Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Σύνθεση μουσικής με δυνατότητα επιλογής μουσικού είδους

Μελέτη και υλοποίηση

Διπλωματικη Εργασια

του

ΜΙΧΑΗΛ Γ. ΜΕΓΓΙΣΟΓΛΟΥ

Επιβλέπων: Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π

Αθήνα, Νοέμβριος 2020

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Σύνθεση μουσικής με δυνατότητα επιλογής μουσικού είδους

Μελέτη και υλοποίηση

Διπλωματικη Εργασια

του

ΜΙΧΑΗΛ Γ. ΜΕΓΓΙΣΟΓΛΟΥ

Επιβλέπων: Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12 Νοεμβρίου 2020.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Ανδρέας-Γεώργιος Σταφυλοπάτης	Γεώργιος Στάμου	Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π	Αν. Καθηγητής Ε.Μ.Π	Καθηγητής Ε.Μ.Π

Αθήνα, Νοέμβριος 2020

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....
Μιχαήλ Μεγγίσογλου

10 Νοεμβρίου 2020

Περίληψη

Η παραγωγή δεδομένων από νευρωνικά δίκτυα που στηρίζονται σε γεννητικά μοντέλα έχει γνωρίσει μεγάλη ανάπτυξη τα τελευταία χρόνια. Τα δίκτυα αυτά εκπαιδεύονται σε σύνολα δεδομένων και κατασκευάζουν κατανομές πιθανοτήτων που τα περιγράφουν. Με αυτόν τον τρόπο δύνανται να ανακατασκευάσουν στιγμιότυπα του συνόλου εκπαίδευσης ή να παράξουν νέα δεδομένα, παρόμοια με όσα έχουν συναντήσει κατά την εκπαίδευσή τους. Η παρούσα εργασία χρησιμοποιεί μια παραλλαγή του Variational αυτοκωδικοποιητή (VAE), ενός γεννητικού νευρωνικού δικτύου, και έχει σκοπό τη σύνθεση μουσικής ενός συγκεκριμένου μουσικού είδους, το οποίο παρέχει ο χρήστης. Με βάση το σύστημα που κατασκευάσαμε, εξερευνούμε τις διάφορες πτυχές του VAE και εκμεταλλευόμαστε την εσωτερική του δομή, για να εξάγουμε ενδιαφέροντα συμπεράσματα.

Λέξεις Κλειδιά

Τεχνητή Νοημοσύνη, Νευρωνικά Δίκτυα, Σύνθεση Μουσικής, Επιβλεπόμενη Μάθηση, Μη Επιβλεπόμενη Μάθηση, LSTM, GRU, Variational Inference

Abstract

Generating data from neural networks that are based on genetic models has experienced great growth over the last years. These genetic networks construct probability distributions that describe the datasets, upon which they were trained. Using the probability ditributions, they can reconstruct instances of the training set or produce new data, similar to what they have encountered during their training. The present study uses a variant of the Variational Autoencoder (VAE), a genetic neural network, in order to compose music of a specific genre, which is provided by the user. Based on the system we built, we explore the various aspects of the VAE and take advantage of its internal structure to draw interesting conclusions.

Keywords

Artificial Intelligence, Neural Networks, Music Composition, Supervised Learning, Unsupervised Learning, LSTM, GRU, Variational Inference

στους γονείς μου

Ευχαριστίες

Η εκπόνηση της Διπλωματικής μου εργασίας σηματοδοτεί το τέλος των προπτυχιακών μου σπουδών στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. Η Διπλωματική μου εργασία πραγματοποιήθηκε στα πλαίσια του Εργαστηρίου Ευφυών και Υπολογιστικών Συστημάτων του τομέα Τεχνολογίας Πληροφορικής και Υπολογιστών, με επιβλέποντα καθηγητή τον κ. Ανδρέα Σταφυλοπάτη, τον οποίο θα ήθελα αρχικά να ευχαριστήσω για την ευκαιρία που μου έδωσε να ασχοληθώ με το συγκεκριμένο επιστημονικό πεδίο. Ιδιαιτέρως θα ήθελα να ευχαριστήσω τον δρ Γεώργιο Σιόλα και τον Έντι Ντερβάκο οι οποίοι μου έδωσαν από την αρχή τις κατάλληλες συμβουλές και κατευθύνσεις, προκειμένου να έχουμε το επιθυμητό αποτέλεσμα στο παρόν έργο. Θα ήθελα επίσης να ευχαριστήσω τους καθηγητές Γεώργιο Στάμου και Στέφανο Κόλλια που συμμετείχαν στην τριμελή εξεταστική επιτροπή της διπλωματικής μου εργασίας. Τέλος, θα ήθελα να ευχαριστήσω βαθύτατα την οικογένειά μου και τους φίλους μου για τη στήριξη που μου έχουν προσφέρει όλα αυτά τα χρόνια.

Αθήνα, Νοέμβριος 2020

Μιχαήλ Μεγγίσογλου

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

1.1 Αντικείμενο της διπλωματικής

Η εργασία μας έχει ως αντικείμενο την κατασκευή ενός γεννητικού συστήματος νευρωνικών δικτύων με απώτερο σκοπό την παραγωγή μουσικής ενός συγκεκριμένου μουσικού είδους που καθορίζεται από το χρήστη. Προς αυτό το σκοπό, μελετάμε διάφορα ζητήματα που προκύπτουν τόσο στη θεωρία όσο και στην υλοποίηση.

Καλούμαστε αρχικά να βρούμε μια έξυπνη και αποδοτική αναπαράσταση της μουσικής, κατάλληλη για επεξεργασία από το σύστημα. Η αναπαράσταση της πληροφορίας σε ένα σύστημα μηχανικής μάθησης παίζει καθοριστικό ρόλο στην επίτευξη υψηλής απόδοσης. Στη συνέχεια, κατασκευάζουμε το κυρίως σύστημα βάσει του οποίου γίνεται η σύνθεση της μουσικής. Το σύστημα αυτό πρέπει να είναι γεννητικό, να έχει δηλαδή την ικανότητα να παράγει νέα δεδομένα, παρόμοια με όσα έχει επεξεργαστεί κατά την εκπαίδευση. Αυτό το σύστημα, που είναι ένας Variational αυτοκωδικοποιητής, μαθαίνει την αναπαράσταση των μουσικών κομματιών με τη μορφή κατανομών πιθανοτήτων. Εκμεταλλευόμαστε αυτήν την αναπαράσταση στα πειράματά μας τόσο για τη σύνθεση μουσικής, όσο και για τη μελέτη των ιδιοτήτων και των διάφορων πτυχών της πληροφορίας που ενθυλακώνεται στο VAE.

1.2 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε οκτώ κεφάλαια. Στο Κεφάλαιο 2 δίνουμε εν συντομία κάποια θεωρητικά στοιχεία σχετικά με τη μουσική. Στο Κεφάλαιο 3 παρουσιάζουμε κάποιες θεωρητικές έννοιες που αφορούν τη Μηχανική Μάθηση ως επιστημονικό κλάδο. Στο Κεφάλαιο 4 αναλύουμε διεξοδικά όλα τα χαρακτηριστικά των νευρωνικών δικτύων που χρησιμοποιούμε στην εργασία. Αρχίζουμε με τα δίκτυα πρόσθιας τροφοδότησης, τα αναδρομικά νευρωνικά δίκτυα και τους αυτοκωδικοποιητές και ολοκληρώνουμε το κεφάλαιο με μια μνεία στις συναρτήσεις ενεργοποίησης και στην εκπαίδευση ενός νευρωνικού δικτύου. Στο Κεφάλαιο 5 παρουσιάζουμε τα σύνολα δεδομένων που χρησιμοποιούμε και τις λεπτομέρειες της αναπαράστασής τους σε μορφή κατάλληλη για επεξεργασία από τα νευρωνικά δίκτυα. Στο Κεφάλαιο 6 μελετάμε το σύστημα που υλοποιήσαμε στα πλαίσια αυτής της εργασίας. Πιο συγκεκριμένα, αναλύουμε σε βάθος τις συνιστώσες του και τον τρόπο λειτουργίας του. Τα αποτελέσματα των πειραμάτων μας, που ως επί το πλείστον είναι μουσικές συνθέσεις καθώς και οι μετρικές αξιολόγησης του συστήματός μας περιγράφονται στο Κεφάλαιο 7. Τέλος, στο

Κεφάλαιο 8 συνοψίζουμε το υλικό που αναπτύξαμε σε αυτήν τη διπλωματική και δίνουμε κάποιες ενδιαφέρουσες μελλοντικές επεκτάσεις του συστήματος που υλοποιήσαμε.

Μέρος I

Θεωρητικό Μέρος

Κεφάλαιο 2

Στοιχεία Μουσικής Θεωρίας

Στο κεφάλαιο αυτό παρουσιάζουμε θεωρητικές έννοιες της μουσικές που θα βοηθήσουν τον αναγνώστη να κατανοήσει καλύτερα τα αποτελέσματα αυτής της εργασίας. Θα παρουσιάσουμε μια ιστορική αναδρομή της αλγοριθμικής σύνθεσης της μουσικής, όπως αυτή προσεγγίστηκε από τους πρωτοπόρους στον τομέα και έπειτα θα δώσουμε μια σύντομη περιγραφή των βασικών μουσικοθεωρητικών εννοιών, που απαιτούνται για την πλήρη κατανόηση του περιεχομένου της παρούσας εργασίας.

2.1 Ιστορική Αναδρομή της Αλγοριθμικής Σύνθεσης

Η ιστορία της θεμελίωσης ενός συστηματικού τρόπου περιγραφής των μουσικών εννοιών και άρα ενός τρόπου αυτοματοποίησης της διαδικασίας της σύνθεσης ξεκινά στην αρχαία Ελλάδα. Οι μεγάλοι φιλόσοφοι και μαθηματικοί της αρχαιότητας, όπως ο Πυθαγόρας, ο Πτολεμαίος και ο Πλάτων προσπάθησαν να δημιουργήσουν ένα μαθηματικό σύστημα θεωρίας με σκοπό να εξηγήσουν τα μουσικά φαινόμενα. Είναι γνωστή εξάλλου στην αρχαία Ελλάδα η αλληλένδετη σχέση μεταξύ αριθμητικής και μουσικής. Όμως, οι φορμαλισμοί που δόθηκαν από τους αρχαίους Έλληνες αφορούσαν αμιγώς την εγκαθίδρυση ενός θεωρητικού υποβάθρου, η εφαρμογή του οποίου στην πράξη είναι αμφισβητούμενη, αφού το μεγαλύτερο μέρος μιας τότε μουσικής παράστασης στηριζόταν στον αυτοσχεδιασμό. Έτσι, δεν μπορούμε να χαρακτηρίσουμε τη μουσική της αρχαίας Ελλάδας ως αλγοριθμική με την αυστηρή έννοια, παρ' όλα αυτά η συνεισφορά των Ελλήνων είναι ιστορικής σημασίας για τη μετέπειτα ανάπτυξη πιο προηγμένων διαδικασιών αυτοματοποιημένης σύνθεσης.

Μια εξέλιξη προς τη συστηματικοποίηση της μουσικής έγινε στα τέλη του 15ου αιώνα με την εισαγωγή του μουσικού κανόνα. Σε αυτό το πλαίσιο, ο συνθέτης δημιουργεί έναν μουσικό πυρήνα - μια μουσική φράση ή ένα τμήμα - και παύει να παρεμβαίνει. Το μουσικό έργο έπειτα συγκροτείται από τους μουσικούς κατά την εκτέλεση με τη μέθοδο της μίμησης, δηλαδή τη διαδικασία επανάληψης μιας φράσης με χρονική καθυστέρηση.

Αξιοσημείωτη είναι η συνεισφορά του Wolfgang Amadeus Mozart (1756-1791) στην αλγοριθμική παραγωγή μουσικής. Ο μεγάλος μουσικός χρησιμοποίησε τεχνικές αυτοματοποίησης της σύνθεσης στο έργο του 'Musikalisches Wurfelspiel' (Μουσική Ζαριών). Σε αυτό το έργο συγκέντρωσε πολλά μικρά μουσικά τμήματα, τα οποία αν συνδεθούν τυχαία μεταξύ τους, σύμφωνα με τη ρίψη ενός ζαριού, δημιουργούν ένα πλήρες μουσικό κομμάτι. Αυτή είναι ίσως η πρώτη φορά στην ιστορία όπου η σύνθεση της μουσικής γίνεται τυχαιοκρατικά

και χωρίς την παρουσία του μουσικοσυνθέτη - πέραν της αρχικής του μόνο παρέμβασης.

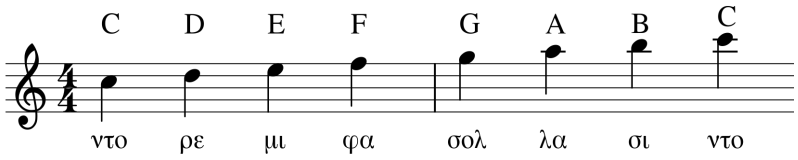
Το επόμενο ορόσημο στην ιστορία της αλγοριθμικής σύνθεσης είναι η καθιέρωση του ηλεκτρονικού υπολογιστή ως μέσο εκτέλεσης υπολογισμών. Πρωτοπόροι στη χρήση του υπολογιστή για τη σύνθεση μουσικής ήταν οι Lejaren Hiller και Robert Baker που δημιούργησαν το πρόγραμμα 'MUSICOMP'. Αυτό το πρόγραμμα στηρίζεται σε ένα σύνολο κανόνων και υπορουτίνων που μπορεί να επικαλείται ο προγραμματιστής, για να καθοδηγήσει τη σύνθεση σύμφωνα με τις δικές του προτιμήσεις. Το 'MUSICOMP' συνέθεσε στα τέλη της δεκαετίας του 1950 το μουσικό έργο 'Computer Cantata'.

Άλλος ένας πρωτοπόρος στη χρήση του ηλεκτρονικού υπολογιστή ήταν ο Ιάννης Ξενάκης. Το πρόγραμμα που κατασκεύασε - τις αρχές του οποίου εξηγεί λεπτομερώς στο [1] - χρησιμοποιούσε πυκνότητες πιθανοτήτων και άλλα στοχαστικά εργαλεία, για να συμπεράνει τους ήχους που έπρεπε να ακούγονται σε κάθε στιγμή του μουσικού κομματιού. Η μέθοδος του Ξενάκη, όμως, έχει κύριο στόχο να βοηθήσει τον συνθέτη στη σύνθεση κι όχι να παράξει μουσικά έργα με αυτόνομο τρόπο.

Τέλος, τα τελευταία χρόνια όλο και περισσότερο γίνεται χρήση τεχνικών μηχανικής μάθησης και νευρωνικών δικτύων, που έχουν ως σκοπό τη σύνθεση μουσικής. Μεγάλο ενδιαφέρον παρουσιάζει η δουλειά των [2] και [3], όπου γίνεται μια εκτενής προσπάθεια αναπαράστασης του λεξιλογίου της μουσικής με έναν τρόπο, κατάλληλο για επεξεργασία από τα νευρωνικά συστήματα. Παράλληλα τίθενται οι βάσεις για την αλγοριθμική παραγωγή μουσικής με χρήση νευρωνικών δικτύων, πρακτική η οποία σχετίζεται σε μεγάλο βαθμό με το αντικείμενο της παρούσας εργασίας.

2.2 Θεωρία της Μουσικής

Η μουσική, όπως αναπτύχθηκε στον Δυτικό κόσμο, αποτελεί μια πλούσια γλώσσα, ικανή να εκφράσει μια μεγάλη ποικιλία συναισθημάτων. Στο μεγαλύτερο μέρος της στηρίζεται σε απλές δομικές μονάδες, οι οποίες μπορούν εύκολα να συνδυαστούν, για να δώσουν τα περίπλοκα ηχητικά αποτελέσματα που επιδιώκουν να πετύχουν οι συνθέτες. Σε αυτήν την παράγραφο παρουσιάζουμε τις βασικές δομές της μουσικής που είναι απαραίτητες για την περαιτέρω κατανόηση του περιεχομένου της εργασίας.



Σχήμα 2.1: *Νότες στο πεντάγραμμο*

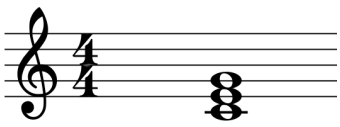
Οι θεμελιώδεις μονάδες της μουσικής είναι οι νότες, οι οποίες τοποθετούνται πάνω στο πεντάγραμμο, όπως φαίνεται για παράδειγμα στο σχήμα 2.1. Η θέση μιας νότας στο πεντάγραμμο προσδιορίζει μονοσήμαντα τον τόνο της, ενώ το σχήμα του συμβόλου προσδιορίζει την αξία της, δηλαδή τη χρονική διάρκεια κατά την οποία η νότα ηχεί. Στον πίνακα 2.1 παρουσιάζουμε τα πιο συνηθισμένα μουσικά σύμβολα ανάλογα με την αξία τους. Η ονομασία του τόνου μιας νότας μπορεί να ακολουθεί ένα από τα δύο καθιερωμένα συστήματα.

Στην Ευρώπη συνήθως χρησιμοποιείται η σημειολογία ντο, ρε, μι φα, σολ, λα, σι, ενώ στην Αμερική χρησιμοποιούνται αντίστοιχα τα γράμματα A, B, C, D, E, F, G, με τον τόνο A να αντιστοιχεί στο λα. Σε κάθε περίπτωση ένας τόνος μπορεί να εμφανίζει μια αλλοίωση με τη μορφή ύφεσης (♭) ή δίεσης (♯).

Νότα	Ονομασία
𝅝	Ολόκληρο
𝅗𝅥	Μισό
♩	Τέταρτο
♪	Όγδοο
𝅘𝅥𝅯	Δέκατο Έκτο
♩.	Τέταρτο Παρεστιγμένο

Πίνακας 2.1: *Συνήθεις χρονικές αξίες στη μουσική*

Όταν δύο νότες εμφανίζουν επικάλυψη κατά την ήχησή τους, τότε λέμε ότι συνηχούν. Σε αυτήν την περίπτωση έχουμε πολυφωνία και η σημειολογία φαίνεται στο σχήμα 2.2. Εάν πιο συγκεκριμένα, συνηχούν τρεις ή περισσότερες νότες, τότε η δομή που δημιουργείται ονομάζεται συγχορδία. Η συγχορδία παίζει καθοριστικό ρόλο στον τομέα της μουσικής, αφού η ταυτόχρονη ήχηση διαφορετικών τόνων, δίνει διαφορετικά ηχητικά αποτελέσματα και επιτρέπει στο συνθέτη να δώσει στο έργο του το ύφος που επιθυμεί.



Σχήμα 2.2: *Η συγχορδία ντο μείζονα*

Τέλος, πάντα στην αρχή κάθε μουσικού κομματιού - ενδεχομένως και σε άλλα σημεία του ανάλογα με τις ανάγκες - υπάρχη η ρυθμική σφραγίδα. Τόσο το σχήμα 2.1 όσο και το σχήμα 2.2 είναι γραμμένα σε ρυθμό τεσσάρων τετάρτων, όπως φαίνεται από την αντίστοιχη επιγραφή. Η μουσική σφραγίδα καθορίζει τη συνολική χρονική αξία που πρέπει να υπάρχει μέσα σε ένα μουσικό μέτρο και ουσιαστικά ρυθμίζει τη ροή του κομματιού. Δεν είναι λίγα τα μουσικά είδη εξάλλου που έχουν διαμορφωθεί εξ ολοκλήρου βάσει μιας ρυθμικής σφραγίδας - όπως για παράδειγμα το βαλς.

Κεφάλαιο 3

Μηχανική Μάθηση

Η μηχανική μάθηση είναι ένα πεδίο της Επιστήμης της Τεχνητής Νοημοσύνης που ασχολείται με την αναγνώριση προτύπων και μοτίβων σε σύνολα δεδομένων με σκοπό τη δημιουργία αναπαραστάσεων ικανών να προβλέψουν νέα δεδομένα. Ο ορισμός που πρότεινε ο Tom M. Mitchell για τη Μηχανική Μάθηση είναι ο ακόλουθος: 'Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από μια εμπειρία Ε ως προς μια κλάση εργασιών Τ και ένα μέτρο επίδοσης Ρ, αν η επίδοση του σε εργασίες της κλάσης Τ, όπως αποτιμάται από το μέτρο Ρ, βελτιώνεται με την εμπειρία Ε' [4]. Υπό αυτή τη σκοπιά, η μηχανική μάθηση και οι εφαρμογές της δεν στηρίζονται σε ένα σαφώς καθορισμένο σύνολο εντολών που έχει δοθεί από έναν προγραμματιστή, αλλά βασίζονται στην σταδιακή εύρεση μιας αναπαράστασης που εξάγεται από εμπειρικά παραδείγματα, τέτοιας που να βελτιστοποιεί την επίδοση ως προς το δοθέν κριτήριο. Είναι επομένως φυσικό να αναμένουμε από τα εν λόγω συστήματα, αποτελέσματα που ικανοποιούν στατιστικά κριτήρια και όχι ακριβείς απαντήσεις, όπως αυτές που θα έδινε ένα τυπικό πρόγραμμα υπολογιστή. Εξάλλου η φύση της διαδικασίας είναι επαγωγική. Εισάγουμε δείγματα στα συστήματα και απαιτούμε από αυτά να βρουν τις χαρακτηριστικές τους ιδιότητές και να γενικεύσουν, να παράξουν, δηλαδή, ικανοποιητική έξοδο ακόμα και για πρωτοφανείς εισόδους.

Η μηχανική μάθηση ως εργαλείο επίλυσης προβλημάτων, κατέχει μια μοναδική θέση. Συγκεκριμένα, ο κλάδος αυτός δίνει απαντήσεις - έστω και προσεγγιστικές - σε προβλήματα που δύσκολα μπορούμε να εκφράσουμε με αυστηρούς μαθηματικούς όρους, αλλά για τα οποία διαθέτουμε πλούσια εμπειρία υπό τη μορφή δειγμάτων εισόδου-εξόδου. Δύνανται επίσης τα προγράμματα μηχανικής μάθησης να εντοπίζουν στατιστικές συσχετίσεις ανάμεσα σε μεγάλους όγκους δεδομένων, οι οποίες δεν είναι προφανείς στον άνθρωπο ούτε έπειτα από ενδελεχή έρευνα.

Σε κάθε πρόβλημα μηχανικής μάθησης υπάρχει το σύνολο δεδομένων, η εμπειρία δηλαδή, πάνω στην οποία στηρίζεται το πρόγραμμα για να βελτιστοποιήσει την επίδοσή του. Ενώ ένα πρόγραμμα μηχανικής μάθησης έχει συγκεκριμένες προδιαγραφές για το είδος των δεδομένων που μπορεί να επεξεργαστεί, εντούτοις με κατάλληλους μετασχηματισμούς μας δίνεται μεγάλη ελευθερία ως προς το είδος των δεδομένων που μπορούμε να δώσουμε σε αυτά τα συστήματα για επεξεργασία. Στις επόμενες παραγράφους εξετάζουμε τις διάφορες κατηγορίες ειδών δεδομένων και τρόπων μάθησης που διαθέτουμε.

3.1 Μέθοδοι Μηχανικής Μάθησης

Η πρώτη κατηγοριοποίηση που παραθέτουμε αφορά στον τρόπο εκμετάλλευσης της διαθέσιμης εμπειρίας, βάσει του προβλήματος εκμάθησης που θέλουμε να λύσουμε, αλλά και των διαθέσιμων δεδομένων. Στην παρούσα εργασία χρησιμοποιούμε δύο είδη μηχανικής μάθησης, τα οποία αναλύουμε συνοπτικά παρακάτω. Για μια πιο λεπτομερή ανάλυση, παραπέμπουμε τον αναγνώστη στο [5].

Επιβλεπόμενη Μάθηση Σε αυτήν την κατηγορία, η εμπειρία πρέπει να είναι διαθέσιμη στη μορφή ζευγών εισόδου-εξόδου. Το πρόγραμμα μηχανικής μάθησης χρησιμοποιεί τα εν λόγω ζεύγη με σκοπό να δημιουργήσει ένα μαθηματικό μοντέλο, ικανό να παράγει την προκαθορισμένη έξοδο, όταν διεγείρεται από την αντίστοιχη είσοδο. Η διαδικασία επιτυγχάνεται με τη βελτιστοποίηση μιας συνάρτησης κόστους, που συσχετίζει την έξοδο του μαθηματικού μοντέλου και την επιθυμητή έξοδο όπως επιβάλλεται από την εμπειρία. Εν τέλει, το μαθηματικό μοντέλο πρέπει να είναι ικανό να παράγει κατάλληλες εξόδους - όπως αυτές καθορίζονται από τη συνάρτηση κόστους - τόσο για γνωστές, όσο και για άγνωστες εισόδους.

Μη Επιβλεπόμενη Μάθηση Σε αυτήν την κατηγορία, η εμπειρία χρειάζεται να έχει μόνο εισόδους. Το πρόγραμμα μηχανικής μάθησης λαμβάνει τις εισόδους και προσπαθεί να εξάγει εγγενή χαρακτηριστικά της εμπειρίας με σκοπό να τα ομαδοποιήσει, να τα ταξινομήσει κλπ. Στη μη επιβλεπόμενη μάθηση επομένως, δεν χρειάζεται η εκ των προτέρων επεξεργασία των δεδομένων με σκοπό την τεκμηρίωση κατάλληλων εξόδων - μια διαδικασία ιδιαίτερα χρονοβόρα και δύσκολα αυτοματοποιήσιμη. Αντίθετα, το σύστημα εντοπίζει πρότυπα και μοτίβα στα δεδομένα, ουσιαστικά μοντελοποιεί δηλαδή την υποβόσκουσα κατανομή τους.

3.2 Τύποι Δεδομένων

Άλλη μία κατηγοριοποίηση που οφείλουμε να παραθέσουμε, έχει να κάνει με τον τύπο των διαθέσιμων δεδομένων. Συγκεκριμένα διακρίνουμε δύο κατηγορίες, τα αριθμητικά και τα κατηγορικά δεδομένα. Ανάλογα με την κατηγορία με την οποία εργαζόμαστε κάθε φορά και φυσικά βάσει των περιορισμών που θέτει το σύστημα μηχανικής μάθησης, ενδέχεται να χρειαστούν μετατροπές των δεδομένων σε άλλη μορφή, πριν αυτά δοθούν ως είσοδο στο σύστημα ή αφού λάβουμε τη συστημική έξοδο.

Αριθμητικά Δεδομένα Σε αυτήν την κατηγορία ανήκουν δεδομένα, τα οποία εκφράζονται σε όρους αριθμών χωρίς κάποιο γνωστό περιορισμό. Τα δεδομένα αυτά, δηλαδή, μπορεί να εκφράζονται από ακέραιους αριθμούς, οπότε έχουμε διακριτά αριθμητικά δεδομένα, ή μπορεί να εκφράζονται από πραγματικούς αριθμούς, οπότε έχουμε συνεχή αριθμητικά δεδομένα. Συνήθως, τα αριθμητικά δεδομένα προκύπτουν από μετρήσεις, όπως για παράδειγμα το ύψος ή το βάρος ενός ανθρώπου, η αξία ενός σπιτιού κλπ. Σημειώνουμε, εδώ, ότι τα αριθμητικά δεδομένα μπορούν να δοθούν αυτούσια προς επεξεργασία στα συστήματα μηχανικής μάθησης, αν και συνήθως υπόκεινται σε κάποια διαδικασία προεπεξεργασίας, με σκοπό την κανονικοποίησή τους.

Κατηγορικά Δεδομένα Σε αυτήν την κατηγορία ανήκουν δεδομένα, των οποίων οι διαφορετικές τιμές είναι πεπερασμένες. Το σύνολο των διαφορετικών τιμών που μπορεί να πάρει ένα κατηγορικό σύνολο δεδομένων ονομάζεται λεξιλόγιο. Συνήθως αντιστοιχίζουμε κάθε μέλος του λεξιλογίου σε έναν ακέραιο αριθμό, για να πάρουμε μια μορφή δεδομένων κατάλληλη για επεξεργασία από το σύστημα μηχανικής μάθησης - έναντι για παράδειγμα μιας συμβολοσειράς ή μιας αφηρημένης έννοιας. Κατηγορικά δεδομένα είναι συνήθως ποιοτικά χαρακτηριστικά όπως το γένος ενός ανθρώπου, η αξιολόγηση σε συγκεκριμένη κλίμακα (πχ 1 έως 5), οι λέξεις μιας φυσικής γλώσσας κλπ. Στην παρούσα εργασία αντιμετωπίζουμε αποκλειστικά τέτοιου είδους δεδομένα. Πιο συγκεκριμένα, όπως θα αναλύσουμε στο κεφάλαιο 5, μοντελοποιούμε τις μουσικές νότες και τις χρονικές διάρκειες των δεδομένων μας ως μέλη ενός συνόλου δυνατών τιμών - το σύνολο των διαφορετικών νοτών και των επιτρεπτών χρονικών διαρκειών.

Κεφάλαιο 4

Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα είναι υπολογιστικά μοντέλα εμπνευσμένα από τα βιολογικά νευρωνικά δίκτυα των έμβιων οργανισμών. Πρόκειται για μη γραμμικά συστήματα τα οποία είναι της μορφής $\boldsymbol{y} = f(\boldsymbol{x};\boldsymbol{\Theta})$, όπου $\boldsymbol{x}$ και $\boldsymbol{y}$ είναι αντίστοιχα η είσοδος και η έξοδος του νευρωνικού δικτύου, f είναι η συνάρτηση αντιστοίχησης και $\boldsymbol{\Theta}$ οι παράμετροι του συστήματος. Τα νευρωνικά δίκτυα είναι εκπαιδεύσιμα, υπό την έννοια ότι οι παράμετροι $\boldsymbol{\Theta}$ μεταβάλλονται, αλλάζοντας τη συστημική έξοδο $\boldsymbol{y}$, με απώτερο σκοπό την ελαχιστοποίηση κάποιας συνάρτησης κόστους. Με αυτόν τον τρόπο πετυχαίνουν τη βελτίωση της επίδοσής τους στην εργασία που τους έχει ανατεθεί και έτσι ικανοποιούν τον ορισμό της μηχανικής μάθησης που δόθηκε στο προηγούμενο κεφάλαιο.

Σημαντική για την επιτυχή λειτουργία ενός νευρωνικού δικτύου είναι η αρχιτεκτονική του, δηλαδή η εσωτερική διαμόρφωση και διασύνδεση των συνιστωσών του. Η αρχιτεκτονική ενός νευρωνικού δικτύου καθορίζει την ικανότητά του να μαθαίνει ισχυρές αναπαραστάσεις της εισόδου και ως εκ τούτου να προβλέπει και να γενικεύει πιο αποδοτικά σε άγνωστα δεδομένα. Τις περισσότερες φορές η φύση των δεδομένων εισόδου επιβάλλει τη χρήση μιας συγκεκριμένης αρχιτεκτονικής, η οποία έχει φανεί να ανταποκρίνεται καλύτερα από άλλες σε τέτοιου είδους δεδομένα. Γι' αυτό είναι χρήσιμο κανείς να είναι εξοικειωμένος με την πληθώρα αρχιτεκτονικών που έχουν αναπτυχθεί. Παρακάτω θα παρουσιάσουμε όλα τα αρχιτεκτονικά στοιχεία που είναι απαραίτητα για τις ανάγκες της παρούσας εργασίας.

4.1 Νευρωνικά Δίκτυα Πρόσθιας Τροφοδότησης

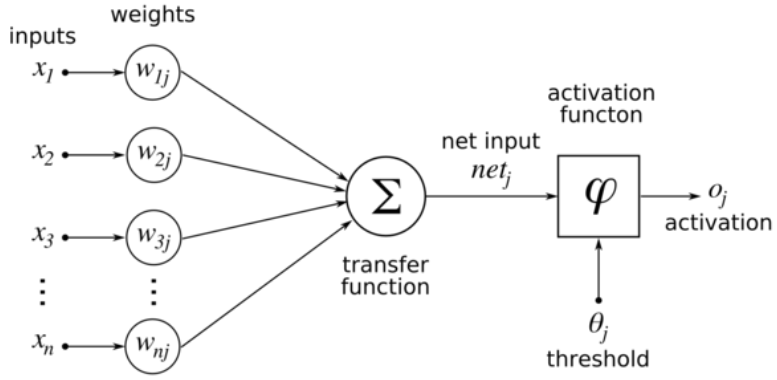
Τα νευρωνικά δίκτυα πρόσθιας τροφοδότησης είναι μαθηματικά μοντέλα που αντιστοιχίζουν την είσοδο τους σε κάποια έξοδο χωρίς να υπάρχει ανατροφοδότηση. Η είσοδος σε ένα τέτοιο νευρωνικό δίκτυο δεν μπορεί να είναι χρονοσειρά. Τα δίκτυα αυτά είναι απλά στη λειτουργία τους και χρησιμοποιούνται συνήθως για την εύρεση ενός κατάλληλου μετασχηματισμού των δεδομένων ή για τη μείωση της διαστατικότητάς τους.

4.1.1 Ο απλός τεχνητός νευρώνας

Ο απλός τεχνητός νευρώνας αποτελεί τη θεμελιώδη μονάδα των νευρωνικών δικτύων. Στο Σχήμα 4.1 απεικονίζεται μια τέτοια υπολογιστική μονάδα. Η βασική επεξεργασία έχει ως εξής: Κάθε συνιστώσα, x_i της εισόδου $\boldsymbol{x}$ πολλαπλασιάζεται με μια συνιστώσα βάρους, w_i.

Στο αποτέλεσμα προστίθεται ένας σταθερός όρος, b, ο οποίος ονομάζεται πόλωση. Τέλος, το συνολικό άθροισμα εισάγεται σε μια συνάρτηση, συνήθως μη γραμμική, που ονομάζεται συνάρτηση ενεργοποίησης και το αποτέλεσμα της οποίας αποτελεί την έξοδο του νευρώνα. Η έξοδος του νευρωνικού δικτύου δίδεται επομένως από την ακόλουθη εξίσωση

$$y = f(\boldsymbol{x}^\intercal \boldsymbol{w} + b) \tag{4.1}$$



Σχήμα 4.1: *Ο απλός τεχνητός νευρώνας*

4.1.2 Πολυεπίπεδα Δίκτυα Απλών Νευρώνων

Τα πολυεπίπεδα δίκτυα απλών νευρώνων είναι δίκτυα, στα οποία περισσότεροι του ενός απλοί νευρώνες συνδέονται μεταξύ τους με οργανωμένο τρόπο. Πιο συγκεκριμένα, όπως φαίνεται στο Σχήμα 4.2, οι απλοί νευρώνες οργανώνονται σε επίπεδα και συνδέονται μεταξύ τους με τον εξής απλό τρόπο. Η έξοδος ενός νευρώνα που βρίσκεται σε κάποιο επίπεδο γίνεται είσοδος σε κάθε νευρώνα του αμέσως επόμενου επιπέδου.
Διαχωρίζουμε τα επίπεδα σε τρεις κατηγορίες.

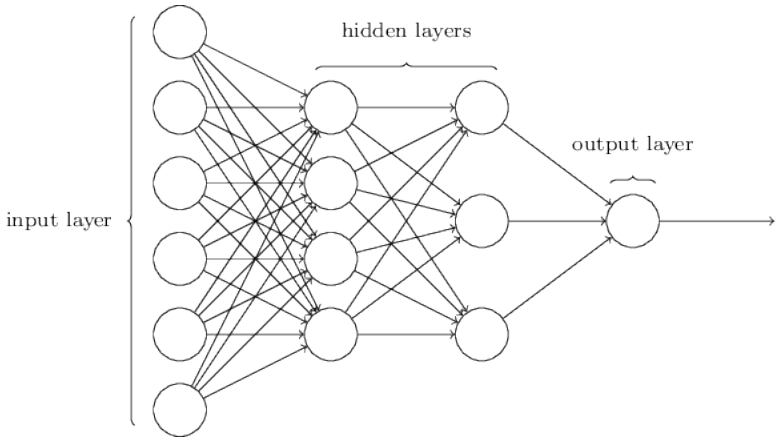
- Επίπεδο εισόδου. Οι νευρώνες αυτού του επιπέδου λαμβάνουν τις εισόδους τους απευθείας από τα δεδομένα εισόδου. Είναι υποχρεωτικό το πλήθος των συνιστωσών των δειγμάτων εισόδου να ισούται με το πλήθος των νευρώνων του επιπέδου αυτού.

- Επίπεδο εξόδου. Η έξοδος που λαμβάνεται από τους νευρώνες αυτού του επιπέδου αποτελεί την συνολική έξοδο του νευρωνικού δικτύου. Είναι υποχρεωτικό το πλήθος των νευρώνων του επιπέδου αυτού να ισούται με το πλήθος των συνιστωσών της επιθυμητής εξόδου.

- Κρυφό επίπεδο. Είναι το επίπεδο μεταξύ των επιπέδων εισόδου και εξόδου. Στην πράξη τα νευρωνικά δίκτυα περιλαμβάνουν περισσότερα του ενός κρυφά επίπεδα - σε αυτήν την περίπτωση τα νευρωνικά δίκτυα καλούνται βαθιά. Δεν υπάρχει κάποιος περιορισμός ως προς το πλήθος νευρώνων που υπάρχουν σε αυτό το επίπεδο, ούτε ως προς το συνολικό πλήθος των επιπέδων αυτών.

Συνοπτικά, για ένα δίκτυο l επιπέδων ισχύει

$$\boldsymbol{o}^{(i)} = g^{(i)}\left(\boldsymbol{W}^{(i)}\boldsymbol{o}^{(i-1)} + \boldsymbol{b}^{(i)}\right) \tag{4.2}$$



Σχήμα 4.2: *Πολυεπίπεδο νευρωνικό δίκτυο*

όπου $\boldsymbol{o}^{(i)}$ είναι η έξοδος του επιπέδου i, $g^{(i)}$ είναι η συνάρτηση ενεργοποίησης του επιπέδου και $\boldsymbol{W}^{(i)}$ και $\boldsymbol{b}^{(i)}$ είναι ο πίνακας των βαρών και το διάνυσμα της πόλωσης αντίστοιχα. Έχουμε θεωρήσει επίσης ότι ισχύει

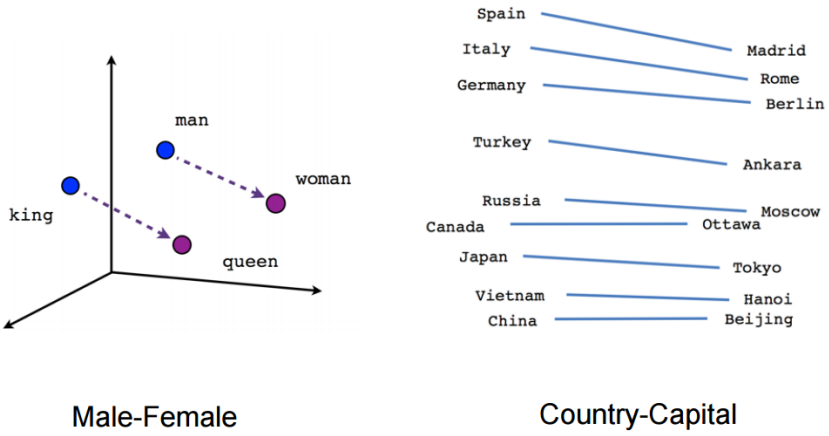
$$\boldsymbol{o}^{(0)} = g^{(0)}\left(\boldsymbol{W}^{(0)}\boldsymbol{x} + \boldsymbol{b}^{(0)}\right)$$

για το πρώτο επίπεδο του δικτύου.

4.1.3 Επίπεδα Εμφύτευσης (Embedding Layers)

Τα επίπεδα εμφύτευσης είναι συγγενικά των επιπέδων τεχνητών νευρώνων. Η κύρια λειτουργία τους είναι να συμπυκνώνουν την πληροφορία που υπάρχει στα δείγματα εισόδου. Χρησιμοποιούνται κυρίως, όταν τα δεδομένα εισόδου είναι κατηγορικά. Σε αυτήν την περίπτωση, όπως έχουμε ήδη αναφέρει, μπορούμε να αντιστοιχίσουμε κάθε μέλος του λεξιλογίου σε έναν ακέραιο αριθμό. Αυτή η πρακτική, όμως, δεν είναι ιδιαίτερα αποτελεσματική, αφού η αντιστοίχιση γίνεται αυθαίρετα κι έτσι είναι πιθανό τα μέλη του λεξιλογίου που είναι συγγενικά, να αντιστοιχιστούν σε μακρινούς αριθμούς. Κάτι τέτοιο δυσχεραίνει τη διαδικασία εκπαίδευσης του συστήματος. Θα προτιμούσαμε, επομένως, αντί της αυθαίρετης αντιστοίχισης να έχουμε μια πιο οργανωμένη διαδικασία, η οποία να ομαδοποιεί τα συγγενικά μέλη του λεξιλογίου σε γειτονικά σημεία. Το επίπεδο εμφύτευσης επιτελεί ακριβώς αυτόν το ρόλο. Βρίσκεται συνήθως στα πρώτα επίπεδα ενός νευρωνικού δικτύου και έχει ως σκοπό να διαμορφώσει ένα χώρο στον οποίο τα συγγενικά μέλη του λεξιλογίου ομαδοποιούνται σε κοντινές αποστάσεις. Αξίζει εδώ να διευκρινίσουμε την έννοια της συγγένειας, μια έννοια που δεν είναι καθόλου σαφής σε μαθηματικούς όρους. Πράγματι, δεν υπάρχει αυστηρή μαθηματική θεμελίωση για την εγγύτητα δύο μελών του λεξιλογίου, αυτή όμως μπορεί να εξαχθεί από τις εμφανίσεις των μελών στα δείγματα του συνόλου δεδομένων.

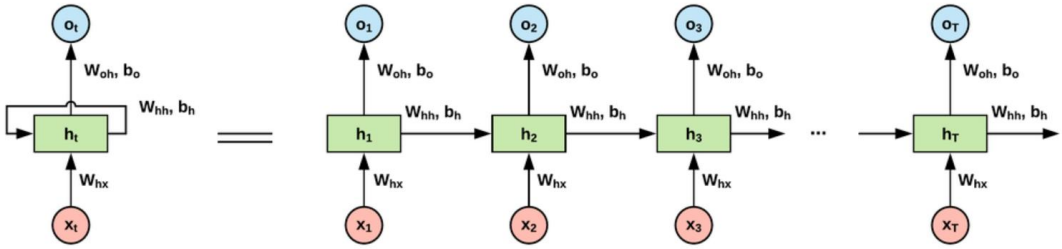
Την ίδια λειτουργία μπορούμε να επιτύχουμε και με ένα επίπεδο απλών τεχνητών νευρώνων, εάν μετασχηματίσουμε κατάλληλα τα δεδομένα εισόδου σε one-hot διανύσματα - αυτό είναι εφικτό, αφού οι διαφορετικές τιμές των δεδομένων εισόδου είναι πεπερασμένες. Όμως, η όλη διαδικασία είναι εμφανώς υπολογιστικά πιο αργή σε σχέση με τη χρήση επιπέδου εμφύτευσης και το αποτέλεσμα πανομοιότυπο.



Σχήμα 4.3: *Απεικονίσεις προτύπων στο χώρο που δημιουργεί ένα επίπεδο εμφύτευσης. Στα αριστερά τα δεδομένα έχουν διαχωριστεί βάσει γένους και αξιώματος με το ζεύγος king-woman να έχει τη μεγαλύτερη απόσταση. Στα δεξιά τα δεδομένα εισόδου έχουν διαχωριστεί σε χώρες και πρωτεύουσες.*

4.2 Αναδρομικά Νευρωνικά Δίκτυα

Τα αναδρομικά νευρωνικά δίκτυα, είναι τεχνητά νευρωνικά δίκτυα τα οποία διαθέτουν μνήμη, χάρη στην εσωτερική κατάσταση που διατηρούν. Οι μονάδες των δικτύων αυτών συνδέονται σειριακά μεταξύ τους με σκοπό να επεξεργαστούν χρονοσειρές, δηλαδή ακολουθίες δεδομένων εισόδου διατεταγμένες ως προς κάποια παράμετρο, όπως για παράδειγμα το χρόνο.



Σχήμα 4.4: *Αναδρομικό νευρωνικό δίκτυο σε συμπτυγμένη (αριστερά) και αναπτυγμένη (δεξιά) μορφή.*

Όπως φαίνεται στο Σχήμα 4.4, σε κάθε βήμα, $t \in \{1, 2, ...T\}$, εκτελούνται οι εξής υπολογισμοί

- Για την κρυφή κατάσταση, h_t, του εκάστοτε νευρώνα του δικτύου:

$$h_t = f(W_{hh}h_{t-1} + b_h + W_{hx}x_t)$$

όπου f είναι η συνάρτηση ενεργοποίησης για την κρυφή κατάσταση, W_{hh} και W_{hx} είναι οι εκπαιδεύσιμες παράμετροι, b_h είναι η προκατάληψη και τέλος x_t είναι η είσοδος στο δίκτυο.

- Για την έξοδο, o_t, του εκάστοτε νευρώνα του δικτύου:

$$o_t = g(W_{oh}h_t + b_o)$$

όπου g είναι η συνάρτηση ενεργοποίησης για την έξοδο, W_{oh} είναι οι εκπαιδεύσιμες παράμετροι και b_o είναι η προκατάληψη.

Αξίζει σε αυτό το σημείο να παρατηρήσουμε ότι οι εκπαιδεύσιμες παράμετροι του αναδρομικού νευρωνικού δικτύου είναι κοινές για κάθε χρονικό βήμα. Έχουμε, δηλαδή, μια περίπτωση δικτύου με μοιραζόμενες παραμέτρους.

Τα αναδρομικά νευρωνικά δίκτυα της παραπάνω μορφής αποτελούν ισχυρά εργαλεία μοντελοποίησης χρονοσειρών, θεωρητικά οποιουδήποτε μήκους. Η πράξη, όμως, δείχνει ότι τα δίκτυα αυτά αδυνατούν να μάθουν τις εξαρτήσεις μεταξύ δειγμάτων της εισόδιας χρονοσειράς, όταν αυτά είναι χρονικά πολύ απομακρυσμένα μεταξύ τους. Αδυνατούν με άλλα λόγια να αναγνωρίσουν και να εκφράσουν, μέσω των εκπαιδεύσιμων παραμέτρων, μακροχρόνιες εξαρτήσεις μεταξύ των δεδομένων εισόδου. Το πρόβλημα αυτό είναι καίριο για την εργασία μας, καθώς τα μουσικά κείμενα ενέχουν τέτοιου είδους μακροχρόνιες εξαρτήσεις και η προσέγγιση της σύνθεσης με τη χρήση του απλού αναδρομικού νευρωνικού δικτύου είναι καταδικασμένη να αποτύχει. Στις επόμενες παραγράφους θα αναλύσουμε δύο δομές που επιλύουν αυτό το πρόβλημα σε ικανοποιητικό βαθμό.

4.2.1 Νευρωνικά Δίκτυα Long Short-Term Memory

Τα νευρωνικά δίκτυα Long Short-Term Memory (LSTM) κατασκευάστηκαν με σκοπό την επεξεργασία ακολουθιών εισόδου μεγάλου μήκους. Ποιοτικά, μία μονάδα LSTM διαθέτει

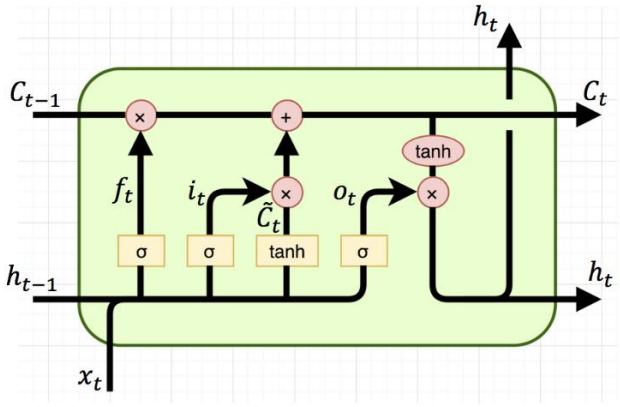
- μια πύλη εισόδου, που ελέγχει τη ροή νέας πληροφορίας στο εσωτερικό της
- μια πύλη απώλειας, που επιτρέπει ή αποτρέπει την παραμονή της υπάρχουσας πληροφορίας στο εσωτερικό της
- μια πύλη εξόδου, που ελέγχει τη ροή της υπάρχουσας πληροφορίας προς το εξωτερικό περιβάλλον

Αναλυτικότερα και όπως φαίνεται στο Σχήμα 4.5, μια μονάδα LSTM εκτελεί τους ακόλουθους υπολογισμούς:

$$\begin{aligned}
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c) \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}$$

όπου ως $\odot$ εννοούμε το γινόμενο Hadamard, δηλαδή το γινόμενο στοιχείο προς στοιχείο.

Είναι φανερό πως οι υπολογισμοί που εκτελεί μια μονάδα LSTM είναι πολλοί και υπολογιστικά κοπιαστικοί. Πράγματι, είναι ευρέως αποδεκτό ότι οι εν λόγω μονάδες είναι αργές και απαιτούν μεγάλη επεξεργαστική ισχύ.



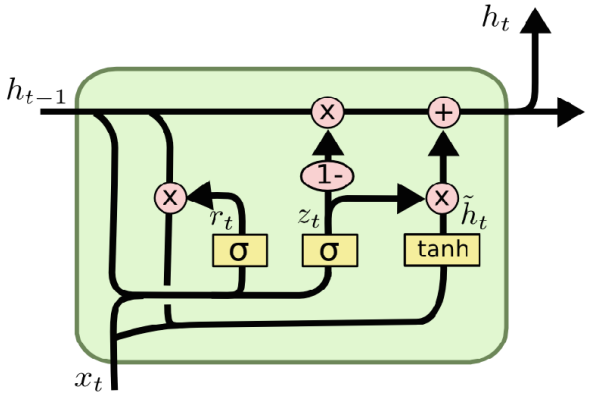
Σχήμα 4.5: *Μονάδα δικτύου LSTM*

4.2.2 Νευρωνικά Δίκτυα Gated Recurrent Unit

Τα νευρωνικά δίκτυα Gated Recurrent Unit (GRU) κατασκευάστηκαν με σκοπό την επεξεργασία ακολουθιών εισόδου μεγάλου μήκους χωρίς να χρειάζονται τη μεγάλη επεξεργαστική ισχύ που απαιτούν οι μονάδες LSTM. Πρόκειται, δηλαδή, για μια απλοποιημένη μονάδα, η οποία σε ορισμένες εργασίες διατηρεί την αποτελεσματικότητά της. Ποιοτικά, μία μονάδα GRU διαθέτει

- μια πύλη ανανέωσης, που ελέγχει τη ροή της υπάρχουσας πληροφορίας προς το εξωτερικό περιβάλλον. Αυτή η πύλη είναι αντίστοιχη της πύλης εξόδου μιας μονάδας LSTM.

- μια πύλη επαναφοράς, που ρυθμίζει τον όγκο πληροφορίας που η μονάδα πρέπει να ξεχάσει. Αυτή η πύλη αποτελεί τον συνδυασμό των πυλών εισόδου και απώλειας του LSTM.



Σχήμα 4.6: *Μονάδα δικτύου GRU*

Αναλυτικότερα και όπως φαίνεται στο Σχήμα 4.6, μια μονάδα GRU εκτελεί τους ακόλου-

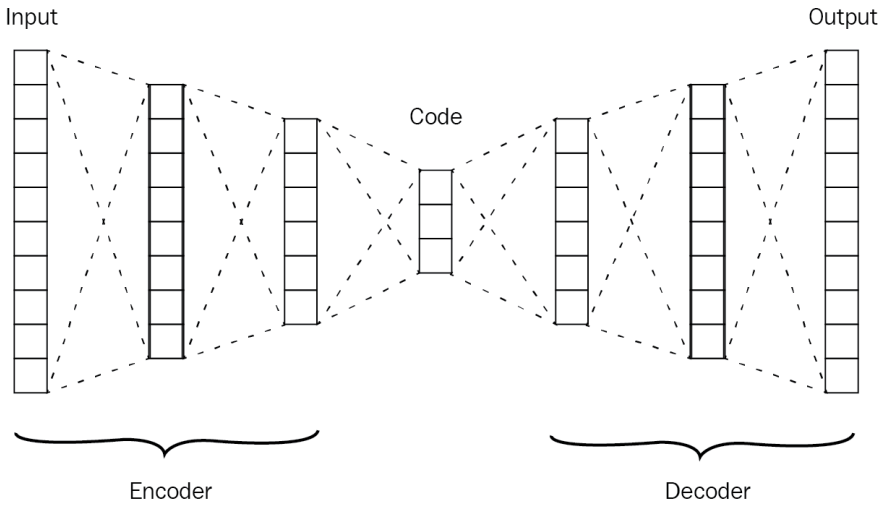
θους υπολογισμούς:

$$
\begin{aligned}
z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z) \\
r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r) \\
\tilde{h}_t &= \tanh(W_h x_t + r_t \odot (U_h h_{t-1}) + b_h) \\
h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t
\end{aligned}
$$

όπου και πάλι ως $\odot$ εννοούμε το γινόμενο Hadamard.

4.3 Αυτοκωδικοποιητές

Οι αυτοκωδικοποιητές αποτελούν μια πολύ ενδιαφέρουσα κατηγορία τεχνητών νευρωνικών δικτύων. Τα δίκτυα αυτά λειτουργούν με μη επιβλεπόμενο τρόπο, δεν απαιτείται δηλαδή, η οργάνωση του συνόλου δεδομένων σε ζεύγη εισόδου-εξόδου. Αντίθετα, ένας αυτοκωδικοποιητής θεωρεί ως επιθυμητή έξοδο, την είσοδο που του δόθηκε. Εκ πρώτης όψεως ένα τέτοιο δίκτυο φαίνεται να είναι άχρηστο, όμως θα δούμε πως η ικανότητα του να γενικεύει στα δεδομένα που του παρέχουμε και να κάνει προβλέψεις το καθιστά ελκυστικό σε πολλές εφαρμογές.



Σχήμα 4.7: *Αναπαράσταση της αρχιτεκτονικής ενός αυτοκωδικοποιητή*

Μια τυπική αρχιτεκτονική ενός αυτοκωδικοποιητή φαίνεται στο Σχήμα 4.7. Συγκεκριμένα, ο αυτοκωδικοποιητής είναι ένα σύστημα που έχει τις ακόλουθες τρεις συνιστώσες.

- Ο κωδικοποιητής. Πρόκειται για ένα νευρωνικό δίκτυο που κωδικοποιεί τα δεδομένα εισόδου, μετατρέπει, δηλαδή, τα δεδομένα εισόδου στην εσωτερική του αναπαράσταση. Η αναπαράσταση αυτή είναι συνήθως μικρότερης διαστατικότητας από αυτή των δεδομένων εισόδου.

- Το κρύφο επίπεδο ή επίπεδο κώδικα. Περιλαμβάνει τις εκπαιδεύσιμες παραμέτρους που σχετίζονται με την εσωτερική αναπαράσταση των δεδομένων εισόδου που διατηρεί

ο αυτοκωδικοποιητής. Αυτές οι παράμετροι ιδανικά ενέχουν όλη την πληροφορία που χρειάζεται για την ανακατασκευή ενός στιγμιοτύπου των δεδομένων εισόδου.

- Ο αποκωδικοποιητής. Πρόκειται για ακόμα ένα νευρωνικό δίκτυο που λαμβάνει ένα στιγμιότυπο της εσωτερικής αναπαράστασης του αυτοκωδικοποιητή και ανακατασκευάζει το στιγμιότυπο εισόδου που αρχικά είχε κωδικοποιηθεί με αυτήν την αναπαράσταση.

Μια ενδιαφέρουσα εφαρμογή του αυτοκωδικοποιητή είναι η ανακατασκευή ενός στιγμιοτύπου των δεδομένων εισόδου με γνώση μόνο της εσωτερικής αναπαράστασης και χρήση του αποκωδικοποιητή. Πιο συγκεκριμένα, μετά την εκμάθηση της εσωτερικής αναπαράστασης, διαγράφουμε τον κωδικοποιητή και διατηρούμε στο δίκτυο μόνο το κρυφό επίπεδο και τον αποκωδικοποιητή. Έτσι, αναμένουμε η είσοδος που πλέον δίνουμε στο δίκτυο να αρκεί για να ανακατασκευαστεί πλήρως ή έστω με παραπλήσιο τρόπο ένα στιγμιότυπο του αρχικού συνόλου δεδομένων. Εάν πετύχουμε κάτι τέτοιο, τότε μπορούμε να επιλέγουμε τυχαία σημεία της εσωτερικής αναπαράστασης του αυτοκωδικοποιητή και να παίρνουμε μια έξοδο κοντινή στο αρχικό σύνολο δεδομένων.

Δυστυχώς, κάτι τέτοιο δεν είναι τόσο απλό να επιτευχθεί με την έως τώρα διάταξη. Πιο συγκεκριμένα, πρέπει να κατανοήσουμε ότι οι παράμετροι του κρυφού επιπέδου ενός αυτοκωδικοποιητή ορίζουν έναν χώρο στον οποίο δεν έχουμε καμία δυνατότητα πλοήγησης, ούτε διαθέτουμε τον κωδικοποιητή, για να μας καθοδηγήσει. Επομένως, μια αυθαίρετη είσοδος στο κρυφό επίπεδο σε καμία περίπτωση δεν μπορεί να οδηγήσει με στατιστική βεβαιότητα στην επιθυμητή έξοδο. Θα υπάρχουν περιπτώσεις όπου επιλέγουμε κάποιο σημείο του χώρου αναπαράστασης που δεν έχει αντιστοιχιστεί με κανένα στιγμιότυπο του αρχικού συνόλου δεδομένων, ούτε προσεγγίζει κάποιο, με αποτέλεσμα η έξοδος του νευρωνικού δικτύου να είναι απροσδιόριστη.

Για να πετύχουμε τον παραπάνω στόχο, χρειάζεται να εισάγουμε έναν ακόμη περιορισμό για τις παραμέτρους του κρυφού επιπέδου. Έναν περιορισμό στο χώρο αναπαράσταση που διαμορφώνεται, ώστε αυτός να πληροί κάποιες εκ των προτέρων γνωστές συνθήκες, βάσει των οποίων κάθε αυθαίρετη ή έστω τυχαία είσοδος στο κρυφό επίπεδο να παράγει έξοδο που προσεγγίζει κάποιο από τα στιγμιότυπα εισόδου. Η τεχνική αυτή περιγράφεται λεπτομερώς στην επόμενη παράγραφο.

4.3.1 Variational Αυτοκωδικοποιητές (VAE)

Οι variational αυτοκωδικοποιητές είναι νευρωνικά δίκτυα, όμοια στην αρχιτεκτονική με τους αυτοκωδικοποιητές που παρουσιάσαμε στην προηγούμενη παράγραφο, διαφέρουν όμως από αυτούς στη μαθηματική μοντελοποίηση. Πρόκειται για γεννητικά μοντέλα, τα οποία προσπαθούν να ανακαλύψουν την εσωτερική δομή των δεδομένων εισόδου, με σκοπό να παράγουν παρόμοια δεδομένα. Στην εργασία μας, αυτή η λειτουργία είναι απαραίτητη, καθώς στοχεύουμε να παράγουμε νέα μουσικά κείμενα, βάσει όσων προηγουμένως παρατηρήθηκαν. Στη συνέχεια παρουσιάζουμε συντόμως το μαθηματικό υπόβαθρο πάνω στο οποίο στηρίζονται αυτοί οι αυτοκωδικοποιητές. Μια πλήρης πραγματεία μπορεί να βρεθεί στο [6].

Κεντρικό ρόλο στον variational αυτοκωδικοποιητή παίζει το κρυφό του επίπεδο. Θεωρούμε ότι οι τιμές των μεταβλητών αυτού του επιπέδου, έστω z_i, προκύπτουν από δειγματοληψία

μιας κατανομής, $p(z)$. Ισχύει, δηλαδή, $z_i \sim p(z)$. Θεωρούμε ακόμη ότι τα δεδομένα εισόδου, έστω x_i, προκύπτουν από την παρατήρηση των τιμών z_i, δηλαδή $x_i \sim p(x \mid z)$.

Σε αυτό το πλαίσιο, σκοπός μας είναι να βρούμε κατάλληλες τιμές z_i, σύμφωνα με τα δεδομένα εισόδου που διαθέτουμε και με τα οποία τροφοδοτούμε το σύστημα. Αναζητούμε, με άλλα λόγια, μια έκφραση για την ποσότητα $p(z \mid x)$.

Από το θεώρημα του Bayes γνωρίζουμε ότι

$$p(z \mid x) = \frac{p(x \mid z)}{p(x)} p(z) \tag{4.3}$$

Άγνωστος στην εξίσωση είναι ο παρονομαστής του κλάσματος, $p(x)$, ο οποίος βέβαια μπορεί να εκφραστεί ως $\int p(x|z)p(z)dz$. Αυτή η μορφή δυστυχώς απαιτεί εκθετικό χρόνο για να υπολογιστεί και έτσι η εύρεση της επιθυμητής ποσότητας με αυτόν τον τρόπο καθίσταται ανέφικτη. Αντ' αυτού θα προσεγγίσουμε την κατανομή $p(z \mid x)$ με μια οικογένεια κατανομών $q_{\lambda}(z \mid x)$, όπου λ είναι οι παράμετροι που προσδιορίζουν μια συγκεκριμένη κατανομή από την οικογένεια. Έπειτα, θα χρησιμοποιήσουμε την απόκλιση Kullback-Leibler, $D\left(q_{\lambda}(z \mid x) \parallel p(z \mid x)\right)$, για να μετρήσουμε την διαφορά μεταξύ των δύο κατανομών και να πάρουμε μια εκτίμηση της αποτελεσματικότητας της προσέγγισης που εφαρμόζουμε. Στόχος μας τώρα είναι να βρούμε κατάλληλες παραμέτρους λ, οι οποίες να ελαχιστοποιούν την απόσταση Kullback-Leibler. Προσπαθούμε να λύσουμε το πρόβλημα

$$q^*_{\lambda}(z \mid x) = \arg\min_{\lambda} D\left(q_{\lambda}(z \mid x) \parallel p(z \mid x)\right) \tag{4.4}$$

Και πάλι εάν κανείς μελετήσει τη μαθηματική έκφραση της απόκλισης Kullback-Leibler, θα εντοπίσει την ποσότητα $p(x)$, η οποία όπως είπαμε είναι δαπανηρό να υπολογιστεί. Όμως μπορούμε να φράξουμε την εν λόγω απόκλιση από την ELBO και να μεγιστοποιήσουμε αυτή. Περισσότερες λεπτομέρειες δεν μας είναι χρήσιμες και παραλείπονται. Αυτό που έχει αξία είναι το γεγονός ότι έχουμε μια αποδοτική μέθοδο επίλυσης της εξίσωσης 4.4.

Ας εξετάσουμε τώρα τη σύνδεση της αρχιτεκτονικής του νευρωνικού δικτύου ενός αυτοκωδικοποιητή με το παραπάνω μαθηματικό πλαίσιο. Θεωρούμε ότι ο κωδικοποιητής του νευρωνικού δικτύου παίρνει τα δεδομένα εισόδου και παράγει τις παραμέτρους, λ, και μοντελοποιεί την κατανομή προσέγγισης $q_{\theta}(z \mid x, \lambda)$, όπου θ είναι οι εσωτερικές παράμετροι του κωδικοποιητή. Το στιγμιότυπο, z, λαμβάνεται από τη δειγματοληψία αυτής της κατανομής και έπειτα δίδεται στον αποκωδικοποιητή. Αυτός μοντελοποιεί την κατανομή $p_{\phi}(x \mid z)$, όπου ϕ είναι οι εσωτερικές παράμετροί του. Τα δείγματα αυτής της κατανομής αποτελούν την έξοδο του συστήματος.

Για να πάρουμε δείγματα από το σύστημα χωρίς να δώσουμε κάποια είσοδο, απλώς δειγματολειπτούμε ένα $z \sim p(z)$ και το τροφοδοτούμε στον αποκωδικοποιητή. Αυτός με τη σειρά του θα μας δώσει ένα δείγμα της $p_{\phi}(x \mid z)$ που είναι κατάλληλα ρυθμισμένη να προσεγγίζει την κατανομή των δεδομένων μας.

4.4 Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης χρησιμοποιούνται κατά κόρον στα νευρωνικά δίκτυα και είναι αυτές που προσδίδουν τον έντονα μη γραμμικό χαρακτήρα σε αυτά τα συστήματα, αλλά και επιτρέπουν την αποδοτική εύρεση βέλτιστων αναπαραστάσεων των εισόδων. Οι συναρτήσεις αυτές επιδρούν στους - συνήθως γραμμικούς - μετασχηματισμούς που επιβάλλουν τα νευρωνικά δίκτυα στις εισόδους τους. Η έξοδός μιας συνάρτησης ενεργοποίησης ποιοτικά εκφράζει το βαθμό στον οποίο ο νευρώνας πρέπει να ενεργοποιηθεί.

Ορισμένες συνήθεις συναρτήσεις ενεργοποίησης που χρησιμοποιούνται είναι οι ακόλουθες:

- Η βηματική συνάρτηση Heaviside:

$$H(x) = \begin{cases} 0, & \text{αν } x < 0 \\ 1, & \text{αν } x \geq 0 \end{cases}$$

- Η σιγμοειδής ή λογιστική συνάρτηση:

$$sigmoid(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

  Η συνάρτηση αυτή έχει μια πολύ βολική ιδιότητα όσον αφορά στην παράγωγό της. Συγκεκριμένα ισχύει

$$sigmoid'(x) = sigmoid(x)(1 - sigmoid(x))$$

  Όπως θα αναλύσουμε στο κεφάλαιο 4.5, η ευκολία στον υπολογισμό της παραγώγου της συνάρτησης ενεργοποίησης είναι μεγάλης σημασίας για την αποδοτική, σε όρους ταχύτητας, εκπαίδευση του νευρωνικού δικτύου.

- Η συνάρτηση υπερβολικής εφαπτομένης:

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

  Ας παρατηρήσουμε, εδώ, ότι $tanh(x) = 2sigmoid(2x) - 1$, επομένως και η συνάρτηση υπερβολικής εφαπτομένης έχει πρώτη παράγωγο που είναι εύκολη στον υπολογισμό.

- Η ανορθωμένη γραμμική συνάρτηση (ReLU):

$$ReLU(x) = x^+ = max(0, x)$$

- Η συνάρτηση softmax:

$$\sigma(\boldsymbol{z})_i = \frac{e^z_i}{\sum_{j=1}^{K} e^z_j}, \ i = 1, ..., K \text{ και } \boldsymbol{z} = (z_1, ..., z_K)$$

  Η συνάρτηση αυτή χρησιμοποιείται κατά κόρον για την αντιστοίχιση των εξόδων ενός νευρωνικού δικτύου σε έγκυρες τιμές πιθανοτήτων. Συχνά, μάλιστα, αναφέρεται ως

επίπεδο softmax, επειδή η είσοδος σε αυτή τη συνάρτηση είναι ένα διάνυσμα που προέρχεται από το τελευταίο επίπεδο ενός νευρωνικού δικτύου.

Η επιλογή της κατάλληλης συνάρτησης ενεργοποίησης δεν είναι ούτε προφανής, ούτε εύκολη. Συνήθως στηριζόμαστε στα αποτελέσματα υπάρχουσας έρευνας ή κάνουμε δοκιμές και επιλέγουμε τη συνάρτηση που αποδίδει καλύτερα. Δύο ζητήματα θα πρέπει να έχουμε υπόψη, όταν επιλέγουμε μια συνάρτηση ενεργοποίησης. Αρχικά, πρέπει να λογαριάσουμε το υπολογιστικό κόστος. Εύκολα διαπιστώνει κανείς, με απλή επισκόπηση των μαθηματικών τύπων, ότι κάποιες συναρτήσεις ενεργοποίησης είναι απλούστερες στη μορφή τους από άλλες. Αυτές επιταχύνουν τη διαδικασία εκπαίδευσης του νευρωνικού δικτύου. Άλλο ένα κριτήριο που θα πρέπει να λάβουμε υπόψη είναι το σύνολο τιμών της συνάρτησης ενεργοποίησης. Δεδομένου ότι η έξοδος ενός νευρωνικού δικτύου είναι συνήθως η έξοδος της συνάρτησης ενεργοποίησής του, τότε θα πρέπει να φροντίσουμε, ώστε το σύνολο τιμών της συνάρτησης ενεργοποίησης να είναι υπερσύνολο των τιμών των επιθυμητών εξόδων. Σε αντίθετη περίπτωση, οι έξοδοι του νευρωνικού δικτύου θα έχουν πάντα απόκλιση από τις επιθυμητές εξόδους.

4.5 Εκπαίδευση Νευρωνικών Δικτύων

Μέχρι τώρα έχουμε αναλύσει σε βάθος τη λειτουργία όλων των συνιστωσών ενός συστήματος νευρωνικών δικτύων που θα μας χρειαστούν για το τελικό σύστημα που θα αναπτύξουμε. Καλούμαστε σε αυτήν την ενότητα να προσδιορίσουμε τον τρόπο με τον οποίο θα εκπαιδεύσουμε το νευρωνικό δίκτυο, δηλαδή, τον τρόπο με τον οποίο θα μεταβάλουμε τις εσωτερικές παραμέτρους του, έτσι ώστε η έξοδος του συστήματος να είναι όσο το δυνατόν παραπλήσια της επιθυμητής εξόδου.

Ας θυμηθούμε τον ορισμό των νευρωνικών δικτύων που δώσαμε στην αρχή του παρόντος κεφαλαίου και συγκεκριμένα ας εξετάσουμε τη γενική μαθηματική έκφραση ενός νευρωνικού δικτύου που αποδώσαμε ως $\hat{\boldsymbol{y}} = f(\boldsymbol{x}; \boldsymbol{\Theta})$. Γίνεται πλέον φανερό, πως ο μόνος τρόπος με τον οποίο μπορούμε να μεταβάλουμε δυναμικά το σύστημα είναι μέσω των παραμέτρων $\boldsymbol{\Theta}$, αφού η συνάρτηση f αφορά στην αρχιτεκτονική του συστήματος και δεν μπορεί να μεταβληθεί. Επομένως, η εκπαίδευση ενός νευρωνικού δικτύου ανάγεται στην εύρεση εκείνων των παραμέτρων $\boldsymbol{\Theta}$ που θα δώσουν συστημική έξοδο ικανοποιητικά παραπλήσια της επιθυμητής εξόδου, για κάθε είσοδο, είτε γνωστή, είτε άγνωστη.

Ξεκινούμε την πραγματεία μας με την καθιέρωση μιας μετρικής που προσδιορίζει το σφάλμα μεταξύ της επιθυμητής εξόδου, $\boldsymbol{y}$, και της συστημικής εξόδου, $\hat{\boldsymbol{y}}$. Η μετρική αυτή θα πρέπει να λαμβάνει την ελάχιστη τιμή της, όταν $\hat{\boldsymbol{y}} = \boldsymbol{y}$. Ορίζουμε τη συνάρτηση κόστους $\mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y})$ για ακριβώς αυτό το σκοπό. Ανάλογα με το είδος του νευρωνικού δικτύου, η συνάρτηση $\mathcal{L}$ ποικίλλει. Συνήθως, όμως, επιλέγεται, ως συνάρτηση κόστους, το μέσο τετραγωνικό σφάλμα για αριθμητικά δεδομένα και η διασταυρούμενη εντροπία για κατηγορικά δεδομένα. Αξίζει να σημειώσουμε, εδώ, ότι

$$\mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \mathcal{L}\left(f(\boldsymbol{x}; \boldsymbol{\Theta}), \boldsymbol{y}\right) = \mathcal{L}(\boldsymbol{\Theta})$$

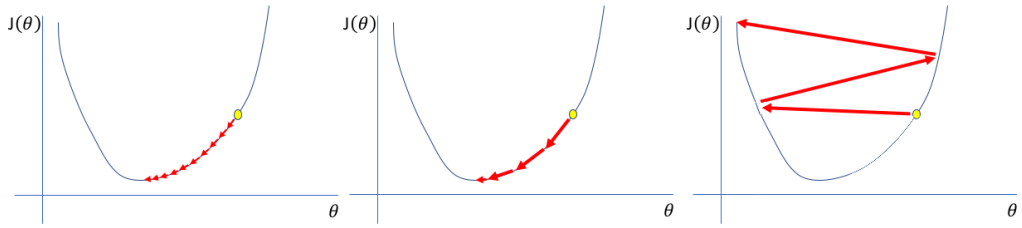
Επομένως, η διαδικασία εκπαίδευσης του νευρωνικού δικτύου έχει πλέον πάρει μαθηματική

μορφή και συγκεκριμένα συνίσταται στην ελαχιστοποίηση της συνάρτησης κόστους ως προς τις παραμέτρους $\boldsymbol{\Theta}$.

Για να ελαχιστοποιήσουμε τη συνάρτηση κόστους, θα χρησιμοποιήσουμε τον επαναληπτικό αλγόριθμο κατάβασης κλίσης Ο αλγόριθμος αυτός είναι ευρέως χρησιμοποιούμενος στο πεδίο των νευρωνικών δικτύων και έχουν προταθεί πολλές παραλλαγές του. Για μια πλήρη επισκόπηση των διαφορετικών αλγορίθμων βελτιστοποίησης, παραπέμπουμε τον αναγνώστη στο [7]. Σύμφωνα με τον απλό αλγόριθμο κατάβασης κλίσης, η τροποποίηση των παραμέτρων, $\boldsymbol{\Theta}$, του συστήματος γίνεται σε βήματα ως εξής:

$$\boldsymbol{\Theta}_{t+1} = \boldsymbol{\Theta}_t - \eta \frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} \tag{4.5}$$

όπου η ο ρυθμός μάθησης της διαδικασίας εκπαίδευσης. Ο ρυθμός μάθησης καθορίζει το μέγεθος της μεταβολής των παραμέτρων του δικτύου και πρέπει να επιλέγεται με μεγάλη προσοχή. Ας θυμηθούμε ότι ο στόχος της εκπαίδευσης είναι η προσέγγιση του ελαχίστου της συνάρτησης κόστους σε εύλογο χρονικό διάστημα. Η επιλογή μικρού ρυθμού εκπαίδευσης θα οδηγήσει σε πολύ αργή χρονικά σύγκλιση. Αντίθετα, ένας πολύ μεγάλος ρυθμός εκπαίδευσης θα οδηγήσει ενδεχομένως σε καθολική αποτυχία σύγκλισης, όπως φαίνεται στο Σχήμα 4.8.



Σχήμα 4.8: *Διάφοροι ρυθμοί εκπαίδευσης. Στα αριστερά ο ρυθμός εκπαίδευσης είναι πολύ μικρός και η σύγκλιση είναι αργή. Στα δεξιά ο ρυθμός εκπαίδευσης είναι πολύ μεγάλος με αποτέλεσμα η εκπαίδευση να αποκλίνει. Στο κέντρο ο ρυθμός εκπαίδευσης είναι ιδανικός.*

Για να συνοψίσουμε, η διαδικασία εκπαίδευσης ενός νευρωνικού δικτύου αποτελείται από τα παρακάτω βήματα:

- Υπολογισμός της εξόδου $\hat{\boldsymbol{y}} = f(\boldsymbol{x}; \boldsymbol{\Theta})$ και της τιμής συνάρτησης κόστους $\mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y})$.
- Αλλαγή της τιμής κάθε εσωτερικής παραμέτρου κατά $\Delta\theta = -\eta \frac{\partial \mathcal{L}}{\partial \theta}$.

Σε αυτό το σημείο πρέπει να τονίσουμε ότι η σχέση μεταξύ της τιμής της συνάρτησης κόστους και των επιμέρους εσωτερικών παραμέτρων του νευρωνικού δικτύου δεν είναι προφανής και άρα η μερική παραγώγιση ως προς αυτές χρήζει περαιτέρω επεξήγησης. Πιο συγκεκριμένα, η συνάρτηση κόστους υπολογίζεται από την έξοδο του συστήματος, όμως οι εσωτερικές παράμετροι βρίσκονται σε όλα τα επίπεδα του δικτύου. Η συσχέτιση όλων των εσωτερικών παραμέτρων με τη συνάρτηση κόστους γίνεται εύκολα αν κανείς λάβει υπόψη το γεγονός ότι η έξοδος ενός επιπέδου του νευρωνικού δικτύου γίνεται είσοδος στο αμέσως επόμενο επίπεδο. Το σκεπτικό αυτό οργανώνεται και χρησιμοποιείται στην τεχνική της οπίσθιας διάδοσης του σφάλματος [8], η οποία καθίσταται εφικτή χάρη στον κανόνα αλυσίδας

των παραγώγων. Οι μαθηματικές λεπτομέρειες είναι ελάσσονος σημασίας και παραλείπονται. Αξίζει μόνο να παραθέσουμε ένα σημείο-κλειδί που ενέχει αυτή η τεχνική και μας επιτρέπει να εκπαιδεύουμε αποτελεσματικά μεγάλα νευρωνικά δίκτυα. Η διάδοση του σφάλματος γίνεται αποκλειστικά από την έξοδο προς την είσοδο του συστήματος. Έτσι, η ανανέωση των παραμέτρων που βρίσκονται σε ένα επίπεδο του δικτύου, δεν εξαρτάται από τις παραμέτρους των προηγούμενων επιπέδων.

Μέρος II

Πρακτικό Μέρος

Κεφάλαιο 5

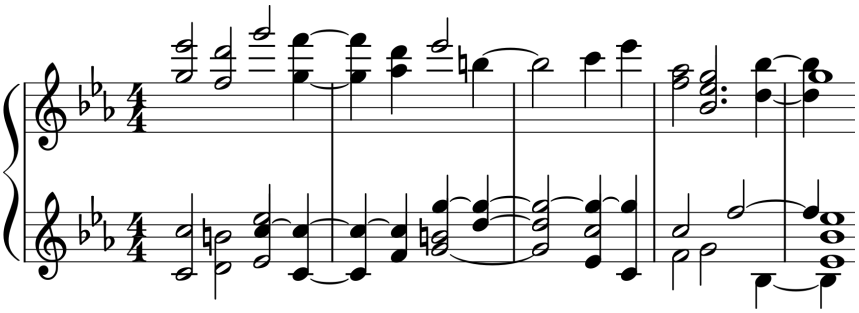
Σύνολα Δεδομένων και Αναπαραστάσεις

Στο κεφάλαιο αυτό παρουσιάζουμε όλα τα σχετικά ζητήματα με τα σύνολα δεδομένων που χρησιμοποιήσαμε σε αυτήν την εργασία καθώς και τον τρόπο αναπαράστασής τους. Είναι πλέον εύκολα κατανοητό πως στο πεδίο της Μηχανικής Μάθησης τα δεδομένα κατέχουν πρωταρχικό ρόλο. Έτσι, τόσο η επιλογή τους, η επεξεργασία τους αλλά και η χρήση ενός κατάλληλου συστήματος αναπαράστασής τους είναι καίριας σημασίας. Στις επόμενες παραγράφους θα αναλύσουμε σε βάθος όλα αυτά τα θέματα.

5.1 Σύνολα Δεδομένων

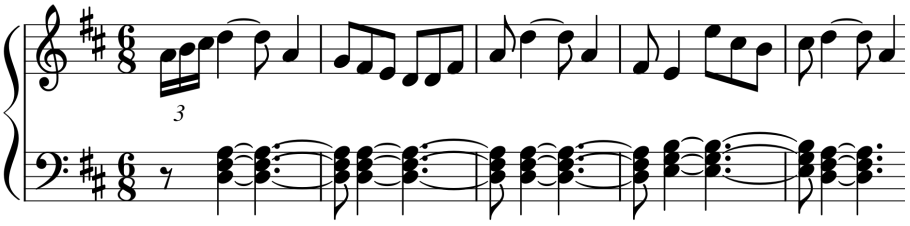
Η επιτυχής λειτουργία ενός νευρωνικού δικτύου στηρίζεται σε μεγάλο βαθμό στην ποιότητα των δεδομένων με τα οποία το εκπαιδεύουμε. Γι' αυτό έχει μεγάλη σημασία η επιλογή συνόλων δεδομένων τα οποία είναι ευρέως χρησιμοποιούμενα, ώστε να εξασφαλίσουμε ότι τυχόν σφάλματα ή παρατυπίες έχουν ήδη διορθωθεί από την κοινότητα. Χρησιμοποιούμε τα δύο ακόλουθα, πολύ γνωστά σύνολα δεδομένων, για τους σκοπούς αυτής της εργασίας.

JSB Chorales Αυτό το σύνολο δεδομένων περιλαμβάνει 382 κοράλ που συνέθεσε ο Γερμανός συνθέτης Γιόχαν Σεμπάστιαν Μπαχ (1685-1750). Πρόκειται για έργα κυρίως εκκλησιαστικά, τα περισσότερα από τα οποία γράφτηκαν για τέσσερις φωνές. Ο Μπαχ αναζητούσε τη θρησκευτική τελειότητα μέσα από τα έργα του και έτσι η μουσική του είναι οργανωμένη με έναν πολύ αυστηρό τρόπο, που δεν έχει σκοπό να προσεγγίσει το συναίσθημα του ακροατή, αλλά προσπαθεί με κάθε κόστος να ακολουθήσει ένα σύνολο κανόνων - που ο ίδιος ο συνθέτης ως επί το πλείστον θέσπισε. Σε αυτά τα έργα, επομένως, υπάρχει μικρή ελευθερία, υπό την έννοια ότι κάθε νότα έχει τοποθετηθεί σε μια συγκεκριμένη θέση και επιτελεί έναν αυστηρά καθορισμένο αρμονικό ρόλο. Όπως φαίνεται στο σχήμα 5.1, κάθε φωνή κινείται σε περιορισμένο εύρος, χωρίς πολλά άλματα και εκπλήξεις για τον ακροατή. Ο ρυθμός είναι ευδιάκριτος, καθώς σε κάθε ρυθμικό κτύπο έχουμε συνήθως την εισαγωγή μιας νέας συγχορδίας ή τη διατήρηση της προηγούμενης, ενώ σχεδόν όλες οι νότες έχουν ίσες χρονικές αξίες. Ένα ακόμα χαρακτηριστικό των έργων αυτών είναι η πολυπλοκότητα τους, όπως γίνεται κατανοητό από την ταυτόχρονη κίνηση κάποιων ή όλων των φωνών ανοδικά ή καθοδικά.



Σχήμα 5.1: *Απόσπασμα του έργου BWV 253 του Γιόχαν Σεμπάστιαν Μπαχ*

Nottingham Database (NMD) Πρόκειται για ένα σύνολο 1200 Βρετανικών και Αμερικάνικων φολκ κομματιών. Αυτά τα κομμάτια παρουσιάζουν μεγαλύτερη ελευθερία ως προς τη διάταξη των νοτών, καθώς ο σκοπός τους είναι η ευχαρίστηση του ακροατή και η κίνηση του ενδιαφέροντός του. Έτσι, η μελωδία είναι πιο ρευστή και υπάρχει μεγαλύτερη ρυθμική ποικιλία. Κύριο χαρακτηριστικό αυτού του συνόλου δεδομένων είναι η συστηματική συνοδεία της μελωδίας από μια τρίφωνη συγχορδία. Αυτή η δομή καθιστά το σύνολο NMD απλούστερο από το JSB.



Σχήμα 5.2: *Απόσπασμα ενός κομματιού του συνόλου Nottingham*

5.2 Αναπαράσταση Μουσικής στον Υπολογιστή

Η μουσική ως γνωστόν αποτελεί πλούσια γλώσσα έκφρασης συναισθημάτων και διήγησης ιστοριών. Δεν παύει όμως να αποτελεί και ένα μέσο επικοινωνίας των ανθρώπων που απαιτεί έναν πολύ αυστηρό κώδικα, ώστε όλοι οι συμμετέχοντες στην επικοινωνία να αντιλαμβάνονται το ίδιο μήνυμα. Έτσι, με το πέρασμα του χρόνου αναπτύχθηκε μια εκτενής σημειολογία που αποτελεί την κοινή γλώσσα της μουσικής. Τις τελευταίες δεκαετίες, με την καθιέρωση των υπολογιστών, δημιουργήθηκε η ανάγκη καταγραφής της μουσικής σε ψηφιακά μέσα. Για το λόγο αυτό αναπτύχθηκαν διάφορα πρωτόκολλα που επιτρέπουν την αναπαράσταση ενός μουσικού κειμένου με ψηφιακό τρόπο. Αυτά τα πρωτόκολλα ορίζουν με αυστηρό τρόπο όλα εκείνα τα στοιχεία που αποτελούν τη γλώσσα της μουσικής.

Ένα από τα πιο γνωστά πρωτόκολλα αναπαράστασης της μουσικής γλώσσας με ψηφιακό τρόπο είναι το MIDI - μια συνοπτική παρουσίασή του γίνεται στο [9]. Για πρώτη φορά προτάθηκε τον Οκτώβριο του 1981 από τους Dave Smith και Chet Wood και είχε σκοπό να καθιερώσει μια κοινώς αποδεκτή διεπαφή μεταξύ τον υπολογιστή και των διαφόρων ηλεκτρονικών μουσικών οργάνων της εποχής και να δώσει τη δυνατότητα ηχογράφησής τους.

Αυτό το πρωτόκολλα κωδικοποιεί όλα τα μουσικά σύμβολα ως γεγονότα που συμβαίνουν στο χρόνο. Τα γεγονότα μπορεί να είναι για παράδειγμα η αρχή και το τέλος μιας νότας, η επισήμανση κάποιου επιπέδου δυναμικής, κάποιας ρυθμικής σφραγίδας, κάποιου κλειδιού και γενικότερα οποιουδήποτε στοιχείου εμπλέκεται στο μουσικό κομμάτι και το επηρεάζει. Περισσότερες λεπτομέρειες του πρωτοκόλλου MIDI δεν μας ενδιαφέρουν άμεσα, αφού υπάρχουν διαθέσιμες πολλές βιβλιοθήκες κώδικα που καθιστούν εύκολη την επεξεργασία και την εξαγωγή πληροφοριών από τα αρχεία αυτού του τύπου.

5.3 Αναπαράσταση Μουσικής για Μηχανική Μάθηση

Εκτός από την ψηφιακή μορφή που έχουν τα δεδομένα που χρησιμοποιούμε, μπορούμε να επιλέξουμε μια πιο βολική αναπαράσταση, εξειδικευμένη στο πρόβλημα που επιλύουμε. Η αναπαράσταση που θα επιλέξουμε να χρησιμοποιήσουμε παίζει πολύ σημαντικό ρόλο για την επιτυχή εκπαίδευση του νευρωνικού δικτύου. Ακόμα και ένα ισχυρό σύστημα είναι καταδικασμένο να αποτύχει, αν η αναπαράσταση των δεδομένων με τα οποία το τροφοδοτούμε είναι κακής ποιότητας. Επιλέγουμε σε αυτήν την εργασία, μια διανυσματική αναπαράσταση που προτάθηκε από το [10]. Σε αυτήν την αναπαράσταση, κάθε νότα του κομματιού εκφράζεται ως ένα διάνυσμα $\boldsymbol{X}$, που έχει τρεις συνιστώσες (P, dt, D) καθεμία εκ των οποίων είναι ένα one-hot διάνυσμα.

- Το P (pitch) αναπαριστά τον τόνο της νότας. Θεωρούμε συνολικά 89 διαφορετικούς τόνους, 88 εκ των οποίων είναι οι διαθέσιμοι τόνοι σε ένα πιάνο και έναν ακόμη χρησιμοποιούμε για τη σηματοδότηση του τέλους μιας μουσικής φράσης - λέξημα τέλους. Επομένως $P \in \{0, 1\}^{89}$.

- Το dt εκφράζει το χρόνο που πέρασε από την προηγούμενη νότα. Τονίζουμε εδώ, ότι αν $dt = 0$, τότε η τρέχουσα νότα ηχεί ταυτόχρονα με την προηγούμενή της. Με αυτόν τον κομψό τρόπο αναπαριστούμε τις συγχορδίες. Εφόσον ο χρόνος είναι συνεχής ποσότητα, χρησιμοποιούμε κβάντωση της χρονικής αξίας μιας πλήρους νότας σε 33 ζώνες. Επομένως $dt \in \{0, 1\}^{33}$.

- Το D (duration) εκφράζει τη χρονική διάρκεια κατά την οποία ηχεί η τρέχουσα νότα. Χρησιμοποιούμε την ίδια κβάντωση και έχουμε ομοίως ότι $D \in \{0, 1\}^{33}$.

Η επιλογή 33 χρονικών ζωνών μας επιτρέπει να εκφράσουμε όλες τις μουσικές διάρκειες που χρησιμοποιούνται συχνά, όπως το όγδοο ή το παρεστιγμένο, καθώς και πιο περίπλοκες διάρκειες όπως τα τρίηχα και οι τρίλιες.

Αξίζει σε αυτό το σημείο να αιτιολογήσουμε τα πλεονεκτήματα αυτή της αναπαράστασης και το λόγο για τον οποίο αυτή είναι ελκυστική. Η μουσική, ως αντικείμενο μελέτης της Μηχανικής Μάθησης, είναι παρεμφερής με τη φυσική γλώσσα. Και στις δύο περιπτώσεις υπάρχει ένα λεξιλόγιο και κάθε στιγμιότυπο του συνόλου δεδομένων αποτελείται από μια χρονοσειρά μελών του λεξιλογίου. Υπάρχει όμως μια σημαντική διαφορά μεταξύ των δύο που καθιστά την μελέτη της μουσικής πιο περίπλοκη. Η μουσική είναι εγγενώς δι-αξονική στη φύση της, υπό την έννοια ότι εκτυλίσσεται τόσο στο χρόνο όσο και στο χώρο της αρμονίας,

με τη μορφή συγχορδιών. Κύριο αίτιο αυτής της διαφορά είναι η δυνατότητα ταυτόχρονης ήχησης νοτών στη μουσική, ενώ κάτι αντίστοιχο δεν συναντάται στη φυσική γλώσσα. Η χρήση αυτής της αναπαράστασης μας επιτρέπει να εκφράζουμε τη χωρική συνιστώσα της μουσικής στο πεδίο του χρόνου (θέτοντας $dt = 0$ όπως περιγράψαμε προηγουμένως). Έτσι καταλήγουμε σε μια αναπαράσταση που συμπυκνώνει όλη την πληροφορία σε έναν μόνο άξονα και μπορούμε να στηριχθούμε στην έρευνα πάνω στην επεξεργασία φυσικής γλώσσας για τα πειράματά μας. Ταυτόχρονα, με αυτήν την αναπαράσταση, το σύνολο των δεδομένων μας καθίσταται κατάλληλο για χρήση στα αναδρομικά δίκτυα που έχουμε αναφέρει και τα οποία δύνανται να επεξεργαστούν χρονοσειρές μίας μόνο διάστασης.

Επιπλέον του μουσικού περιεχομένου, πρέπει να αναπαραστήσουμε και το μουσικό είδος ενός κομματιού. Αυτό γίνεται εύκολα αν θεωρήσουμε ακόμα ένα one-hot διάνυσμα $\boldsymbol{\mathcal{Z}_{cat}} \in \{0, 1\}^s$, όπου s το πλήθος των διαφορετικών μουσικών ειδών που χρησιμοποιούμε.

Ως τελευταίο βήμα, για τους σκοπούς της εργασίας, χωρίζουμε ένα μουσικό κομμάτι σε τμήματα καθένα εκ των οποίων αποτελείται από k μη επικαλυπτόμενα μουσικά μέτρα - ή λιγότερα αν πρόκειται για το τελευταίο τμήμα. Στην παρούσα εργασία ασχολούμαστε μόνο με τις περιπτώσεις $k = 4$ ή $k = 8$. Από εδώ και πέρα θεωρούμε ως στιγμιότυπα του συνόλου δεδομένων αυτά τα τμήματα k μουσικών μέτρων και όχι εξ ολοκλήρου τα κομμάτια που το απαρτίζουν. Ορίζουμε n_T, το πλήθος των νοτών σε ένα τμήμα. Είμαστε πλέον σε θέση να ορίσουμε αυστηρά τη μαθηματική μορφή ενός στιγμιοτύπου του συνόλου δεδομένων μας ως ένα διάνυσμα νοτών $\boldsymbol{\mathcal{X}} = (\boldsymbol{X_1}, \boldsymbol{X_2} \ldots, \boldsymbol{X_{n_T}})$ και ένα διάνυσμα μουσικού είδους $\boldsymbol{\mathcal{Z}_{cat}}$.

Σκοπός του νευρωνικού μας δικτύου, όπως θα δούμε στο επόμενο κεφάλαιο, είναι η ανακατασκευή του διανύσματος νοτών και η πρόβλεψη του μουσικού είδους βάσει των δοθέντων νοτών. Αυτή η πρακτική θα δώσει μετέπειτα στο νευρωνικό μας δίκτυο τη δυνατότητα συμπερασμού μουσικού περιεχομένου με συγκεκριμένο είδος.

Κεφάλαιο 6

Υλοποίηση του Συστήματος

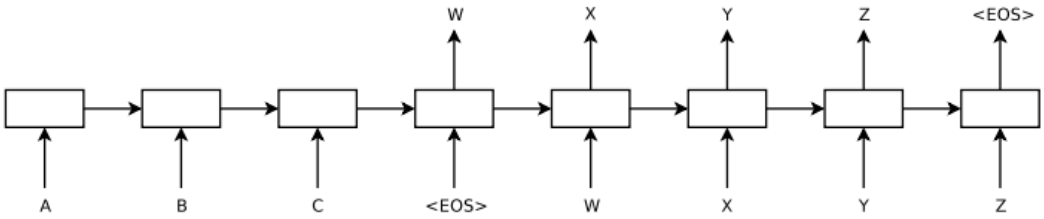
Σε αυτό το κεφάλαιο αναλύουμε λεπτομερώς το νευρωνικό δίκτυο που αναπτύξαμε στα πλαίσια της παρούσας διπλωματικής εργασίας. Στις επόμενες παραγράφους θα περιγράψουμε τις συνιστώσες του συστήματος, τον τρόπο εκπαίδευσής του καθώς και την απαραίτητη επεξεργασία των δεδομένων πριν την είσοδο και μετά την έξοδο. Το σύστημα είναι κατά κύριο λόγο εμπνευσμένο από τη δουλειά που παρουσιάζεται στο [11].

Στόχος του συστήματος είναι η παραγωγή μουσικής με βάση το επιλεγμένο είδος και χωρίς καμία παρέμβαση από τον συνθέτη. Με άλλα λόγια, η μόνη είσοδος στο σύστημα κατά τη λειτουργία του θα είναι το μουσικό είδος και πέρα από αυτήν δεν θα δίνεται καμία άλλη πληροφορία. Επιθυμητή έξοδος θα είναι μια μουσική φράση συγκεκριμένης διάρκειας, η οποία αντικατοπτρίζει το μουσικό είδος που έχει επιλεγεί. Η διάρκεια της μουσικής φράσης είναι ρητά ορισμένη από τη διάρκεια των μουσικών τμημάτων που αποτελούν το σύνολο δεδομένων. Όπως αναφέραμε στο προηγούμενο κεφάλαιο πειραματιζόμαστε με διάρκειες $k = 4$ και $k = 8$ μουσικών μέτρων. Είναι εμφανές ότι όσο μεγαλύτερη είναι η διάρκεια ενός μουσικού τμήματος, τόσο μεγαλύτερη είναι η χρονοσειρά νοτών που το αντιπροσωπεύει και επομένως η εκπαίδευση του νευρωνικού δικτύου καθίσταται δυσκολότερη.

6.1 Αρχιτεκτονική Seq2Seq

Προτού μελετήσουμε τις λεπτομέρειες υλοποίησης του νευρωνικού δικτύου της εργασίας, αξίζει να αναφερθούμε στο γενικότερο πρόβλημα το οποίο προσπαθούμε να επιλύσουμε. Πιο συγκεκριμένα και εάν αγνοήσουμε τη δυνατότητα επιλογής μουσικού είδους από το χρήστη, είμαστε αντιμέτωποι με ένα πρόβλημα seq2seq [12]. Σε αυτό το πρόβλημα έχουμε μια ακολουθία εισόδου και μια ακολουθία επιθυμητής εξόδου - οι οποίες πιθανώς έχουν διαφορετικά μήκη. Σκοπός του νευρωνικού δικτύου είναι η σωστή πρόβλεψη της επιθυμητής εξόδου με βάση την είσοδο. Αυτό το πρόβλημα δίνει ένα αρκετά γενικό θεωρητικό υπόβαθρο για προβλήματα όπως η αυτόματη μετάφραση από μια φυσική γλώσσα σε μια άλλη, ή η αυτόματη ολοκλήρωση μιας φράσης.

Η βασική ιδέα μιας αρχιτεκτονικής seq2seq στηρίζεται σε ένα αναδρομικό νευρωνικό δίκτυο που κωδικοποιεί την είσοδο σε ένα διάνυσμα σταθερού μήκους, το οποίο μετέπειτα χρησιμοποιείται από ένα επίσης αναδρομικό νευρωνικό δίκτυο που έχει σκοπό να εξάγει την επιθυμητή έξοδο. Συνήθως προσαρτούμε ένα λέξημα που προσδιορίζει την αρχή και ένα που προσδιορίζει το τέλος της πρότασης τόσο στην ακολουθία εισόδου όσο και στην ακολουθία



Σχήμα 6.1: *Γενική δομή ενός προβλήματος seq2seq*

εξόδου. Με αυτόν τον τρόπο, οριοθετούμε τις προτάσεις, ώστε κατά την πρόβλεψη να ξέρουμε πότε αρχίζει και πότε τελειώνει η έξοδος του νευρωνικού δικτύου.

Για τους σκοπούς της παρούσας εργασίας, χρησιμοποιούμε αυτό το θεωρητικό υπόβαθρο με ελαφρές τροποποιήσεις. Πιο συγκεκριμένα, η αρχιτεκτονική μας - που όπως θα δούμε στην επόμενη παράγραφο είναι ένας variational αυτοκωδικοποιητής - διαθέτει δύο αναδρομικά νευρωνικά δίκτυα για την κωδικοποίηση της εισόδιας ακολουθίας και την μετέπειτα εξαγωγή της επιθυμητής εξόδου. Υπενθυμίζουμε ότι η επιθυμητή έξοδος, στο πλαίσιο ενός αυτοκωδικοποιητή, ταυτίζεται με την είσοδο. Προσαρτούμε στην επιθυμητή έξοδο ένα εναρκτήριο λέξημα, ώστε αυτό να σηματοδοτεί την αρχή του μουσικού τμήματος. Αυτό είναι απαραίτητο στην εφαρμογή μας, καθώς κατά το συμπερασμό το δίκτυο δεν έχει καμιά είσοδο σχετική με το μουσικό περιεχόμενο. Έτσι, ορίζουμε εμείς αυτό το λέξημα ως την αρχική είσοδο στο νευρωνικό δίκτυο, το οποίο μετέπειτα θα συνεχίσει την παραγωγή μουσικής βήμα προς βήμα. Περισσότερες λεπτομέρειες δίνονται στις επόμενες παραγράφους.

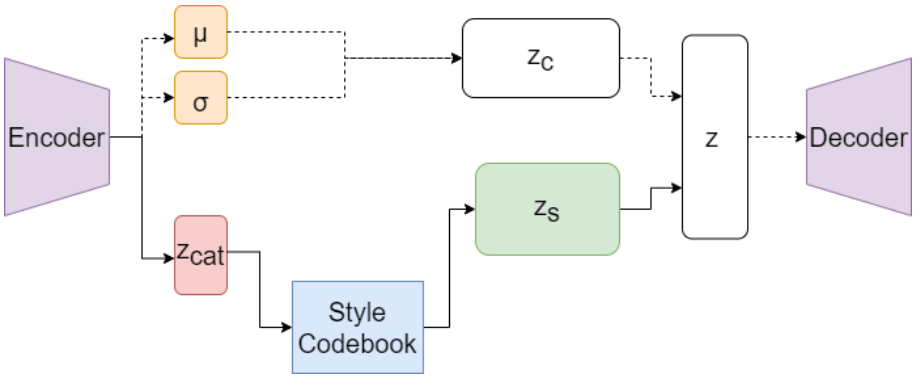
6.2 Αρχιτεκτονική του Νευρωνικού Δικτύου

Το νευρωνικό δίκτυο της εργασίας μας είναι ένας variational αυτοκωδικοποιητής, οι συνιστώσες του οποίου είναι είναι ικανές να επεξεργαστούν χρονοσειρές. Επιλέγουμε έναν variational αυτοκωδικοποιητή, γιατί μας ενδιαφέρει η ικανότητά του να αποτυπώνει εσωτερικά την κατανομή των δεδομένων εισόδου και να παράγει νέα δεδομένα βάσει αυτής.

Όπως έχουμε αναφέρει στο κεφάλαιο 4.3, ένας αυτοκωδικοποιητής έχει ένα δίκτυο κωδικοποιητή και ένα δίκτυο αποκωδικοποιητή. Χρησιμοποιούμε και στις δύο περιπτώσεις αναδρομικά νευρωνικά δίκτυα και συγκεκριμένα επιλέγουμε τα δίκτυα LSTM ή GRU. Αυτή η επιλογή δικαιολογείται από το γεγονός ότι τόσο η είσοδος στο δίκτυο μας όσο και η έξοδος από αυτό είναι χρονοσειρές σχετικά μακράς διάρκειας και όπως αναλύθηκε, αυτά τα αναδρομικά νευρωνικά δίκτυα είναι πολύ αποτελεσματικά στην επεξεργασία τέτοιου είδους δεδομένων. Σημειώνουμε εδώ ότι, πριν το αναδρομικό δίκτυο του κωδικοποιητή, έχουμε τοποθετήσει ένα επίπεδο εμφύτευσης με σκοπό να μειώσουμε τη διαστατικότητα των στιγμιοτύπων της εισόδου.

Το κρυφό επίπεδο που χρησιμοποιούμε στον αυτοκωδικοποιητή είναι ίσως το πιο περίπλοκο. Αυτό το επίπεδο χωρίζεται σε δύο υποδίκτυα, τα οποία εν τέλει συνενώνονται και αποτελούν το τελικό επίπεδο κώδικα, z, που διαμορφώνει το χώρο αναπαράστασης της πληροφορίας του αυτοκωδικοποιητή. Οι δύο συνιστώσες αντιστοιχούν στις αναπαραστάσεις των δύο λειτουργιών που επιδιώκουμε να εξάγουμε από τα δεδομένα.

Η μία συνιστώσα, z_c, αφορά στην εξαγωγή του μουσικού περιεχομένου, δηλαδή ασχο-



Σχήμα 6.2: *Αρχιτεκτονική του νευρωνικού δικτύου*

λείται με την εύρεση μιας κατάλληλης αναπαράστασης, ώστε το σύστημα να μπορεί να αναπαράξει αποτελεσματικά την είσοδο. Αυτή η συνιστώσα ακολουθεί τη δομή ενός τυπικού variational αυτοκωδικοποιητή όπου η οικογένεια κατανομών προσέγγισης $q_\phi(z \mid x, \hat{n})$ είναι η κανονική κατανομή (μια συνοπτική ανάλυση του μαθηματικού υποβάθρου του variational αυτοκωδικοποιητή γίνεται στην παράγραφο 4.3.1). Επομένως, η συνιστώσα αυτή αποτελείται από δύο επίπεδα, μ και σ, ίσου πλήθους απλών τεχνητών νευρώνων, τα οποία μοντελοποιούν τις μέσες τιμές και τις τυπικές αποκλίσεις των κατανομών Gauss του αυτοκωδικοποιητή. Για να είναι υπολογίσιμες οι παράγωγοι, εφαρμόζουμε το - σύνηθες στους variational αυτοκωδικοποιητές - τρικ επαναπροσδιορισμού των παραμέτρων [6] σύμφωνα με το οποίο

$$z_c = \mu + \epsilon\sigma$$

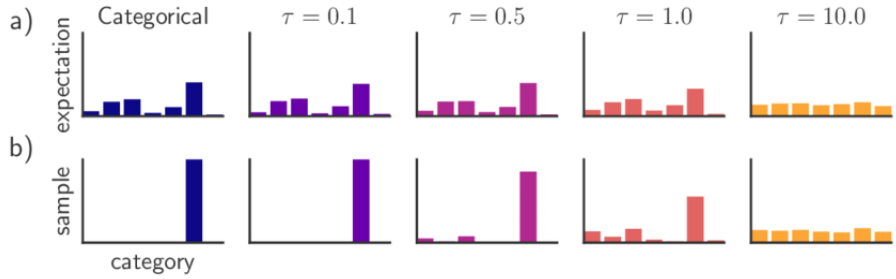
όπου ϵ είναι ένας τυχαίος θόρυβος Gauss.

Η δεύτερη συνιστώσα, z_s, σχετίζεται με την εκμάθηση του μουσικού είδους, δηλαδή τη σωστή πρόβλεψή του από το μουσικό τμήμα που δόθηκε ως είσοδος. Αρχικά ο κωδικοποιητής παράγει το διάνυσμα z_{cat} που αντιπροσωπεύει το μουσικό είδος της εισόδου και το οποίο θεωρούμε ότι είναι one-hot. Η διακριτή φύση του z_{cat} καθιστά ανέφικτο τον υπολογισμό των παραγώγων που απαιτούνται κατά την εκπαίδευση του δικτύου, γι' αυτό χρησιμοποιούμε το μετασχηματισμό βάσει της κατανομής Gumbel [13]:

$$G = -\log(-\log(\text{Unif}[0,1]))$$
$$z'_{cat} = \text{softmax}\left((z_{cat} + G)/\tau_{gumbel}\right)$$

όπου G είναι ο θόρυβος και τ_{gumbel} η θερμοκρασία Gumbel. Όσο πιο κοντά στο μηδέν είναι η θερμοκρασία, τόσο το διάνυσμα z'_{cat} προσεγγίζει το one-hot αντίστοιχό του. Χρησιμοποιούμε το μετασχηματισμένο διάνυσμα ως είσοδο σε ένα επίπεδο εμφύτευσης. Έτσι, παίρνουμε μια συμπαγή αναπαράσταση του μουσικού είδους, η οποία εμπειρικά αποδίδει καλύτερα από την one-hot αναπαράσταση. Αυτή η αναπαράσταση συνιστά το z_s.

Από κοινού οι συνιστώσες z_c και z_s συνενώνονται και αποτελούν το κρυφό επίπεδο z, το οποίο εν τέλει εμπεριέχει όλη την απαραίτητη πληροφορία για την αυτόνομη παραγωγή ενός μουσικού τμήματος συγκεκριμένου είδους από το νευρωνικό δίκτυο.



Σχήμα 6.3: *Μετασχηματισμός Gumbel ενός διανύσματος για διάφορες θερμοκρασίες*

6.3 Λειτουργία του Συστήματος

Για την καλύτερη κατανόηση των εσωτερικών συνιστωσών του νευρωνικού δικτύου που αναπτύξαμε, θα εξετάσουμε τις δύο περιπτώσεις στις οποίες λειτουργεί ένας variational αυτοκωδικοποιητής. Αρχικά, θα μελετήσουμε την πορεία των δεδομένων από την είσοδο έως την έξοδο κατά την εκπαίδευση του συστήματος. Υπενθυμίζουμε ότι κατά την εκπαίδευση, το σύστημα είναι εφοδιασμένο τόσο με τον κωδικοποιητή όσο και με τον αποκωδικοποιητή. Έπειτα, θα εξετάσουμε τη λειτουργία συμπερασμού, δηλαδή, τη διαδικασία παραγωγής μουσικών τμημάτων, όπου έχουμε απομακρύνει τον κωδικοποιητή και η μόνη είσοδος στο σύστημα είναι το μουσικό είδος που επιλέγει ο χρήστης.

6.3.1 Λειτουργία κατά την Εκπαίδευση

Έχουμε ήδη ορίσει ως είσοδο στο σύστημα κατά την εκπαίδευση τα δύο διανύσματα $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2} \ldots, \boldsymbol{X_{n_T}})$ και $\boldsymbol{\mathcal{Z}_{cat}}$. Θα μελετήσουμε τώρα την αλληλεπίδραση αυτών με το νευρωνικό μας δίκτυο. Το διάνυσμα $\boldsymbol{X}$ είναι αυτό που περνάει μέσα από τα επίπεδα του νευρωνικού δικτύου και παρέχει την πληροφορία τόσο για την εκμάθηση του μουσικού περιεχομένου, όσο και για την πρόβλεψη του μουσικού είδους. Το διάνυσμα $\boldsymbol{\mathcal{Z}_{cat}}$ χρησιμοποιείται μόνο στην έξοδο για την επιβεβαίωση της σωστής πρόβλεψης του μουσικού είδους.

Η πορεία που ακολουθεί το διάνυσμα $\boldsymbol{X}$ μέσα στο νευρωνικό δίκτυο έχει ως εξής. Αρχικά, τροφοδοτούμε τη συνιστώσα P καθενός $\boldsymbol{X_i} \in \boldsymbol{X}$ στο επίπεδο εμφύτευσης, με σκοπό να μειώσουμε τη διαστατικότητά της σε μια τιμή κοντά σε αυτές των dt και D. Έχουμε επιλέξει την τιμή 32 και έτσι παίρνουμε το διάνυσμα νοτών $\boldsymbol{X_i'} = (P', dt, D)$, όπου πλέον $P' \in \mathbb{R}^{32}$ και $dt, D \in \{0, 1\}$[33]. Έπειτα τροφοδοτούμε το $\boldsymbol{X_i'}$ στον κωδικοποιητή, από τον οποίο παίρνουμε τις μέσες τιμές και τις τυπικές αποκλίσεις των κατανομών Gauss που μοντελοποιεί ο αυτοκωδικοποιητής, καθώς και μια πρόβλεψη, $\widehat{\boldsymbol{\mathcal{Z}}}_{\boldsymbol{cat}}$, για το μουσικό είδος. Εφαρμόζουμε τους δύο μετασχηματισμούς που αναλύσαμε στην προηγούμενη παράγραφο και δίνουμε το παραχθέν $\widehat{\boldsymbol{\mathcal{Z}}}'_{\boldsymbol{cat}}$ στο επίπεδο εμφύτευσης του μουσικού είδους. Έτσι, έχουμε τις εσωτερικές αναπαραστάσεις z_c και z_s τις οποίες συνενώνουμε στο τελικό διάνυσμα z.

Τέλος, χρησιμοποιούμε το z ως αρχική κατάσταση του αναδρομικού δικτύου του αποκωδικοποιητή και τροφοδοτούμε σε αυτόν το εναρκτήριο λέξημα. Συγκεντρώνουμε τις εξόδους που δίνει ο αποκωδικοποιητής βήμα προς βήμα, έως ότου βρούμε το λέξημα τέλους. Φτιάχνουμε έτσι την ακολουθία νοτών εξόδου, $\widehat{\boldsymbol{X}}$, του συστήματος. Η συνολική έξοδος του δικτύου είναι η $\hat{\boldsymbol{y}} = \left(\widehat{\boldsymbol{X}}, \widehat{\boldsymbol{\mathcal{Z}}}_{\boldsymbol{cat}}\right)$.

Αλγοριθμος 6.1: *Εμπρόσθια τροφοδότηση της εισόδου στο νευρωνικό δίκτυο*

Είσοδος: $\boldsymbol{X}$ (*το διάνυσμα νοτών*)
Έξοδος: $\widehat{\boldsymbol{X}}, \widehat{\boldsymbol{Z}}_{\boldsymbol{cat}}$ (*το ανακατασκευασμένο διάνυσμα νοτών και η πρόβλεψη του μουσικού είδους*)
Θέσε $(P, dt, D) = \boldsymbol{X}$
Θέσε $\boldsymbol{X'} = (\text{embed}(P), dt, D)$
Θέσε $\mu, \sigma, \widehat{\boldsymbol{Z}}_{\boldsymbol{cat}} = \text{encode}(\boldsymbol{X'})$

Θέσε $z_c = \mu + \epsilon \cdot \sigma$
Θέσε $\hat{z}'_{cat} = \text{softmax}\left((\hat{z}_{cat} + G)/\tau_{gumbel}\right)$ (*G είναι ο θόρυβος Gumbel*)
Θέσε $z_s = \text{embed}(\hat{z}'_{cat})$
Θέσε $z = \text{concat}(z_c, z_s)$

Θέσε $\widehat{\boldsymbol{X}} = \text{decode}(z, start\_token)$

Επέστρεψε $\widehat{\boldsymbol{X}}, \widehat{\boldsymbol{Z}}_{\boldsymbol{cat}}$

6.3.2 Λειτουργία Συμπερασμού

Για τη λειτουργία συμπερασμού απομακρύνουμε τον κωδικοποιητή και παίρνουμε ως είσοδο από το χρήστη απευθείας το one-hot διάνυσμα $\boldsymbol{Z}_{\boldsymbol{cat}}$. Σε αυτή τη λειτουργία δεν υπολογίζουμε παραγώγους και επομένως δεν χρειάζεται να μετασχηματίσουμε το $\boldsymbol{Z}_{\boldsymbol{cat}}$. Δίνουμε αυτό το διάνυσμα στο επίπεδο εμφύτευσης που σχετίζεται με το μουσικό είδος και παίρνουμε από αυτό την εσωτερική αναπαράσταση, z_s, του μουσικού είδους που έχει επιλεγεί. Έπειτα, δειγματολειπτούμε μια τυπική κανονική κατανομή και συνενώνουμε το αποτέλεσμα αυτό με το z_s. Παίρνουμε έτσι το διάνυσμα του κρυφού επιπέδου, z.

Το z αποτελεί την αρχική κατάσταση του αναδρομικού δικτύου του αποκωδικοποιητή, τον οποίο τροφοδοτούμε με το εναρκτήριο λέξημα. Συγκεντρώνουμε τις εξόδους που δίνει ο αποκωδικοποιητής βήμα προς βήμα, έως ότου βρούμε το λέξημα που σηματοδοτεί το τέλος του μουσικού τμήματος. Φτιάχνουμε έτσι την ακολουθία νοτών εξόδου του συστήματος και την μετατρέπουμε σε μορφή MIDI, ώστε να είναι αναγνώσιμη από τον χρήστη.

6.4 Εκπαίδευση του Νευρωνικού Δικτύου

Σε αυτήν την παράγραφο θα εξετάσουμε τον τρόπο εκπαίδευσης του νευρωνικού δικτύου. Το πρώτο βήμα προς αυτήν την κατεύθυνση είναι η εύρεση μιας κατάλληλης συνάρτησης κόστους. Η συνάρτηση κόστους έχει αρκετές συνιστώσες, ώστε να καλύπτει όλες τις πλευρές του συστήματος μας. Έπειτα θα δούμε πως χειριζόμαστε τις υπερπαραμέτρους που προκύπτουν στο σύστημα.

Συνάρτηση Κόστους Πρωταρχικό ρόλο στη συνάρτηση κόστους παίζει το κόστος ανακατασκευής ενός μουσικού τμήματος, δηλαδή μια ποσοτικοποίηση της διαφοράς της εξόδου, $\widehat{\boldsymbol{X}}$, του νευρωνικού δικτύου από την αναμενόμενη έξοδο, που στο πλαίσιο του αυτοκωδικοποιητή ταυτίζεται με την είσοδο, $\boldsymbol{X}$. Χρησιμοποιούμε τη διασταυρούμενη εντροπία αθροιστικά για

κάθε συνιστώσα των $\widehat{\boldsymbol{X}}$ και $\boldsymbol{X}$.

$$L_{r_i} = -\left(\sum_{j=1}^{89} P_j \log(\widehat{P_j}) + \sum_{j=1}^{33} dt_j \log(\widehat{dt_j}) + \sum_{j=1}^{33} D \log(\widehat{D_j})\right) \tag{6.1}$$

$$L_r = \sum_{i=1}^{n_T} L_{r_i} \tag{6.2}$$

Έπειτα, θεωρούμε το κόστος απόκλισης Kullback–Leibler όπως δικαιολογήθηκε στην παράγραφο 4.3.1. Εδώ χρησιμοποιούμε την τυπική κανονική κατανομή, οπότε ο όρος αυτό γίνεται

$$L_{KL} = -\beta \cdot D\left(q_\partial(z \mid x, \mu, \sigma) \parallel \mathcal{N}(0, \boldsymbol{I})\right) \tag{6.3}$$

Χρησιμοποιούμε την παράμετρο β, για να καθορίσουμε το βαθμό στον οποίο η εν λόγω απόκλιση συμμετέχει στη συνάρτηση κόστους. Όσο λιγότερο συμμετέχει, τόσο πιο εύκολη είναι η εκπαίδευση. Εμπειρικά είναι χρήσιμο στην αρχή της εκπαίδευσης, η απόκλιση να μην εμπλέκεται, αφού τόσο ο κωδικοποιητής όσο και ο αποκωδικοποιητής δεν είναι ακόμα καλά εκπαιδευμένοι και η παρουσία του εν λόγω όρου δυσχεραίνει τη διαδικασία. Χρησιμοποιούμε τη μέθοδο της ανόπτησης πάνω στο β και αυξάνουμε σταδιακά τη συνεισφορά του όρου L_{KL} στη συνάρτηση κόστους.

Οι αυτοκωδικοποιητές χρονοσειρών είναι ιδιαίτερα ευάλωτοι σε ένα πρόβλημα που αποκαλείται posterior collapse [14]. Σε αυτήν την περίπτωση και λόγω της αποτελεσματικότητας των αναδρομικών νευρωνικών δικτύων στην εύρεση αναπαραστάσεων, ο αποκωδικοποιητής αγνοεί το κρυφό επίπεδο και ο όρος KL μηδενίζεται εξαιτίας του μηδενισμού των μέσων τιμών που μοντελοποιούμε. Αυτό δεν είναι επιθυμητό, αφού εν τέλει στηριζόμαστε στο κρυφό επίπεδο για την παραγωγή μουσικών τμημάτων και ο μηδενισμός αυτού το καθιστά άχρηστο. Προς αποφυγή αυτής της δυσμενούς κατάστασης, εισάγουμε έναν ακόμη παράγοντα κόστους, ο οποίος θα αποδίδει ποινές, όταν η διασπορά των μέσων τιμών των κατανομών Gauss που μοντελοποιούμε είναι πολύ μικρή. Ο όρος αυτός ονομάζεται μ-forcing και δίνεται από την ακόλουθη έκφραση:

$$L_\mu = \max\left(0, \beta_\mu - \frac{1}{2N}\sum_{n=1}^{N}\left(\mu^n - \bar{\mu}\right)^T(\mu^n - \bar{\mu})\right) \tag{6.4}$$

Η παράμετρος β_μ είναι υπερπαράμετρος και καθορίζει το επίπεδο στο οποίο θα σταθεροποιηθεί η διασπορά των μέσω τιμών.

Τέλος, έχουμε το κόστος που προκύπτει από την εσφαλμένη πρόβλεψη του μουσικού είδους, το οποίο εκφράζουμε πάλι σε όρους διασταυρούμενης εντροπίας:

$$L_s = -\sum_{s=1}^{S} z_{cat_s} \log \widehat{z_{cat_s}} \tag{6.5}$$

Η τελική συνάρτηση κόστους του νευρωνικού δικτύου είναι το άθροισμα όλων των προηγούμενων όρων

$$L(\hat{\boldsymbol{y}}, \boldsymbol{y}) = L_r + L_{KL} + L_\mu + L_s \tag{6.6}$$

και χρησιμοποιείται από τον αλγόριθμο βελτιστοποίησης για την εύρεση των βέλτιστων εσωτερικών παραμέτρων του δικτύου.

Υπερπαράμετροι Οι υπερπαράμετροι είναι όλες εκείνες οι παράμετροι του συστήματος που δεν είναι εκπαιδεύσιμες, δηλαδή οι βέλτιστες τιμές τους δεν βρίσκονται αυτόματα από τον αλγόριθμο βελτιστοποίησης κατά την εκπαίδευση αλλά επιλέγονται από εμάς και μένουν σταθερές κατά την εκπαίδευση. Είναι αναγκαίο να γνωρίζουμε επακριβώς ποιες είναι οι υπερπαράμετροι του συστήματος, αφού η επιλογή τους επηρεάζει δραστικά τη διαδικασία εκπαίδευσης και κατ' επέκταση την επίδοση του νευρωνικού δικτύου. Οι υπερπαράμετροι του συστήματος παρουσιάζονται στον πίνακα 6.1.

Υπερπαράμετρος	Περιγραφή	Πιθανές Τιμές
rnn_dim	Διάσταση των αναδρομικών δικτύων	128, 256, 512
rnn_type	Τύπος αναδρομικών δικτύων	LSTM, GRU
enc_dropout	Dropout για τον κωδικοποιητή	0.4, 0.5, 0.6, 0.7
dec_dropout	Dropout για τον αποκωδικοποιητή	0.4, 0.5, 0.6, 0.7
cont_dim	Διάσταση κρυφού επιπέδου z	20, 50, 120, 200, 400
mu_force	Παράμετρος β_μ του όρου μ-forcing	1.3
t_gumbel	Θερμοκρασία Gumbel	0.0005, 0.001, 0.02, 0.1
style_dim	Διάσταση επιπέδου εμφύτευσης του είδους	20, 80, 150, 300
beta_anneal	Βήματα ανόπτησης της απόκλισης KL	1000, 2500, 5000
lr	Αρχικός ρυθμός μάθησης	5e-4, 5.5e-4, 6e-4, 8e-4, 1e-3
decay	Ρυθμός πτώσης του ρυθμού μάθησης	0.85, 0.93, 0.95, 0.97

Πίνακας 6.1: *Υπερπαράμετροι του νευρωνικού δικτύου*

Η βελτιστοποίηση των υπερπαραμέτρων του συστήματος συνίσταται στην εξαντλητική αναζήτηση του χώρου των πιθανών τιμών τους. Κάτι τέτοιο, δυστυχώς, είναι υπολογιστικά ανέφικτο, γι' αυτό χρησιμοποιούμε τη μέθοδο Hyperband [15]. Αυτή η μέθοδος αποδίδει τιμές στις υπερπαραμέτρους αρχικά με τυχαίο τρόπο και εκπαιδεύει τα διαφορετικά μοντέλα για λίγες εποχές. Στη συνέχεια, διατηρεί τα μοντέλα που απέδωσαν καλύτερα και τα εκπαιδεύει για περισσότερες εποχές. Με αυτόν τον τρόπο, παίρνουμε εν τέλει μια προσέγγιση της βέλτιστης επιλογής υπερπαραμέτρων. Αξίζει να σημειώσουμε ότι η μέθοδος Hyperband αποδίδει εκθετικά περισσότερους πόρους (όπως ο χρόνος εκπαίδευσης) στα καλύτερα μοντέλα, γεγονός το οποίο εν μέρει αιτιολογεί την επίδοση αυτής της μεθόδου στη βελτιστοποίηση των υπερπαραμέτρων.

Υπερεκπαίδευση (overfitting) Η υπερεκπαίδευση είναι μια κατάσταση κατά την οποία το νευρωνικό δίκτυο έχει μάθει τόσο τα δεδομένα εισόδου, όσο και τον θόρυβο που κρύβεται σε αυτά, με αποτέλεσμα να αδυνατεί να γενικεύσει σε πρωτοφανή δεδομένα και να μειώνεται κατ' επέκταση η επίδοση του. Στο σύστημα της παρούσας εργασίας χρησιμοποιούμε τρία αντίμετρα, για να αποφύγουμε την υπερεκπαίδευση.

- Τυχαιοποιούμε τη σειρά με την οποία παρουσιάζουμε τα δεδομένα εισόδου στο νευρωνικό δίκτυο. Με αυτόν τον τρόπο βοηθάμε τη διαδικασία ελαχιστοποίησης της συνάρ-

τησης κόστους και αποφεύγουμε την παγίδευση του αλγορίθμου βελτιστοποίησης σε τοπικά ελάχιστα.

- Αλλοιώνουμε ορισμένους τόνους σε κάθε μουσικό τμήμα - ουσιαστικά εισάγουμε θόρυβο στα δεδομένα. Έτσι, υποχρεώνουμε το νευρωνικό δίκτυο να μάθει γενικότερες αναπαραστάσεις των δεδομένων με αποτέλεσμα να γενικεύει καλύτερα.

- Χρησιμοποιούμε δύο επίπεδα dropout, ένα στον κωδικοποιητή και ένα στον αποκωδικοποιητή. Τα επίπεδα αυτά αποκλείουν από την εκπαίδευση ορισμένους νευρώνες του δικτύου με τυχαίο τρόπο και έτσι εξαναγκάζουν όλους τους νευρώνες του δικτύου να εμπλακούν στη διαδικασία της μάθησης.

Κεφάλαιο 7

Πειραματικά Αποτελέσματα

7.1 Μετρικές Αξιολόγησης του Συστήματος

Ξεκινούμε την πραγματεία των πειραμάτων μας αρχικά με τη μελέτη των μετρικών αξιολόγησης του συστήματος. Η εγκαθίδρυση μιας κατάλληλης σειράς μετρικών για την παρακολούθηση του συστήματος είναι απαραίτητη, ώστε να έχουμε σε κάθε στιγμή μια ξεκάθαρη εικόνα της αποτελεσματικότητάς του. Όπως αναλύουμε στη συνέχεια, διαχωρίζουμε τις μετρικές σε δύο κατηγορίες ανάλογα με τη χρήση τους. Από τη μια πλευρά έχουμε τις μετρικές τις οποίες υπολογίζουμε τοπικά πάνω σε ένα μουσικό τμήμα και από την άλλη έχουμε όσες εξάγονται βάσει ενός συνόλου τμημάτων που ανήκουν στο ίδιο μουσικό είδος.

Οι ακόλουθες δύο μετρικές χρησιμοποιούνται πάνω σε ένα μουσικό τμήμα και προσδιορίζουν άμεσα την αποτελεσματικότητα του νευρωνικού μας δικτύου. Οι μετρικές αυτές υπολογίζονται σε κάθε βήμα της εκπαίδευσης και για κάθε δείγμα εισόδου.

Συνάρτηση κόστους Η συνάρτηση κόστους αποτελεί την πρωταρχική μετρική που χρησιμοποιεί το σύστημά μας κατά την εκπαίδευση. Όσο πιο μικρή είναι η τιμή της συνάρτησης αυτής, τόσο καλύτερα αναμένουμε να πηγαίνει το νευρωνικό δίκτυο στην εργασία που εκτελεί, δηλαδή στη σύνθεση μουσικής. Όπως αναλύσαμε στην παράγραφο 6.4, η συνάρτηση κόστους είναι η ακόλουθη:

$$L = L_r + L_{KL} + L_\mu + L_s$$

Ακρίβεια Χρησιμοποιούμε αυτή τη μετρική για να αξιολογήσουμε τόσο την επιτυχία στην ανακατασκευή του μουσικού τμήματος, όσο και την ορθότητα στην πρόβλεψη του μουσικού είδους του. Η ακρίβεια δίνεται από τον παρακάτω τύπο:

$$\text{Ακρίβεια} = \frac{\text{Πλήθος σωστών προβλέψεων}}{\text{Συνολικό πλήθος προβλέψεων}}$$

Η ακρίβεια όσον αφορά στην πρόβλεψη του μουσικού είδους είναι εύκολα κατανοητή, αφού πρόκειται για ένα τυπικό πρόβλημα ταξινόμησης σε πολλές κλάσεις. - εν προκειμένω έχουμε μόνο δύο κατηγορίες, τα μουσικά τμήματα του συνόλου JSB και αυτά του NMD. Η ακρίβεια σχετικά με την ανακατασκευή του μουσικού περιεχομένου είναι πιο περίπλοκη. Σε αυτήν την περίπτωση επιτυγχάνουμε 100% ακρίβεια όταν $\hat{\boldsymbol{X}} = \boldsymbol{X}$. Αυτό πρακτικά επιβάλλει την

ακόλουθη απαίτηση για τις εξόδους του νευρωνικού δικτύου:

$$\begin{cases} \widehat{\boldsymbol{X_1}} = \boldsymbol{X_1} & \Leftrightarrow (\widehat{P_1}, \widehat{dt_1}, \widehat{D_1}) = (P_1, dt_1, D_1) \\ & \vdots \\ \widehat{\boldsymbol{X_{n_T}}} = \boldsymbol{X_{n_T}} & \Leftrightarrow (\widehat{P_{n_T}}, \widehat{dt_{n_T}}, \widehat{D_{n_T}}) = (P_{n_T}, dt_{n_T}, D_{n_T}) \end{cases}$$

Είναι προφανές πως η μεγιστοποίηση της ακρίβειας που σχετίζεται με την ανακατασκευή είναι πολύ δύσκολη, αν όχι ανέφικτη. Αξίζει, όμως, να τονίσουμε εδώ ότι ο κύριος σκοπός μας δεν είναι η μεγιστοποίηση της ακρίβειας, αλλά η επίτευξη ενός επιπέδου σύνθεσης, ώστε τα μουσικά τμήματα που παράγονται να είναι αρεστά και εύηχα.

Οι επόμενες δύο μετρικές εξάγονται βάσει ενός συνόλου μουσικών τμημάτων που ανήκουν σε ένα είδος και προσδιορίζουν εν μέρει την ταυτότητά του. Η χρησιμότητά τους γίνεται φανερή κατά τη μαζική παραγωγή μουσικών τμημάτων από το σύστημα και τη μετέπειτα σύγκριση των τιμών των μετρικών από το παραχθέν και το αρχικό σύνολο.

Μοναδικοί τόνοι Αυτή η μετρική υπολογίζει το πλήθος των μοναδικών μουσικών τόνων που εμφανίζονται σε ένα κομμάτι. Ουσιαστικά πρόκειται για ένα μέτρο της ποικιλομορφίας ενός κομματιού. Η χρησιμότητά της μετρικής προκύπτει από το γεγονός ότι τα διαφορετικά μουσικά είδη έχουν και διαφορετικές ποικιλομορφίες ως προς τους μοναδικούς μουσικούς τόνους. Για παράδειγμα, σε ένα κομμάτι κλασσικής μουσικής συναντάμε κατά μέσο όρο πολύ μικρότερο πλήθος μοναδικών τόνων απ' ότι σε ένα κομμάτι μουσικής τζαζ.

Πολυφωνία Η πολυφωνία αποτελεί μια πολύ ενδιαφέρουσα μετρική που χρησιμοποιείται στο [16]. Όπως γίνεται φανερό και από το όνομα αυτής της μετρική, η πολυφωνία μας δίνει μια εκτίμηση για το πόσες νότες ηχούν ταυτόχρονα σε ένα μουσικό κομμάτι. Αυτό είναι πολύ σημαντικό, καθώς μπορούμε βάσει αυτού να εξάγουμε ενδιαφέροντα συμπεράσματα για το σύνολο δεδομένων μας. Η πολυφωνία n νοτών δίνεται από τον ακόλουθο τύπο:

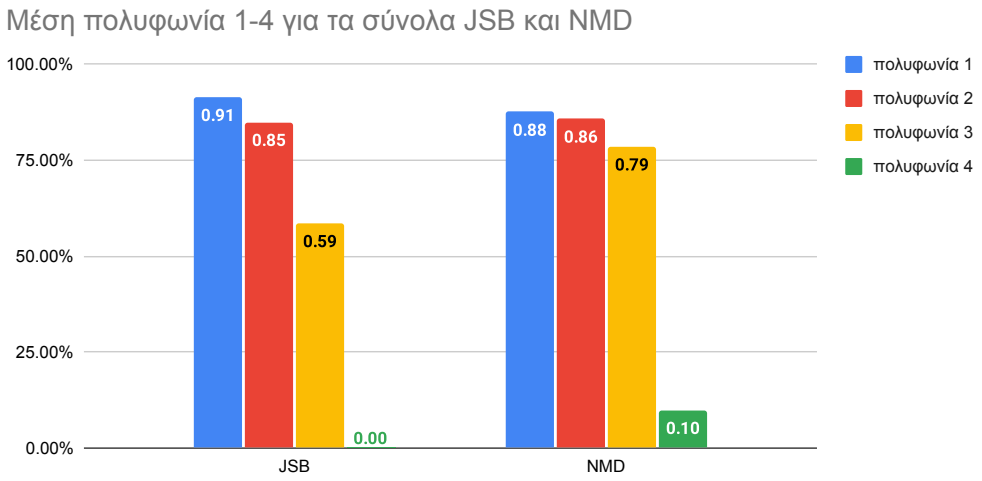
$$\text{πολυφωνία}(n) = \frac{\text{χρονικά βήματα όπου περισσότερες από } n \text{ νότες ηχούν ταυτόχρονα}}{\text{συνολικός αριθμός χρονικών βημάτων}}$$

7.2 Μετρικές για τα Σύνολα Δεδομένων

Σε αυτήν την παράγραφο παρουσιάζουμε μια ποσοτική περιγραφή των συνόλων δεδομένων που χρησιμοποιούμε. Υπενθυμίζουμε ότι χρησιμοποιούμε τα σύνολα JSB και NMD, τα κομμάτια των οποίων έχουμε χωρίσει σε τμήματα των $k = 4$ και $k = 8$ μουσικών μέτρων. Τα διάφορα ποσοτικά στοιχεία καθώς και οι μετρικές που αφορούν στα σύνολα αυτά, παρουσιάζονται στους πίνακες 7.1, 7.2, 7.3 και στο σχήμα 7.1.

Σύνολο Δεδομένων	Πλήθος Κομματιών	Χαμηλότερος Τόνος	Ψηλότερος Τόνος
JSB	382	43 (G2)	96 (C7)
NMD	1037	36 (C2)	88 (E6)

Πίνακας 7.1: *Χαρακτηριστικές μετρικές των συνόλων δεδομένων*



Σχήμα 7.1: *Πολυφωνία των συνόλων δεδομένων*

Το σύνολο NMD έχει σχεδόν τρεις φορές περισσότερα κομμάτια, γι' αυτό κατά τη διαδικασία της εκπαίδευσης δεν χρησιμοποιούμε εξ ολοκλήρου αυτό το σύνολο δεδομένων, αλλά παίρνουμε ένα ποσοστό αυτού, τέτοιο ώστε τα δύο σύνολα να είναι εξισορροπημένα.

Στα πλαίσια της εργασίας μας εξετάζουμε τις πολυφωνίες για $n = 1, 2, 3, 4$, δηλαδή για ταυτόχρονη ήχηση $2, 3, 4$ ή 5 νοτών. Από τη γραφική παράσταση 7.1 μπορούμε να εξάγουμε πολλά συμπεράσματα, τα οποία συμφωνούν με τις λεπτομέρειες των συνόλων δεδομένων που δώσαμε στο κεφάλαιο 5. Πιο συγκεκριμένα, για το JSB η πολυφωνία 4 είναι μηδενική, επομένως ποτέ δεν συνηχούν 5 νότες. Πράγματι τα έργα αυτά είναι γραμμένα για τέσσερις φωνές. Σχετικά με το NMD, παρατηρούμε ότι οι τρεις πρώτες πολυφωνίες είναι σχεδόν ίσες μεταξύ τους, ενώ η πολυφωνία 4 είναι πολύ χαμηλή. Άρα κυρίως σε αυτά τα κομμάτια ηχούν ταυτόχρονα 4 νότες, γεγονός το οποίο συνάδει με τη μουσική δομή αυτού του συνόλου, ότι δηλαδή, υπάρχει μια μελωδία που συνοδεύεται από μια τρίφωνη συγχορδία.

Το μέσο πλήθος νοτών ανά τμήμα - ουσιαστικά η πυκνότητα νοτών - αποτελεί μια έκφραση της δυσκολίας του συνόλου δεδομένων ως προς την εκμάθηση. Όσο περισσότερες νότες ανά τμήμα έχουμε τόσες πιο μακρές είναι οι χρονοσειρές των διανυσμάτων εισόδου $\boldsymbol{X}$. Προφανώς, για $k = 8$ αναμένουμε διπλασιασμό αυτής της μετρικής - τον οποίο πράγματι παρατηρούμε - από την περίπτωση $k = 4$. Αξίζει να σημειώσουμε το γεγονός ότι οι μετρικές αυτές για τα δύο σύνολα δεδομένων είναι κοντινές και αυτό εξασφαλίζει ότι η δυσκολία των δύο συνόλων είναι παραπλήσια.

Σύνολο Δεδομένων	Πλήθος τμημάτων	Μέσο πλήθος νοτών ανά τμήμα	Μέσο πλήθος μοναδικών τόνων
JSB	1820	30,9	17,1
NMD	9523	32,9	13,8

Πίνακας 7.2: *Μετρικές πάνω στο σύνολο των δεδομένων για $k = 4$*

Σύνολο Δεδομένων	Πλήθος τμημάτων	Μέσο πλήθος νοτών ανά τμήμα	Μέσο πλήθος μοναδικών τόνων
JSB	1019	55,3	20,0
NMD	4789	65,5	16,3

Πίνακας 7.3: *Μετρικές πάνω στο σύνολο των δεδομένων για $k = 8$*

7.3 Σύνθεση Μουσικής

Σε αυτήν την παράγραφο παρουσιάζουμε τα πειραματικά αποτελέσματα που πήραμε από την εκπαίδευση και τη χρήση του νευρωνικού δικτύου για τη σύνθεση μουσικών τμημάτων μήκους 4 και 8 μέτρων. Και στις δύο περιπτώσεις, εκτελούμε πρώτα τον αλγόριθμο Hyperband με μέγιστο αριθμό 250 εποχών, ώστε να βρούμε τις βέλτιστες υπερπαραμέτρους και έπειτα εκπαιδεύουμε το βέλτιστο μοντέλο για 400 εποχές. Ο πίνακας 7.4 συνοψίζει τις βέλτιστες τιμές για τις υπερπαραμέτρους του συστήματός μας.

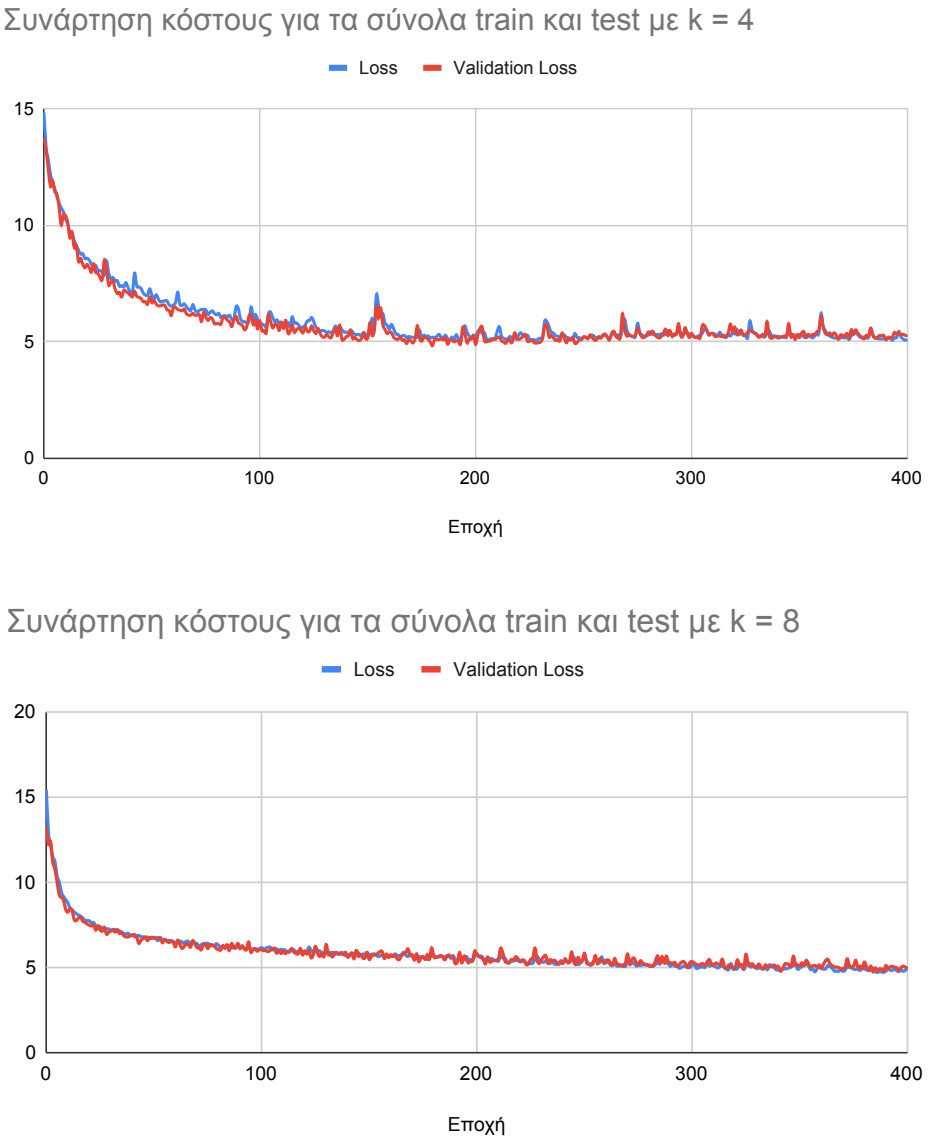
Υπερπαράμετρος	Βέλτιστη τιμή για $k = 4$	Βέλτιστη τιμή για $k = 8$
rnn_dim	512	512
rnn_type	GRU	GRU
enc_dropout	0.4	0.5
dec_dropout	0.4	0.6
cont_dim	50	200
mu_force	1.3	1.3
t_gumbel	0.1	0.02
style_dim	80	300
beta_anneal	1800	2500
lr	5.5e-4	1.0e-3
decay	0.95	0.85

Πίνακας 7.4: *Βέλτιστες τιμές των υπερπαραμέτρων του νευρωνικού δικτύου*

Οι γραφικές παραστάσεις 7.2, 7.3 και 7.4 συνοψίζουν τις μετρικές με τις οποίες αξιολογούμε το σύστημα, τόσο στα δεδομένα εκπαίδευσης, όσο και στα δεδομένα ελέγχου. Υπενθυμίζουμε ότι δεν χρησιμοποιούμε ξεχωριστά σύνολα test και validation, έτσι οι δύο όροι, όταν χρησιμοποιούνται, έχουν την ίδια σημασία. Για την εξαγωγή της πολυφωνίας χρησιμοποιούμε το νευρωνικό δίκτυο, για να παράγουμε 50 δείγματα από κάθε μουσικό είδος. Με αυτόν τον τρόπο δημιουργούμε το σύνολο που απαιτείται για να πάρουμε μια κατάλληλη μέτρηση της μέσης τιμής της πολυφωνίας.

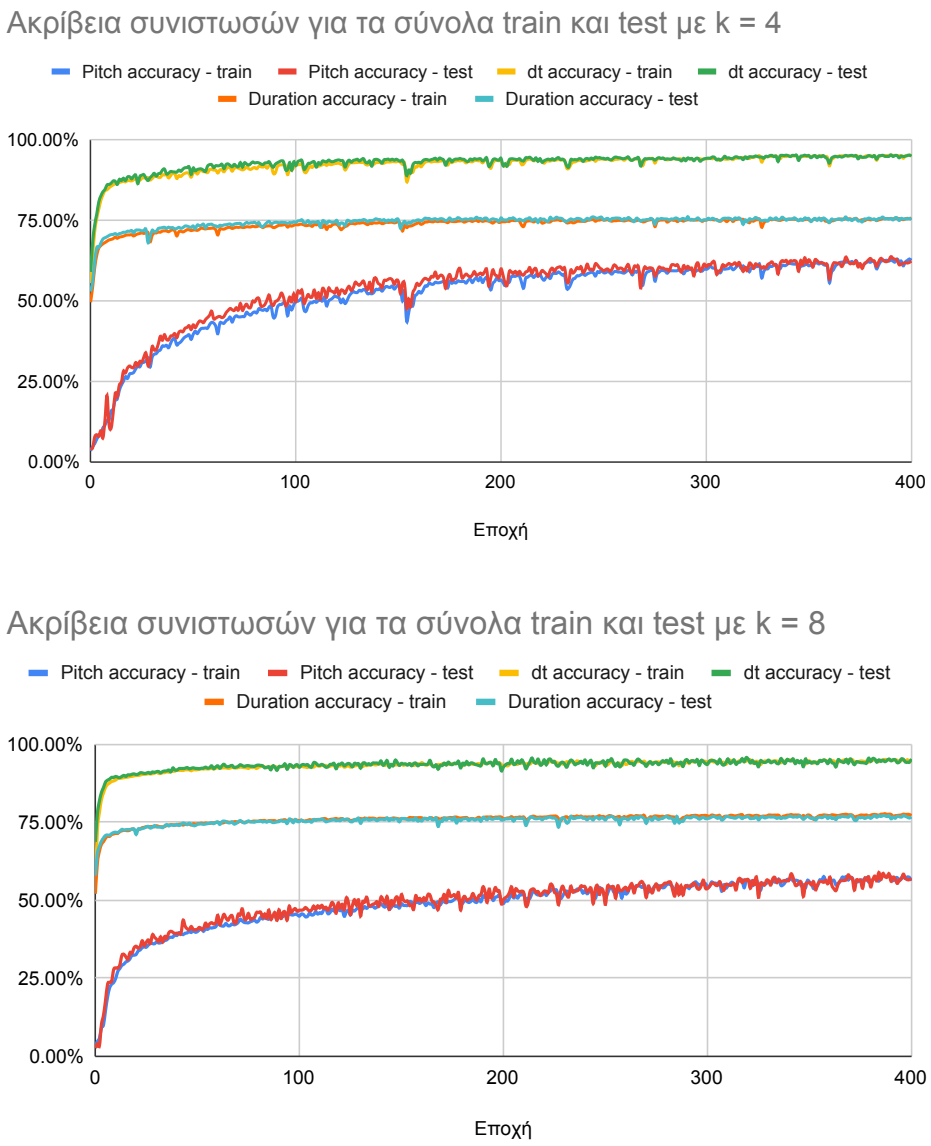
Παρατηρούμε ότι η συνάρτηση κόστους μειώνεται έως ότου φτάσει σε ένα πλατώ. Αντίστοιχη συμπεριφορά εμφανίζει και η γραφική παράσταση της ακρίβειας καθεμιάς από τις τρεις συνιστώσες του διανύσματος νοτών. Πιο συγκεκριμένα, η ακρίβεια αυξάνεται έντονα μέχρι μια τιμή, έπειτα από την οποία η κλίση της καμπύλης της μειώνεται. Παρατηρούμε ακόμη ότι η ακρίβεια ανακατασκευής της συνιστώσας του μουσικού τόνου παίρνει τις χαμηλότερες τιμές. Αυτό είναι αναμενόμενο, αφού ο μουσικός τόνος αποτελεί την πιο σύνθετη συνιστώσα από τις τρεις. Πράγματι, η συνιστώσα P μπορεί να πάρει 89 διαφορετικές τιμές

ενώ οι άλλες δύο συνιστώσες μπορούν να πάρουν μόνο 33 - όπως έχουμε αναφέρει στο κεφάλαιο 5. Επιπλέον, τα μουσικά κομμάτια εμφανίζουν πολύ μεγαλύτερη ποικιλία σε τόνους απ' ότι σε χρονικές αξίες, γεγονός το οποίο επιβάλλει την εντατική χρήση μεγάλου μέρους του συνόλου των επιτρεπτών τόνων, σε αντίθεση με το σύνολο των επιτρεπτών χρονικών αξιών, το οποίο εν μέρει υποχρησιμοποιείται. Αυτό αποτελεί εμπόδιο στην αποδοτική εκμάθηση της συνιστώσας P από το νευρωνικό δίκτυο και αποτυπώνεται στις γραφικές παραστάσεις. Αρκεί, όμως, αυτή η επίδοση ως προς την ακρίβεια για την ικανοποιητική παραγωγή μουσικής στα πλαίσια αυτής της εργασίας.



Σχήμα 7.2: *Εξέλιξη της συνάρτησης κόστους για τα σύνολα train και test ως προς τις εποχές*

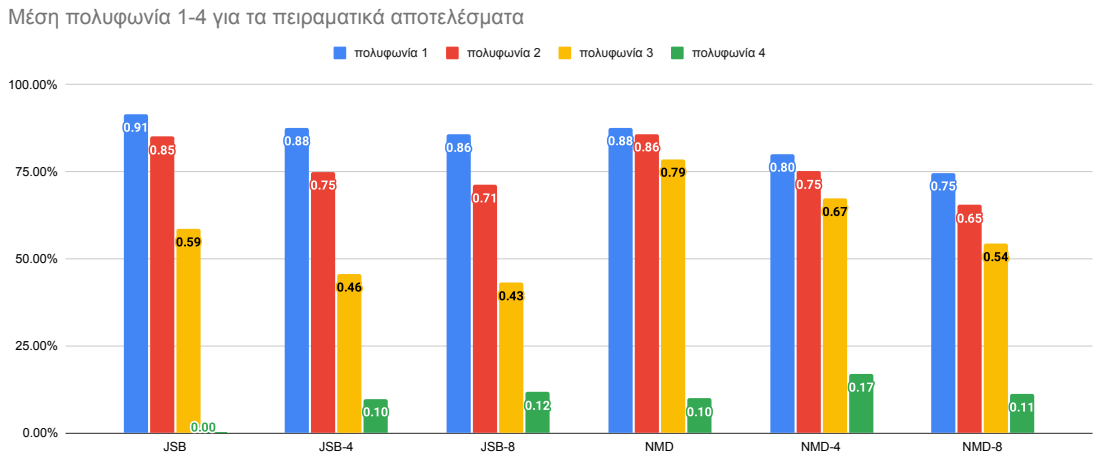
Αξίζει να σημειώσουμε ότι σε κάθε περίπτωση οι αντίστοιχες καμπύλες για τα δεδομένα εκπαίδευσης και ελέγχου είναι παραπλήσιες, γεγονός από το οποίο συμπεραίνουμε ότι το σύστημά μας δεν πάσχει από υπερεκπαίδευση. Αυτό φυσικά οφείλεται στα αντίμετρα που έχουμε χρησιμοποιήσει, όπως η χρήση dropout και η επιβολή τυχαίας σειράς στα δεδομένα.



Σχήμα 7.3: *Εξέλιξη της ακρίβειας των τριών συνιστωσών του διανύσματος νοτών για τα σύνολα train και test ως προς τις εποχές*

Οι τιμές της πολυφωνίας, όπως φαίνονται στο σχήμα 7.4, συμφωνούν με τις τιμές που πήραμε για ολόκληρο το σύνολο δεδομένων τόσο ως προς τις απόλυτες τιμές, όσο και ως προς τα ποιοτικά χαρακτηριστικά. Πιο συγκεκριμένα, η πολυφωνία 4 του JSB είναι σχεδόν μηδενική, ενώ οι τρεις πρώτες πολυφωνίες του NMD είναι παραπλήσιες και η τέταρτη πολύ χαμηλότερη.

Στον πίνακα 7.5 παραθέτουμε μια σύνοψη των αριθμητικών αποτελεσμάτων μας. Αξίζει να τονίσουμε ότι σε κάθε μια από τις δύο περιπτώσεις εκπαίδευσης, η ακρίβεια πρόβλεψης του μουσικού είδους ήταν σχεδόν 100%, γεγονός που δείχνει την ευκολία στην εκμάθηση αυτής της πτυχής των δεδομένων μας, αλλά και βάσει του οποίου μπορούμε να υποστηρίξουμε την ξεκάθαρη διαφορά μεταξύ των συνθέσεων των δύο ειδών όπως θα δούμε στη συνέχεια.

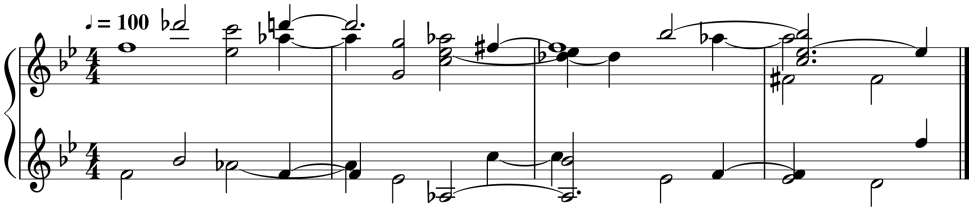


Σχήμα 7.4: *Πολυφωνία 1 έως 4 για τα αρχικά σύνολα δεδομένων και τα πειραματικά αποτελέσματα για μήκη 4 και 8 μέτρων*

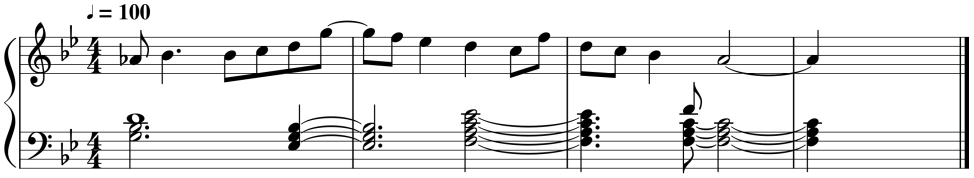
Μήκος τμήματος	Κόστος	KL	Ακρίβεια P (τόνος)	Ακρίβεια dt	Ακρίβεια D (διάρκεια)
4	4,83	0,89	0,66	0,96	0,76
8	4,91	1,03	0,61	0,96	0,77

Πίνακας 7.5: *Σύνοψη πειραματικών αποτελεσμάτων*

Τέλος, στα σχήματα 7.5 και 7.6 παρουσιάζουμε δύο μουσικές συνθέσεις 4 και 8 μέτρων αντίστοιχα που παρήχθησαν από το νευρωνικό δίκτυο. Οι διαφορές στις δύο συνθέσεις είναι εμφανείς από τα σχήματα και συνάδουν με τις περιγραφές που δώσαμε στο κεφάλαιο 5 για τα δύο σύνολα δεδομένων. Περισσότερες συνθέσεις παραθέτουμε στο παράρτημα Α΄.1.

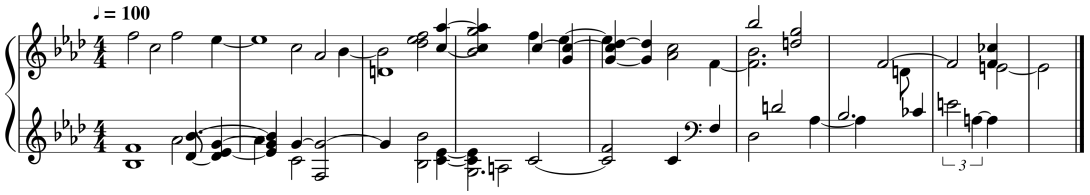


(α΄) *Σύνθεση με βάση το σύνολο JSB*

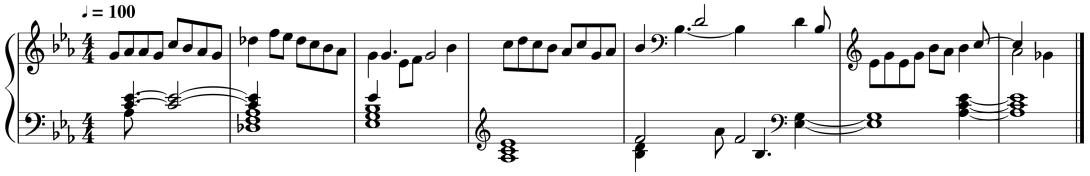


(β΄) *Σύνθεση με βάση το είδος NMD*

Σχήμα 7.5: *Μουσική σύνθεση του νευρωνικού δικτύου για 4 μέτρα*



(α') *Σύνθεση με βάση το σύνολο JSB*



(β') *Σύνθεση με βάση το είδος NMD*

Σχήμα 7.6: *Μουσική σύνθεση του νευρωνικού δικτύου για 8 μέτρα*

7.4 Γραμμική Παρεμβολή στο Κρυφό Επίπεδο του VAE

Το κρυφό επίπεδο, z, του Variational αυτοκωδικοποιητή, όπως έχουμε αναφέρει, διαμορφώνει έναν χώρο από τον οποίο λαμβάνουμε ένα δείγμα και το μετατρέπουμε σε μουσικό τμήμα μέσω του αποκωδικοποιητή. Σε αυτήν την παράγραφο μελετάμε μια επέκταση αυτής της διαδικασίας, τη γραμμική παρεμβολή μεταξύ δύο σημείων στο χώρο z, δηλαδή τη λήψη δειγμάτων ανά τακτά διαστήματα και την μετ' έπειτα αποκωδικοποιήσή τους. Σε αυτό το πείραμα χρησιμοποιούμε εξ ολοκλήρου τη διάταξη του αυτοκωδικοποιητή. Πιο αναλυτικά, επιλέγουμε δύο μουσικά τμήματα από τα δεδομένα εισόδου, έστω τα $\boldsymbol{X_a}$ και $\boldsymbol{X_b}$, και εφαρμόζουμε τη διαδικασία της κωδικοποίησης, για να πάρουμε τις εικόνες τους στο κρυφό επίπεδο, z_a και z_b. Το τελικό σημείο, z, το οποίο τροφοδοτούμε στον αποκωδικοποιητή και βάσει του οποίου παίρνουμε το μουσικό αποτέλεσμα είναι το

$$z = (1 - a) \cdot z_a + a \cdot z_b$$

όπου $a \in [0, 1]$ είναι η παράμετρος της γραμμικής παρεμβολής και ορίζει το ποσοστό συμμετοχής των χαρακτηριστικών κάθε όρου στο τελικό αποτέλεσμα. Μπορούμε να ορίσουμε το πλήθος των δειγμάτων που θέλουμε ως N και τότε το n-οστό δείγμα δίνεται από τη σχέση

$$z_n = (1 - \frac{n}{N-1}) \cdot z_a + \frac{n}{N-1} \cdot z_b \tag{7.1}$$

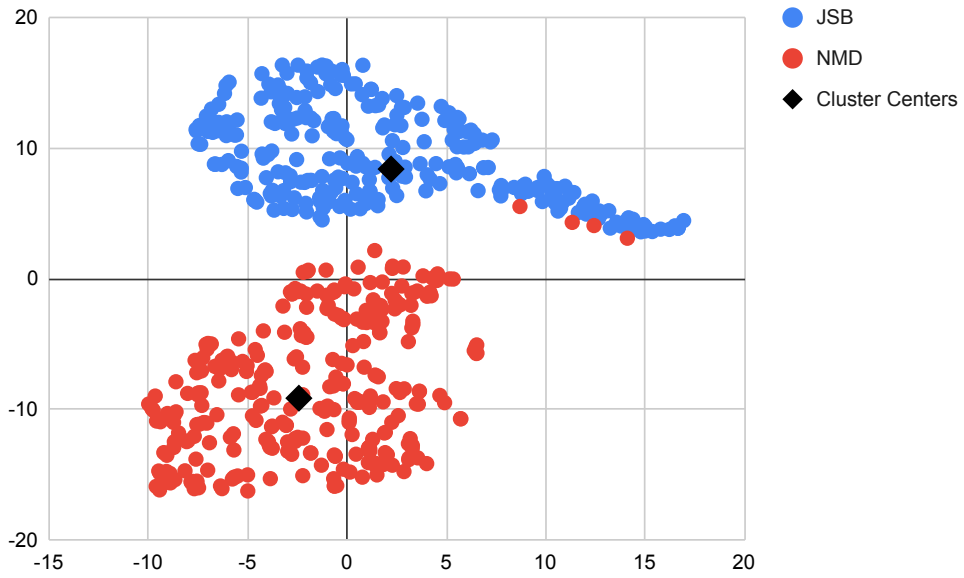
για κάθε $n = 0, 1, \ldots N - 1$.

Με βάση την εξίσωση 7.1, χρησιμοποιούμε το εκπαιδευμένο σύστημα, για να παράγουμε έξι μουσικά τμήματα μήκους 8 μέτρων. Ως $\boldsymbol{X_a}$ θεωρούμε ένα τυχαίο δείγμα του συνόλου JSB και ως $\boldsymbol{X_b}$ ένα τυχαίο δείγμα του συνόλου NMD. Έτσι, θα παρακολουθήσουμε την εξέλιξη των δειγμάτων μεταξύ των δύο μουσικών ειδών. Το αποτέλεσμα της γραμμικής παρεμβολής παρουσιάζεται στο παράρτημα Α΄.2.

7.5 Μελέτη Ανεξαρτησίας των Συνιστωσών του VAE

Μια ενδιαφέρουσα μελέτη που αξίζει να κάνουμε είναι να εξετάσουμε την ανεξαρτησία μεταξύ των δύο συνιστωσών z_c και z_s του νευρωνικού δικτύου. Με τον όρο ανεξαρτησία εννοούμε κατά πόσο μπορούμε να χρησιμοποιήσουμε τη μία συνιστώσα για να κάνουμε προβλέψεις για την άλλη. Στα πλαίσια αυτής της εργασίας θα εξετάσουμε κατά πόσο η πληροφορία που σχετίζεται με το μουσικό είδος ενός κομματιού - και άρα αποτελεί αρμοδιότητα της συνιστώσας z_s - αποτυπώνεται στη συνιστώσα z_c. Η ανεξαρτησία των συνιστωσών φαίνεται να ενισχύει την επίδοση ενός VAE όπως τονίζεται στα [17], [18]. Τονίζουμε ότι στην παρούσα εργασία δεν έχουμε χρησιμοποιήσει κανέναν παράγοντα που να επιβάλει την ανεξαρτησία μεταξύ των δύο συνιστωσών, αξίζει όμως να την μελετήσουμε.

Προς αυτή τη μελέτη χρησιμοποιούμε τη μέθοδο t-SNE, για να απεικονίσουμε στο επίπεδο τις τιμές που παίρνει η συνιστώσα z_c, όταν τροφοδοτείται με ένα υποσύνολο των δεδομένων εισόδου. Η μέθοδος t-SNE είναι μια διαδικασία μείωσης της διαστατικότητας ενός συνόλου δεδομένων με σκοπό τη διατήρηση των βασικότερων χαρακτηριστικών, δηλαδή όσων χαρακτηριστικών εμφανίζουν τη μικρότερη συσχέτιση. Η μέθοδος t-SNE είναι παρόμοια της PCA, δίνει όμως καλύτερα αποτελέσματα από αυτήν σε βάρος μεγαλύτερης επεξεργαστικής πολυπλοκότητας. Μια λεπτομερής παρουσίαση της μεθόδου t-SNE γίνεται στο [19]. Το αποτέλεσμα της εφαρμογής της μεθόδου επί των δεδομένων μας φαίνεται στο σχήμα 7.7.



Σχήμα 7.7: *Διδιάστατο διάγραμμα της συνιστώσας z_c με τη μέθοδο t-SNE*

Είναι εμφανής ο διαχωρισμός των σημείων σε δύο συστάδες, ανάλογα με το είδος τους. Από αυτό μπορούμε να συμπεράνουμε ότι πράγματι στη συνιστώσα z_c έχει αποτυπωθεί πληροφορία σχετική με το είδος του μουσικού κειμένου, γεγονός το οποίο είναι ανεπιθύμητο. Στα πλαίσιο της εργασίας αυτής, όμως, δεν προχωράμε σε πιο σύνθετες μεθόδους, ικανές να ανεξαρτητοποιήσουν τις δύο συνιστώσες.

Μέρος III

Επίλογος

Κεφάλαιο 8

Επίλογος

8.1 Σύνοψη

Στην παρούσα εργασία ασχοληθήκαμε εκτενώς με την κατασκευή ενός γεννητικού νευρωνικού δικτύου τύπου VAE με σκοπό τη σύνθεση μουσικής. Το ιδιαίτερο χαρακτηριστικό του έργου είναι η ικανότητα παραγωγής μουσικής σύμφωνα με ένα εξωτερικά ορισμένο είδος. Αυτό επέβαλε το διαχωρισμό της δομής του VAE σε δύο συνιστώσες, καθεμία εκ των οποίων είναι αρμόδια για την εκμάθηση μιας πτυχής των δεδομένων. Από τη μια πλευρά η συνιστώσα του μουσικού περιεχομένου ασχολείται με την εκμάθηση μιας αναπαράστασης των νοτών που δίνονται ως είσοδος. Από την άλλη η συνιστώσα του μουσικού είδους προσπαθεί να προβλέψει σωστά το είδος του κομματιού που δόθηκε ως είσοδος. Από κοινού αυτές οι συνιστώσες συνενώνονται και αποτελούν το σύστημα VAE που υλοποιήσαμε.

Τα πειράματά μας περιλαμβάνουν αρχικά τη σύνθεση μουσικής 4 και 8 μέτρων σύμφωνα με τα δύο μουσικά είδη που διαθέτουμε, το JSB και το NMD. Μέσω αυτών διαπιστώνουμε την αποτελεσματικότητα του VAE ως μοντέλο αναπαράστασης πληροφορίας αλλά και παραγωγής νέων δεδομένων. Στη συνέχεια, εκμεταλλευόμαστε την εσωτερική κατανομή πιθανοτήτων που διατηρεί το σύστημά μας και η οποία περιγράφει τα σύνολα δεδομένων και κάνουμε μια γραμμική παρεμβολή μεταξύ δύο σημείων αυτής. Έτσι, παρατηρούμε την εξέλιξη ενός μουσικού κομματιού που ξεκινά από το ένα μουσικό είδος και καταλήγει στο άλλο. Τέλος, κάνουμε μια σύντομη μελέτη της ανεξαρτησίας των δύο συνιστωσών που αποτελούν το VAE. Έχει δειχθεί ότι όταν οι συνιστώσες δεν εμφανίζουν συσχέτιση, τότε το σύστημα τείνει να αποδίδει καλύτερα. Στη δική μας εργασία δεν έχουμε λάβει κάποια πρόληψη έναντι αυτού του προβλήματος κι έτσι παρατηρούμε ότι οι συνιστώσες του συστήματός μας είναι αρκετά εξαρτημένες, με πιθανό αποτέλεσμα τον περιορισμό της αποτελεσματικότητας.

8.2 Μελλοντικές Επεκτάσεις

Το σύστημα που αναπτύξαμε έχει πολλές προοπτικές επεκτάσεων, ορισμένες από τις οποίες αναφέρουμε παρακάτω:

- Ενσωμάτωση περισσότερων συνόλων δεδομένων για περισσότερα μουσικά είδη. Αν και η μελέτη της μουσικής από τη σκοπιά των νευρωνικών δικτύων υστερεί, σε σχέση για παράδειγμα με την επεξεργασία φυσικής γλώσσας, εντούτοις υπάρχουν αρκετά σύνολα δεδομένων που είναι διαθέσιμα και τα οποία μπορούν να χρησιμοποιηθούν

για τον εμπλουτισμό του συστήματος. Σίγουρα, όμως, η αύξηση του συνόλου των δεδομένων απαιτεί και αύξηση των δυνατοτήτων του νευρωνικού δικτύου σε πλήθος νευρώνων αλλά και σε επίπεδα.

- Χρήση τεχνικών για επιβολή ανεξαρτησίας μεταξύ των δύο συνιστωσών, περιεχομένου και μουσικού είδους. Όπως έχουμε αναφέρει, όταν οι συνιστώσες ενός Variational αυτοκωδικοποιητή είναι ανεξάρτητες, τότε η απόδοσή του αυξάνεται. Η επιστημονική κοινότητα έχει αναπτύξει μεθόδους που τροποποιούν τις συνιστώσες ενός τέτοιου αυτοκωδικοποιητή και ενσωματώνουν όρους κανονικοποίησης στο σύστημα με σκοπό την ανεξαρτητοποίηση των συνιστωσών.

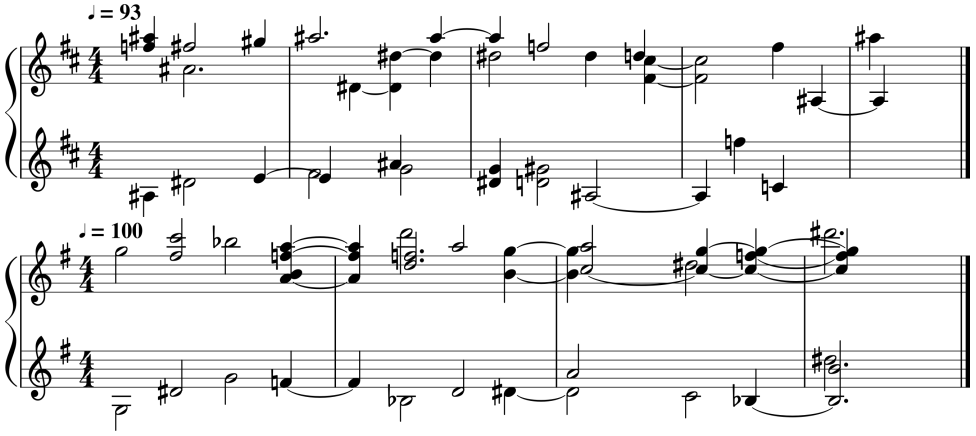
- Χρήση των δικτύων Transformers [20]. Τα δίκτυα αυτά αποτελούν μια πολύ ενδιαφέρουσα εναλλακτική των αναδρομικών νευρωνικών δικτύων και η χρήση τους φαίνεται να είναι πολλά υποσχόμενη. Ειδικό ενδιαφέρον παρουσιάζει ο συνδυασμός αυτών των δικτύων με τα δίκτυα VAE που χρησιμοποιούμε ως επί το πλείστον σε αυτήν την εργασία.

Παραρτήματα

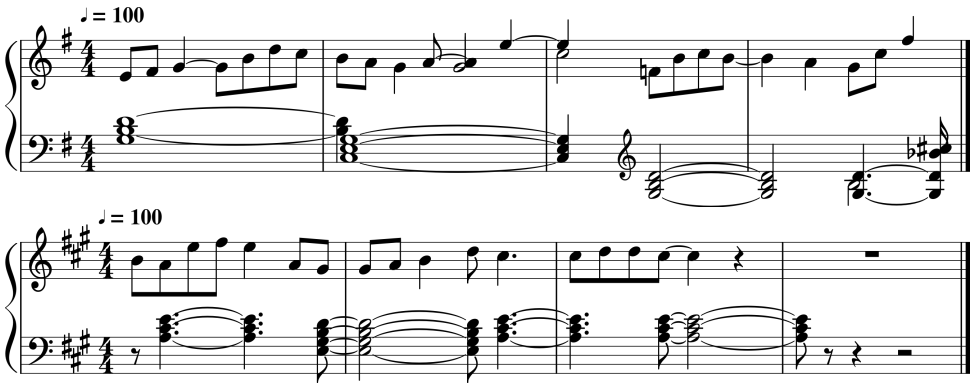
Παράρτημα Α΄

Σχήματα Μουσικών Συνθέσεων

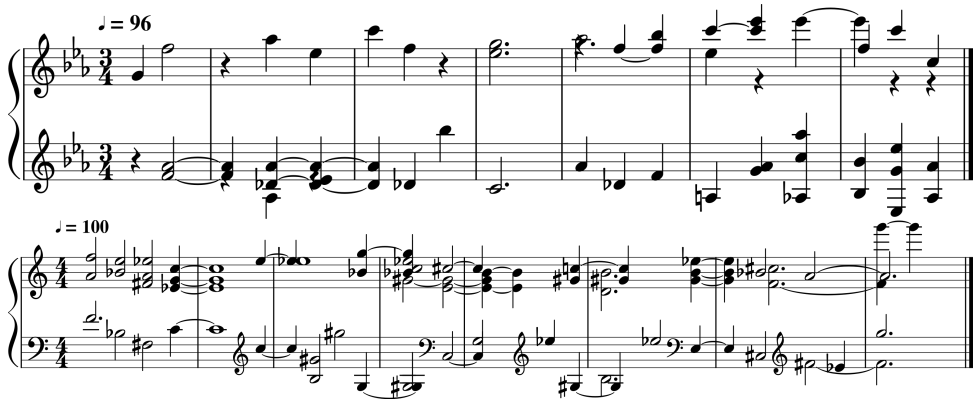
Α΄.1 Μουσικές Συνθέσεις 4 και 8 μέτρων



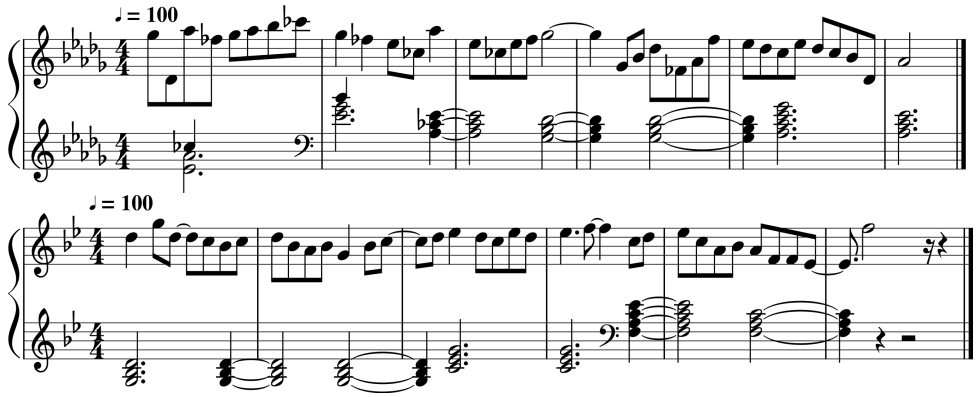
Σχήμα Α΄.1: *Σύνθεση του νευρωνικού δικτύου για 4 μέτρα σύμφωνα με το σύνολο JSB*



Σχήμα Α΄.2: *Σύνθεση του νευρωνικού δικτύου για 4 μέτρα σύμφωνα με το σύνολο NMD*

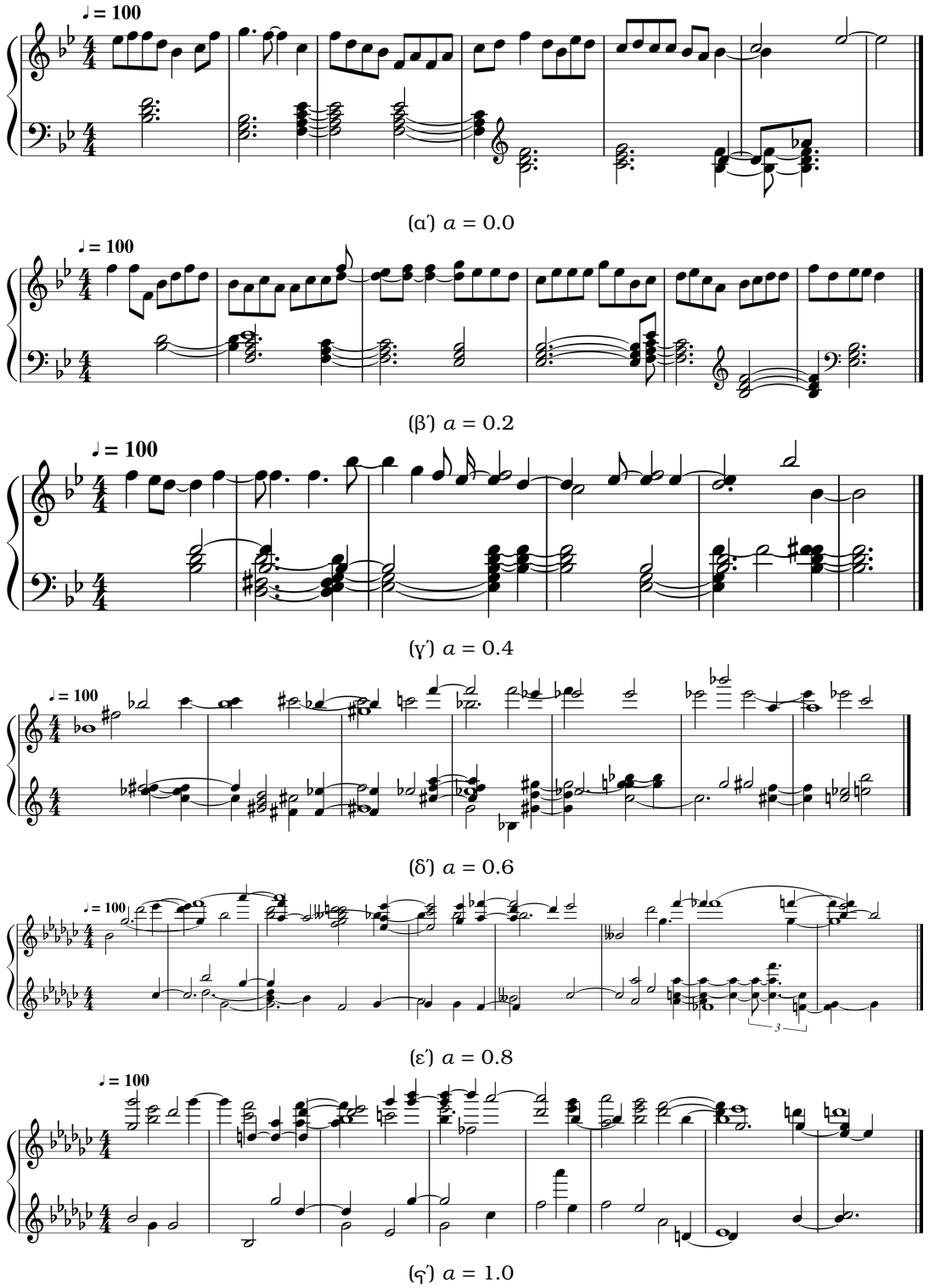


Σχήμα Α΄.3: *Σύνθεση του νευρωνικού δικτύου για 8 μέτρα σύμφωνα με το σύνολο JSB*



Σχήμα Α΄.4: *Σύνθεση του νευρωνικού δικτύου για 8 μέτρα σύμφωνα με το σύνολο NMD*

Α΄.2 Αποτελέσματα Γραμμικής Παρεμβολής



Σχήμα Α΄.5: *Γραμμική παρεμβολή στο κρυφό χώρο μεταξύ ενός δείγματος του συνόλου NMD (a = 0.0) και ενός του JSB (a = 1.0)*

Βιβλιογραφία

[1] Iannis Xenakis. *Formalized Music: Thought and Mathematics in Composition*. Indiana University Press, 1971.

[2] Feynman T. Liang, Mark Gotham, M. Johnson και J. Shotton. *Automatic Stylistic Composition of Bach Chorales with Deep LSTM. ISMIR*, 2017.

[3] Nicolas Boulanger-Lewandowski, Yoshua Bengio και Pascal Vincent. *Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription. arXiv e-prints*, 2012.

[4] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.

[5] Simon S. Haykin. *Neural networks and learning machines*. Pearson Education, Upper Saddle River, NJ, 3η έκδοση, 2009.

[6] Diederik P Kingma και Max Welling. *Auto-Encoding Variational Bayes. arXiv e-prints*, 2013.

[7] Jiawei Zhang. *Gradient Descent based Optimization Algorithms for Deep Learning Models Training. arXiv e-prints*, 2019.

[8] *Derivation of Backpropagation*. `https://www.cs.swarthmore.edu/~meeden/cs81/s10/BackPropDeriv.pdf`. Ημερομηνία πρόσβασης: 22-10-2020.

[9] *David's MIDI Spec*. `https://www.cs.cmu.edu/~music/cmsip/readings/davids-midi-spec.htm`. Ημερομηνία πρόσβασης: 29-10-2020.

[10] Florian Colombo και Wulfram Gerstner. *BachProp: Learning to Compose Music in Multiple Styles. arXiv e-prints*, 2018.

[11] Y. Q. Lim, C. S. Chan και F. Y. Loo. *Style-Conditioned Music Generation. 2020 IEEE International Conference on Multimedia and Expo (ICME)*, 2020.

[12] Ilya Sutskever, Oriol Vinyals και Quoc V Le. *Sequence to sequence learning with neural networks. Advances in neural information processing systems*, σελίδες 3104–3112, 2014.

[13] Eric Jang, Shixiang Gu και Ben Poole. *Categorical Reparameterization with Gumbel-Softmax. arXiv e-prints*, 2016.

[14] Chaochao Yan, Sheng Wang, Jinyu Yang, Tingyang Xu και Junzhou Huang. *Re-balancing Variational Autoencoder Loss for Molecule Sequence Generation. arXiv e-prints*, 2019.

[15] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh και Ameet Talwalkar. *Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. arXiv e-prints*, 2016.

[16] Hao Wen Dong, Wen Yi Hsiao, Li Chia Yang και Yi Hsuan Yang. *MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment. arXiv e-prints*, 2017.

[17] Hyunjik Kim και Andriy Mnih. *Disentangling by Factorising. arXiv e-prints*, 2018.

[18] Yingzhen Li και Stephan Mandt. *Disentangled Sequential Autoencoder. arXiv e-prints*, 2018.

[19] Laurensvan der Maaten και Geoffrey Hinton. *Viualizing data using t-SNE. Journal of Machine Learning Research*, 9:2579–2605, 2008.

[20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser και Illia Polosukhin. *Attention Is All You Need. arXiv e-prints*, 2017.