ένας διαφορίσιμος και αντιστρέψιμος μετασχηματισμός μίας τυχαίας μεταβλητής ϵ , με την τυχαία αυτή μεταβλητή να είναι ανεξάρτητη των \mathbf{x} και ϕ , δηλαδή:

$$\mathbf{z} = \mathbf{g}(\epsilon, \phi, \mathbf{x}) \quad (4.14)$$

Η αναμενόμενη τιμή μπορεί να γραφεί ως:

$$\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [f(\mathbf{z})] = \mathbb{E}_{p(\epsilon)} [f(\mathbf{z})] \quad (4.15)$$

με τους τελεστές της αναμενόμενης τιμής και της παραγώγου να μπορούν να αντιμετατεθούν. Επομένως:

$$\nabla_{\phi} \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [f(\mathbf{z})] = \nabla_{\phi} \mathbb{E}_{p(\epsilon)} [f(\mathbf{z})] \quad (4.16)$$

$$= \mathbb{E}_{p(\epsilon)} \nabla_{\phi} [f(\mathbf{z})] \quad (4.17)$$

Από την παραπάνω σχέση μπορεί να προκύψει ο απλός εκτιμητής $\nabla_{\phi} [f(\mathbf{z})]$, ο οποίος είναι παραγωγίσιμος ως προς και τις δύο παραμέτρους.

Συνδυάζοντας τους δύο εκτιμητές, προκύπτει ο τελικός εκτιμητής^[15] για το ELBO ως:

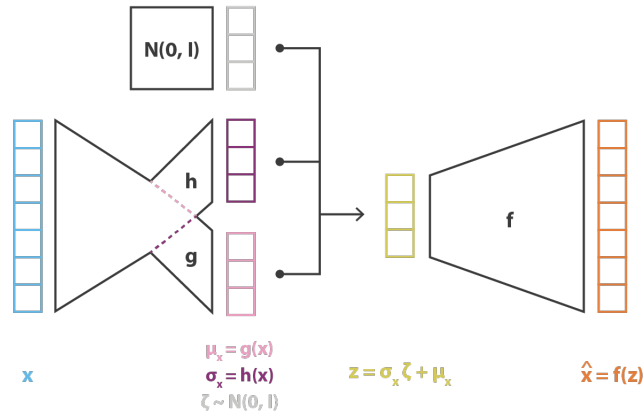
$$\begin{aligned} \tilde{\mathcal{L}}_{\theta, \phi}(\mathbf{x}) &= \log p_{\theta}(\mathbf{x}, \mathbf{z}) - \log q_{\phi}(\mathbf{z}|\mathbf{x}) \\ &\text{με } \mathbf{z} = \mathbf{g}(\phi, \epsilon, \mathbf{x}) \text{ και } \epsilon \sim p(\epsilon) \end{aligned} \quad (4.18)$$

4.4 Η υλοποίηση ενός Variational Autoencoder

Σε αυτή την ενότητα θα περιγράψουμε μια υλοποίηση του παραπάνω στατιστικού μοντέλου, με τη χρήση δύο νευρωνικών δικτύων, για τον υπολογισμό της μεταγενέστερης κατανομής και της συνάρτησης πιθανοφάνειας με βάση την τεχνική της Variational Inference.

Αρχικά, θα κάνουμε τις εξής υποθέσεις. Θεωρούμε ότι η πρότερη κατανομή $p(\mathbf{z})$ είναι μια κατανομή Gauss πολλών μεταβλητών, με $p(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I})$. Η επιλογή της πιθανοφάνειας $p_{\theta}(\mathbf{x}|\mathbf{z})$ είναι πιο σύνθετη. (π.χ. Gauss για τις περιπτώσεις δεδομένων με πραγματικές τιμές μπορεί ή Bernulli για τις περιπτώσεις δεδομένων με δυαδικές τιμές). Ο υπολογισμός των παραμέτρων θ ισοδυναμεί με τον υπολογισμό των παραμέτρων του νευρωνικού δικτύου (για παράδειγμα MLP ή CNN) που θα χρησιμοποιηθεί στον αποκωδικοποιητή.

Τέλος, για την βελτιστοποίηση του ELBO αξιοποιείται ο εκτιμητής της εξίσωσης 4.18. Σημειώνεται ότι επειδή οι κατανομές $q(\mathbf{z}|\mathbf{x})$ και $p(\mathbf{z})$ ακολουθούν κανονική κατανομή, τότε η απόκλιση Kullback Leibler μπορεί εναλλακτικά να υπολογισθεί με αναλυτικό τρόπο.



$$\text{loss} = C \|x - \hat{x}\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)] = C \|x - f(z)\|^2 + \text{KL}[N(g(x), h(x)), N(0, I)]$$

4.4.1 Η πλευρά του κωδικοποιητή

Η πλευρά του κωδικοποιητή υλοποιεί την μεταγενέστερη κατανομή. Θα υποθέσουμε ότι η προσέγγιση $q_\phi(\mathbf{z}|\mathbf{x})$ της δυσεπίλυτης προγενέστερης κατανομής, είναι και αυτή κατανομή Gauss πολλών μεταβλητών. Για λόγους απλούστευσης, θεωρούμε ότι ο πίνακας συνδιακύμανσης είναι διαγώνιος. Οι παράμετροι ϕ αντίστοιχα βελτιστοποιούνται μέσα από τη βελτιστοποίηση των παραμέτρων του δικτύου του κωδικοποιητή.

$$\log q_\phi(\mathbf{z}|\mathbf{x}^{(i)}) = \log \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{2(i)}\mathbf{I}) \quad (4.19)$$

Με τις μεταβλητές $\boldsymbol{\mu}^{(i)}$ και $\boldsymbol{\sigma}^{(i)}$ να είναι οι έξοδοι του κωδικοποιητή.

4.4.2 Η προηγούμενη κατανομή και η δειγματοληψία της τυχαίας μεταβλητής \mathbf{z}

Το διάνυσμα της μεταβλητής \mathbf{z} προσδιορίζεται μέσα από τη χρήση του τριχ επαναπαραμετροποίησης, αξιοποιώντας την τυχαία μεταβλητή $\epsilon \sim \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I})$. Έτσι τελικά για τη δημιουργία του διανύσματος της κρυφής μεταβλητής $\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x}^{(i)})$ δειγματοληπτούμε χρησιμοποιώντας την βοηθητική τυχαία μεταβλητή ϵ :

$$\mathbf{z}^{(i,l)} = g_\phi(\mathbf{x}^{(i)}, \epsilon^{(l)}) = \boldsymbol{\mu}^{(i)} + \boldsymbol{\sigma}^{(i)} \odot \epsilon^{(l)} \quad (4.20)$$

4.4.3 Η πλευρά του αποκωδικοποιητή

Η πλευρά του αποκωδικοποιητή υλοποιεί την συνάρτηση πιθανοφάνειας. Όπως έχει περιγραφεί από τη συνάρτηση στόχου του δικτύου, ο σκοπός είναι η μεγιστοποίηση της ή ισοδύναμα, η ελαχιστοποίηση της διασταυρούμενης εντροπίας μεταξύ των αποτελεσμάτων του δικτύου και της εισόδου. Η επιλογή της κατανομής της πιθανοφάνειας εξαρτάται κάθε φορά με το πρόβλημα που πρέπει να επιλυθεί.

1. Αν η έξοδος αποτελεί συνεχή μεταβλητή, μία αρκετά διαδεδομένη επιλογή είναι η (πολυδιάστατη) γκαουσιανή κατανομή.
2. Αν η έξοδος παίρνει δυαδική τιμή, τότε μπορεί να υποτεθεί μία κατανομή Bernoulli.
3. Αν η έξοδος ανήκει σε μία από πολλές κλάσεις, τότε μπορεί να υποτεθεί μία πολυωνυμική κατανομή.

4.4.4 Η διαδικασία της εκπαίδευσης

Έστω το τυχαίο υποσύνολο \mathbf{X}^M , με $\{x^{(i)}\}_{i=1}^M$ του συνόλου δεδομένων.

Η εκπαίδευση του Αυτοκωδικοποιητή Διακύμανσης πραγματοποιείται μέσα από τη μεγιστοποίηση του ELBO. Με χρήση της (4.18), η μεγιστοποίηση του ELBO υπολογίζεται μέσα από τον Monte Carlo εκτιμητή:

$$\tilde{\mathcal{L}}_{\theta,\phi}(\mathbf{X}^M) = \frac{1}{M} \sum_{i=1}^M \tilde{\mathcal{L}}_{\theta,\phi}(x^{(i)}) \quad (4.21)$$

Με τον υπολογισμό των παραγώγων, χρησιμοποιείται ένας αλγόριθμος βελτιστοποίησης (π.χ. Adam) για τον από κοινού υπολογισμό των παραμέτρων των δύο νευρωνικών δικτύων.

Κεφάλαιο 5

Σύνολο Δεδομένων

Εισαγωγή

Το σύνολο δεδομένων που χρησιμοποιείται οργανώνει την πληροφορία σε λίστες αναπαραγωγής που περιέχουν μουσικά κομμάτια. Ωστόσο, ο τρόπος αυτής της οργάνωσης είχε ως αποτέλεσμα την αδυναμία εύρεσης ενός έτοιμου συνόλου δεδομένων, αφού όσα παρεμφερή βρέθηκαν, οργάνωναν την πληροφορία με διαφορετικό τρόπο (για παράδειγμα, αντί για λίστες αναπαραγωγής υπήρχαν καλλιτέχνες κ.ο.κ.).

Σε αυτό το κεφάλαιο, θα περιγραφεί η διαδικασία για την εύρεση, την αποθήκευση, την οργάνωση και την επεξεργασία των δεδομένων καθώς και η δημιουργία του συνόλου δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση των μοντέλων.

5.1 Εύρεση λιστών αναπαραγωγής

Το πρώτο βήμα για την κατασκευή του συνόλου δεδομένων και τη συλλογή των απαραίτητων χαρακτηριστικών αποτέλεσε η συλλογή των μοναδικών αριθμών αναφοράς (Identity Document-ID) των λιστών αναπαραγωγής από το Spotify. Μέσω της χρήσης της Διεπαφής Προγραμματισμού Εφαρμογών (API), έγινε αναζήτηση με κριτήριο το όνομα των λιστών για κάθε πιθανό συνδυασμό δύο αλφαριθμητικών της αγγλικής γλώσσας.

Η αναζήτηση δεν έγινε με το πιο απλό κριτήριο της χρήσης μόνο ενός γράμματος, λόγω του περιορισμού των αποτελεσμάτων που επιστρέφει η διεπαφή του Spotify για κάθε ερώτημα (1000 λίστες). Με το παραπάνω κριτήριο αναζήτησης διαμορφώθηκε μία λίστα με 1 εκατομμύριο μοναδικούς αριθμούς αναφοράς.

Για την χρήση της διεπαφής του Spotify χρησιμοποιήθηκε η βιβλιοθήκη της Python `spotify`.

5.2 Συλλογή και αποθήκευση των δεδομένων

Για την αποθήκευση και καλύτερη οργάνωση της αντλούμενης πληροφορίας, χρησιμοποιήθηκε μία βάση δεδομένων `Mongodb`. Η `Mongodb` χαρακτηρίζεται ως μία βάση `NoSQL` που οργανώνει την πληροφορία σε έγγραφα (documents) μορφής `Binary JSON (BSON)`, και οργανώνει τα έγγραφα σε συλλογές (collections). Ένα από τα μεγάλα πλεονεκτήματα αυτής της βάσης είναι η ευελιξία στο σχήμα (schema) που χρησιμοποιείται για την οργάνωση των δεδομένων μέσα στις συλλογές. Για την επαφή με τη βάση χρησιμοποιήθηκε η βιβλιοθήκη της Python, `pymongo`.

Το σχήμα που επιλέχθηκε για την οργάνωση της πληροφορίας αποτυπώνεται παρακάτω:

```

  _id: "0YtYcGZCo4cXEWebdUntJ1"
  name: "When Hell Freezes Over"
  tracks: Array
    0: "0i74VE2iKY9ynQaBPzkzAm"
    1: "4gvea7UldkAvsJBpZAd4oB"
    2: "78Pny505RDkHJNPzf1QcFs"
    3: "00hb20GQiyWVPSMprmbTm"
    4: "2vIRpCpsekN5TiWq3Ua2Ht"
    5: "2wvMC5EyaayQwBfiwY2xE"
    6: "1NSjv4Ji9MktuQDXHAH4k0"
    7: "7yk8I8Z10EW2KsEsPoxMUI"
    8: "3vhJw6V39Gh1PpEMHjiMFO"
    9: "2v7qe4TkWOM0upTIDcSnd3"
    10: "6zodLniefdaHk6QoreGwKT"

  _id: "0JPlgRBDGRteFR8G4FeCG5"
  name: "Starless"
  playlists: Array
    0: "08Im3ExQugNOgcls56c10x"
    1: "095sRx1bQLh2SVVldsslPU"
    2: "0Py2Ch0fxMRFz5AWIVPLgb"
    3: "0En0MaEmpeoqesQB3RV3if"
    4: "0VSNIMDmqowcAjKqIrSX5b"
    5: "05w0PAbC9wsteIq0PDV1on"
    6: "0Ds0g7x1wVRPk0JNp3GM2h"
    7: "0FBJRlUXWapJgyC3YwD2pV"
    8: "0BuAJGe3B1Uj7y3oPscGw9"
    9: "0AFSyRLbw7m37TZqPHeZj4"
    10: "0NM1MHWk8zBwzCbJxGfrz9"

```

(α) Ένα έγγραφο λίστας αναπαραγωγής (β) Ένα έγγραφο μουσικού κομματιού

Σχήμα 5.1: Παραδείγματα εγγράφων στο MongoDB Compass

Για κάθε λίστα αναπαραγωγής, μέσα από τον μοναδικό αριθμό ID και τη χρήση της διεπαφής API του Spotify, συλλέχθηκαν τα πεδία του ονόματος και των μουσικών κομματιών που περιέχονται σε αυτή και καταχωρήθηκαν σε μία συλλογή. Κάθε έγγραφο περιέχει τις εξής πληροφορίες:

- `_id`: Τον μοναδικό αριθμό ID του Spotify, που χρησιμοποιείται και ως ο δείκτης αναφοράς.
- `name`: Το όνομα της λίστας αναπαραγωγής.
- `tracks`: Οι μοναδικοί αριθμοί ID του Spotify των μουσικών κομματιών που περιέχονται σε αυτή.

Για κάθε κομμάτι της λίστας αναπαραγωγής, εξετάστηκε η προϋπαρξη αντίστοιχης καταχώρησης στην αντίστοιχη συλλογή των μουσικών κομματιών. Σε περίπτωση απουσίας του, γινόταν νέα καταχώρηση, ενώ στην περίπτωση επιτυχούς αναζήτησης του κομματιού, ενημερωνόταν το πεδίο `playlists` της εγγραφής, με την προσθήκη του μοναδικού αριθμού αναφοράς ID της λίστας αναπαραγωγής. Κάθε έγγραφο της συλλογής περιέχει:

- `_id`: Τον μοναδικό αριθμό ID του Spotify, που χρησιμοποιείται και ως ο δείκτης αναφοράς.
- `name`: Το όνομα του κομματιού.
- `playlists`: Οι μοναδικοί αριθμοί ID του Spotify των λιστών αναπαραγωγής που το συγκεκριμένο μουσικό κομμάτι περιέχεται.

Ο τελικός αριθμός των αποθηκευμένων στη βάση λιστών αναπαραγωγής είναι 650175, ενώ τα αποθηκευμένα κομμάτια είναι πάνω από 13 εκατομμύρια. Στη διαδικασία που περιγράφηκε παραπάνω, υπήρχαν αρκετοί κωδικοί μοναδικής αναφοράς μουσικών κομματιών και λιστών αναπαραγωγής που δεν αντιστοιχούσαν σε πραγματικά δεδομένα και απορρίπτονταν.

Το μέγεθος της αποθηκευμένης πληροφορίας είναι πάνω από 9 GigaByte.

5.3 Το σύνολο δεδομένων

Το σύνολο δεδομένων που χρησιμοποιείται στη διαδικασία εκπαίδευσης και ελέγχου των μοντέλων έχει τη μορφή πίνακα. Οι γραμμές του αντιστοιχούν στις λίστες αναπαραγωγής ενώ οι στήλες του αντιστοιχούν στα μουσικά

κομμάτια. Για κάθε μουσικό κομμάτι που υπάρχει σε μία λίστα, η αντίστοιχη τιμή του πίνακα τίθεται 1, ενώ σε διαφορετική περίπτωση τίθεται 0.

Βάσει αυτού του κριτηρίου σχηματισμού του συνόλου δεδομένων προκύπτουν τα εξής χαρακτηριστικά:

1. Το σύνολο δεδομένων αποτελείται από δεδομένα έμμεσης αλληλεπίδρασης (implicit data). Ως άμεσης αλληλεπίδρασης θεωρούνται τα δεδομένα που δημιουργούνται λόγω της ηθελημένης και ενεργητικής συμμετοχής ενός χρήστη με τη χρήση κάποιας μεθόδου. Παραδείγματα αυτών αποτελεί η βαθμολόγηση βάσει μίας κλίμακας, η ταξινόμηση από το καλύτερο μέχρι το χειρότερο, οι αναζητήσεις του χρήστη κ.ο. Αντιθέτως, η συλλογή χαρακτηριστικών με έμμεσο τρόπο πραγματοποιείται με κριτήρια που βασίζονται στην έμμεση συμπεριφορά του χρήστη. Τέτοια κριτήρια είναι η παρακολούθηση των αντικειμένων που ένας χρήστης βλέπει (για παράδειγμα, τα προϊόντα μέσα σε ένα ηλεκτρονικό κατάστημα), η ανάλυση του χρόνου που ένας χρήστης βλέπει ένα αντικείμενο, το ιστορικό αγορών ή των βίντεο που βλέπει και η ανάλυση του κοινωνικού δικτύου.
2. η φύση των δεδομένων δημιουργεί ένα πάρα πολύ αραιό πίνακα, που προκαλέσει εν δυνάμει προβλήματα στην εκπαίδευση των μοντέλων και στη διαχείριση της μνήμης.

Δύο σύνολα δεδομένων

Στην παρούσα εργασία κατασκευάστηκαν δύο σύνολα δεδομένων. Το πρώτο σύνολο δεδομένων που δημιουργήθηκε είχε σχετικά λίγες λίστες αναπαραγωγής και έναν περιορισμένο αριθμό μουσικών κομματιών και χρησιμοποιήθηκε για την ανάπτυξη των πρώτων μοντέλων. Το δεύτερο σύνολο δεδομένων αξιοποίησε πολλές περισσότερες λίστες αναπαραγωγής, με τα μουσικά κομμάτια που αυτό περιέχει να είναι πάνω από τριπλάσια. Στόχος ήταν η βελτίωση της απόδοσης των μοντέλων και η ανάπτυξη της δυνατότητας εκπαίδευσης τους με δεδομένα που δεν χωρούν εξολοκλήρου στη μνήμη. Τονίζεται ότι, ενώ τα περισσότερα μουσικά κομμάτια είναι και κοινά, το μικρό σύνολο δεδομένων δεν αποτελεί υποσύνολο του μεγάλου συνόλου.

Η διαδικασία κατασκευής του συνόλου δεδομένων

Το πρώτο βήμα της κατασκευής του συνόλου δεδομένων πραγματοποιείται μέσα από ένα ερώτημα στη βάση δεδομένων για την αναζήτηση όλων των μουσικών κομματιών που εμφανίζονται στις λίστες αναπαραγωγής, περισσότερες φορές από ένα καθορισμένο όριο. Για το μεγάλο σύνολο δεδομένων ο αριθμός αυτός τέθηκε στις 2500 λίστες αναπαραγωγής, ενώ για το μικρό σύνολο δεδομένων, το όριο τέθηκε στις 200.

Και για τα δύο σύνολα δεδομένων σχηματίζονται:

1. Μία λίστα με όλους τους μοναδικούς αριθμούς ID τις λίστες αναπαραγωγής που περιέχουν έστω και ένα από τα παραπάνω κομμάτια.
2. Ένα λεξικό που περιέχει ως κλειδιά τους μοναδικούς αριθμούς ID των μουσικών κομματιών και ως τιμές μία λίστα με τους μοναδικούς αριθμούς ID των λιστών αναπαραγωγής που αυτή περιέχει.

Για το μεγάλο σύνολο δεδομένων, ο αριθμός των μοναδικών λιστών αναπαραγωγής είναι 482.284 και των μουσικών κομματιών είναι 5.546. Στη συνέχεια, γίνεται τυχαία δειγματοληψία 100.000 λιστών αναπαραγωγής, για τη μείωση των διαστάσεων και της μνήμης και αφαιρούνται λίστες αναπαραγωγής που έχουν λιγότερα από 5 μουσικά κομμάτια.

Για το μικρό σύνολο δεδομένων, ο αριθμός των μοναδικών λιστών αναπαραγωγής είναι 17.086 και ο αριθμός των μουσικών κομματιών είναι 1.890.

Για την παραπάνω διαδικασία χρησιμοποιήθηκαν οι βιβλιοθήκες Pandas και Numpy της Python.

Η ολοκλήρωση του συνόλου δεδομένων μπορεί να πραγματοποιηθεί με δύο τρόπους:

Πυκνή Μορφή

Ο πρώτος τρόπος είναι η υλοποίηση του σε πυκνή (dense) μορφή: Αξιοποιώντας το λεξικό που περιγράφηκε παραπάνω είναι εύκολο να κατασκευαστεί ένα πλαίσιο δεδομένων (Dataframe) της Pandas. Στη συνέχεια, πραγματοποιείται ο διαχωρισμός του σε τρία μέρη:

- Το σύνολο δεδομένων εκπαίδευσης (Train Dataset), που αποτελεί το 80% του συνολικού και χρησιμοποιείται για την εκπαίδευση των μοντέλων.
- Το σύνολο δεδομένων επικύρωσης (Validation Dataset), που αποτελεί το 10% του συνολικού και χρησιμοποιείται ως κριτήριο εκπαίδευσης και δεν χρησιμοποιείται άμεσα για την εκπαίδευση των βαρών. Η εκπαίδευση σταματάει όταν τα αποτελέσματα του μοντέλου με βάση αυτό το σύνολο δεδομένων αρχίζουν και χειροτερεύουν.
- Το σύνολο δεδομένων ελέγχου (Test Dataset), που αποτελεί το τελικό 10% του συνολικού και χρησιμοποιείται για τον τελικό έλεγχο των αποτελεσμάτων του μοντέλου. Αποτελούν δεδομένα που δεν έχουν χρησιμοποιηθεί είτε με άμεσο είτε με έμμεσο τρόπο στη διαδικασία εκπαίδευσης του μοντέλου.

Αραιή Μορφή

Ο δεύτερος τρόπος είναι η υλοποίηση σε αραιή (sparse) μορφή. Η αραιή μορφή αποθηκεύει μόνο τα μη μηδενικά στοιχεία ενός πίνακα:

- Ο διαχωρισμός του συνόλου δεδομένων γίνεται με τον διαχωρισμό της λίστας των μοναδικών λιστών αναπαγωγής σε τρία μέρη (80%, 10%, 10%) που αντιπροσωπεύουν το διαχωρισμό σε εκπαίδευση, επικύρωση και έλεγχο.
- Δημιουργείται μία λίστα με όλα τα μουσικά κομμάτια. Μπορεί να δημιουργηθεί σχετικά εύκολα αξιοποιώντας τα κλειδιά του λεξικού που περιγράφηκε πιο πάνω. Για κάθε λίστα που αντιστοιχεί σε ένα από τα υποσύνολα δεδομένων, εντοπίζονται οι θέσεις (indexes) των μουσικών κομματιών. Για παράδειγμα, αν ένα μουσικό κομμάτι με index 3 στη λίστα των μουσικών κομματιών βρίσκεται στην τέταρτη λίστα αναπαγωγής, τότε δημιουργείται το ζεύγος (4,3).
- Το σύνολο όλων αυτών των ζευγών σχηματίζει με εύκολο τρόπο μία αναπαράσταση αραιού πίνακα με τη χρήση της βιβλιοθήκης Scipy της Python.

Το βασικό μειονέκτημα της υλοποίησης με χρήση πυκνής μορφής είναι το μεγάλο μέγεθος των πινάκων που δημιουργούνται. Είναι ενδεικτικό ότι το μέγεθος του συνόλου δεδομένων εκπαίδευσης χρειάστηκε 1.3 GB για την αποθήκευση σε πυκνή μορφή και μόλις 9.23 MB σε αραιή μορφή.

Αντίστοιχα, το βασικό μειονέκτημα της υλοποίησης με χρήση αραιής μορφής είναι ότι δημιουργείται δυσκολία στον χειρισμό και επεξεργασία των δεδομένων. Ένας τρόπος αντιμετώπισης αυτού του ζητήματος είναι η χρήση παρτίδων (batches) για την εκπαίδευση των μοντέλων, μέσα από το πλαίσιο διαχείρισης δεδομένων που προσφέρει η πλατφόρμα του Tensorflow.

Οι λίστες αναπαγωγής που είχαν τα τελικά σύνολα δεδομένων ήταν 63.460 για το σύνολο εκπαίδευσης και από 7.933 για τα σύνολα επικύρωσης και ελέγχου.

Διαχωρισμός των μουσικών κομματιών στα σύνολα δεδομένων επικύρωσης και ελέγχου

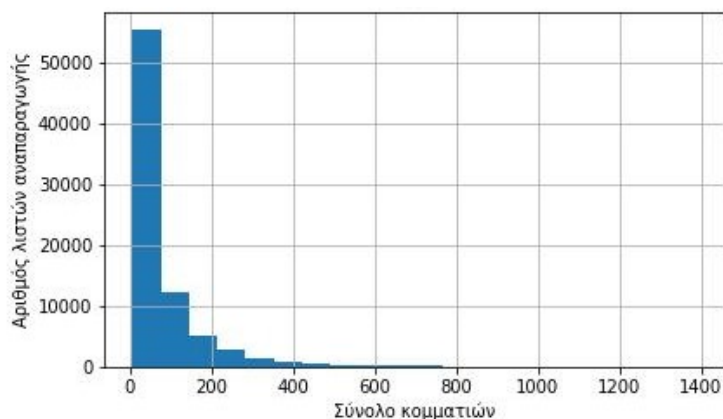
Για τη αξιολόγηση των αποτελεσμάτων των μοντέλων, υλοποιήθηκε εκ νέου διαχωρισμός των μουσικών κομματιών στα σύνολα δεδομένων και επικύρωσης. Για κάθε λίστα αναπαγωγής του συνόλου δεδομένων, επιλέγεται το 20% των μουσικών κομματιών, το οποίο αφαιρείται από αυτό και δημιουργείται ένα νέο σύνολο δεδομένων.

Αυτή η διαδικασία πραγματοποιείται για τον υπολογισμό των μετρικών. Το σύνολο με τα περισσότερα δεδομένα χρησιμοποιείται ως είσοδος στο μοντέλο, και τα αποτελέσματα συγκρίνονται με το δεύτερο σύνολο, δηλαδή τα δεδομένα που έχουν κρατηθεί. Με αυτό τον τρόπο εξετάζεται η δυνατότητα των μοντέλων να προτείνουν μουσικά κομμάτια που υπήρχαν στις αρχικές λίστες αναπαραγωγής.

Στα παρακάτω γραφήματα παρουσιάζονται ο διαχωρισμός σε κλάσεις των λιστών αναπαραγωγής, βάσει του αριθμού των μουσικών κομματιών που αυτές περιέχουν, και παρομοίως, ο διαχωρισμός σε κλάσεις των μουσικών κομματιών βάσει του αριθμού των λιστών αναπαραγωγής που αυτά περιέχονται και δίνονται οι αντίστοιχοι πίνακες.

Κλάσεις	[5, 144]	(144, 282]	(282, 420]	(420, 558]	(558, 697]	(697, 835]	(835, 1388]
Αριθμός λιστών αναπαραγωγής	67890	7976	2177	793	282	118	89

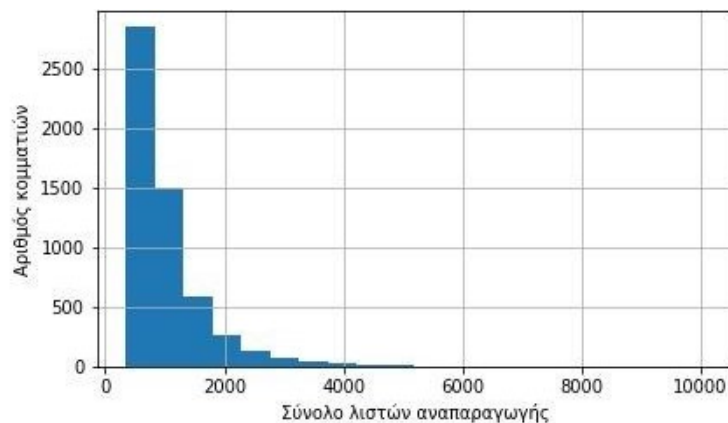
Πίνακας 5.1: Διαχωρισμός των λιστών αναπαραγωγής του μεγάλου συνόλου δεδομένων βάσει του αριθμού των μουσικών κομματιών που περιέχουν.



Σχήμα 5.2: Ο αριθμός των λιστών αναπαραγωγής στο μεγάλο σύνολο δεδομένων βάσει του αριθμού των μουσικών κομματιών που περιέχουν.

Κλάσεις	(332, 1313]	(1313, 2284]	(2284, 3255]	(3255, 4226]	(4226, 5197]	(5197, 10053]
Αριθμός μουσικών κομματιών	4340	863	212	81	28	22

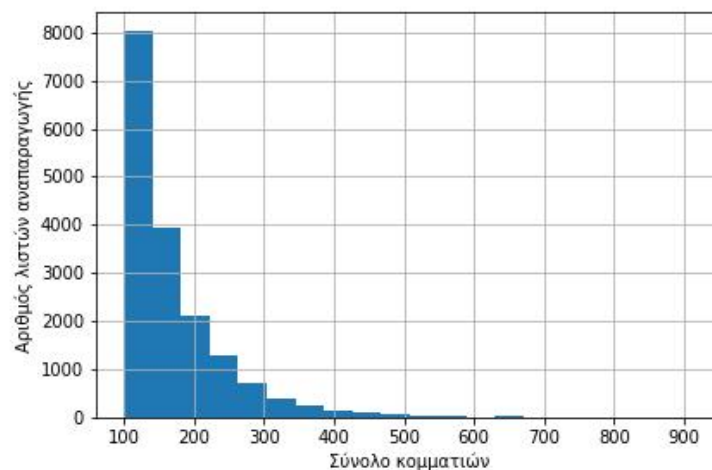
Πίνακας 5.2: Διαχωρισμός των μουσικών κομματιών του μεγάλου συνόλου δεδομένων βάσει του αριθμού των λιστών αναπαραγωγής που αυτά περιέχονται



Σχήμα 5.3: Ο αριθμός των κομματιών στο μεγάλο σύνολο δεδομένων βάσει του αριθμού των λιστών αναπαραγωγής που αυτά περιέχονται.

Κλάσεις	(100, 182]	(182, 263]	(263, 344]	(344, 426]	(426, 507]	(507, 588]	(588, 914]
Αριθμός λιστών αναπαραγωγής	11978	3401	1114	381	135	46	33

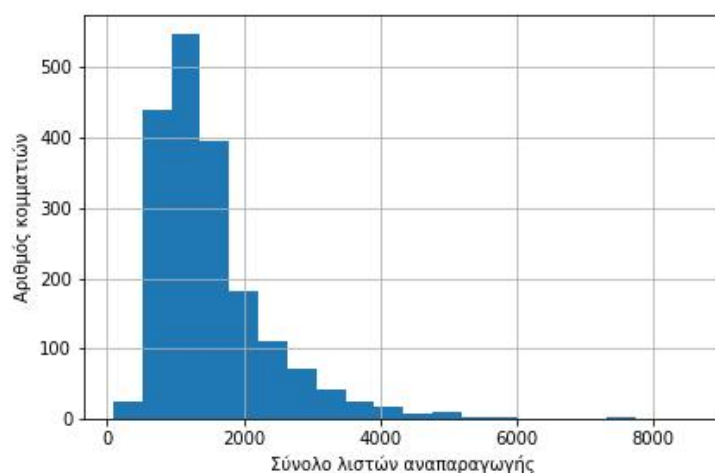
Πίνακας 5.3: Διαχωρισμός των λιστών αναπαραγωγής του μικρού συνόλου δεδομένων βάσει του αριθμού των μουσικών κομματιών που περιέχονται.



Σχήμα 5.4: Ο αριθμός των λιστών αναπαραγωγής στο μικρό σύνολο δεδομένων βάσει το σύνολο των κομματιών που περιέχονται.

Κλάσεις	(83, 941]	(941, 1791]	(1791, 2640]	(2640, 3490]	(3490, 4339]	(4339, 8587]
Αριθμός μουσικών κομματιών	466	943	292	115	41	33

Πίνακας 5.4: Διαχωρισμός των μουσικών κομματιών του μικρού συνόλου δεδομένων βάσει του αριθμού των λιστών αναπαραγωγής που αυτά περιέχονται.



Σχήμα 5.5: Ο αριθμός των κομματιών στο μικρό σύνολο δεδομένων βάσει του αριθμού των λιστών αναπαραγωγής που περιέχονται.

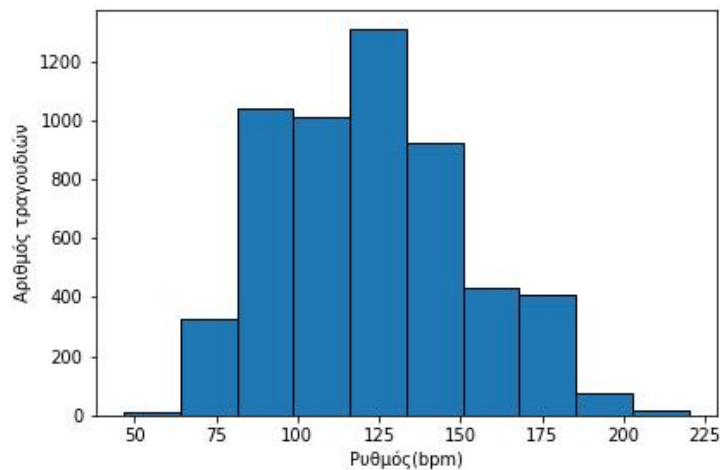
5.4 Χαρακτηριστικά Μουσικών Κομματιών (Track Features)

Τα χαρακτηριστικά των μουσικών κομματιών αποτελούν ποιοτικές σχέσεις που προκύπτουν με την ανάλυση στον ήχο τους από το Spotify. Διαμορφώνουν κριτήρια γύρω από τη διάθεση, τις ιδιότητες και το πλαίσιο και τα ακουστικά χαρακτηριστικά των μουσικών κομματιών. Στόχος είναι η αξιοποίηση τους για την ποιοτική ανάλυση των προτάσεων των μοντέλων που αναπτύχθηκαν. Η συλλογή των χαρακτηριστικών πραγματοποιήθηκε μέσα από την διεπαφή (API) του Spotify. Τα χαρακτηριστικά που περιγράφονται παρακάτω αφορούν τα μουσικά κομμάτια του μεγάλου συνόλου δεδομένων.

Ταχύτητα (Tempo)

Η συνολική εκτιμώμενη ταχύτητα του μουσικού κομματιού σε χτύπους ανά λεπτό. Η ταχύτητα ενός μουσικού κομματιού καθορίζει τον ρυθμό ενός κομματιού ως προς την διαδοχή ενός παλμού ή χτύπου από τον επόμενο, με τις μικρότερες τιμές να σημαίνουν πιο αργή διαδοχή, ενώ αντίθετα οι μεγάλες τιμές υποδηλώνουν ένα γρήγορο ρυθμικά κομμάτι.

Κλάσεις	(46, 64]	(64, 81]	(81, 98]	(98, 116]	(116, 133]	(133, 150]	(150, 168]	(168, 185]	(185, 202]	(202, 220]
Αριθμός τραγουδιών	10	329	1038	1009	1310	921	432	408	76	13

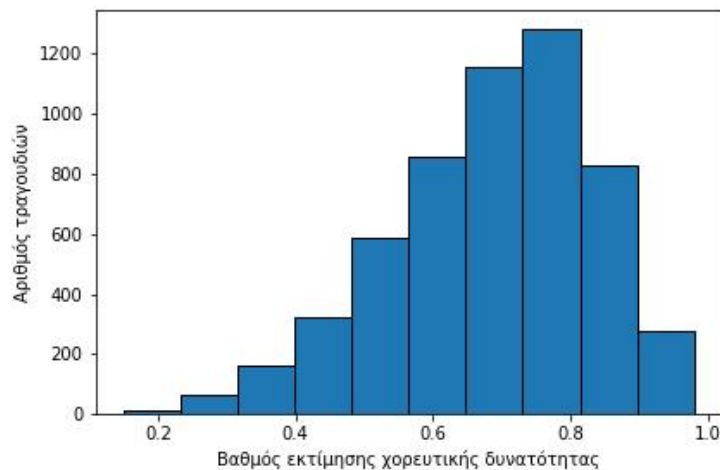


Σχήμα 5.6: Διαχωρισμός των μουσικών κομματιών ως προς την ταχύτητα

Χορευτική Δυνατότητα (Danceability)

Η χορευτική δυνατότητα περιγράφει το πόσο κατάλληλο είναι ένα κομμάτι για χορό και συνυπολογίζει τον ρυθμό, την σταθερότητα ως προς τις εναλλαγές και τη δύναμη των χτύπων. Μία τιμή 0 υποδεικνύει ότι το μουσικό κομμάτι δεν είναι καθόλου χορευτικό, ενώ μία τιμή 1 σημαίνει ότι είναι κατάλληλο για χορό.

Κλάσεις	(0.149, 0.233]	(0.233, 0.316]	(0.316, 0.399]	(0.399, 0.482]	(0.482, 0.565]	(0.565, 0.648]	(0.648, 0.731]	(0.731, 0.814]	(0.814, 0.897]	(0.897, 0.98]
Αριθμός τραγουδιών	15	64	160	330	591	846	1171	1278	813	278

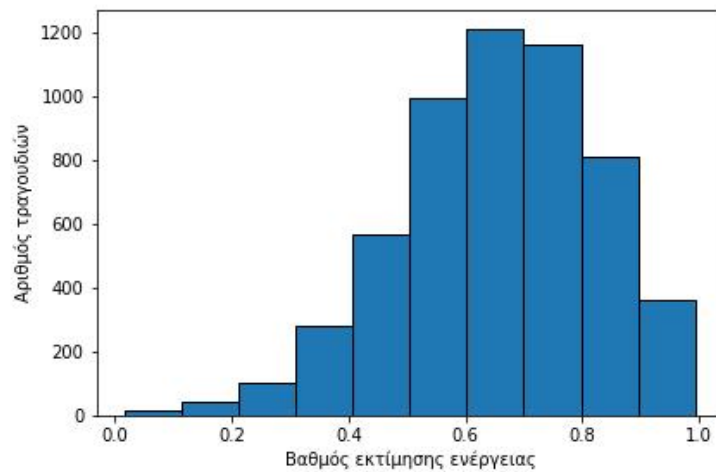


Σχήμα 5.7: Διαχωρισμός των μουσικών κομματιών ως προς τη χορευτική δυνατότητα

Ενέργεια (Energy)

Η ενέργεια είναι μία μέτρηση από το 0 μέχρι το 1 και ποσοτικοποιεί την ενεργητικότητα ενός κομματιού. Δηλαδή, η τιμή 1 αντιπροσωπεύει ένα πολύ έντονο, δυναμικό κομμάτι (για παράδειγμα πολλά είδη Metal) ενώ η τιμή 0 αντιπροσωπεύει ένα πολύ ήρεμο κομμάτι (για παράδειγμα ένα πρελούδιο κλασικής μουσικής).

Κλάσεις	(0.0157, 0.114]	(0.114, 0.212]	(0.212, 0.31]	(0.31, 0.407]	(0.407, 0.505]	(0.505, 0.602]	(0.602, 0.7]	(0.7, 0.798]	(0.798, 0.895]	(0.895, 0.993]
Αριθμός τραγουδιών	13	40	104	280	566	993	1212	1163	811	364

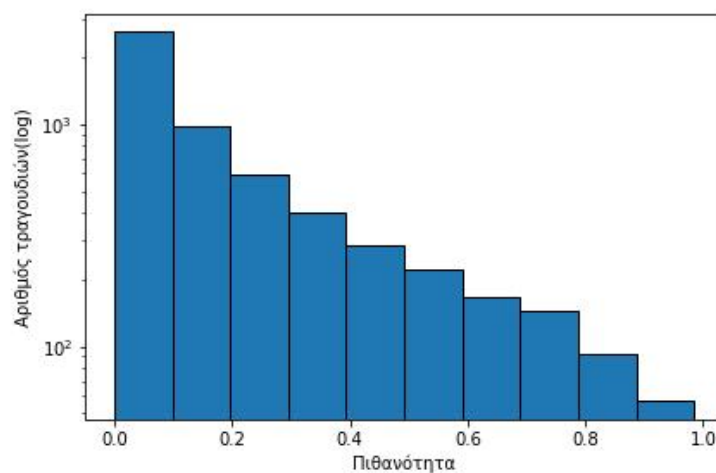


Σχήμα 5.8: Διαχωρισμός των μουσικών κομματιών ως προς την ενέργεια

Ακουστικότητα (Acousticness)

Η ακουστικότητα αποτελεί μία μέτρηση της χρήσης μόνο ακουστικών οργάνων. Το εύρος τιμών είναι από 0 μέχρι 1.

Κλάσεις	(0, 0.09]	(0.09, 0.19]	(0.19, 0.29]	(0.29, 0.39]	(0.39, 0.49]	(0.49, 0.59]	(0.59, 0.69]	(0.69, 0.78]	(0.78, 0.88]	(0.88, 0.98]
Αριθμός τραγουδιών	2603	973	597	403	287	220	170	144	92	57



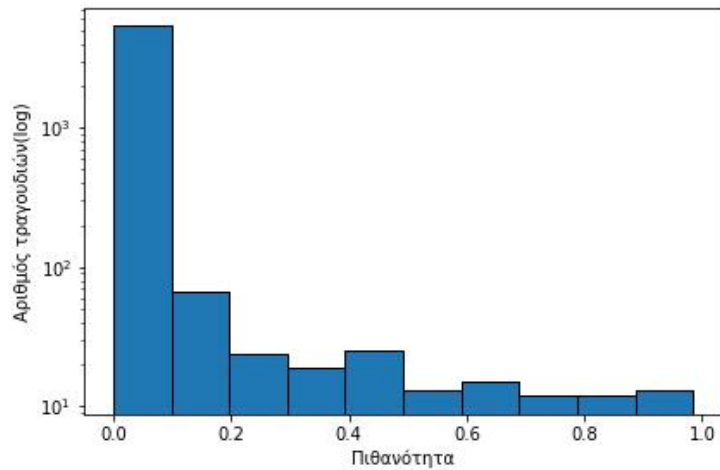
Σχήμα 5.9: Διαχωρισμός των μουσικών κομματιών ως προς την ακουστικότητα

Ορχηστρικότητα (Instrumentalness)

Η ορχηστρικότητα αποτελεί μία πρόβλεψη του κατά πόσο ένα μουσικό κομμάτι δεν περιέχει φωνητικά. Για παράδειγμα, ένα Rap κομμάτι, θα έχει τιμή κοντά στο 0, ενώ αντιθέτως ένα κομμάτι ηλεκτρονικής ή κλασσικής

μουσικής θα έχει τιμή κοντά στο 1.

Κλάσεις	[0, 0.098]	(0.098, 0.197]	(0.197, 0.296]	(0.296, 0.394]	(0.394, 0.493]	(0.493, 0.591]	(0.591, 0.69]	(0.69, 0.788]	(0.788, 0.887]	(0.887, 0.985]
Αριθμός τραγουδιών	5346	67	24	19	25	13	15	12	12	13

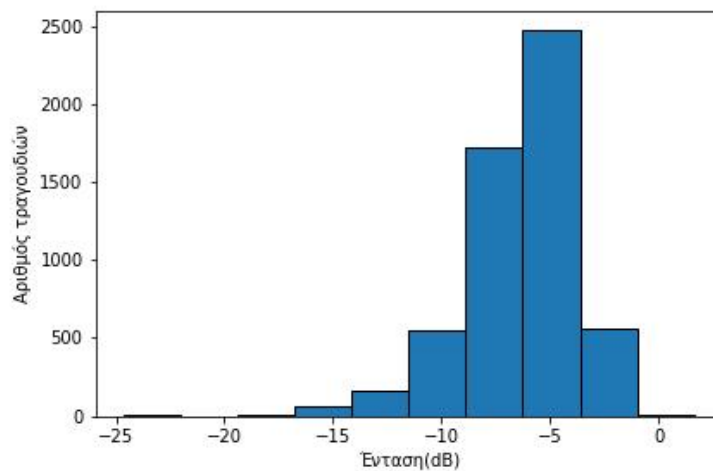


Σχήμα 5.10: Διαχωρισμός των μουσικών κομματιών ως προς την ορχηστρικότητα

Ένταση (Loudness)

Το χαρακτηριστικό της έντασης αποτελεί τον μέσο όρο έντασης του ήχου στο σύνολο του μουσικό κομμάτι και μετριέται σε decibel (dB), με τις μετρήσεις αυτές να βρίσκονται μεταξύ -60dB έως 0.

Κλάσεις	(-24.66, -22.01]	(-22.01, -19.38]	(-19.38, -16.75]	(-16.75, -14.13]	(-14.13, -11.5]	(-11.5, -8.87]	(-8.87, -6.24]	(-6.24, -3.62]	(-3.62, -0.99]	(-0.993, 1.63]
Αριθμός τραγουδιών	4	2	12	58	165	548	1721	2473	560	3

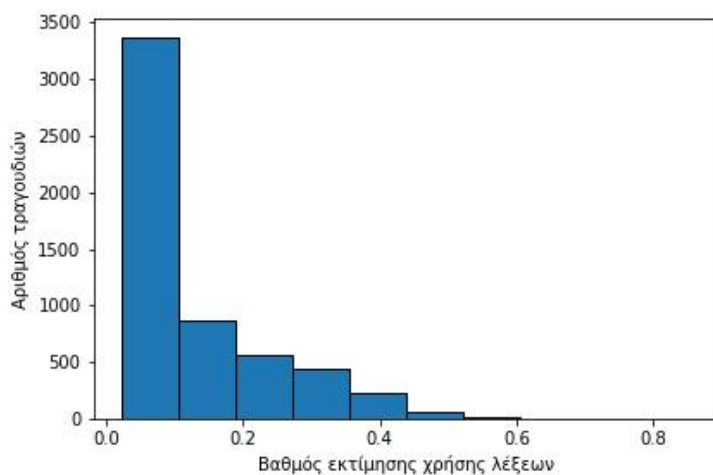


Σχήμα 5.11: Διαχωρισμός των μουσικών κομματιών ως προς την ένταση

Χρήση λέξεων (Speechiness)

Το χαρακτηριστικό ανιχνεύει την ομηλία. Καταχωρήσεις όπως συνεντεύξεις ή ποιήματα έχουν τιμή κοντά στο 1, ενώ τα ορχηστρικά κομμάτια λαμβάνουν τιμές κοντά στο 0. Τα Rap κομμάτια έχουν υψηλές τιμές.

Κλάσεις	(0.022, 0.106]	(0.106, 0.189]	(0.189, 0.273]	(0.273, 0.356]	(0.356, 0.439]	(0.439, 0.523]	(0.523, 0.606]	(0.606, 0.689]	(0.689, 0.773]	(0.773, 0.856]
Αριθμός τραγουδιών	3358	862	561	439	227	71	18	5	4	1

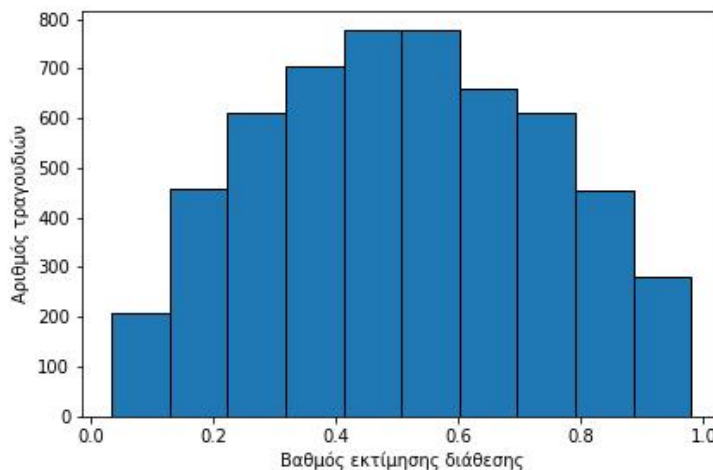


Σχήμα 5.12: Διαχωρισμός των μουσικών κομματιών ως προς τη χρήση λέξεων

Διάθεση (Valence)

Το χαρακτηριστικό αυτό περιγράφει τη διάθεση που προκαλεί ένα μουσικό κομμάτι. Η τιμή 1 συνδέεται με τα θετικά συναισθήματα, όπως η χαρά, ενώ τιμές κοντά στο 0 συνδέονται με αρνητικά συναισθήματα όπως για παράδειγμα η λύπη ή ο θυμός.

Κλάσεις	(0.032, 0.128]	(0.128, 0.223]	(0.223, 0.318]	(0.318, 0.412]	(0.412, 0.507]	(0.507, 0.602]	(0.602, 0.696]	(0.696, 0.791]	(0.791, 0.885]	(0.885, 0.98]
Αριθμός τραγουδιών	209	457	612	704	777	778	659	612	456	282



5.5 Συλλογή ετικετών (tags) μουσικών κομματιών

Για την ανάλυση των αποτελεσμάτων και την αναζήτηση ποιοτικών χαρακτηριστικών στις προτάσεις που παράγονται από τα μοντέλα, ήταν απαραίτητη η εύρεση ετικετών που περιγράφουν τα μουσικά είδη που περιγράφουν ένα κομμάτι. Το Spotify δεν παρέχει τέτοιου είδους πληροφορία για τα μουσικά κομμάτια.

Για την επίλυση του παραπάνω προβλήματος αξιοποιήθηκε η διεπαφή (API) του LastFM, μίας δημοφιλούς πλατφόρμας ροής (streaming) μουσικής, ενώ ελέγχθηκε και η δυνατότητα συλλογής της απαραίτητης πληροφορίας από το MusicBrainz, που αποτελεί μία ανοιχτή εγκυκλοπαίδεια καταγραφής και ταξινόμησης μουσικής.

Ένα βασικό ζήτημα που προέκυψε είναι ότι δεν υπήρχε ενιαία αναφορά των IDs καθώς, σε αντίθεση με το Spotify, το LastFm χρησιμοποιεί τους κωδικούς αναφοράς MBID (MusicBrainz Identifier).

Για να ξεπεραστεί η ασυμβατότητα μεταξύ των τρόπων αναφοράς στα μουσικά κομμάτια, κατασκευάστηκε ένα ερώτημα προς τη διεπαφή του LastFM για τις καλύτερες ετικέτες κάθε κομματιού, με κριτήριο την αναζήτηση ως προς το όνομα του κομματιού και των καλλιτεχνών. Από αυτή τη διαδικασία, επιστράφηκαν ετικέτες σε 4696 μουσικά κομμάτια του μεγάλου σύνολου δεδομένων.

Η συλλογή των ετικετών από το Musicbrainz πραγματοποιήθηκε με τον ακόλουθο τρόπο: Για κάθε μουσικό κομμάτι, έγινε κλήση στην διεπαφή Musicbrainz, βάσει το όνομα του κομματιού και τον καλλιτεχνών που το ερμηνεύουν. Η απάντηση ήταν ένα σύνολο καταχωρήσεων, καθώς ένα κομμάτι καταγράφεται στην εγκυκλοπαίδεια αυτή αρκετές φορές (ζωντανές ηχογραφήσεις, επανεκδόσεις, Remasters). Κάθε καταχώρηση περιείχε ένα ποσοστό ομοιότητας με τα κριτήρια της αναζήτησης. Για κάθε καταχώρηση με μεγάλο ποσοστό ομοιότητας, αναζητήθηκαν ετικέτες που αφορούσαν τα μουσικά κομμάτια.

Ενώ η αναζήτηση ετικετών από το Musicbrainz ήταν λειτουργική, κρίθηκε αρκετά αναποτελεσματική, καθώς, σε αντίθεση με τη συλλογή ετικετών από το LastFM, βρέθηκαν καταχωρήσεις για μόνο 300 μουσικά κομμάτια.

Στη συνέχεια, τα μουσικά κομμάτια χωρίστηκαν σε 8 μεγάλες κατηγορίες. Η συχνότητα εμφάνισης της κάθε κατηγορίας φαίνεται παρακάτω:

hip hop	1644
pop	860
rock	853
r&b	316
ηλεκτρονική	164
country	35
jazz	19

Πίνακας 5.5: Πίνακας του διαχωρισμού των μουσικών κομματιών ανά κατηγορία μουσικού είδους

Η χρήση αυτών των ετικετών δεν προορίζεται για την εκπαίδευση των μοντέλων αλλά μόνο για την εκτίμηση της ποιότητας των αποτελεσμάτων. Σημειώνεται επιπλέον ότι παρά τις προσπάθειες εξασφάλισης της ποιότητας της πληροφορίας που παρέχεται μέσα από τις ετικέτες, αυτή είναι επιρρεπής τόσο σε λανθασμένες καταχωρήσεις όσο και στην υποκειμενική κρίση του ανθρώπινου παράγοντα.

Κεφάλαιο 6

Υλοποίηση Μοντέλων

6.1 Παραγοντοποίηση Πινάκων με Εναλλασσόμενα Τετράγωνα

6.1.1 Πίνακες χαρακτηριστικών/παραγόντων

Για την κατασκευή των πινάκων χρηστών-παραγόντων (λίστών αναπαραγωγής) και αντικειμένων-παραγόντων (μουσικών κομματιών) αξιοποιούνται οι μεταβλητές του Tensorflow και αρχικοποιούνται με τυχαίο τρόπο από το διάστημα -0.05 έως 0.05 . Οι διαστάσεις των πινάκων αυτών για το μικρό σύνολο δεδομένων είναι $(10251, n_f)$ και $(1890, n_f)$, με 10.251 να είναι οι λίστες αναπαραγωγής του συνόλου εκπαίδευσης, 1890 τα μουσικά κομμάτια ενώ με n_f συμβολίζεται ο αριθμός των παραγόντων/χαρακτηριστικών. Αντίστοιχα για το μεγάλο σύνολο δεδομένων, οι αντίστοιχες διαστάσεις των δύο πινάκων είναι $(63460, n_f)$ και $(5546, n_f)$.

6.1.2 Η διαδικασία εκπαίδευσης

Για την εκπαίδευση του μοντέλου, το σύνολο δεδομένων εκπαίδευσης διαχωρίστηκε σε παρτίδες (batches) με τον εξής τρόπο: Αρχικά, για την εκπαίδευση του πίνακα χρηστών-παραγόντων, ο διαχωρισμός των παρτίδων έγινε βάσει των λίστών αναπαραγωγής. Για την εκπαίδευση του πίνακα αντικειμένων-παραγόντων, ο διαχωρισμός πραγματοποιήθηκε βάσει των μουσικών κομματιών. Με αυτό τον τρόπο καθίσταται δυνατό να αντιμετωπιστούν τα προβλήματα μνήμης που ενδέχεται να δημιουργηθούν.

Η διαδικασία εκπαίδευσης πραγματοποιείται επαναληπτικά για έναν αριθμό εποχών. Για κάθε εποχή, περιγράφονται τα παρακάτω βήματα:

Υπολογισμός του πίνακα χρηστών-παραγόντων

Για κάθε παρτίδα του συνόλου δεδομένων που χωρίστηκε με βάση τις λίστες αναπαραγωγής, υπολογίζονται οι χρήστες του πίνακα χρηστών-παραγόντων μέσα από την εξίσωση:

$$x_u = [Y^T(Y \circ C'_u) + \lambda I]^{-1} Y^T p(u) \circ c_u \quad (6.1)$$

Με Y συμβολίζεται ο πίνακας αντικειμένων-παραγόντων με r_u το διάνυσμα των βαθμολογιών που αντιστοιχεί στη γραμμή μίας λίστας αναπαραγωγής του συνόλου εκπαίδευσης και με c_u συμβολίζεται το διάνυσμα εμπιστοσύνης του συγκεκριμένου χρήστη, με C'_u που περιέχει το διάνυσμα c_u , n_f φορές (Broadcasting).

Κάθε διάνυσμα x_u μπορεί να υπολογισθεί ανεξάρτητα, κάτι που επιτρέπει τη δημιουργία γράφου στο Tensorflow για όλη την παρτίδα και, εν τέλει, της εκτέλεσης της διαδικασίας στη GPU.

Υπολογισμός του πίνακα αντικειμένων-παραγόντων

Με ακριβώς ανάλογο τρόπο, υπολογίζονται τα στοιχεία του πίνακα αντικειμένων-παραγόντων. Για κάθε παρτίδα του συνόλου δεδομένων που χωρίστηκε με βάση τα μουσικά κομμάτια, υπολογίζονται τα αντικείμενα του πίνακα αντικειμένων-παραγόντων μέσα από την εξίσωση:

$$y_i = [X^T(X \circ C'_i) + \lambda I]^{-1} X^T p(i) \circ c_i \quad (6.2)$$

Με X συμβολίζεται ο πίνακας αντικειμένων-παραγόντων με r_i το διάνυσμα των βαθμολογιών που αντιστοιχεί στη στήλη ενός μουσικού κομματιού στο σύνολο εκπαίδευσης και με c_u συμβολίζεται το διάνυσμα εμπιστοσύνης του συγκεκριμένου χρήστη, με C'_i που περιέχει το διάνυσμα c_i , n_f φορές (Broadcasting).

Κάθε διάνυσμα y_i μπορεί να υπολογισθεί ανεξάρτητα, κάτι που επιτρέπει τη δημιουργία γράφου στο Tensorflow για όλη την παρτίδα και, εν τέλει, της εκτέλεσης της διαδικασίας στη GPU.

6.1.3 Υπολογισμός Προτάσεων

Ο υπολογισμός της πρότασης (recommendation) ενός μουσικού κομματιού για μία λίστα αναπαραγωγής, αντιστοιχεί στο εσωτερικό γινόμενο ενός ανάστροφου διανύσματος-γραμμής του πίνακα αντικειμένων-παραγόντων με ένα διάνυσμα-γραμμή του πίνακα χρηστών-παραγόντων.

Ο υπολογισμός των προτάσεων βάσει μίας λίστας αναπαραγωγής για όλα τα μουσικά κομμάτια αποτελεί τον πολλαπλασιασμό του πίνακα αντικειμένων-παραγόντων με το διάνυσμα της λίστας αναπαραγωγής.

6.1.4 Υπολογισμός των Μετρικών στα Σύνολα Δεδομένων Επικύρωσης και Ελέγχου

Για κάθε λίστα αναπαραγωγής από τα σύνολα επικύρωσης και ελέγχου έχει αφαιρεθεί, με τυχαίο τρόπο, το 20% των μουσικών κομματιών. Το πρώτο κομμάτι των δύο συνόλων χρησιμοποιείται για τον προσδιορισμό δύο νέων πινάκων χρηστών-παραγόντων, έναν που να αντιστοιχεί στις λίστες αναπαραγωγής του πρώτου συνόλου και έναν που να αντιστοιχεί στο δεύτερο, με τη μέθοδο που παρουσιάστηκε παραπάνω και υπολογίζονται οι προτάσεις που αντιστοιχούν σε κάθε χρήστη. Για το σύνολο εκπαίδευσης δεν κρίθηκε αναγκαία η εφαρμογή της παραπάνω διαδικασίας, καθώς ο λόγος για τον υπολογισμό των μετρικών σε αυτό το σύνολο είναι ο οπτικός προσδιορισμός του σημείου υπερεκπαίδευσης.

Για τον υπολογισμό των μετρικών Recall@50 και NDCG@50 για τα σύνολα δεδομένων επικύρωσης και ελέγχου, χρησιμοποιούνται:

- Οι προτάσεις του πρώτου κομματιού των συνόλων, όπως υπολογίστηκαν από το μοντέλο
- Τα μουσικά κομμάτια που αφαιρέθηκαν, ως τα προσδοκώμενα αποτελέσματα.

Ενώ για τον υπολογισμό των μετρικών Recall@50 και NDCG@50 στο σύνολο δεδομένων εκπαίδευσης:

- Οι προτάσεις προκύπτουν από την έξοδο του νευρωνικού δικτύου με είσοδο το σύνολο εκπαίδευσης.
- Το σύνολο εκπαίδευσης, ως τα προσδοκώμενα αποτελέσματα.

Με αυτό τον τρόπο βαθμολογείται η δυνατότητα του μοντέλου να δίνει προτάσεις κομματιών που, ενώ ανήκουν στην πραγματικότητα στη λίστα αναπαραγωγής, έχουν αποκρυφθεί από τη λίστα πριν παρουσιαστούν σε αυτό.

6.1.5 Το κριτήριο επιλογής βέλτιστου μοντέλου

Για την παύση της εκπαίδευσης σε μία συγκεκριμένη τοπολογία αξιοποιείται το κριτήριο Early Stopping, όταν δηλαδή οι βαθμολογίες της Recall@50 στο σύνολο επικύρωσης αρχίζουν να μειώνονται και αποθηκεύονται οι παράμετροι του μοντέλου.

Για την επιλογή της βέλτιστης τοπολογίας του μοντέλου, χρησιμοποιήθηκε η μεγαλύτερη τιμή της Recall@50 για το σύνολο δεδομένων ελέγχου. Για τις προτάσεις μουσικών κομματιών το μοντέλο περιγράφεται πλήρως από τον πίνακα Αντικειμένων-Παραγόντων.

Η μέθοδος αυτός συγκλίνει αρκετά γρήγορα, με την πάροδο μόλις πέντε εποχών. Με την περαιτέρω εκπαίδευση ενώ υπήρχε μία μικρή βελτίωση στα αποτελέσματα των μετρικών, αυτή δεν ήταν ιδιαίτερα σημαντική.

6.1.6 Βελτιστοποίηση υπερπαραμέτρων με τη χρήση Grid Search

Για την εύρεση των βέλτιστων τιμών των εν λόγω υπερπαραμέτρων πραγματοποιήθηκε Grid Search σε διαφορετικούς συνδυασμούς. Ως υπερπαραμέτροι του μοντέλου θεωρούνται οι μεταβλητές:

- α , που χρησιμοποιείται στην κατασκευή του πίνακα εμπιστοσύνης C , μέσα από την εξίσωση $c_{ui} = 1 + \alpha r_{ui}$, για τις τιμές 0, 5 και 20.
- λ , που αποτελεί τον L2 παράγοντα κανονικοποίησης. Ερευνήθηκαν οι τιμές 0.5, 1 και 2.
- f , που αντιστοιχεί στον αριθμό των παραγόντων των πινάκων χρηστών και αντικειμένων. Τα μεγέθη που εξετάστηκαν ήταν 20, 50 και 100

Για το μικρό σύνολο δεδομένων, ο καλύτερος συνδυασμός υπερπαραμέτρων προσδιορίζεται από τις τιμές: $\alpha=5$, $\lambda=2$ και $features=20$. Για το μεγάλο σύνολο δεδομένων ο συνδυασμός με τα βέλτιστα αποτελέσματα δίνεται από τον εξής συνδυασμό: $\alpha=5$, $\lambda=1$ και $features=100$.

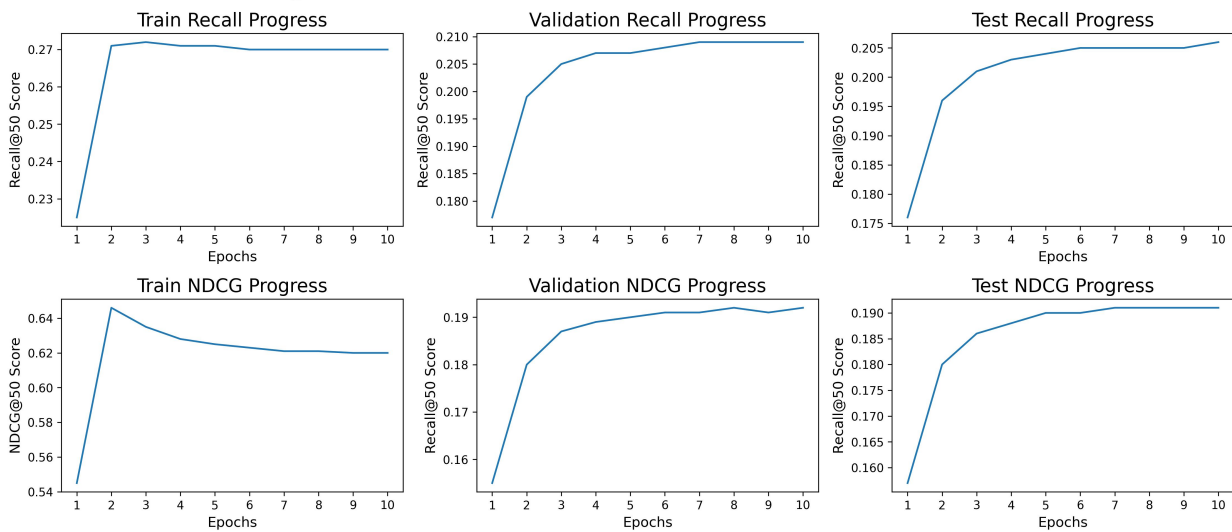
Τα αποτελέσματα της αναζήτησης παρατίθενται στους παρακάτω πίνακες, για τα μοντέλα που εκπαιδεύθηκαν στο μεγάλο και στο μικρό σύνολο δεδομένων καθώς και τα γραφήματα της εξέλιξης της εκπαίδευσης του βέλτιστου μοντέλου σε κάθε σύνολο. Στα γραφήματα που παρουσιάζονται, επιτράπηκε στο μοντέλο να ολοκληρώσει όλες τις εποχές, μόνο για την καταγραφή της εξέλιξης της εκπαίδευσης. Το μοντέλο που αποθηκεύεται είναι αυτό με την καλύτερη απόδοση στο σύνολο επικύρωσης.

Οι πολύ μεγάλες τιμές της NDCG@50 για το σύνολο εκπαίδευσης ερμηνεύονται από τη χρήση ολόκληρου του συνόλου κατά τη διάρκεια υπολογισμού της μετρικής.

λ	features	α	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
0.5	20	5	0.175	0.156	0.177	0.156	0.189	0.406
0.5	20	0	0.152	0.125	0.155	0.126	0.166	0.344
0.5	20	20	0.185	0.173	0.187	0.173	0.198	0.438
0.5	50	0	0.180	0.157	0.183	0.156	0.212	0.460
0.5	50	5	0.200	0.186	0.202	0.184	0.232	0.520
0.5	50	20	0.204	0.195	0.205	0.192	0.236	0.535
0.5	100	0	0.187	0.165	0.192	0.166	0.249	0.560
0.5	100	5	0.205	0.192	0.209	0.192	0.270	0.620
0.5	100	20	0.205	0.197	0.207	0.196	0.270	0.627
1.0	20	0	0.152	0.125	0.155	0.126	0.167	0.344
1.0	20	5	0.175	0.156	0.177	0.156	0.189	0.406
1.0	20	20	0.185	0.173	0.188	0.173	0.198	0.438
1.0	50	0	0.180	0.157	0.183	0.156	0.212	0.461
1.0	50	5	0.199	0.185	0.203	0.185	0.233	0.521
1.0	50	20	0.203	0.194	0.205	0.192	0.236	0.534
1.0	100	0	0.188	0.165	0.192	0.166	0.248	0.560
1.0	100	5	0.206	0.191	0.209	0.192	0.270	0.620
1.0	100	20	0.205	0.196	0.207	0.195	0.271	0.628
2.0	20	0	0.151	0.124	0.154	0.125	0.166	0.343
2.0	20	5	0.175	0.156	0.177	0.157	0.189	0.406
2.0	20	20	0.185	0.174	0.188	0.173	0.199	0.440
2.0	50	0	0.180	0.157	0.182	0.156	0.212	0.459
2.0	50	5	0.200	0.186	0.203	0.185	0.233	0.522
2.0	50	20	0.203	0.195	0.205	0.192	0.236	0.536
2.0	100	0	0.187	0.165	0.192	0.167	0.248	0.559
2.0	100	5	0.205	0.191	0.209	0.192	0.270	0.620
2.0	100	20	0.206	0.190	0.206	0.194	0.271	0.629

Πίνακας 6.1: Τα αποτελέσματα της μεθόδου παραγοντοποίησης πινάκων στην αναζήτηση Grid για το μεγάλο σύνολο δεδομένων

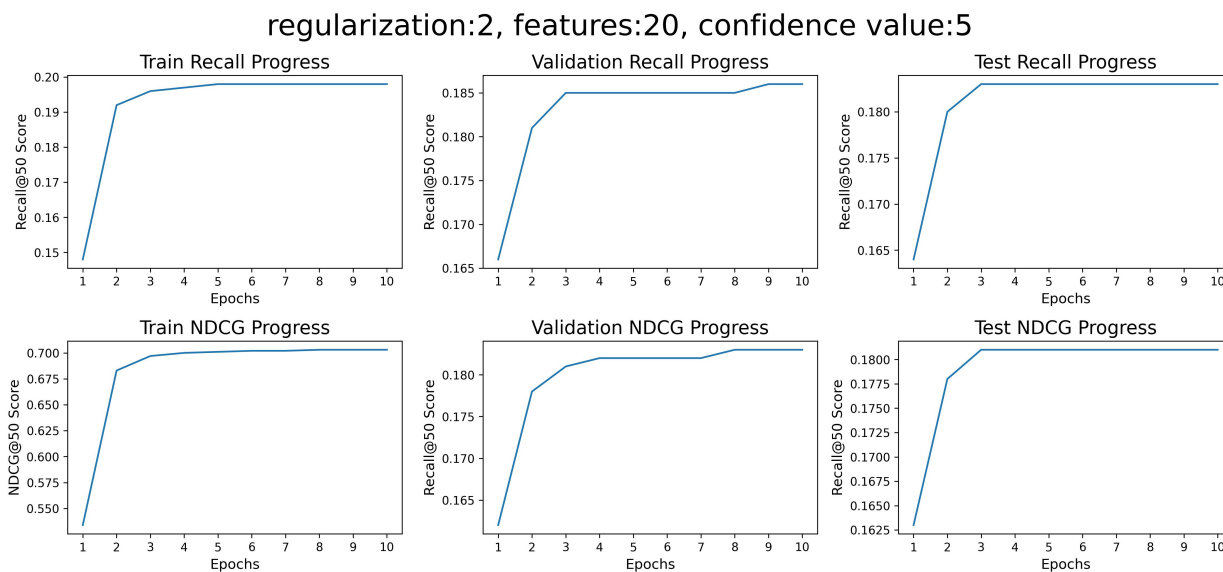
regularization:1, features:100, confidence value:5



Σχήμα 6.1: Διάγραμμα εκπαίδευσης του βέλτιστου μοντέλου παραγοντοποίησης πινάκων για το μεγάλο σύνολο δεδομένων

λ	features	α	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
0.5	20	0	0.176	0.174	0.180	0.175	0.197	0.705
0.5	20	5	0.183	0.179	0.186	0.182	0.198	0.703
0.5	20	20	0.178	0.176	0.180	0.176	0.185	0.646
0.5	50	0	0.176	0.166	0.177	0.167	0.224	0.794
0.5	50	5	0.183	0.180	0.183	0.179	0.218	0.764
0.5	50	20	0.176	0.175	0.177	0.175	0.193	0.657
0.5	100	0	0.150	0.133	0.150	0.133	0.247	0.870
0.5	100	5	0.165	0.162	0.165	0.161	0.235	0.815
0.5	100	20	0.163	0.167	0.162	0.166	0.199	0.646
1.0	20	0	0.176	0.174	0.180	0.175	0.197	0.705
1.0	20	5	0.182	0.181	0.186	0.183	0.198	0.703
1.0	20	20	0.177	0.175	0.179	0.176	0.185	0.647
1.0	50	0	0.176	0.167	0.178	0.168	0.224	0.794
1.0	50	5	0.182	0.179	0.183	0.179	0.218	0.765
1.0	50	20	0.175	0.175	0.175	0.173	0.194	0.658
1.0	100	0	0.149	0.132	0.151	0.133	0.247	0.870
1.0	100	5	0.164	0.160	0.164	0.160	0.236	0.815
1.0	100	20	0.163	0.168	0.163	0.168	0.199	0.650
2.0	20	0	0.177	0.174	0.180	0.175	0.197	0.706
2.0	20	5	0.183	0.181	0.186	0.183	0.198	0.703
2.0	20	20	0.177	0.174	0.180	0.175	0.185	0.648
2.0	50	0	0.176	0.166	0.178	0.168	0.224	0.794
2.0	50	5	0.183	0.180	0.183	0.179	0.218	0.765
2.0	50	20	0.176	0.175	0.176	0.174	0.194	0.659
2.0	100	0	0.151	0.134	0.151	0.134	0.247	0.869
2.0	100	5	0.164	0.161	0.166	0.161	0.236	0.816
2.0	100	20	0.162	0.167	0.162	0.166	0.200	0.652

Πίνακας 6.2: Τα αποτελέσματα της μεθόδου παραγοντοποίησης πινάκων στην αναζήτηση Grid για το μικρό σύνολο δεδομένων



Σχήμα 6.2: Διάγραμμα εκπαίδευσης του βέλτιστου μοντέλου παραγοντοποίησης πινάκων για το μικρό σύνολο δεδομένων

6.2 Αυτοκωδικοποιητές

6.2.1 Απλοί Αυτοκωδικοποιητές και Αυτοκωδικοποιητές Αποθορυβοποίησης

Λόγω της κοινής αρχιτεκτονικής μεταξύ των Αυτοκωδικοποιητών και των Αυτοκωδικοποιητών Αποθορυβοποίησης, ο τρόπος υλοποίησης τους θα περιγραφεί από κοινού.

Κωδικοποιητής

Ο κωδικοποιητής αποτελείται από ένα Multilayer Perceptron. Το επίπεδο εισόδου περιέχει νευρώνες ίσους με τον αριθμό των μουσικών κομματιών. Το μοντέλο δέχεται ως είσοδο ένα υποσύνολο διανυσμάτων από το σύνολο δεδομένων. Κάθε διάνυσμα εισόδου αντιστοιχεί σε μία λίστα αναπαραγωγής. Η διάσταση της εξόδου του κωδικοποιητή καθορίζεται από τις διαστάσεις της κρυφής μεταβλητής που αποτελεί υπερπαραμέτρο του μοντέλου και υπολογίζεται πειραματικά.

Στην περίπτωση του Αυτοκωδικοποιητή Αποθορυβοποίησης, η είσοδος περνάει από ένα στάδιο αλλοίωσης, μέσα από την εφαρμογή dropout 50%.

6.2.2 Αποκωδικοποιητής

Η είσοδος του αποκωδικοποιητή είναι η έξοδος του κωδικοποιητή. Η διάσταση του σταδίου εξόδου είναι ίδια με την είσοδο του κωδικοποιητή, δηλαδή ο αριθμός των μουσικών κομματιών. Ως συνάρτηση ενεργοποίησης του σταδίου εξόδου επιλέγεται η Softmax, με όλα τα στοιχεία της εξόδου να ανήκουν στο διάστημα (0, 1) και το άθροισμα τους ισούται με 1.

Συνάρτηση Στόχου

Το κριτήριο βελτιστοποίησης της εκπαίδευσης είναι η εκτίμηση μέγιστης πιθανοφάνειας της πολυωνυμικής κατανομής: $\log p_{\theta}(\mathbf{x}|\mathbf{z}) = \sum_i x_i \log y_i$, με y συμβολίζεται το διάνυσμα εξόδου.

6.2.3 Διαδικασία Εκπαίδευσης

Κατασκευάστηκε το βήμα εκπαίδευσης από την αρχή. Αναλυτικά: Το σύνολο δεδομένων χωρίστηκε σε παρτίδες (batches). Για κάθε εποχή, κάθε τμήμα εισάγεται στο νευρωνικό δίκτυο, υπολογίζεται η συνάρτηση κόστους και τα αποτελέσματα των μετρικών. Στη συνέχεια, πραγματοποιείται η αναπροσαρμογή των βαρών του δικτύου με τη χρήση του αλγόριθμου βελτιστοποίησης Adam.

Κατά τη διάρκεια της εκπαίδευσης των μοντέλων στο μεγάλο σύνολο δεδομένων, χρειάστηκε η αξιοποίηση της αραιής μορφής αναπαράστασης του συνόλου εκπαίδευσης, λόγω προβλημάτων στη διαχείριση της μνήμης (Out of Memory).

Υπολογισμός Προτάσεων

Οι προτάσεις προκύπτουν από την ταξινόμηση του διανύσματος εξόδου του αποκωδικοποιητή του νευρωνικού δικτύου.

Υπολογισμός των Μετρικών στα Σύνολα Δεδομένων Επικύρωσης και Ελέγχου

Για κάθε λίστα αναπαραγωγής από τα σύνολα επικύρωσης και ελέγχου έχει αφαιρεθεί, με τυχαίο τρόπο, το 20% των μουσικών κομματιών. Το πρώτο κομμάτι των δύο συνόλων χρησιμοποιείται για τον υπολογισμό των προτάσεων που αντιστοιχούν σε κάθε χρήστη. Για το σύνολο εκπαίδευσης δεν κρίθηκε αναγκαία η εφαρμογή της παραπάνω διαδικασίας, καθώς ο λόγος για τον υπολογισμό των μετρικών σε αυτό το σύνολο είναι ο οπτικός προσδιορισμός του σημείου υπερεκπαίδευσης.

Συγκεκριμένα, για τον υπολογισμό των μετρικών Recall@50 και NDCG@50 για τα σύνολα δεδομένων επικύρωσης και ελέγχου, χρησιμοποιούνται:

- Οι προτάσεις του πρώτου κομματιού των συνόλων, όπως υπολογίστηκαν από το μοντέλο
- Τα μουσικά κομμάτια που αφαιρέθηκαν, ως τα προσδοκώμενα αποτελέσματα.

ενώ για τον υπολογισμό των μετρικών Recall@50 και NDCG@50 στο σύνολο δεδομένων εκπαίδευσης:

- Οι προτάσεις προκύπτουν από την έξοδο του νευρωνικού δικτύου με είσοδο το σύνολο εκπαίδευσης.
- Το σύνολο εκπαίδευσης, ως τα προσδοκώμενα αποτελέσματα.

Κριτήριο Επιλογής Βέλτιστου Μοντέλου

Για την παύση της εκπαίδευσης σε μία συγκεκριμένη τοπολογία αξιοποιείται το κριτήριο Early Stopping, όταν δηλαδή οι βαθμολογίες της Recall@50 στο σύνολο επικύρωσης αρχίζουν να μειώνονται και αποθηκεύονται οι παράμετροι του μοντέλου.

Για την επιλογή της βέλτιστης τοπολογίας του μοντέλου, χρησιμοποιήθηκε η μεγαλύτερη τιμή της Recall@50 για το σύνολο δεδομένων ελέγχου.

Σύγκριση Διαφορετικών Τοπολογιών - Βελτιστοποίηση Υπερπαραμέτρων

Για την εύρεση του βέλτιστου μοντέλου πραγματοποιήθηκε Gridsearch στις εξής παραμέτρους του μοντέλου:

- Για το πρώτο επίπεδο, το επίπεδο δοκιμάστηκε με 600 και 1000 νευρώνες
- Δοκιμάστηκαν δίκτυα χωρίς ενδιάμεσο επίπεδο ή με ένα ενδιάμεσο επίπεδο 500 νευρώνων.
- Το επίπεδο εξόδου υλοποιήθηκε με 50 και 200 νευρώνες.
- Ως προς τη συνάρτηση ενεργοποίησης των επιπέδων δοκιμάστηκαν η tanh και η ReLU.

Απλοί Κωδικοποιητές

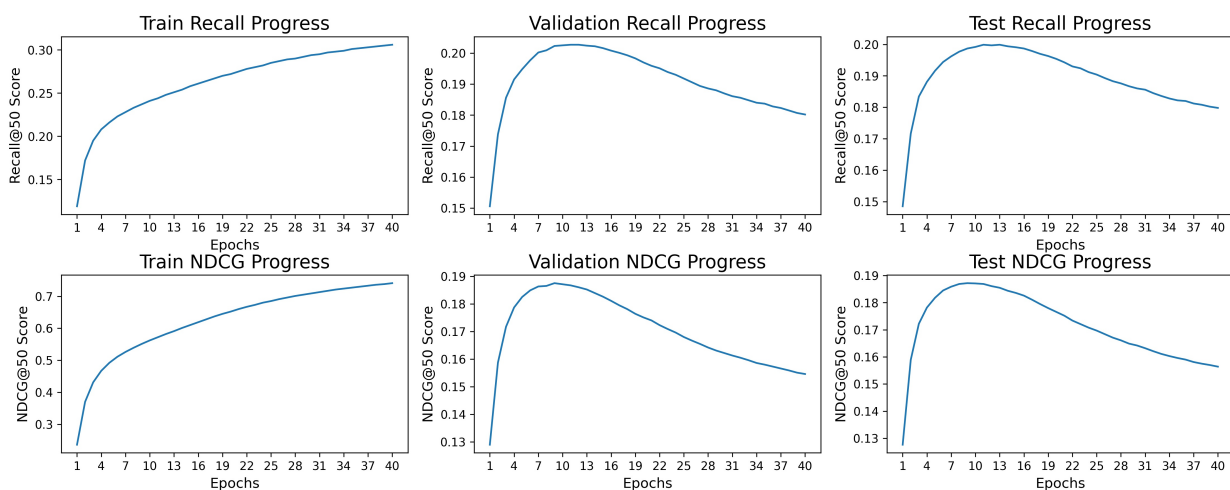
Οι τιμές των υπερπαραμέτρων του μοντέλου με τη καλύτερη απόδοση είναι οι εξής: Για το μοντέλο που εκπαιδεύτηκε στο μεγάλο σύνολο δεδομένων:

- 2 επίπεδα. Το πρώτο επίπεδο έχει μέγεθος 600 νευρώνων και το κρυφό επίπεδο 50 νευρώνων.
- επιλέχθηκε η tanh συνάρτηση ενεργοποίησης.

First Layer	Interminent Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	50	relu	0.1976	0.1832	0.2011	0.182	0.31	0.743
1000	-	50	tanh	0.1991	0.1868	0.2019	0.1866	0.313	0.756
1000	-	100	relu	0.1962	0.1817	0.1994	0.1813	0.346	0.828
1000	-	100	tanh	0.1986	0.1859	0.2019	0.1852	0.351	0.844
1000	[500]	50	relu	0.1958	0.1819	0.2002	0.1818	0.298	0.721
1000	[500]	50	tanh	0.1994	0.1849	0.2018	0.1857	0.296	0.716
1000	[500]	100	relu	0.1953	0.1825	0.1987	0.1815	0.317	0.771
1000	[500]	100	tanh	0.199	0.1854	0.202	0.1851	0.319	0.776
600	-	50	relu	0.1968	0.1825	0.2004	0.1819	0.29	0.691
600	-	50	tanh	0.2003	0.1873	0.2022	0.1864	0.293	0.708
600	-	100	relu	0.1956	0.1807	0.1983	0.1802	0.313	0.752
600	-	100	tanh	0.2	0.1865	0.203	0.1865	0.319	0.774
600	[500]	50	relu	0.1971	0.1825	0.1994	0.1827	0.272	0.651
600	[500]	50	tanh	0.2003	0.1873	0.2027	0.1871	0.27	0.642
600	[500]	100	relu	0.1959	0.1817	0.199	0.1816	0.291	0.707
600	[500]	100	tanh	0.2	0.187	0.2039	0.1863	0.284	0.683

Πίνακας 6.3: Τα αποτελέσματα των Απλών Αυτοκωδικοποιητών στην αναζήτηση Grid για το μεγάλο σύνολο δεδομένων

First Layer:600, Hidden Layer:50,Activation:tanh



Σχήμα 6.3: Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή για το μεγάλο σύνολο δεδομένων

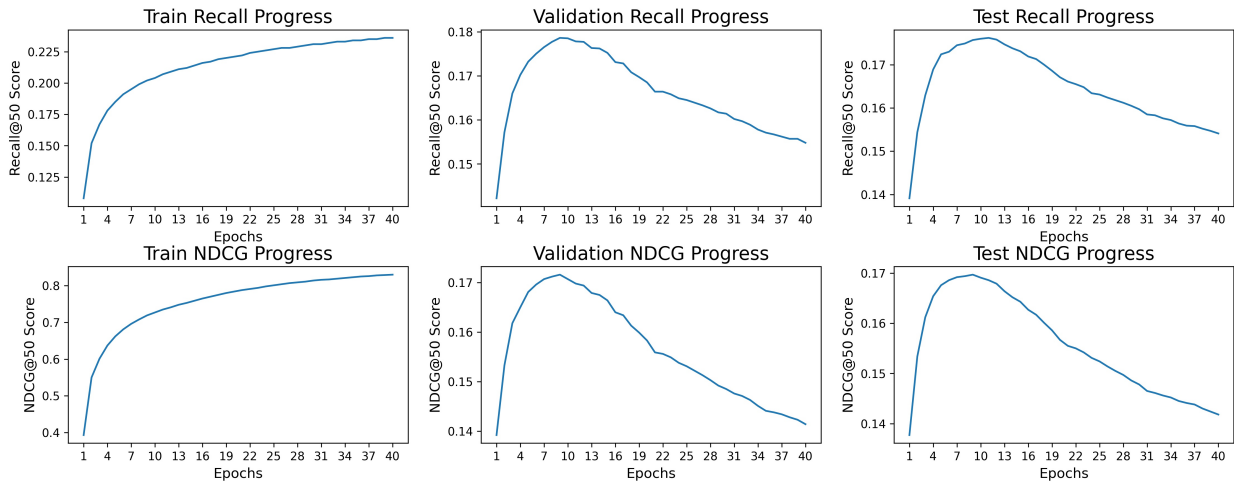
Για το μοντέλο που εκπαιδεύτηκε στο μικρό σύνολο δεδομένων:

- 3 επίπεδα. Το πρώτο επίπεδο έχει μέγεθος 600 νευρώνων, το ενδιάμεσο 500 νευρώνων και το κρυφό επίπεδο 50 νευρώνων.
- επιλέχθηκε η tanh συνάρτηση ενεργοποίησης.

First Layer	Interminent Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	100	tanh	0.1724	0.1666	0.1741	0.1654	0.263	0.913
1000	-	100	relu	0.1702	0.164	0.174	0.1666	0.258	0.898
1000	-	50	tanh	0.1746	0.1686	0.1769	0.1699	0.241	0.845
1000	-	50	relu	0.1738	0.1677	0.1752	0.1687	0.239	0.838
1000	500	100	tanh	0.1736	0.1673	0.1751	0.1679	0.26	0.904
1000	500	100	relu	0.1708	0.1645	0.1732	0.166	0.248	0.868
1000	500	50	tanh	0.1759	0.1696	0.1776	0.1702	0.241	0.846
1000	500	50	relu	0.1734	0.1672	0.1749	0.1693	0.236	0.828
600	-	100	tanh	0.1734	0.1681	0.1749	0.1672	0.256	0.892
600	-	100	relu	0.1689	0.1633	0.1717	0.1639	0.25	0.874
600	-	50	tanh	0.1755	0.1701	0.1778	0.1708	0.237	0.834
600	-	50	relu	0.1722	0.1667	0.1741	0.1668	0.233	0.821
600	500	100	tanh	0.1743	0.1698	0.1761	0.1701	0.251	0.877
600	500	100	relu	0.1703	0.1647	0.1732	0.1658	0.238	0.837
600	500	50	tanh	0.1762	0.1697	0.1786	0.1716	0.236	0.83
600	500	50	relu	0.1735	0.1677	0.176	0.1688	0.228	0.804

Πίνακας 6.4: Τα αποτελέσματα των Απλών Αυτοκωδικοποιητών στην αναζήτηση Grid για το μικρό σύνολο δεδομένων

First Layer:600, Interminent Layers:[500],Hidden Layer:50,Activation:tanh



Σχήμα 6.4: Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή για το μικρό σύνολο δεδομένων

Κωδικοποιητές Αποθρομβοποίησης

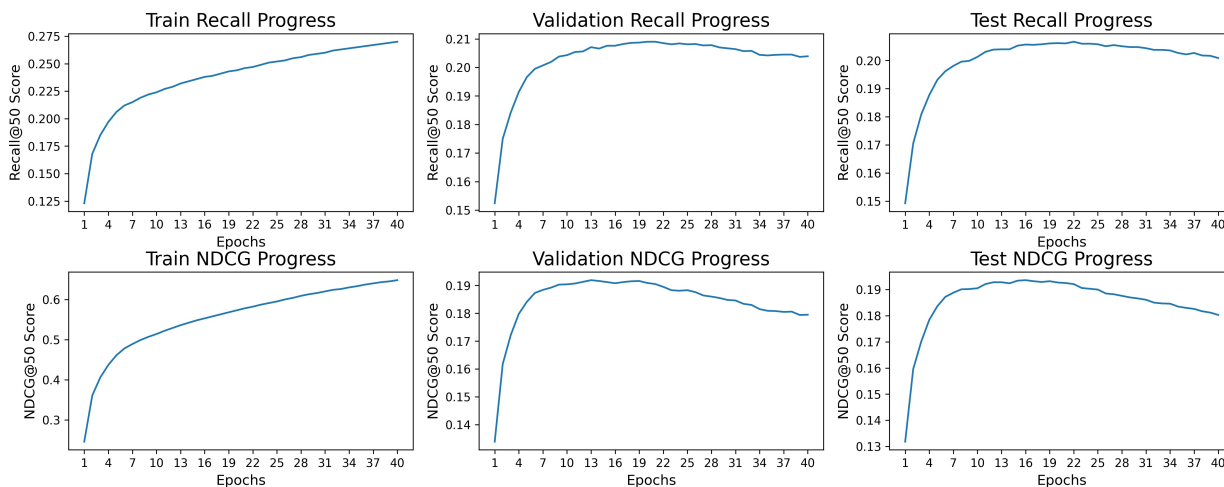
Οι τιμές των υπερπαραμέτρων του μοντέλου με τη καλύτερη απόδοση είναι οι εξής: Για το μοντέλο που εκπαιδεύτηκε στο μεγάλο σύνολο δεδομένων:

- 3 επίπεδα. Το πρώτο επίπεδο έχει μέγεθος 1000 νευρώνων, το ενδιάμεσο 500 νευρώνων και το κρυφό επίπεδο 100 νευρώνων.
- επιλέχθηκε η tanh συνάρτηση ενεργοποίησης.

First Layer	Interminent Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	50	relu	0.2041	0.191	0.2056	0.1891	0.263	0.618
1000	-	50	tanh	0.2038	0.191	0.2063	0.191	0.262	0.623
1000	-	100	relu	0.2005	0.1881	0.2037	0.1869	0.278	0.662
1000	-	100	tanh	0.2033	0.1897	0.2058	0.1896	0.276	0.665
1000	500	50	relu	0.2028	0.191	0.2067	0.1913	0.255	0.598
1000	500	50	tanh	0.2071	0.1943	0.2083	0.1926	0.25	0.587
1000	500	100	relu	0.203	0.1902	0.2052	0.1895	0.264	0.627
1000	500	100	tanh	0.2072	0.1942	0.2089	0.1924	0.258	0.615
600	-	50	relu	0.2036	0.1909	0.2056	0.1895	0.25	0.582
600	-	50	tanh	0.2056	0.1933	0.2072	0.1922	0.248	0.581
600	-	100	relu	0.2019	0.1887	0.2045	0.1876	0.26	0.613
600	-	100	tanh	0.2047	0.1917	0.2065	0.1903	0.257	0.611
600	500	50	relu	0.2041	0.1925	0.207	0.1919	0.241	0.561
600	500	50	tanh	0.2062	0.1959	0.2083	0.1939	0.235	0.544
600	500	100	relu	0.2035	0.1916	0.2067	0.1908	0.248	0.58
600	500	100	tanh	0.2068	0.1942	0.2099	0.1938	0.242	0.562

Πίνακας 6.5: Τα αποτελέσματα των Αυτοκωδικοποιητών Αποθρομβοποίησης στην αναζήτηση Grid για το μεγάλο σύνολο δεδομένων

First Layer:1000, Interminet Layers:[500],Hidden Layer:100,Activation:tanh



Σχήμα 6.5: Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή Αποθυροβοποίησης για το μεγάλο σύνολο δεδομένων

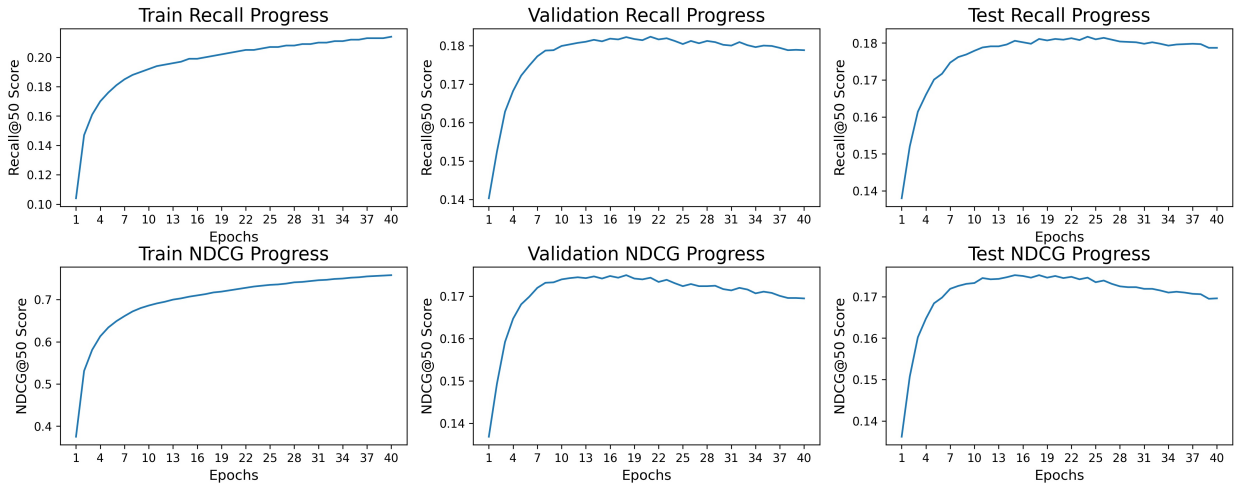
Για το μοντέλο που εκπαιδεύτηκε στο μικρό σύνολο δεδομένων:

- 3 επίπεδα. Το πρώτο επίπεδο έχει μέγεθος 600 νευρώνων, το ενδιάμεσο 500 νευρώνων και το κρυφό επίπεδο 50 νευρώνων.
- επιλέχθηκε η tanh συνάρτηση ενεργοποίησης.

First Layer	Interminet Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	100	tanh	0.1757	0.1688	0.1774	0.1691	0.234	0.826
1000	-	100	relu	0.1724	0.1652	0.1738	0.1653	0.229	0.81
1000	-	50	tanh	0.1787	0.1726	0.1801	0.1731	0.222	0.785
1000	-	50	relu	0.1756	0.1685	0.1777	0.1704	0.219	0.777
1000	500	100	tanh	0.1784	0.1721	0.1803	0.1726	0.229	0.808
1000	500	100	relu	0.1737	0.1672	0.176	0.1681	0.223	0.791
1000	500	50	tanh	0.1788	0.1733	0.1815	0.1748	0.22	0.78
1000	500	50	relu	0.1763	0.1706	0.1782	0.1707	0.216	0.765
600	-	100	tanh	0.1763	0.1696	0.178	0.1715	0.225	0.798
600	-	100	relu	0.1731	0.1658	0.1746	0.1669	0.221	0.784
600	-	50	tanh	0.1795	0.1735	0.1804	0.173	0.217	0.77
600	-	50	relu	0.1768	0.1701	0.1777	0.1707	0.213	0.756
600	500	100	tanh	0.1799	0.1735	0.1808	0.1743	0.219	0.776
600	500	100	relu	0.1744	0.1683	0.1773	0.1694	0.216	0.765
600	500	50	tanh	0.1817	0.1752	0.1823	0.175	0.214	0.758
600	500	50	relu	0.1765	0.1702	0.1791	0.1711	0.209	0.743

Πίνακας 6.6: Τα αποτελέσματα των Αυτοκωδικοποιητών Αποθυροβοποίησης στην αναζήτηση Grid για το μικρό σύνολο δεδομένων

First Layer:600, Intermiment Layers:[500],Hidden Layer:50,Activation:tanh



Σχήμα 6.6: Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή Αποθορυβοποίησης για το μικρό σύνολο δεδομένων

6.2.4 Αυτοκωδικοποιητές Διακύμανσης

Ενώ οι Variational Autoencoders αποτελούν παραγωγικά (ή αναγεννητικά) δίκτυα που έχουν χρησιμοποιηθεί ευρέως για την παραγωγή προσώπων, κειμένων κ.ο.κ., έχουν αναπτυχθεί αρκετά ενδιαφέρουσες προσεγγίσεις [16] που αξιοποιούν τις δυνατότητες των Variational Autoencoders ως σύστημα προτάσεων με αρκετά καλές επιδόσεις.

Κωδικοποιητής

Ο κωδικοποιητής μοντελοποιήθηκε έτσι ώστε η έξοδος του να αναπαριστά μία πολυδιάστατη κατανομή Gauss. Το μοντέλο δέχεται ως είσοδο ένα υποσύνολο διανυσμάτων από το σύνολο δεδομένων. Κάθε διάνυσμα αντιστοιχεί σε μία λίστα αναπαραγωγής. Με 1 συμβολίζονται τα μουσικά κομμάτια που υπάρχουν στη λίστα, ενώ με 0 τα κομμάτια που απουσιάζουν από αυτή.

Η είσοδος του κωδικοποιητή περνάει από ένα στάδιο dropout με σκοπό τον περιορισμό της υπερεκπαίδευσης του μοντέλου. Τα ενδιάμεσα επίπεδα του δικτύου περιγράφονται από ένα Multilayer Perceptron με το πλήθος των νευρώνων κάθε επόμενου σταδίου να μειώνεται και να καταλήγει σε δύο κρυφές εξόδους ίδιας διάστασης, με την πρώτη να αντιστοιχεί στο διάνυσμα που αντιπροσωπεύει τη μέση τιμή μ και τη δεύτερη στο διάνυσμα του λογαρίθμου της διακύμανσης¹.

Τριχ Επαναπαραμετροποίησης

Με το τριχ επαναπαραμετροποίησης, δειγματοληπτείται το διάνυσμα \mathbf{z} . Πιο συγκεκριμένα, κατασκευάζεται αρχικά ένα βοηθητικό διάνυσμα ϵ μέσα από τη δειγματοληψία μίας τυποποιημένης κανονικής κατανομής (δηλαδή, με μηδενικό μέσο όρο και διακύμανση ίση με 1). Το διάνυσμα που δημιουργείται είναι ίδιας διάστασης με τα αντίστοιχα διανύσματα του κρυφού επιπέδου, που αντιστοιχούν στη μέση τιμή και το λογάριθμο της διακύμανσης. Μέσα από το συνδυασμό τους προκύπτει το διάνυσμα \mathbf{z} ως $\mu \times \epsilon + \sigma$, με $\epsilon \in [0, 1]$.

¹Επιλέγεται ο λογάριθμος της διακύμανσης και όχι απλά η διακύμανση λόγω αριθμητικής ευστάθειας και εξασφάλισης ότι η διακύμανση θα είναι θετική.

Αποκωδικοποιητής

Η είσοδος του αποκωδικοποιητή είναι το διάνυσμα \mathbf{z} . Τα ενδιάμεσα επίπεδα του δικτύου περιγράφονται από ένα Multilayer Perceptron με το πλήθος των νευρώνων κάθε σταδίου να αυξάνεται σε σχέση με το προηγούμενο. Η διάσταση του σταδίου εξόδου είναι ίδια με την είσοδο του κωδικοποιητή, δηλαδή ο αριθμός των μουσικών κομματιών. Ως συνάρτηση ενεργοποίησης του σταδίου εξόδου επιλέγεται η Softmax, με όλα τα στοιχεία της εξόδου να ανήκουν στο διάστημα $(0,1)$ και το άθροισμα τους να ισούται με 1.

Σημειώνεται ότι δοκιμάστηκαν αρχιτεκτονικές με συμμετρικές διαστάσεις των επιπέδων του κωδικοποιητή και του αποκωδικοποιητή.

Συνάρτηση Στόχου

Το κριτήριο εκπαίδευσης του μοντέλου είναι η μεγιστοποίηση του Evidence Lower Bound (ELBO). Αυτό προκύπτει, υπολογίζοντας διακριτά:

- Το reconstruction loss ως η εκτίμηση μέγιστης λογαριθμικής πιθανοφάνειας (log likelihood) της πολυωνυμικής κατανομής:

$$\log p_{\theta}(\mathbf{x}|\mathbf{z}) = \sum_i x_i \log y_i, \text{ με } y \text{ συμβολίζεται το διάνυσμα εξόδου.}$$

Όπου με \mathbf{x} συμβολίζεται το διάνυσμα εισόδου, ενώ με \mathbf{y} το διάνυσμα εξόδου, ενώ ο δείκτης i συμβολίζει τη θέση ενός στοιχείου στο διάνυσμα.

Στις αρχικές υλοποιήσεις του μοντέλου πραγματοποιήθηκαν δοκιμές με τη χρήση κατανομής Bernoulli. Ωστόσο, τα αποτελέσματα που προέκυπταν ήταν λανθασμένα, με το μοντέλο να προκρίνει για κάθε είσοδο τα μουσικά κομμάτια με τη μεγαλύτερη συχνότητα εμφάνισης στις λίστες αναπαραγωγής.

- Η απόκλιση Kullback Leibler μεταξύ της κατανομής που ακολουθεί η μεταβλητή \mathbf{z} και της κατανομής της προσέγγισης $q_{\phi}(\mathbf{z}|\mathbf{x})$. Επειδή έχει θεωρηθεί ότι η μεταβλητή \mathbf{z} ακολουθεί την τυποποιημένη κανονική κατανομή και επειδή η $q_{\phi}(\mathbf{z}|\mathbf{x})$ αντίστοιχα θεωρείται ότι ακολουθεί κανονική κατανομή, ο υπολογισμός της απόκλισης K-L γράφεται αναλυτικά ως:

$$D_{KL}(q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) = 1 + \log \sigma^2 - \mu^2 - \exp \log \sigma^2$$

Όπως έχει ήδη αναφερθεί στη θεωρία, οι συναρτήσεις αυτές είναι παραγωγίσιμες και συνεπώς μπορεί να χρησιμοποιηθεί ένας optimizer για τον υπολογισμό των παραγώγων και την εκπαίδευση του δικτύου. Στην συγκεκριμένη υλοποίηση χρησιμοποιήθηκε ο Adam.

Η διαδικασία εκπαίδευσης

Κατασκευάστηκε το βήμα εκπαίδευσης από την αρχή. Αναλυτικά: Το σύνολο δεδομένων χωρίστηκε σε παρτίδες (batches). Για κάθε εποχή, κάθε τμήμα εισάγεται στο νευρωνικό δίκτυο, υπολογίζεται η συνάρτηση κόστους και τα αποτελέσματα των μετρικών. Στη συνέχεια, πραγματοποιείται η αναπροσαρμογή των βαρών του δικτύου με τη χρήση του αλγόριθμου βελτιστοποίησης Adam.

Κατά τη διάρκεια της εκπαίδευσης των μοντέλων στο μεγάλο σύνολο δεδομένων, χρειάστηκε η αξιοποίηση της αραιής μορφής αναπαράστασης του συνόλου εκπαίδευσης, λόγω προβλημάτων στη διαχείριση της μνήμης (Out of Memory). Η μετατροπή αυτή, σε συνδυασμό με τη χρήση των παρτίδων επιτρέπει την χρήση πολύ μεγάλων σε μέγεθος συνόλων δεδομένων. Μία εναλλακτική προσέγγιση, θα ήταν η μετατροπή των δεδομένων σε αρχεία TF Records και η αποθήκευσή τους στο σκληρό δίσκο. Το μειονέκτημα αυτής της μεθόδου αποτελεί η καθυστέρηση στις εντολές εγγραφής/ανάγνωσης του πιο αργού αποθηκευτικού μέσου. Στον αντίποδα, το πλεονέκτημα της είναι η δυνατότητα χρήσης συσκευών με μεγάλη αποθηκευτική δυνατότητα.

Υπολογισμός Προτάσεων

Ο υπολογισμός των προτάσεων (recommendations) των μουσικών κομματιών για μία λίστα αναπαραγωγής, αντιστοιχεί στην έξοδο του αποκωδικοποιητή και εκφράζονται ως την πιθανότητα κάθε κομματιού να είναι το πιο κατάλληλο για επιλογή. Με την ταξινόμηση των πιθανοτήτων αυτών, προκύπτουν οι βέλτιστες προτάσεις.

Υπολογισμός των Μετρικών στα Σύνολα Δεδομένων Επικύρωσης και Ελέγχου

Για κάθε λίστα αναπαραγωγής από τα σύνολα επικύρωσης και ελέγχου έχει αφαιρεθεί, με τυχαίο τρόπο, το 20% των μουσικών κομματιών. Το πρώτο κομμάτι των δύο συνόλων χρησιμοποιείται για τον υπολογισμό των προτάσεων που αντιστοιχούν σε κάθε χρήστη. Για το σύνολο εκπαίδευσης δεν κρίθηκε αναγκαία η εφαρμογή της παραπάνω διαδικασίας, καθώς ο λόγος για τον υπολογισμό των μετρικών σε αυτό το σύνολο είναι ο οπτικός προσδιορισμός του σημείου υπερεκπαίδευσης.

Συγκεκριμένα, για τον υπολογισμό των μετρικών Recall@50 και NDCG@50 για τα σύνολα δεδομένων επικύρωσης και ελέγχου, χρησιμοποιούνται:

- Οι προτάσεις του πρώτου κομματιού των συνόλων, όπως υπολογίστηκαν από το μοντέλο
- Τα μουσικά κομμάτια που αφαιρέθηκαν, ως τα προσδοκώμενα αποτελέσματα.

ενώ για τον υπολογισμό των μετρικών Recall@50 και NDCG@50 στο σύνολο δεδομένων εκπαίδευσης:

- Οι προτάσεις προκύπτουν από την έξοδο του νευρωνικού δικτύου με είσοδο το σύνολο εκπαίδευσης.
- Το σύνολο εκπαίδευσης, ως τα προσδοκώμενα αποτελέσματα.

Κριτήριο Επιλογής Βέλτιστου Μοντέλου

Για την παύση της εκπαίδευσης σε μία συγκεκριμένη τοπολογία αξιοποιείται το κριτήριο Early Stopping, όταν δηλαδή οι βαθμολογίες της Recall@50 στο σύνολο επικύρωσης αρχίζουν να μειώνονται και αποθηκεύονται οι παράμετροι του μοντέλου.

Για την επιλογή της βέλτιστης τοπολογίας του μοντέλου, χρησιμοποιήθηκε η μεγαλύτερη τιμή της Recall@50 για το σύνολο δεδομένων ελέγχου.

Σύγκριση Διαφορετικών Τοπολογιών - Βελτιστοποίηση Υπερπαραμέτρων

Για την εύρεση του βέλτιστου μοντέλου πραγματοποιήθηκε Gridsearch στις εξής παραμέτρους του μοντέλου:

- Για το πρώτο επίπεδο, το επίπεδο δοκιμάστηκε με 600 και 1000 νευρώνες
- Δοκιμάστηκαν δίκτυα χωρίς ενδιάμεσο επίπεδο ή με ένα ενδιάμεσο επίπεδο 500 νευρώνων.
- Το επίπεδο εξόδου υλοποιήθηκε με 50 και 200 νευρώνες.
- Ως προς τη συνάρτηση ενεργοποίησης των επιπέδων δοκιμάστηκαν η tanh και η ReLU.
- Για τον πολλαπλασιαστή β στον όρο Kullback-Leibler πραγματοποιήθηκαν πειράματα:
 - Με το β να αυξάνεται σταδιακά από το 0 μέχρι το 0.2.(annealing)
 - Με το β να αυξάνεται σταδιακά από το 0 μέχρι το 1.

– Με το β να είναι σταθερό με τιμή 20.

Οι τιμές των υπερπαραμέτρων του μοντέλου με τη καλύτερη απόδοση είναι οι εξής: Για το μοντέλο που εκπαιδεύτηκε στο μικρό σύνολο δεδομένων:

- 2 επίπεδα. Το πρώτο επίπεδο έχει μέγεθος 600 νευρώνων και το κρυφό επίπεδο 50 νευρώνων.
- επιλέχθηκε η tanh συνάρτηση ενεργοποίησης.
- Ο πολλαπλασιαστής β ισούται με 1.

Για το μοντέλο που εκπαιδεύτηκε στο μεγάλο σύνολο δεδομένων:

- 3 επίπεδα. Το πρώτο επίπεδο έχει μέγεθος 600 νευρώνων, το δεύτερο 500 νευρώνων και το κρυφό αποτελείται από 50 νευρώνες.
- επιλέχθηκε η tanh συνάρτηση ενεργοποίησης.
- Ο πολλαπλασιαστής β ισούται με 2.

Τα αποτελέσματα της αναζήτησης παρατίθενται στους παρακάτω πίνακες, για τους Αυτοκωδικοποιητές Διακύμανσης που εκπαιδεύθηκαν στο μεγάλο και στο μικρό σύνολο δεδομένων καθώς και τα γραφήματα της εξέλιξης της εκπαίδευσης του βέλτιστου μοντέλου σε κάθε σύνολο. Στα γραφήματα που παρουσιάζονται, επιτράπηκε στο μοντέλο να ολοκληρώσει όλες τις εποχές, μόνο για την καταγραφή της εξέλιξης της εκπαίδευσης. Για κάθε τοπολογία, το μοντέλο που αποθηκεύεται είναι αυτό με την καλύτερη απόδοση στο σύνολο επικύρωσης.

Οι πολύ μεγάλες τιμές της NDCG@50 για το σύνολο εκπαίδευσης ερμηνεύονται από τη χρήση ολόκληρου του συνόλου κατά τη διάρκεια υπολογισμού της μετρικής.

Για το μεγάλο σύνολο δεδομένων:

First Layer	Interim Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	200	tanh	0.20258	0.18847	0.2051	0.189	0.30112	0.72501
1000	-	200	relu	0.19908	0.1851	0.20263	0.18522	0.30651	0.73136
1000	-	50	tanh	0.20481	0.19228	0.20757	0.19118	0.26963	0.64133
1000	-	50	relu	0.20357	0.19059	0.20574	0.19004	0.27202	0.64202
1000	500	200	tanh	0.20626	0.19407	0.20811	0.19238	0.26952	0.64602
1000	500	200	relu	0.20112	0.18907	0.20466	0.18946	0.28438	0.68324
1000	500	50	tanh	0.20755	0.19432	0.21006	0.19347	0.25574	0.6017
1000	500	50	relu	0.20337	0.19113	0.20603	0.191	0.26605	0.6299
600	-	200	tanh	0.20532	0.19198	0.20695	0.1907	0.27361	0.65898
600	-	200	relu	0.19969	0.18647	0.20357	0.18632	0.28365	0.67878
600	-	50	tanh	0.2048	0.1924	0.20817	0.19189	0.25443	0.60065
600	-	50	relu	0.20301	0.1903	0.20654	0.18959	0.25947	0.60857
600	500	200	tanh	0.20823	0.19541	0.2103	0.19455	0.24804	0.58078
600	500	200	relu	0.203	0.19079	0.20594	0.19071	0.26546	0.63432
600	500	50	tanh	0.20635	0.19589	0.2094	0.19555	0.24014	0.55591
600	500	50	relu	0.20379	0.19222	0.20673	0.19215	0.2503	0.58669

Πίνακας 6.7: Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μεγάλο σύνολο δεδομένων, με σταδιακή αύξηση του β από το 0 έως το 0.2.

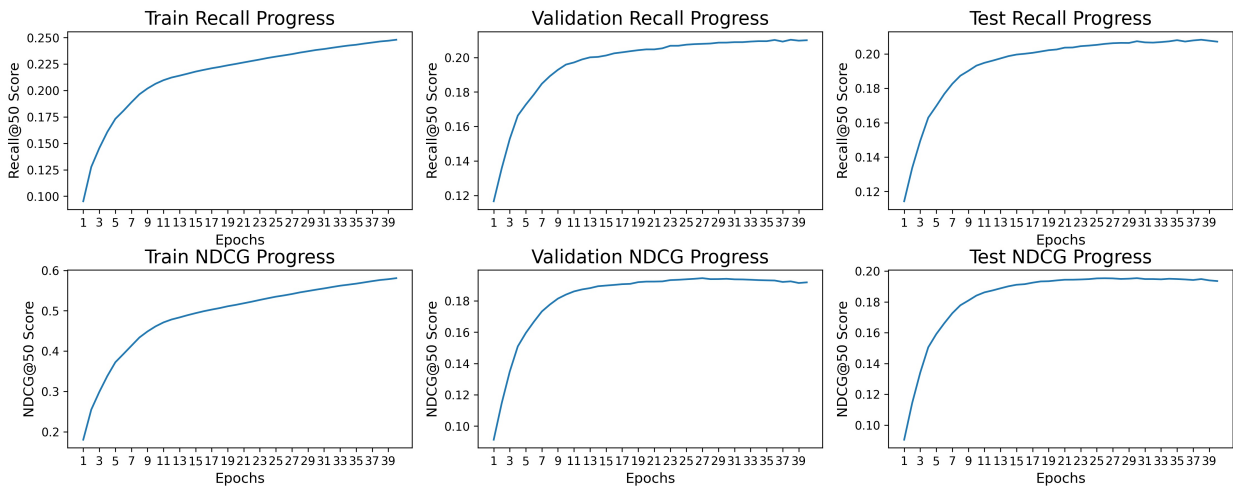
First Layer	Interminent Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	200	tanh	0.20246	0.18892	0.20502	0.1883	0.30125	0.72537
1000	-	200	relu	0.19925	0.18486	0.20167	0.18446	0.30669	0.73209
1000	-	50	tanh	0.20309	0.19148	0.20655	0.19031	0.26987	0.64212
1000	-	50	relu	0.20247	0.1893	0.20552	0.18919	0.27202	0.64249
1000	500	200	tanh	0.20648	0.19416	0.20988	0.19312	0.26926	0.64569
1000	500	200	relu	0.20177	0.18886	0.20492	0.18884	0.28503	0.68539
1000	500	50	tanh	0.2074	0.19461	0.21005	0.19447	0.25533	0.59988
1000	500	50	relu	0.20405	0.19156	0.20666	0.19105	0.26565	0.62858
600	-	200	tanh	0.20329	0.19038	0.20715	0.19111	0.27432	0.66029
600	-	200	relu	0.20012	0.18649	0.20255	0.18583	0.28379	0.67901
600	-	50	tanh	0.20587	0.19314	0.20811	0.1921	0.25497	0.60126
600	-	50	relu	0.20332	0.19036	0.20608	0.18992	0.25916	0.60765
600	500	200	tanh	0.20758	0.1949	0.2101	0.19352	0.249	0.58429
600	500	200	relu	0.20313	0.19079	0.20552	0.18998	0.26582	0.63551
600	500	50	tanh	0.2066	0.19589	0.20928	0.19493	0.24022	0.55583
600	500	50	relu	0.20464	0.19273	0.20716	0.1924	0.25081	0.5877

Πίνακας 6.8: Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μεγάλο σύνολο δεδομένων, με σταδιακή αύξηση του β από το 0 έως το 1.

First Layer	Interminent Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	200	tanh	0.20178	0.18723	0.20557	0.18659	0.30568	0.734
1000	-	200	relu	0.19892	0.18472	0.20241	0.18523	0.307	0.73226
1000	-	50	tanh	0.20402	0.19012	0.20681	0.19078	0.27218	0.64773
1000	-	50	relu	0.20279	0.19033	0.20644	0.18986	0.27191	0.64225
1000	500	200	tanh	0.2067	0.19368	0.20834	0.1925	0.27248	0.65427
1000	500	200	relu	0.2015	0.18789	0.20533	0.18823	0.28391	0.68302
1000	500	50	tanh	0.2066	0.1936	0.2099	0.19279	0.25975	0.61221
1000	500	50	relu	0.2043	0.19181	0.20681	0.1918	0.2641	0.62451
600	-	200	tanh	0.20414	0.19069	0.20723	0.18973	0.27816	0.67044
600	-	200	relu	0.19998	0.1861	0.20243	0.1857	0.28292	0.67645
600	-	50	tanh	0.20606	0.19352	0.20903	0.19166	0.25845	0.61125
600	-	50	relu	0.20355	0.19029	0.20633	0.19012	0.25957	0.60885
600	500	200	tanh	0.2075	0.19478	0.2105	0.19458	0.25142	0.59125
600	500	200	relu	0.20213	0.18987	0.20561	0.19001	0.26553	0.63513
600	500	50	tanh	0.20682	0.19459	0.20914	0.19409	0.24467	0.56869
600	500	50	relu	0.20429	0.19293	0.20882	0.19262	0.24969	0.58428

Πίνακας 6.9: Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μεγάλο σύνολο δεδομένων, με β σταθερό και ίσο με 20.

First Layer:600, Interminent Layers:[500],Hidden Layer:200,Activation:tanh



Σχήμα 6.7: Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή Διακύμανσης για το μεγάλο σύνολο δεδομένων

Για το μικρό σύνολο δεδομένων:

First Layer	Interminent Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	200	tanh	0.17253	0.16577	0.17486	0.16692	0.24536	0.86347
1000	-	200	relu	0.16836	0.16042	0.168	0.15996	0.24245	0.85425
1000	-	50	tanh	0.17804	0.17214	0.18072	0.17313	0.22134	0.78345
1000	-	50	relu	0.17614	0.17016	0.178	0.16991	0.21959	0.77757
1000	500	200	tanh	0.1781	0.17225	0.18023	0.17396	0.23061	0.81546
1000	500	200	relu	0.17179	0.1658	0.17278	0.165	0.22879	0.80924
1000	500	50	tanh	0.18078	0.17596	0.18287	0.17643	0.21865	0.77431
1000	500	50	relu	0.17591	0.17022	0.1775	0.16977	0.21619	0.76555
600	-	200	tanh	0.17545	0.16896	0.17636	0.16903	0.231	0.8177
600	-	200	relu	0.16788	0.16059	0.16966	0.16015	0.23076	0.81714
600	-	50	tanh	0.17955	0.174	0.18206	0.17444	0.21606	0.76612
600	-	50	relu	0.17633	0.16975	0.17764	0.17009	0.21371	0.75844
600	500	200	tanh	0.17922	0.17392	0.18048	0.17387	0.21858	0.775
600	500	200	relu	0.17247	0.16681	0.17402	0.16615	0.21904	0.77731
600	500	50	tanh	0.18091	0.17488	0.18336	0.17651	0.21106	0.74863
600	500	50	relu	0.17715	0.17086	0.17864	0.17104	0.20967	0.74371

Πίνακας 6.10: Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μικρό σύνολο δεδομένων, με σταδιακή αύξηση του β από το 0 έως το 0.2.

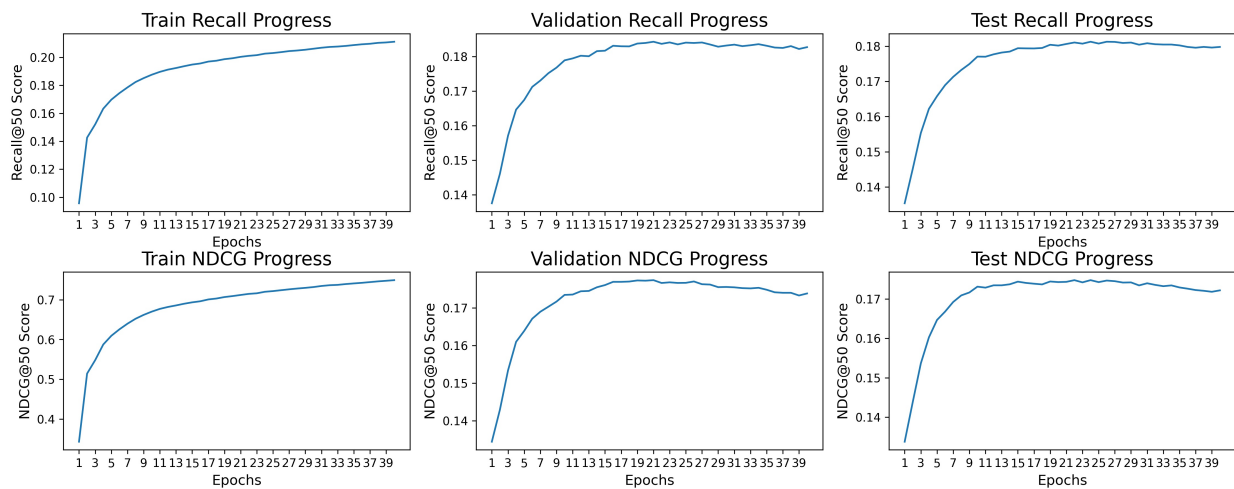
First Layer	Interminent Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	200	tanh	0.17292	0.16652	0.17551	0.16622	0.24563	0.86424
1000	-	200	relu	0.16759	0.15905	0.16875	0.16	0.24256	0.85468
1000	-	50	tanh	0.17759	0.17174	0.18078	0.17352	0.22171	0.78502
1000	-	50	relu	0.17585	0.17058	0.17793	0.16992	0.2192	0.77662
1000	500	200	tanh	0.17788	0.17297	0.17945	0.17165	0.23082	0.81606
1000	500	200	relu	0.17206	0.16614	0.17278	0.16569	0.22957	0.81196
1000	500	50	tanh	0.17956	0.17471	0.18217	0.17513	0.21889	0.77546
1000	500	50	relu	0.17661	0.17022	0.17848	0.17062	0.21587	0.765
600	-	200	tanh	0.17508	0.16922	0.17692	0.1691	0.23103	0.81792
600	-	200	relu	0.16855	0.16166	0.16947	0.16077	0.23123	0.81845
600	-	50	tanh	0.17936	0.17318	0.18112	0.17412	0.21598	0.76589
600	-	50	relu	0.17584	0.17011	0.17694	0.17037	0.2142	0.75994
600	500	200	tanh	0.17949	0.1731	0.18089	0.17386	0.21879	0.77563
600	500	200	relu	0.17376	0.16731	0.17528	0.16728	0.21917	0.7778
600	500	50	tanh	0.18139	0.17571	0.18251	0.17575	0.21166	0.75089
600	500	50	relu	0.17712	0.17103	0.17972	0.17195	0.20951	0.74355

Πίνακας 6.11: Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μικρό σύνολο δεδομένων, με σταδιακή αύξηση του β από το 0 έως το 1.

First Layer	Interminent Layers	Hidden Layer	Activation	Test Recall@50	Test NDCG@50	Validation Recall@50	Validation NDCG@50	Train Recall@50	Train NDCG@50
1000	-	200	tanh	0.17208	0.16575	0.17518	0.16616	0.24622	0.86584
1000	-	200	relu	0.16686	0.1596	0.16898	0.15965	0.24221	0.85351
1000	-	50	tanh	0.17798	0.17236	0.18023	0.17318	0.2212	0.78329
1000	-	50	relu	0.17587	0.1698	0.17817	0.17006	0.21914	0.77636
1000	500	200	tanh	0.17797	0.17167	0.18078	0.17277	0.23151	0.81838
1000	500	200	relu	0.17148	0.16495	0.17278	0.16454	0.22951	0.8115
1000	500	50	tanh	0.18083	0.17451	0.18255	0.17541	0.2191	0.77594
1000	500	50	relu	0.17626	0.16962	0.17781	0.16988	0.21583	0.76466
600	-	200	tanh	0.1748	0.16891	0.17685	0.16904	0.23151	0.81952
600	-	200	relu	0.16717	0.15949	0.16967	0.16037	0.23043	0.81605
600	-	50	tanh	0.17954	0.17394	0.18144	0.17425	0.21633	0.76745
600	-	50	relu	0.17637	0.16962	0.17731	0.17017	0.21343	0.75722
600	500	200	tanh	0.17906	0.17301	0.18208	0.1748	0.21979	0.77945
600	500	200	relu	0.1724	0.16515	0.17375	0.16506	0.21924	0.77765
600	500	50	tanh	0.18031	0.17614	0.18257	0.17626	0.21203	0.75209
600	500	50	relu	0.1777	0.17155	0.18014	0.17242	0.20921	0.74255

Πίνακας 6.12: Τα αποτελέσματα των Αυτοκωδικοποιητών Διακύμανσης για το μικρό σύνολο δεδομένων, με β σταθερό και ίσο με 20.

First Layer:600, Interim Layers:[500],Hidden Layer:50,Activation:tanh



Σχήμα 6.8: Διάγραμμα εκπαίδευσης του βέλτιστου Αυτοκωδικοποιητή Διακύμανσης για το μικρό σύνολο δεδομένων

6.2.5 Η επίδραση του σταδίου Dropout και του πολλαπλασιαστή β της απόκλισης Kullback Leibler

Με την εισαγωγή ενός σταδίου dropout 50%, παρατηρείται βελτίωση στην απόδοση των μοντέλων και ανθεκτικότητα στην υπερεκπαίδευση, με τις μετρικές στα σύνολα ελέγχου και επικύρωσης να φθίνουν με πιο αργό ρυθμό.

Η επίδραση του πολλαπλασιαστή οδήγησε σε μία πολύ μικρή αύξηση της απόδοσης του μοντέλου.

Κεφάλαιο 7

Αποτελέσματα και Οπτικοποίηση Δεδομένων

Στο παρόν κεφάλαιο παρουσιάζονται τα αποτελέσματα των νευρωνικών δικτύων που εκπαιδεύθηκαν, στο μικρό και το μεγάλο σύνολο δεδομένων ελέγχου. Επιπροσθέτως, πραγματοποιείται οπτικοποίηση των δεδομένων με βάση το μουσικό είδος και η αναζήτηση σχέσεων μεταξύ των παραγόντων των μουσικών κομματιών και των χαρακτηριστικών.

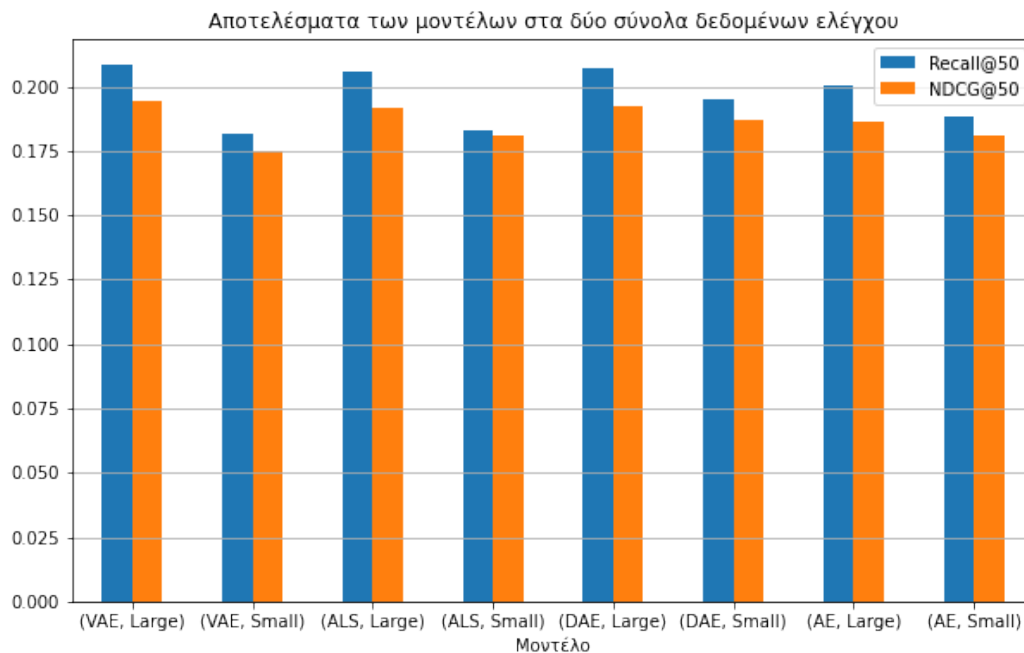
7.1 Αποτελέσματα

Για κάθε σύνολο δεδομένων, θεωρήθηκαν τα μοντέλα που κατά τη διάρκεια της εκπαίδευσης είχαν το καλύτερο αποτέλεσμα της μετρικής Recall@50 σύνολο δεδομένων επικύρωσης. Οι αρχιτεκτονικές των βέλτιστων μοντέλων περιγράφηκαν στο προηγούμενο κεφάλαιο.

Παρακάτω παρουσιάζονται τα αποτελέσματα των μοντέλων.

model dataset	VAE		ALS		DAE		AE	
	Large	Small	Large	Small	Large	Small	Large	Small
Recall@50	0.208167	0.181369	0.205509	0.18265	0.207115	0.194745	0.200329	0.188602
NDCG@50	0.194584	0.174733	0.191436	0.181049	0.19243	0.187061	0.186236	0.180928

Πίνακας 7.1: Αποτελέσματα στα δύο σύνολα δεδομένων ελέγχου



Σχήμα 7.1: Αποτελέσματα στα δύο σύνολα δεδομένων ελέγχου

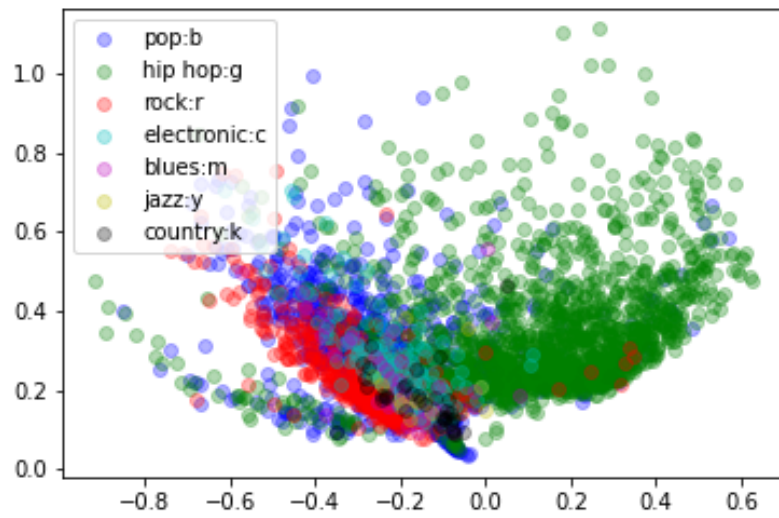
Το μοντέλο με την καλύτερη απόδοση είναι ο Αποκωδικοποιητής Διακύμανσης στο μεγάλο σύνολο δεδομένων, με τον Αυτοκωδικοποιητή Αποθρομβοποίησης να δίνει επίσης καλά αποτελέσματα, τα οποία είναι λίγο καλύτερα από τα αντίστοιχα του μοντέλου ALS. Οι βασικοί παράγοντες που βελτιώνουν τα αποτελέσματα στα μοντέλα είναι η αύξηση των παραδειγμάτων εκπαίδευσης, ενώ επιπρόσθετα, για τους αυτοκωδικοποιητές, η χρήση του σταδίου dropout.

7.2 Οπτικοποίηση δεδομένων με βάση το μουσικό είδος των κομματιών

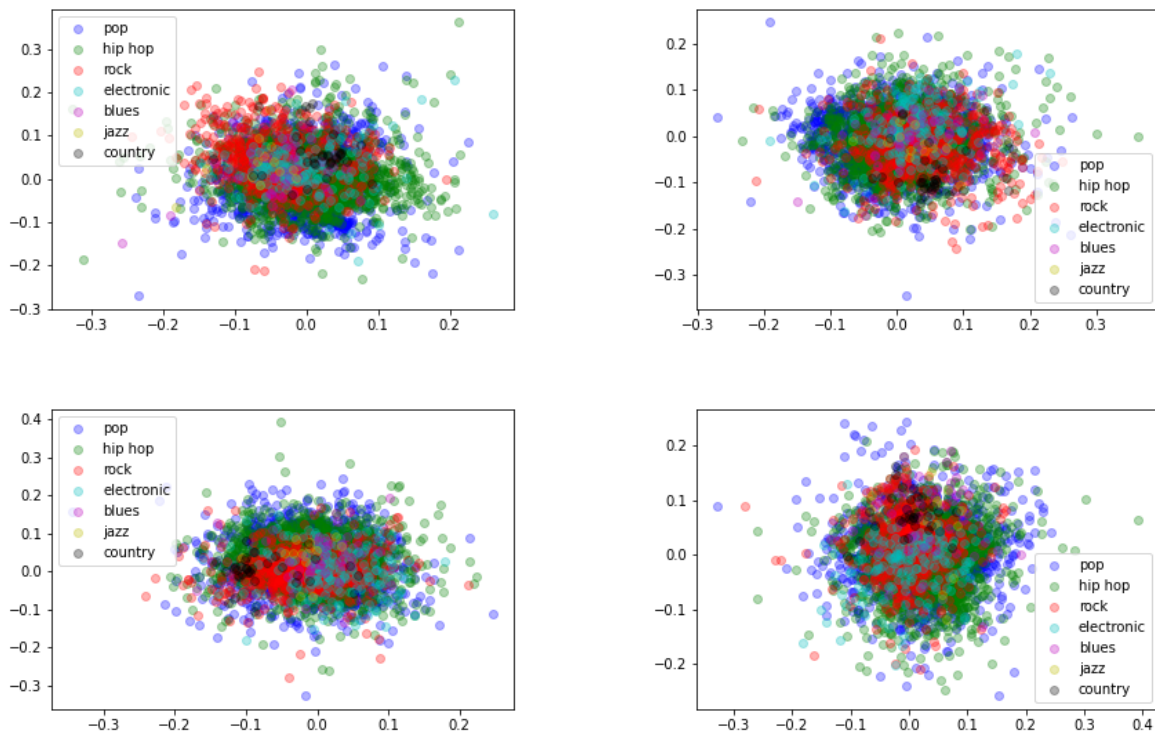
Για την οπτικοποίηση των κομματιών ανά μουσικό είδος πραγματοποιήθηκε η εξής διαδικασία:

Για τα μοντέλα παραγωγής πινάκων, κάθε γραμμή του πίνακα αντικείμενων-χαρακτηριστικών/παραγόντων (Item-Features Matrix) αποτελεί την απεικόνιση του μουσικού κομματιού στον πολυδιάστατο χώρο που ορίζεται από τα χαρακτηριστικά. Η οπτικοποίηση για δύο διαφορετικά μοντέλα πραγματοποιήθηκε ως εξής:

- Στην πρώτη περίπτωση, χρησιμοποιείται ένα μοντέλο δύο μόνο παραγόντων. Η απεικόνιση πραγματοποιείται μέσα από την δημιουργία ενός διαγράμματος διασποράς στις δύο διαστάσεις. Ενώ το μοντέλο φαίνεται να απεικονίζει μουσικά κομμάτια του ίδιου είδους σε κοντινή απόσταση, η απόδοση του μοντέλου δεν είναι ικανοποιητική.
- Στη δεύτερη περίπτωση, χρησιμοποιήθηκε το βέλτιστο μοντέλο 100 παραγόντων. Με τυχαίες επιλογές δύο παραγόντων (καθώς δεν είναι δυνατή η αποτύπωση πολλών διαστάσεων) μπορεί να δοθεί μία εικόνα για τον τρόπο που το μοντέλο απεικονίζει τα μουσικά κομμάτια στον χώρο των 100 διαστάσεων. Το βέλτιστο μοντέλο, σε αντίθεση με το προηγούμενο, δίνει αρκετά ικανοποιητικές προτάσεις. Ο χώρος που διαμορφώνεται είναι πιο ενιαίος, και τα μουσικά κομμάτια που ανήκουν στο ίδιο μουσικό είδος βρίσκονται κοντά μεταξύ τους.



Σχήμα 7.2: Οπτικοποίηση του μοντέλου παραγωγής πινάκων δύο χαρακτηριστικών

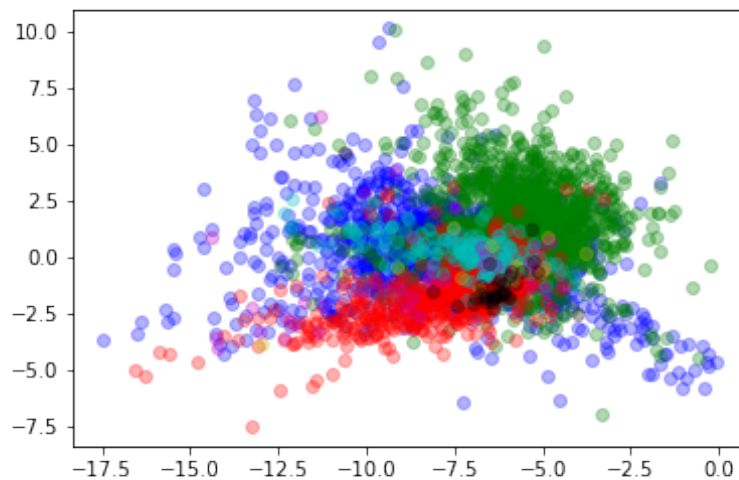


Σχήμα 7.3: Οπτικοποιήσεις τυχαίων χαρακτηριστικών του βέλτιστου μοντέλου παραγοντοποίησης πινάκων

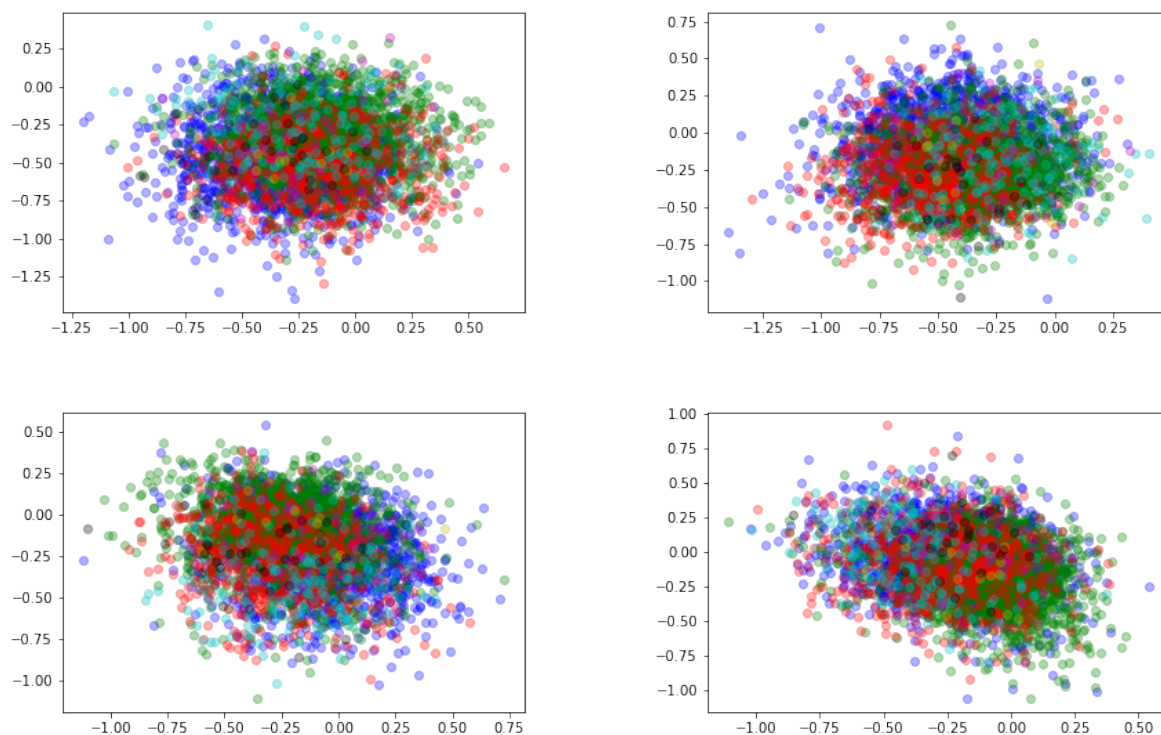
Για τους Αυτοκωδικοποιητές κατασκευάστηκε ένας μοναδιαίος διαγώνιος πίνακας ίσος με το πλήθος των μουσικών κομματιών που λειτούργησε ως είσοδος του δικτύου. Για τους απλούς Αυτοκωδικοποιητές και τους Αυτοκωδικοποιητές διακύμανσης, για την απεικόνιση των αποτελεσμάτων, χρησιμοποιήθηκε η έξοδος του κωδικοποιητή, ενώ για τους Αυτοκωδικοποιητές Διακύμανσης χρησιμοποιήθηκε η έξοδος του κωδικοποιητή που αντιστοιχεί στους μέσους όρους της πρότερης κατανομής. Πιο συγκεκριμένα:

Για κάθε δίκτυο εκπαιδεύτηκαν και χρησιμοποιήθηκαν:

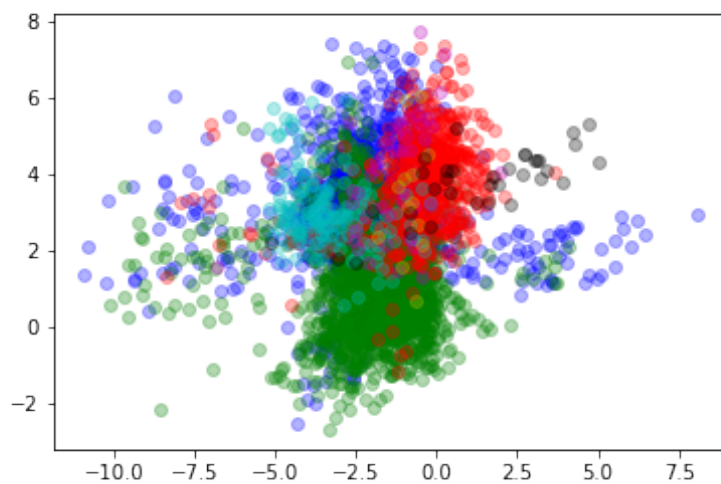
- Ένας κωδικοποιητής με κρυφό επίπεδο δύο διαστάσεων.
- Ο κωδικοποιητής των, κάθε φορά, βέλτιστων παραμέτρων.



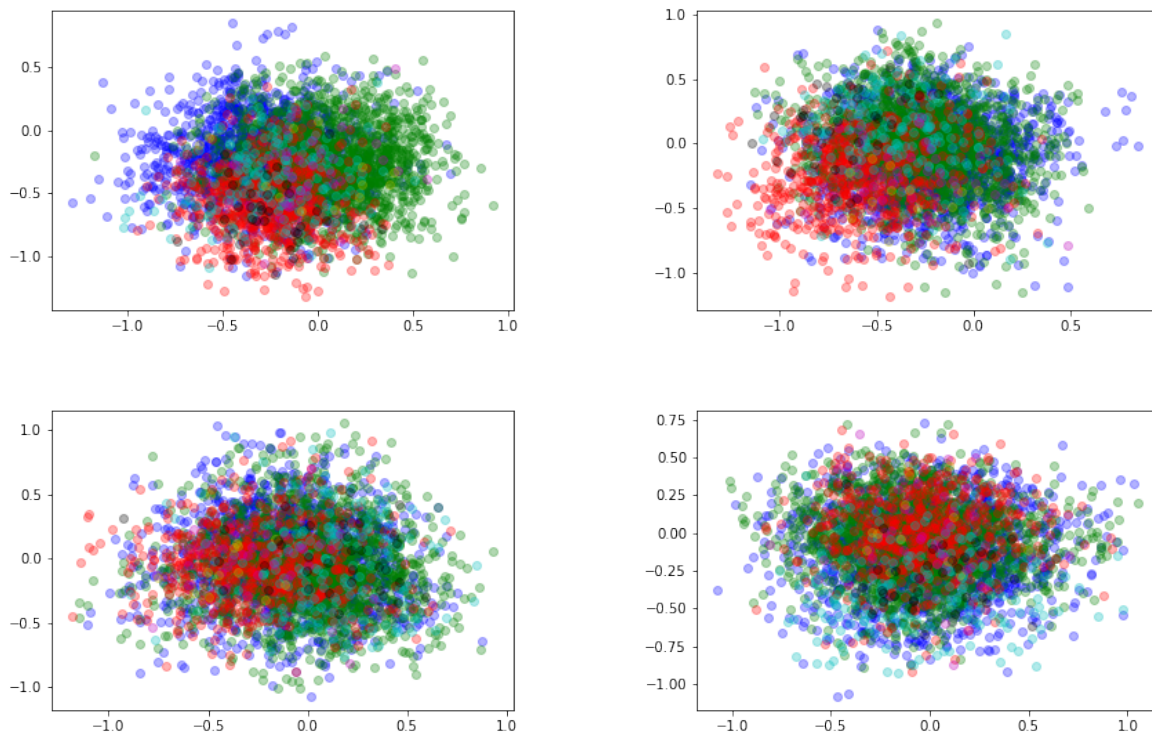
Σχήμα 7.4: Οπτικοποίηση του μοντέλου δύο διαστάσεων ενός απλού αυτοκωδικοποιητή



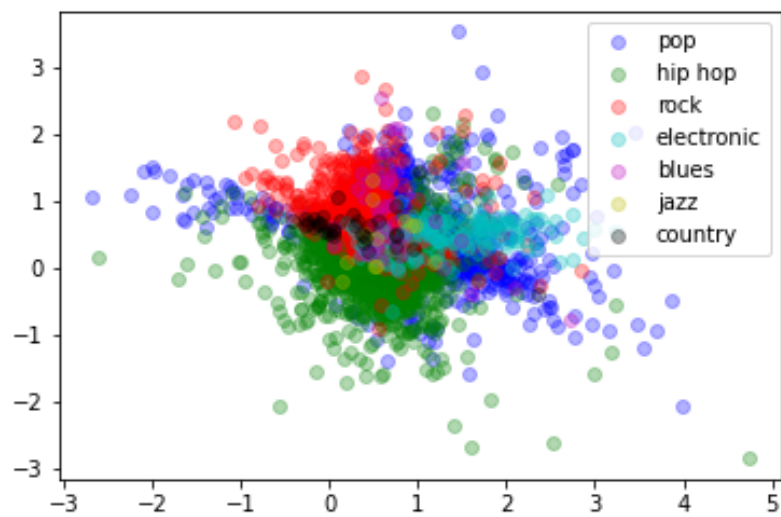
Σχήμα 7.5: Οπτικοποιήσεις τυχαίων χαρακτηριστικών του βέλτιστου μοντέλου απλού αυτοκωδικοποιητή



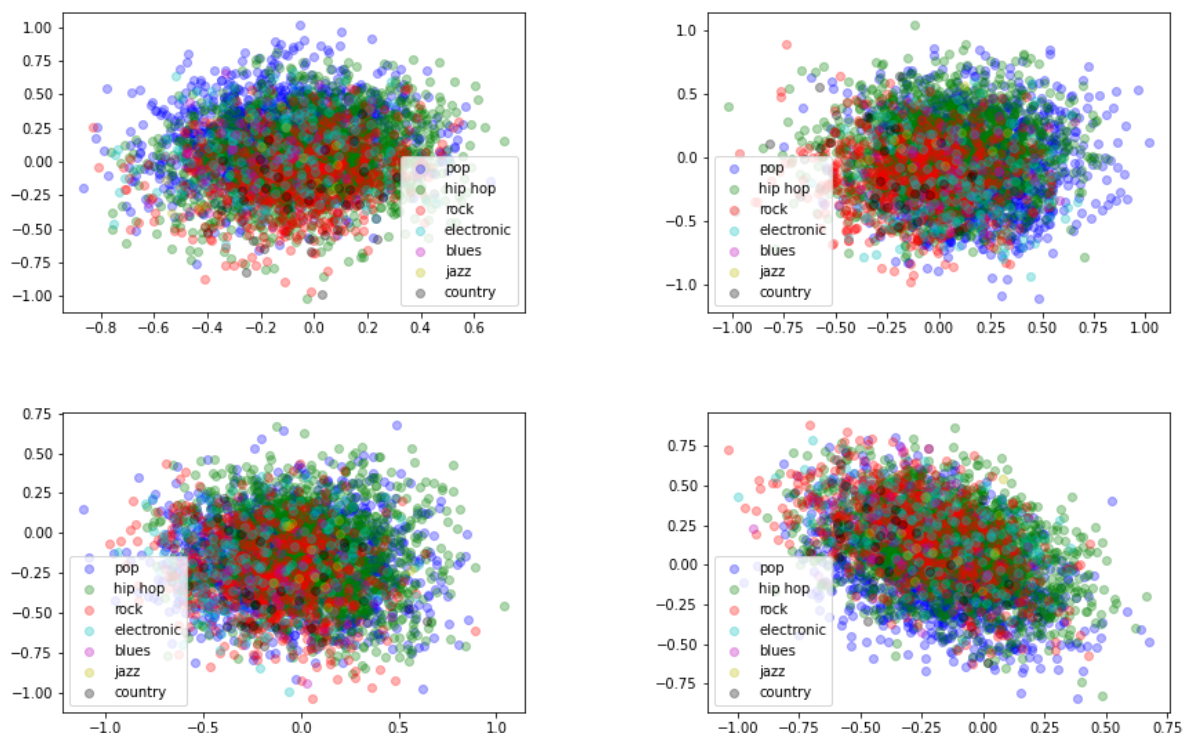
Σχήμα 7.6: Οπτικοποίηση ενός μοντέλου δύο διαστάσεων DAE



Σχήμα 7.7: Οπτικοποιήσεις τυχαίων χαρακτηριστικών του βέλτιστου μοντέλου DAE



Σχήμα 7.8: Οπτικοποίηση του μοντέλου VAE



Σχήμα 7.9: Οπτικοποιήσεις τυχαίων χαρακτηριστικών του βέλτιστου μοντέλου VAE

Για όλα τα δίκτυα Αυτοκωδικοποιητών, η αύξηση των διαστάσεων οδηγεί σε πιο ενιαίο χώρο και σαφώς καλύτερα αποτελέσματα στις προτάσεις. Επιπλέον, ακόμα και στα δίκτυα των δύο διαστάσεων φαίνεται να υπάρχει ένας σαφής διαχωρισμός των μουσικών κομματιών.

7.3 Ανάλυση χαρακτηριστικών

Παρακάτω περιγράφονται τα βήματα για την διερεύνηση της ύπαρξης γραμμικών σχέσεων μεταξύ των παραγόντων του εκάστοτε μοντέλου και των χαρακτηριστικών μουσικής ανάλυσης του Spotify, με σκοπό, αφενός την ποιοτική ερμηνεία της συμπεριφοράς των παραγόντων και αφετέρου την δυνατότητα παρέμβασης των προτάσεων προς κάποια επιθυμητή κατεύθυνση. Το πρώτο βήμα αποτελεί ο μετασχηματισμός σε κύριες συνιστώσες για τη μείωση των διαστάσεων. Στη συνέχεια υπολογίζονται οι συντελεστές συσχέτισης Pearson για την αναζήτηση γραμμικών σχέσεων μεταξύ των κύριων συνιστωσών και των μουσικών χαρακτηριστικών. Τέλος, πραγματοποιείται συσταδοποίηση των χαρακτηριστικών για την αναζήτηση διαχωρισμών στα χαρακτηριστικά.

7.3.1 Ανάλυση σε Κύριες Συνιστώσες

Το πρώτο στάδιο που πραγματοποιήθηκε ήταν η ανάλυση σε Κύριες Συνιστώσες (Principal Components Analysis). Ο αριθμός των συνιστωσών επιλέχθηκε στις 10. Παρακάτω παρατίθεται το ποσοστό της διακύμανσης που εξηγείται από τις 10 κύριες συνιστώσες για κάθε μοντέλο, που αποτελεί το μέτρο της αποτελεσματικότητας της μεθόδου :

model	ALS	VAE	DAE	AE
variance	17%	70.3%	61,8%	58%

Table 7.2: Ποσοστό της διακύμανσης που εξηγείται από την ανάλυση σε 10 κύριες συνιστώσες

7.3.2 Συντελεστές Γραμμικής Συσχέτισης

Στη συνέχεια, για κάθε παράγοντα υπολογίζεται ο συντελεστής συσχέτισης γινομένου ροπής Pearson (Pearson Coefficient Correlation) για την αναζήτηση γραμμικών σχέσεων μεταξύ των παραγόντων και των χαρακτηριστικών. Ισχυρή συσχέτιση που παρουσιάζει ανάλογη συμπεριφορά ισούται με 1, ισχυρή συσχέτιση που παρουσιάζει αντιστρόφως ανάλογη συμπεριφορά ισούται με -1, ενώ η απουσία συσχέτισης ισοδυναμεί με 0.

Παρακάτω παρουσιάζεται ο πίνακας των μεγαλύτερων και μικρότερων βαθμών συσχέτισης για τις βέλτιστες τοπολογίες των διαφορετικών μεθόδων:

	Min Values	Min Value Index	Max Values	Max Value Index
danceability	-0.225	3	0.314	0
energy	-0.376	3	0.083	8
acousticness	-0.071	4	0.372	3
instrumentalness	-0.072	9	0.153	8
valence	-0.223	0	0.195	7
tempo	-0.085	2	0.073	0
speechiness	-0.136	3	0.450	0
loudness	-0.306	3	0.199	1

Πίνακας 7.3: Συντελεστές συσχέτισης Pearson για το βέλτιστο μοντέλο ALS

	Min Values	Min Value Index	Max Values	Max Value Index
danceability	-0.294	6	0.246	7
energy	-0.193	9	0.324	7
acousticness	-0.296	7	0.210	9
instrumentalness	-0.074	5	0.125	8
valence	-0.163	6	0.231	4
tempo	-0.043	0	0.092	7
speechiness	-0.273	1	0.216	2
loudness	-0.210	9	0.183	3

Πίνακας 7.4: Συντελεστές συσχέτισης Pearson για το βέλτιστο μοντέλο AE

	Min Values	Min Value Index	Max Values	Max Value Index
danceability	-0.367	0	0.250	1
energy	-0.224	7	0.347	5
acousticness	-0.320	5	0.227	7
instrumentalness	-0.046	3	0.108	7
valence	-0.199	2	0.202	8
tempo	-0.062	0	0.111	5
speechiness	-0.503	0	0.078	1
loudness	-0.226	7	0.207	5

Πίνακας 7.5: Συντελεστές συσχέτισης Pearson για το βέλτιστο μοντέλο DAE

	Min Values	Min Value Index	Max Values	Max Value Index
danceability	-0.403	0	0.257	2
energy	-0.215	1	0.268	4
acousticness	-0.213	6	0.229	3
instrumentalness	-0.108	7	0.152	5
valence	-0.176	1	0.246	2
tempo	-0.080	0	0.073	4
speechiness	-0.499	0	0.060	8
loudness	-0.193	3	0.289	2

Πίνακας 7.6: Συντελεστές συσχέτισης Pearson για το βέλτιστο μοντέλο VAE

Στους παραπάνω πίνακες εντοπίζονται κάποιες ασθενείς συσχετίσεις κάποιων συνιστωσών και χαρακτηριστικών. Ενδεικτικά, αναφέρεται η ασθενής συσχέτιση του Αυτοκωδικοποιητή Αποθορυβοποίησης (DAE) που εμφανίζει αρνητική συσχέτιση μεταξύ της χρήσης λέξεων και της πρώτης συνιστώσας (που έχει δείκτη 0) και η χορευτική δυνατότητα ενός μουσικού κομματιού, με την πρώτη συνιστώσα στον Αποκωδικοποιητή Διακύμανσης.

7.3.3 Ποιοτική Ανάλυση με Συστατοδοποίηση

Για την περαιτέρω διερεύνηση της ύπαρξης σχέσεων μεταξύ των αποτελεσμάτων και των μουσικών χαρακτηριστικών πραγματοποιήθηκε συστατοδοποίηση (clustering) για τον βέλτιστο συνδυασμό δύο συνιστωσών. Η μεθοδολογία περιγράφεται παρακάτω:

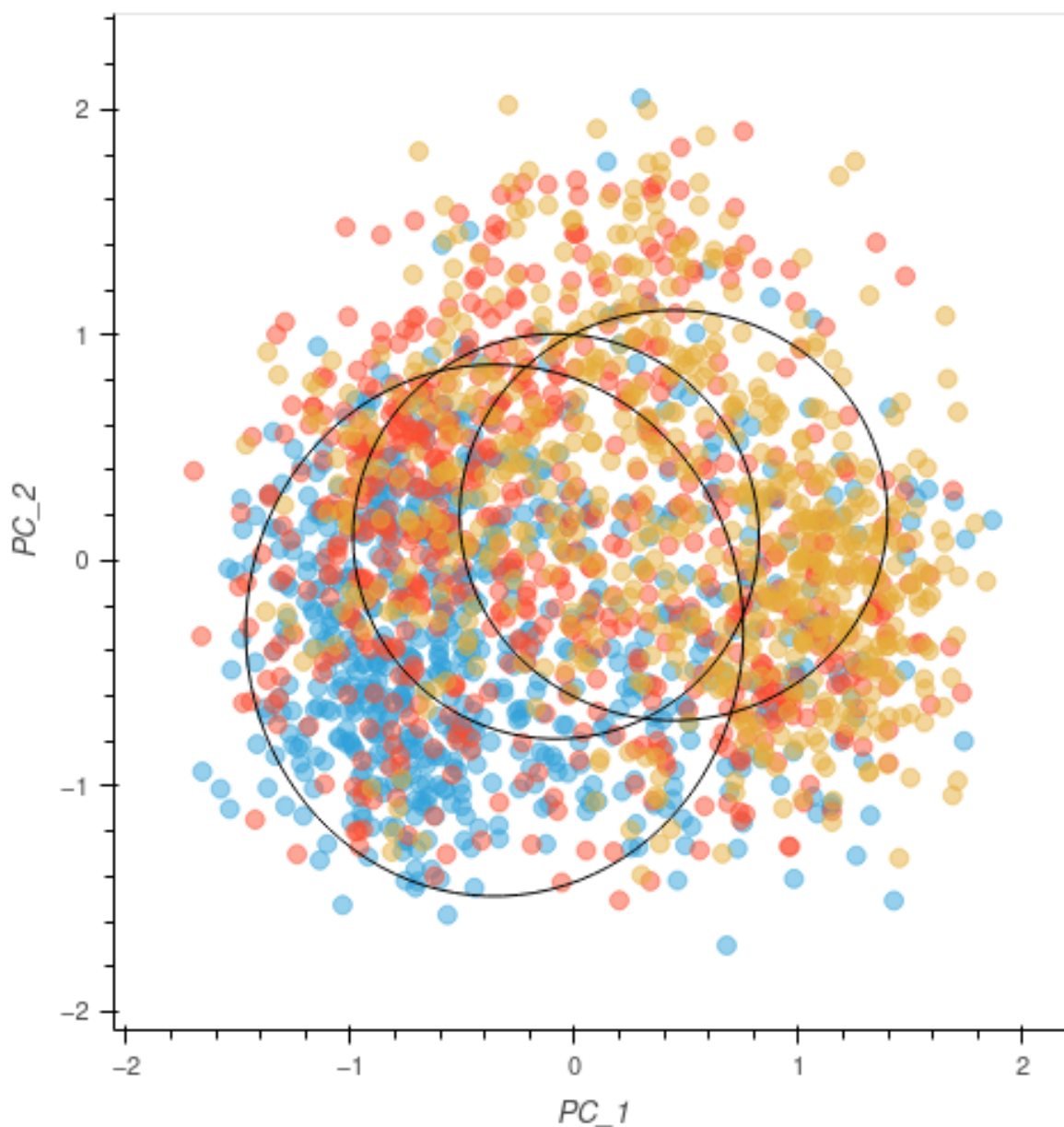
- Μέσα από τον υπολογισμό του ελάχιστου αθροίσματος Τετραγωνικού Σφάλματος (Sum of Squared Error), αναζητείται ο καλύτερος συνδυασμός δύο Συνιστωσών.
- Δειγματοληπτούνται 2000 από το σύνολο των μουσικών κομματιών.
- Χωρίζονται σε 3 ισοπληθής κλάσεις, βάσει τις τιμές των χαρακτηριστικών.
- Για κάθε κλάση, σχεδιάζεται μία έλλειψη:
 - Υπολογίζεται το σημείο της μέσης τιμής (ως προς τις δύο διαστάσεις).
 - Υπολογίζονται η μικρή και η μεγάλη ακτίνα μίας έλλειψης, ως το μισό της μέγιστης απόστασης ως προς κάθε άξονα. Δηλαδή:

$$\alpha = \max|\mu_{\mathbf{x}} - \mathbf{x}|/2 \quad (7.1)$$

$$\beta = \max|\mu_{\mathbf{y}} - \mathbf{y}|/2 \quad (7.2)$$

- Αφού υπολογιστούν οι ακτίνες είναι δυνατό να σχεδιαστεί η έλλειψη.

Ενδεικτικά, για τη χορευτική δυνατότητα στο βέλτιστο μοντέλο VAE προκύπτει η έλλειψη:



Σχήμα 7.10: Διαχωρισμός σε κλάσεις ενός χαρακτηριστικού στις δύο βέλτιστες συνιστώσες

Τα αποτελέσματα δεν καταδεικνύουν ισχυρή δυνατότητα συσταδοποίησης των χαρακτηριστικών των μουσικών κομματιών με τις κύριες συνιστώσες.

7.4 Κατεύθυνση Προτάσεων με Ανάλυση Κύριων Συνιστωσών

Για τα μοντέλα των Αυτοκωδικοποιητών που η διακύμανση εξηγείται σε ικανοποιητικό βαθμό, πραγματοποιείται προσπάθεια κατεύθυνσης των προτάσεων μέσα από την μεταβολή στις τιμές των συνιστωσών και τη δημιουργία του νέου διανύσματος εισόδου στον αποκωδικοποιητή, μέσα από τον αντίστροφο μετασχηματισμό PCA. Με αυτό τον τρόπο ερευνάται η δυνατότητα παρέμβασης στις προτάσεις που δίνονται από τα μοντέλα και διερεύνησης της δυνατότητας εξαγωγής γνώσης από αυτά.

Ενδεικτικά, για τον βέλτιστο Αυτοκωδικοποιητή Διακύμανσης, θέτοντας τις δύο πρώτες κύριες συνιστώσες στη τιμή 20, οι προτάσεις κατευθύνονται προς Pop μουσικά κομμάτια με περιορισμένη χρήση λέξεων: (Speechiness)

Name	Artists	Tags	danc...	energy	acou...	ins...	valence	tempo	speechiness	loudness	parsed_tags	Title
Memories	[Maroon 5]	[2010s', 'pop', 'ir	0.764	0.32	0.837	0.0	0.575	91.019	0.0546	-7.209	pop	Memories
I Like Me Better	[Lauv]	[indie', 'linedanc	0.752	0.505	0.535	0.000003	0.419	91.97	0.253	-7.621	pop	I Like Me Better
Way Back Home (feat. Conor Maynard) - Sam Feldt Edit	[SHAUN', 'Cono	[sam feldt', 'elec	0.737	0.747	0.269	0.0	0.447	105.943	0.0323	-4.818	pop	Way Back Home (feat. Conor Maynard)
Rewrite The Stars	[Zac Efron', 'Zer	[Soundtrack', 'm	0.684	0.619	0.0716	0.0	0.284	125.046	0.0386	-7.005	other	Rewrite The Stars
10,000 Hours (with Justin Bieber)	[Dan + Shay', 'Ji	[justin bieber', 'b	0.654	0.63	0.153	0.0	0.43	89.991	0.0259	-4.644	other	10,000 Hours (with Justin Bieber)
Boy With Luv (feat. Halsey)	[BTS', 'Halsey]	[bots', 'k-pop', 'p	0.645	0.862	0.0923	0.0	0.798	119.991	0.0965	-4.757	pop	Boy With Luv (feat. Halsey)
Don't Give Up On Me - (From "Five Feet Apart")	[Andy Grammer]	[Soundtrack', 'B	0.641	0.537	0.0875	0.000009	0.0681	112.864	0.0282	-5.452	other	Don't Give Up On Me - (From "Five Feet Apart")
Euphoria	[BTS]	[k-pop', 'jungkoo	0.637	0.799	0.394	0.0	0.562	104.995	0.0338	-4.519	pop	Euphoria
Perfect	[Ed Sheeran]	[pop', 'Ballad', 's	0.599	0.448	0.163	0.0	0.168	95.05	0.0232	-6.312	pop	Perfect
Midnight (feat. Liam Payne)	[Alesso', 'Liam F	[liam payne]	0.587	0.648	0.0561	0.0	0.281	144.028	0.0439	-6.927	other	Midnight (feat. Liam Payne)
Falling	[Harry Styles]	[pop', 'rock', 'alt	0.567	0.267	0.839	0.000001	0.0592	110.011	0.0299	-6.502	pop	Falling
What A Time (feat. Niall Horan)	[Julia Michaels',	[pop', 'justin tran	0.516	0.467	0.662	0.0	0.386	132.861	0.0302	-6.184	pop	What A Time (feat. Niall Horan)
Someone You Loved	[Lewis Capaldi]	[2010s', 'alternat	0.501	0.405	0.751	0.0	0.446	109.891	0.0319	-5.679	rock	Someone You Loved
Never Not	[Lauv]	[indie pop', 'hair	0.47	0.221	0.693	0.0	0.264	133.714	0.0371	-9.243	pop	Never Not
Sweet Night	[V]	[taewon class',	0.462	0.339	0.495	0.000008	0.149	92.106	0.025	-7.943	hip hop	Sweet Night
Before You Go	[Lewis Capaldi]	[alternative', 'ind	0.459	0.575	0.604	0.0	0.183	111.881	0.0573	-4.858	pop	Before You Go
Before You Go	[Lewis Capaldi]	[alternative', 'ind	0.459	0.575	0.604	0.0	0.183	111.881	0.0573	-4.858	pop	Before You Go
Lover	[Taylor Swift]	[pop', 'Ballad', 'c	0.359	0.543	0.492	0.000016	0.453	68.534	0.0919	-7.582	pop	Lover
Say You Won't Let Go	[James Arthur]	[pop]	0.358	0.557	0.695	0.0	0.494	85.043	0.059	-7.398	pop	Say You Won't Let Go

Σχήμα 7.11: Η επίδραση στις προτάσεις των δύο πρώτων συνιστωσών με τιμή 20

Απεναντίας, όταν οι ίδιες συνιστώσες τίθενται στο -20, οι προτάσεις που προκύπτουν κατευθύνονται προς το Hip Hop με αισθητά πιο ψηλές τιμές στη χρήση λέξεων:

Name	Artists	Tags	d...	e...	aco...	instru...	vale...	tempo	speechi...	lou...	pars...	Title
Who Shot Ya? - 2005 Remaster	[The Notorious B.I.P.]		0.831	0.682	0.256	0.0	0.6	91.771	0.238	-5.697	other	Who Shot Ya? - 2005 Remaster
Survival of the Fittest	[Mobb Deep]	[Hip-Hop', 'mobt	0.813	0.703	0.125	0.000001	0.241	94.828	0.268	-5.077	hip hop	Survival of the Fittest
Real Muthaphuckkin' G's	[Eazy-E]	[Gangsta Rap',	0.724	0.853	0.0164	0.0	0.775	170.51	0.283	-3.895	hip hop	Real Muthaphuckkin' G's
Hard To Choose One	[Future]		0.884	0.436	0.011	0.0	0.254	119.013	0.255	-9.922	other	Hard To Choose One
Brand New Draco	[21 Savage', 'Me	[Gangsta Rap',	0.697	0.709	0.041	0.000296	0.439	155.023	0.298	-6.642	hip hop	Brand New Draco
Rap Saved Me (feat. Quavo)	[21 Savage', 'Of	[trap]	0.845	0.399	0.00525	0.741	0.148	143.049	0.312	-7.509	hip hop	Rap Saved Me (feat. Quavo)
N.Y. State of Mind	[Nas]	[Hip-Hop', 'rap',	0.665	0.91	0.0394	0.0	0.887	84.099	0.223	-4.682	hip hop	N.Y. State of Mind
Slidin	[21 Savage', 'Me		0.774	0.706	0.00163	0.000001	0.382	152.891	0.355	-6.494	other	Slidin
EA (feat. 21 Savage)	[Young Nudy', '2	[armed robbery	0.451	0.876	0.261	0.0	0.227	87.306	0.414	-5.205	other	EA (feat. 21 Savage)
My Dawg	[21 Savage', 'Me	[Gangsta Rap',	0.831	0.532	0.0652	0.0	0.364	140.105	0.301	-5.319	hip hop	My Dawg
No Auto (feat. Lil Durk)	[Lil Uzi Vert', 'Lil	[LIL DURK', 'Lil	0.713	0.681	0.00286	0.0	0.635	159.107	0.351	-7.645	other	No Auto (feat. Lil Durk)
Choppa Won't Miss (feat. Young Thug)	[Playboi Carti',	[trap', 'rap]	0.857	0.661	0.0182	0.0	0.612	163.077	0.319	-6.544	hip hop	Choppa Won't Miss (feat. Young Thug)
Many Men	[21 Savage', 'Me		0.635	0.793	0.000647	0.747	0.281	77.594	0.0322	-4.333	other	Many Men
No Opp Left Behind	[21 Savage', 'Me	[Gangsta Rap',	0.823	0.703	0.0161	0.000001	0.26	178.043	0.298	-2.691	hip hop	No Opp Left Behind
Glock In My Lap	[21 Savage', 'Me		0.847	0.733	0.00574	0.000001	0.202	130.029	0.16	-6.439	other	Glock In My Lap
Shook Ones, Pt. II	[Mobb Deep]	[Hip-Hop', 'rap',	0.763	0.786	0.0146	0.0114	0.504	93.857	0.229	-6.472	hip hop	Shook Ones, Pt. II
NBAYOUNGBOAT	[Lil Yachty', 'You	[Hip-Hop', 'trap',	0.769	0.807	0.0515	0.0	0.364	82.422	0.306	-2.966	hip hop	NBAYOUNGBOAT
My Choppa Hate Niggas	[21 Savage', 'Me	[experimental hi	0.862	0.274	0.0974	0.185	0.154	77.492	0.416	-5.818	hip hop	My Choppa Hate Niggas
New Choppa	[Playboi Carti',	[experimental hi	0.773	0.605	0.174	0.0	0.425	128.16	0.0579	-4.223	hip hop	New Choppa
Russian Cream	[Key Glock]		0.867	0.774	0.11	0.0	0.549	130.007	0.315	-5.791	other	Russian Cream

Σχήμα 7.12: Η επίδραση στις προτάσεις των δύο πρώτων συνιστωσών με τιμή -20

7.5 Σύστημα Προτάσεων

Αναπτύχθηκε ένα σύστημα προτάσεων που δέχεται ως είσοδο μουσικά κομμάτια που σχηματίζουν μία λίστα αναπαραγωγής και παράγει προτάσεις με βάση αυτές. Για κάθε νέο τραγούδι που προστίθεται, οι προτάσεις για το επόμενο μουσικό κομμάτι υπολογίζονται εκ νέου. Για παράδειγμα, σχηματίστηκε μία λίστα αναπαραγωγής με Rock μουσικά κομμάτια, στο βέλτιστο μοντέλο VAE:

Filter selected options

Good Times Bad Times - 1993 Remaster ^
 Immigrant Song - Remaster
 Sultans Of Swing
 Bohemian Rhapsody

Score ▲	Name ▲	Artists ▲	Tags ▲	d
0.032012	Sultans Of Swing	[Dire Straits]	[classic rock', 'ro	
0.022491	Immigrant Song - Remaster	[Led Zeppelin]	[hard rock', 'devc	
0.021264	Don't Stop Me Now - Remastered 2011	[Queen]	[rock', 'Queen']	
0.018884	Bohemian Rhapsody - Remastered 2011	[Queen]	[classic rock', 'h	
0.016242	Highway to Hell	[AC/DC]	[hard rock', 'rock	
0.014991	Another One Bites The Dust - Remastered 2011	[Queen]	[classic rock', 'ro	
0.014082	Whole Lotta Love - 1990 Remaster	[Led Zeppelin]	[proto-metal]	
0.012865	Back In Black	[AC/DC]	[hard rock', 'rock	
0.012799	Sweet Child O' Mine	[Guns N' Roses]	[rock', 'hard rock	
0.012753	Cocaine	[Eric Clapton]	[classic rock', 'ro	
0.012622	La Grange - 2005 Remaster	[ZZ Top]	[]	
0.012597	Good Times Bad Times - 1993 Remaster	[Led Zeppelin]	[classic rock', 'ro	
0.010412	Under Pressure - Remastered 2011	[Queen', 'David E	[rock', '80s', 'Pop	
0.010312	We Will Rock You - Remastered 2011	[Queen]	[classic rock', 'ro	
0.010021	Sweet Home Alabama	[Lynyrd Skynyrc	[classic rock', 'S	
0.009205	Radio Ga Ga - Remastered 2011	[Queen]	[classic rock', 'ro	
0.00911	Welcome To The Jungle	[Guns N' Roses]	[hard rock', 'rock	
0.008994	Smoke On The Water - Remastered 2012	[Deep Purple]	[hard rock', 'clas	
0.008837	Thunderstruck	[AC/DC]	[hard rock', 'rock	
0.008417	Paradise City	[Guns N' Roses]	[hard rock', 'rock	

Κεφάλαιο 8

Επίλογος

Κλείνοντας την εργασία, γίνεται η συγκεφαλαίωση των συμπερασμάτων που προέκυψαν μέσα από την διαδικασία εκπόνησης της, σε σχέση με τους στόχους που τέθηκαν στην εισαγωγή. Επιπροσθέτως, παρατίθενται κάποιες κατευθύνσεις και προβληματισμοί για τη μελλοντική έρευνα, γύρω από τις εφαρμογές που αναπτύχθηκαν και τις τεχνικές που χρησιμοποιήθηκαν.

8.1 Συμπεράσματα

Στη διπλωματική σχεδιάστηκαν και αναπτύχθηκαν διαφορετικές μεταξύ τους μέθοδοι μοντέλων συστάσεων περιεχομένου με συνεργατικό φιλτράρισμα. Όλες τους είχαν τη δυνατότητα να παράγουν ποιοτικές προτάσεις, που παρουσίαζαν μεγάλη συνάφεια με τα μουσικά κομμάτια που δέχονταν ως είσοδοι.

Όλοι οι Αυτοκωδικοποιητές που δοκιμάστηκαν, ως νευρωνικά δίκτυα βαθιάς μάθησης, αποδείχθηκε ότι έχουν τη δυνατότητα να αποτελέσουν τη βάση ενός συστήματος συστάσεων με συνεργατικό φιλτράρισμα, καθώς η επίδοσή τους είναι συγκρίσιμη, ή ακόμα και λίγο καλύτερη, με την αντίστοιχη της αρκετά διαδεδομένης τεχνικής παραγοντοποίησης πινάκων. Παρουσιάζουν, ωστόσο, κάποιες αδυναμίες όσον αφορά τον χρόνο εκπαίδευσης, αφού η σύγκλιση τους επιτυγχάνεται σε περισσότερες εποχές.

Για την εκπαίδευση των μοντέλων κατασκευάστηκαν εξαρχής δύο διαφορετικά σύνολα δεδομένων, με το πρώτο να είναι μικρότερο τόσο σε πλήθος μουσικών κομματιών, όσο και σε αναλογία κομματιών/λιστών αναπαραγωγής. Η εκπαίδευση των μοντέλων στο δεύτερο, μεγαλύτερο σύνολο οδήγησε σε σαφώς καλύτερα αποτελέσματα συγκριτικά με την εκπαίδευση στο μικρότερο. Ωστόσο, η αύξηση του μεγέθους είχε ως επακόλουθο την ανάδειξη προβλημάτων που σχετίζονται με τη διαχείριση μνήμης, τα οποία και αντιμετωπίστηκαν με επιτυχία.

Μέσα από τη διαδικασία της οπτικοποίησης, πραγματοποιήθηκε η ανάλυση της απεικόνισης των μουσικών κομματιών στον πολυδιάστατο χώρο των μοντέλων. Για τη τεχνική που βασίζεται στην παραγοντοποίηση πινάκων, η απεικόνιση αυτή πραγματοποιήθηκε μέσα από τη χρήση του πίνακα αντικειμένου-παραγόντων/χαρακτηριστικών, ενώ για τα δίκτυα των Αυτοκωδικοποιητών, αξιοποιήθηκε το διάγραμμα του κρυφού επιπέδου στην έξοδο των κωδικοποιητών. Στα βέλτιστα μοντέλα κάθε τεχνικής, ο χώρος που σχηματίστηκε ήταν ενιαίος, με μουσικά κομμάτια παρόμοιων μουσικών ειδών να τοποθετούνται κοντά μεταξύ τους.

Επιπλέον, πραγματοποιήθηκε ανάλυση σε κύριες συνιστώσες (PCA) των παραγόντων των μουσικών κομματιών και στη συνέχεια αναζητήθηκαν γραμμικές συσχετίσεις μεταξύ των συνιστωσών και των ποιοτικών κατηγοριών του Spotify (λ.χ. η ταχύτητα ή η διάθεση), χωρίς να εντοπιστούν ισχυρές γραμμικές σχέσεις μεταξύ τους. Επιπροσθέτως, έγινε προσπάθεια κατηγοριοποίησης των χαρακτηριστικών σε συστάδες, με μεικτά αποτελέσματα.

Τέλος, δημιουργήθηκαν δύο μικρές εφαρμογές. Ένα λειτουργικό σύστημα προτάσεων, στο οποίο εισάγεται ένα σύνολο μουσικών κομματιών και σχηματίζει εύστοχες προτάσεις, και μία εφαρμογή που επιτρέπει την παρέμβαση

στις κύριες συνιστώσες και, συνεπώς, την τροποποίηση των προτάσεων του εκάστοτε μοντέλου.

8.2 Μελλοντική Έρευνα

Κατά τη διάρκεια εκπόνησης της εργασίας αναδείχθηκε πληθώρα πεδίων και τεχνικών ζητημάτων, που άπτονται της παρούσας εργασίας αλλά ξεφεύγουν από τον σκοπό και την κλίμακα της. Παρακάτω παρουσιάζονται μερικά από αυτά.

Για το σύνολο δεδομένων, ένα πεδίο μελέτης αφορά την περαιτέρω αύξηση του μεγέθους του και την ανάπτυξη στρατηγικών χρήσης του, που να αντιμετωπίζουν τα προβλήματα διαχείρισης μνήμης.

Ιδιαίτερη αναφορά πρέπει να γίνει στο The Million Playlist Dataset, ένα σύνολο δεδομένων που χρησιμοποιήθηκε στο διαγωνισμό RecSys Challenge 2018 που συνδιοργάνωσε το Spotify, το πανεπιστήμιο της Μασαχουσέτης στο Amherst και το πανεπιστήμιο Johannes Kepler του Linz, το οποίο, ενώ προσφέρει την πληροφορία σε αξιοποίηση μορφή, εντούτοις δεν ήταν δημόσια διαθέσιμο την περίοδο κατασκευής του συνόλου δεδομένων που τελικά χρησιμοποιήθηκε.

Όσον αφορά τις τεχνικές που χρησιμοποιήθηκαν, υπάρχει δυνατότητα βελτιστοποίησης των υλοποιήσεων σε διάφορους τομείς, όπως η βελτίωση των χρόνων εκπαίδευσης, τα καλύτερα αποτελέσματα τους βάσει των μετρικών κ.α.. Σε αυτή την περίπτωση κινείται, για παράδειγμα, η ανάπτυξη της τεχνικής παραγοντοποίησης πινάκων απευθείας σε GPU, αξιοποιώντας εργαλεία όπως η πλατφόρμα CUDA, ενώ επιπροσθέτως μπορεί να διερευνηθεί η δυνατότητα χρήσης εναλλακτικών εφαρμογών με βάση τους Αυτοκωδικοποιητές Διακύμανσης.

Μία ακόμα κατεύθυνση με ιδιαίτερο ερευνητικό ενδιαφέρον αποτελεί η ανάπτυξη διαφορετικών μοντέλων βαθιάς μηχανικής μάθησης (όπως για παράδειγμα τα Generative Adversarial Networks ως εφαρμογές συνεργατικού φιλτραρίσματος για τη σύγκριση τους με τις παγιωμένες μεθόδους).

Βιβλιογραφία

- [1] David L. Poole, Alan K. Mackworth. *Artificial Intelligence, Foundations of Computational Agents*. 2010.
- [2] Stuart Russell and Peter Norvig. *Artificial Intelligence, A modern approach, 3rd edition*. 2010.
- [3] Zachary Susskind and Bryce Arden and Lizy K. John and Patrick Stockton and Eugene B. John. «Neuro-Symbolic AI: An Emerging Class of AI Workloads and their Characterization». στο: (2021).
- [4] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. 2006.
- [5] Simon Haykin. *Neural Networks and Learning Machines. 3rd Ed.* 2009.
- [6] Ian Goodfellow and Yoshua Bengio and Aaron Courville. *Deep Learning*. 2016.
- [7] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) (cs.LG).
- [8] B. Sarwar and G. Karypis and J. Konstan and J. Riedl. «Item-Based Collaborative Filtering Recommendation Algorithms». στο: *WWW10*. 2001, σσ. 285–295.
- [9] Su, Xiaoyuan and Khoshgoftaar, Taghi. «A Survey of Collaborative Filtering Techniques». στο: *Adv. Artificial Intelligence* 2009 (Οκτ. 2009). ΔΟΙ: [10.1155/2009/421425](https://doi.org/10.1155/2009/421425).
- [10] Charu C. Aggarwal. *Recommender Systems, The Textbook*. 2016.
- [11] Dacrema, Maurizio Ferrari and Cremonesi, Paolo and Jannach, Dietmar. «Are we really making much progress? A worrying analysis of recent neural recommendation approaches». στο: *Proceedings of the 13th ACM Conference on Recommender Systems* (2019). ΔΟΙ: [10.1145/3298689.3347058](https://doi.org/10.1145/3298689.3347058).
- [12] Hu Yifan and Koren Yehuda and Volinsky Chris. «Collaborative Filtering for Implicit Feedback Datasets». στο: *Δεχ.* 2008, σσ. 263–272.
- [13] Radford M. Neal. *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Technical Report CRG-TR-93-1. Department of Computer Science, University of Toronto.
- [14] Diederik P Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2013. arXiv: [1312.6114](https://arxiv.org/abs/1312.6114).
- [15] Kingma, Diederik P. and Welling, Max. «An Introduction to Variational Autoencoders». στο: *Foundations and Trends® in Machine Learning* 12.4 (2019), σσ. 307–392. ISSN: 1935-8245. ΔΟΙ: [10.1561/22000000056](https://doi.org/10.1561/22000000056).
- [16] Dawen Liang and Rahul G. Krishnan and Matthew D. Hoffman and Tony Jebara. *Variational Autoencoders for Collaborative Filtering*. 2018. arXiv: [1802.05814](https://arxiv.org/abs/1802.05814).