ΠΑΝΕΠΙΣΤΉΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ

Διπλωματική Εργασία

Ηχητική Κατάτμηση Σημάτων με Μεθόδους Μηχανικής Μάθησης

Συγγραφέας

Απόλλωνας Χαραλάμπους

ΑΜ : 222017102

Επιβλέπων

Δημήτριος Κάντζος

Αθήνα, Σεπτέμβριος , 2024

UNIVERSITY OF WEST ATTICA

SCHOOL OF ENGINEERING

DEPARTMENT OF INDUSTRIAL DESIGN AND PRODUCTION ENGINEERING

Diploma Thesis

Audio Signal Segmentation Using Machine Learning Methods

Author

Apollonas Charalampous

Registration Number : 222017102

Supervisor

Dimitrios Kantzos

Athens, September , 2024

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ

Ηχητική κατάτμηση σημάτων με μεθόδους μηχανικής μάθησης

Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι Εξεταστική Επιτροπή:

Α/α	ΟΝΟΜΑΤΕΠΩΝΥΜΟ	ΒΑΘΜΙΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
1	Δημήτριος Κάντζος	Καθηγητής	DIMITRIOS KANTZOS Digitally signed by DIMITRIOS KANTZOS Date: 2024.10.02 10:44:22 +03'00'
2	Ελένη Αικατερίνη Λελίγκου	Καθηγήτρια	ELENI AIKATERINI LELIGKOU Digitally signed by ELENI AIKATERINI LELIGKOU Date: 2024.10.02 10:55:07 +03'00'
3	Γρηγόριος Νικολάου	Λέκτορας	Grigorios Nikolaou Digitally signed by Grigorios Nikolaou Date: 2024.10.02 11:06:19 +03'00'

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

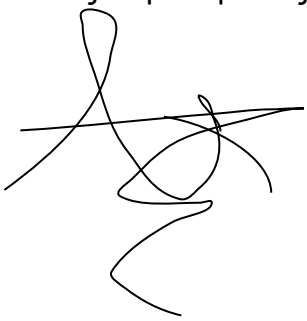
Ο κάτωθι υπογεγραμμένος Απόλλωνας Χαραλάμπους, με αριθμό μητρώου 222017102 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου.»

Ο Δηλών,

Απόλλωνας Χαραλάμπους



Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Ευχαριστίες

Θα ήθελα να εκφράσω την βαθύτατη ευγνωμοσύνη μου προς τον καθηγητή μου, κ. Δημήτριο Κάντζο για την πολύτιμη καθοδήγηση και την υποστήριξη του, που υπήρξαν καθοριστικές για την επιτυχή ολοκλήρωση της μελέτης μου. Επιπλέον, θα ήθελα να ευχαριστήσω θερμά τους γονείς μου για την αδιάκοπη στήριξη και την αγάπη τους. Η ηθική και υλική τους βοήθεια υπήρξε θεμέλιο για την ακαδημαϊκή μου πορεία και για την επίτευξη αυτού του στόχου.

Περίληψη

Η παρούσα διπλωματική εργασία διερευνά την αναγνώριση συναισθημάτων από την ομιλία, χρησιμοποιώντας το σύνολο δεδομένων RAVDESS και σύγχρονες τεχνικές μηχανικής μάθησης. Η εργασία είναι δομημένη σε πέντε κεφάλαια, τα οποία καλύπτουν τη θεωρητική βάση και την πειραματική διαδικασία.

Στο πρώτο κεφάλαιο, παρουσιάζονται οι βασικές αρχές της επιστήμης του ήχου και των ψηφιακών δεδομένων ήχου. Εξετάζουμε τις φυσικές ιδιότητες του ήχου, την ψηφιοποίηση και τις μορφές αποθήκευσης των ηχητικών δεδομένων, προκειμένου να κατανοήσουμε πώς τα ηχητικά σήματα μετατρέπονται σε δεδομένα που μπορούν να επεξεργαστούν οι υπολογιστές.

Το δεύτερο κεφάλαιο επικεντρώνεται στην εξαγωγή χαρακτηριστικών από τα ηχητικά δεδομένα. Περιγράφονται διάφορες τεχνικές και μεθοδολογίες για την ανάλυση των ηχητικών σημάτων, όπως η Ανάλυση Συχνοτήτων και τα Μελ-Φίλτρα Συχνοτήτων, που χρησιμοποιούνται για την εξαγωγή σημαντικών πληροφοριών από τα ηχητικά σήματα.

Στο τρίτο κεφάλαιο, εξετάζουμε διάφορους αλγορίθμους και μεθόδους μηχανικής μάθησης. Αναλύουμε τους κύριους αλγορίθμους επιβλεπόμενης και μη επιβλεπόμενης μάθησης, όπως τα Νευρωνικά Δίκτυα και οι Υποστηριζόμενες Διανυσματικές Μηχανές, εξηγώντας τις βασικές αρχές λειτουργίας τους και τα πλεονεκτήματα που προσφέρουν.

Το τέταρτο κεφάλαιο ασχολείται με την αναγνώριση συναισθημάτων από την ομιλία. Παρουσιάζουμε τις διάφορες προσεγγίσεις και τεχνικές που χρησιμοποιούνται στον τομέα αυτό, τις προκλήσεις που αντιμετωπίζονται και την σημασία της αναγνώρισης συναισθημάτων για τις εφαρμογές τεχνητής νοημοσύνης και τις ανθρώπινες-υπολογιστικές αλληλεπιδράσεις.

Στο πέμπτο κεφάλαιο, προχωράμε στην πειραματική διαδικασία, όπου εφαρμόζουμε και συγκρίνουμε δύο διαφορετικούς ταξινομητές, τους LSTM (Long Short-Term Memory) και MLP (Multi-Layer Perceptron). Χρησιμοποιώντας το σύνολο δεδομένων RAVDESS, εκπαιδεύουμε και αξιολογούμε τα μοντέλα μας για την ταξινόμηση συναισθημάτων, διερευνώντας τις δυνατότητες και τις προκλήσεις που προκύπτουν από κάθε προσέγγιση.

Λέξεις-Κλειδία : Τεχνητή Νοημοσύνη, Βαθία Μάθηση, Νευρωνικά Δίκτυα, MFCCs, Ηχητική Κατάτμηση , Αναγνώριση Συναισθημάτων από Ομιλία, LSTM

ABSTRACT

This thesis explores speech emotion recognition using the RAVDESS dataset and modern machine learning techniques. The thesis is structured into five chapters, covering the theoretical foundations and the experimental process.

The first chapter presents the fundamental principles of the science of sound and digital audio data. We examine the physical properties of sound, digitization, and storage formats of audio data, to understand how audio signals are transformed into data that computers can process.

The second chapter focuses on the extraction of features from audio data. Various techniques and methodologies for audio signal analysis are described, such as Frequency Analysis and Mel-Frequency Cepstral Coefficients, which are used to extract significant information from audio signals.

In the third chapter, we examine various machine learning algorithms and methods. We analyze the main supervised and unsupervised learning algorithms, such as Neural Networks and Support Vector Machines, explaining their fundamental operational principles and the advantages they offer.

The fourth chapter deals with speech emotion recognition. We present the different approaches and techniques used in this field, the challenges encountered, and the importance of emotion recognition for artificial intelligence applications and human-computer interactions.

In the fifth chapter, we proceed with the experimental process, where we apply and compare two different classifiers, LSTM (Long Short-Term Memory) and MLP (Multi-Layer Perceptron). Using the RAVDESS dataset, we train and evaluate our models for emotion classification, exploring the capabilities and challenges that arise from each approach.

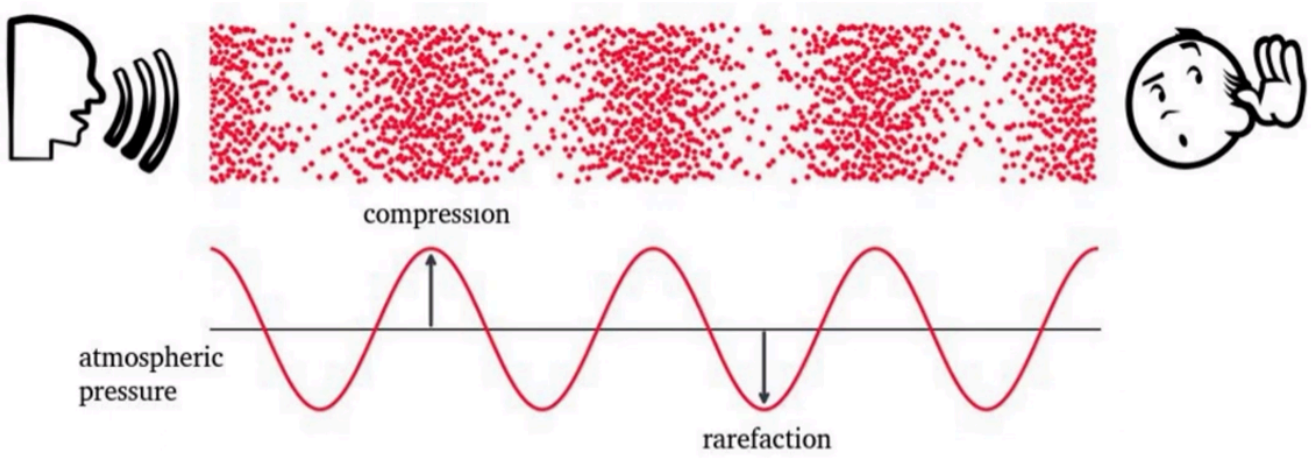
Keywords: Artificial Intelligence, Deep Learning, Neural Networks, MFCCs, Sound Segmentation, Speech Emotion Recognition, LSTM

Κεφάλαιο 1. Εισαγωγή στα ηχητικά σήματα

1.1 Ο Ήχος

Ο ήχος είναι ένα πολύπλοκο φαινόμενο το οποίο βασίζεται στην διάδοση των μηχανικών κυμάτων εντός ενός μέσου, κυρίως του αέρα. Κατά την δόνηση ενός αντικειμένου το δονούμενο σώμα κάνει το μέσο γύρω του να δονείται και αυτό στην αντίστοιχη συχνότητα παράγοντας ήχο.

Τα ηχητικά κύματα αποτελούνται από περιοχές υψηλής και χαμηλής πίεσης που ονομάζονται συμπιέσεις και αραιώσεις αντίστοιχα και γίνονται αντιληπτά μόλις φτάσουν σε ένα αισθητήριο όργανο (π.χ. αυτί, μικρόφωνο).

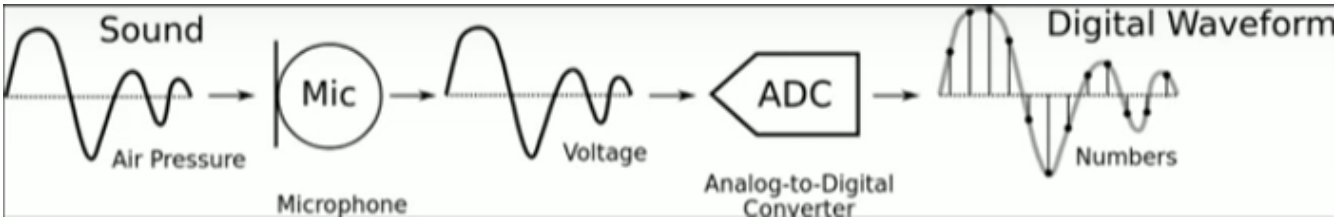


Σχήμα 1.1 Παράδειγμα διάδοσης ηχητικού κύματος

Βασικά στοιχεία ενός ηχητικού σήματος:

- **Συχνότητα** : Είναι ο αριθμός των δονήσεων ή των κύκλων ανά δευτερόλεπτο και μετριέται σε Hertz (**Hz**). Οι υψηλότερες συχνότητες γίνονται αντιληπτές ως ήχοι υψηλότερου τόνου ενώ οι χαμηλότερες ως ήχοι χαμηλότερου τόνου.
- **Πλάτος** : Αναφέρεται στην ισχύ ή την ένταση του ηχητικού κύματος και συχνά σχετίζεται με την ένταση του ήχου. Το πλάτος μετριέται σε decibel (**dB**).
- **Μήκος κύματος**: Αφορά την απόσταση μεταξύ δύο διαδοχικών σημείων που βρίσκονται σε φάση, όπως για παράδειγμα δύο διαδοχικά σημεία συμπίεσης ή αραίωσης. Το μήκος κύματος είναι αντιστρόφως ανάλογο της συχνότητας, συνεπώς τα κύματα υψηλότερων συχνοτήτων έχουν μικρότερα μήκη κύματος.

1.2 Μετατροπή αναλογικού ηχητικού σήματος σε ψηφιακό



Σχήμα 1.2 Ροή διαδικασίας μετατροπής αναλογικού σήματος

Η μετατροπή του αναλογικού σήματος σε αντίστοιχο ψηφιακό περιλαμβάνει μια σχολαστική διαδικασία που καταγράφει τις συνεχείς διακυμάνσεις των ηχητικών κυμάτων και τις μετατρέπει σε διακριτές, ψηφιακές αναπαραστάσεις.

Στην αρχή, το αναλογικό σήμα, το οποίο αντικατοπτρίζει τη συνεχή φύση των ακουστικών κυμάτων, υποβάλλεται σε δειγματοληψία, μέσω της οποίας γίνεται η λήψη τακτικών στιγμιότυπων ή δειγμάτων του αναλογικού σήματος σε συγκεκριμένα χρονικά διαστήματα. Η ουσία της δειγματοληψίας είναι περίπλοκα συνδεδεμένη με το **θεώρημα Nyquist**, μια βασική αρχή στην ψηφιακή επεξεργασία σήματος, το οποίο μαθηματικά διατυπώνεται ως εξής:

Αν έχουμε ένα σήμα x(t) που περιέχει συχνότητες μέχρι fmax τότε το σήμα αυτό μπορεί να ανακατασκευαστεί πλήρως από τα δείγματά του $x[n] = x(nT)$, όπου T είναι η περίοδος δειγματοληψίας, εφόσον: $fs > 2f$ όπου $\frac{fs=1}{T}$ είναι η συχνότητα δειγματοληψίας.

Το θεώρημα στηρίζεται στη θεωρία της αναπαράστασης σήματος μέσω των Σειρών Fourier. Συγκεκριμένα, ένα σήμα που περιορίζεται σε συχνότητα μπορεί να αναπαρασταθεί ως άθροισμα ημιτονοειδών κυματομορφών μέχρι τη μέγιστη συχνότητά του. Η δειγματοληψία αυτού του σήματος με ρυθμό που είναι τουλάχιστον διπλάσιος από τη μέγιστη συχνότητα εξασφαλίζει ότι τα δείγματα περιέχουν επαρκείς πληροφορίες για την ανακατασκευή του αρχικού σήματος χωρίς απώλεια πληροφορίας.

Εάν ένα συνεχές σήμα δειγματοληπτείται με συχνότητα μικρότερη από τη διπλάσια της μέγιστης συχνότητας του σήματος, τότε οι συχνότητες του αρχικού σήματος που είναι υψηλότερες από τη μισή συχνότητα δειγματοληψίας (συχνότητα Nyquist) αναδιπλώνονται (**aliasing**) και εμφανίζονται ως χαμηλότερες συχνότητες, προκαλώντας παραμόρφωση στο δειγματοληπτημένο σήμα.

Μετά τη δειγματοληψία, οι τιμές συνεχούς πλάτους του αναλογικού σήματος διακριτοποιούνται μέσω κβαντοποίησης. Η κβαντοποίηση είναι η διαδικασία μετατροπής ενός συνεχούς φάσματος τιμών σε ένα διακριτό σύνολο επιπέδων, εκχωρώντας αριθμητικές τιμές σε κάθε δείγμα μέσα σε ένα προκαθορισμένο εύρος και καθορίζει την ανάλυση της ψηφιακής αναπαράστασης. Ένα υψηλότερο βάθος bit στην κβαντοποίηση οδηγεί σε λεπτότερη ανάλυση, διατηρώντας περισσότερες λεπτομέρειες στο σήμα.

Η διαδικασία κβαντισμού μπορεί να χωριστεί σε δύο κύρια στάδια, τη στρογγυλοποίηση (Rounding ή Truncation) και την Αντιστοίχιση (Mapping). Κατά την στρογγυλοποίηση κάθε δειγματοληπτημένη τιμή του αναλογικού σήματος συγκρίνεται με τα επίπεδα κβαντισμού και ανατίθεται στο πλησιέστερο επίπεδο. Για παράδειγμα, αν τα επίπεδα κβαντισμού είναι 0, 1, 2, 3 και η δειγματοληπτημένη τιμή είναι 2.7 η τιμή αυτή θα στρογγυλοποιηθεί στο 3. Στην αντιστοίχιση κάθε τιμή αντιστοιχίζεται σε έναν δυαδικό αριθμό, δηλαδή σε μια ψηφιακή αναπαράσταση. Αυτή η ψηφιακή τιμή είναι η έξοδος του ADC και μπορεί να χρησιμοποιηθεί για περαιτέρω επεξεργασία ή αποθήκευση.

Υπάρχουν διάφορες μέθοδοι κβαντισμού, ανάλογα με την εφαρμογή και τις απαιτήσεις ακρίβειας:

1. **Ομοιόμορφος Κβαντισμός (Uniform Quantization):**
   - Τα επίπεδα κβαντισμού είναι ισομερώς κατανεμημένα, δηλαδή οι αποστάσεις μεταξύ των διαδοχικών επιπέδων είναι ίσες.
   - Είναι απλός και εύκολος στην υλοποίηση, αλλά μπορεί να μην είναι αποτελεσματικός για σήματα με μεγάλη δυναμική περιοχή.
2. **Μη Ομοιόμορφος Κβαντισμός (Non-uniform Quantization):**
   - Τα επίπεδα κβαντισμού δεν είναι ισομερώς κατανεμημένα και μπορεί να προσαρμόζονται ανάλογα με τις ιδιότητες του σήματος.
   - Συνήθως χρησιμοποιείται σε εφαρμογές όπως η επεξεργασία ήχου, όπου η ανθρώπινη ακοή είναι πιο ευαίσθητη σε χαμηλές εντάσεις.

Ο κβαντισμός εισάγει αναπόφευκτα σφάλμα, γνωστό ως σφάλμα κβαντισμού. Το σφάλμα κβαντισμού είναι η διαφορά μεταξύ της πραγματικής δειγματοληπτημένης τιμής και της κβαντισμένης τιμής. Μπορεί να περιγραφεί ως θόρυβος και επηρεάζει την ακρίβεια του ψηφιακού σήματος.

Το σφάλμα αυτό μπορεί να μειωθεί με:

- **Αύξηση του αριθμού των επιπέδων κβαντισμού**: Χρησιμοποιώντας περισσότερα bits για την αναπαράσταση των κβαντισμένων τιμών, μπορούμε να αυξήσουμε τον αριθμό των επιπέδων κβαντισμού, μειώνοντας έτσι το μέγεθος του σφάλματος.
- **Χρήση μη ομοιόμορφου κβαντισμού**: Προσαρμόζοντας τα επίπεδα κβαντισμού ώστε να ταιριάζουν καλύτερα με την κατανομή του σήματος.

Τέλος η κωδικοποίηση (**encoding**) είναι το τελικό στάδιο στη διαδικασία μετατροπής ενός αναλογικού σήματος σε ψηφιακό, όπου οι κβαντισμένες τιμές μετατρέπονται σε δυαδικούς αριθμούς, δηλαδή σε ψηφιακή μορφή που μπορεί να επεξεργαστεί από υπολογιστές και άλλα ψηφιακά συστήματα. Οι κβαντισμένες τιμές που έχουν ήδη αναπαρασταθεί σε ένα προκαθορισμένο σύνολο επιπέδων μεταφράζονται σε μια σειρά από bits (δυαδικούς αριθμούς) που αποτελούνται από 0 και 1.
Ο αριθμός των bits που χρησιμοποιείται εξαρτάται από τον αριθμό των επιπέδων κβαντισμού, με περισσότερα bits να προσφέρουν μεγαλύτερη ακρίβεια αλλά και να απαιτούν περισσότερη αποθηκευτική ικανότητα και υπολογιστική ισχύ. Για παράδειγμα, ένας 8-bit ADC μπορεί να αναπαραστήσει 256 διαφορετικά επίπεδα, καθώς $2^8 = 256$
. Αυτή η ψηφιακή αναπαράσταση επιτρέπει τη μεταφορά των δεδομένων μέσω δικτύων, την

αποθήκευση σε ψηφιακά μέσα και την επεξεργασία από αλγόριθμους, καθιστώντας δυνατές πολλές σύγχρονες τεχνολογικές εφαρμογές. Η κωδικοποίηση διασφαλίζει ότι κάθε αναλογική είσοδος μπορεί να αποθηκευτεί και να ανακτηθεί με ακρίβεια, παρέχοντας την υποδομή για τις τεχνολογίες που χρησιμοποιούμε καθημερινά, από τη μουσική και τις φωτογραφίες μέχρι την ιατρική απεικόνιση και τις επικοινωνίες.

1.3 Ψηφιακά ηχητικά σήματα

Μετά την ψηφιοποίηση ενός ηχητικού κύματος, λαμβάνουμε μια ψηφιακή αναπαράσταση του ηχητικού σήματος. Αυτή η αναπαράσταση τυπικά αποτελείται από μια ακολουθία διακριτών αριθμητικών τιμών, καθεμία από τις οποίες αντιπροσωπεύει το πλάτος του ηχητικού κύματος σε μια συγκεκριμένη χρονική στιγμή. Αυτές οι αριθμητικές τιμές συνήθως αποθηκεύονται ως δυαδικά δεδομένα σε ένα ψηφιακό αρχείο ήχου.

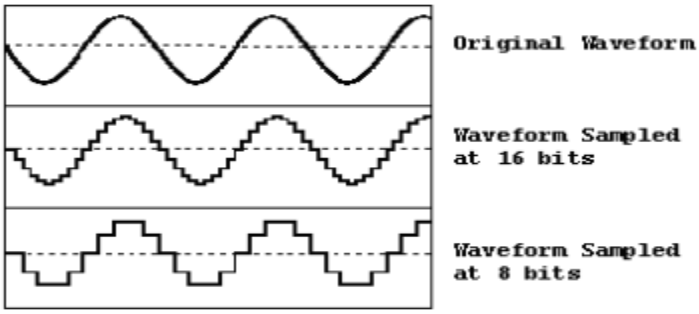
Στοιχεία ενός ψηφιακού σήματος :

Ρυθμός δειγματοληψίας: Ο ρυθμός δειγματοληψίας μετρημένος σε Hertz (Hz), αναφέρεται στον αριθμό των δειγμάτων που λαμβάνονται ανά δευτερόλεπτο κατά τη διάρκεια της διαδικασίας μετατροπής αναλογικού σε ψηφιακό. Καθορίζει τη χρονική ανάλυση του ψηφιακού σήματος ήχου. Οι συνήθεις ρυθμοί δειγματοληψίας περιλαμβάνουν 44,1 kHz (χρησιμοποιείται σε CD), 48 kHz (χρησιμοποιείται σε DVD και στις περισσότερες μορφές ψηφιακού ήχου) και υψηλότερους ρυθμούς για εξειδικευμένες εφαρμογές όπως ήχος υψηλής ανάλυσης.

Βάθος bit: Το βάθος bit αναφέρεται στον αριθμό των bit που χρησιμοποιούνται για την αναπαράσταση κάθε δείγματος στο ψηφιακό ηχητικό σήμα. Καθορίζει το δυναμικό εύρος του σήματος ή το εύρος των πλατών που μπορούν να αναπαρασταθούν με ακρίβεια. Τα κοινά βάθη bit περιλαμβάνουν 16 bit (χρησιμοποιείται σε CD), 24 bit (χρησιμοποιείται σε πολλές επαγγελματικές εφαρμογές ήχου) και μεγαλύτερα βάθη bit για ήχο υψηλής πιστότητας.

Κανάλια: Τα κανάλια αναφέρονται στον αριθμό των ανεξάρτητων σημάτων ήχου που συνδυάζονται για τη δημιουργία του ψηφιακού σήματος ήχου. Οι συνήθεις διαμορφώσεις περιλαμβάνουν μονοφωνικές μορφές (1 κανάλι), στερεοφωνικό (2 κανάλια) και μορφές ήχου surround (όπως κανάλια 5.1 ή 7.1).

Μορφή αρχείου: Το ψηφιακό σήμα ήχου συνήθως αποθηκεύεται σε μια συγκεκριμένη μορφή αρχείου, η οποία καθορίζει τον τρόπο οργάνωσης και κωδικοποίησης των δεδομένων ήχου. Οι κοινές μορφές αρχείων περιλαμβάνουν WAV (Waveform Audio File Format), AIFF (Audio Interchange File Format), MP3 (MPEG Audio Layer III), AAC (Advanced Audio Coding) και FLAC (Free Lossless Audio Codec), μεταξύ άλλων. Κάθε μορφή μπορεί να έχει διαφορετικούς αλγόριθμους συμπίεσης, υποστήριξη μεταδεδομένων και συμβατότητα με διαφορετικές συσκευές αναπαραγωγής και λογισμικό.



Σχήμα 1.3 Ηχητικό σήμα υπό διαφορετικό ρυθμό δειγματοληψίας

Κεφάλαιο 2. Εξαγωγή ηχητικών χαρακτηριστικών

2.1 Εισαγωγή

Για την εκπαίδευση οποιουδήποτε μοντέλου στατιστικής ή μηχανικής μάθησης είναι απαραίτητη η εξαγωγή σχετικών χαρακτηριστικών από τα ηχητικά σήματα. Αυτή η διαδικασία είναι γνωστή ως εξαγωγή ηχητικών χαρακτηριστικών (audio feature extraction) και εστιάζει στον χειρισμό και την επεξεργασία των σημάτων ήχου. Περιλαμβάνει εργασίες όπως η αφαίρεση θορύβου και η εναρμόνιση των περιοχών χρόνου-συχνότητας μέσω της μετατροπής τόσο των ψηφιακών όσο και των αναλογικών σημάτων. Η εξαγωγή χαρακτηριστικών ήχου περιστρέφεται κυρίως γύρω από υπολογιστικές τεχνικές που στοχεύουν στην αλλαγή των χαρακτηριστικών του ήχου.

Τα κοινά ηχητικά χαρακτηριστικά που είναι χρήσιμα για τη μοντελοποίηση περιλαμβάνουν μια σειρά από περιγραφικούς δείκτες που καταγράφουν διάφορες πτυχές του ήχου, επιτρέποντας την ανάπτυξη έξυπνων μοντέλων μέσω στατιστικών ή μηχανικών προσεγγίσεων εκμάθησης. Αυτές οι δυνατότητες βρίσκουν εφαρμογή σε εργασίες όπως η ταξινόμηση ήχου, η αναγνώριση ομιλίας, η επισήμανση μουσικής, η τμηματοποίηση, ο διαχωρισμός πηγών, η λήψη δακτυλικών αποτυπωμάτων, η διαγραφή θορύβων και η ανάκτηση πληροφοριών μουσικής.

2.2 Κατηγοριοποίηση ηχητικών χαρακτηριστικών:

2.2.1 Στατιστικά Χαρακτηριστικά

Όπως και με κάθε άλλο αριθμητικό σύνολο έτσι και με τον ψηφιακό ήχο μπορούμε να υπολογίσουμε βασικά στατιστικά χαρακτηριστικά όπως:

i) Μέση Τιμή (Mean Value)

Η μέση τιμή ενός σήματος είναι η συνολική αθροιστική τιμή των δειγμάτων διαιρούμενη με τον αριθμό των δειγμάτων. Αποτελεί μια ένδειξη της κεντρικής τάσης του σήματος.

ii) Διακύμανση (Variance)

Η διακύμανση μετρά τη διασπορά των τιμών του σήματος γύρω από τη μέση τιμή. Υψηλή διακύμανση υποδεικνύει ότι τα δείγματα του σήματος έχουν μεγάλη απόκλιση από τη μέση τιμή.

iii) Ασυμμετρία (Skewness)

Η ασυμμετρία μετρά την ασυμμετρία της κατανομής των τιμών του σήματος. Εάν η ασυμμετρία είναι μηδενική, η κατανομή είναι συμμετρική.

iv) Κύρτωση (Kurtosis)

Η κύρτωση μετρά την "αιχμηρότητα" της κατανομής των τιμών του σήματος. Μια υψηλή τιμή κύρτωσης υποδεικνύει ότι η κατανομή έχει αιχμηρές κορυφές.

v) Μέγιστη και ελάχιστη τιμή (Min/Max Values)

vi) Εύρος (Range)

Το εύρος είναι η διαφορά μεταξύ της μέγιστης και της ελάχιστης τιμής του σήματος.

vii) Διάμεσος (Median)

Η διάμεσος είναι η μεσαία τιμή του σήματος όταν τα δείγματα είναι ταξινομημένα σε αύξουσα σειρά.

viii) Διασπορά (Standard Deviation)

Η διασπορά είναι η τετραγωνική ρίζα της διακύμανσης και δίνει μια αίσθηση της τυπικής απόκλισης των δειγμάτων από τη μέση τιμή.

ix) Ενέργεια (Energy)

Η ενέργεια ενός σήματος είναι το άθροισμα των τετραγώνων των τιμών των δειγμάτων.

x) Ενέργεια Ριζικής Μέσης Τετραγωνικής Τιμής (Root Mean Square Energy - RMSE)

Η RMSE είναι ένα μέτρο της ενέργειας του σήματος και υπολογίζεται ως η τετραγωνική ρίζα του μέσου όρου των τετραγώνων των τιμών των δειγμάτων του σήματος. Αντικατοπτρίζει την ισχύ του σήματος και είναι ιδιαίτερα χρήσιμη για την αξιολόγηση της έντασης του ήχου.

2.2.2 Ρυθμός Μηδενικής Διέλευσης (Zero-Crossing Rate – ZCR)

Ο ρυθμός μηδενικής διέλευσης (ZCR) μετρά τον αριθμό των φορών που το σήμα διασχίζει τον άξονα του μηδενός ανά μονάδα χρόνου.

2.2.3 Φασματική Πυκνότητα Ισχύος (Power Spectral Density – PSD)

Η Φασματική Πυκνότητα Ισχύος (Power Spectral Density - PSD) είναι μια βασική μέθοδος στην ανάλυση σημάτων που επιτρέπει τον προσδιορισμό της ισχύος ενός σήματος σε σχέση με τις συχνότητες. Αυτή η ανάλυση είναι κρίσιμη για την κατανόηση του τρόπου με τον οποίο η ισχύς διανέμεται στις διάφορες συχνότητες ενός σήματος.

Η PSD μετρά την ενεργειακή κατανομή ενός σήματος ως συνάρτηση της συχνότητας. Αυτή η μέτρηση επιτυγχάνεται μέσω του μετασχηματισμού Fourier της αυτοσυσχέτισης του

σήματος. Η αυτοσυσχέτιση ενός σήματος x(t), είναι ο υπολογισμός της μέσης τιμής του γινομένου του σήματος με μια μετατοπισμένη εκδοχή του εαυτού του και ορίζεται μαθηματικά ως :

$$R_{xx}(\tau) = E[x(t)\, x(t+\tau)] \tag{2.1}$$

Ο μετασχηματισμός Fourier αυτής της συνάρτησης δίνει τη φασματική πυκνότητα ισχύος και ορίζεται ως:

$$S_{xx}(f) = \int_{-\infty}^{+\infty} R_{xx}(\tau)\, e^{-j2\pi f \tau} d\tau \tag{2.2}$$

Η σημασία της PSD έγκειται στην ικανότητά της να αποκαλύπτει τη συχνοτική περιεκτικότητα του σήματος. Εάν ένα σήμα περιέχει συχνοτικές συνιστώσες που επαναλαμβάνονται, η PSD θα παρουσιάσει αιχμές σε αυτές τις συχνότητες, δίνοντας μια σαφή ένδειξη των περιοδικών χαρακτηριστικών του σήματος. Αυτό είναι ιδιαίτερα χρήσιμο στην ανάλυση θορύβου, καθώς επιτρέπει τον διαχωρισμό των θορυβωδών στοιχείων από το χρήσιμο σήμα.

Ένα πρακτικό παράδειγμα χρήσης της PSD είναι στην επεξεργασία ήχου, όπου μπορεί να χρησιμοποιηθεί για την αναγνώριση και την εξάλειψη θορύβων από ηχογραφήσεις. Στις τηλεπικοινωνίες, η PSD βοηθά στην ανάλυση και τον σχεδιασμό συστημάτων μετάδοσης, εξασφαλίζοντας ότι τα σήματα μεταδίδονται με την ελάχιστη δυνατή απώλεια και παρεμβολή.

2.2.4 Φασματική Εντροπία (Spectral Entropy)

Η φασματική εντροπία είναι ένα μέτρο της αταξίας ή της τυχαιότητας στο φάσμα συχνοτήτων ενός σήματος δηλαδή ένα μέτρο της πολυπλοκότητας και της διαταραχής στο συχνοτικό πεδίο. Η φασματική εντροπία βασίζεται στην έννοια της εντροπίας στην θεωρία πληροφοριών, που εισήχθη από τον Claude Shannon, και παρέχει μια μέτρηση της αβεβαιότητας ή της πληροφορίας ενός σήματος. Χρησιμοποιείται συχνά στην ανάλυση σημάτων, όπως τα βιολογικά σήματα (EEG, ECG), καθώς και σε άλλες εφαρμογές επεξεργασίας σήματος και τηλεπικοινωνιών.

Η φασματική εντροπία αναλύει ένα σήμα στο συχνοτικό πεδίο, ξεκινώντας με τον μετασχηματισμό Fourier, που μετατρέπει το σήμα x(t) από το χρονικό στο συχνοτικό πεδίο μέσω του εξής τύπου:

$$X(f) = \int_{-\infty}^{+\infty} x(t)\, e^{-j2\pi f t} dt \tag{2.3}$$

Επιπρόσθετα υπολογίζουμε το φάσμα ισχύος ως το τετράγωνο του μέτρου του μετασχηματισμού Fourier:

$$P(f) = X(f)^2 \tag{2.4}$$

Έπειτα εφαρμόζουμε κανονικοποίηση στο φάσμα ισχύος έτσι ώστε το άθροισμα της πυκνότητας ισχύος σε όλες τις συχνότητες να ισούται με 1.

$$P_n(f) = \frac{P(f)}{\int_{-\infty}^{\infty} P(f)\, df} \tag{2.5}$$

Η διαδικασία αυτή είναι απαραίτητη για να εξασφαλιστεί ότι το συνολικό άθροισμα των φασματικών συνιστωσών ισούται με 1. Αυτό είναι κρίσιμο διότι επιτρέπει την ορθολογική σύγκριση της κατανομής ισχύος μεταξύ διαφορετικών σημάτων ή μεταξύ διαφορετικών τμημάτων του ίδιου σήματος. Χωρίς αυτή την κανονικοποίηση, τα σήματα με διαφορετική συνολική ισχύ δεν θα μπορούσαν να συγκριθούν απευθείας ως προς την εντροπία τους.

2.2.5 Συντελεστές Μελ-Συχνότητας (Mel-Frequency Cepstral Coefficients- MFCC)

Τα MFCCs είναι από τα πιο κοινά εξαγώγιμα ηχητικά στοιχεία για εφαρμογές μηχανικής μάθησης. Χρησιμοποιούνται σε μια πλειάδα εφαρμογών όπως η αναγνώριση φωνής και η αναγνώριση φύλου.
Για να εξάγουμε ένα MFCC θα πρέπει να ακολουθήσουμε μια συγκεκριμένη διαδικασία **5 σταδίων:**

Στάδιο 1ο : Προέμφαση

Η προέμφαση είναι μία από τις πιο κοινές πρακτικές προεπεξεργασίας ηχητικών δεδομένων. Χρησιμοποιείται για να αντισταθμίσει την υψηλότερη συχνότητα του σήματος η οποία κατεστάλη κατά τη διάρκεια παραγωγής του σήματος. Η προέμφαση αποτελεί το πρώτο βήμα στην εξαγωγή ενός MFCC και υλοποιείται με την απλή εφαρμογή ενός υψιπερατού φίλτρου στο αρχικό σήμα.
Η διαδικασία αυτή συνήθως θα επηρεάσει και την ενεργειακή κατανομή μεταξύ των συχνοτήτων όπως και το συνολικό ενεργειακό επίπεδο του σήματος

Στάδιο 2ο : Πλαίσιο και παράθυρο σήματος

Το πρώτο βήμα είναι η τμηματοποίηση του συνεχούς σήματος σε μικρότερα κομμάτια τα οποία ονομάζονται πλαίσια. Αυτή η διαδικασία είναι απαραίτητη επειδή τα χαρακτηριστικά των σημάτων ομιλίας αλλάζουν με την πάροδο του χρόνου, με αποτέλεσμα η ανάλυση του σήματος σε μικρότερα πλαίσια να μας επιτρέπει να καταγράψουμε αυτές τις αλλαγές με μεγαλύτερη ακρίβεια.

Αναλυτικότερα το σήμα χωρίζεται σε επικαλυπτόμενα πλαίσια. Συνήθως το κάθε καρέ έχει χρονική διάρκεια 20-40 χιλιοστών του δευτερολέπτου (ms). Αυτό το μήκος επιλέγεται καθώς τα 20ms είναι ο χρόνος που χρειάζεται η ανθρώπινη γλωττίδα για να παράξει ήχο. Συνεπώς υποθέτουμε ότι οι ιδιότητες του σήματος σε κάθε πλαίσιο παραμένουν σχετικά σταθερές αλλά παραμένει αρκετά μεγάλο ώστε να περιέχει αρκετά δεδομένα για ανάλυση.

Επιπρόσθετα τα πλαίσια αυτά επικαλύπτονται με την διαδοχική μετατόπιση. Τα διαδοχικά καρέ συνήθως επικαλύπτονται κατά 50% ή και περισσότερο. Η επικάλυψη βοηθά στην διασφάλιση της ομαλότητας κατά την μετάβαση μεταξύ των πλαισίων ενώ ταυτόχρονα δεν χάνονται τα σημαντικά χαρακτηριστικά του αρχικού σήματος. Για παράδειγμα εάν το χρονικό μήκος του αρχικού μας καρέ είναι 50ms και επιθυμούμε επικάλυψη 50% τότε η μετατόπιση του επόμενου καρέ θα είναι 25ms.

Μόλις χωρίσουμε το σήμα μας σε πλαίσια πολλαπλασιάζουμε το κάθε καρέ με μια συνάρτηση παραθύρου (window function). Η δημιουργία παραθύρων βοηθάει στην μείωση της φασματικής διαρροής η οποία συμβαίνει λόγω του πεπερασμένου μήκους των πλαισίων όταν αυτά μετασχηματίζονται στον τομέα της συχνότητας.

Η συνάρτηση αυτή εφαρμόζεται σε κάθε πλαίσιο μηδενίζοντας τις άκρες του. Με αυτό τον τρόπο εξομαλύνει τις ασυνέχειες στα όρια του κάθε καρέ. Οι 3 πιο κοινές συναρτήσεις είναι η συνάρτηση Hamming, η συνάρτηση Hanning αλλά και το ορθογώνιο παράθυρο.

Ορθογώνιο Παράθυρο

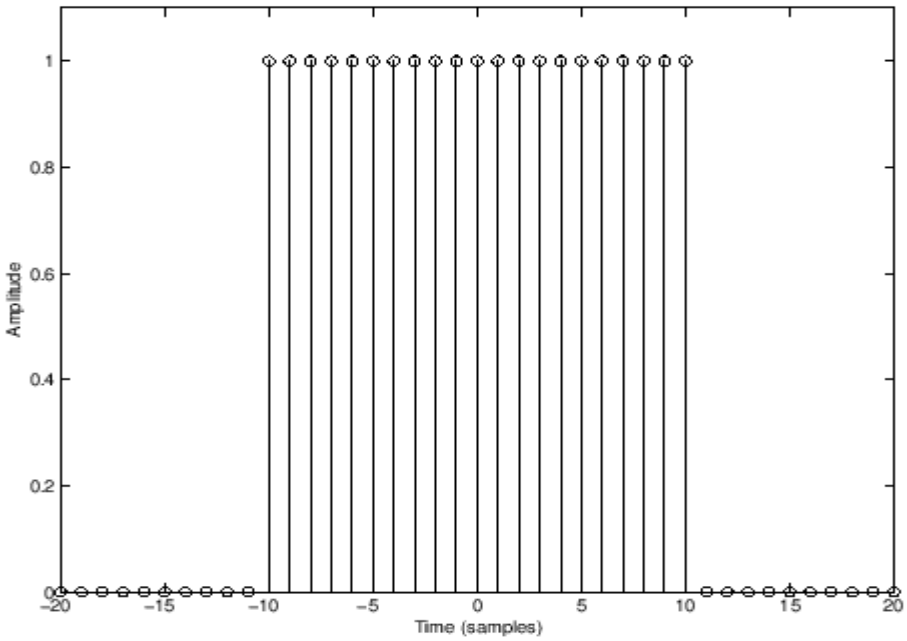
Το πιο απλό παράθυρο, το ορθογώνιο παράθυρο, είναι μια σειρά από μονάδες που κόβει το σήμα σε ένα καθορισμένο μήκος χωρίς να αλλάζει τις τιμές του εντός του παραθύρου. Αν και εύκολο στην εφαρμογή, έχει υψηλή διαρροή καθώς δεν μειώνει σταδιακά τις τιμές στις άκρες του σήματος.

Μπορούμε να το ορίσουμε μαθηματικά ως :

Για ένα σήμα x(n) και ένα παράθυρο μήκους N, το ορθογώνιο παράθυρο w(n) ορίζεται ως:

$$w(n) = 1, \text{για } 0 \leq n < N \text{ αλλιως } w(n) = 0 \qquad (2.6)$$

.



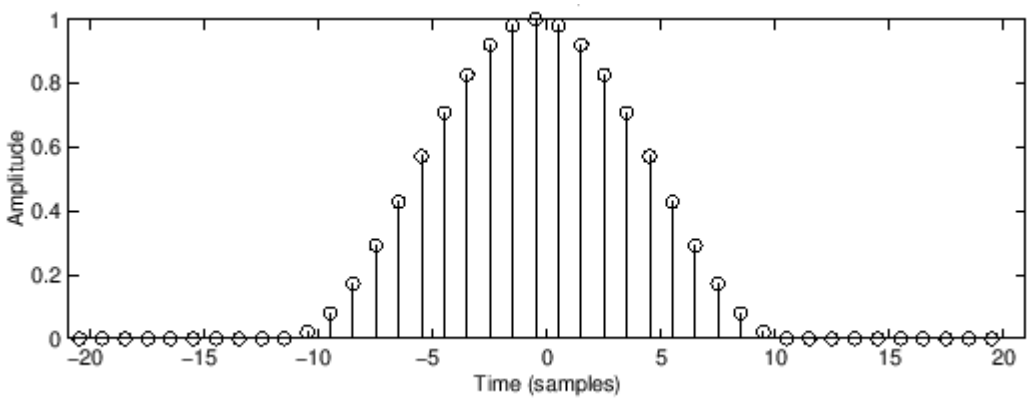
Σχήμα 2.1 Το ορθογώνιο παράθυρο

Παράθυρο Hanning

Το παράθυρο Hanning είναι ένα ημιτονοειδές παράθυρο που μειώνει σταδιακά τις τιμές προς τις άκρες, δίνοντας μια ομαλότερη μετάβαση από και προς το μηδέν. Αυτό μειώνει τα φαινόμενα διαρροής σε σύγκριση με το ορθογώνιο παράθυρο.

Το ορίζουμε μαθηματικά ως :

$$w(n) = 0.5\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right) \tag{2.7}$$



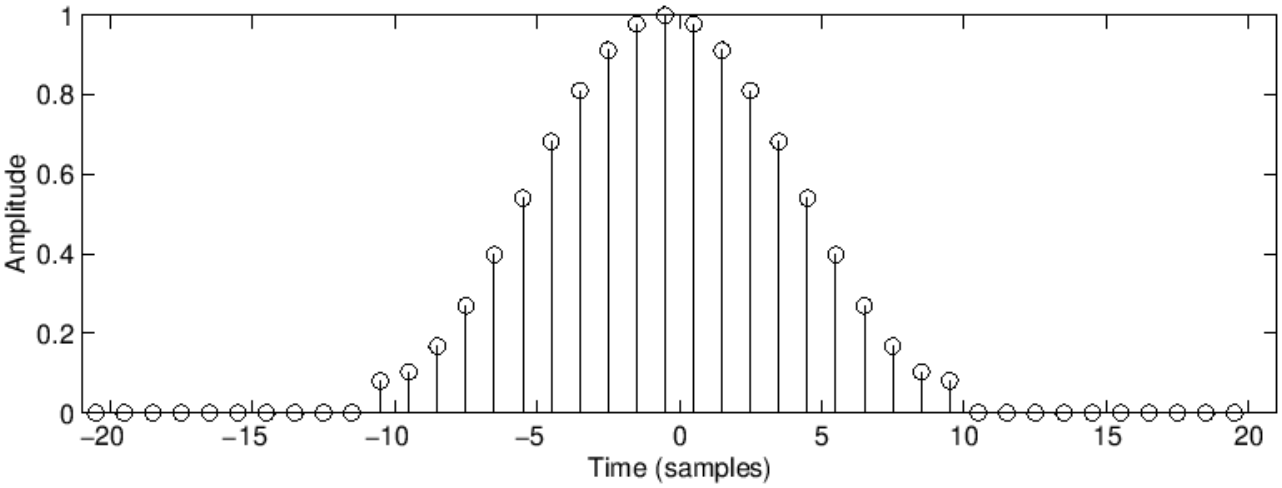
Σχήμα 2.2 Το παράθυρο Hanning

Παράθυρο Hamming

Το παράθυρο Hamming είναι παρόμοιο με το παράθυρο Hanning αλλά χρησιμοποιεί διαφορετικούς συντελεστές για να μειώσει τα φαινόμενα διαρροής ακόμη περισσότερο.

Το μαθηματικό του μοντέλο είναι :

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad (2.8)$$



Σχήμα 2.3 Το παράθυρο Hamming

Στάδιο 3ο : Φάσμα Ισχύος

Το φάσμα ισχύος ορίζεται ως η κατανομή της ισχύς των συχνοτικών στοιχείων που συνθέτουν το σήμα. Παραδοσιακά χρησιμοποιείται ο Διακριτός Μετασχηματισμός Fourier (DFT) για τον υπολογισμό του και προσδιορίζεται για το κάθε πλαίσιο ξεχωριστά με την παρακάτω εξίσωση.

$$x(k) = \sum_{n=0}^{N-1} x(n)\, e^{-\frac{2\pi jnk}{N}} \quad (2.9)$$

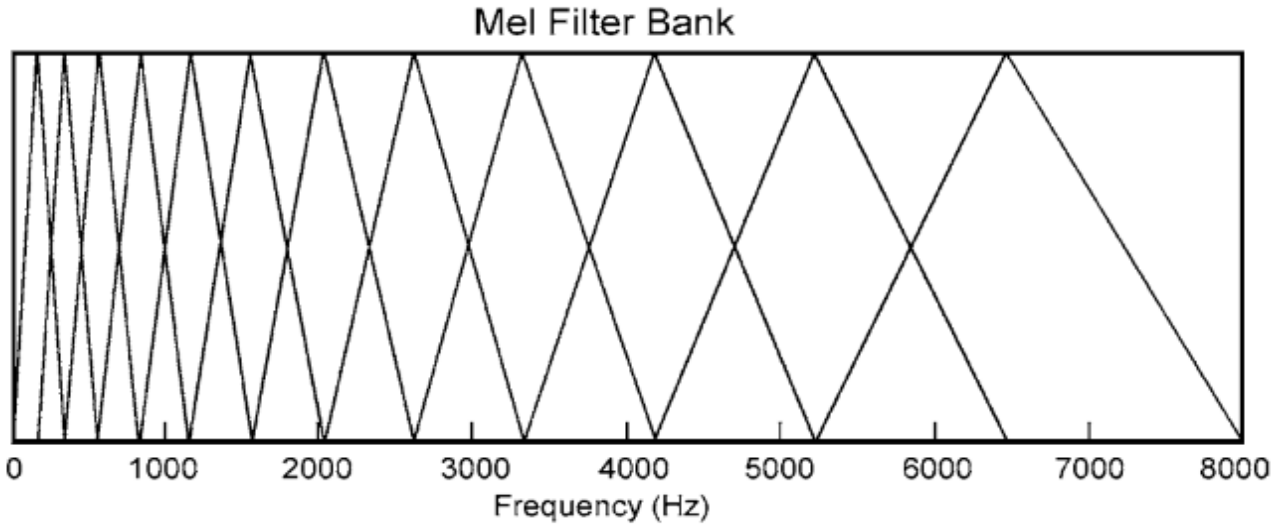
Όπου x(n) είναι το διακριτό σήμα και N το μήκος του σήματος.

Στάδιο 4ο : Συστοιχία φίλτρων Mel

Το ζωνοπερατό φίλτρο Mel είναι μια συστοιχία 40 τριγωνικών φίλτρων η οποία είναι κατασκευασμένη με βάση την αντίληψη του τόνου. Το φίλτρο Mel αναπτύχθηκε αρχικά για την ανάλυση της ομιλίας. Στοχεύει στην αντίληψη της ομιλίας όπως γίνεται από το ανθρώπινο αυτί, εξάγοντας μη γραμμικές αναπαραστάσεις του ηχητικού σήματος. Αυτά τα φίλτρα εφαρμόζονται στο φάσμα ισχύος του σήματος, που συνήθως λαμβάνεται μέσω ενός μετασχηματισμού Fourier των βραχυχρόνιων πλαισίων του σήματος ήχου. Κάθε φίλτρο δίνει

έμφαση σε ένα συγκεκριμένο εύρος συχνοτήτων και τα φίλτρα είναι σχεδιασμένα έτσι ώστε να έχουν υψηλότερη ανάλυση σε χαμηλότερες συχνότητες και χαμηλότερη ανάλυση σε υψηλότερες συχνότητες. Αυτό επιτυγχάνεται τοποθετώντας τα φίλτρα πιο κοντά μεταξύ τους στο εύρος χαμηλής συχνότητας και τοποθετώντας τα σε μεγαλύτερη απόσταση μεταξύ τους στο εύρος υψηλών συχνοτήτων.

Η έξοδος κάθε φίλτρου είναι το σταθμισμένο άθροισμα των στοιχείων του φάσματος ισχύος εντός του εύρους αυτού του φίλτρου. Αυτές οι έξοδοι αντιπροσωπεύουν την ισχύ του σήματος σε κάθε ζώνη συχνοτήτων σε κλίμακα Mel, αποτυπώνοντας αποτελεσματικά τις φασματικές ιδιότητες του ήχου με τρόπο που ευθυγραμμίζεται με την ανθρώπινη ακουστική αντίληψη. Το ημερολόγιο αυτών των ενεργειών της τράπεζας φίλτρου Mel λαμβάνεται συνήθως για να συμπιέσει το δυναμικό εύρος και για να προσεγγίσει τη λογαριθμική αντίληψη της έντασης του ανθρώπινου αυτιού.



Σχήμα 2.4 Συστοιχία φίλτρων Mel

<u>Στάδιο 5ο : Διακριτός Μετασχηματισμός Συνημιτόνου (DCT)</u>

Το DCT είναι ένας μαθηματικός μετασχηματισμός που χρησιμοποιείται στην επεξεργασία σήματος και στη συμπίεση δεδομένων. Ξεχωρίζει για την ικανότητά του να μετατρέπει μια ακολουθία τιμών σε ένα άθροισμα συναρτήσεων συνημιτόνου που ταλαντώνονται σε διαφορετικές συχνότητες. Το DCT είναι ιδιαίτερα αποτελεσματικό στη συμπαγή αναπαράσταση του περιεχομένου πληροφοριών ενός σήματος, καθιστώντας το βασικό εργαλείο σε διάφορες εφαρμογές όπως η συμπίεση εικόνας και ήχου, συμπεριλαμβανομένων των μορφών JPEG και MP3.

Επιπλέον στο πλαίσιο της εξαγωγής χαρακτηριστικών από τα MFCCs ο DCT κρίνεται ιδιαίτερα χρήσιμος. Μετά το φιλτράρισμα του φάσματος ισχύος ενός ηχητικού σήματος μέσω της των φίλτρων Mel και την εφαρμογή λογαριθμικής κλιμάκωσης, οι ενέργειες που

προκύπτουν από την συστοιχία πρέπει να μετατραπούν σε μια μορφή που είναι πιο κατάλληλη για περαιτέρω εργασίες επεξεργασίας και αναγνώρισης προτύπων. Το DCT παίρνει τις ενέργειες της συστοιχίας φίλτρου log Mel και τις μετατρέπει σε ένα σύνολο συντελεστών. Αυτός ο μετασχηματισμός εξυπηρετεί δύο πρωταρχικούς σκοπούς:

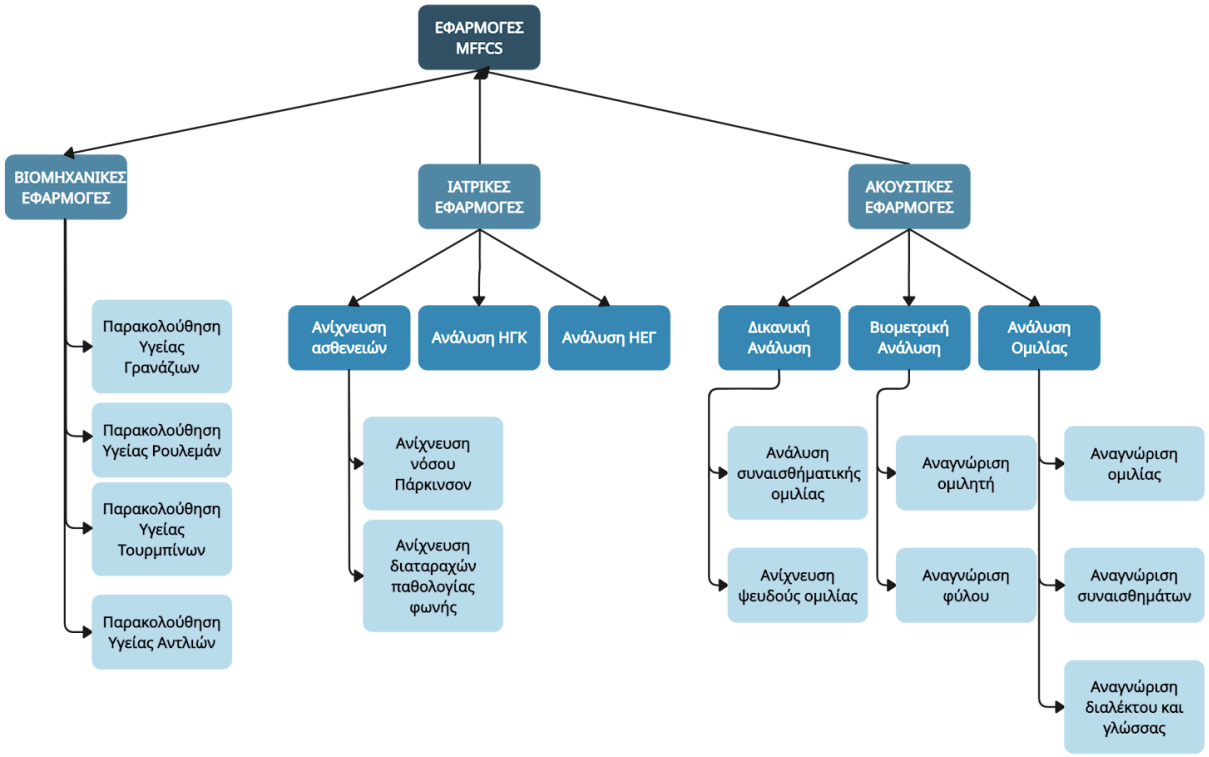
i) Συμπίεση ενέργειας: Το DCT τείνει να συγκεντρώνει την ενέργεια του σήματος στους πρώτους συντελεστές. Αυτή η ιδιότητα σημαίνει ότι οι περισσότερες σημαντικές πληροφορίες σχετικά με τα φασματικά χαρακτηριστικά του σήματος συλλαμβάνονται στους πρώτους συντελεστές DCT, επιτρέποντας τη μείωση της διάστασης. Αυτό καθιστά δυνατή την απόρριψη συντελεστών υψηλότερης τάξης χωρίς σημαντική απώλεια πληροφοριών, μειώνοντας έτσι την πολυπλοκότητα των επόμενων σταδίων επεξεργασίας.

ii) Αποσυσχέτιση: Το DCT αποσυσχετίζει τις ενέργειες της τράπεζας φίλτρου καταγραφής Mel. Με άλλα λόγια, μετατρέπει τις συσχετισμένες τιμές ενέργειας σε ένα σύνολο μη συσχετισμένων συντελεστών. Αυτό είναι επωφελές για αλγόριθμους μηχανικής μάθησης, όπως αυτοί που χρησιμοποιούνται στην επεξεργασία ομιλίας και ήχου, επειδή τα μη συσχετισμένα χαρακτηριστικά συχνά οδηγούν σε καλύτερη απόδοση και πιο αποτελεσματική μάθηση.

Μαθηματικά ο DCT ορίζεται ως :

$$X[k] = \sum_{n=0}^{N-1} x[n] \cos\left(\frac{n(2n+1)(k)}{2N} \right) \tag{2.10}$$

Στον υπολογισμό των MFCC, τυπικά διατηρούνται μόνο οι πρώτοι 12-13 συντελεστές DCT, καθώς αυτοί καταγράφουν τα πιο σημαντικά φασματικά χαρακτηριστικά του ηχητικού σήματος. Αυτοί οι συντελεστές, που τώρα αναφέρονται ως MFCC, σχηματίζουν μια συμπαγή αναπαράσταση του φασματικού φακέλου του ηχητικού σήματος, το οποίο είναι ανθεκτικό στις διακυμάνσεις της ομιλίας και του περιβαλλοντικού θορύβου, καθιστώντας τους εξαιρετικά αποτελεσματικούς για εργασίες όπως η αναγνώριση ομιλίας και η ανίχνευση συναισθημάτων.



Σχήμα 2.5 Εφαρμογές των MFCCs

2.2.6 Ρυθμικά Χαρακτηριστικά (Rhythmic Features)

Τα ρυθμικά χαρακτηριστικά ενός ηχητικού σήματος περιγράφουν τις χρονικές ιδιότητες και τα μοτίβα του σήματος που επαναλαμβάνονται σε συγκεκριμένα χρονικά διαστήματα. Αυτά τα χαρακτηριστικά παίζουν καθοριστικό ρόλο στην αντίληψη της μουσικής και της ομιλίας, καθορίζοντας τη δομή και τη δυναμική του ήχου. Μερικά από τα βασικά ρυθμικά χαρακτηριστικά περιλαμβάνουν τον ρυθμό (tempo), την περιοδικότητα, το beat, το μέτρο (meter) και τα μοτίβα (rhythmic patterns).

Ρυθμός (Tempo)

Ο ρυθμός είναι η ταχύτητα με την οποία εκτελείται ένα μουσικό κομμάτι και εκφράζεται σε παλμούς ανά λεπτό (beats per minute - BPM). Ο ρυθμός επηρεάζει την αίσθηση και την ενέργεια του κομματιού, με ταχύτερους ρυθμούς να προκαλούν αισθήματα ενθουσιασμού και πιο αργούς ρυθμούς να δημιουργούν χαλαρή ή στοχαστική ατμόσφαιρα.

Beat

Το beat είναι η θεμελιώδης μονάδα χρόνου σε ένα μουσικό κομμάτι, η οποία συχνά αντιστοιχεί στους παλμούς που ο ακροατής μπορεί να κουνήσει το κεφάλι ή το πόδι του. Το beat διαχωρίζεται σε ισχυρά και ασθενή, δημιουργώντας τη βάση για το μέτρο και τα ρυθμικά μοτίβα.

Μέτρο (Meter)

Το μέτρο καθορίζει το μοτίβο των ισχυρών και ασθενών beats σε ένα μουσικό κομμάτι, συνήθως σε ομάδες των δύο, τριών ή τεσσάρων beats. Το πιο κοινό μέτρο είναι το τετραμερές (4/4), όπου τέσσερα beats δημιουργούν έναν κύκλο, με το πρώτο beat να είναι το ισχυρότερο. Το μέτρο επηρεάζει τη ροή και τη δομή του κομματιού.

Ρυθμικά Μοτίβα (Rhythmic Patterns)

Τα ρυθμικά μοτίβα είναι οι επαναλαμβανόμενες ακολουθίες των beats και των σιωπών που δημιουργούν τη ρυθμική δομή ενός μουσικού κομματιού. Αυτά τα μοτίβα μπορούν να είναι απλά ή σύνθετα και διαδραματίζουν σημαντικό ρόλο στη δημιουργία ποικιλίας και ενδιαφέροντος στη μουσική.

Περιοδικότητα

Η περιοδικότητα αναφέρεται στην επαναλαμβανόμενη φύση των ρυθμικών μοτίβων. Σε ένα ηχητικό σήμα, η περιοδικότητα μπορεί να αναλυθεί χρησιμοποιώντας μεθόδους όπως η Ανάλυση Fourier για να προσδιοριστεί η κυρίαρχη συχνότητα των επαναλαμβανόμενων μοτίβων. Αυτή η πληροφορία είναι ιδιαίτερα χρήσιμη στην ανάλυση της μουσικής δομής και του συναισθηματικού της αντίκτυπου.

2.2.7 Ποιότητα Φωνής (Voice Quality)

Η ποιότητα της φωνής μπορεί να θεωρηθεί ως ένα σημαντικό χαρακτηριστικό ήχου για εξαγωγή σε εργασίες μηχανικής μάθησης που σχετίζονται με την επεξεργασία και ανάλυση του λόγου. Η ποιότητα της φωνής αναφέρεται στα αντιληπτικά χαρακτηριστικά της φωνής ενός ομιλητή, τα οποία περιλαμβάνουν χαρακτηριστικά όπως η τονικότητα, το ηχόχρωμα (timbre), η ένταση και η καθαρότητα. Στην μηχανική μάθηση, η ποιότητα της φωνής μπορεί να ποσοτικοποιηθεί και να εξαχθεί από τα ηχητικά σήματα χρησιμοποιώντας διάφορες τεχνικές επεξεργασίας σήματος και μεθόδους εξαγωγής χαρακτηριστικών.

i) Η τονικότητα αναφέρεται στο ποιοτικό χαρακτηριστικό της φωνής που σχετίζεται με την υψηλότητα ή χαμηλότητα των ήχων που παράγει ένας ομιλητής. Αναπαριστά την παράμετρο της συχνότητας του ήχου και καθορίζει εάν μια φωνή ακούγεται μεγαλύτερη ή μικρότερη σε σχέση με μια αναφερόμενη νότα. Η τονικότητα είναι σημαντική για την κατανόηση της μελωδικής δομής της φωνής και τη διάκριση μεταξύ διαφορετικών παρουσιάσεων λόγου.

ii) Το ηχόχρωμα αφορά τον χαρακτηριστικό τόνο της φωνής που καθορίζει την ποιοτική της αντίληψη και τη διακριτική της φύση. Είναι το συνολικό αποτέλεσμα των συχνοτήτων και των εντάσεων των διαφόρων συνιστωσών της φωνής, καθώς και της απόκρισης του ακουστικού περιβάλλοντος, συνεπώς παρέχει πληροφορίες σχετικά με την πλούσια ποικιλία των ήχων που παράγονται από μια φωνή, και μπορεί να χρησιμοποιηθεί για να αναγνωριστούν οι ιδιαιτερότητες της φωνής ενός ομιλητή.

(iii) Η ένταση αφορά την ισχύ της φωνής, δηλαδή πόσο δυνατά ακούγεται ο ήχος που παράγεται από έναν ομιλητή. Είναι ένας κρίσιμος παράγοντας που καθορίζει το επίπεδο αντίληψης και προσοχής του ακροατή. Μια μεγάλη ένταση μπορεί να χρησιμοποιηθεί για να

επισημανθεί η σημασία ενός μηνύματος ή να μεταδοθεί μια έντονη συναισθηματική κατάσταση.

(iv) Η καθαρότητα αναφέρεται στην ομαλότητα και την ευκρίνεια της φωνής, που επηρεάζεται από παράγοντες όπως η προφορά, η εκφορά και η ποιότητα του ήχου. Μια καθαρή φωνή είναι ευκολότερο να αναγνωριστεί και να κατανοηθεί από τους ακροατές, ενώ μια μη καθαρή φωνή μπορεί να οδηγήσει σε παρερμηνείες και παραπλανητική ερμηνεία του λόγου.

Ένας συνηθισμένος τρόπος εξαγωγής χαρακτηριστικών ποιότητας φωνής περιλαμβάνει την ανάλυση του φάσματος του ήχου. Τεχνικές όπως η ανάλυση Fourier και ο μετασχηματισμός Wavelet μπορούν να χρησιμοποιηθούν για την αποσύνθεση του ηχητικού σήματος σε τις συνιστώσες συχνότητας, αποκαλύπτοντας πληροφορίες σχετικά με την τονικότητα και τα φασματικά χαρακτηριστικά της φωνής.

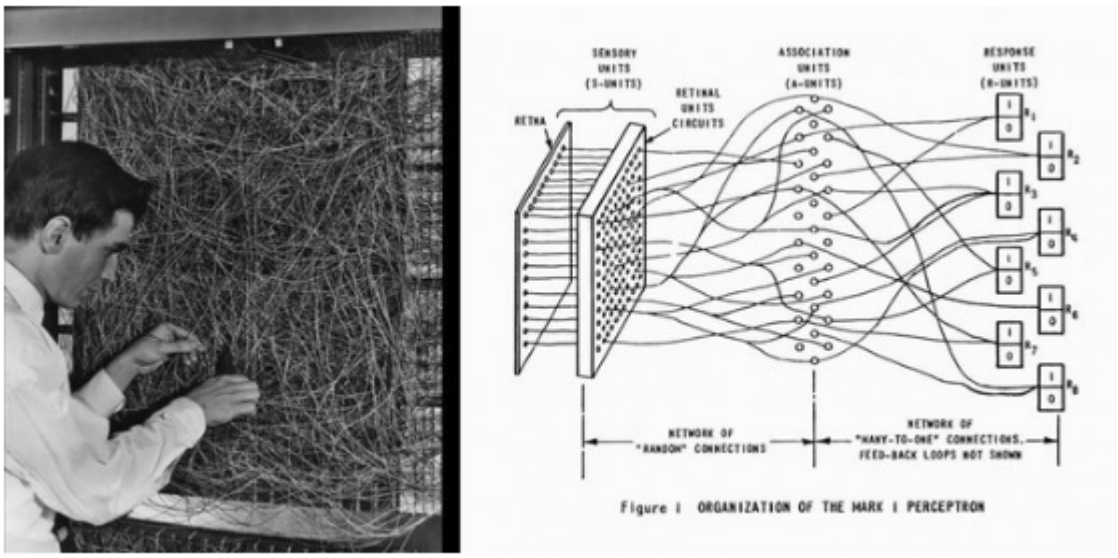
Επιπλέον, τα χαρακτηριστικά που προέρχονται από τον χώρο του χρόνου μπορούν να καταγράψουν δυναμικές πτυχές της ποιότητας της φωνής με το πέρασμα του χρόνου. Αυτά τα χαρακτηριστικά μπορεί να περιλαμβάνουν μετρήσεις της μεταβλητότητας του σήματος, όπως η ταλάντωση (μεταβολή συχνότητας κύκλου προς κύκλο) και η τρεμοπαίδωση (μεταβολή της ταλάντωσης κύκλου προς κύκλο), που είναι ενδεικτικές της ευστάθειας και της ομαλότητας της φωνής.

Κεφάλαιο 3. Μηχανική Μάθηση

3.1 Ιστορική αναδρομή

Η μηχανική μάθηση είναι η επιστήμη που δίνει τη δυνατότητα στους υπολογιστές να μαθαίνουν από δεδομένα και να βελτιώνονται με την πάροδο του χρόνου χωρίς να προγραμματίζονται ρητά, έχει μια πλούσια και συναρπαστική ιστορία που εκτείνεται σε αρκετές δεκαετίες. Από την ταπεινή αρχή της που βασίζεται στη μελέτη των νευρωνικών δικτύων μέχρι την τρέχουσα εποχή της βαθιάς μάθησης και της τεχνητής νοημοσύνης, η μηχανική μάθηση έχει υποστεί αξιοσημείωτη εξέλιξη με γνώμονα την πρόοδο στη θεωρία, τους αλγόριθμους, την υπολογιστική ισχύ και τη διαθεσιμότητα δεδομένων.

Οι απαρχές της μηχανικής μάθησης εντοπίζονται στα μέσα του 20ου αιώνα, με πρωτοποριακή εργασία στον τομέα των νευρωνικών δικτύων. Εμπνευσμένοι από τους βιολογικούς νευρώνες του ανθρώπινου εγκεφάλου, οι πρώτοι ερευνητές όπως ο Warren McCulloch και ο Walter Pitts έθεσαν τις βάσεις για υπολογιστικά μοντέλα νευρωνικών δικτύων τη δεκαετία του 1940. Αυτή η εποχή είδε την ανάπτυξη θεμελιωδών εννοιών όπως το perceptron, ένα απλό νευρωνικό δίκτυο ικανό να μάθει να ταξινομεί τα δεδομένα εισόδου σε δυαδικές κατηγορίες, που εισήχθη από τον Frank Rosenblatt το 1957.



Σχήμα 3.1 Ο Frank Rosenblatt με το Mark I perceptron του (αριστερά) και μια γραφική αναπαράστασή του (δεξιά).

Ωστόσο, η πρόοδος στη μηχανική μάθηση ήταν αργή κατά τη διάρκεια των δεκαετιών του 1960 και του 1970, καθώς οι ερευνητές επικεντρώθηκαν κυρίως σε συμβολικές προσεγγίσεις της τεχνητής νοημοσύνης (AI), όπως τα έμπειρα συστήματα και η λογική βασισμένη σε κανόνες. Μόλις τη δεκαετία του 1980 αναζωπυρώθηκε το ενδιαφέρον για τα νευρωνικά δίκτυα με την ανακάλυψη του αλγορίθμου backpropagation, ο οποίος επέτρεψε την αποτελεσματική εκπαίδευση πολυστρωματικών perceptrons. Αυτό σηματοδότησε την αρχή μιας νέας εποχής στη μηχανική μάθηση, που χαρακτηρίζεται από την αναβίωση των προσεγγίσεων που βασίζονται σε νευρωνικά δίκτυα.

Κατά τη διάρκεια της δεκαετίας του 1990 η μηχανική μάθηση γνώρισε σημαντικές προόδους στους αλγόριθμους και τις τεχνικές. Οι μηχανές διανυσμάτων υποστήριξης (SVM), που αναπτύχθηκαν από τον Vladimir Vapnik και άλλους, αναδείχθηκαν ως ισχυρά εργαλεία για εργασίες ταξινόμησης και παλινδρόμησης, ενώ οι μέθοδοι συνόλου όπως τα Τυχαία Δάση κέρδισαν δημοτικότητα για την ικανότητά τους να βελτιώνουν την προγνωστική απόδοση συνδυάζοντας πολλαπλά μοντέλα. Επιπρόσθετα, το πεδίο είδε την άνοδο των δικτύων Bayes για πιθανοτικό συλλογισμό και λήψη αποφάσεων υπό αβεβαιότητα.

Η αλλαγή της χιλιετίας εγκαινίασε μια περίοδο άνευ προηγουμένου ανάπτυξης και καινοτομίας στη μηχανική μάθηση. Οι καινοτομίες στην υπολογιστική ισχύ, που τροφοδοτήθηκαν από τον νόμο του Moore και την εμφάνιση των μονάδων γραφικής επεξεργασίας (GPU), επέτρεψαν την εκπαίδευση μεγαλύτερων και πιο περίπλοκων μοντέλων σε τεράστιες ποσότητες δεδομένων. Η βαθιά μάθηση, ένα υποπεδίο της μηχανικής μάθησης που επικεντρώνεται σε νευρωνικά δίκτυα με πολλαπλά επίπεδα (εξ ου και ο όρος "βαθύ"), αναδείχθηκε ως κυρίαρχο παράδειγμα, φέρνοντας επανάσταση σε τομείς όπως η υπολογιστική όραση, η επεξεργασία φυσικής γλώσσας και η αναγνώριση ομιλίας.

Η δεκαετία του 2010 σηματοδότησε τη χρυσή εποχή της βαθιάς μάθησης, με την ευρεία υιοθέτηση των συνελικτικών νευρωνικών δικτύων (CNN) για εργασίες ταξινόμησης εικόνων, των επαναλαμβανόμενων νευρωνικών δικτύων (RNN) για διαδοχική ανάλυση δεδομένων και των μοντέλων μετασχηματιστών για την κατανόηση και τη δημιουργία γλώσσας. Αυτές οι εξελίξεις οδηγήθηκαν από ανακαλύψεις σε αλγοριθμικές καινοτομίες, όπως η εισαγωγή διορθωμένων γραμμικών μονάδων (ReLU) ως συναρτήσεων ενεργοποίησης, η ανάπτυξη προηγμένων τεχνικών βελτιστοποίησης όπως ο Adam και η επεκτασιμότητα κατανεμημένων πλαισίων εκπαίδευσης όπως το TensorFlow και το PyTorch.

Τα τελευταία χρόνια, η μηχανική μάθηση έχει γίνει ολοένα και περισσότερο συνυφασμένη με άλλες τεχνολογίες, όπως τα μεγάλα δεδομένα, το cloud computing και το edge computing. Οι εφαρμογές της μηχανικής μάθησης καλύπτουν ένα ευρύ φάσμα τομέων, όπως η υγειονομική περίθαλψη, τα οικονομικά, οι μεταφορές, η ψυχαγωγία και πολλά άλλα. Από εξατομικευμένα συστήματα συστάσεων έως αυτόνομα οχήματα, από ιατρική διάγνωση έως ανίχνευση απάτης, η μηχανική μάθηση έχει αλλάξει τον τρόπο με τον οποίο αλληλεπιδρούμε με την τεχνολογία και τον κόσμο γύρω μας.

3.2 Βασικές διεργασίες και τεχνικές

Προεπεξεργασία δεδομένων:

Η δημιουργία ενός ισχυρού μοντέλου μηχανικής μάθησης ξεκινά με την προεπεξεργασία δεδομένων, ένα βήμα κατά το οποίο τα ακατέργαστα δεδομένα μετασχηματίζονται και προετοιμάζονται για ανάλυση. Αυτό περιλαμβάνει τον καθαρισμό των δεδομένων για την αφαίρεση του θορύβου, τον χειρισμό τιμών που λείπουν μέσω καταλογισμού ή διαγραφής και την κωδικοποίηση κατηγορικών μεταβλητών σε αριθμητικές αναπαραστάσεις. Επιπλέον, μπορούν να εφαρμοστούν τεχνικές μηχανικής χαρακτηριστικών για την εξαγωγή σχετικών πληροφοριών και τη δημιουργία νέων χαρακτηριστικών που ενισχύουν την προγνωστική ισχύ του μοντέλου. Διασφαλίζοντας την ποιότητα και τη συνάφεια των δεδομένων εισόδου.

Επιλογή μοντέλου:

Η επιλογή του καταλληλότερου αλγορίθμου ή αρχιτεκτονικής μοντέλου είναι μια βασική απόφαση που επηρεάζει σημαντικά την απόδοση και τις δυνατότητες γενίκευσης ενός συστήματος μηχανικής μάθησης. Αυτή η διαδικασία περιλαμβάνει την κατανόηση του τομέα του προβλήματος, την εξερεύνηση των χαρακτηριστικών του συνόλου δεδομένων και τον πειραματισμό με διαφορετικούς αλγόριθμους για τον εντοπισμό του μοντέλου με την καλύτερη απόδοση. Παράγοντες όπως η φύση των δεδομένων (π.χ. δομημένα ή μη), το μέγεθος του συνόλου δεδομένων και η πολυπλοκότητα του προβλήματος επηρεάζουν την επιλογή του αλγορίθμου, είτε πρόκειται για δέντρα απόφασης, μηχανές υποστήριξης διανυσμάτων, νευρωνικά δίκτυα ή μέθοδοι συνόλου . Η επιλογή μοντέλου είναι ταυτόχρονα τέχνη και επιστήμη, που απαιτεί βαθιά κατανόηση των αρχών της μηχανικής μάθησης και πρακτική εμπειρία στην αξιολόγηση και σύγκριση μοντέλων.

Εκπαίδευση:

Μόλις επιλεγεί ένα κατάλληλο μοντέλο, το επόμενο βήμα είναι να εκπαιδευτεί χρησιμοποιώντας τα διαθέσιμα δεδομένα. Η εκπαίδευση περιλαμβάνει τη βελτιστοποίηση των παραμέτρων του μοντέλου για την ελαχιστοποίηση μιας επιλεγμένης αντικειμενικής συνάρτησης, συνήθως μέσω επαναληπτικών αλγορίθμων όπως η κατάβαση κλίσης ή η αντίστροφη διάδοση. Κατά τη διάρκεια της εκπαίδευσης, το μοντέλο μαθαίνει να αποτυπώνει μοτίβα και σχέσεις μέσα στα δεδομένα, προσαρμόζοντας τις εσωτερικές του παραμέτρους για να κάνει ακριβείς προβλέψεις σε αόρατα παραδείγματα. Η αποτελεσματικότητα της εκπαιδευτικής διαδικασίας εξαρτάται από παράγοντες όπως η επιλογή του αλγορίθμου βελτιστοποίησης, η ποιότητα και η ποσότητα των δεδομένων εκπαίδευσης και η πολυπλοκότητα της αρχιτεκτονικής του μοντέλου.

Εκτίμηση:

Η αξιολόγηση της απόδοσης ενός εκπαιδευμένου μοντέλου είναι απαραίτητη για τη μέτρηση της αποτελεσματικότητάς του και τον εντοπισμό τομέων προς βελτίωση. Οι μετρήσεις αξιολόγησης όπως η ακρίβεια, η ανάκληση, η βαθμολογία F1 και η περιοχή κάτω από την καμπύλη ROC παρέχουν ποσοτικές μετρήσεις της απόδοσης του μοντέλου σε διαφορετικά κριτήρια αξιολόγησης. Ανάλογα με τις ειδικές απαιτήσεις του τομέα προβλήματος, μπορεί να δοθεί προτεραιότητα σε διαφορετικές μετρήσεις για τη βελτιστοποίηση του μοντέλου για συγκεκριμένους στόχους, όπως η ελαχιστοποίηση των ψευδών θετικών αποτελεσμάτων στην ιατρική διάγνωση ή η μεγιστοποίηση των εσόδων από την πρόβλεψη πωλήσεων. Η αυστηρή αξιολόγηση διασφαλίζει ότι οι προβλέψεις του μοντέλου ευθυγραμμίζονται με τα επιθυμητά αποτελέσματα και ανταποκρίνονται στις προσδοκίες των ενδιαφερομένων.

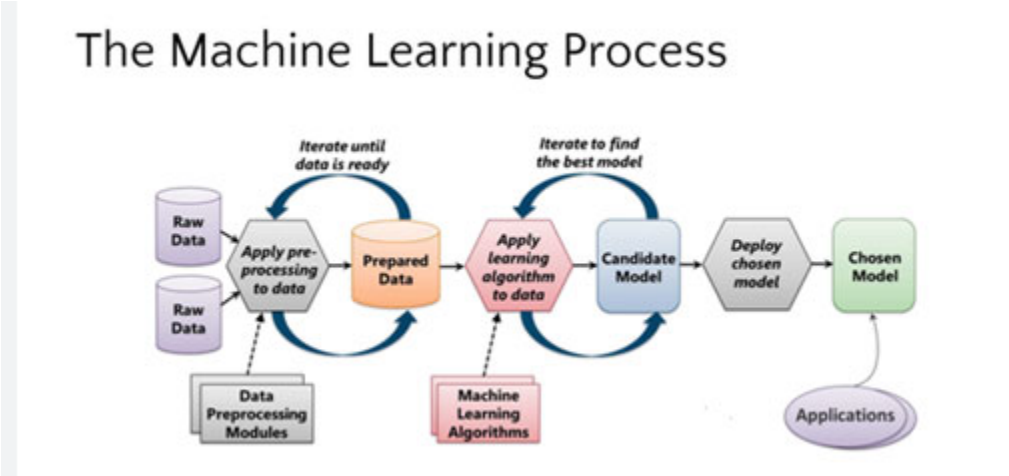
Ανάπτυξη:

Η εισαγωγή ενός εκπαιδευμένου μοντέλου στην παραγωγή περιλαμβάνει την ανάπτυξή του σε συστήματα πραγματικού κόσμου και την ενσωμάτωσή του με την υπάρχουσα υποδομή λογισμικού. Τα ζητήματα ανάπτυξης περιλαμβάνουν την επεκτασιμότητα, την αξιοπιστία και

την καθυστέρηση, καθώς και την ανάγκη για δυνατότητες έκδοσης και παρακολούθησης για την παρακολούθηση της απόδοσης του μοντέλου σε περιβάλλοντα παραγωγής. Τεχνολογίες όπως η αρχιτεκτονική εμπορευματοκιβωτίων και μικροϋπηρεσιών διευκολύνουν την ανάπτυξη μοντέλων μηχανικής εκμάθησης σε κλίμακα, επιτρέποντας την απρόσκοπτη ενσωμάτωση με εφαρμογές Ιστού, εφαρμογές για κινητές συσκευές και συσκευές IoT. Η επιτυχής ανάπτυξη διασφαλίζει ότι τα οφέλη της μηχανικής μάθησης υλοποιούνται σε πρακτικές εφαρμογές, οδηγώντας την αξία και την καινοτομία σε όλους τους κλάδους.

Παρακολούθηση και Συντήρηση:

Η ανάπτυξη όμως ενός μοντέλου δεν αποτελεί το τελικό στάδιο στην διαδικασία παραγωγής του. Αντίθετα, απαιτεί συνεχή παρακολούθηση και συντήρηση για να διασφαλιστεί η συνεχής συνάφεια και αποτελεσματικότητά του με την πάροδο του χρόνου. Η παρακολούθηση περιλαμβάνει την παρακολούθηση βασικών δεικτών απόδοσης, όπως η ακρίβεια της πρόβλεψης, η μετατόπιση του μοντέλου και η ποιότητα των δεδομένων, και η λήψη προληπτικών μέτρων για την αντιμετώπιση των ζητημάτων που προκύπτουν. Οι εργασίες συντήρησης μπορεί να περιλαμβάνουν την επανεκπαίδευση του μοντέλου σε νέα δεδομένα, τη λεπτομερή ρύθμιση υπερπαραμέτρων και την ενημέρωση της αρχιτεκτονικής του μοντέλου για προσαρμογή στις μεταβαλλόμενες απαιτήσεις ή περιβάλλοντα. Καθιερώνοντας ισχυρές διαδικασίες παρακολούθησης και συντήρησης, οι οργανισμοί μπορούν να μεγιστοποιήσουν τη μακροζωία και τον αντίκτυπο των επενδύσεών τους στη μηχανική μάθηση, διατηρώντας ένα ανταγωνιστικό πλεονέκτημα σε ένα συνεχώς εξελισσόμενο τοπίο.



Σχήμα 3.2 Διάγραμμα ροής διαδικασίας της Μηχανικής Μάθησης

3.3 Μέθοδοι μηχανικής μάθησης

Κατά την μηχανική μάθηση οι αλγόριθμοι ταξινομούνται ανάλογα με την διαδικασία κατά την οποία λαμβάνεται η μάθηση ή σχετικά με την δομή του συστήματος ανατροφοδότησης του.

Υπάρχουν δύο βασικές μέθοδοι που χρησιμοποιούνται.

3.3.1 Επιβλεπόμενη μάθηση

Η επιβλεπόμενη μάθηση είναι μια προσέγγιση της μηχανικής μάθησης που ορίζεται από τη χρήση επισημασμένων συνόλων δεδομένων. Αυτά τα σύνολα δεδομένων έχουν σχεδιαστεί για να εκπαιδεύουν ή να εποπτεύουν αλγόριθμους για την ταξινόμηση δεδομένων ή την ακριβή πρόβλεψη των αποτελεσμάτων. Χρησιμοποιώντας εισόδους και εξόδους με ετικέτα, το μοντέλο μπορεί να μετρήσει την ακρίβειά του και να μάθει με την πάροδο του χρόνου.

Η επιβλεπόμενη μάθηση είναι καλή σε προβλήματα όπως :

(i) Αναγνώριση εικόνων και αντικειμένων: Οι αλγόριθμοι μάθησης με επίβλεψη μπορούν να χρησιμοποιηθούν για την ανίχνευση,το διαχωρισμό και την ταξινόμηση αντικειμένων, βίντεο και εικόνων και είναι χρήσιμοι όταν εφαρμόζονται σε διάφορες τεχνικές ανάλυσης εικόνας και υπολογιστικής όρασης.

(ii) Προβλεπτική ανάλυση: Μια ευρέως διαδεδομένη περίπτωση χρήσης των μοντέλων μάθησης με επίβλεψη είναι η κατασκευή συστημάτων προβλεπτικής ανάλυσης που παρέχουν βαθιές γνώσεις σε διάφορα σημεία επιχειρηματικών δεδομένων. Αυτό επιτρέπει στις εταιρείες να προβλέπουν συγκεκριμένα αποτελέσματα με βάση δεδομένες μεταβλητές εξόδου, βοηθώντας τους επιχειρηματικούς ηγέτες να δικαιολογήσουν αποφάσεις ή να κάνουν στροφή προς όφελος του οργανισμού.

(iii) Ανάλυση συναισθήματος πελατών: Χρησιμοποιώντας αλγόριθμους μηχανικής μάθησης με επίβλεψη, οι εταιρείες μπορούν να εξάγουν και να κατηγοριοποιούν σημαντικές πληροφορίες από μεγάλες ποσότητες δεδομένων. Αυτό μπορεί να είναι πολύ χρήσιμο για την καλύτερη κατανόηση των αλληλεπιδράσεων των πελατών και μπορεί επίσης να χρησιμοποιηθεί για τη βελτίωση της πιστότητας της μάρκας.

(iv) Ανίχνευση ανεπιθύμητης αλληλογραφίας: Η ανίχνευση ανεπιθύμητης αλληλογραφίας είναι ένα άλλο παράδειγμα μοντέλου μάθησης με επίβλεψη. Με τη χρήση αλγορίθμων ταξινόμησης με επίβλεψη, οι εταιρείες μπορούν να εκπαιδεύσουν τις βάσεις δεδομένων τους ώστε να εντοπίζουν μοτίβα και ανωμαλίες σε νέα δεδομένα,διαχωρίζοντας αποτελεσματικά τα μηνύματα spam από τα μη.

Επιπρόσθετα υπάρχουν δύο κύριες κατηγορίες εποπτευόμενης μάθησης. Η ταξινόμηση (classification) όπου οι τιμές εξόδου είναι κατηγορικές και η παλινδρόμηση (regression) όπου οι τιμές εξόδου είναι αριθμητικές.

α) Αλγόριθμοι ταξινόμησης

Ένας αλγόριθμος ταξινόμησης στοχεύει στην ταξινόμηση των εισόδων σε έναν δεδομένο αριθμό κατηγοριών ή κλάσεων, με βάση τα επισημασμένα δεδομένα στα οποία εκπαιδεύτηκε. Οι αλγόριθμοι ταξινόμησης μπορούν να χρησιμοποιηθούν για δυαδικές ταξινομήσεις, όπως το

φιλτράρισμα μηνυμάτων ηλεκτρονικού ταχυδρομείου σε ανεπιθύμητα ή μη ανεπιθύμητα και η κατηγοριοποίηση των σχολίων των πελατών ως θετικών ή αρνητικών. Η αναγνώριση χαρακτηριστικών, όπως η αναγνώριση χειρόγραφων γραμμάτων και αριθμών ή η ταξινόμηση των φαρμάκων σε πολλές διαφορετικές κατηγορίες, είναι ένα άλλο πρόβλημα ταξινόμησης που επιλύεται με την επιβλεπόμενη μάθηση.

β) Αλγόριθμοι παλινδρόμησης

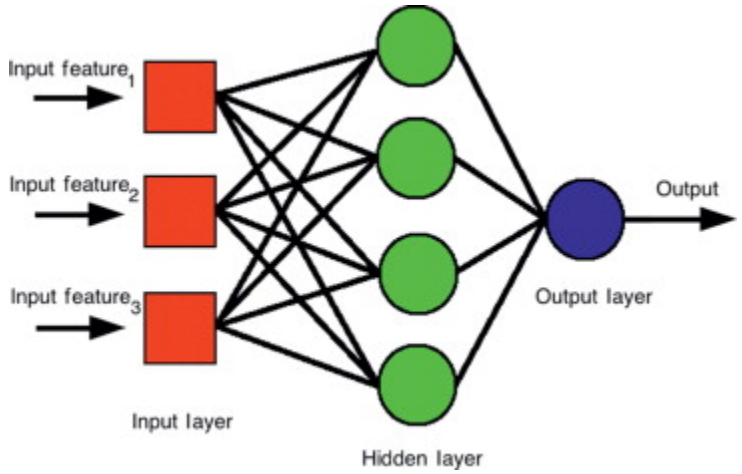
Οι εργασίες παλινδρόμησης είναι διαφορετικές, καθώς αναμένουν από το μοντέλο να παράγει μια αριθμητική σχέση μεταξύ των δεδομένων εισόδου και εξόδου. Παραδείγματα μοντέλων παλινδρόμησης περιλαμβάνουν την πρόβλεψη τιμών ακινήτων με βάση την τοποθεσία ή την πρόβλεψη ποσοστών προβολών σε τηλεοπτικά κανάλια σε σχέση με την ώρα της ημέρας ή τον προσδιορισμό του ποσού των τηλεθεατών που παρακολουθούν το κάθε πρόγραμμα με βάση την ηλικία τους.

3.3.2 Παραδείγματα αλγορίθμων επιβλεπόμενης μάθησης

- **Νευρωνικά δίκτυα (MLP):**

Χρησιμοποιούνται κυρίως για αλγόριθμους βαθιάς μάθησης, τα νευρωνικά δίκτυα επεξεργάζονται δεδομένα εκπαίδευσης μιμούμενοι τη διασύνδεση του ανθρώπινου εγκεφάλου μέσω στρωμάτων κόμβων. Κάθε κόμβος αποτελείται από εισόδους, βάρη, μεροληψία (ή κατώφλι) και έξοδο.

Εάν αυτή η τιμή εξόδου υπερβεί ένα δεδομένο όριο, «πυροδοτεί» ή ενεργοποιεί τον κόμβο, περνώντας δεδομένα στο επόμενο επίπεδο του δικτύου. Τα νευρωνικά δίκτυα μαθαίνουν αυτή τη λειτουργία χαρτογράφησης μέσω της εποπτευόμενης μάθησης, προσαρμόζοντας με βάση τη συνάρτηση απώλειας μέσω της διαδικασίας gradient descent. Όταν η συνάρτηση κόστους είναι στο μηδέν ή κοντά στο μηδέν,μπορούμε να είμαστε σίγουροι για την ακρίβεια του μοντέλου για να δώσουμε τη σωστή απάντηση.



Σχήμα 3.3 Αρχιτεκτονική ενός νευρωνικού δικτύου

Ένα βασικό νευρωνικό δίκτυο έχει διασυνδεδεμένους τεχνητούς νευρώνες σε τρία στρώματα. Αρχικά έχουμε το επίπεδο εισόδου στο οποίο οι πληροφορίες από τον έξω κόσμο εισέρχονται στο τεχνητό νευρωνικό δίκτυο από το στρώμα εισόδου. Οι κόμβοι εισόδου επεξεργάζονται τα δεδομένα, τα αναλύουν ή τα κατηγοριοποιούν και τα μεταβιβάζουν στο επόμενο επίπεδο.

Επιπρόσθετα έχουμε το κρυφό επίπεδο. Τα κρυφά επίπεδα λαμβάνουν την είσοδό τους από το επίπεδο εισόδου ή άλλα κρυφά επίπεδα. Τα τεχνητά νευρωνικά δίκτυα μπορούν να έχουν μεγάλο αριθμό κρυφών στρωμάτων. Κάθε κρυφό επίπεδο αναλύει την έξοδο από το προηγούμενο επίπεδο, την
επεξεργάζεται περαιτέρω και τη μεταβιβάζει στο επόμενο επίπεδο.

Τέλος το επίπεδο εξόδου δίνει το τελικό αποτέλεσμα όλης της επεξεργασίας δεδομένων από το τεχνητό νευρωνικό δίκτυο. Μπορεί να έχει μονούς ή πολλαπλούς κόμβους. Για παράδειγμα, αν έχουμε ένα δυαδικό πρόβλημα ταξινόμησης, το επίπεδο εξόδου θα έχει έναν κόμβο εξόδου, ο οποίος θα δώσει το αποτέλεσμα ως 1 ή 0. Ωστόσο, εάν έχουμε ένα πρόβλημα ταξινόμησης πολλαπλών κλάσεων, το επίπεδο εξόδου μπορεί να αποτελείται από περισσότερους από έναν κόμβους εξόδου.

- **<u>Naive Bayes:</u>**

Οι ταξινομητές Naive Bayes είναι μια οικογένεια απλών πιθανοτικών ταξινομητών που εφαρμόζουν το θεώρημα του Bayes με ισχυρές παραδοχές ανεξαρτησίας μεταξύ των χαρακτηριστικών. Παρά την απλότητά του, αποδίδει εκπληκτικά καλά σε διάφορες εφαρμογές όπως η ταξινόμηση κειμένων και η ιατρική διάγνωση.

Το θεώρημα του Bayes παρέχει έναν τρόπο για την ενημέρωση της εκτίμησης πιθανότητας για μια υπόθεση καθώς γίνεται διαθέσιμη περισσότερη απόδειξη. Μαθηματικά, εκφράζεται ως:

$$P(H|E) = P(E|H) \cdot \frac{P(H)}{P(E)} \tag{3.1}$$

όπου P(H|E) είναι η εκ των υστέρων πιθανότητα της υπόθεσης H δεδομένης της απόδειξης P, P(E|H) είναι η πιθανότητα της απόδειξης E δεδομένης της υπόθεσης H, P(H) είναι η προηγούμενη πιθανότητα της υπόθεσης H και P(E) είναι η οριακή πιθανότητα της απόδειξης E.

Ο ταξινομητής Naive Bayes υποθέτει ότι όλα τα χαρακτηριστικά είναι υπό συνθήκη ανεξάρτητα δεδομένης της ετικέτας της κατηγορίας κάτι που απλοποιεί τον υπολογισμό των εκ των υστέρων πιθανοτήτων. Υπάρχουν διάφοροι τύποι ταξινομητών Naive Bayes: ο Gaussian Naive Bayes υποθέτει ότι τα συνεχόμενα χαρακτηριστικά ακολουθούν μια κανονική κατανομή,

ο Multinomial Naive Bayes είναι κατάλληλος για δεδομένα διακριτών καταμετρήσεων και ο Bernoulli Naive Bayes είναι κατάλληλος για δυαδικά/Boolean χαρακτηριστικά.

Η εκπαίδευση του ταξινομητή περιλαμβάνει τον υπολογισμό των προηγούμενων πιθανοτήτων για κάθε κατηγορία και των πιθανοτήτων των χαρακτηριστικών δεδομένων κάθε κατηγορίας. Η πρόβλεψη περιλαμβάνει τον υπολογισμό των εκ των υστέρων πιθανοτήτων για κάθε κατηγορία για μια δεδομένη περίπτωση και την επιλογή της κατηγορίας με τη μεγαλύτερη εκ των υστέρων πιθανότητα.

Στην ταξινόμηση κειμένων όπως η ανίχνευση ανεπιθύμητων μηνυμάτων, κάθε μήνυμα ηλεκτρονικού ταχυδρομείου αναπαρίσταται ως ένα σύνολο λέξεων (χαρακτηριστικά), και ο στόχος είναι να ταξινομηθεί ως "ανεπιθύμητο" ή "όχι ανεπιθύμητο" . Κατά την εκπαίδευση, ο ταξινομητής υπολογίζει τις προηγούμενες πιθανότητες και την πιθανότητα κάθε λέξης δεδομένης της κατηγορίας. Για ένα νέο μήνυμα ηλεκτρονικού ταχυδρομείου, ο ταξινομητής υπολογίζει τις εκ των υστέρων πιθανότητες για τις δύο κατηγορίες και το ταξινομεί βάσει της μεγαλύτερης εκ των υστέρων πιθανότητας.

Ο ταξινομητής Naive Bayes έχει πολλά πλεονεκτήματα, συμπεριλαμβανομένης της απλότητας, της ευκολίας υλοποίησης και της αποτελεσματικότητας με δεδομένα υψηλής διάστασης. Ωστόσο, η υπόθεσή του για συνθήκη ανεξαρτησίας μεταξύ των χαρακτηριστικών συχνά δεν ισχύει σε πραγματικά σενάρια, κάτι που μπορεί να περιορίσει την απόδοσή του. Επιπλέον, ο ταξινομητής μπορεί να αποδώσει άσχημα εάν παραβιαστεί αυτή η υπόθεση ανεξαρτησίας.

- **<u>Γραμμική παλινδρόμηση:</u>**

Η Γραμμική Παλινδρόμηση είναι μια θεμελιώδης τεχνική μηχανικής μάθησης και στατιστικής που χρησιμοποιείται για την πρόβλεψη της τιμής μιας μεταβλητής στόχου, βασιζόμενη στις τιμές μίας ή περισσότερων εξηγητικών μεταβλητών. Η βασική ιδέα πίσω από τη γραμμική παλινδρόμηση είναι η εύρεση της καλύτερης δυνατής γραμμής που περνάει μέσα από τα δεδομένα και ελαχιστοποιεί την απόσταση μεταξύ των παρατηρούμενων τιμών και των προβλεπόμενων τιμών.

Η πιο απλή μορφή γραμμικής παλινδρόμησης είναι η απλή γραμμική παλινδρόμηση, η οποία χρησιμοποιείται όταν υπάρχει μόνο μία εξηγητική μεταβλητή. Το μοντέλο της απλής γραμμικής παλινδρόμησης έχει τη μορφή :

$$y = \beta_0 + \beta_1 X \tag{3.2}$$

όπου y είναι η μεταβλητή στόχος, x είναι η εξηγητική μεταβλητή, β0 είναι η τομή με τον άξονα y και β1 είναι η κλίση της γραμμής. Η διαδικασία εκπαίδευσης του μοντέλου περιλαμβάνει την εκτίμηση των παραμέτρων β0 και β1 έτσι ώστε να ελαχιστοποιηθεί το άθροισμα των τετραγωνικών σφαλμάτων μεταξύ των παρατηρούμενων και των προβλεπόμενων τιμών.

Στην πολυμεταβλητή γραμμική παλινδρόμηση, χρησιμοποιούνται περισσότερες από μία

εξηγητικές μεταβλητές και το μοντέλο έχει τη μορφή:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (3.3)$$

όπου x1, x2, ..., xn είναι οι εξηγητικές μεταβλητές. Η διαδικασία εκπαίδευσης περιλαμβάνει την εκτίμηση των παραμέτρων β0, β1,...,βn χρησιμοποιώντας μεθόδους όπως η Μέθοδος των Ελαχίστων Τετραγώνων, η οποία επιδιώκει να ελαχιστοποιήσει το άθροισμα των τετραγωνικών σφαλμάτων.

Η γραμμική παλινδρόμηση υποθέτει ότι υπάρχει μια γραμμική σχέση μεταξύ της μεταβλητής στόχου και των εξηγητικών μεταβλητών, ότι τα σφάλματα ακολουθούν κανονική κατανομή με μέση τιμή μηδέν και σταθερή διακύμανση (ομοσκεδαστικότητα) και ότι οι παρατηρήσεις είναι ανεξάρτητες. Ωστόσο, στην πράξη, αυτές οι υποθέσεις μπορεί να παραβιάζονται και διάφορες τεχνικές, όπως η κανονικοποίηση (regularization) και η ανάλυση υπολειμμάτων, χρησιμοποιούνται για τη βελτίωση της ακρίβειας και της αξιοπιστίας των μοντέλων γραμμικής παλινδρόμησης.

- **<u>Λογιστική παλινδρόμηση:</u>**

Ενώ η γραμμική παλινδρόμηση χρησιμοποιείται όταν οι εξαρτημένες μεταβλητές είναι συνεχείς, η λογιστική παλινδρόμηση επιλέγεται όταν η εξαρτημένη μεταβλητή είναι κατηγορική, που σημαίνει ότι έχουν δυαδικές εξόδους, όπως "true" και "false" ή "ναι" και "no". Αντί να μοντελοποιεί μια γραμμική σχέση, όπως η γραμμική παλινδρόμηση, η λογιστική παλινδρόμηση χρησιμοποιεί τη συνάρτηση sigmoid για να περιορίσει τις προβλέψεις μεταξύ 0 και 1, οι οποίες αντιπροσωπεύουν τις πιθανότητες των δύο κατηγοριών.Το μοντέλο λογιστικής παλινδρόμησης έχει τη μορφή :

$$P(Y=1|X) = \frac{1}{1 + e^{-\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n\right)}} \quad (3.4)$$

όπου P(Y=1|X) είναι η πιθανότητα το αποτέλεσμα να είναι 1, δεδομένων των εξηγητικών μεταβλητών X1,X2,...,Xn, β0, είναι η σταθερά και β1,β2,...,βn είναι οι συντελεστές των εξηγητικών μεταβλητών.

Η διαδικασία εκπαίδευσης του μοντέλου λογιστικής παλινδρόμησης περιλαμβάνει την εκτίμηση των παραμέτρων β1,β2,...,βn μέσω της μεγιστοποίησης της συνάρτησης πιθανοφάνειας (likelihood function), που μετρά την πιθανότητα παρατήρησης των δεδομένων δεδομένων των παραμέτρων του μοντέλου. Η μέθοδος μεγιστοποίησης πιθανοφάνειας χρησιμοποιεί συχνά αλγόριθμους όπως η Στοχαστική Κατηφορική Βαθμίδωση (Stochastic Gradient Descent, SGD) για την εύρεση των βέλτιστων παραμέτρων που μεγιστοποιούν την πιθανοφάνεια.

Η λογιστική παλινδρόμηση υποθέτει ότι τα δεδομένα είναι ανεξάρτητα και ότι υπάρχει μια γραμμική σχέση μεταξύ των εξηγητικών μεταβλητών και του λογιστικού μετασχηματισμού της πιθανότητας του αποτελέσματος. Παρόλο που η γραμμική αυτή σχέση μπορεί να είναι μια απλοποιητική υπόθεση, η λογιστική παλινδρόμηση είναι ιδιαίτερα αποτελεσματική για πολλά πραγματικά προβλήματα.

- **Support Vector Machines (SVM):**

Τα Support Vector Machines (SVMs) χρησιμοποιούνται για ταξινόμηση και παλινδρόμηση. Στοχεύουν στην εύρεση του βέλτιστου υπερπλάνου που διαχωρίζει τα δεδομένα σε διαφορετικές κατηγορίες με το μέγιστο περιθώριο. Το υπερπλάνο είναι μια γραμμή (σε δύο διαστάσεις), ένα επίπεδο (σε τρεις διαστάσεις) ή ένα υπερεπίπεδο (σε περισσότερες από τρεις διαστάσεις) που χωρίζει τα δεδομένα.

Η βασική ιδέα πίσω από τα SVMs είναι να βρεθεί ένα υπερπλάνο που όχι μόνο διαχωρίζει τα δεδομένα αλλά το κάνει με το μεγαλύτερο δυνατό περιθώριο, δηλαδή την μεγαλύτερη απόσταση μεταξύ των πιο κοντινών σημείων δεδομένων των δύο κατηγοριών, τα οποία ονομάζονται υποστηρικτικά διανύσματα. Η γενική μορφή του προβλήματος βελτιστοποίησης που λύνουν τα SVMs είναι η μεγιστοποίηση του περιθωρίου ενώ ταυτόχρονα ελαχιστοποιείται το λάθος ταξινόμησης.

Για γραμμικά διαχωρίσιμα δεδομένα, ένα SVM αναζητά το βέλτιστο γραμμικό υπερπλάνο. Ωστόσο, πολλά πραγματικά προβλήματα δεν είναι γραμμικά διαχωρίσιμα. Σε αυτές τις περιπτώσεις, τα SVMs μπορούν να χρησιμοποιήσουν μια τεχνική που ονομάζεται πυρηνική μέθοδος (kernel trick) για να χαρτογραφήσουν τα δεδομένα σε έναν υψηλότερης διάστασης χώρο όπου είναι πιθανότερο να είναι γραμμικά διαχωρίσιμα. Κοινά χρησιμοποιούμενοι πυρήνες περιλαμβάνουν τον γραμμικό πυρήνα, τον πολυωνυμικό πυρήνα, τον πυρήνα RBF (Radial Basis Function) και τον πυρήνα sigmoid.

Η διαδικασία εκπαίδευσης ενός SVM περιλαμβάνει την επίλυση ενός προβλήματος βελτιστοποίησης. Για τα γραμμικά SVMs, αυτό το πρόβλημα μπορεί να επιλυθεί χρησιμοποιώντας μεθόδους όπως ο αλγόριθμος του Lagrange και οι μέθοδοι διαδοχικών μικρών βημάτων (Sequential Minimal Optimization, SMO). Για τα μη γραμμικά SVMs, οι πυρηνικές μέθοδοι επιτρέπουν την αποδοτική υπολογιστική επεξεργασία των εσωτερικών γινομένων στο υψηλότερης διάστασης χώρο χωρίς την ανάγκη για ρητή υπολογιστική μετάβαση σε αυτόν τον χώρο.

Τα SVMs είναι γνωστά για την καλή γενίκευση τους, δηλαδή την ικανότητά τους να αποδίδουν καλά σε νέα, αόρατα δεδομένα. Αυτή η ιδιότητα οφείλεται στο γεγονός ότι το μοντέλο εστιάζει μόνο στα υποστηρικτικά διανύσματα, τα πιο ενημερωτικά δείγματα από τα δεδομένα εκπαίδευσης, και όχι σε όλα τα δεδομένα. Ωστόσο, η απόδοση των SVMs μπορεί να επηρεαστεί από την επιλογή των παραμέτρων, όπως το κανονιστικό παράμετρο (C) και οι παράμετροι του πυρήνα, κάτι που συχνά απαιτεί προσεκτική ρύθμιση μέσω διαδικασιών όπως η cross-validation.

- **K-nearest neighbors (KNN):**

Το KNN, είναι ένας μη παραμετρικός αλγόριθμος που ταξινομεί τα σημεία δεδομένων με βάση την εγγύτητα και τη συσχέτισή τους με άλλα διαθέσιμα δεδομένα. Βασίζεται στην υπόθεση ότι παρόμοια δεδομένα βρίσκονται κοντά το ένα στο άλλο στον χώρο των χαρακτηριστικών. Το KNN δεν κάνει υποθέσεις για την κατανομή των δεδομένων και είναι γνωστός ως αλγόριθμος μη παραμετρικός, πράγμα που σημαίνει ότι δεν κάνει ρητές υποθέσεις για τη μορφή της συνάρτησης από την οποία προέρχονται τα δεδομένα.

Στην ταξινόμηση με KNN, για να προβλεφθεί η κατηγορία ενός νέου δείγματος, ο αλγόριθμος υπολογίζει τις αποστάσεις του από όλα τα δείγματα στο σύνολο εκπαίδευσης και επιλέγει τους K πλησιέστερους γείτονες. Η κατηγορία του νέου δείγματος καθορίζεται από την πλειοψηφία των κατηγοριών αυτών των γειτόνων. Η πιο κοινή μέθοδος μέτρησης της απόστασης είναι η Ευκλείδεια απόσταση, αν και μπορούν να χρησιμοποιηθούν και άλλες μετρικές αποστάσεων όπως η Μανχάταν ή η απόσταση Minkowski.

Στην παλινδρόμηση με KNN, η πρόβλεψη της τιμής ενός νέου δείγματος γίνεται με βάση τις μέσες τιμές των K πλησιέστερων γειτόνων. Αυτό σημαίνει ότι αντί να χρησιμοποιείται η πλειοψηφία για την ταξινόμηση, ο KNN υπολογίζει τον μέσο όρο των τιμών των γειτόνων για να προβλέψει την τιμή της εξαρτημένης μεταβλητής.

Η επιλογή του αριθμού των γειτόνων, K, είναι κρίσιμη για την απόδοση του αλγόριθμου. Ένα μικρό K μπορεί να κάνει το μοντέλο ευαίσθητο στο θόρυβο των δεδομένων εκπαίδευσης, ενώ ένα μεγάλο K μπορεί να καταστήσει το μοντέλο υπερβολικά απλό και να μην αποτυπώνει καλά τις υποκείμενες δομές των δεδομένων. Συνήθως, η βέλτιστη τιμή του K επιλέγεται μέσω διαδικασιών όπως η cross-validation.

Ο αλγόριθμος KNN είναι γνωστός για την απλότητά του και την ευκολία κατανόησής του, αλλά έχει ορισμένα μειονεκτήματα. Ένα από τα κύρια μειονεκτήματα είναι η υπολογιστική του πολυπλοκότητα κατά την πρόβλεψη, ειδικά για μεγάλα σύνολα δεδομένων, αφού πρέπει να υπολογίσει τις αποστάσεις για όλα τα δείγματα του συνόλου εκπαίδευσης. Επιπλέον, η απόδοση του KNN μπορεί να επηρεαστεί από την κλίμακα των χαρακτηριστικών, γι' αυτό συχνά είναι αναγκαία η προεπεξεργασία των δεδομένων μέσω κανονικοποίησης ή τυποποίησης.

- **Τυχαίο δάσος:**

Τα τυχαία δάση (Random Forests) είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται τόσο για ταξινόμηση όσο και για παλινδρόμηση. Πρόκειται για μια μέθοδο ensemble που συνδυάζει πολλαπλά δέντρα απόφασης για να βελτιώσει την ακρίβεια των προβλέψεων και να μειώσει την υπερπροσαρμογή. Η βασική ιδέα πίσω από τα τυχαία δάση είναι να δημιουργηθούν πολλαπλά δέντρα απόφασης κατά τη διάρκεια της εκπαίδευσης και να

χρησιμοποιηθεί η πλειοψηφία των ψήφων (ή ο μέσος όρος στην περίπτωση της παλινδρόμησης) για την τελική πρόβλεψη.

Κάθε δέντρο απόφασης στο δάσος κατασκευάζεται χρησιμοποιώντας έναν διαφορετικό τυχαίο υποσύνολο των δεδομένων εκπαίδευσης και των χαρακτηριστικών. Αυτό επιτυγχάνεται μέσω δύο κύριων τεχνικών: της δειγματοληψίας με επανατοποθέτηση (bagging) και της τυχαίας δειγματοληψίας χαρακτηριστικών. Η δειγματοληψία με επανατοποθέτηση σημαίνει ότι κάθε δέντρο εκπαιδεύεται σε ένα τυχαίο υποσύνολο των δεδομένων εκπαίδευσης, με τα δείγματα να επιλέγονται τυχαία με επανατοποθέτηση. Η τυχαία δειγματοληψία χαρακτηριστικών σημαίνει ότι κατά τη διαδικασία διάσπασης κάθε κόμβου σε ένα δέντρο, επιλέγεται ένας τυχαίος υποσύνολο χαρακτηριστικών και χρησιμοποιείται το καλύτερο χαρακτηριστικό από αυτό το υποσύνολο για τη διάσπαση.

Η διαδικασία εκπαίδευσης των τυχαίων δασών περιλαμβάνει τη δημιουργία πολλών δέντρων απόφασης, το καθένα από τα οποία είναι "αδύναμος μαθητής" αλλά το σύνολο των δέντρων λειτουργεί ως "ισχυρός μαθητής". Κατά την πρόβλεψη, τα τυχαία δάση συνδυάζουν τις προβλέψεις όλων των δέντρων και χρησιμοποιούν την πλειοψηφία των ψήφων για την τελική απόφαση στην περίπτωση της ταξινόμησης ή τον μέσο όρο των προβλέψεων στην περίπτωση της παλινδρόμησης.

Τα τυχαία δάση έχουν πολλά πλεονεκτήματα, όπως η υψηλή ακρίβεια, η αντοχή στην υπερπροσαρμογή, και η δυνατότητα χειρισμού μεγάλου αριθμού χαρακτηριστικών. Ωστόσο, έχουν και κάποια μειονεκτήματα, όπως η ανάγκη για περισσότερους πόρους υπολογιστικής ισχύος και μνήμης, ειδικά όταν ο αριθμός των δέντρων είναι μεγάλος. Επίσης, η ερμηνευσιμότητα των μοντέλων τυχαίων δασών είναι συνήθως χαμηλότερη σε σύγκριση με τα μοντέλα ενός μόνο δέντρου απόφασης.

3.4 Μη επιβλεπόμενη μάθηση

Στην διερευνητική ανάλυση δεδομένων, συχνά δεν γνωρίζουμε τα πραγματικά “labels” ή μπορεί να θέλουμε να εξετάσουμε τα φυσικά αναδυόμενα μοτίβα στα δεδομένα. Για το σκοπό αυτό μπορούμε να χρησιμοποιήσουμε μη εποπτευόμενες μεθόδους μάθησης όπως η ομαδοποίηση(**clustering**) και η συχνή ανίχνευση μοτίβων και μείωση διαστάσεων (**dimensionality reduction**).

3.4.1 Ομαδοποίηση (Clustering)

Ο στόχος της εφαρμογής μεθόδων ομαδοποίησης είναι ο προσδιορισμός σχετικών υποομάδων σε ένα
δεδομένο σύνολο δεδομένων, χωρίς να υπάρχει μια προκαθορισμένη υπόθεση για τις ιδιότητες που μπορεί να έχουν οι υποομάδες. Οι αλγόριθμοι ομαδοποίησης χρησιμοποιούνται για την επεξεργασία ακατέργαστων, μη ταξινομημένων αντικειμένων δεδομένων σε ομάδες που αντιπροσωπεύονται από δομές ή μοτίβα στις πληροφορίες, ενώ κατηγοριοποιούνται σε αποκλειστικούς, επικαλυπτόμενους, ιεραρχικούς και πιθανολογικούς.

Ένα σύμπλεγμα(cluster) είναι ένα υποσύνολο των δεδομένων που είναι "παρόμοια" μεταξύ τους.Υπάρχουν πολλές προσεγγίσεις για την ομαδοποίηση που χρησιμοποιούν διαφορετικούς υποκείμενους αλγόριθμους για να ομαδοποιήσουν τα σημεία δεδομένων με βάση την "ομοιότητά" τους. Όλα αυτά έχουν πλεονεκτήματα και μειονεκτήματα και πρέπει να επιλέγονται προσεκτικά ανάλογα με την εφαρμογή και τις ιδιότητες των δεδομένων.

α) Αποκλειστική και επικαλυπτόμενη ομαδοποίηση

Η αποκλειστική ομαδοποίηση είναι μια μορφή ομαδοποίησης που ορίζει ότι ένα σημείο δεδομένων μπορεί να υπάρχει μόνο σε ένα σύμπλεγμα. Αυτό μπορεί επίσης να αναφέρεται ως «σκληρή» ομαδοποίηση. Ο αλγόριθμος ομαδοποίησης K-means είναι ένα παράδειγμα αποκλειστικής ομαδοποίησης. Η ομαδοποίηση K-means είναι ένα κοινό παράδειγμα μιας αποκλειστικής μεθόδου ομαδοποίησης όπου τα σημεία δεδομένων εκχωρούνται σε ομάδες K, όπου το K αντιπροσωπεύει τον αριθμό των ομάδων με βάση την απόσταση από το κέντρο κάθε ομάδας. Τα σημεία δεδομένων που βρίσκονται πιο κοντά σε ένα δεδομένο κέντρο θα συγκεντρωθούν στην ίδια κατηγορία. Μια μεγαλύτερη τιμή K θα είναι ενδεικτική για μικρότερες ομαδοποιήσεις με μεγαλύτερη ευαισθησία, ενώ μια μικρότερη τιμή K θα έχει μεγαλύτερες ομαδοποιήσεις και λιγότερη ευαισθησία. Η ομαδοποίηση K-means χρησιμοποιείται συνήθως στην τμηματοποίηση της αγοράς, τη ομαδοποίηση εγγράφων, την τμηματοποίηση εικόνας και τη συμπίεση εικόνας. Τα επικαλυπτόμενα συμπλέγματα διαφέρουν από την αποκλειστική ομαδοποίηση στο ότι επιτρέπει στα σημεία δεδομένων να ανήκουν σε πολλαπλά συμπλέγματα με ξεχωριστούς βαθμούς συμμετοχής. Η "μαλακή" ή ασαφής ομαδοποίηση K-means είναι ένα παράδειγμα αλληλοεπικαλυπτόμενης ομαδοποίησης.

β) Ιεραρχική ομαδοποίηση

Η ιεραρχική ομαδοποίηση γνωστή και ως (HCA) είναι ένας αλγόριθμος ομαδοποίησης χωρίς επίβλεψη που μπορεί να κατηγοριοποιηθεί σε αθροιστικό ή διαιρετικό.

Η συγκεντρωτική ομαδοποίηση θεωρείται μια bottom-up προσέγγιση. Τα σημεία δεδομένων του απομονώνονται αρχικά ως ξεχωριστές ομαδοποιήσεις και στη συνέχεια συγχωνεύονται μεταξύ τους επαναληπτικά με βάση την ομοιότητα μέχρι να επιτευχθεί ένα σύμπλεγμα.
Τέσσερις διαφορετικές μέθοδοι χρησιμοποιούνται συνήθως για τη μέτρηση της ομοιότητας:

Σύνδεση Ward: Αυτή η μέθοδος δηλώνει ότι η απόσταση μεταξύ δύο συστάδων ορίζεται από την αύξηση του αθροίσματος του τετραγώνου μετά τη συγχώνευση των συστάδων.

Μέση σύνδεση: Αυτή η μέθοδος ορίζεται από τη μέση απόσταση μεταξύ δύο σημείων σε κάθε σύμπλεγμα.

Πλήρης (ή μέγιστη) σύνδεση: Αυτή η μέθοδος ορίζεται από τη μέγιστη απόσταση μεταξύ δύο σημείων σε κάθε σύμπλεγμα.

Ενιαία (ή ελάχιστη) σύνδεση: Αυτή η μέθοδος ορίζεται από την ελάχιστη απόσταση μεταξύ δύο σημείων σε κάθε σύμπλεγμα.

Η Ευκλείδεια απόσταση είναι η πιο κοινή μέτρηση που χρησιμοποιείται για τον υπολογισμό αυτών των αποστάσεων. Ωστόσο, άλλες μετρήσεις όπως η απόσταση του Μανχάταν, αναφέρονται επίσης στη βιβλιογραφία ομαδοποίησης. Η διαιρετική ομαδοποίηση μπορεί να οριστεί ως το αντίθετο της συσσωρευτικής ομαδοποίησης. Αντίθετα, ακολουθεί μια προσέγγιση «από πάνω προς τα κάτω». Σε αυτήν την περίπτωση, ένα μοναδικό σύμπλεγμα δεδομένων χωρίζεται με βάση τις διαφορές μεταξύ των σημείων δεδομένων. Η διαιρετική ομαδοποίηση δεν χρησιμοποιείται συνήθως, αλλά αξίζει να σημειωθεί στο πλαίσιο της ιεραρχικής ομαδοποίησης. Αυτές οι διαδικασίες ομαδοποίησης συνήθως οπτικοποιούνται χρησιμοποιώντας ένα δενδρογράφημα, ένα διάγραμμα που μοιάζει με δέντρο που τεκμηριώνει τη συγχώνευση ή τον διαχωρισμό σημείων δεδομένων σε κάθε επανάληψη.

γ) Πιθανολογική ομαδοποίηση

Ένα πιθανοτικό μοντέλο είναι μια τεχνική χωρίς επίβλεψη που μας βοηθά να λύσουμε προβλήματα εκτίμησης πυκνότητας ή «μαλακής» ομαδοποίησης. Στην πιθανολογική ομαδοποίηση, τα σημεία δεδομένων ομαδοποιούνται με βάση την πιθανότητα ότι ανήκουν σε μια συγκεκριμένη κατανομή. Το Gaussian Mixture Model (**GMM**) είναι μια από τις πιο συχνά χρησιμοποιούμενες πιθανοτικές μεθόδους ομαδοποίησης. Τα μοντέλα Gaussian Mixture ταξινομούνται ως μοντέλα μείγματος, πράγμα που σημαίνει ότι αποτελούνται από έναν απροσδιόριστο αριθμό συναρτήσεων κατανομής πιθανότητας. Τα GMM χρησιμοποιούνται κυρίως για να προσδιοριστεί σε ποια κατανομή πιθανότητας ανήκει ένα δεδομένο σημείο δεδομένων (Γκαουσιανή ή κανονική). Εάν ο μέσος όρος ή η διακύμανση είναι γνωστοί, τότε μπορούμε να προσδιορίσουμε σε ποια κατανομή ανήκει ένα σημείο δεδομένων. Ωστόσο, στα GMM, αυτές οι μεταβλητές δεν είναι γνωστές, επομένως υποθέτουμε ότι υπάρχει μια λανθάνουσα ή κρυφή μεταβλητή για την κατάλληλη ομαδοποίηση σημείων δεδομένων. Αν και δεν απαιτείται η χρήση του αλγόριθμου Προσδοκίας-Μεγιστοποίησης (EM), χρησιμοποιείται συνήθως για την εκτίμηση των πιθανοτήτων εκχώρησης για ένα δεδομένο σημείο δεδομένων σε ένα συγκεκριμένο σύμπλεγμα δεδομένων.

3.4.2 Μείωση διαστάσεων (dimensionality reduction)

Ενώ περισσότερα δεδομένα αποφέρουν γενικά πιο ακριβή αποτελέσματα, μπορεί επίσης να επηρεάσει την απόδοση των αλγορίθμων μηχανικής εκμάθησης (π.χ. υπερπροσαρμογή) και μπορεί επίσης να δυσκολέψει την οπτικοποίηση των συνόλων δεδομένων.

Η μείωση των διαστάσεων (Dimensionality Reduction) είναι μια σημαντική τεχνική στη μηχανική μάθηση και την ανάλυση δεδομένων που χρησιμοποιείται για τη μείωση του αριθμού των χαρακτηριστικών (διαστάσεων) σε ένα σύνολο δεδομένων, διατηρώντας όσο το δυνατόν περισσότερες από τις σχετικές πληροφορίες. Οι τεχνικές μείωσης των διαστάσεων είναι χρήσιμες για την αντιμετώπιση της "κατάρας της διαστατικότητας", τη βελτίωση της αποδοτικότητας των αλγορίθμων, και την αποκατάσταση της απεικόνισης των δεδομένων.

Η μείωση των διαστάσεων μπορεί να χωριστεί σε δύο κύριες κατηγορίες: γραμμικές και μη γραμμικές τεχνικές. Οι γραμμικές τεχνικές μετασχηματίζουν τα δεδομένα σε έναν χαμηλότερης διάστασης χώρο χρησιμοποιώντας γραμμικούς μετασχηματισμούς. Η πιο γνωστή γραμμική μέθοδος είναι η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis, PCA). Η PCA

βρίσκει τους κύριους άξονες του χώρου των δεδομένων και προβάλει τα δεδομένα σε αυτούς τους άξονες, οι οποίοι είναι οι κατευθύνσεις με τη μέγιστη διακύμανση. Οι κύριες συνιστώσες είναι γραμμικοί συνδυασμοί των αρχικών χαρακτηριστικών που διατηρούν όσο το δυνατόν περισσότερη από την αρχική διακύμανση.

Οι μη γραμμικές τεχνικές, από την άλλη πλευρά, χρησιμοποιούνται όταν η σχέση μεταξύ των χαρακτηριστικών δεν μπορεί να αποδοθεί γραμμικά. Μια δημοφιλής μη γραμμική μέθοδος είναι η Ενσωμάτωση Ενδοπολλαπλασιασμού (Isometric Mapping, Isomap), η οποία χρησιμοποιεί γεωδαιτικές αποστάσεις για να διατηρήσει τις μη γραμμικές δομές των δεδομένων. Άλλες μη γραμμικές μέθοδοι περιλαμβάνουν τον Αλγόριθμο t-Distributed Stochastic Neighbor Embedding (t-SNE) και την Ανάλυση Πολυδιάστατης Κλίμακας (Multidimensional Scaling, MDS).

Η μείωση των διαστάσεων μπορεί να επιτευχθεί και μέσω επιλογής χαρακτηριστικών (feature selection) ή εξαγωγής χαρακτηριστικών (feature extraction). Στην επιλογή χαρακτηριστικών, επιλέγονται υποσύνολα των αρχικών χαρακτηριστικών που θεωρούνται τα πιο σημαντικά για την ανάλυση, ενώ στην εξαγωγή χαρακτηριστικών, τα αρχικά χαρακτηριστικά μετασχηματίζονται σε έναν νέο, χαμηλότερης διάστασης χώρο.

3.5 Μέθοδοι Βελτιστοποίησης στη Μηχανική Μάθηση

Οι μέθοδοι βελτιστοποίησης αποτελούν τον πυρήνα της μηχανικής μάθησης, καθώς είναι το εργαλείο μέσω του οποίου τα μοντέλα μαθαίνουν από τα δεδομένα και βελτιώνουν την απόδοσή τους. Οι μέθοδοι αυτές μας επιτρέπουν να ελαχιστοποιούμε τις συναρτήσεις κόστους, οι οποίες μετρούν την απόκλιση των προβλέψεων του μοντέλου από τις πραγματικές τιμές, οδηγώντας σε πιο ακριβείς και αξιόπιστες προβλέψεις. Οι διαφορετικές μέθοδοι βελτιστοποίησης, όπως η Στοχαστική Κατάβαση Κλίσης, η Κατάβαση Κλίσης με Μικρά Δέσμη, η RMSProp και η Adam, παρέχουν ποικίλες προσεγγίσεις για την προσαρμογή των παραμέτρων του μοντέλου, λαμβάνοντας υπόψη διαφορετικές πτυχές των δεδομένων και της συνάρτησης κόστους. Καθεμία από αυτές τις μεθόδους προσφέρει συγκεκριμένα πλεονεκτήματα που μπορούν να βελτιώσουν την ταχύτητα σύγκλισης, τη σταθερότητα και την αποδοτικότητα της εκπαίδευσης των μοντέλων. Η κατανόηση και η σωστή εφαρμογή των μεθόδων βελτιστοποίησης είναι καθοριστικής σημασίας για την ανάπτυξη ισχυρών και αποτελεσματικών αλγορίθμων μηχανικής μάθησης.

3.5.1 Κατάβαση Κλίσης (Gradient Descent - GD)

Η μέθοδος Gradient Descent αποτελεί ένα από τα θεμέλια των αλγορίθμων βελτιστοποίησης στη μηχανική μάθηση, σχεδιασμένη για την εύρεση του ελάχιστου της συνάρτησης κόστους J(θ), όπου θ είναι το διάνυσμα των παραμέτρων του μοντέλου. Αρχικά, ορίζεται η συνάρτηση κόστους ως:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 \quad (3.5)$$

όπου $h\theta\left(x^{(i)}\right)$είναι η πρόβλεψη του μοντέλου για την είσοδο $x^{(i)}$, $y^{(i)}$ είναι η πραγματική τιμή και m ο αριθμός των παραδειγμάτων εκπαίδευσης. Η μέθοδος Gradient Descent προσπαθεί να ελαχιστοποιήσει τη συνάρτηση κόστους προσαρμόζοντας επαναληπτικά τις παραμέτρους θ.

Τρόπος Λειτουργίας :

i) Γίνεται υπολογισμός της κλίσης της συνάρτησης κόστους ως προς το διάνυσμα παραμέτρων θ. Η κλίση ∇θJ(θ) υπολογίζεται ως:

$$\nabla_\theta J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \{h_\theta(x^{(i)}) - y^{(i)}) x^{(i)} \quad (3.6)$$

όπου x^(i) είναι το διάνυσμα χαρακτηριστικών του i-οστού παραδείγματος εκπαίδευσης.

ii) Οι παράμετροι θ ενημερώνονται κατά τη διάρκεια της εκπαίδευσης σύμφωνα με την εξίσωση ενημέρωσης:

$$\theta := \theta - \alpha \nabla_\theta J(\theta) \quad (3.7)$$

όπου α είναι ο ρυθμός μάθησης (learning rate), ο οποίος καθορίζει το μέγεθος του βήματος προς την κατεύθυνση της αντίθετης κλίσης. Ένας καλά επιλεγμένος ρυθμός μάθησης είναι κρίσιμος για την ταχύτητα σύγκλισης του αλγορίθμου.

iii) Η μέθοδος συνεχίζει την ενημέρωση των παραμέτρων μέχρι να συγκλίνει, δηλαδή η τιμή της συνάρτησης κόστους να μην μπορεί να μειωθεί περαιτέρω ή να μην αλλάζει σημαντικά.

3.5.2 Στοχαστική Κατάβαση Κλίσης (Stochastic Gradient Descent – SGD)

Η μέθοδος Stochastic Gradient Descent (SGD) αποτελεί μια παραλλαγή της μεθόδου Gradient Descent που εφαρμόζεται ευρέως στη μηχανική μάθηση για την εκπαίδευση μοντέλων, ιδίως όταν οι συλλογές δεδομένων είναι μεγάλες και όταν το κόστος υπολογισμού της κλίσης για όλα τα δεδομένα είναι υψηλό.

Τρόπος Λειτουργίας :

i) Όπως και στην κλασική μέθοδο Gradient Descent, στην SGD επιδιώκουμε να ελαχιστοποιήσουμε τη συνάρτηση κόστους J(θ). Η συνάρτηση κόστους ορίζεται ως:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 \tag{3.8}$$

όπου $h\theta\left(x^{(i)}\right)$ είναι η πρόβλεψη του μοντέλου για την είσοδο $x^{(i)}$, $y^{(i)}$είναι η πραγματική τιμή και m ο αριθμός των παραδειγμάτων εκπαίδευσης.

ii) Ο ρυθμός μάθησης α παραμένει κρίσιμος και στη μέθοδο SGD. Αυτός καθορίζει το μέγεθος του βήματος που κάνουμε κατά την ενημέρωση των παραμέτρων του μοντέλου.

iii) Η βασική διαφορά της SGD είναι ότι δεν υπολογίζει την κλίση ως το σύνολο των δεδομένων εκπαίδευσης, αλλά μόνο για ένα τυχαία επιλεγμένο υποσύνολο (ή ακόμα και ένα μόνο παράδειγμα). Αυτό συνεπάγεται ότι η ενημέρωση των παραμέτρων γίνεται πιο συχνά και με μικρότερο κόστος υπολογισμού.

iv) Ο υπολογισμός της κλίσης $\nabla\boldsymbol{\theta J(\theta)}$ γίνεται τώρα για το τυχαία επιλεγμένο υποσύνολο δεδομένων:

$$\nabla_\theta J(\theta) = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \left(h_\theta(x^{(i)} - y^{(i)} \right) x^{(i)} \tag{3.9}$$

όπου B είναι το τυχαία επιλεγμένο υποσύνολο δεδομένων.

v) Οι παράμετροι θ ενημερώνονται κατά τη διάρκεια της εκπαίδευσης σύμφωνα με την εξίσωση:

$$\theta := \theta - \alpha \nabla_\theta J(\theta) \tag{3.10}$$

όπου α είναι ο ρυθμός μάθησης και $\nabla\boldsymbol{\theta J(\theta)}$ η κλίση που υπολογίζεται με βάση το τρέχον υποσύνολο δεδομένων B

iv) Η μέθοδος SGD επαναλαμβάνει αυτήν τη διαδικασία για έναν ορισμένο αριθμό εποχών ή μέχρι η συνάρτηση κόστους να συγκλίνει σε μια αποδεκτή τιμή.

3.5.3 Κατάβαση Κλίσης με Μικρά Δέσμη (Mini-Batch Gradient Descent)

Η μέθοδος Mini-Batch Gradient Descent αποτελεί μια ενδιάμεση προσέγγιση μεταξύ της κλασικής Gradient Descent και της Stochastic Gradient Descent, συνήθως χρησιμοποιούμενη στη μηχανική μάθηση για την εκπαίδευση μοντέλων σε μεγάλα σύνολα δεδομένων.

Τρόπος Λειτουργίας :

i) Η συνάρτηση κόστους παραμένει η ίδια με την κλασική μέθοδο Gradient Descent:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 \tag{3.11}$$

ii) Ο ρυθμός μάθησης α είναι η παράμετρος που καθορίζει το μέγεθος των βημάτων κατά την ενημέρωση των παραμέτρων του μοντέλου.

iii) Στη μέθοδο Mini-Batch Gradient Descent, υπολογίζουμε την κλίση όχι για όλα τα δεδομένα εκπαίδευσης (όπως στην κλασική Gradient Descent), ούτε για ένα μόνο δείγμα (όπως στην SGD), αλλά για ένα μικρό υποσύνολο δεδομένων, το οποίο ονομάζεται mini-batch. Το mini-batch περιέχει τυχαία επιλεγμένα παραδείγματα από το σύνολο δεδομένων.

iv) Η κλίση $\nabla \boldsymbol{\theta} \boldsymbol{J(\theta)}$ υπολογίζεται για το mini-batch ως:

$$\nabla_\theta J(\theta) = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \left(h_\theta(x^{(i)} - y^{(i)} \right) x^{(i)} \tag{3.12}$$

όπου B είναι το mini-batch τυχαία επιλεγμένων παραδειγμάτων, και |B|| ο αριθμός των παραδειγμάτων στο mini-batch.

v) Οι παράμετροι θ ενημερώνονται κατά τη διάρκεια της εκπαίδευσης σύφμωνα με το (3.7)

vi) Η διαδικασία επαναλαμβάνεται για πολλά mini-batches μέχρι η συνάρτηση κόστους να συγκλίνει σε μια αποδεκτή τιμή ή μέχρι να ολοκληρωθεί ο αριθμός των εποχών εκπαίδευσης.

3.5.4 Κινητικότητα (Momentum)

Η τεχνική της Κινητικότητας (Momentum) αποτελεί βελτίωση της μεθόδου Gradient Descent που στοχεύει στην επιτάχυνση της σύγκλισης και την αποφυγή των τοπικών ελαχίστων.

Η ιδέα βασίζεται στην εισαγωγή ενός όρου κινητικότητας που λειτουργεί ως "αδράνεια" του βηματισμού κατά την ενημέρωση των παραμέτρων.

Τρόπος Λειτουργίας :

i) Η Κινητικότητα προστίθεται στην ενημέρωση των παραμέτρων για να βελτιώσει τη σταθερότητα και την ταχύτητα σύγκλισης. Η ενημέρωση των παραμέτρων γίνεται ως εξής:

$$u := \beta u - \alpha \nabla_{\theta} J(\theta)$$
$$\theta := \theta + u \tag{3.13}$$

Εδώ, β είναι ο όρος κινητικότητας, που καθορίζει το ποσοστό του προηγούμενου βήματος που θα προστεθεί στο τρέχον βήμα ενημέρωσης. Αυτός ο όρος βοηθά στο να δίνεται προτεραιότητα στις κατευθύνσεις που έχουν συνεχίσει να παρουσιάζουν μεγάλη κλίση. Όλα τα υπόλοιπα βήματα είναι τα ίδια με τις προηγούμενες τεχνικές.

3.5.5 AdaGrad

Η μέθοδος Adagrad (Adaptive Gradient Algorithm) είναι μια παραλλαγή του Gradient Descent που προσαρμόζει δυναμικά το ρυθμό μάθησης για κάθε παράμετρο του μοντέλου, επιτρέποντας την αποδοτική εκπαίδευση σε σενάρια όπου τα δεδομένα έχουν σπάνιες ή μη συχνές ιδιότητες.

Τρόπος Λειτουργίας :

Η συνάρτηση κόστους υπολογίζεται όπως ακριβώς γίνεται στην GD ενώ το ίδιο ισχύει και για την κλίση της. Η διαφόρα του όμως έγκειται στό ρυθμό μάθησης αλλά και στην συσσώρευση των τετραγώνων των κλίσεων. Ο αρχικός ρυθμός μάθησης α προσαρμόζεται δυναμικά κατά τη διάρκεια της εκπαίδευσης, λαμβάνοντας υπόψη την ιστορική κλίση κάθε παραμέτρου. Ταυτόχρονα κάθε παράμετρος **θj** έχει έναν αθροιστή **Gj** που συσσωρεύει τα τετράγωνα των κλίσεων της αντίστοιχης παραμέτρου:

$$G_j = \sum_{t=1}^{T} (\nabla \theta j J(\theta))^2 \tag{3.14}$$

Κάθε παράμετρος ενημερώνεται ως εξής :

$$\theta_j := \theta_j - \frac{\alpha}{\sqrt{G_j} + e} \nabla_\theta J(\theta) \tag{3.15}$$

όπου e είναι μια πολύ μικρή σταθερά που προστίθεται για να αποφευχθεί η διαίρεση με το μηδέν.

Η Adagrad είναι ιδιαίτερα αποτελεσματική σε περιπτώσεις όπου τα δεδομένα είναι σπάνια ή όπου διαφορετικές παράμετροι απαιτούν διαφορετικούς ρυθμούς μάθησης. Η προσαρμογή του ρυθμού μάθησης ανά παράμετρο επιτρέπει στην Adagrad να δίνει μεγαλύτερα βήματα στις παραμέτρους με σπάνιες ή χαμηλές κλίσεις, και μικρότερα βήματα στις παραμέτρους με συχνές ή υψηλές κλίσεις, καθιστώντας τη μέθοδο πιο σταθερή και αποδοτική σε τέτοιες περιπτώσεις.

3.5.6 RMSProp

Η RMSprop είναι ιδιαίτερα αποτελεσματική σε σενάρια όπου οι κλίσεις μπορούν να ποικίλλουν σημαντικά σε διαφορετικές διαστάσεις ή κατά τη διάρκεια διαφορετικών σταδίων εκπαίδευσης. Με την προσαρμογή των ρυθμών μάθησης ανά παράμετρο με βάση τις ιστορικές τους κλίσεις, η RMSprop βοηθά στην επιτάχυνση της σύγκλισης και βελτιώνει τη σταθερότητα της εκπαίδευσης νευρωνικών δικτύων.

Τρόπος Λειτουργίας :

Η συνάρτηση κόστους **J(θ)** παραμένει η ίδια όπως στις κλασικές μεθόδου. Η RMSprop προσαρμόζει δυναμικά τον ρυθμό μάθησης για κάθε παράμετρο βασιζόμενη στον μέσο όρο των πρόσφατων τετραγώνων των κλίσεων της. Διατηρεί ένα κινούμενο μέσο όρο των τετραγώνων των κλίσεων για κάθε παράμετρο. Για κάθε παράμετρο **θj** η RMSprop διατηρεί ένα εκθετικά υπολογισμένο μέσο όρο των προηγούμενων τετραγώνων των κλίσεων, **E[g2]**:

$$E\left[g^2\right]_t = \beta E\left[g^2\right]_{t-1} + (1-\beta)\left(\nabla \theta_j J(\theta)\right)^2 \tag{3.16}$$

όπου β είναι ένας ρυθμός υποβάθμισης (συνήθως κοντά στο 1). Οι παράμετροι ενημερώνονται χρησιμοποιώντας τον προσαρμοσμένο ρυθμό μάθησης:

$$\theta_j := \theta_j - \frac{a}{\sqrt{E\left[g^2\right] + e}} \nabla \theta_j J(\theta) \tag{3.17}$$

όπου e είναι μια μικρή σταθερά που προστίθεται για να αποφευχθεί η διαίρεση με το μηδέν.

3.5.7 Adam (Adaptive Moment Estimation)

Το Adam είναι ένας αλγόριθμος βελτιστοποίησης σχεδιασμένος να συνδυάζει τα πλεονεκτήματα δύο άλλων επεκτάσεων της SGD: του AdaGrad και του RMSProp. Ο αλγόριθμος Adam υπολογίζει προσαρμοστικά ποσοστά μάθησης για κάθε παράμετρο χρησιμοποιώντας εκτιμήσεις των πρώτων και δεύτερων στιγμών των παραγώγων. Ο αλγόριθμος διατηρεί έναν εκθετικά φθίνοντα μέσο όρο των παρελθοντικών παραγώγων (την πρώτη στιγμή), που σημειώνεται ως **mt**, και έναν εκθετικά φθίνοντα μέσο όρο των παρελθοντικών τετραγωνικών παραγώγων (τη δεύτερη στιγμή), που σημειώνεται ως **vt**. Για να αντιμετωπιστούν οι προκαταλήψεις που εισάγονται από αυτούς τους μέσους όρους, ειδικά κατά τις αρχικές επαναλήψεις, το Adam περιλαμβάνει όρους διόρθωσης προκαταλήψεων.

Τρόπος Λειτουργίας :

Ο αλγόριθμος ξεκινά με την αρχικοποίηση του χρόνου **t=0**, του διανύσματος πρώτης στιγμής **m0=0**, του διανύσματος δεύτερης στιγμής **v0=0** και των παραμέτρων **θ0** προς βελτιστοποίηση. Επίσης, ορίζει διάφορες υπερπαραμέτρους: το ποσοστό μάθησης **α** (τυπικά **0.001**), τους ρυθμούς αποσύνθεσης για τις πρώτες και δεύτερες στιγμές **β1** συνήθως (**0.9**) και **β2** συνήθως (**0.999**), και μια μικρή σταθερά **e** (συνήθως 10^{-8}) για την αποφυγή διαίρεσης με το μηδέν. Για κάθε επανάληψη **t**, ο χρόνος αυξάνεται, και οι παράγωγοι της συνάρτησης απώλειας ως προς τις παραμέτρους υπολογίζονται και σημειώνονται ως **gt**. Η μεροληπτική πρώτη στιγμή ενημερώνεται ως:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{3.18}$$

Έπειτα η μεροληπτική δεύτερη στιγμή ενημερώνεται ως :

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{3.19}$$

Στη συνέχεια, η διορθωμένη πρώτη στιγμή υπολογίζεται ως:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{3.20}$$

και η διορθωμένη δεύτερη στιγμή ως:

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{3.21}$$

Τέλος, οι παράμετροι ενημερώνονται χρησιμοποιώντας τον κανόνα:

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon} \tag{3.22}$$

Το Adam συνδυάζει αποτελεσματικά τα πλεονεκτήματα της ορμής (πρώτη στιγμή) και του RMSProp (δεύτερη στιγμή) για να παρέχει μια ανθεκτική και αποδοτική τεχνική βελτιστοποίησης. Είναι υπολογιστικά αποδοτικό, έχει χαμηλές απαιτήσεις μνήμης και είναι κατάλληλο για προβλήματα μεγάλης κλίμακας και δεδομένα με σπανιότητα. Η προσαρμοστική δυνατότητα του ποσοστού μάθησης του επιτρέπει να αποδίδει καλά σε μη σταθερούς στόχους, καθιστώντας τον δημοφιλή επιλογή για τη βελτιστοποίηση μοντέλων βαθιάς μάθησης.

Κεφάλαιο 4. Speech Emotion Recognition (SER)

4.1 Εισαγωγή

Η Αναγνώριση Συναισθημάτων από Ομιλία (SER) είναι ένα αναπτυσσόμενο πεδίο της επεξεργασίας ομιλίας και της υπολογιστικής συναισθηματικής ανάλυσης, αφιερωμένο στην ταυτοποίηση και κατηγοριοποίηση των ανθρώπινων συναισθημάτων από την ομιλία. Αυτή η τεχνολογία έχει σημαντικές επιπτώσεις στην αλληλεπίδραση ανθρώπου-υπολογιστή, επιτρέποντας στα μηχανήματα να ανταποκρίνονται με ενσυναίσθηση στις ανθρώπινες συναισθηματικές καταστάσεις. Οι εφαρμογές της SER εκτείνονται σε διάφορους τομείς, όπως η εξυπηρέτηση πελατών, η ψυχική υγεία και η κοινωνική ρομποτική, όπου η κατανόηση των ανθρώπινων συναισθημάτων αποτελεί καίριο παράγοντα.

Η διαδικασία κατασκευής ενός συστήματος SER περιλαμβάνει διάφορα στάδια ξεκινώντας με τη συλλογή δεδομένων και την προεπεξεργασία τους. Το βασικό στοιχείο κάθε συστήματος SER είναι το dataset, ένα ισχυρό και ποικίλο σύνολο δεδομένων που περιλαμβάνει δείγματα ομιλίας με ετικέτες συναισθημάτων όπως χαρά, λύπη, θυμό, φόβο και ουδετερότητα. Αυτά τα σύνολα δεδομένων είναι απαραίτητα για την εκπαίδευση των μοντέλων μηχανικής μάθησης ώστε να αναγνωρίζουν διαφορετικές συναισθηματικές καταστάσεις. Η συλλογή δεδομένων συνήθως περιλαμβάνει την ηχογράφηση ανθρώπινης ομιλίας σε ελεγχόμενα περιβάλλοντα για να εξασφαλιστεί ο υψηλής ποιότητας ήχος, αλλά μπορεί επίσης να περιλαμβάνει πιο αυθόρμητη ροή λόγου από πραγματικά σενάρια για να βελτιώσει την ανθεκτικότητα του συστήματος.

Κατά την προεπεξεργασία των συλλεγμένων ηχητικών δεδομένων γίνεται ενίσχυση της ποιότητας και της συνέπειας της εισόδου στο σύστημα αναγνώρισης. Αυτό το στάδιο περιλαμβάνει διάφορες εργασίες, όπως μείωση θορύβου, κανονικοποίηση και τμηματοποίηση. Οι τεχνικές μείωσης θορύβου εφαρμόζονται για την ελαχιστοποίηση των θορύβων του περιβάλλοντος και άλλων άσχετων ήχων που θα μπορούσαν να επηρεάσουν την αναγνώριση συναισθημάτων. Η κανονικοποίηση εξασφαλίζει ότι τα ηχητικά δείγματα έχουν συνεπή μορφή και ένταση, διευκολύνοντας την ακριβέστερη ανάλυση. Η τμηματοποίηση περιλαμβάνει τη διαίρεση του ήχου σε μικρότερα, διαχειρίσιμα τμήματα, τα οποία στη συνέχεια αναλύονται ξεχωριστά για να ανιχνευθούν συναισθηματικές ενδείξεις.

΄Επειτα γίνεται η εξαγωγή χαρακτηριστικών από τα δεδομένα. Αυτή η διαδικασία περιλαμβάνει την απόσπαση σημαντικών αναπαραστάσεων του ηχητικού σήματος που μπορούν να χρησιμοποιηθούν για τη διάκριση μεταξύ διαφορετικών συναισθηματικών καταστάσεων. Τα ακουστικά χαρακτηριστικά αποτελούν το κύριο επίκεντρο στη SER, περιλαμβάνοντας προσωδικά, φασματικά και ποιοτικά χαρακτηριστικά φωνής. Τα προσωδικά χαρακτηριστικά, όπως το ύψος, η ενέργεια και η διάρκεια, συλλαμβάνουν την τονικότητα και τον ρυθμό της ομιλίας, τα οποία συχνά επηρεάζονται από τη συναισθηματική κατάσταση του ομιλητή. Τα φασματικά χαρακτηριστικά, όπως τα MFCCs, παρέχουν λεπτομερή αναπαράσταση του φάσματος ισχύος του ηχητικού σήματος σε μια μη γραμμική κλίμακα Mel,

η οποία ευθυγραμμίζεται περισσότερο με την ανθρώπινη ακουστική αντίληψη. Τα χαρακτηριστικά ποιότητας φωνής, όπως το jitter και το shimmer, περιγράφουν τις μεταβολές στη συχνότητα και την ένταση, προσφέροντας πληροφορίες για την εκφραστικότητα της φωνής.

Εκτός από τα ακουστικά χαρακτηριστικά, τα γλωσσικά χαρακτηριστικά που προέρχονται από το περιεχόμενο του κειμένου της ομιλίας μπορούν επίσης να είναι πληροφοριακά. Αυτά τα χαρακτηριστικά εξάγονται χρησιμοποιώντας τεχνικές επεξεργασίας φυσικής γλώσσας (NLP), οι οποίες αναλύουν την επιλογή των λέξεων, τη δομή των προτάσεων και το σημασιολογικό περιεχόμενο για να ανιχνεύσουν συναισθηματικές ενδείξεις. Ο συνδυασμός ακουστικών και γλωσσικών χαρακτηριστικών μπορεί να βελτιώσει σημαντικά την ακρίβεια των συστημάτων αναγνώρισης συναισθημάτων.

Επιπρόσθετα η επιλογή χαρακτηριστικών είναι ένα απαραίτητο βήμα μετά την εξαγωγή χαρακτηριστικών, καθώς δεν είναι όλα τα εξαγόμενα χαρακτηριστικά εξίσου πληροφοριακά για την αναγνώριση συναισθημάτων. Τα περιττά ή άσχετα χαρακτηριστικά μπορούν να εισάγουν θόρυβο και να μειώσουν την απόδοση του συστήματος. Διάφορες στατιστικές και τεχνικές μηχανικής μάθησης, όπως η ανάλυση κύριων συνιστωσών (PCA) και η αναδρομική εξάλειψη χαρακτηριστικών (RFE), χρησιμοποιούνται για να εντοπίσουν και να διατηρήσουν τα πιο σχετικά χαρακτηριστικά, βελτιώνοντας έτσι την αποτελεσματικότητα και την ακρίβεια του μοντέλου.

Το κεντρικό κομμάτι όμως ενός συστήματος SER βρίσκεται στο στάδιο της ταξινόμησης, όπου οι αλγόριθμοι μηχανικής μάθησης εκπαιδεύονται να αντιστοιχούν τα εξαγόμενα χαρακτηριστικά σε συγκεκριμένες ετικέτες συναισθημάτων. Οι παραδοσιακές προσεγγίσεις μηχανικής μάθησης, όπως οι μηχανές διανυσματικής υποστήριξης (SVMs), τα δέντρα αποφάσεων και οι πλησιέστεροι γείτονες (k-NN), έχουν χρησιμοποιηθεί ευρέως στη SER. Ωστόσο, με την έλευση της βαθιάς μάθησης, πιο προηγμένα μοντέλα, όπως τα συνελικτικά νευρωνικά δίκτυα (CNNs) και τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs), έχουν δείξει ανώτερη απόδοση. Τα CNNs είναι ιδιαίτερα αποτελεσματικά στη σύλληψη χωρικών ιεραρχιών στα φασματικά χαρακτηριστικά, ενώ τα RNNs, ιδιαίτερα τα δίκτυα μακράς βραχείας μνήμης (LSTM), είναι κατάλληλα για τη μοντελοποίηση χρονικών εξαρτήσεων στο ηχητικό σήμα.

4.2 Συλλογή και Ανάθεση Δεδομένων

Η βάση κάθε συστήματος αναγνώρισης συναισθημάτων από ομιλία (SER) έγκειται στην ποιότητα και την ποικιλία των δεδομένων που χρησιμοποιούνται για την εκπαίδευση των μοντέλων. Η συλλογή και η ανάθεση τέτοιων δεδομένων.

Ένα σώμα δεδομένων είναι μια συλλογή καταγεγραμμένων δεδομένων ομιλίας που χρησιμοποιούνται για την εκπαίδευση και την αξιολόγηση συστημάτων SER. Αρκετά γνωστά σώματα δεδομένων έχουν αναπτυχθεί ειδικά για την αναγνώριση συναισθημάτων. Για παράδειγμα, η Διαδραστική Βάση Δεδομένων Συναισθηματικής Δυαδικής Κίνησης (IEMOCAP) περιλαμβάνει περίπου 12 ώρες δεδομένων ήχου και βίντεο με αυθόρμητες και σεναριακές αλληλεπιδράσεις μεταξύ ηθοποιών, τα οποία έχουν ανατεθεί σε διάφορα συναισθήματα όπως η χαρά, ο θυμός, η θλίψη και η ουδετερότητα. Ένα άλλο παράδειγμα

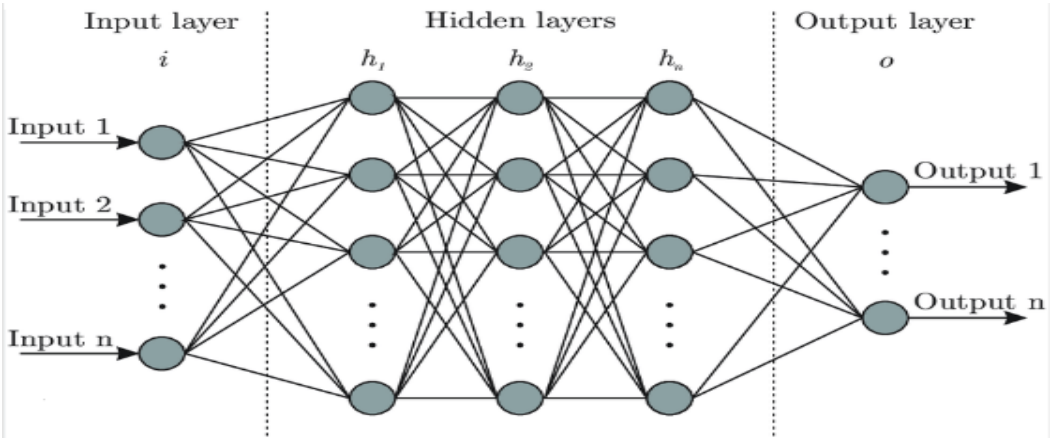
είναι η Βάση Δεδομένων Ήχου και Βίντεο Συναισθηματικής Ομιλίας και Τραγουδιού του Ryerson (RAVDESS), η οποία περιλαμβάνει ηχογραφήσεις 24 επαγγελματιών ηθοποιών που εκφωνούν δύο λεξικά ισοδύναμες δηλώσεις με ουδέτερη Βορειοαμερικανική προφορά, με κάθε έκφραση διαθέσιμη σε δύο εντάσεις και πρόσθετες ηχογραφήσεις συναισθηματικού τραγουδιού. Η Βάση Δεδομένων Συναισθημάτων του Βερολίνου (Emo-DB) αποτελείται από δέκα επαγγελματίες Γερμανούς ηθοποιούς που εκτέλεσαν 10 προτάσεις σε επτά διαφορετικές συναισθηματικές καταστάσεις, συμπεριλαμβανομένων του θυμού, της πλήξης, της αηδίας, της ανησυχίας/φόβου, της χαράς, της θλίψης και της ουδετερότητας. Αυτές οι βάσεις δεδομένων αποτελούν πολύτιμους πόρους για την ανάπτυξη και τη σύγκριση συστημάτων SER, παρέχοντας ένα τυποποιημένο σημείο αναφοράς.

Η ανάθεση συναισθηματικών δεδομένων περιλαμβάνει την επισήμανση των ηχογραφήσεων ομιλίας με τις κατάλληλες συναισθηματικές κατηγορίες, μια διαδικασία που είναι κρίσιμη αλλά και δύσκολη λόγω της υποκειμενικής φύσης των συναισθημάτων. Συνήθως, ανθρώπινοι ακροατές, ακούνε τις ηχογραφήσεις και επισημαίνουν κάθε τμήμα με το αντιληπτό συναίσθημα. Για τη βελτίωση της αξιοπιστίας, πολλοί ακροατές συχνά επισημαίνουν τα ίδια δεδομένα και η συμφωνία μεταξύ τους μετράται χρησιμοποιώντας μετρήσεις όπως το Κ του Cohen. Πρόσφατα, έχουν αναπτυχθεί αυτοματοποιημένα εργαλεία και αλγόριθμοι για να βοηθήσουν στην ανάθεση, παρέχοντας προκαταρκτικές ετικέτες που στη συνέχεια επαληθεύουν και βελτιώνουν οι άνθρωποι. Η διαδικασία ανάθεσης πρέπει να λαμβάνει υπόψη διάφορους παράγοντες, όπως το πλαίσιο στο οποίο εμφανίζεται ένα τμήμα ομιλίας, καθώς τα συναισθήματα συχνά εξαρτώνται από το πλαίσιο, την ένταση των συναισθημάτων και τις πολιτισμικές διαφορές που μπορούν να επηρεάσουν την αντίληψη των συναισθημάτων.

Αρκετές προκλήσεις σχετίζονται με τη συλλογή και την ανάθεση δεδομένων στον τομέα του SER. Τα συναισθήματα είναι εγγενώς υποκειμενικά και διαφορετικοί αναθέτες μπορεί να αντιλαμβάνονται το ίδιο τμήμα ομιλίας διαφορετικά, οδηγώντας σε ασυνέπειες στη βάση δεδομένων. Η πολιτισμική μεταβλητότητα μπορεί να περιπλέξει τη δημιουργία συστημάτων SER που εφαρμόζονται παγκοσμίως, καθώς τα συναισθήματα μπορεί να εκφράζονται και να γίνονται αντιληπτά διαφορετικά σε διάφορους πολιτισμούς. Επιπλέον, οι ηχογραφήσεις στον πραγματικό κόσμο συχνά περιέχουν θόρυβο στο παρασκήνιο, ο οποίος μπορεί να παρεμποδίσει την αναγνώριση συναισθημάτων. Η συλλογή ηχογραφήσεων υψηλής ποιότητας σε ελεγχόμενα περιβάλλοντα μπορεί να μετριάσει αυτό το ζήτημα, αλλά μπορεί να περιορίσει την ποικιλία της βάσης δεδομένων. Επιπλέον, οι συναισθηματικές εκφράσεις δεν κατανέμονται ομοιόμορφα στην πραγματική ζωή, οδηγώντας σε ανισορροπία βάσεων δεδομένων όπου κάποια συναισθήματα υποεκπροσωπούνται, γεγονός που μπορεί να προκαλέσει προκατάληψη του μοντέλου προς τα συχνότερα εμφανιζόμενα συναισθήματα. Η αντιμετώπιση αυτών των προκλήσεων απαιτεί προσεκτικό σχεδιασμό και εφαρμογή βέλτιστων πρακτικών στη συλλογή και την ανάθεση δεδομένων, εξασφαλίζοντας την ανάπτυξη αξιόπιστων και γενικεύσιμων συστημάτων SER.

4.3 Τεχνικές Αναγνώρισης Συναισθημάτων

Οι αλγόριθμοι αναγνώρισης συναισθημάτων στη φωνή (SER) χρησιμοποιούνται ευρέως σε πρακτικές εφαρμογές για την ερμηνεία των συναισθημάτων που εκφράζονται μέσω της προφορικής γλώσσας. Παραδοσιακές μέθοδοι όπως τα GMMs και τα SVMs έχουν εφαρμοστεί με επιτυχία σε εργασίες ανάλυσης συναισθημάτων, όπως η ανάλυση αισθημάτων σε αλληλεπιδράσεις εξυπηρέτησης πελατών. Για παράδειγμα, τα GMMs έχουν επιτύχει ακρίβειες περίπου 66% στην αναγνώριση συναισθημάτων όπως χαρά, λύπη και θυμό σε ηχογραφήσεις κέντρων εξυπηρέτησης (Eyben et al., 2010). Τα SVMs, γνωστά για την ικανότητά τους να κατηγοριοποιούν ακριβώς, έχουν επιδείξει ακρίβειες που κυμαίνονται από 70% έως 80% σε σύνολα δεδομένων όπως το Interactive Emotional Dyadic Motion Capture (IEMOCAP), επιτρέποντας την ακριβή ανίχνευση συναισθημάτων όπως η ουδέτερη αντίδραση, η χαρά, η λύπη και ο θυμός.



Σχήμα 4.1 Τυπική αρχιτεκτονική ενός DNN

Τα μοντέλα βαθιάς μάθησης, όπως τα Συνελικτικά Νευρωνικά Δίκτυα (CNNs) και τα Αναδρομικά Νευρωνικά Δίκτυα (RNNs), έχουν σημαντικά επεκτείνει τις δυνατότητες του SER σε πραγματικές εφαρμογές. Οι CNNs εξορύσσουν χωρικά χαρακτηριστικά από σπεκτρογράμματα ή MFCCs, επιτυγχάνοντας ακρίβειες που κυμαίνονται από 75% έως 85%. Για παράδειγμα, συστήματα SER βασισμένα σε CNN έχουν ενσωματωθεί σε συσκευές ελέγχου φωνής, επιτυγχάνοντας υψηλές ακρίβειες στην ανίχνευση συναισθημάτων όπως χαρά και θυμός σε εντολές χρηστών, βελτιώνοντας έτσι την ικανοποίηση και την αλληλεπίδραση του χρήστη.

Αντίστοιχα, τα RNNs είναι σχεδιασμένα για να αντιλαμβάνονται τις χρονικές εξαρτήσεις σε ακολουθιακά δεδομένα, ξεπερνούν τις παραδοσιακές μεθόδους με ακρίβειες που υπερβαίνουν το 80% σε σύνολα δεδομένων όπως το IEMOCAP, όπου αναγνωρίζουν αποτελεσματικά μια ευρεία γκάμα συναισθημάτων, συμπεριλαμβανομένων της έκπληξης, του φόβου και της αηδίας. Οι υβριδικές προσεγγίσεις όπως οι Συνελικτικά Αναδρομικά Νευρωνικά Δίκτυα (CRNNs) συνδυάζουν τα πλεονεκτήματα των CNNs για την εξαγωγή χωρικών χαρακτηριστικών και των RNNs για την σειριακή μοντελοποίηση, επιτυγχάνοντας βελτιωμένη ακρίβεια και προσαρμοστικότητα σε πολύπλοκες εργασίες αναγνώρισης συναισθημάτων.

Αυτά τα μοντέλα ενσωματώνονται όλο και περισσότερο σε εφαρμογές όπως τα συστήματα διαλόγου που αναγνωρίζουν τα συναισθηματικά κείμενα και τις διηγηματικές πλατφόρμες, εμπλουτίζοντας τις εμπειρίες των χρηστών με την εξατομίκευση των αλληλεπιδράσεων βασισμένη στην ανίχνευση των αναγνωρισμένων συναισθημάτων.
Οι τεχνικές αυτές ενσωματώνουν πολλαπλούς ταξινομητές ή μοντέλα, εφαρμόζονται σε πολύγλωσσα συστήματα SER για διεθνείς συνδιασκέψεις ή παγκόσμιες επιχειρηματικές λειτουργίες εξυπηρέτησης πελατών. Ένα DNN συνήθως αποτελείται από ένα στρώμα εισόδου, πολλαπλά κρυφά στρώματα και ένα στρώμα εξόδου. Το στρώμα εισόδου λαμβάνει ακατέργαστα δεδομένα, ενώ κάθε κρυφό στρώμα εφαρμόζει μια σειρά μετασχηματισμών στα εισερχόμενα δεδομένα. Το στρώμα εξόδου παράγει την τελική πρόβλεψη ή ταξινόμηση. Το βάθος ενός DNN αναφέρεται στον αριθμό των κρυφών στρωμάτων με τα βαθύτερα δίκτυα να είναι ικανά να καταγράφουν πιο σύνθετα πρότυπα αλλά επίσης να απαιτούν περισσότερους υπολογιστικούς πόρους και προσεκτική κανονικοποίηση για να αποφεύγεται η υπερπροσαρμογή. Η εκπαίδευση ενός DNN περιλαμβάνει τη βελτιστοποίηση των βαρών και των μεροληψιών των νευρώνων για την ελαχιστοποίηση μιας συνάρτησης απώλειας που μετρά τη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών εξόδων. Αυτή η διαδικασία γίνεται συνήθως χρησιμοποιώντας μια τεχνική που ονομάζεται οπισθοδιάδοση (backpropagation), σε συνδυασμό με έναν αλγόριθμο βελτιστοποίησης όπως η Στοχαστική Κατηφορική Βαθμίδωση (Stochastic Gradient Descent, SGD). Κατά τη διάρκεια της προώθησης προς τα εμπρός, τα δεδομένα εισόδου διέρχονται από το δίκτυο στρώμα προς στρώμα, για να ληφθούν οι προβλέψεις εξόδου. Η συνάρτηση απώλειας (π.χ. Μέσο Τετραγωνικό Σφάλμα για παλινδρόμηση, Απώλεια Διασταυρούμενης Εντροπίας για ταξινόμηση) υπολογίζει το σφάλμα μεταξύ των προβλεπόμενων και των πραγματικών εξόδων. Η οπισθοδιάδοση υπολογίζει την κλίση της συνάρτησης απώλειας σε σχέση με κάθε βάρος στο δίκτυο, εφαρμόζοντας τον κανόνα της αλυσίδας του διαφορικού λογισμού και ο αλγόριθμος βελτιστοποίησης ενημερώνει τα βάρη χρησιμοποιώντας τις υπολογισμένες κλίσεις. Υπάρχουν διάφορες αρχιτεκτονικές DNNs, καθεμία σχεδιασμένη να χειρίζεται συγκεκριμένους τύπους δεδομένων και εργασιών πιο αποτελεσματικά. Τα Εμπρόσθια Νευρωνικά Δίκτυα (Feedforward Neural Networks, FNNs) είναι η απλούστερη μορφή DNN όπου οι συνδέσεις μεταξύ των κόμβων δεν σχηματίζουν κύκλο. Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks, CNNs) είναι σχεδιασμένα για την επεξεργασία δεδομένων δομημένων σε πλέγμα, όπως εικόνες και χρησιμοποιούν συνελικτικά στρώματα για να μαθαίνουν αυτόματα και προσαρμοστικά τις ιεραρχίες των χαρακτηριστικών. Τα Επαναληπτικά Νευρωνικά Δίκτυα (Recurrent Neural Networks, RNNs) είναι κατάλληλα για σειριακά δεδομένα, όπως χρονοσειρές ή φυσική γλώσσα, με συνδέσεις που σχηματίζουν κύκλους, επιτρέποντας την διατήρηση πληροφοριών με την πάροδο του χρόνου. Παραλλαγές όπως το Long Short-Term Memory (LSTM) και το Gated Recurrent Unit (GRU) χρησιμοποιούνται για να αντιμετωπίσουν ζητήματα εξαφάνισης και έκρηξης των κλίσεων στα παραδοσιακά RNNs. Τα Γεννητικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks, GANs) αποτελούνται από δύο δίκτυα, έναν γεννήτορα και έναν διακριτή, που ανταγωνίζονται μεταξύ τους και χρησιμοποιούνται για τη δημιουργία ρεαλιστικών δειγμάτων δεδομένων όπως εικόνες, βίντεο και ήχο. Τα DNNs έχουν εφαρμοστεί επιτυχώς σε διάφορους τομείς

συμπεριλαμβανομένης της αναγνώρισης εικόνας, της επεξεργασίας φυσικής γλώσσας (NLP), της αναγνώρισης ομιλίας και των παιχνιδιών. Στην αναγνώριση εικόνας τα CNNs έχουν θέσει πρότυπα σε εργασίες όπως η ταξινόμηση εικόνων, η ανίχνευση αντικειμένων και η αναγνώριση προσώπου. Στην NLP, τα RNNs και οι παραλλαγές τους χρησιμοποιούνται ευρέως σε εργασίες όπως η μοντελοποίηση γλώσσας, η μηχανική μετάφραση και η ανάλυση συναισθήματος, με μοντέλα όπως το Transformer και το BERT να προωθούν την τεχνολογική πρωτοπορία. Στην αναγνώριση ομιλίας τα DNNs χρησιμοποιούνται για τη μετατροπή του προφορικού 36 λόγου σε κείμενο με υψηλή ακρίβεια, χρησιμοποιούμενα σε εικονικούς βοηθούς όπως το Siri και το Google Assistant. Στα παιχνίδια, η Βαθιά Ενισχυμένη Μάθηση (Deep Reinforcement Learning) ένας συνδυασμός DNNs και ενισχυτικής μάθησης έχει χρησιμοποιηθεί για τη δημιουργία πρακτόρων που επιτυγχάνουν υπεράνθρωπη απόδοση σε παιχνίδια όπως το Go (AlphaGo) και το Dota 2. Παρά την ευρεία χρήση τους τα DNNs αντιμετωπίζουν αρκετές προκλήσεις όπως οι απαιτήσεις σε υπολογιστική ισχύ και μνήμη, η ανάγκη για μεγάλα σύνολα δεδομένων με ετικέτες για αποτελεσματική εκπαίδευση, οι δυσκολίες στην ερμηνευσιμότητα και η πιθανή υπερπροσαρμογή. Η μελλοντική έρευνα στα DNNs επικεντρώνεται στη βελτίωση της ερμηνευσιμότητας, στη μείωση των υπολογιστικών απαιτήσεων και στην ανάπτυξη μεθόδων για την εκπαίδευση μοντέλων με λιγότερα δεδομένα.

4.3.1 Long Short-Term Memory (LSTM)

Στην παρούσα εργασία θα χρησιμοποιήσουμε έναν αλγόριθμο LSTM για την ταξινόμηση των συναισθηματικών καταστάσεων. Οι Long Short-Term Memory, είναι ένας τύπος αναδρομικού νευρωνικού δικτύου (Recurrent Neural Network) που έχει σχεδιαστεί για να μαθαίνει αλληλουχίες δεδομένων. Εισήχθησαν από τους Hochreiter και Schmidhuber το 1997 και έχουν αποδειχθεί εξαιρετικά επιτυχημένα σε διάφορες εφαρμογές όπως αναγνώριση ομιλίας, μετάφραση γλώσσας και πρόβλεψη σειρών.

Η βασική διαφορά μεταξύ των LSTM και των παραδοσιακών RNN έγκειται στη δυνατότητα των LSTM να αποθηκεύουν πληροφορίες για μεγαλύτερες χρονικές περιόδους. Αυτό επιτυγχάνεται μέσω της δομής των κυψελών LSTM που περιλαμβάνει τρεις τύπους "πυλών": την πύλη εισόδου, την πύλη εξόδου και την πύλη λήθης.

1. **Πύλη Εισόδου**: Αποφασίζει ποιες νέες πληροφορίες θα αποθηκευτούν στην κυψέλη. Λαμβάνει την τρέχουσα είσοδο και την προηγούμενη κατάσταση κρυφής μνήμης και εφαρμόζει μια συνάρτηση ενεργοποίησης για να καθορίσει ποιες πληροφορίες είναι σημαντικές και πρέπει να προστεθούν στην κατάσταση της κυψέλης.
2. **Πύλη Λήθης**: Αποφασίζει ποιες πληροφορίες θα διαγραφούν από την κυψέλη. Λαμβάνει την τρέχουσα είσοδο και την προηγούμενη κατάσταση κρυφής μνήμης και εφαρμόζει μια συνάρτηση ενεργοποίησης για να καθορίσει ποιες πληροφορίες δεν είναι πλέον χρήσιμες και μπορούν να αφαιρεθούν.
3. **Πύλη Εξόδου**: Αποφασίζει ποια θα είναι η έξοδος της κυψέλης. Λαμβάνει την τρέχουσα είσοδο και την κατάσταση της κυψέλης και εφαρμόζει μια συνάρτηση ενεργοποίησης για να καθορίσει ποιο μέρος της κατάστασης της κυψέλης θα

χρησιμοποιηθεί για την έξοδο.

Αυτές οι πύλες συνεργάζονται για να διατηρούν και να τροποποιούν την κατάσταση της κυψέλης με τρόπους που επιτρέπουν στο LSTM να μαθαίνει και να θυμάται αλληλουχίες πληροφοριών για μεγάλες χρονικές περιόδους, αποφεύγοντας τα προβλήματα εξαφάνισης και έκρηξης των βαθμίδων που παρατηρούνται στα παραδοσιακά RNN.

Τρόπος Λειτουργίας :

Η βασική μονάδα των LSTM, η κυψέλη LSTM, περιλαμβάνει τρεις τύπους "πυλών": την πύλη εισόδου, την πύλη λήθης και την πύλη εξόδου, οι οποίες συνεργάζονται για να διατηρούν και να τροποποιούν την κατάσταση της κυψέλης. Η πύλη λήθης αποφασίζει ποιες πληροφορίες από την προηγούμενη κατάσταση της κυψέλης πρέπει να ξεχαστούν. Η διαδικασία υπολογισμού της πύλης λήθης είναι η εξής:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{4.1}$$

όπου ft είναι το διάνυσμα εξόδου της πύλης λήθης, σ είναι η συνάρτηση σιγμοειδούς, **Wf** είναι ο πίνακας βαρών της πύλης λήθης, **ht−1** είναι το διάνυσμα της κρυφής κατάστασης από την προηγούμενη χρονική στιγμή, **xt** είναι το διάνυσμα εισόδου και bf είναι το διάνυσμα των μεροληψιών της πύλης λήθης.

Η πύλη εισόδου αποφασίζει ποιες νέες πληροφορίες θα αποθηκευτούν στην κυψέλη, με την εξίσωση :

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{4.2}$$

όπου **it** είναι το διάνυσμα εξόδου της πύλης εισόδου, **Wi** είναι ο πίνακας βαρών της πύλης εισόδου και **bi** είναι το διάνυσμα των μεροληψιών της πύλης εισόδου. Η πύλη των υποψήφιων αξιών της κυψέλης δημιουργεί τις υποψήφιες τιμές που μπορούν να προστεθούν στην κατάσταση της κυψέλης:

$$C_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{4.3}$$

όπου **Ct** είναι το διάνυσμα των υποψήφιων τιμών της κυψέλης, **WC** είναι ο πίνακας βαρών της κυψέλης και **bC** είναι το διάνυσμα των μεροληψιών της κυψέλης. Η νέα κατάσταση της κυψέλης υπολογίζεται ως:

$$C_t^n = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4.4}$$

όπου **Ctn** είναι η νέα κατάσταση της κυψέλης και **Ct−1** είναι η παλιά κατάσταση της κυψέλης. Η πύλη εξόδου αποφασίζει ποιο μέρος της κατάστασης της κυψέλης θα χρησιμοποιηθεί για την έξοδο, με την εξίσωση:

$$o_t = \sigma\left(W_o . \left[h_{t-1}, x_t \right] + b_0 \right) \quad (4.5)$$

όπου **ot** είναι το διάνυσμα εξόδου της πύλης εξόδου, **Wo**, είναι ο πίνακας βαρών της πύλης εξόδου και **bo** είναι το διάνυσμα των μεροληψιών της πύλης εξόδου.
Τέλος, η κρυφή κατάσταση υπολογίζεται ως:

$$h_t = o_t \cdot tanh\left(C_t \right) \quad (4.6)$$

όπου **ht** είναι το διάνυσμα της κρυφής κατάστασης.

Συνοψίζοντας, η συνολική διαδικασία για κάθε χρονική στιγμή περιλαμβάνει τον υπολογισμό της πύλης λήθης, της πύλης εισόδου, των υποψήφιων τιμών της κυψέλης, την ενημέρωση της κατάστασης της κυψέλης, τον υπολογισμό της πύλης εξόδου και της κρυφής κατάστασης.

4.3.2 Καταλληλότητα των LSTM για SER

Η αναγνώριση συναισθημάτων από την ομιλία αποτελεί μια περίπλοκη διαδικασία που απαιτεί την ανάλυση χρονοσειρών δεδομένων με συχνές διακυμάνσεις και πολυπλοκότητες. Τα LSTM είναι ιδιαίτερα κατάλληλα για αυτή την εργασία για διάφορους λόγους:

1. **Αντιμετώπιση της Εξάρτησης Μακράς Διαρκείας**: Τα συναισθήματα στην ομιλία δεν εξαρτώνται μόνο από τις τρέχουσες λέξεις ή φράσεις αλλά και από το γενικότερο πλαίσιο της συνομιλίας. Τα LSTM μπορούν να διατηρούν πληροφορίες από προηγούμενες χρονικές στιγμές και να χρησιμοποιούν αυτές τις πληροφορίες για να κατανοήσουν καλύτερα τα συναισθήματα που εκφράζονται.
2. **Ευελιξία και Ανθεκτικότητα**: Η αρχιτεκτονική των LSTM τα καθιστά ικανά να διαχειρίζονται δυναμικά και αβέβαια μοτίβα δεδομένων που είναι συνηθισμένα στην ανθρώπινη ομιλία. Οι πύλες εισόδου, λήθης και εξόδου επιτρέπουν στο δίκτυο να προσαρμόζεται και να μαθαίνει αποτελεσματικά από περίπλοκα δεδομένα.
3. **Διαχείριση Μεταβλητότητας**: Τα συναισθήματα στην ομιλία μπορεί να μεταβάλλονται γρήγορα και απότομα. Τα LSTM είναι σε θέση να αντιλαμβάνονται και να προσαρμόζονται σε αυτές τις γρήγορες μεταβολές, επιτρέποντας μια πιο ακριβή αναγνώριση των συναισθημάτων.
4. **Εξαγωγή Χαρακτηριστικών**: Τα LSTM μπορούν να εξάγουν πολύπλοκα χαρακτηριστικά από τα δεδομένα της ομιλίας που μπορεί να είναι σημαντικά για την αναγνώριση συναισθημάτων. Αυτά τα χαρακτηριστικά μπορούν να περιλαμβάνουν τον τόνο, την ένταση, τον ρυθμό και άλλες παραμέτρους της ομιλίας που είναι δύσκολο να αναλυθούν με παραδοσιακές μεθόδους.

4.3.3 Χρήση των LSTM για SER

Η χρήση των LSTM στην αναγνώριση συναισθημάτων από την ομιλία έχει επιφέρει σημαντικές βελτιώσεις στις επιδόσεις των συστημάτων αυτών. Παρακάτω αναφέρονται ορισμένα παραδείγματα εφαρμογών:

1. **Ανάλυση Συναισθηματικού Περιεχομένου Κλήσεων**: Σε κέντρα εξυπηρέτησης πελατών, τα LSTM χρησιμοποιούνται για την ανάλυση των συναισθημάτων των πελατών κατά τη διάρκεια τηλεφωνικών κλήσεων. Αυτό επιτρέπει στις εταιρείες να κατανοήσουν καλύτερα τις ανάγκες και τις ανησυχίες των πελατών τους και να βελτιώσουν την ποιότητα της εξυπηρέτησης.
2. **Ανάλυση Συναισθημάτων στα Μέσα Κοινωνικής Δικτύωσης**: Τα LSTM χρησιμοποιούνται για την ανάλυση των συναισθημάτων στα βίντεο και τις ηχητικές αναρτήσεις στα μέσα κοινωνικής δικτύωσης. Αυτό βοηθά τις εταιρείες να κατανοήσουν τις αντιδράσεις και τα συναισθήματα των χρηστών απέναντι στα προϊόντα και τις υπηρεσίες τους.
3. **Εκπαιδευτικά Εργαλεία**: Στην εκπαίδευση, τα LSTM μπορούν να χρησιμοποιηθούν για την ανάλυση των συναισθημάτων των μαθητών κατά τη διάρκεια διαδικτυακών μαθημάτων ή διαλέξεων. Αυτό επιτρέπει στους εκπαιδευτικούς να προσαρμόσουν τις μεθόδους διδασκαλίας τους για να βελτιώσουν την εμπειρία μάθησης των μαθητών.
4. **Υποστήριξη Ψυχικής Υγείας**: Τα LSTM μπορούν να βοηθήσουν στην παρακολούθηση και ανάλυση των συναισθημάτων των ατόμων που αντιμετωπίζουν ψυχικά προβλήματα. Αυτό μπορεί να βοηθήσει τους ψυχολόγους και τους θεραπευτές να παρέχουν πιο αποτελεσματική υποστήριξη και συμβουλευτική.

Κεφάλαιο 5 : Πρακτική Εφαρμογή

5.1 Πρόλογος

Η αναγνώριση συναισθημάτων ομιλίας (SER) έχει ως στόχο την αναγνώριση των ανθρώπινων συναισθημάτων από τα σήματα ομιλίας. Η ακριβής SER μπορεί να βελτιώσει την αλληλεπίδραση ανθρώπου-υπολογιστή κάνοντας τα συστήματα πιο ευαίσθητα και προσαρμοστικά στις συναισθηματικές καταστάσεις των χρηστών. Σε αυτό το κεφάλαιο, παρουσιάζουμε μια λεπτομερή ανάλυση και σύγκριση δύο αλγορίθμων μηχανικής μάθησης, του Multilayer Perceptron (MLP) και των δικτύων Long Short-Term Memory (LSTM), που εφαρμόζονται στην αναγνώριση συναισθημάτων ομιλίας χρησιμοποιώντας τη βάση δεδομένων RAVDESS. Έπειτα θα εξετάσουμε τις διαφορές μεταξύ των 2 αυτών ταξινομητών κατά την εισαγωγή διαφορετικών ηχητικών χαρακτηριστικών, με σκοπό την κατανόηση της καταλληλότητας του κάθε συνδυασμού.

5.2 Μέθοδοι και εργαλεία

Για την πλήρη υλοποίηση της εφαρμογής θα χρησιμοποιήσουμε το ολοκληρωμένο περιβάλλον ανάπτυξης **Spyder**, την γλώσσα προγραμματισμού **Python** αλλά και τις βιβλιοθήκες **Tensorflow**, **Keras** και **Librosa.**

5.2.1 Γλώσσα προγραμματισμού Python

Η Python είναι μια ευέλικτη και υψηλού επιπέδου γλώσσα προγραμματισμού, έχει κερδίσει τεράστια δημοτικότητα στο χώρο της μηχανικής μάθησης για πολλούς επιτακτικούς λόγους.

Η απλότητα και η αναγνωρισιμότητα της την καθιστούν προσβάσιμη τόσο σε αρχάριους όσο και σε έμπειρους προγραμματιστές, διευκολύνοντας την ταχεία δημιουργία πρωτοτύπων και τους αποτελεσματικούς κύκλους ανάπτυξης. Η Python διαθέτει ένα εκτεταμένο οικοσύστημα βιβλιοθηκών και πλαισίων, προσαρμοσμένων ειδικά για εργασίες μηχανικής μάθησης, όπως το TensorFlow, το PyTorch και το scikit-learn τα οποία παρέχουν ισχυρά εργαλεία για χειρισμό δεδομένων, εκπαίδευση μοντέλων και αξιολόγηση.

Επιπλέον, η δυναμική πληκτρολόγηση και η ερμηνευτική φύση της Python συμβάλλουν στην ευελιξία της, επιτρέποντας στους προγραμματιστές να επαναλάβουν γρήγορα και να πειραματιστούν με διαφορετικούς αλγόριθμους και προσεγγίσεις.

Η ισχυρή κοινοτική της υποστήριξη προωθεί τη συνεργασία και την ανταλλαγή γνώσεων, οδηγώντας στη συνεχή εξέλιξη και τελειοποίηση των τεχνικών μηχανικής μάθησης. Επιπλέον, η απρόσκοπτη ενσωμάτωση της Python με άλλες τεχνολογίες και η συμβατότητά της σε όλες τις πλατφόρμες την καθιστούν ιδανική επιλογή για την κατασκευή επεκτάσιμων και έτοιμα για παραγωγή συστημάτων μηχανικής μάθησης.

Συνολικά, η ευκολία χρήσης, το πλούσιο οικοσύστημα και η ευελιξία της Python την

τοποθετούν ως τη κύρια γλώσσα επιλογής για την αντιμετώπιση της πολυπλοκότητας των σύγχρονων εφαρμογών μηχανικής εκμάθησης.

5.2.2 Spyder IDE

Το Spyder είναι ακρωνύμιο για το Scientific PYthon Development EnviRonment και ξεχωρίζει ως ένα ισχυρό ολοκληρωμένο περιβάλλον ανάπτυξης (IDE) ειδικά προσαρμοσμένο για εργασίες επιστημονικού υπολογισμού και ανάλυσης δεδομένων στην Python.

Η διεπαφή χρήστη και το ολοκληρωμένο σύνολο χαρακτηριστικών του το καθιστούν αγαπημένο μεταξύ επιστημόνων, μηχανικών και ερευνητών που αναζητούν ένα παραγωγικό περιβάλλον για τη βέλτιστη χρήση της Python. Το Spyder παρέχει μια σειρά από βασικά εργαλεία για αποτελεσματική κωδικοποίηση, εντοπισμό σφαλμάτων και εξερεύνηση δεδομένων, συμπεριλαμβανομένης μιας διάταξης πολλαπλών παραθύρων για ταυτόχρονη προβολή σεναρίων, κονσολών, εξερευνητών μεταβλητών και γραφικών παραστάσεων.

Η ενσωμάτωσή του με δημοφιλείς επιστημονικές βιβλιοθήκες όπως οι NumPy, SciPy και Matplotlib διευκολύνει την απρόσκοπτη επεξεργασία δεδομένων, ανάλυση και οπτικοποίηση ροών εργασιών. Η διαδραστική κονσόλα της Spyder και η ενσωμάτωση IPython προσφέρουν ένα διαδραστικό περιβάλλον υπολογιστών ιδανικό για γρήγορη δημιουργία πρωτοτύπων και πειραματισμό.

Επιπλέον, ο ισχυρός επεξεργαστής κώδικα της Spyder διαθέτει χαρακτηριστικά όπως επισήμανση σύνταξης, συμπλήρωση κώδικα και αυτόματη εσοχή, βελτιώνοντας την αναγνωσιμότητα και την παραγωγικότητα του κώδικα. Η ενσωματωμένη υποστήριξή του για συστήματα ελέγχου εκδόσεων όπως το Git επιτρέπει τη συνεργατική ανάπτυξη και διαχείριση έργων.

Τέλος, η επεκτασιμότητα του Spyder μέσω προσθηκών και προσαρμόσιμων διαμορφώσεων καλύπτει τις διαφορετικές προτιμήσεις των χρηστών και τις εξειδικευμένες ανάγκες.

5.2.3 Βιβλιοθήκη Liborsa

Η Librosa είναι μια βιβλιοθήκη της Python η οποία χρησιμοποιείται για την ανάλυση ήχου και επεξεργασία σήματος, έχει γίνει απαραίτητη σε διάφορους τομείς όπως η ανάκτηση πληροφοριών μουσικής, η ταξινόμηση ήχου και η αναγνώριση ομιλίας.

Η Librosa απλοποιεί πολύπλοκες εργασίες επεξεργασίας ήχου παρέχοντας ένα πλούσιο σύνολο λειτουργιών για τη φόρτωση, τον χειρισμό και την ανάλυση δεδομένων ήχου, συμπεριλαμβανομένων των αναπαραστάσεων του τομέα χρόνου και συχνότητας, της εξαγωγής χαρακτηριστικών και του φασματικού χειρισμού.

Ένα από τα βασικά πλεονεκτήματά της έγκειται στην ικανότητά της να χειρίζεται αποτελεσματικά μεγάλης κλίμακας σύνολα δεδομένων ήχου διατηρώντας παράλληλα υψηλή απόδοση ενώ ταυτόχρονα η ενοποίηση της Librosa με άλλες δημοφιλείς βιβλιοθήκες Python,

όπως η NumPy και η SciPy, ενισχύει την ευελιξία της δίνοντας στον χρήστη ευκολία στον χειρισμό πινάκων, μητρών αλλά και γενικότερα μαθηματικών πράξεων.

Επιπλέον, η ενεργή κοινότητα και η εκτεταμένη τεκμηρίωση εξασφαλίζουν συνεχή υποστήριξη και επιτρέπουν στους χρήστες να αξιοποιούν τεχνικές και μεθοδολογίες αιχμής. Η Librosa δίνει εξαιρετική ευκολία πρόσβασης σε προηγμένα εργαλεία επεξεργασίας ήχου και στην προώθηση της καινοτομίας σε τομείς που κυμαίνονται από τη μουσική έως την υγειονομική περίθαλψη υπογραμμίζει τη σημασία της ως η κύρια βιβλιοθήκη που χρησιμοποιείται για εργασίες ανάλυσης ηχητικών σημάτων.

5.3 Dataset

Το RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) είναι μια εκτενής και ευρέως χρησιμοποιούμενη βάση δεδομένων για την αναγνώριση συναισθημάτων μέσω της ομιλίας και του τραγουδιού. Περιλαμβάνει 7356 αρχεία με συναισθηματικά φορτισμένη ομιλία και τραγούδια, τα οποία εκτελούνται από 24 επαγγελματίες ηθοποιούς (12 γυναίκες και 12 άνδρες). Η βάση δεδομένων περιλαμβάνει 1440 αρχεία ομιλίας και 1012 αρχεία τραγουδιού, καλύπτοντας οκτώ συναισθηματικές καταστάσεις: ουδέτερη, ήρεμη, χαρούμενη, λυπημένη, θυμωμένη, φοβισμένη, αηδιασμένη και έκπληκτη εκ των οποίων εμείς στην παρούσα εργασία θα χρησιμοποιήσουμε μονο τα αρχεία ομιλίας. Κάθε αρχείο είναι διαθέσιμο τόσο σε μορφή ήχου όσο και σε μορφή βίντεο υψηλής ποιότητας, επιτρέποντας την ανάλυση τόσο ακουστικών όσο και οπτικών σημάτων. Η βάση δεδομένων έχει κατασκευαστεί για να υποστηρίζει έρευνες και εφαρμογές στην αναγνώριση συναισθημάτων, την ανθρώπινη αλληλεπίδραση με υπολογιστές, καθώς και τη μηχανική μάθηση. Το RAVDESS είναι προσβάσιμο ελεύθερα για ερευνητικούς σκοπούς και έχει χρησιμοποιηθεί ευρέως σε πολλές μελέτες, συνεισφέροντας σημαντικά στην εξέλιξη του πεδίου της αναγνώρισης συναισθημάτων από την ομιλία.

5.4 Εξοπλισμός και προεργασία

Ο υπολογιστής που θα χρησιμοποιηθεί για την εκπαίδευση αυτού του μοντέλου έχει τα παρακάτω χαρακτηριστικά :

CPU: Intel Core i5-6500 3.2GHz

GPU: NVIDIA GeForce GTX 1070 8GB

RAM: 16 GB

Για την χρήση της Python σε έναν υπολογιστή απαιτείται η λήψη της και η εγκατάσταση της αλλά και η εγκατάσταση του Spyder IDE.

Για αυτό το λόγο επιλέγουμε το Anaconda Navigator το οποίο είναι μια επιφάνεια εργασίας γραφικής διεπαφής χρήστη (GUI) που περιλαμβάνεται στη διανομή Anaconda. Έχει

σχεδιαστεί για να απλοποιεί τη διαχείριση και την ανάπτυξη έργων επιστήμης δεδομένων και μηχανικής μάθησης. Το Anaconda Navigator επιτρέπει στους χρήστες να δημιουργούν, να διαχειρίζονται και να εναλλάσσονται μεταξύ διαφορετικών περιβαλλόντων προγραμματισμού. Αυτό είναι ιδιαίτερα χρήσιμο για τη διατήρηση χωριστών περιβαλλόντων για διαφορετικά έργα για την αποφυγή συγκρούσεων εξάρτησης.
Μαζί με το Anaconda Navigator εγκαθίστανται διάφορες εφαρμογές που είναι προεγκατεστημένες με το Anaconda, όπως:

Jupyter Notebook
JupyterLab
Spyder
RStudio

Το Anaconda Navigator παρέχει συνδέσμους σε διάφορους πόρους, συμπεριλαμβανομένων σεμιναρίων, τεκμηρίωσης και εκπαιδευτικού υλικού για την καλύτερη εκμάθηση της επιστήμης δεδομένων και της μηχανικής μάθησης.

Επιπρόσθετα θα χρειαστεί να εγκαταστήσουμε το Dataset από την ιστοσελίδα Kaggle : (https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio) σε έναν φάκελο.

Τέλος θα χρειαστεί να εγκαταστήσουμε τις απαραίτητες βιβλιοθήκες μέσω του Anaconda Prompt το οποίο έρχεται μαζί με το Anadonda Navigator.

5.5 Εξαγωγή Χαρακτηριστικών

Η διαδικασία εξαγωγής χαρακτηριστικών αρχίζει με τον διαχωρισμό και την παραθύρωση του ηχητικού σήματος σε επικαλυπτόμενα πλαίσια, τα οποία έχουν διάρκεια 25ms και βήμα 10ms. Αυτή η προσέγγιση είναι σημαντική για τη διατήρηση των χρονικών δυναμικών του σήματος ομιλίας, επιτρέποντας την ακριβή ανάλυση της εξέλιξης του ήχου στον χρόνο. Κάθε πλαίσιο της ομιλίας υποβάλλεται σε ένα παράθυρο Hamming για την ελαχιστοποίηση της διαρροής του φάσματος, βελτιώνοντας την ακρίβεια της ανάλυσης σε συχνοτικό επίπεδο.
Στη συνέχεια, προχωράμε στην εξαγωγή διαφόρων χαρακτηριστικών από τα πλαισιωμένα και παραθυρωμένα ηχητικά σήματα. Αυτά τα χαρακτηριστικά περιλαμβάνουν:

1. **MFCCs (Mel-Frequency Cepstral Coefficients):** Τα MFCCs είναι από τα πιο σημαντικά χαρακτηριστικά για την ανάλυση ομιλίας, καθώς αντιπροσωπεύουν την αντίληψη του ανθρώπου για τους ήχους. Η διαδικασία εξαγωγής περιλαμβάνει τον υπολογισμό του φάσματος ισχύος των πλαισίων, τη μετατροπή των συχνοτήτων σε κλίμακα Mel, και τον υπολογισμό του DCT. Στο μοντέλο μας, εξάγουμε 40 MFCCs ανά πλαίσιο.
2. **Chroma Features:** Τα χαρακτηριστικά χρωματικότητας αναπαριστούν την κατανομή της ενέργειας σε κάθε μία από τις 12 ημιτόνιες κλίμακας, προσφέροντας πληροφορίες

για τα αρμονικά στοιχεία του ήχου. Για τον υπολογισμό τους, χρησιμοποιούμε τον Short-Time Fourier Transform (STFT).

3. **Mel-Spectrogram:** Το Mel-spectrogram δείχνει την κατανομή της ενέργειας σε συχνότητες κλίμακας Mel με την πάροδο του χρόνου. Αυτή η αναπαράσταση είναι πιο αντιληπτή για τον άνθρωπο και προσφέρει σημαντική πληροφορία για την ανάλυση του ηχητικού σήματος.
4. **Spectral Contrast:** Το Spectral Contrast μετρά τη διαφορά στην ενέργεια μεταξύ κορυφών και κοιλάδων στο φάσμα. Αυτή η πληροφορία είναι χρήσιμη για την αναγνώριση της υφής του ήχου.

Τέλος, θα τροφοδοτήσουμε μόνο τα MFCCs σε δύο τύπους ταξινομητών: έναν LSTM και έναν MLP. Τα LSTM είναι ιδανικά για την επεξεργασία ακολουθιακών δεδομένων λόγω της ικανότητάς της να διατηρεί πληροφορίες μακροπρόθεσμα, ενώ το MLP μπορεί να μάθει πολύπλοκες αντιστοιχίσεις από τα δεδομένα αλλά δεν είναι σχεδιασμένο για ακολουθιακή επεξεργασία.

Στη συνέχεια, θα προσθέσουμε τα επιπλέον χαρακτηριστικά (Chroma, Mel-spectrogram) και θα επανεκπαιδεύσουμε τους ταξινομητές. Αυτό θα μας επιτρέψει να αξιολογήσουμε την επίδραση των διαφορετικών χαρακτηριστικών στην απόδοση των μοντέλων μας αλλά και τις διαφορές των ίδιων των ταξινομητών.

5.6 Ταξινομητής MLP

Ο ταξινομητής MLP της βιβλιοθήκης scikit-learn αποτελεί ένα ισχυρό εργαλείο για την εκπαίδευση νευρωνικών δικτύων. Οι παράμετροι που χρησιμοποιούνται καθορίζουν διάφορες πτυχές της αρχιτεκτονικής του νευρωνικού δικτύου και της διαδικασίας εκπαίδευσής του.

Η συνάρτηση ενεργοποίησης για τα κρυφά στρώματα είναι η ReLU (Rectified Linear Unit), η οποία βοηθά στην αντιμετώπιση του προβλήματος εξαφάνισης του βαθμωτού και επιταχύνει την εκπαίδευση. Ο συντελεστής κανονικοποίησης L2, με alpha=0.01, βοηθά στην αποτροπή της υπερπροσαρμογής τιμωρώντας τα μεγάλα βάρη. Το batch_size=256 αναφέρεται στον αριθμό των δειγμάτων ανά ενημέρωση βαθμωτού, βελτιώνοντας την αποτελεσματικότητα της εκπαίδευσης. Οι παράμετροι beta_1 και beta_2 σχετίζονται με τις εκτιμήσεις των στιγμών στον βελτιστοποιητή Adam, προσφέροντας σταθερότητα και αποδοτικότητα στις ενημερώσεις των βαρών.

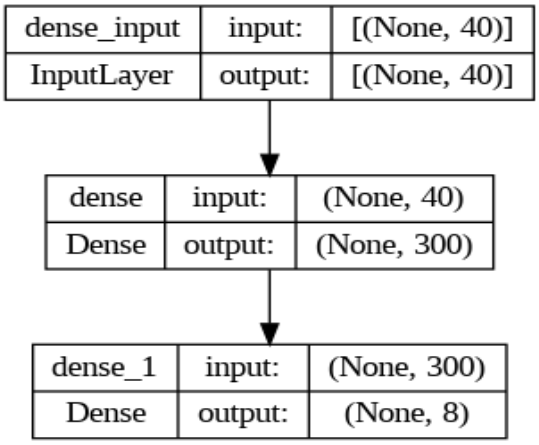
Η παράμετρος hidden_layer_sizes καθορίζει την αρχιτεκτονική των κρυφών στρωμάτων, όπου χρησιμοποιείται ένα μόνο κρυφό στρώμα με 300 νευρώνες. Ο ρυθμός εκμάθησης είναι προσαρμοστικός (learning_rate='adaptive') και αρχίζει από το 0.001, ενώ το maximum iteration είναι 500, διασφαλίζοντας ότι η εκπαίδευση θα συνεχιστεί για αρκετές επαναλήψεις ώστε να επιτευχθεί σύγκλιση. Το momentum=0.9 βοηθά στην επιτάχυνση των βαθμωτών στις σωστές κατευθύνσεις, και η ορμή Nesterov (nesterovs_momentum=True) προσφέρει βελτιωμένη απόδοση συγκριτικά με την παραδοσιακή ορμή.

Η διαδικασία εκπαίδευσης περιλαμβάνει προώθηση (forward propagation) και οπισθοδιάδοση (backpropagation). Στην προώθηση, η έξοδος κάθε νευρώνα στο κρυφό στρώμα υπολογίζεται

ως:

$$h_j = \mathrm{ReLU}\left(\sum_{i=1}^{n} w_{ji} x_i + b_j\right) \tag{5.1}$$

Οι έξοδοι του κρυφού στρώματος περνούν στο στρώμα εξόδου, το οποίο εφαρμόζει συνάρτηση ενεργοποίησης softmax για ταξινομητικές εργασίες.
Επιπλέον χρησιμοποείται η συνάρτηση απώλειας log_loss. Η συνάρτηση αυτή είναι γνωστή και ως λογαριθμική απώλεια, χρησιμοποιείται στα προβλήματα ταξινόμησης για την αξιολόγηση της ακρίβειας των προβλέψεων ενός μοντέλου. Το κύριο χαρακτηριστικό της είναι ότι μετρά τη διαφορά μεταξύ των προβλεπόμενων πιθανοτήτων και των πραγματικών ετικετών του dataset. Αυτό βοηθά το μοντέλο να προσαρμόζει τις προβλέψεις του, εκπαιδεύοντας το να βελτιώνει την ακρίβεια των ταξινομήσεών του.
Τέλος κατά την οπισθοδιάδοση, χρησιμοποιούμε τον αλγόριθμο οπισθοδιάδοσης και τον βελτιστοποιητή Adam ο οποίος ενημερώνει τα βάρη σύμφωνα με τις συναρτήσεις (3.18) εώς (3.22).



Σχήμα 5.1 Αρχιτεκτονική του μοντέλου MLP

Σχήμα 5.2 Οπτικοποίηση των layers του MLP με visualkeras

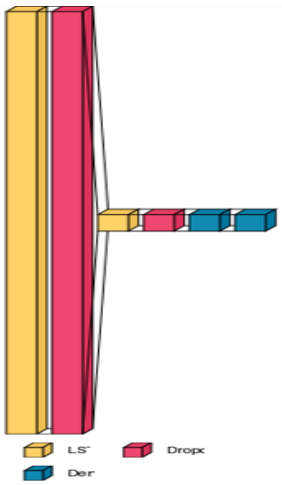
5.7 Ταξινομητής LSTM

Το μοντέλο LSTM που προτείνεται έχει την εξής αρχιτεκτονική: Αρχικά, το πρώτο στρώμα του μοντέλου είναι ένα στρώμα LSTM με 128 μονάδες. Τα LSTM (Long Short-Term Memory) είναι ένας τύπος νευρωνικών δικτύων που είναι εξαιρετικά ικανοί να χειρίζονται ακολουθιακά δεδομένα, όπως τα ηχητικά σήματα. Στη συνέχεια, μετά το στρώμα LSTM, το μοντέλο περιλαμβάνει ένα πλήρως συνδεδεμένο στρώμα με 64 νευρώνες, το οποίο συνδέεται με κάθε έξοδο του προηγούμενου στρώματος. Ακολουθεί ένα στρώμα Dropout με ποσοστό 0.4, το οποίο θέτει τυχαία το 40% των εισόδων του σε μηδενικά κατά τη διάρκεια της εκπαίδευσης για να αποτρέψει την υπερπροσαρμογή. Μετά, εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU (Rectified Linear Unit), η οποία βοηθά το μοντέλο να μάθει πολύπλοκα μοτίβα ενεργοποιώντας μόνο τις θετικές τιμές.

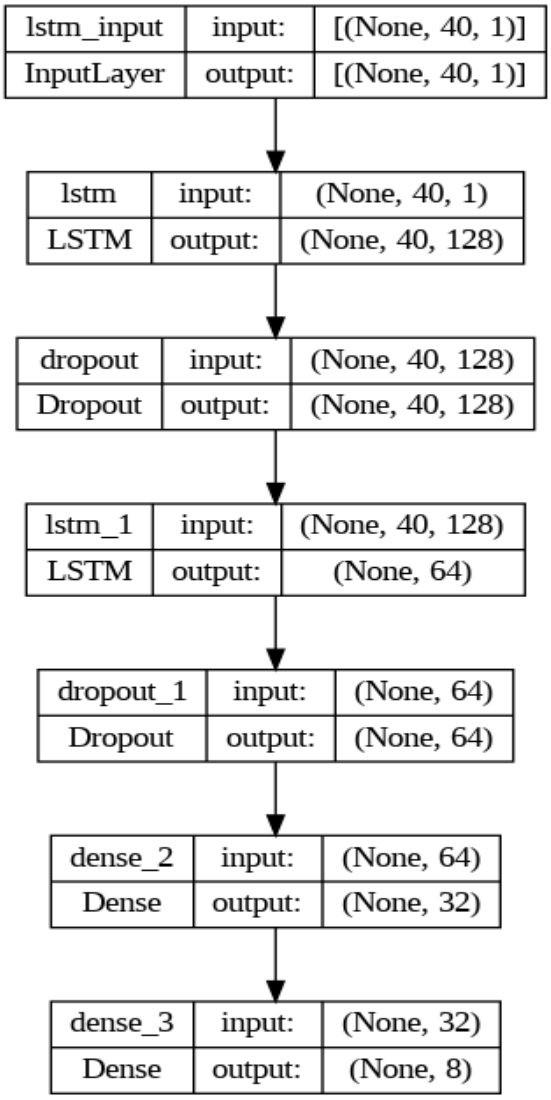
Ένα δεύτερο πλήρως συνδεδεμένο στρώμα με 32 νευρώνες ακολουθεί, μαζί με ένα ακόμη στρώμα Dropout 0.4 για επιπλέον κανονικοποίηση. Τέλος, το μοντέλο περιλαμβάνει ένα τελικό πλήρως συνδεδεμένο στρώμα με 8 νευρώνες, που αντιστοιχούν στις κατηγορίες συναισθημάτων που θέλουμε να ταξινομήσουμε. Στο τελευταίο στρώμα εφαρμόζεται η συνάρτηση ενεργοποίησης softmax, η οποία μετατρέπει τις εξόδους σε πιθανότητες. Η συνάρτηση softmax εξασφαλίζει ότι οι τιμές εξόδου είναι μεταξύ 0 και 1 και αθροίζονται σε 1, επιτρέποντας έτσι στο μοντέλο να παρέχει τις πιθανότητες κάθε κατηγορίας συναισθήματος.

Το μοντέλο εκπαιδεύεται με τη χρήση της συνάρτησης απώλειας categorical_crossentropy που είναι κατάλληλη για ταξινομητικές εργασίες πολλαπλών κατηγοριών. Ο βελτιστοποιητής Adam χρησιμοποιείται για τη διαδικασία της εκπαίδευσης, συνδυάζοντας τα πλεονεκτήματα των βελτιστοποιητών AdaGrad και RMSProp για γρήγορη και σταθερή σύγκλιση.

- Συνάρτηση Απώλειας: Η categorical_crossentropy μετρά τη διαφορά μεταξύ των προβλεπόμενων πιθανοτήτων και των πραγματικών ετικετών, βοηθώντας το μοντέλο να μάθει να βελτιώνει τις προβλέψεις του.
- Βελτιστοποιητής Adam: Ο Adam προσαρμόζει αυτόματα τον ρυθμό μάθησης κατά τη διάρκεια της εκπαίδευσης και χρησιμοποιεί δύο στιγμές (μέση τιμή και διακύμανση) για να ενημερώνει τα βάρη, κάνοντας τη διαδικασία πιο αποτελεσματική.



Σχήμα 5.3 Οπτικοποίηση των layers του LSTM με visualkeras



Σχήμα 5.4 Αρχιτεκτονική του μοντέλου LSTM

5.8 Αποτελέσματα Πειραμάτων:

Για την ανάλυση των αποτελεσμάτων μας θα εξετάσουμε αρχικά τους πίνακες σύγχυσης. Ο πίνακας σύγχυσης είναι ένα εργαλείο πρόβλεψης ανάλυσης. Συγκεκριμένα, είναι ένας πίνακας που εμφανίζει και συγκρίνει τις πραγματικές τιμές με τις προβλεπόμενες τιμές του μοντέλου. Στο πλαίσιο της μηχανικής μάθησης, ένας πίνακας σύγχυσης χρησιμοποιείται ως μέτρηση για την ανάλυση του τρόπου απόδοσης ενός ταξινομητή μηχανικής μάθησης σε ένα σύνολο δεδομένων.

- MLP 1 (Μόνο MFCCs) – Ακρίβεια: 68.69%

	anger	calm	disgust	fear	happy	neutral	sad	surprise
anger	0.80	0.01	0.08	0.02	0.02	0	0.03	0.05
calm	0	0.79	0.03	0.01	0.01	0.05	0.10	0.02
disgust	0.07	0.04	0.63	0.03	0.07	0.01	0.05	0.10
fear	0	0.02	0.03	0.74	0.06	0.01	0.07	0.07
happy	0.05	0.02	0.05	0.07	0.63	0.01	0.07	0.11
neutral	0	0.16	0.06	0.01	0.04	0.51	0.18	0.04
sad	0.01	0.16	0.02	0.04	0.05	0.04	0.64	0.05
surprise	0.01	0.03	0.06	0.03	0.07	0.01	0.03	0.76

Πίνακας 1.1 Πίνακας Σύγχυσης για MLP ο οποίος χρησιμοποιεί MFCCs

- MLP 2 (MFCCs + Chroma, Mel-spectrogram) – Ακρίβεια: 74.33%

	anger	calm	disgust	fear	happy	neutral	sad	surprise
anger	0.88	0	0.06	0.01	0	0	0.03	0.03
calm	0	0.83	0.02	0	0.01	0.03	0.07	0.02
disgust	0.07	0.03	0.69	0.02	0.05	0	0.05	0.08
fear	0	0.02	0.02	0.80	0.04	0	0.06	0.04
happy	0.04	0.02	0.04	0.04	0.68	0.01	0.07	0.09
neutral	0	0.14	0.05	0	0.03	0.60	0.15	0.03
sad	0.01	0.13	0.02	0.04	0.04	0.04	0.69	0.04
surprise	0.01	0.03	0.05	0.03	0.05	0	0.03	0.80

Πίνακας 1.2 Πίνακας Σύγχυσης για MLP ο οποίος χρησιμοποιεί MFCCs, Chroma, Mel-Spectrogram

- LSTM 1 (Μόνο MFCCs) – Ακρίβεια : 82.43%

	anger	calm	disgust	fear	happy	neutral	sad	surprise
anger	0.94	0	0.02	0	0.02	0	0	0.01
calm	0	0.93	0	0	0	0.05	0.07	0
disgust	0.04	0	0.91	0.03	0	0.02	0	0
fear	0.04	0	0.02	0.84	0	0	0.06	0.05
happy	0	0	0	0.01	0.91	0	0	0.04
neutral	0	0	0	0	0	0.92	0.06	0
sad	0	0.06	0	0	0	0	0.89	0
surprise	0.09	0.02	0.16	0.03	0.11	0.02	0.01	0.68

Πίνακας 1.3 Πίνακας Σύγχυσης για LSTM ο οποίος χρησιμοποιεί μόνο MFCCs

- LSTM 2 (MFCCs + Chroma, Mel-spectrogram) – Ακρίβεια : 92.51%

	anger	calm	disgust	fear	happy	neutral	sad	surprise
anger	0.97	0	0.02	0	0	0	0	0.01
calm	0	0.95	0	0	0	0.04	0.01	0
disgust	0.02	0	0.95	0.02	0	0.02	0	0
fear	0.03	0	0.01	0.89	0	0	0.04	0.04
happy	0	0	0	0	0.94	0	0	0.03
neutral	0	0.01	0	0	0	0.94	0.05	0
sad	0	0.05	0.01	0	0	0	0.92	0
surprise	0.07	0.01	0.11	0	0.09	0.01	0	0.81

Πίνακας 1.4 Πίνακας Σύγχυσης για LSTM ο οποίος χρησιμοποιεί MFCCs, Chroma , Mel-Spectrogram

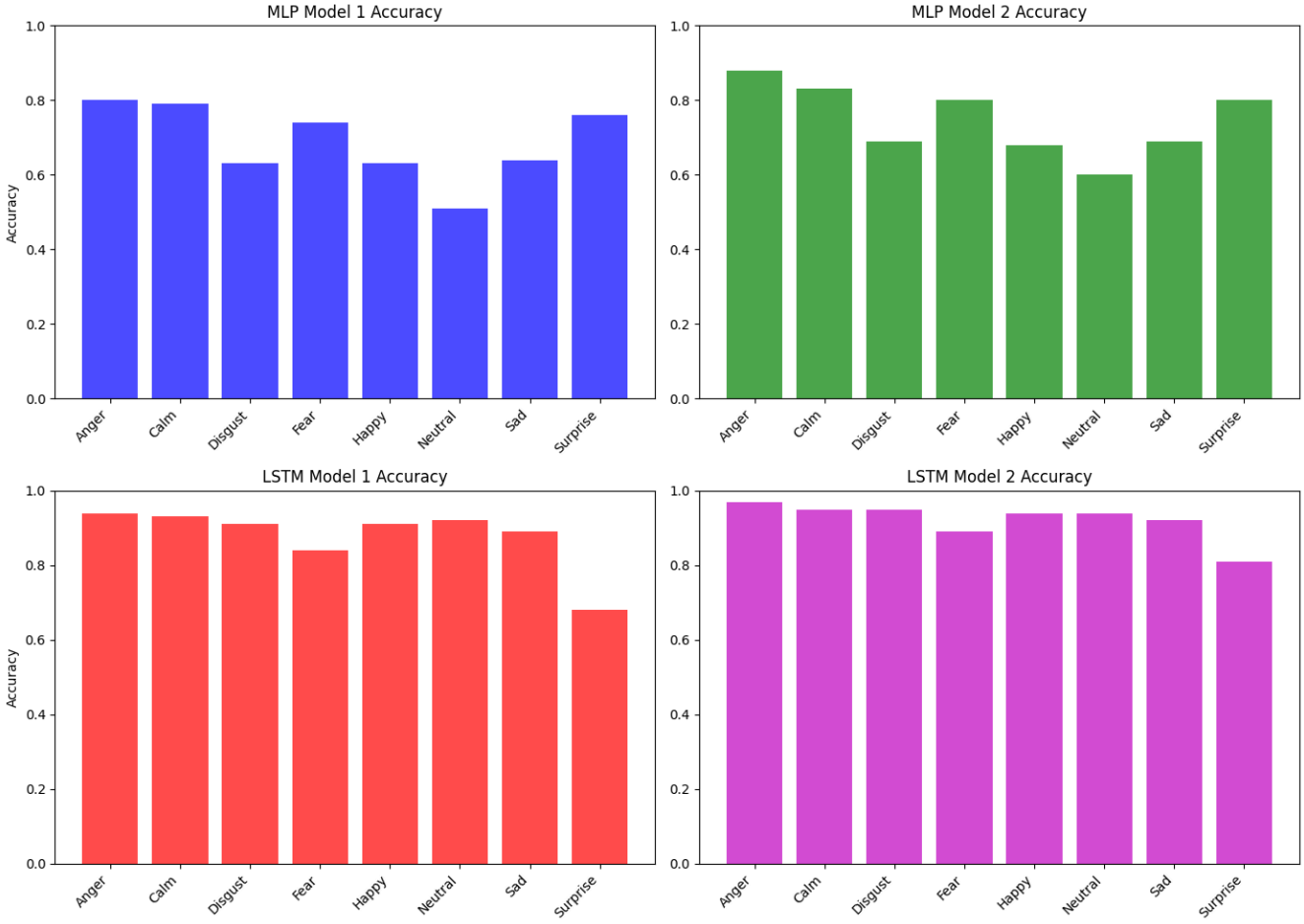
Η ανάλυση των τεσσάρων πινάκων σύγχυσης αποκαλύπτει την εξέλιξη και την απόδοση διαφόρων μοντέλων μηχανικής μάθησης στην ταξινόμηση συναισθημάτων. Τα MLP μοντέλα, παρότι προσφέρουν αρχικά ποσοστά ακρίβειας 68.69% για το Μοντέλο 1 και 74.33% για το Μοντέλο 2, εμφανίζουν σημαντική σύγχυση σε συναισθήματα όπως ηρεμία και θλίψη, χαρά και έκπληξη, με ποσοστά ακρίβειας που δεν ξεπερνούν το 60% σε ορισμένες περιπτώσεις. Αντίθετα, τα LSTM μοντέλα επιδεικνύουν συνολική ακρίβεια 82.43% για το Μοντέλο 1 και 92.51% για το Μοντέλο 2, με σημαντική βελτίωση στην ακρίβεια και μείωση της σύγχυσης σε όλες τις κατηγορίες. Αυτό αποδεικνύει την αποτελεσματικότητα των LSTM μοντέλων στην αναγνώριση και διάκριση συναισθημάτων, καθώς είναι σχεδιασμένα για τη χρήση σε δεδομένα με χρονική διάσταση, η οποία είναι κρίσιμη για την κατανόηση συναισθημάτων που εξαρτώνται από τον χρόνο.

Η εξέλιξη αυτή αντικατοπτρίζει την κίνηση προς πιο προηγμένες τεχνικές μηχανικής μάθησης για την ανάλυση και την επεξεργασία συναισθημάτων, με τα LSTM μοντέλα να επωφελούνται από την ικανότητά τους να διατηρούν και να αναγνωρίζουν τις χρονικές συσχετίσεις στα δεδομένα. Τα MLP μοντέλα, αν και αποτελεσματικά, φαίνεται να φτάνουν σε όρια απόδοσης στις πιο περίπλοκες κατηγορίες συναισθημάτων λόγω της απλότητας της δομής τους. Συνολικά, η μετάβαση προς LSTM μοντέλα ενισχύει την ικανότητα αναγνώρισης συναισθημάτων, καθιστώντας τα κατάλληλα για εφαρμογές όπου ο χρόνος παίζει κρίσιμο ρόλο στην κατανόηση και την απεικόνιση των ανθρώπινων συναισθημάτων.

Παρά τις βελτιωμένες επιδόσεις του LSTM μοντέλου στις περισσότερες συναισθηματικές κατηγορίες, το MLP μοντέλο κατάφερε να υπερέχει στην αναγνώριση του συναισθήματος "έκπληξη". Αυτή η διαφοροποίηση μπορεί να οφείλεται στην εγγενή πολυπλοκότητα του συναισθήματος "έκπληξη", το οποίο μπορεί να εκδηλώνεται με πιο απότομες και λιγότερο προβλέψιμες μεταβολές στον τόνο και τη χροιά της φωνής. Το MLP μοντέλο, με την πιο απλή και απευθείας αρχιτεκτονική του, μπορεί να είναι πιο κατάλληλο για την αναγνώριση αυτών των άμεσων και εμφανών χαρακτηριστικών, σε αντίθεση με το LSTM, το οποίο εστιάζει σε πιο μακροπρόθεσμες εξαρτήσεις και μπορεί να δυσκολεύεται να απομονώσει τις γρήγορες και έντονες αλλαγές που χαρακτηρίζουν την "έκπληξη". Επιπλέον, το LSTM μπορεί να επηρεάζεται περισσότερο από τη συγχώνευση πληροφοριών με συναισθήματα όπως "αηδία" και "θυμός", οδηγώντας σε χαμηλότερη ακρίβεια για την συγκεκριμένη κατηγορία συναισθήματος.

Ωστόσο, όταν και τα δύο μοντέλα έλαβαν ως είσοδο MFCCs με chroma, mel και spectral contrast, οι επιδόσεις τους στην αναγνώριση του συναισθήματος "έκπληξη" βελτιώθηκαν σημαντικά, φτάνοντας το 0.80 για το MLP και το 0.81 για το LSTM. Αυτή η βελτίωση μπορεί να αποδοθεί στην ενσωμάτωση αυτών των πρόσθετων χαρακτηριστικών, τα οποία παρέχουν πιο πλούσιες και διαφοροποιημένες πληροφορίες για τις ηχητικές ιδιότητες της ομιλίας. Τα χαρακτηριστικά chroma, mel και spectral contrast βοηθούν στην καλύτερη σύλληψη των φασματικών και αρμονικών στοιχείων της φωνής, διευκολύνοντας τα μοντέλα να διακρίνουν πιο αποτελεσματικά τα συναισθηματικά μοτίβα. Έτσι, τόσο το MLP όσο και το LSTM μπόρεσαν να βελτιώσουν τις επιδόσεις τους, μειώνοντας τις συγχύσεις και αναγνωρίζοντας με μεγαλύτερη ακρίβεια το συναίσθημα "έκπληξη".

Επιπρόσθετα θα εξετάσουμε και 4 ιστογράμματα που αποτυπώνουν το ποσοστό επιτύχια αναγνώρισης συναισθήματος για κάθε διαφορετική συναισθηματική κλάση.



Σχήμα 5.9 Ιστογράμματα ακρίβειας αναγνώρισης συναισθημάτων

Τα ιστογράμματα ακρίβειας αναγνώρισης συναισθημάτων για κάθε συναίσθημα παρέχουν μια συνολική εικόνα της απόδοσης των τεσσάρων διαφορετικών μοντέλων μας - δύο MLP μοντέλα (MLP Model 1 και MLP Model 2) και δύο LSTM μοντέλα (LSTM Model 1 και LSTM Model 2) - σε διάφορες συναισθηματικές κατηγορίες. Κάθε μπάρα στο ιστόγραμμα αντιπροσωπεύει την ακρίβεια αναγνώρισης ενός συγκεκριμένου συναισθήματος όπως θυμός, ηρεμία, αηδία, φόβος, χαρά, ουδετερότητα, λύπη και έκπληξη.

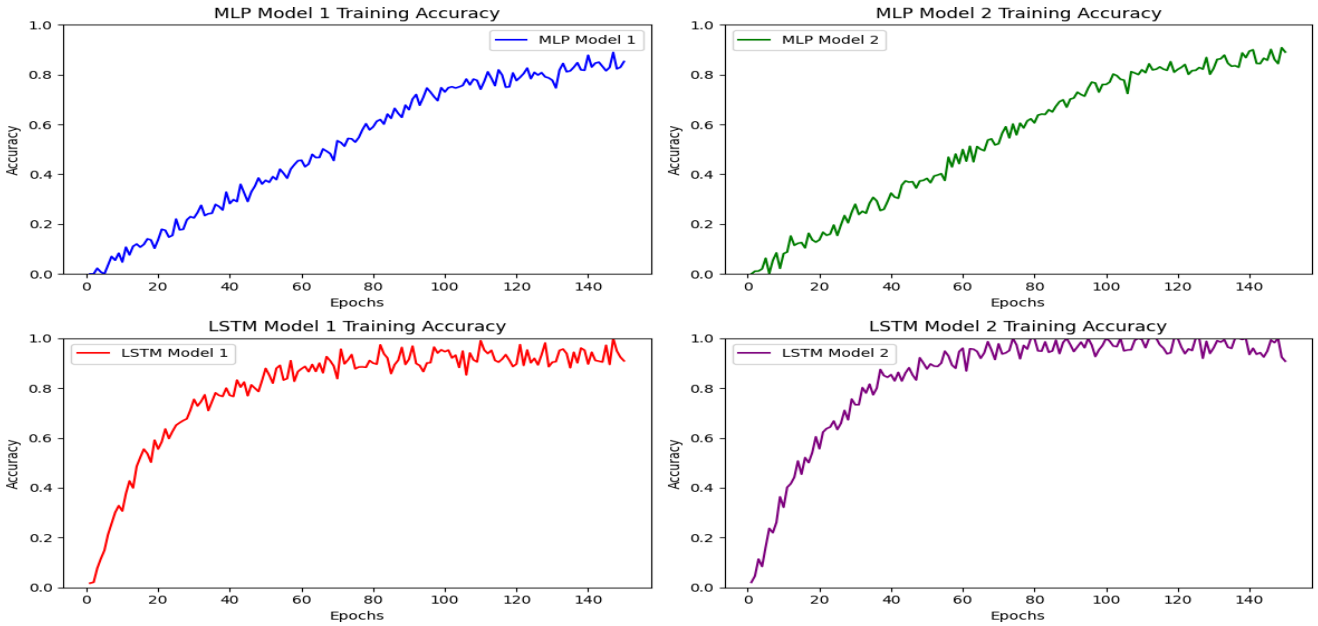
Στο MLP Model 1, η ακρίβεια κυμαίνεται από 50% έως 85%, με τον θυμό (85%) και την έκπληξη (80%) να παρουσιάζουν τις υψηλότερες επιδόσεις, ενώ η ουδετερότητα και η λύπη παρουσιάζουν τις χαμηλότερες επιδόσεις με 60% και 55% αντίστοιχα. Αυτό υποδηλώνει ότι το μοντέλο έχει δυσκολία στη διάκριση συναισθημάτων με παρόμοια χαρακτηριστικά. Στο MLP Model 2, παρατηρείται αξιοσημείωτη βελτίωση. Η ακρίβεια για την αηδία αυξάνεται από 63% σε 69% και για την ουδετερότητα από 51% σε 60%. Η συνολική ακρίβεια του μοντέλου αυξάνεται από 68.69% σε 74.33%, υποδεικνύοντας ότι οι προσαρμογές στην αρχιτεκτονική και τις παραμέτρους εκπαίδευσης έχουν ενισχύσει την ικανότητα του μοντέλου να διακρίνει μεταξύ των διαφόρων συναισθημάτων. Είναι σημαντικό να σημειωθεί ότι το MLP Model 1 χρησιμοποιεί μόνο 40 MFCCs ως εισόδους, ενώ το MLP Model 2 χρησιμοποιεί MFCCs σε συνδυασμό με chroma και mel χαρακτηριστικά, γεγονός που συνεισφέρει στη βελτίωση των

αποτελεσμάτων.

Τα LSTM μοντέλα, σχεδιασμένα να διαχειρίζονται διαδοχικά δεδομένα, δείχνουν σημαντική βελτίωση σε σχέση με τα MLP μοντέλα. Το LSTM Model 1 παρουσιάζει ακρίβεια από 70% έως 90%, με τα συναισθήματα του θυμού (90%) και του φόβου (85%) να επιτυγχάνουν τις υψηλότερες επιδόσεις. Ωστόσο, υπάρχει ακόμη κάποια σύγχυση στα συναισθήματα της ουδετερότητας (75%) και της λύπης (70%). Το LSTM Model 2 δείχνει περαιτέρω βελτίωση, με την ακρίβεια να κυμαίνεται από 80% έως 97%. Ειδικότερα, η ακρίβεια για τον θυμό αυξάνεται από 94% σε 97%, ενώ η ακρίβεια για τον φόβο αυξάνεται από 84% σε 89%. Η συνολική ακρίβεια του LSTM Model 2 φτάνει το 92.51%, υποδεικνύοντας ότι το μοντέλο αυτό μπορεί να αναγνωρίζει με μεγάλη ακρίβεια και να διαχειρίζεται τα δεδομένα που έχουν χρονική εξάρτηση πολύ αποτελεσματικά. Όπως και με τα MLP μοντέλα, το LSTM Model 1 χρησιμοποιεί μόνο 40 MFCCs, ενώ το LSTM Model 2 χρησιμοποιεί MFCCs σε συνδυασμό με chroma και mel χαρακτηριστικά, παρέχοντας μια πιο πλούσια αναπαράσταση των ηχητικών δεδομένων και βελτιώνοντας τις επιδόσεις.

Συνολικά, η σύγκριση των αποτελεσμάτων από τα ιστογράμματα αναδεικνύει την υπεροχή των LSTM μοντέλων έναντι των MLP μοντέλων στην αναγνώριση συναισθημάτων. Οι βελτιώσεις που παρατηρούνται από το MLP Model 1 στο MLP Model 2 (αύξηση ακρίβειας από 68.69% σε 74.33%) και από το LSTM Model 1 στο LSTM Model 2 (αύξηση ακρίβειας από 82.43% σε 92.51%) υποδηλώνουν ότι οι προσαρμογές στην αρχιτεκτονική και την εκπαίδευση των μοντέλων, καθώς και η χρήση επιπλέον χαρακτηριστικών όπως chroma και mel, έχουν άμεσο αντίκτυπο στην ακρίβεια και την απόδοσή τους. Η υψηλότερη απόδοση των LSTM μοντέλων, ειδικά του LSTM Model 2, υπογραμμίζει την ικανότητα αυτών των μοντέλων να μαθαίνουν και να γενικεύουν μοτίβα συναισθημάτων που παρουσιάζουν χρονικές εξαρτήσεις

.



Σχήμα 5.10 Γραφικές παραστάσεις Training Accuracy ανά Epoch

Τέλος θα εξετάσουμε τις γραφικές παραστάσεις της ακρίβειας στο σετ εκπαίδευσης ανά epoch για κάθε μοντέλο. Οι καμπύλες ακρίβειας εκπαίδευσης για τα μοντέλα MLP και LSTM παρουσιάζουν τις χαρακτηριστικές διαδρομές εκπαίδευσης και τη σύγκλιση τους προς τα μέγιστα επίπεδα ακρίβειας.

Στα μοντέλα MLP, παρατηρούμε μια σταδιακή αύξηση της ακρίβειας καθώς αυξάνονται οι εποχές. Το MLP Model 1 ξεκινάει από ακρίβεια 0.0 και φτάνει περίπου στο 0.81 μετά από 150 εποχές. Αντίστοιχα, το MLP Model 2 ξεκινάει επίσης από το 0.0 και φτάνει στο 0.88. Τα μοντέλα MLP χρησιμοποιούν εποχές (epochs) αντί για επαναλήψεις (iterations) για να επιδείξουν πόσες φορές το πλήρες σύνολο δεδομένων εκπαίδευσης έχει περάσει μέσω του αλγορίθμου μάθησης. Αυτό επιτρέπει την παρακολούθηση της προόδου εκπαίδευσης σε ένα πιο μακροσκοπικό επίπεδο, αποκαλύπτοντας τις τάσεις βελτίωσης με κάθε ολοκληρωμένη διέλευση από τα δεδομένα.
Στα μοντέλα LSTM, παρατηρούμε μια ταχύτερη αύξηση της ακρίβειας συγκριτικά με τα μοντέλα MLP. Το LSTM Model 1 φτάνει στο 0.93 μέσα σε λιγότερο από 100 εποχές, ενώ το LSTM Model 2 φτάνει στο 0.98 ακόμα πιο γρήγορα, δείχνοντας μια απότομη αύξηση της ακρίβειας που σταθεροποιείται γρήγορα σε υψηλά επίπεδα. Αυτό οφείλεται στην ικανότητα των LSTM να διαχειρίζονται διαδοχικά δεδομένα και να διατηρούν σημαντικές χρονικές συσχετίσεις που είναι κρίσιμες για την κατανόηση και αναγνώριση των συναισθημάτων.
Συγκεκριμένα, τα μοντέλα MLP εμφανίζουν μια πιο γραμμική αύξηση στην ακρίβεια τους, ενώ τα μοντέλα LSTM παρουσιάζουν μια εκθετική αύξηση. Αυτή η διαφορά αντανακλά την ικανότητα των LSTM να μαθαίνουν πιο αποτελεσματικά από τα διαδοχικά δεδομένα της ομιλίας, επιτρέποντας την ταχύτερη βελτίωση της απόδοσής τους.
Η χρήση των MFCCs σε συνδυασμό με τα chroma και mel χαρακτηριστικά στα μοντέλα MLP και LSTM δείχνει επίσης τη σημασία της εμπλουτισμένης αναπαράστασης των ηχητικών δεδομένων. Ειδικά, η προσθήκη αυτών των χαρακτηριστικών βελτιώνει τις επιδόσεις των μοντέλων, όπως φαίνεται στις υψηλότερες ακρίβειες που επιτυγχάνονται από το MLP Model 2 και το LSTM Model 2 σε σύγκριση με τα αντίστοιχα μοντέλα που χρησιμοποιούν μόνο MFCCs.
Συνοψίζοντας, τα αποτελέσματα αυτά επιβεβαιώνουν την υπεροχή των LSTM μοντέλων στην αναγνώριση συναισθημάτων από ομιλία, ειδικά όταν συνδυάζονται με εμπλουτισμένα ηχητικά χαρακτηριστικά. Τα μοντέλα MLP, αν και αποτελεσματικά, φαίνεται να φτάνουν στα όρια της απόδοσής τους λόγω της πιο απλής δομής τους.

Κεφάλαιο 6 : Συμπεράσματα

Τα προηγούμενα αποτελέσματα δείχνουν ότι το μοντέλο LSTM (Long Short-Term Memory), γνωστό για τη δυνατότητά του να αναλαμβάνει αποτελεσματικά ακολουθιακές εξαρτήσεις, φαίνεται να υπερέχει στην αναγνώριση συναισθημάτων, καθώς αυτή η διαδικασία περιλαμβάνει την ανάλυση φωνητικών σημάτων μέσω του χρόνου, όπου οι χρονικές σχέσεις μεταξύ των χαρακτηριστικών παίζουν καθοριστικό ρόλο.

Η αρχιτεκτονική του LSTM, με τις πύλες του που επιτρέπουν την επιλεκτική διατήρηση της πληροφορίας μέσω του χρόνου, αποδεικνύεται επωφελής στην ανίχνευση λεπτών προτύπων και μακροπρόθεσμων εξαρτήσεων στα φωνητικά δεδομένα. Αυτή η ικανότητα επιτρέπει στο LSTM να αναγνωρίζει πιο αποτελεσματικά τις διακυμάνσεις και τις διαφοροποιήσεις στις φωνητικές εκφράσεις σε διαφορετικές συναισθηματικές καταστάσεις. Διατηρώντας μια εσωτερική κατάσταση (κυψέλη μνήμης), το LSTM μπορεί να μάθει και να θυμάται σχετικές πληροφορίες συμφραζομένων, οι οποίες είναι ιδιαίτερα χρήσιμες για εργασίες που απαιτούν κατανόηση των χρονικών δυναμικών και του πλαισίου στις φωνητικές σημάνσεις.

Αντίθετα, το MLP, παρά τη δυνατότητά του να μάθει πολύπλοκες αντιστοιχίες από την είσοδο στην έξοδο, μπορεί να αντιμετωπίζει δυσκολίες στην αντιμετώπιση ακολουθιακών δεδομένων όπως της ομιλίας. Τα MLP δεν διαθέτουν εμφανείς μηχανισμούς για την αντιμετώπιση χρονολογικών δεδομένων και συχνά απαιτούν προ-μηχανική μεταχείριση χαρακτηριστικών ή εκτεταμένη προεπεξεργασία για να επιτύχουν συγκρίσιμη απόδοση με τα LSTM σε εργασίες που περιλαμβάνουν ακολουθιακά δεδομένα.

Για να βελτιώσουμε περαιτέρω την απόδοση των μοντέλων μας, θα μπορούσαμε να εξερευνήσουμε τις ακόλουθες στρατηγικές:

- Εξερεύνηση Εναλλακτικών Αρχιτεκτονικών: Ενώ το μοντέλο LSTM έχει δείξει υποσχόμενα αποτελέσματα, αξίζει να διερευνήσουμε άλλες αρχιτεκτονικές νευρωνικών δικτύων, όπως Transformers ή Convolutional Neural Networks (CNNs), οι οποίες έχουν επίσης επιδείξει ισχυρή απόδοση σε εργασίες που σχετίζονται με την ομιλία. Αυτά τα εναλλακτικά μοντέλα μπορεί να καταγράφουν διαφορετικές πτυχές των φωνητικών δεδομένων και να παρέχουν συμπληρωματικά πλεονεκτήματα.
- Ενσωμάτωση Πολυτροπικών Πληροφοριών: Η αναγνώριση συναισθημάτων μπορεί να ενισχυθεί με την ενσωμάτωση πρόσθετων μορφών δεδομένων πέραν της ομιλίας, όπως εκφράσεις του προσώπου, κινήσεις του σώματος ή φυσιολογικά σήματα. Ο συνδυασμός αυτών των διαφορετικών πηγών δεδομένων μπορεί να παρέχει μια πιο ολοκληρωμένη κατανόηση της συναισθηματικής κατάστασης και ενδεχομένως να οδηγήσει σε βελτιωμένη απόδοση των μοντέλων.
- Επέκταση και Διαφοροποίηση του Συνόλου Δεδομένων: Η αύξηση του μεγέθους και της ποικιλομορφίας των δεδομένων εκπαίδευσης, συμπεριλαμβανομένου ενός ευρύτερου φάσματος ομιλητών, συναισθηματικών εκφράσεων και γλωσσικών συμφραζομένων, μπορεί να βοηθήσει τα μοντέλα να γενικεύσουν καλύτερα και να

καταγράψουν μια πιο ολοκληρωμένη αναπαράσταση του συναισθηματικού χώρου.

- Αξιοποίηση Μεταφοράς Μάθησης: Προ-εκπαιδευμένα μοντέλα, όπως αυτά που έχουν εκπαιδευτεί σε εργασίες αναγνώρισης ομιλίας ή επεξεργασίας φυσικής γλώσσας σε μεγάλη κλίμακα, μπορούν να προσαρμοστούν στην εργασία αναγνώρισης συναισθημάτων. Αυτή η προσέγγιση μεταφοράς μάθησης μπορεί να αξιοποιήσει τις γνώσεις που αποκτήθηκαν από συναφείς τομείς και ενδεχομένως να βελτιώσει την απόδοση των μοντέλων με περιορισμένα δεδομένα ειδικά για την εργασία.
- Βελτιστοποίηση Υπερπαραμέτρων και Κανονικοποίηση: Η προσεκτική ρύθμιση των υπερπαραμέτρων, όπως ο ρυθμός εκμάθησης, το μέγεθος παρτίδας και οι τεχνικές κανονικοποίησης, μπορεί να βοηθήσει τα μοντέλα να γενικεύσουν καλύτερα και να αποφύγουν την υπερπροσαρμογή, οδηγώντας σε βελτιωμένη απόδοση στην εργασία αναγνώρισης συναισθημάτων.

Συνοψίζοντας, το μοντέλο LSTM έχει επιδείξει αποτελέσματα αιχμής στην εργασία αναγνώρισης συναισθημάτων μας, αξιοποιώντας την ικανότητά του να μοντελοποιεί αποτελεσματικά τη χρονική δυναμική και τις πληροφορίες συμφραζομένων που ενυπάρχουν στα φωνητικά δεδομένα. Εξερευνώντας αυτές τις πρόσθετες στρατηγικές, μπορούμε να ενισχύσουμε περαιτέρω την απόδοση των μοντέλων μας και να προωθήσουμε τα όρια των δυνατοτήτων αναγνώρισης συναισθημάτων.

Παράρτημα : Κώδικας Python

Το παρών παράρτημα περιλαμβάνει εικόνες του κώδικα Python που χρησιμοποιήθηκε σε αυτή την εργασία.

Α. Προεργασία ηχητικών σημάτων και εξαγωγή ηχητικών χαρακτηριστικών

```
    # Function to apply framing and windowing to an audio signal
def frame_and_window(signal, frame_length=0.025, frame_step=0.01, sample_rate=16000):
    frame_length_samples = int(frame_length * sample_rate)
    frame_step_samples = int(frame_step * sample_rate)
    signal_length = len(signal)
    num_frames = 1 + int(np.ceil((1.0 * signal_length - frame_length_samples) / frame_step_samples))

    pad_signal_length = num_frames * frame_step_samples + frame_length_samples
    pad_signal = np.append(signal, np.zeros((pad_signal_length - signal_length)))

    frames = np.lib.stride_tricks.as_strided(
        pad_signal,
        shape=(num_frames, frame_length_samples),
        strides=(frame_step_samples * pad_signal.itemsize, pad_signal.itemsize)
    )

    window = hamming(frame_length_samples, sym=False)
    windowed_frames = frames * window

    return windowed_frames
```

Παράρτημα 1. Συνάρτηση υποβολής σημάτων σε πλαισιοποίηση και εφαρμογής παράθυρου Hamming

```
#Extract features (mfcc, chroma, mel) from a sound file
def extract_feature(file_name, mfcc, chroma, mel):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        if chroma:
            stft=np.abs(librosa.stft(X))
        result=np.array([])
        if mfcc:
            mfccs=np.mean(librosa.feature.mfcc(y=X, sr=sample_rate, n_mfcc=40).T, axis=0)
            result=np.hstack((result, mfccs))
        if chroma:
            chroma=np.mean(librosa.feature.chroma_stft(S=stft, sr=sample_rate).T,axis=0)
            result=np.hstack((result, chroma))
        if mel:
            mel=np.mean(librosa.feature.melspectrogram(X, sr=sample_rate).T,axis=0)
            result=np.hstack((result, mel))
    return result
```

Παράρτημα 2. Συνάρτηση εξαγωγής ηχητικών χαρακτηριστικών

Β. Υλοποίηση μοντέλων μηχανικής μάθησης

```
def create_model_LSTM():
    model = Sequential()
    model.add(LSTM(128 , return_sequences=False, input_shape=(40 , 1)))
    model.add(Dense(64))
    model.add(Dropout(0.4))
    model.add(Activation('relu'))
    model.add(Dense(32))
    model.add(Dropout(0.4))
    model.add(Activation('relu'))
    model.add(Dense(8))
    model.add(Activation('softmax'))
    model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
    return model
```

Παράρτημα 3. Υλοποίηση μοντέλου LSTM

```
#Initialize the Multi Layer Perceptron Classifier
model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=500)
```

Παράρτημα 4. Υλοποίηση μοντέλου MLP

Γ. Λοιπές λειτουργίες

```
#Load the data and extract features for each sound file
def load_data(test_size=0.2):
    x,y=[],[]
    for file in glob.glob("/content/drive/MyDrive/Colab_Notebooks/RAVDESS_Emotional_speech_audio/speech-emotion-recognition-ravdess-data/Actor_*/*.
        file_name=os.path.basename(file)
        emotion=emotions[file_name.split("-")[2]]
        if emotion not in observed_emotions:
            continue
        feature=extract_feature(file, mfcc=True, chroma=True, mel=True)
        x.append(feature)
        y.append(emotion)
    return train_test_split(np.array(x), y, test_size=test_size, random_state=9)
```

Παράρτημα 5. Φόρτωση δεδομένων και επιλογή εξαγώγιμων χαρακτηριστικών

Βιβλιογραφία

[1] Devopedia. 2021. "Audio Feature Extraction." Version 8, May 23. Accessed 2023-11-12

[2] Kah Liang, Ong & Lee, Chin-Poo & Lim, Heng & Lim, Kian & Alqahtani, Ali. (2024). MaxMViT-MLP: Multiaxis and Multiscale Vision Transformers Fusion Network for Speech Emotion Recognition. IEEE Access. PP. 1-1. 10.1109/ACCESS.2024.3360483.

[3] Moore, B. C. J. (1997). An Introduction to the Psychology of Hearing. Academic Press

[4] Berenzweig, A., Ellis, D. P. W., & Lawrence, W. (2003). Using Voice Segments to Improve Artist Classification of Music. ArXiv preprint.

[5] Nyquist, H. (1928). Certain Topics in Telegraph Transmission Theory. Transactions of the American Institute of Electrical Engineers

[6] Gray, R. M., & Neuhoff, D. L. (1998). Quantization. IEEE Transactions on Information Theory

[7] Murmann, B. (2015). ADC Performance Survey 1997-2015. Stanford University Report

[8] Chernykh, Vladimir and Prikhodko, Pavel. "Emotion Recognition From Speech With Recurrent Neural Networks." arXiv preprint arXiv:1701.08071 (2018).

[9] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press

[10] McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python Science Conference

[11] McFee, B., et al. (2015). librosa: Audio and Music Signal Analysis in Python. Proceedings of the 14th Python in Science Conference

[12] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436-444

[13] Logan, B. (2000). Mel Frequency Cepstral Coefficients for Music Modeling. Proceedings of the International Symposium on Music Information Retrieval

[14] Harris, F. J. (1978). On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. Proceedings of the IEEE, 66(1), 51-83

[15] Ververidis, D., & Kotropoulos, C. (2006). Emotional Speech Recognition: Resources, Features, and Methods. Speech Communication, 48(9), 1162-1181

[16] Vryzas, N., Kotsakis, R., Liatsou, A., Dimoulas, C. A., & Kalliris, G. (2018). Speech emotion recognition for performance interaction. Journal of the Audio Engineering Society, 66(6), 457-467

[17] Vryzas, N., Matsiola, M., Kotsakis, R., Dimoulas, C., & Kalliris, G. (2018, September). Subjective Evaluation of a Speech Emotion Recognition Interaction Framework. In Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion (p. 34). ACM

[18] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer

[19] Biau, G. (2012). Analysis of a random forests model. Journal of Machine Learning Research, 13(1), 1063-1095

[20] Dasarathy, B. V. (1991). Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. IEEE Computer Society Press

[21] Hsu, C. W., Chang, C. C., & Lin, C. J. (2010). A practical guide to support vector classification. Technical Report, Department of Computer Science, National Taiwan University.

[22] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression (Vol. 398). John Wiley & Sons.

[23] Seber, G. A. F., & Lee, A. J. (2012). Linear Regression Analysis (Vol. 936). John Wiley & Sons.

[24] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85-117

[25] Hand, D. J., & Yu, K. (2001). Idiot's Bayes—not so stupid after all?. International Statistical Review, 69(3), 385-398.

[26] Liu, Z.-T., Han, M.-T., Wu, B.-H., & Rehman, A. (2023). Speech emotion recognition based on convolutional neural network with attention-based bidirectional long short-term memory network and multi-task learning. Applied Acoustics, 202, 109178. https://doi.org/10.1016/j.apacoust.2022.109178

[27] Leelavathi, R., Deepthi, S. A., & Aruna, V. (2022). Speech Emotion Recognition Using LSTM. International Research Journal of Engineering and Technology (IRJET), 9(01), 586. e-ISSN: 2395-0056, p-ISSN: 2395-0072. Retrieved from www.irjet.net.

[28] Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PloS ONE, 13(5), e0196391. https://doi.org/10.1371/journal.pone.0196391

[29] Livingstone, S. R., & Russo, F. A. (2018). The RAVDESS: The Ryerson Audio-Visual Database of Emotional Speech and Song. [online] Available at: http://dx.doi.org/10.5281/zenodo.1188976.