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

Μελέτη Μουσικής Ομοιότητας με τη Χρήση Ευφυών Συστημάτων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δήμητρα - Δέσποινα Α. Μαούτσα

Επιβλέπων : Ανδρέας – Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2013

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

Μελέτη Μουσικής Ομοιότητας με τη Χρήση Ευφυών Συστημάτων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δήμητρα - Δέσποινα Α. Μαούτσα

Επιβλέπων : Ανδρέας – Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 17 Ιουλίου 2013.

.....	.....	.....
Ανδρέας - Γεώργιος Σταφυλοπάτης	Στέφανος Κόλλιας	Γεώργιος Στάμου

Αθήνα, Ιούλιος 2013

.....

Δήμητρα - Δέσποινα Α. Μαούτσα

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

Περίληψη

Ο σκοπός της παρούσας διπλωματικής εργασίας ήταν η μελέτη της έννοιας της μουσικής ομοιότητας και η σύγκριση διαφόρων συστημάτων ταξινόμησης ως προς την επίδοση τους στην εκμάθηση ενός μέτρου απόστασης, το οποίο εκφράζει την ανομοιότητα μεταξύ δύο μουσικών κομματιών.

Χρησιμοποιώντας το σύνολο δεδομένων MagnaTagATune για εξαγωγή των μουσικών χαρακτηριστικών, και μετατρέποντας τις αξιολογήσεις των χρηστών που περιέχονται σε αυτό υπό την μορφή ψήφων ανομοιοτήτων σε περιορισμούς σχετικών αποστάσεων κατασκευάσαμε το πρόβλημα δυαδικής ταξινόμησης, το οποίο στοχεύει στην μάθηση του μέτρου απόστασης, που αντιστοιχεί στην ανομοιότητα των κομματιών.

Για την επίλυση του προβλήματος ταξινόμησης χρησιμοποιήθηκαν τόσο Τεχνητά Νευρωνικά Δίκτυα, όσο και Μηχανές Διανυσμάτων Υποστήριξης (SVM), καθώς και μια παραλλαγή Νευρωνικού Δικτύου επηρεασμένη από τα SVM, το SVNN.

Τέλος προκειμένου να αποκτήσουμε μια ποιοτική έποψη του αποτελέσματος του Νευρωνικού Δικτύου επιχειρήθηκε μια τρισδιάστατη απεικόνιση των μουσικών κομματιών της βάσης με χρήση της μεθόδου Multidimensional Scaling (MDS) καθώς και μια προσπάθεια ερμηνείας της σημαντικότητας των μουσικών χαρακτηριστικών που χρησιμοποιήθηκαν, βασιζόμενοι στο διάνυσμα συναπτικών βαρών του Νευρωνικού Δικτύου.

Λέξεις Κλειδιά:

Μουσική Ομοιότητα, Περιορισμοί Σχετικών Αποστάσεων, Εξόρυξη Μουσικής Πληροφορίας, MagnaTagATune, EchoNest, ENT Χαρακτηριστικά, TagATune, MFCC, Ένταση, Ενέργεια Βραχέως Χρόνου, Ρυθμός, ZCR, Roll off, Ταξινόμηση, Τεχνητά Νευρωνικά Δίκτυα, Μηχανές Διανυσμάτων Υποστήριξης, wSVM, Νευρωνικό Δίκτυο Διανυσμάτων Υποστήριξης, SVNN, MDS

Abstract

The purpose of this thesis was the study of the notion of music similarity and the comparison of the efficiency of several classification algorithms in learning a distance metric, that represents the dissimilarity between two music pieces.

Having used the MagnaTagATune dataset for the aid of feature extraction and after converting the user's dissimilarity ratings, that accompany the tracks in the dataset, into relative distance constraints we define the binary classification problem, that aims to learn the distance metric, that represents the dissimilarities of the songs.

For the implementation of the classification problem Artificial Neural Networks (ANN) and Support Vectors Machines (SVM) have been used, as well as a Neural Network variation influenced by SVM's, called SVNN.

Finally, in order to obtain a qualitative point of view of the output of the Neural Network model, we attempted a three-dimensional embedding of the musical clips included in the database using Multidimensional Scaling (MDS) and an attempt to interpret the significance of the musical features, based on the vector of synaptic weights of the Neural Network.

Key Words:

Music Similarity, Relative Distance Constraints, Music Information Retrieval, MagnaTagATune, EchoNest, ENT Features, TagATune, MFCC, Loudness, RMS, Tempo, ZCR, Roll off, Classification, Artificial Neural Networks, Support Vectors Machine, wSVM, Support Vector Neural Network, SVNN, Multidimensional Scaling , MDS

Περιεχόμενα

1. **ΕΙΣΑΓΩΓΗ**
   - 1.1 Εισαγωγή....10
   - 1.2 Διάρθρωση της εργασίας....10
2. **MUSIC INFORMATION RETRIEVAL**
   - 2.1 Εισαγωγή....12
   - 2.2 Μουσική Ομοιότητα....13
   - 2.3 Music Information Retrieval....14
   - 2.4 Περιοχές Έρευνας του MIR....15
   - 2.5 Περιγραφείς Μουσικών Κομματιών....17
   - 2.5.1 Μεταδεδομένα Κειμένου....17
   - 2.5.2 Πληροφορίες Μουσικού Περιεχομένου....18
   - 2.5.2.1 Υπολογισμός των MFCC....26
3. **ΤΟΠΟΘΕΤΗΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ**
   - 3.1 Εισαγωγή....31
   - 3.2 Relative Constraints....32
   - 3.3 Κατασκευή Γράφου Περιορισμών....33
   - 3.4 Διατύπωση του Προβλήματος Μάθησης Απόστασης....33
   - 3.5 Μετασχηματισμός σε Πρόβλημα Ταξινόμησης....35
4. **ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ**
   - 4.1 Ιστορική Αναδρομή....37
   - 4.2 Βιολογικό Πρότυπο....38
   - 4.3 Δομή και Λειτουργία Τεχνητού Νευρώνα....39
   - 4.4 Τύποι Συναρτήσεων Ενεργοποίησης....41
   - 4.5 Perceptron....43
   - 4.6 Πολυστρωματικά Νευρωνικά Δικτυα (Multilayer Perceptron)....44
   - 4.7 Εκπαίδευση....46
   - 4.7.1 Backpropagation....47
   - 4.8 Κανόνες Μάθησης....52
5. **ΜΗΧΑΝΕΣ ΔΙΑΝΥΣΜΑΤΩΝ ΥΠΟΣΤΗΡΙΞΗΣ**
   - 5.1 Εισαγωγή....56
   - 5.2 Θεωρία Στατιστικής Μάθησης....56
   - 5.3 Η Διάσταση VC....58
   - 5.4 Αρχή Ελαχιστοποίησης του Δομικού Κινδύνου....58
   - 5.5 Ταξινομητές Υπερεπιπέδου....58
   - 5.6 Η ικανότητα γενίκευσης του τέλεια εκπαιδευμένου SVM....62
   - 5.7 Soft Margin Classification....63
   - 5.8 Χρήση Συναρτήσεων Πυρήνα....64
   - 5.9 Εκπαίδευση SVM....67
   - 5.10 wSVM....67

6. **Support Vector Neural Network – SVNN**
6.1 Εισαγωγή....68
6.2 Eigenvalue Decay....68
6.2.1 Ανάλυση της Eigenvalue Decay....69
6.3 Ταξινόμηση με το SVNN....73
7. **ΜΕΤΡΙΚΕΣ ΑΞΙΟΛΟΓΗΣΗΣ**
7.1 Εισαγωγή....75
7.2 Confusion Matrix....75
7.3 Receiver Operator Characteristics - ROC....77
8. **VIZUALIZATION**
8.1 Εισαγωγή....79
8.2 Multidimensional Scaling (MDS)....80
8.2.1 Μετρική Κλασσική MDS....81
8.2.2 Μη Μετρική MDS....82
9. **ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ**
9.1 Επιλογή Dataset....84
9.2 Κατασκευή Διανύσματος Χαρακτηριστικών....85
9.2.1 EchoNest Analyzer....85
9.2.2 EchoNest Timbral (ENT) Χαρακτηριστικά....87
9.2.3 Αναπαράσταση ENT Χαρακτηριστικών....88
9.2.4 Επιλογή Επιπλέον Χαρακτηριστικών....88
9.3 Γράφος Ανομοιότητας....89
9.4 Data Partitioning....91
9.5 Ταξινομητές....91
9.5.1 NNeural Network....91
9.5.2 SVM....94
9.5.3 SVNN....94
9.6 Αξιολόγηση Αποτελεσμάτων....101
9.7 Visualization....103
9.7.1 MDS Αναπαράσταση....106
9.8 Ερμηνεία και Ανάλυση Πίνακα Συναπτικών Βαρών....115
10. **ΣΥΜΠΕΡΑΣΜΑΤΑ....117**
11. **ΒΙΒΛΙΟΓΡΑΦΙΑ....121**

Ευχαριστίες:

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της εργασίας αυτής, κ. Γ. Σταφυλοπάτη για την ευκαιρία που μου έδωσε για να μελετήσω σε βάθος το θέμα της μουσικής ομοιότητας, καθώς και τον Γ. Σιόλα για τις χρήσιμες συμβουλές του και την πολύτιμη βοήθεια του στην αντιμετώπιση των προβλημάτων που συνάντησα.

Ακόμα θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την υπομονή και την κατανόηση που έδειξαν απέναντι μου κατά την διάρκεια της εκπόνησης της παρούσας εργασίας.

ΚΕΦΑΛΑΙΟ 1:

1.1 Εισαγωγή

Καθώς ο όγκος της διαθέσιμης μουσικής πληροφορίας συνεχώς αυξάνεται, αναδεικνύεται η αναγκαιότητα της χρήσης αυτοματοποιημένων συστημάτων για εφαρμογές διαχείρισης και ανακάλυψης μουσικού περιεχομένου, τα οποία θα μπορούν να προσεγγίζουν την έννοια της μουσικής ομοιότητας με τρόπο παρόμοιο με την ανθρώπινη αντίληψη. Ο ορισμός όμως της μουσικής ομοιότητας είναι ασθενώς ορισμένος και δεν είναι καθολικός, με αποτέλεσμα να διαφαίνεται σχεδόν αδύνατη η υλοποίηση ενός συστήματος που να την αναπαριστά αποδοτικά.

Στην εργασία αυτή μελετώνται οι τρόποι προσέγγισης της μουσικής ομοιότητας με τη χρήση ευφυών συστημάτων και επιχειρείται μια αναπαράσταση αυτής.

1.2 Διάρθρωση της Εργασίας:

Στο δεύτερο κεφάλαιο ξεκινάμε εξετάζοντας τις έννοιες της ομοιότητας και της μουσικής ομοιότητας ειδικότερα, διατυπώνοντας τα προβλήματα ορισμού που είναι συνυφασμένα με τις έννοιες αυτές. Στην συνέχεια παρουσιάζουμε το ερευνητικό πεδίο και την εξέλιξη της Εξόρυξης Μουσικών Πληροφοριών (Music Information Retrieval – MIR) καθώς και τις εφαρμογές του κλάδου αυτού. Τέλος αναφέρονται οι πληροφορίες και τα χαρακτηριστικά, που χρησιμοποιούνται συνήθως σε MIR εφαρμογές και γίνεται μια σύντομη περιγραφή του τρόπου υπολογισμού των Mel Frequency Cepstral Coefficients (MFCCs), που αποτελούν τη βάση των χαρακτηριστικών που χρησιμοποιήθηκαν σε αυτήν την εργασία.

Στο τρίτο κεφάλαιο περιγράφονται οι τρόποι συλλογής δεδομένων μουσικής ομοιότητας και στη συνέχεια αναλύεται ο τρόπος μετατροπής δεδομένων τριαδικής σύγκρισης σε περιορισμούς σχετικών αποστάσεων για τα τρία κομμάτια στα οποία αναφέρεται η τριαδική σύγκριση. Στη συνέχεια περιγράφεται ο τρόπος κατασκευής του γράφου περιορισμών με τη βοήθεια του οποίου θα εξαλειφθούν οι ασυνέπειες που παρουσιάζονται στο σύνολο δεδομένων και τέλος αναλύεται ο τρόπος με τον οποίο μετασχηματίζονται οι περιορισμοί σχετικών αποστάσεων σε δεδομένα εισόδου ενός δυαδικού προβλήματος ταξινόμησης.

Στο τέταρτο κεφάλαιο ύστερα από μια ιστορική αναδρομή της εξέλιξης των Τεχνητών Νευρωνικών δικτύων και μια σύντομη περιγραφή της λειτουργίας του βιολογικού τους προτύπου, περιγράφουμε την λειτουργία του τεχνητού νευρώνα και του perceptron και στη συνέχεια επεκτεινόμαστε στα πολυστρωματικά δίκτυα και αναλύουμε την εκπαίδευση με όπισθεν διάδοση σφάλματος.

Στο πέμπτο κεφάλαιο επικεντρωνόμαστε στη θεωρία που διέπει τις μηχανές Διανυσμάτων Υποστήριξης και αφού εισάγουμε τις έννοιες τις στατιστικής μάθησης και της Ελαχιστοποίησης του δομικού κινδύνου, εξηγούμε τον τρόπο λειτουργίας των γραμμικών και μη γραμμικών ταξινομητών Support Vector Machines (SVM) και κάνουμε αναφορά στη παραλλαγή wSVM που θα χρησιμοποιηθεί στη συνέχεια, η οποία επιτρέπει την απόδοση διαφορετικής βαρύτητας σε κάθε στιγμιότυπο εκπαίδευσης του ταξινομητή.

Στο έκτο κεφάλαιο παρουσιάζουμε τον τρόπο λειτουργίας του SVNN, μιας παραλλαγής του Νευρωνικού Δικτύου που χρησιμοποιήθηκε, η οποία έχει υλοποιηθεί κατά την εκπόνηση της διδακτορικής διατριβής [Ludwig, 2012]. Αναλύεται το κριτήριο Eigenvalue Decay το οποίο χρησιμοποιείται από τον ταξινομητή αυτόν καθώς και ο τρόπος με τον οποίο επιτυγχάνει μεγάλο εύρος ταξινόμησης.

Στο έβδομο κεφάλαιο γίνεται αναφορά στα μέτρα αξιολόγησης που χρησιμοποιούνται σε εφαρμογές μηχανικής μάθησης και ταξινόμησης ειδικότερα και ορίζονται τα μέτρα αυτά που θα χρησιμοποιηθούν στο πρακτικό μέρος της εργασίας.

Στο όγδοο κεφάλαιο γίνεται παρουσίαση της μεθόδου οπτικοποίησης αποστάσεων αντικειμένων Multidimensional Scaling (MDS) και αναλύονται οι διάφορες παραλλαγές της μεθόδου αυτής.

Στο ένατο κεφάλαιο αναλύεται το πειραματικό μέρος της εργασίας. Αναφέρεται ο τρόπος κατασκευής του συνόλου δεδομένων, μέσω εξαγωγής μουσικών χαρακτηριστικών και κατασκευής ενός γράφου ανομοιοτήτων καθώς και οι διάφοροι προβληματισμοί που ανέκυψαν κατά την υλοποίηση. Στη συνέχεια παρουσιάζονται τα αποτελέσματα των ταξινομητών που χρησιμοποιήθηκαν και γίνεται ανάλυση των αποτελεσμάτων τους. Τέλος παρουσιάζονται οι απεικονίσεις που προέκυψαν μέσω της MDS μεθόδου που εφαρμόστηκε στις αποστάσεις που «έμαθε» το Νευρωνικό Δίκτυο και γίνεται μια προσπάθεια ερμηνείας των απεικονίσεων αυτών, καθώς και του ρόλου των μουσικών χαρακτηριστικών που χρησιμοποιήθηκαν.

ΚΕΦΑΛΑΙΟ 2:

2.1 Εισαγωγή

Ποια χαρακτηριστικά καθορίζουν την ομοιότητα δύο αντικείμενων? Η έννοια της ομοιότητας (similarity) αποτελεί ένα από τα βασικότερα πεδία έρευνας της σύγχρονης αντιληπτικής και γνωσιακής επιστήμης καθώς παίζει καθοριστικό ρόλο σε νοητικές διαδικασίες λήψης αποφάσεων, κατηγοριοποίησης και επίλυσης προβλημάτων. Σύμφωνα με την θεωρία του Gestalt ο νόμος της ομοιότητας αποτελεί έναν από τους βασικούς νόμους ομαδοποίησης αντικειμένων σύμφωνα με τον οποίο παρόμοια μεταξύ τους οπτικά στοιχεία ομαδοποιούνται με αποκλεισμό των ανόμοιων.

Εμπειρικές μελέτες καθώς και προσπάθειες υπολογιστικής μοντελοποίησης έχουν δείξει ότι όταν οι άνθρωποι αξιολογούν την ομοιότητα δύο αντικειμένων διεξάγουν μια διαδικασία παρόμοια με αυτή του αναλογικού συμπερασμού. Συγκεκριμένα, τα μέρη του ενός από τα συγκρινόμενα αντικείμενα τοποθετούνται σε ευθυγράμμιση ή αντιστοιχία με τα μέρη του δεύτερου αντικειμένου και αυτές οι αντιστοιχίες αλληλοεπηρεάζονται.

Ένα άλλο χαρακτηριστικό της έννοιας της ομοιότητας που είναι χρήσιμο να ληφθεί υπόψη αποτελεί το γεγονός ότι είναι *προσαρμοστική* και ανάλογη του εκάστοτε σημασιολογικού πλαισίου. Αυτό καταδεικνύει ότι η έννοια της ομοιότητας κατασκευάζεται κάθε φορά για συγκεκριμένους σκοπούς και κάτω από συγκεκριμένες περιστάσεις και για κανέναν λόγο δεν απομνημονεύεται από τον ανθρώπινο νου. Συνεπώς τα αντικείμενα δεν έχουν σταθερές ομοιότητες μεταξύ τους, τις οποίες ο ανθρώπινος εγκέφαλος απλώς ανακαλύπτει, αλλά για κάθε περίσταση ο ανθρώπινος εγκέφαλος ομαδοποιεί ή καλύτερα «ομοιοποιεί» αντικείμενα.

Ο Goodman ([Goodman, 1972]) παρομοιάζει την ομοιότητα με την έννοια της κίνησης, από την άποψη ότι και για τις δύο έννοιες απαιτείται ο ορισμός ενός πλαισίου αναφοράς. Με παρόμοιο τρόπο όπως παρατηρούμε ότι ένα σώμα κινείται *σε σχέση* με κάποιο άλλο, απαιτείται να διευκρινίσουμε *σε σχέση* με ποιες ιδιότητες είναι όμοια δύο αντικείμενα. Χωρίς τον προσδιορισμό αυτόν όλα τα αντικείμενα μπορούν να λογιστούν ως όμοια υπό μια έννοια και αντίστοιχα μπορούν να θεωρηθούν και διαφορετικά από όλα τα άλλα αντικείμενα υπό μια άλλη έννοια.

Για ένα δεδομένο σύνολο σχετικών ιδιοτήτων η ομοιότητα μπορεί να οριστεί σαν «*μερική ταυτότητα*», δηλαδή δύο οντότητες είναι όμοιες αν μοιράζονται κάποιες ιδιότητες (κατηγορήματα), αλλά όχι απαραίτητα όλες. Ζευγάρια οντοτήτων μπορούν να συγκριθούν και

κάποιο ζευγάρι μπορεί να χαρακτηριστεί ως πιο όμοιο από κάποιο άλλο, αν τα μέλη που το απαρτίζουν έχουν περισσότερες κοινές ιδιότητες σε σχέση με τα μέλη του άλλου ζευγαριού.

Η ομοιότητα μεταξύ δύο οντοτήτων μπορεί να υπολογιστεί απλά μετρώντας πόσες από τις ιδιότητες των στοιχείων είναι κοινές (σε περιπτώσεις μη μετρίσιμων ιδιοτήτων). Διαφορετικά, η ομοιότητα μπορεί να οριστεί σαν μια συνάρτηση των διαφορών μεταξύ όλων των ζευγών των ιδιοτήτων που κατέχουν τα δύο αντικείμενα (σε περιπτώσεις μετρήσιμων ιδιοτήτων). Και στις δύο περιπτώσεις θα πρέπει να εισαχθεί μια συνάρτηση στάθμισης για να αποδοθεί η πρέπουσα αντιληπτική σημασία στην κάθε ιδιότητα.

Πρέπει όμως να καθοριστεί πάνω σε ποιες πτυχές ενός συγκεκριμένου προβλήματος εκτιμάται η συνάρτηση στάθμισης αυτή. Ποιες μαθηματικές ιδιότητες όμως σχετίζονται με την αντίληψη της ομοιότητας? Είναι ο βαθμός της ομοιότητας μεταξύ του A και του B ίδιος με τον βαθμό της ομοιότητας του B με τον A? Υπακούει δηλαδή η ομοιότητα στη συμμετρική ιδιότητα? Υπάρχουν μελέτες που υποστηρίζουν τη μη μεταβατικότητα της ομοιότητας καθώς και άλλες που προτείνουν μη συμμετρικά μοντέλα ([Tversky, 1977], [Ashby et al., 1988]).

2.2 Μουσική Ομοιότητα

Η μουσική ομοιότητα έχει μελετηθεί από διάφορούς τομείς έρευνας (μουσικολογία, πειραματική ψυχολογία, γνωσιακή νευροεπιστήμη,computer science) καθένας από τους οποίους την παρατηρεί και την προσεγγίζει από διαφορετική σκοπιά. Από την μουσικολογική έποψη μελετώνται οι γνωσιακές και αντιληπτικές επιδράσεις των διαφόρων μουσικολογικών χαρακτηριστικών. Από την προσέγγιση των ερευνητών που ασχολούνται με το music information retrieval (MIR) υιοθετούνται σύνθετες στατιστικές και αριθμητικές μέθοδοι για την εξαγωγή πληροφοριών από το μουσικό σήμα. Τα πειραματικά αποτελέσματα όμως είναι αρκετά αποσπασματικά και μια θεωρητική διατύπωση καθίσταται πολύ δύσκολη.

Κατά τους Orpen και Huron [Orpen et al., 1992] τα δύο βασικά ζητήματα που αναδύονται κατά τη μελέτη της μουσικής ομοιότητας είναι η ενδογενής *πολυδιαστασιμότητα* της μουσικής που καθιστά την μουσική ομοιότητα μια πολύπλευρη οντότητα, καθώς και το γεγονός ότι δεν υπάρχει κανένας ακριβής ορισμός για τη μέτρηση της μουσικής ομοιότητας μεταξύ δύο μουσικών αποσπασμάτων.

Εξαιτίας της πολυδιαστασιμότητας της, η μουσική ομοιότητα μπορεί να μελετηθεί από πολλές σκοπιές όπως η μελωδική ομοιότητα ([Müllensiefen et al., 2004]) ή ρυθμική ομοιότητα ([Hofmann, 2002]), οι οποίες είναι μετρήσιμες και περιγράφονται από φυσικές παραμέτρους. Η έρευνα όμως γίνεται πιο πολυσύνθετη όταν μελετάται η ομοιότητα από πιο υποκειμενικές κατευθύνσεις όπως αυτή της διάθεσης([Tolos et al., 2005]). Στη περίπτωση αυτή ο ερευνητής θέτει σαν στόχο την αναγνώριση των σχετικών διαστάσεων της μουσικής παράλληλα με τη μέτρηση της σχετικότητας τους. Στη περίπτωση μελέτης μιας υποκειμενικής παραμέτρου η συνήθης τυπική διαδικασία που υιοθετείται είναι η μελέτη της έννοιας της μουσικής ομοιότητας ολιστικά, με μια πειραματική διαδικασία που έχει σαν στόχο την απομόνωση των σχετικών μουσικών διαστάσεων κατά το στάδιο της ανάλυση των αποτελεσμάτων.

Με ποιο όμως τρόπο θα καθορίσουμε το βαθμό της ομοιότητας μεταξύ δύο μουσικών κομματιών? Μπορούμε να θέσουμε υπό σύγκριση τις μουσικές παρτιτούρες των δύο κομματιών και να σχηματίσουμε ένα μοντέλο που να αξιολογεί τις διαφορές μεταξύ τους. Εναλλακτικά θα μπορούσαμε είτε να εξάγουμε μουσικές περιγραφές ή χαρακτηριστικά από το

ηχητικό μουσικό σήμα και να συγκρίνουμε τις στατιστικές τιμές τους , είτε να βασιστούμε στις κρίσεις των αντικειμένων σε ένα ανθρώπινο πείραμα ακουστικής αντίληψης όπου αξιολογείται η μουσική ομοιότητα μεταξύ δύο κομματιών. Σε αυτές τις τρεις προσεγγίσεις μετράται διαφορετικό είδος ομοιότητας: η ομοιότητα της σύνθεσης, όπου δίνεται έμφαση στα μουσικολογικά χαρακτηριστικά της μουσικής, αγνοώντας την εκτέλεση ή τις τεχνικές ηχογράφησης _ η ακουστική ομοιότητα, η οποία επικεντρώνεται στη μαθηματική περιγραφή του ηχητικού σήματος και η αντιληπτική ομοιότητα, που βασίζεται σε ανθρώπινες κρίσεις για την αξιολόγηση της ομοιότητας.

Ένα άλλο σημείο που όπως αναφέρθηκε προηγουμένως χρήζει προσοχής είναι η εξάρτηση της ομοιότητας από το γενικότερο πλαίσιο [Cambouropoulos, 2009]. Σε μια μινιμαλιστική σύνθεση απαρτιζόμενη από λίγες μουσικές φράσεις που αλλάζουν αργά στον χρόνο, μια μικρή μελωδική παραλλαγή, όπως μια αλλαγή στη τονικότητα μιας νότας, γίνεται αντιληπτή σαν σημαντική διαφορά. Ο καθορισμός της ομοιότητας είναι έγκυρος μόνο μέσα σε ένα συγκεκριμένο πλαίσιο. Για αυτό είναι πολύ σημαντικό σε κάθε μελέτη να καθορίζεται ρητώς το συγκεκριμένο πλαίσιο της έρευνας κάτω από το οποίο θα τεθούν υποθέσεις, ερωτήματα, μεθοδολογίες και συμπεράσματα.

Στην εργασία αυτή μελετάται η μουσική ομοιότητα από τη σκοπιά του music information retrieval.

2.3 Music Information Retrieval

Με την εξάπλωση της ευρυζωνικότητας του διαδικτύου η ποσότητα των διαθέσιμων μουσικών κομματιών προς ακρόαση ολοένα αυξάνεται, ενώ σε συνδυασμό με την ευρεία εξάπλωση των ψηφιακών συσκευών (π.χ. mp3 players, smart phones κτλ), ο τρόπος που άνθρωποι ανακαλύπτουν και αλληλεπιδρούν με τη μουσική μεταβάλλεται συνεχώς. Παρατηρούνται προσπάθειες ψηφιοποίησης μουσικής που είναι ηχογραφημένη σε αναλογικό μέσο, είτε από δισκογραφικές εταιρίες ή άλλους πολιτιστικούς οργανισμούς είτε ακόμα και από μουσικόφιλους χρήστες. Επιπλέον η ανάπτυξη των προγραμμάτων ψηφιακής σύνθεσης και μίξης μουσικών κομματιών σε συνδυασμό με την ευρεία χρήση σελίδων κοινωνικής δικτύωσης έχει επιτρέψει σε αρκετούς καλλιτέχνες να δημιουργήσουν και να δημοσιοποιήσουν τα μουσικά τους έργα χωρίς να αντιμετωπίζουν τις διάφορες οικονομικές ή διαδικαστικές δυσκολίες που θα αντιμετώπιζαν μερικά χρόνια νωρίτερα.

Όπως μπορεί να αντιληφθεί κανείς πολύ εύκολα οι χρήστες του διαδικτύου έχουν πλέον πρόσβαση σε έναν τεράστιο όγκο μουσικής πληροφορίας που ολοένα αυξάνεται με ραγδαίους ρυθμούς. Στο παρελθόν, το μουσικό υλικό που ήταν διαθέσιμο σε ένα άτομο, ήταν περιορισμένο στην προσωπική του συλλογή, η οποία συνήθως αποτελούταν από μερικές εκατοντάδες κομμάτια. Σήμερα, online υπηρεσίες όπως το iTunes, το Youtube, το 8tracks και πολλές άλλες προσφέρουν άμεση πρόσβαση σε μια τεράστια συλλογή μουσικών κομματιών με το πάτημα ενός μόλις κουμπιού. Ενώ παλαιότερα οι μόνοι τρόποι να αποκτήσει κάποιος νέα μουσικά ακούσματα περιοριζόταν στην ακρόαση μουσικών ραδιοφωνικών εκπομπών, στην επίσκεψη δισκοπωλείων ή στη συναναστροφή με άτομα με παρόμοιες μουσικές προτιμήσεις, πλέον με την εξάπλωση της χρήσης του διαδικτύου και των παντός είδους σελίδων κοινωνικής δικτύωσης, αυτό μπορεί να γίνει μέσα σε μόλις λίγα λεπτά, με όποιο θετικό ή αρνητικό επακόλουθο μπορεί να ενέχει αυτή η ευκολία. Ο απίστευτος όγκος μουσικής πληροφορίας με

την οποία έρχεται σχεδόν καθημερινά σε επαφή ένας χρήστης του διαδικτύου, είναι πολύ δύσκολά διαχειρίσιμος και ερευνήσιμος και μπορεί εύκολα να οδηγήσει σε υπερπληροφόρηση. Περιέργως όμως αυτή η τεράστια γκάμα μουσικών επιλογών που διαθέτει ένας χρήστης προς ακρόαση έχει οδηγήσει στο φαινόμενο της *«τυραννίας της επιλογής» ("tyranny of choice")*[Schwartz, 2004]: ο τελικός χρήστης λόγω του εύρους των επιλογών τις οποίες διαθέτει, δυσκολεύεται να λάβει την τελική του απόφαση.

Συνεπώς η ανάπτυξη στρατηγικών που διευκολύνουν τη πρόσβαση σε μουσικές συλλογές, τόσο καινούργιου όσο και αρχειακού υλικού, παρουσιάστηκε μέσω του κλάδου του Music Information Retrieval (Ανάκτηση Μουσικών Πληροφοριών). Το MIR αποτελεί μια αναπτυσσόμενη περιοχή έρευνας αφιερωμένη κυρίως στο να καλύψει τις ανάγκες των χρηστών για πληροφόρηση και αλληλεπίδραση με τη μουσική πληροφορία. Οι κύριες ομάδες κοινού που επωφελούνται από τις έρευνες του κλάδου αυτού είναι τρείς: οι επιχειρήσεις που δραστηριοποιούνται στο χώρο της ηχογράφησης, συγκέντρωσης και διάδοσης μουσικής, οι τελικοί χρήστες που θέλουν να ανακαλύψουν καινούργια μουσική, και οι επαγγελματίες (μουσικοί, καθηγητές, μουσικολόγοι, μουσικοί παραγωγοί και δικηγόροι που ασχολούνται με δισκογραφικά δικαιώματα) [Casey et al., 2008].

2.4 Περιοχές Έρευνας του MIR (MIR Research Tasks)

Το MIR ασχολείται με την εξαγωγή, ανάλυση και χρήση πληροφοριών σχετικών με κάθε είδους μουσική οντότητα (κομμάτια, καλλιτέχνες κτλ). Στη συνέχεια ακολουθεί μια παρουσίαση των ερευνητικών περιοχών και εφαρμογών με τις οποίες καταπιάνεται ο κλάδος του MIR.

- Εξαγωγή Μουσικών Χαρακτηριστικών (Feature Extraction):
- Ακουστική Ομοιότητα (Acoustic Similarity)
- Δομική Ανάλυση (Structural Analysis)
- Ανάλυση Μουσικής Τονικότητας (Music Tonality Analysis)
- Audio Fingerprinting
- Μουσική Αναγνώριση (Music Identification)
- Αναζήτηση και Ανακάλυψη Μουσικής (Music Search and Discovery)
- Αυτόματη Σήμανση (Automatic Tagging)
- Παρακολούθηση Δικαιωμάτων (Copyright Monitoring)
- Ταξινόμηση κατά είδος (Genre Classification)
- Αναγνώριση Καλλιτέχνη (Artist Recognition)
- Αναγνώριση Μουσικού Οργάνου (Instrument Identification)

Είναι σημαντικό να αναφερθεί ότι καθώς εξελίσσονται οι ερευνητικές εργασίες στον τομέα του MIR, τα αντικείμενα στα οποία εστιάζουν γίνονται πιο εξειδικευμένα.

Το MIREX (Music Information Retrieval Evaluation Exchange) ([Downie et al., 2005]) αποτελεί μια ετήσια εκδήλωση που οργανώνεται προκειμένου να εξεταστεί η αποτελεσματικότητα των καινούργιων MIR αλγορίθμων που προτείνονται από διάφορους ερευνητές του κλάδου, παρακινώντας έτσι την ερεύνα προκειμένου να βελτιωθεί η ακρίβεια των υπαρχόντων αλγορίθμων. Προκειμένου να αναδείξουμε την εξέλιξη των ερευνών, στον πίνακα

που ακολουθεί παρουσιάζονται οι περιοχές έρευνας που είχαν προταθεί από το MIREX το 2005, τη πρώτη χρονιά που διοργανώθηκε η εκδήλωση αυτή, και το 2013.

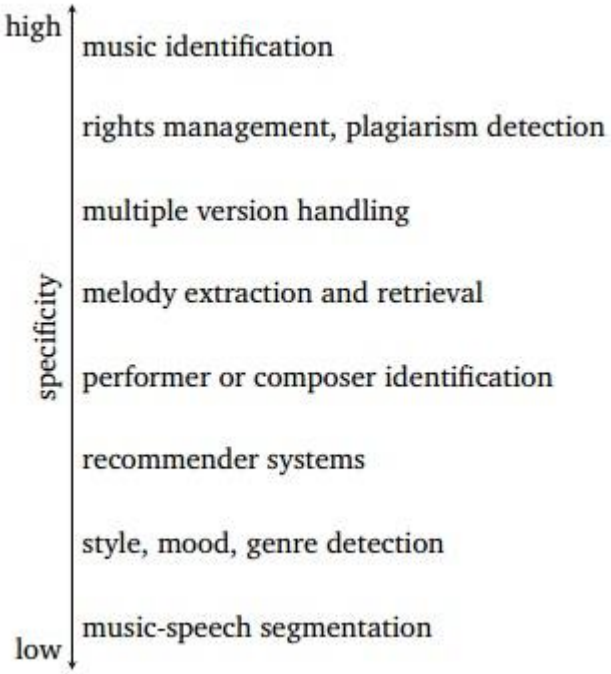
2005	2013
Audio Artist Identification	Audio Cover Song Identification
Audio Drum Detection	Audio Tag Classification
Audio Genre Classification	Audio Music Similarity and Retrieval
Audio Melody Extraction	Symbolic Melodic Similarity
Audio Onset Detection	Audio Onset Detection
Audio Tempo Extraction	Audio Key Detection
Audio and Symbolic Key Finding	Real-time Audio to Score Alignment
Symbolic Genre Classification	Query by Singing/Humming
Symbolic Melodic Similarity	Audio Melody Extraction
	Multiple Fundamental Frequency Estimation & Tracking
	Audio Chord Estimation
	Query by Tapping
	Audio Beat Tracking
	Structural Segmentation
	Audio Tempo Estimation
	Discovery of Repeated Themes & Sections
	Audio US Pop Genre Classification
	Audio Latin Genre Classification
	Audio Music Mood Classification
	Audio Classical Composer Identification

Πίνακας 1 Περιοχές Ενδιαφέροντος του MIREX κατά τα έτη 2005 και 2013[1]

Παρατηρούμε ότι το αντικείμενο ενδιαφέροντος έχει μετατοπιστεί σε σημασιολογικά πιο σύνθετες εφαρμογές σε σχέση με το παρελθόν. (π.χ. η ταξινόμηση των US Pop κομματιών ή των Latin κομματιών, που στοχεύει στην περεταίρω ταξινόμηση των υποειδών).

Ένας διαχωρισμός που μπορεί να γίνει στα MIR συστήματα με βάση τον στόχο που καλούνται να πετύχουν αφορά τον βαθμό *εξειδίκευσης (specificity)* τους ([Casey et al., 2008]), που μπορεί να θεωρηθεί ότι σχετίζεται με την διακριτικότητα του στόχου. Συστήματα με *υψηλή εξειδίκευση* θεωρούνται αυτά που καλούνται να αναγνωρίσουν ένα συγκεκριμένο μουσικό κομμάτι μέσα από μια βάση δεδομένων, δεχόμενα σαν είσοδο μόνο ένα μικρό μουσικό απόσπασμα και σχετίζονται με εφαρμογές μουσικής αναγνώρισης (audio fingerprinting) και διαχείρισης μουσικών δικαιωμάτων. Χαμηλής εξειδίκευσης θεωρούνται τα συστήματα που βασίζονται σε εννοιολογική πληροφορία όπως η ταξινόμηση σε μουσικό είδος ή ύφος. Σε ενδιάμεσες θέσεις εξειδίκευσης τοποθετούνται οι υπόλοιποι στόχοι με βάση το πόσο αυστηρή μπορεί να είναι η οριοθέτηση μεταξύ σωστού ή λάθους αποτελέσματος της ανάκτησης. Ο διαχωρισμός αυτός, αξίζει να σημειωθεί ότι αφορά κυρίως εφαρμογές που διαχειρίζονται πληροφορίες προερχόμενες από το ηχητικό σήμα. Στην επόμενη εικόνα παρουσιάζεται μια απεικόνιση της κλίμακας εξειδίκευσης που μόλις αναφέρθηκε.

[1] Πηγή: [MIREX, 2005] , [MIREX, 2013]



Εικόνα 1 Κατηγοριοποίηση μερικών στόχων του MIR ανάλογα με το βαθμό εξειδίκευσής τους[2]

2.5 Περιγραφείς Μουσικών Κομματιών

Τα περισσότερα MIR συστήματα είτε επιτελούν την ανάλυση ενός μουσικού κομματιού και καθορίζουν μια ετικέτα για αυτό (π.χ. είδος , καλλιτέχνης κτλ.) είτε επιστρέφουν ένα μέτρο ομοιότητας/ανομοιότητας μεταξύ δύο ή περισσοτέρων κομματιών. Και στις δύο περιπτώσεις, τα κομμάτια δεν χρησιμοποιούνται σαν ηχητικό σήμα, αλλά υπόκεινται σε μια εξαγωγή πληροφοριών και χαρακτηριστικών, τα οποία αποτελούν την είσοδο του εκάστοτε συστήματος.

Οι πληροφορίες που χρησιμοποιούνται είναι μεταδεδομένα κειμένου (textual metadata) και πληροφορία μουσικού περιεχομένου (content-based)

2.5.1 Μεταδεδομένα Κειμένου

Οι πληροφορίες μεταδεδομένων κειμένου χρησιμοποιούνταν αρχικά από τα περισσότερα MIR συστήματα, δεδομένου ότι αυτές αποτελούν μια αρκετά πλούσια και εκφραστική περιγραφή για τη μουσική ([Casey et al., 2008]). Πολλές από τις υπηρεσίες που παρέχουν music downloading λειτουργούν αρκετά αποδοτικά χρησιμοποιώντας μόνο πληροφορίες μεταδεδομένων.

Με τον όρο πληροφορίες κειμένου μεταδεδομένων αναφερόμαστε τόσο σε αντικειμενικές πληροφορίες όσο και σε υποκειμενικές πληροφορίες και πληροφορίες συνάφειας (contextual).

[2] Πηγή: [Schwartz, 2010]

- Τα *αντικειμενικά μεταδεδομένα* περιέχουν πληροφορίες όπως ο τίτλος του τραγουδιού, το όνομα του καλλιτέχνη, τη χρονολογία ηχογράφησης, το όνομα του δίσκου.
- Τα *υποκειμενικά μεταδεδομένα* γνωστά και ως tags περιέχουν πληροφορίες σχετικά με το ύφος, το είδος, τα συναισθήματα και τη διάθεση ενός μουσικού κομματιού.
- Οι *πληροφορίες συνάφειας* αφορούν δεδομένα βασισμένα στις προτιμήσεις των χρηστών, όπως ο αριθμός των φορών που αναπαράχθηκαν κάποια κομμάτια, η σειρά εμφάνισης τους σε μια playlist, δεδομένα από αρχεία καταγραφής διαδικτυακών αγορών ή αναζητήσεων καθώς και πληροφορία προερχόμενη από text mining από μουσικά άρθρα, reviews και blogs

Τα περισσότερα σύγχρονα συστήματα χρησιμοποιούν συνδυασμούς των τριών ειδών της προαναφερθείσας πληροφορίας. Για να λειτουργήσουν βέβαια αποδοτικά αυτά τα συστήματα, η πληροφορία που διαχειρίζονται θα πρέπει να είναι ακριβής και το λεξιλόγιο που χρησιμοποιείται καθολικά κατανοητό ([Freed, 2006]). Το Web 2.0, επιτρέποντας σε κοινότητες χρηστών να ψηφίζουν για την ακρίβεια των μεταδεδομένων ενός κομματιού, διασφαλίζει ότι οι περιγραφές αυτές είναι συνεπείς με τη χρήση που επιδέχονται από διάφορες κοινότητες.

Το γεγονός όμως ότι ενδέχεται τέτοιες πληροφορίες κειμένου ορισμένες φορές να μην υπάρχουν, καθιστά αυτές τις μεθόδους αναποτελεσματικές σε κάποιες περιπτώσεις, αφού η χρήση τους απαιτεί γνώση η οποία δεν συνάγεται από απλή ακρόαση και επιπλέον έχουν περιορισμένο πεδίο εφαρμογής, επειδή βασίζονται σε προκαθορισμένους περιγραφείς. Για το λόγο αυτό έχει παρουσιαστεί αρκετό ενδιαφέρον στην ερευνητική κοινότητα για ανεύρεση μεθόδων που θα προσδιορίζουν αυτόματα πληροφορίες τέτοιου είδους ([Bertin-Mahieux et al., 2010]).

.

2.5.2 Πληροφορίες Μουσικού Περιεχομένου

Οι πληροφορίες μουσικού περιεχομένου αποτελούνται από χαρακτηριστικά που έχουν εξαχθεί κατευθείαν από το ηχητικό μουσικό σήμα. Αυτή η διαδικασία εξαγωγής χαρακτηριστικών δείχνει να αποτελεί μια πολύ κοινή ενέργεια για τους ανθρώπους εξαιτίας της τεράστιας υπολογιστικής ικανότητας του ανθρώπινου εγκεφάλου να αξιοποιήσει ένα τεράστιο ποσό σημασιολογικής πληροφορίας για την αναζήτηση ομοιοτήτων και διαφορών μεταξύ ήχων καθώς και για την ομαδοποίηση αυτών των ήχων. Αντιθέτως, αυτή η διαδικασία γίνεται πολύ πιο δύσκολη, όταν πραγματοποιείται από αυτοματοποιημένα υπολογιστικά συστήματα βασιζόμενη μόνο στα χαρακτηριστικά περιεχομένου που εξάγονται, καθώς τα χαρακτηριστικά αυτά έχουν ελάχιστη ή σχεδόν καθόλου σημασιολογική σημασία.

Το πρόβλημα της ανάκτησης μουσικών πληροφοριών βασισμένης στο περιεχόμενο είναι ένα μη καλώς ορισμένο πρόβλημα, διότι εισάγει ένα *σημασιολογικό κενό* ανάμεσα στις έννοιες υψηλού επιπέδου και στις περιγραφές χαμηλού επιπέδου, δηλαδή ανάμεσα στο ηχητικό σήμα και την σημασιολογία των περιεχομένων του. Για παράδειγμα, μια ηχογράφηση της 9ης συμφωνίας του Beethoven αποτελεί μια σειρά από αριθμητικές τιμές (samples) για τον υπολογιστή. Σε ένα υψηλότερο σημασιολογικό επίπεδο η συμφωνία είναι μια σειρά από νότες με συγκεκριμένες διάρκειες. Ένας άνθρωπος μπορεί να εκλάβει τις υψηλού βαθμού έννοιες σαν μουσικές οντότητες (μοτίβα, θέματα) και σαν συναισθήματα (έκσταση, ευφορία).

Οι πληροφορίες μουσικού περιεχομένου διαχωρίζονται σε *υψηλού* και *χαμηλού επιπέδου* μουσικά χαρακτηριστικά (high level and low level features). Τα χαμηλού επιπέδου χαρακτηριστικά σχετίζονται με πληροφορία που εξάγεται κατευθείαν από το ηχητικό σήμα και συνήθως δεν φέρουν καμία σημασιολογική ερμηνεία για τον τελικό χρήστη. Χρησιμοποιούνται ως περιγραφείς του σήματος για να συνθέσουν τα χαρακτηριστικά υψηλού επιπέδου (όπως ο τόνος, η μελωδία, το ηχόχρωμα).

Στη συνέχεια παρατίθενται μερικά από τα χαρακτηριστικά περιεχομένου που χρησιμοποιούνται συνήθως στις MIR σύμφωνα με την ταξινόμηση των [Mitrovic et al., 2010].

➢ Χρονικά χαρακτηριστικά (Temporal Features)

Ο χρόνος είναι το ενδογενές πεδίο ορισμού των ηχητικών σημάτων. Όλα τα χρονικά χαρακτηριστικά έχουν σαν κοινό ότι εξάγονται κατευθείαν από το ηχητικό σήμα, χωρίς να χρειαστεί να προηγηθεί κάποιος μετασχηματισμός.

- **Χαρακτηριστικά Βασισμένα σε Μηδενισμούς (Zero Crossing-Based Features):**
  - Ρυθμός Μηδενισμού (Zero Crossing Rate- ZCR)

    Ορίζεται ως ο αριθμός των μηδενισμών που συμβαίνουν στο πεδίο του χρόνου ανά δευτερόλεπτο. Αποτελεί μέτρο της κύριας συχνότητας του σήματος.

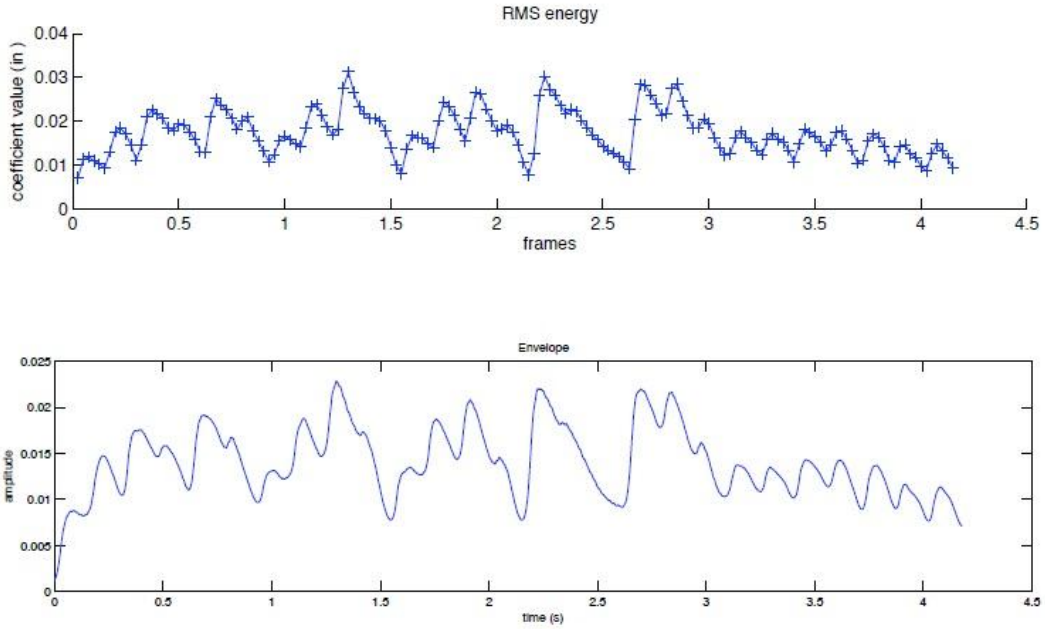
- **Χαρακτηριστικά Βασισμένα σε Ενέργεια (Power-Based Features):**

  Η ενέργεια του σήματος είναι το τετράγωνο του πλάτους της κυματομορφής. Η ενέργεια ενός ήχου είναι η ενέργεια που μεταφέρεται ανά δευτερόλεπτο, συνεπώς είναι το μέσο τετράγωνο του σήματος.
  - Ενέργεια Βραχέως Χρόνου (Short-time Energy – STE or Root-Mean Square Energy - RMS)

    Περιγράφει την περιβάλλουσα ενός σήματος και χρησιμοποιείται ευρέως σε διάφορους τομείς του MIR. Ορίζεται ως η μέση ενέργεια ανά πλαίσιο.

$$x_{STE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

Υπάρχουν βέβαια και άλλοι ορισμοί της STE που λαμβάνουν υπόψη τους την φασματική ενέργεια. Στο σχήμα παρατηρούμε ότι η καμπύλη της ενέργειας προσεγγίζει την περιβάλλουσα του σήματος.



Εικόνα 2 Χρονική εξέλιξη της Ενέργειας του σήματος (πάνω) και της περιβάλλουσας αυτού (κάτω).[3]

- Χρονικό Κέντρο Βάρους (Temporal Centroid)

  Το χρονικό κέντρο βάρους είναι η μέση τιμή στον χρόνο της περιβάλλουσας του σήματος. Είναι το σημείο στο χρόνο όπου εντοπίζεται η περισσότερη ενέργεια του σήματος.

- Λογαριθμικός Χρόνος Attack (Log Attack Time - LAT)

  Χαρακτηρίζει το attack ενός ήχου. Είναι ο λογάριθμος του χρόνου από το ξεκίνημα ενός ηχητικού σήματος μέχρι το σημείο του χρόνου όπου το πλάτος φτάνει σε στο πρώτο τοπικό μέγιστο. Χρησιμοποιείται για κατάταξη μουσικών οργάνων ([Zhang et al., 2006])

➢ Χαρακτηριστικά Φυσικής Συχνότητας(Physical Frequency Features):

Τα ηχητικά χαρακτηριστικά αυτά ορίζονται στο πεδίο της συχνότητας ή της αυτοσυσχέτισης. Υπάρχουν διάφοροι τρόποι για να καταλήξουμε σε αυτά τα πεδία. Ο ποιο συνηθισμένος είναι η εφαρμογή μετασχηματισμού Fourier. Άλλοι συνηθισμένοι τρόποι είναι ο μετασχηματισμός Συνημιτόνου, ο μετασχηματισμός Κυματιδίων και ο constant Q μετασχηματισμός.

- **Χαρακτηριστικά Βασισμένα στον Μετασχηματισμό Fourier Βραχέως Χρόνου (Short – Time Fourier Transform-Based Features):**

  Ο μετασχηματισμός STFT οδηγεί σε πραγματικές και μιγαδικές τιμές. Οι πραγματικές τιμές αντιπροσωπεύουν την κατανομή των συνιστωσών της συχνότητας (φασματική περιβάλλουσα), ενώ οι μιγαδικές τιμές περιέχουν πληροφορία για την φάση των συνιστωσών.

[3] Πηγή: [Lartillot, 2011]

- Φασματική Ροή (Spectral Flux - SF)

  Αποτελεί μια L2 – νόρμα του διανύσματος της διαφοράς του φασματικού πλάτους μεταξύ πλαισίων. Ποσοστικοποιεί χοντρικά αλλαγές στο σχήμα του φάσματος στο χρόνο. Ορίζεται ως η μέση διαφορά μεταξύ δύο συνεχόμενων πλαισίων μετασχηματισμού Fourier βραχέως χρόνου:

$$SF(n) = \frac{\sqrt{\sum_{k=0}^{K/2-1}(|X(k,n)| - |X(k,n-1)|)^2}}{K/2}$$

  Σταθερά σήματα ή αυτά που παρουσιάζουν αργή μεταβολή φασματικών ιδιοτήτων (πχ. θόρυβος) έχουν χαμηλή φασματική ροή.

  Μπορεί να θεωρηθεί σαν μια υποτυπώδης προσέγγιση της τραχύτητας του ήχου.

- Φασματική Κλίση (Spectral Slope)

  Αποτελεί προσέγγιση του φασματικού σχήματος με γραμμική παλινδρόμηση. Αντιπροσωπεύει την μείωση των φασματικών πλατών από τις χαμηλές στις υψηλές συχνότητες.

- Φασματικές Κορυφές (Spectral Peaks)
- Φασματική Εξάπλωση (Spectral Spread)
- Φασματική Ασυμμετρία (Spectral Skewness)
- Φασματική Κύρτωση (Spectral Kurtosis)

➢ Χαρακτηριστικά Αντιληπτικών Συχνοτήτων(Perceptual Frequency Features)

Στην ενότητα αυτή αναλύονται χαρακτηριστικά που έχουν σημασιολογική σημασία στα πλαίσια της ανθρώπινης ακουστικής αντίληψης και ομαδοποιούνται ανάλογα με την ακουστική ιδιότητα που περιγράφουν.

- **Φωτεινότητα (Brightness):**

  Η Φωτεινότητα χαρακτηρίζει τη φασματική κατανομή των συχνοτήτων και περιγράφει εάν σε ένα σήμα κυριαρχούν οι υψηλές ή οι χαμηλές συχνότητες. Ένας ήχος γίνεται πιο φωτεινός καθώς το περιεχόμενο των υψηλών συχνοτήτων κυριαρχεί. Είναι στενά συνδεδεμένη με την αίσθηση της οξύτητας του ήχου.

  - Φασματικό Κεντροειδές (Spectral Centroid - SC)

    Ορίζεται ως το κέντρο βάρους του φασματικού πλάτους. Δείχνει σε ποιο σημείο του φάσματος είναι συγκεντρωμένη περισσότερη ενέργεια και σχετίζεται με την επικρατούσα συχνότητα του σήματος.

$$SC(n) = \frac{\sum_{k=0}^{K/2-1} k \cdot |X(k,n)|^2}{\sum_{k=0}^{K/2-1} |X(k,n)|^2}$$

    Χαμηλές τιμές υποδηλώνουν σημαντικές χαμηλόσυχνες συνιστώσες και αμελητέες υψηλόσυχνες, δηλαδή χαμηλή φωτεινότητα.

  - Οξύτητα (Sharpness)

    Αποτελεί μια διάσταση του ηχοχρώματος που επηρεάζεται από τη κεντρική συχνότητα σημάτων στενής ζώνης και αυξάνει καθώς αυξάνει η ισχύς των υψηλών

συχνοτήτων στο φάσμα. Χρησιμοποιείται σε εφαρμογές μουσικής ομοιότητας (π.χ. [Herre et al., 2003])

- Φασματικό Κέντρο (Spectral Center)

  Ορίζεται ως η συχνότητα εκείνη όπου η μισή ενέργεια του φάσματος βρίσκεται κάτω από αυτήν και η άλλη μισή πάνω από αυτή. Αποτελεί μια προσέγγιση της κατανομής της ενέργειας και χρησιμοποιείται σε εφαρμογές παρακολούθησης ρυθμού.

- **Τονικότητα (Tonality):**

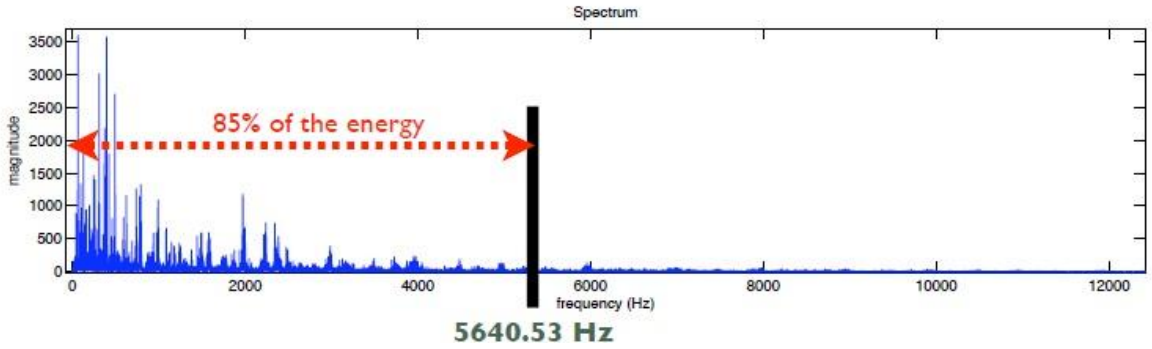
  Η τονικότητα είναι αυτή η ιδιότητα του ήχου που διαχωρίζει τους θορυβώδεις ήχους από τους τονικούς. Οι θορυβώδεις ήχοι έχουν ένα συνεχόμενο φάσμα και συνεπώς ελάχιστη τονικότητα, σε αντίθεση με τους τονικούς ήχους που έχουν γραμμικά φάσματα (line spectra) και υψηλή τονικότητα. Η τονικότητα σχετίζεται με την ένταση του τόνου (pitch strength),η οποία περιγράφει την ένταση του αντιληπτού ήχου.

  - Φασματική Διασπορά (Spectral Dispersion)

    Αποτελεί μέτρο της εξάπλωσης του φάσματος γύρω από το φασματικό κέντρο.

  - Σημείο Φασματικού Roll off (Spectral Rolloff Point)

    Προσδιορίζεται ως η συχνότητα κάτω από την οποία περιέχεται το 85% (ή 95%) της συνολικής ενέργειας. Η τιμή της συχνότητας αυτής αυξάνει αν αυξήσουμε το εύρος ζώνης του σήματος. Χρησιμοποιείται αρκετά συχνά σε MIR εφαρμογές όπως [Mckay, 2005].



Εικόνα 3I Σημείο Φασματικού Roll off[4]

- Φασματική Ομαλότητα (Spectral Flatness)

  Υπολογίζει σε ποιο βαθμό οι συχνότητες του φάσματος είναι κανονικά κατανεμημένες. Χρησιμοποιείται συχνά σε εφαρμογές audio fingerprinting.([Herre et al., 2001])

- Παράγοντας Φασματικής Κορυφογραμμής (Spectral Crest Factor)

  Αποτελεί μέτρο της μη ομαλότητας του φάσματος και χρησιμοποιείται για τη διαφοροποίηση θορυβοδών και τονικών ήχων. Ορίζεται σαν την αναλογία της μέγιστης φασματικής ενέργειας και της μέσης ενέργειας φάσματος μιας υποζώνης. Χρησιμοποιείται αρκετά συχνά σε εφαρμογές fingerprinting. ([Herre et al., 2001])

[4] Πηγή: [Lartillot, 2011]

- **Ένταση (Loudness):**

  Η ένταση είναι αυτό το χαρακτηριστικό της ακουστικής αίσθησης στα πλαίσια του οποίου κατατάσσονται οι ήχοι στη κλίμακα απαλοί – δυνατοί.

  - Αίσθηση Συγκεκριμένης Έντασης (Specific Loudness Sensation - Sone)

    Υπολογίζεται πρώτα το φασματογράφημα στη κλίμακα Bark και στη συνέχεια εφαρμόζεται φασματική συγκάλυψη (masking) και «ισο-εντασιακά» περιγράμματα (εκφρασμένα σε phon), για να μετασχηματιστεί το φάσμα στη συνέχεια αίσθηση συγκεκριμένης έντασης σε sone. Το χαρακτηριστικό αυτό χρησιμοποιείται για ανίχνευση ρυθμικών προτύπων.
  - Ενιαία Ένταση (Integral Loudness)

- **Τόνος (Pitch):**

  Ο τόνος αποτελεί βασική διάσταση του ήχου μαζί με την ένταση, τη διάρκεια και το ηχόχρωμα. Η ακουστική αίσθηση του τόνου ορίζεται ως το χαρακτηριστικό εκείνο στα πλαίσια του οποίου οι ήχοι κατατάσσονται στη κλίμακα χαμηλός/βαθύς – υψηλός/οξύς.

  - Βασική Συχνότητα (Fundamental Frequency)

    Ορίζεται σαν τη χαμηλότερη συχνότητα της αρμονικής σειράς και αποτελεί μια χοντρική προσέγγιση του ψυχοακουστικού τόνου. Χρησιμοποιείται σε εφαρμογές κατάταξης είδους ([])
  - Ιστόγραμμα Τόνου(Pitch Histogram)

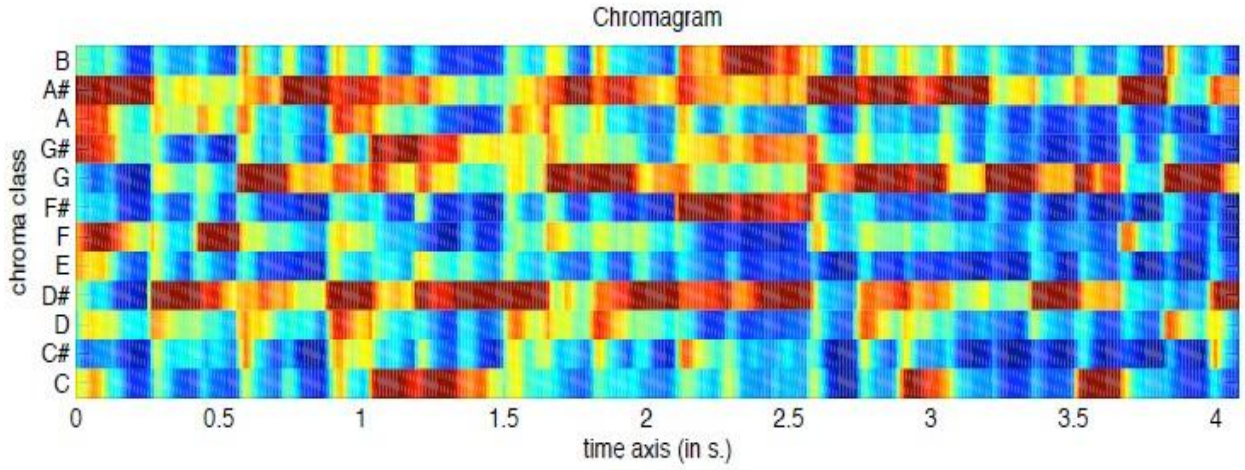
    Περιγράφει το περιεχόμενο τονικότητας του σήματος με αρκετά συμπαγή τρόπο και χρησιμοποιείται συχνά στην κατηγοριοποίηση μουσικού είδους. Στη μουσική ανάλυση ο τόνος συχνά αντιστοιχεί στις μουσικές νότες, έτσι το pitch histogram προσφέρει μια αναπαράσταση της κατανομής των μουσικών νοτών σε ένα μουσικό κομμάτι.

- **Χρώμα / Χροιά (Chroma):**

  Σύμφωνα με τον Shepard ([Shepard, 1964]) η αίσθηση του μουσικού τόνου μπορεί να χαρακτηριστεί από δύο διαστάσεις: *το ύψος του τόνου* και το *χρώμα*. Η διάσταση του ύψους του τόνου διαιρείται στις μουσικές οκτάβες. Οι τόνοι (μουσικές νότες) της ίδιας τονικής κλάσης κατέχουν το ίδιο χρώμα και παράγουν παρόμοια ηχητική αίσθηση.

  - Χρωματογράφημα (Chromagram)

    Αποτελεί ουσιαστικά ένα φασματογράφημα που παρουσιάζει την φασματική ενέργεια για κάθε μια από τις 12 τονικές κλάσεις. Βασίζεται σε λογαριθμικό φάσμα Fourier βραχέως χρόνου. Οι συχνότητες αντιστοιχίζονται (κβαντίζονται) στις 12 τονικές κλάσεις μέσω μιας συσσωρευτικής συνάρτησης. Αντιστοιχίζει ουσιαστικά όλες τις συχνότητες σε μία οκτάβα, το οποίο οδηγεί σε συμπύκνωση του φάσματος, επιτρέποντας έτσι μια πιο συμπαγή περιγραφή των αρμονικών σημάτων.



Εικόνα 4 Απεικόνιση Χρωματογραφήματος[5]

- Τονικό Κεντροειδές (Tonal Centroid)

  Υπολογίζεται το 6-διάστατο διάστημα τονικού κεντροειδούς από το χρωματογράφημα, το οποίο αντιστοιχεί στην προβολή των συγχορδιών στους κύκλους Πέμπτης, Μεγάλης Τρίτης και Μικρής Τρίτης

- Προφίλ Τονικότητας (Pitch Profile)

  Αποτελεί μια αρκετά ακριβή αναπαράσταση του περιεχομένου που σχετίζεται με το ύψος του τόνου καθώς λαμβάνει υπόψη τον κακό συντονισμό του ύψους που εισάγεται από μη-συντονισμένα μουσικά όργανα και είναι αρκετά εύρωστο απέναντι σε θορυβώδεις ήχους κρουστών.

- **Αρμονικότητα (Harmonicity):**

  Η αρμονικότητα αποτελεί μια ιδιότητα που διαφοροποιεί τα περιοδικά σήματα (αρμονικοί ήχοι) από τα μη περιοδικά (μη αρμονικοί και θορυβώδεις ήχοι). Οι αρμονικές είναι συχνότητες, ακέραια πολλαπλάσια της θεμελιώδους συχνότητας και συνεπώς στο αρμονικό φάσμα εμφανίζονται κορυφές στη θεμελιώδη συχνότητα και στα ακέραια πολλαπλάσια της. Τα χαρακτηριστικά αρμονικότητας μπορούν να χρησιμοποιηθούν στην αναγνώριση μουσικών οργάνων (π.χ. τα έγχορδα όργανα παρουσιάζουν πιο αρμονική δομή σε σχέση με τα κρουστά)

  - Αρμονικός Λόγος (Harmonic Ratio)

    Είναι ο λόγος της ενέργειας της θεμελιώδους συχνότητας προς την συνολική ενέργεια του πλαισίου. Είναι ένα μέτρο του βαθμού της αρμονικότητας ενός κομματιού.

  - Μέτρα Μη- Αρμονικότητας (Inharmonicity Measures)
  - Περιγραφείς Φασματικού Ηχοχρώματος (Spectral Timbral Descriptors)

    Σχετίζονται με την αρμονική δομή του ήχου και βασίζονται σε μια εκτίμηση της θεμελιώδους συχνότητας και την ανίχνευση αρμονικών κορυφών στο φάσμα. Αποτελούν στατιστικές ροπές (moments) των αρμονικών συχνοτήτων και του πλάτους αυτών:

[5] Πηγή: [Lartillot, 2011]

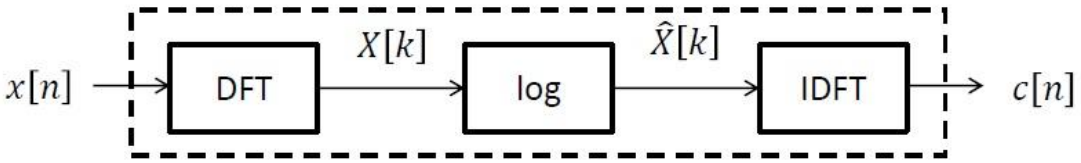
- Harmonic Spectral Centroid - HSC
- Harmonic Spectral Deviation - HSD
- Harmonic Spectral Spread – HSS
- Harmonic Spectral Variation – HSV

➢ Cepstral Χαρακτηριστικά (Cepstral Features)

Η έννοια του cepstrum εισήχθη αρχικά από τους [Bogert et al., 1963] για εφαρμογή στην ανίχνευση ηχούς σε σεισμικά σήματα. Τα cepstral χαρακτηριστικά χρησιμοποιούνται συχνά σαν συχνοτικά εξομαλυμένες αναπαραστάσεις του λογαριθμικού πλάτους του φάσματος και αναδεικνύουν χαρακτηριστικά ηχοχρώματος και τόνου. Επιτρέπουν εφαρμογή Ευκλείδειου μέτρου σαν μέτρο απόστασης λόγω της ορθογώνιας βάσης τους και έτσι διευκολύνουν συγκρίσεις ομοιότητας.

- **Χαρακτηριστικά Βασισμένα σε Συστοιχίες Αντιληπτικών Φίλτρων (Perceptual Filter Bank – Based Features):**

  Το cepstrum ορίζεται ως ο Fourier Μετασχηματισμός του λογαρίθμου του πλάτους του φάσματος του αρχικού σήματος.



Εικόνα 5 Υπολογισμός του Cepstrum

- Συντελεστές Cepstral Συχνότητας Μελ (Mel – Frequency Cepstral Coefficients - MFCCs)

  Εκφράζουν πληροφορία ηχοχρώματος (φασματική περιβάλλουσα) ενός σήματος. Ο υπολογισμός τους αναλύεται στην επόμενη παράγραφο.

- Επεκτάσεις των MFCCs (Extensions of MFCCs)

  Για εφαρμογές μουσικής ομοιότητας συνήθως τα MFCC συνοδεύονται από διανύσματα που εκφράζουν την πρώτη ή και τη δεύτερη στιγμιαία παράγωγο, τα delta-MFCC και τα delta-delta-MFCC ([Wang et al., 2011]).

➢ Χαρακτηριστικά Διαμόρφωσης Συχνότητας (Modulation Frequency Features)

Τα χαρακτηριστικά διαμόρφωσης συχνότητας χαρακτηρίζουν τη πληροφορία διαμόρφωσης χαμηλών συχνοτήτων σε ηχητικά σήματα. Ένα διαμορφωμένο σήμα περιέχει τουλάχιστον δύο συχνότητες: την υψηλή συχνότητα του φέροντος και την, συγκριτικά με το φέρον, χαμηλότερη συχνότητα διαμόρφωσης. Τα διαμορφωμένα σήματα προκαλούν διαφορετικές ηχητικές εντυπώσεις στο ανθρώπινο ακουστικό σύστημα. Οι χαμηλές συχνότητες διαμόρφωσης (μέχρι 20 Hz) προκαλούν την αίσθηση της διακύμανσης

(fluctuation strength). Οι υψηλότερες συχνότητες διαμόρφωσης δημιουργούν την αίσθηση της τραχύτητας του ήχου.

Ο ρυθμός και το τέμπο αποτελούν χαρακτηριστικά του ήχου που συνδέονται στενά με την μακρόχρονη διαμόρφωση. Οι ρυθμικές δομές μπορούν να αποκαλυφθούν αναλύοντας διαμορφώσεις χαμηλών συχνοτήτων στο χρόνο.

- **Ρυθμός (Rhythm):**

  Ο ρυθμός αποτελεί μια ιδιότητα του ηχητικού σήματος που υποδηλώνει ένα μοτίβο αλλαγής του ηχοχρώματος και της ενέργειας στον χρόνο. Σύμφωνα με τους Zwicker και Fastl, η ακουστική αίσθηση του ρυθμού εξαρτάται από τη χρονική μεταβολή της έντασης. Τα ρυθμικά μοτίβα ανευρίσκονται συνήθως αναλύοντας τα πλάτη διαμόρφωσης χαμηλών συχνοτήτων.

  - Μετρική Παλμού (Pulse Metric)

    Μέτρο της ρυθμικότητας του ήχου. Η μετρική παλμού έχει υψηλή τιμή όταν οι αυτοσυσχετίσεις σε όλες τις υποζώνες εμφανίζουν κορυφές σε κοντινές θέσεις, δηλαδή υπάρχει ισχυρή ρυθμική δομή στον ήχο.
  - Περιοδικότητα Ζώνης (Band Periodicity)

    Η περιοδικότητα ζώνης υπολογίζει επίσης την ύπαρξη ρυθμικής δομής στο σήμα.
  - Φάσμα Beat (Beat Spectrum)

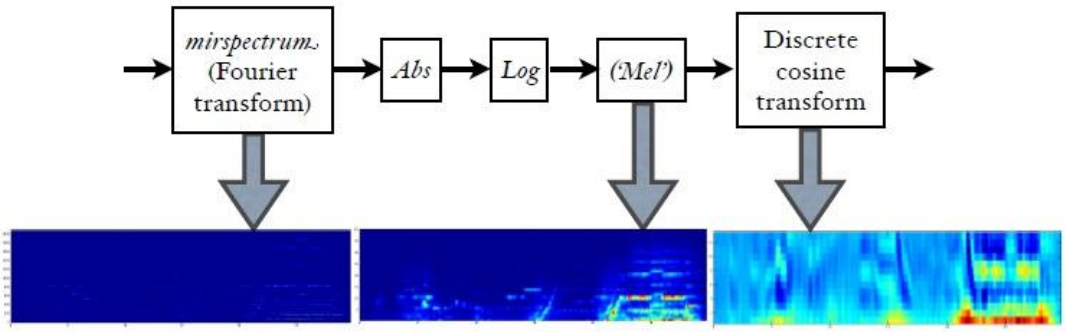
    Υποδεικνύει την αυτοομοιότητα ενός σήματος συναρτήσει διαφορετικών χρονικών καθυστερήσεων. Οι κορυφές στο φάσμα beat καταδεικνύουν δυνατούς παλμούς με συγκεκριμένο ρυθμό επανάληψης. Χρησιμοποιείται συνήθως στην onset detection και στον προσδιορισμό ρυθμικώς όμοιας μουσικής ή ακόμα και στην κατάτμηση ενός κομματιού στα διαφορετικά ρυθμικά κομμάτια που το απαρτίζουν.
  - Κυκλικό Beat Φάσμα (Cyclic Beat Spectrum)
  - Ιστόγραμμα Παλμού (Beat Histogram)

    Αποτελεί μια συμπαγή αναπαράσταση του ρυθμικού περιεχομένου ενός ηχητικού κομματιού, περιγράφοντας τον βαθμό επανάληψης του κυρίου παλμού αλλά και των υπο-παλμών μαζί με την ισχύ τους. Χρησιμοποιείται κυρίως για κατηγοριοποίηση μουσικού είδους.

2.5.2.1 Υπολογισμός MFCC

Τα Mel Frequency Cepstral Coefficients (MFCC) προτάθηκαν σαν μια συμπαγής, βασισμένη στο ανθρώπινο αντιληπτικό μοντέλο αναπαράσταση των πλαισίων φωνής, αλλά έχουν αποδειχτεί ως αρκετά χρήσιμα και σε εφαρμογές εξόρυξης μουσικής πληροφορίας.
Περιγράφουν το φασματικό περιεχόμενο ενός ακουστικού σήματος βραχέως χρόνου (20-30 ms) χρησιμοποιώντας τον Διακριτό Μετασχηματισμό Συνημιτόνου (Discrete Cosine Transform) προκειμένου να αποσυσχετίσουν τα bins ενός φασματικού ιστογράμματος Μελ συχνότητας. Ο υπολογισμός των MFCC περιλαμβάνει την μετατροπή των συντελεστών

Fourier στη Mel[6] κλίμακα, η οποία προσεγγίζει καλύτερα την απόκριση του ανθρώπινου ακουστικού συστήματος σε σχέση με τις ζώνες γραμμικών κατανεμημένων συχνοτήτων.



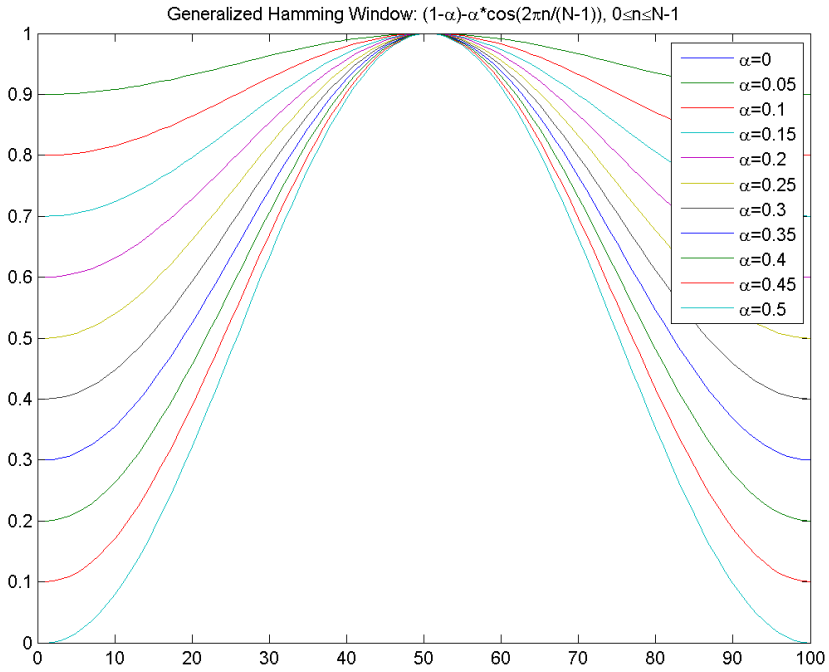
Εικόνα 6 Σχηματική Απεικόνιση του υπολογισμού των συντελεστών MFCC

Τα βήματα για τον υπολογισμό των MFCC είναι τα ακόλουθα ([Rabiner et al., 1993]):

1]. Συνήθως εφαρμόζεται προέμφαση αρχικά στο ηχητικό σήμα, προκειμένου να προσεγγιστεί χοντρικά το φιλτράρισμα που λαμβάνει χώρα στη περιοχή του έξω ότως στο ανθρώπινο ακουστικό σύστημα.

2]. Το ηχητικό σήμα μετατρέπεται από το πεδίο του χρόνου στο πεδίο της συχνότητας. Για το σκοπό αυτό το σήμα τμηματοποιείται σε μικρά αλληλεπικαλυπτόμενα πλαίσια μερικών ms (συνήθως 20-30 ms με 50-75% επικάλυψη) και εφαρμόζεται σε αυτά μια συνάρτηση παραθύρωσης (π.χ. παράθυρο Hamming), προκειμένου να διατηρηθεί η συνέχεια του πρώτου και του τελευταίου σημείου του πλαισίου, δεδομένου ότι για να εφαρμόσουμε τον Fourier Μετασχηματισμό υποθέτουμε ότι το σήμα μέσα σε κάθε πλαίσιο είναι περιοδικό και συνεχές αν αυτό αναδιπλωθεί.

[6] Η μονάδα Mel αποτελεί μονάδα μέτρησης του αντιληπτικού τονικού ύψους ενός σήματος ή τη συχνότητα ενός απλού τόνου. Η σχέση της κλίμακας Mel με τη γραμμική συχνότητα δίνεται από τη σχέση $f_{Mel} = 2595 \cdot log_{10}\left(1 + \frac{f}{700}\right)$



Εικόνα 7 Παράθυρα Hamming για διαφορετικές τιμές της παραμέτρου α

Αν συμβολίσουμε με $x(n), n = 0,1,2, \dots, N$ το σήμα σε κάθε πλάισιο, τότε το αποτέλεσμα της εφαρμογής του παραθύρου Hamming δίνεται από τη σχέση :

$$\overline{x(n)} = x(n) \cdot w(n) \qquad (2.1)$$

,όπου $w(n)$ το παράθυρο Hamming, που ορίζεται ως:

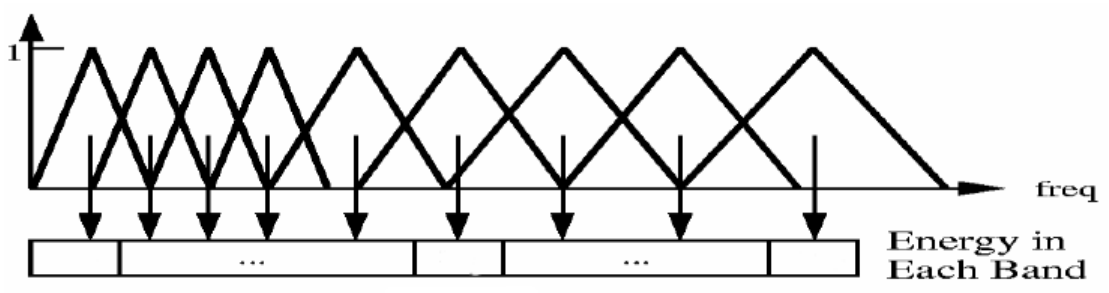
$$\mathrm{w(n,a)} = (1 - \mathrm{a}) - \mathrm{a}\cos(2\pi \mathrm{n}/(\mathrm{N}-1)), \quad 0 \leqq \mathrm{n} \leqq \mathrm{N}-1 \qquad (2.2)$$

Διαφορετικές τιμές του *a* αντιστοιχούν σε διαφορετικές καμπύλες για του παραθύρου όπως φαίνεται καλύτερα και στο προηγούμενο σχήμα.

Στη συνέχεια για να υπολογίσουμε το φάσμα ισχύος (power spectrum), εφαρμόζουμε σε κάθε πλαίσιο Διακριτό Μετασχηματισμό Fourier (DFT) και παίρνουμε το μέτρο $|X(k)|$ αυτού.

$$X(k) = \sum_{n=0}^{N-1} w(n)x(n)\mathrm{e}^{\frac{-\mathrm{j}2\pi kn}{\mathrm{N}}} \qquad (2.3)$$

,για $k = 0,1, \dots, N-1$, όπου το k αντιστοιχεί στη συχνότητα $f(k) = k\frac{f_s}{N}$, και f_sη συχνότητα δειγματοληψίας σε Hz.



Εικόνα 8 Απεικόνιση της συστοιχίας φίλτρων που χρησιμοποιείται κατά τον υπολογισμό των MFCC

3]. Το φάσμα ισχύος απεικονίζεται στην κλίμακα Mel χρησιμοποιώντας μια συστοιχία φίλτρων (filterbanks) $H(k,m)$ αποτελούμενη από $M \ll N$ (συνήθως 20~60) τριγωνικά ζωνοπερατά φίλτρα. Η κλίμακα Μελ είναι προσεγγιστικά γραμμική για χαμηλές συχνότητες (κάτω από 500Hz) και λογαριθμική για τις πιο υψηλές. Κάθε τριγωνικό φίλτρο υπολογίζει το φάσμα μιας ζώνης συχνοτήτων. Γειτονικά τριγωνικά φίλτρα στη συστοιχία παρουσιάζουν επικάλυψη, έτσι ώστε η κεντρική συχνότητα $f_c(m)$ κάθε φίλτρου να αποτελεί την χαμηλότερη συχνότητα του επόμενου φίλτρου και την υψηλότερη του προηγούμενού του. Τα φίλτρα δίνονται από την σχέση:

$$H(k,m) = \begin{cases} 0 & ,για\ f(k) < f_c(m-1) \\ \dfrac{f(k) - f_c(m-1)}{f_c(m) - f_c(m-1)} & ,για\ f_c(m-1) \le f(k) < f_c(m) \\ \dfrac{f(k) - f_c(m+1)}{f_c(m) - f_c(m+1)} & ,για\ f_c(m) \le f(k) < f_c(m+1) \\ 0 & ,για\ f(k) \ge f_c(m+1) \end{cases} \quad (2.4)$$

Οι κεντρικές συχνότητες των φίλτρων στη Mel κλίμακα δίνονται από τη σχέση

$$\varphi_c(m) = m \cdot \Delta\varphi \quad ,για\ m = 1,2,\dots,M \quad (2.5)$$

Η ανάλυση Δφ στη κλίμακα Mel υπολογίζεται χρησιμοποιώντας την σχέση

$$\Delta\varphi = \frac{\varphi_{max} - \varphi_{min}}{M+1} \quad (2.6)$$

,όπου φ_{max} και φ_{min} είναι η υψηλότερη και η χαμηλότερη συχνότητα του φίλτρου στην κλίμακα Mel αντίστοιχα , υπολογισμένες από τη σχέση (2.5) για f_{max} και f_{min} και M ο αριθμός της συστοιχίας φίλτρων.

Για τον υπολογισμό των κεντρικών συχνοτήτων σε Hertz που θα χρησιμοποιηθούν στην εξίσωση (2.4) έχουμε την ακόλουθη εξίσωση:

$$f_c(m) = 700 \cdot \left(10^{\varphi_c(m)/2595} - 1\right) \quad (2.7)$$

4]. Στη συνέχεια υπολογίζεται ο λογάριθμος των ενεργειών των N φίλτρων σε κάθε μία από τις Mel συχνότητες, προκειμένου να προσεγγιστεί η μη γραμμική αντιληπτικότητα της έντασης που παρουσιάζεται στο ανθρώπινο σύστημα ακοής.

$$X'(m) = ln\left(\sum_{k=0}^{N-1} |X(k)| \cdot H(k,m)\right) \tag{2.8}$$

5]. Τέλος εφαρμόζεται ο ανάστροφος Διακριτός Μετασχηματισμός Συνημιτόνου (DCT) με σκοπό να συμπιεστεί το Mel φάσμα ισχύος $X'(m)$. Οι Mel ζώνες συχνοτήτων αναπαρίστανται από μία σειρά συντελεστών (συνήθως μεταξύ 13 και 20). Με τη συμπίεση αυτή το φάσμα ομαλοποιείται κατά μήκος του άξονα των συχνοτήτων, αποτελώντας έτσι μια απλουστευμένη προσέγγιση της φασματικής συγκάλυψης (spectral masking) που λαμβάνει χώρα στο ανθρώπινο ακουστικό σύστημα.
Τα MFCCs υπολογίζονται τελικά από την σχέση:

$$c(l) = \sum_{m=1}^{M} X'(m) \cos\left(l\frac{\pi}{M}\left(m - \frac{1}{2}\right)\right) \tag{2.9}$$

,για $l = 1,2,\dots,M$,όπου $c(l)$ ο l-στος συντελεστής MFCC.

Αναπαριστούν τις ασυσχέτιστες πληροφορίες του φάσματος. Ο πρώτος DCT συντελεστής (άρα και ο πρώτος MFCC) αντιπροσωπεύει την μέση ισχύ του φάσματος. Ο δεύτερος συντελεστής προσεγγίζει χοντρικά το σχήμα του φάσματος και συνδέεται με το φασματικό κεντροειδές. Οι συντελεστές μεγαλύτερης τάξης αντιπροσωπεύουν πιο λεπτά φασματικά χαρακτηριστικά (π.χ. οξύτητα).

ΚΕΦΑΛΑΙΟ 3:

3.1 Εισαγωγή

Ο εντοπισμός ομοιοτήτων μεταξύ μουσικών κομματιών αλλά ακόμα και μέσα στο ίδιο κομμάτι είναι πολύ συχνό φαινόμενο για έναν ακροατή. Μέσα στο ίδιο το κομμάτι οι ακροατές εντοπίζουν στιγμιαία μουσικά δομικά κομμάτια με παρόμοιες λειτουργίες – ρεφραίν , γέφυρες κτλ – εξάγοντας έτσι μια δομική περιγραφή για το κομμάτι. Η ομοιότητα μεταξύ διαφορετικών κομματιών χρησιμοποιείται από τους ακροατές για κατηγοριοποίηση κατά στυλ και είδη και οργάνωση των κομματιών σε συλλογές και λίστες αναπαραγωγής.

Παρόλο που η έννοια της μουσικής ομοιότητας είναι ασαφής εξαιτίας του ότι είναι εξαρτώμενη του πλαισίου και δεν υπάρχει σαφής στοιχειοθέτηση των μουσικών διαστάσεων που επηρεάζουν την ανθρώπινη αντίληψη κατά την αξιολόγηση της μουσικής ομοιότητας, εντούτοις κατά την διεξαγωγή αντιληπτικών πειραμάτων οι συμμετέχοντες μπορούν αρκετά εύκολα να καταλήξουν στο αν δύο μουσικά κομμάτια είναι όμοια ή όχι και μάλιστα μπορούν να το κάνουν με συνέπεια. Αυτό αναδεικνύει το γεγονός ότι παρόλο που αντίληψη της μουσικής ομοιότητας εξαρτάται από διάφορους σύνθετους παράγοντες όπως το ηχόχρωμα, ο ρυθμός, η κουλτούρα, το κοινωνικό πλαίσιο, το προσωπικό παρελθόν και τη διάθεση μεταξύ άλλων, οι ακροατές μπορούν να ερμηνεύουν την έννοια της μουσικής ομοιότητας με συνέπεια. Σε διερώτηση σχετικά με τον παράγοντα που τους οδήγησε να χαρακτηρίσουν δυο μουσικά κομμάτια ως όμοια, συχνά οι ακροατές αναφέρονται σε «επιφανειακά» χαρακτηριστικά, όπως εξέχοντα μουσικά στοιχεία ενός μουσικού κομματιού (π.χ. οι δυναμικές, η ένταση, ο ρυθμός και το ηχόχρωμα)

Βέβαια παρόλη την σύγκλιση αρκετών αντιληπτικών πειραμάτων στο γεγονός ότι τα επιφανειακά μουσικά χαρακτηριστικά επηρεάζουν την αντίληψη της μουσικής ομοιότητας ([Lamont et al., 2001],[McAdams et al., 2004]), δεν έχει υπάρξει ακόμα καμία συμφωνία μεταξύ των ερευνών αυτών σχετικά με το ποια χαρακτηριστικά κατέχουν καθοριστικό ρόλο στη κρίση αυτή. Από κάποιους ερευνητές υποστηρίζεται η θέση ότι αυτή η ασυμφωνία μπορεί να οφείλεται στο πλαίσιο των ερεθισμάτων που χρησιμοποιούνται, δηλαδή ότι κάθε ξεχωριστό σύνολο ερεθισμάτων μπορεί να οργανώνεται αντιληπτικά από ένα συγκεκριμένο σύνολο μεταβλητών ελέγχου.

Ελλείψει μιας γενικευμένης βάσης δεδομένων βασισμένη στη μουσική ομοιότητα που να καλύπτει διάφορα είδη μουσικής, ο έλεγχος των θεωρητικών μοντέλων και των αλγοριθμικών εφαρμογών που σχετίζονται με τη μουσική ομοιότητα καθίσταται προβληματικός. Έτσι πολλοί ερευνητές βασίζουν την εκπαίδευση και την αξιολόγηση των εφαρμογών τους σε διάφορες άλλες πηγές ([Aucouturier et al., 2002b],[Berenzweig, 2003]) όπως μεταδεδομένα και κείμενα διαδικτύου, ή επαφιόμενοι στην εύθραυστη υπόθεση ότι δύο κομμάτια είναι όμοια αν ανήκουν στον ίδιο καλλιτέχνη, ή στο ίδιο άλμπουμ ή εμφανίζονται στην ίδια playlist.

Ένα μεγάλο πρόβλημα στη συλλογή δεδομένων αντίληψης της ομοιότητας αποτελεί η σχέση του αριθμού των ερεθισμάτων και του χρόνου του πειράματος: ακόμα για έναν μικρό αριθμό από ερεθίσματα , ο αριθμός των συγκρίσεων που πρέπει να γίνουν απαιτεί αρκετά μεγάλο χρόνο διεξαγωγής του πειράματος. Στη βιβλιογραφία ([Burton et al., 1976],[Lamont et al., 2001]) απαντώνται κυρίως τρεις μέθοδοι για την εκτίμηση ομοιότητας μουσικών κομματιών με τη χρήση αντιληπτικών πειραμάτων:

- Στην **αξιολόγηση ζευγαριού** (pair rating), οι συμμετέχοντες αποδίδουν μια τιμή ομοιότητας για ένα ζευγάρι μουσικών αποσπασμάτων μέσα σε μια αριθμητική κλίμακα.
- Η **κατάταξη ζεύγους** (pair ranking) αποτελεί μια διαδικασία κατάταξης κατά την οποία οι συμμετέχοντες καλούνται να κατατάξουν ζεύγη μουσικών κομματιών ανάλογα με την ομοιότητα τους.
- Στη περίπτωση της **ομαδοποίησης αντικειμένων** (object grouping), παρουσιάζονται στους συμμετέχοντες μια σειρά από μουσικά κομμάτια και αυτοί στη συνέχεια καλούνται να τα ομαδοποιήσουν ανάλογα με την ομοιότητα που διαβλέπουν.

Σύμφωνα με κάποιες έρευνες η διαδικασία της αξιολόγησης ζευγαριού ενέχει αρκετές δυσκολίες για τους συμμετέχοντες και ίσως καταλήγει σε ανακριβή αποτελέσματα, εξαιτίας του ότι οι συμμετέχοντες αποκτούν οριστική αίσθηση της κλίμακας των ερεθισμάτων αφού εκτεθούν στο σύνολο των ερεθισμάτων, σε αντίθεση με την απλότητα και την ευρωστία της διαδικασίας κατάταξης που επιβάλει η κατάταξη ζεύγους. Όσο για την περίπτωση της ομαδοποίησης αντικειμένων, λόγω των μνημονικών ικανοτήτων που επιβάλλεται να έχουν οι συμμετέχοντες τα παραδείγματα που παρουσιάζονται πρέπει να είναι αριθμητικά λίγα ώστε τα αποτελέσματα να είναι αξιόπιστα.

3.2 Relative Distance Constraints

Η τριαδική σύγκριση, που έχει χρησιμοποιηθεί για την κατασκευή του dataset που θα χρησιμοποιήσουμε σε αυτήν την εργασία, θεωρείται μια ειδική περίπτωση κατάταξης ζεύγους, οπού οι συμμετέχοντες αφού ακούσουν τρία μουσικά αποσπάσματα, καλούνται να απαντήσουν πιο από τα τρία είναι το πιο διαφορετικό από τα υπόλοιπα. Έτσι κατασκευάζεται μια τριάδα μουσικών κομματιών (a,b,c) που αντιπροσωπεύει την γνώση ότι «το κομμάτι a είναι πιο όμοιο με το κομμάτι b από ότι είναι με το κομμάτι c».

Η λογική πίσω από την τριαδική σύγκριση υποθέτει ότι ο άνθρωπος έχει μια γενική γνώση της «κοντινότητας» (ομοιότητας) για ένα υποσύνολο των δεδομένων και αυτή η γνώση μπορεί να είναι αόριστη και μπορεί να στοιχειοθετηθεί μόνο με μια σύγκριση μεταξύ δύο παραδειγμάτων ποιο είναι πιο όμοιο με ένα τρίτο , παρά με τον ορισμό ενός απολύτου μέτρου εγγύτητας. Ανάλογες τριάδες μπορούν να προκύψουν και με έμμεσους τρόπους εξόρυξης γνώσης, όπως από την προσωπική ιεραρχία φακέλων με μουσικά αρχεία.

Κάθε τριάδα (a,b,c) η οποία περιλαμβάνει την γνώση που αναφέραμε πιο πριν, μπορεί να χρησιμοποιηθεί για να αντληθεί η πληροφορία ότι:

$$(a,b) >^{sim} (a,c)$$
$$\text{και}$$
$$(a,b) >^{sim} (b,c)$$

, όπου το σύμβολο $>^{sim}$ ερμηνεύεται ως «πιο όμοια από…» .

Δηλαδή μπορούμε να γράψουμε για την απόσταση των τριών κομματιών τους ακόλουθους *περιορισμούς σχετικών αποστάσεων (relative distance constraints)*:

$$d(a,b) < d(a,c)$$
$$\text{και} \qquad (3.I)$$
$$d(a,b) < d(b,c)$$

3.3 Κατασκευή Γράφου Περιορισμών

Ακολουθώντας την μέθοδο που περιγράφεται στο [McFee., 2009], συγκεντρώνοντας τις σχέσεις (I) για όλο το σύνολο δεδομένων, μπορούμε να κατασκευάσουμε έναν κατευθυνόμενο γράφου όπου κάθε ζεύγος κομματιών αντιστοιχεί σε έναν κόμβο και κάθε ψήφος αναπαρίσταται με μία ακμή, η οποία ξεκινάει από το ζευγάρι εκείνο που θεωρείται πιο όμοιο και καταλήγει στο ζευγάρι που θεωρείται πιο ανόμοιο. Δηλαδή για την παραπάνω τριάδα (a,b,c) εμφανίζονται στο γράφο οι ακμές από τον κόμβο (a,b) προς τους κόμβους (a,c) και (b,c).

Αν ο γράφος που θα προκύψει δεν περιέχει κύκλους, αποτελεί δηλαδή κατευθυνόμενο ακυκλικό γράφημα (directed acyclic graph – DAG), τότε μπορούμε να ορίσουμε μια μερική διάταξη των αποστάσεων μεταξύ των κομματιών, το οποίο συνεπάγεται την ύπαρξη κάποιου χώρου ομοιότητας που είναι συνεπής με τους συγκεντρωμένους ψήφους. Αν εμφανίζονται κύκλοι στο γράφο τότε δεν υπάρχει συνάρτηση ομοιότητας που να μπορεί να ικανοποιήσει όλους τους περιορισμούς που θέτουν οι ψήφοι και συνεπώς το σύνολο θεωρείται ασυνεπές.

Πρακτικά όμως επειδή πάντα η ανθρώπινη αντίληψη της ομοιότητας οδηγεί σε ασυνέπειες, ενώ οι κύκλοι στον γράφο μπορούν να θεωρηθούν "θόρυβοι ετικέτας" (label noise), οι οποίοι μπορεί να επιβαρύνουν το μοντέλο όσον αφορά την πολυπλοκότητα και την γενίκευση, συνήθως εφαρμόζονται τεχνικές εξάλειψης των κύκλων, αφαιρώντας τα κομμάτια εκείνα που δημιουργούν τις ασυνέπειες, προκειμένου να προκύψει μια μερική διάταξη.

3.4 Διατύπωση του Προβλήματος Μάθησης Απόστασης

Η χρήση μιας μετρικής για την αναπαράσταση της μουσικής ομοιότητας συνεπάγεται διάφορες παραδοχές, οι οποίες έχουν αμφισβητηθεί στο παρελθόν από τον Tversky [Tversky, 1977], ο οποίος υποστηρίζει ότι η ομοιότητα δεν αποτελεί μια γραμμική, θετικά ορισμένη και συμμετρική συνάρτηση, η οποία ικανοποιεί την τριγωνική ανισότητα, αλλά αποτελεί μια κατευθυνόμενη σχέση. Οι ιδιότητες, ωστόσο, μιας μετρικής επιτρέπουν την χρήση αποδοτικών και εύρωστων αλγορίθμων μάθησης και επιδέχονται μιας απλής γεωμετρικής ερμηνείας, που επιτρέπει ακόμα και την σύγκριση μεταξύ δύο διαφορετικών μετρικών και συνεπώς σε αυτήν την εργασία ακολουθήθηκε αυτή η προσέγγιση.

Δοσμένου ενός συνόλου αντικειμένων O και ενός συνόλου χαρακτηριστικών F των αντικειμένων του O, έστω ότι ο χώρος που ορίζεται από τις τιμές των χαρακτηριστικών είναι ο S. Μια πτυχή (facet) f ([Wolf et al., 2012b]) ορίζεται από το μέτρο απόστασης πτυχής δ_f σε έναν υποχώρο $S_f \subseteq S$, όπου η δ_f ικανοποιεί τις ακόλουθες σχέσεις για οποιαδήποτε $a, b \in O$:

- $\delta_f(a,b) \geq 0$ και $\delta_f(a,b) = 0$ αν και μόνο αν $a = b$
- $\delta_f(a,b) = \delta_f(b,a)$ (Συμμετρία)

Επιπλέον η δ_f αποτελεί μια μετρική απόστασης αν ικανοποιεί την τριγωνική ανισότητα για κάθε $a, b \in O$:

$$\delta_f(a,c) \leq \delta_f(a,b) + \delta_f(b,c)$$

Προκειμένου να αποφευχθεί η μεροληψία καθώς συναθροίζονται αρκετές μετρήσεις απόστασης πτυχών, θα πρέπει οι τιμές να κανονικοποιηθούν, όπως στο [Stober, 2011b]. Συνεπώς όλες οι αποστάσεις πτυχών θα υποστούν την ακόλουθη κανονικοποίηση ώστε να έχουν μέση τιμή 1:

$$\delta_f' = \frac{\delta_f(a,b)}{\mu_f} \qquad (3.1)$$

,όπου μ_fείναι η μέση απόσταση της πτυχής f::

$$\mu_f = \frac{1}{|\{(a,b) \in O^2\}|} \sum_{(a,b)\in O^2} \delta_f(a,b) \qquad (3.2)$$

Σύμφωνα με τη Τοπική- Καθολική Αρχή(Local – Global Principal[7]), η καθολική απόσταση δύο αντικειμένων μπορεί να υπολογιστεί σαν άθροισμα (aggregation) των τοπικών μέτρων απόστασης:

$$d(a,b) = AGG\left(\delta_{f1}(a,b), \delta_{f2}(a,b) \dots \delta_{fl}(a,b)\right) \qquad (3.3)$$

,όπου AGG ένας κατάλληλος τελεστής συσσωμάτωσης.

Στη παρούσα εργασία το μέτρο απόστασης που επιθυμούμε να βελτιστοποιήσουμε ορίζεται σε επίπεδο κλιπ 29 δευτερολέπτων και για τον υπολογισμό της απόστασης αυτής ανάμεσα στα αντικείμενα $a, b \in O$ όσον αφορά τις πτυχές $f_1, \dots, f_l$ θα χρησιμοποιήσουμε τον γραμμικό συνδυασμό των ξεχωριστών αποστάσεων $\delta_{f1}(a,b), \delta_{f2}(a,b), \dots, \delta_{f3}(a,b)$:

$$d(a,b) = \sum_{i=1}^{l} w_i \delta_{fi}(a,b) \qquad (3.4)$$

Με την εισαγωγή αυτών των βαρών μπορούμε να προσαρμόσουμε την βαρύτητα κάθε πτυχής, ανάλογα με το πρόβλημα που έχουμε να αντιμετωπίσουμε. Τα βάρη αυτά θα πρέπει να

[7] Σύμφωνα με την Τοπική – Καθολική Αρχή (Local Global Principle) μια συγκεκριμένη ιδιότητα είναι αληθής καθολικά αν και μόνο αν είναι αληθής παντού τοπικά.

είναι θετικά ή μηδέν για να διασφαλιστεί η μονοτονικότητα του μέτρου απόστασης (3.3) και να διαθέτουν ένα άνω όριο:

$$w_i \geq 0 \qquad \forall i\,, 1 \leq i \leq l \tag{3.5}$$

$$\sum_{i=1}^{l} w_i = l \tag{3.6}$$

Τα βάρη αυτά μπορούν είτε να καθοριστούν χειροκίνητα είτε να αντληθούν από πληροφορίες προτιμήσεων. Στη παρούσα εργασία οι πληροφορίες προτιμήσεων προέρχονται από τριαδικές συγκρίσεις όπως αυτές αναλύθηκαν πιο πάνω και ανάγονται σε περιορισμούς σχετικών αποστάσεων.

Συνεπώς συνδυάζοντας τις εξισώσεις (3.3.I) και (3.3.4) οι περιορισμοί σχετικών αποστάσεων μπορούν να εκφραστούν ως:

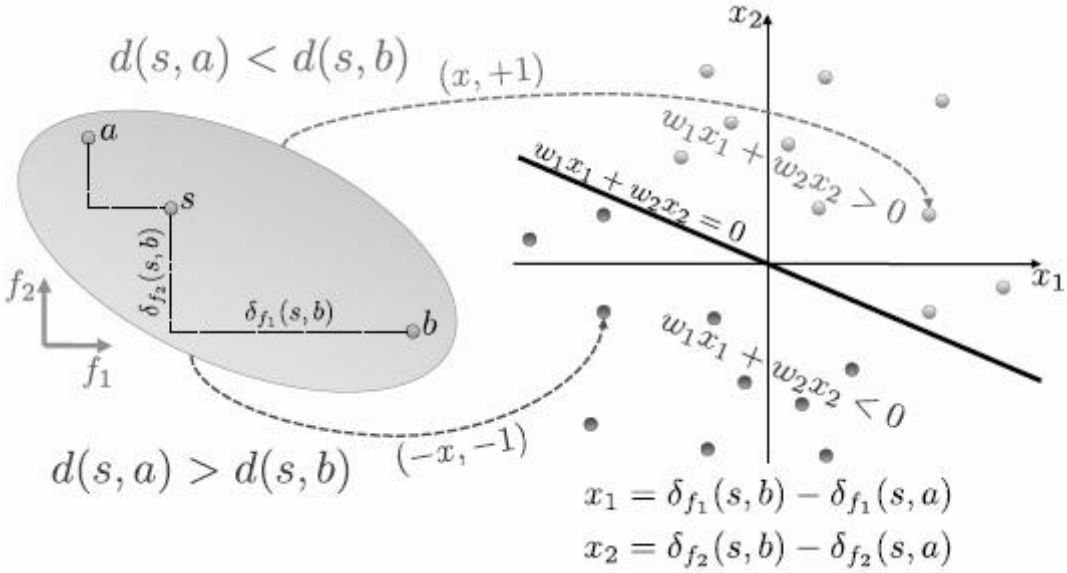
$$\sum_{i=1}^{l} w_i \left(\delta_{fi}(a,c) - \delta_{fi}(a,b)\right) = \sum_{i=1}^{l} w_i x_i > 0 \tag{3.7}$$

, όπου έγινε η αντικατάσταση $x_i = \delta_{fi}(a,c) - \delta_{fi}(a,b)$.

3.5 Μετασχηματισμός σε Πρόβλημα Ταξινόμησης

Για τον υπολογισμό τον βαρών της εξίσωσης (3.7) έχουν χρησιμοποιηθεί αρκετές τεχνικές. Στο [Stober, 2011b] γίνεται χρήση αλγορίθμων βελτιστοποίησης (Gradient Descend και τετραγωνικός Προγραμματισμός), ενώ στο [Wolf et al., 2011] προτείνεται η χρήση ranking αλγορίθμων, και συγκεκριμένα του Metric Learning to Rank (MLR) που παρουσιάζεται στο [McFee et al., 2010]. Από τους [Cheng et al., 2008] προτείνεται η τεχνική της κατασκευής ενός δυαδικού προβλήματος ταξινόμησης, η οποία ακολουθήθηκε στη παρούσα εργασία, και χρήση συνηθισμένων ταξινομητών.

Σύμφωνα με το [Cheng et al., 2008] το πρόβλημα μάθησης απόστασης μπορεί να απλοποιηθεί σε πρόβλημα ταξινόμησης (classification problem) με το $\boldsymbol{x} = (x_1, x_2, \dots, x_l)$ να αποτελεί ένα θετικό παράδειγμα και το $(-\boldsymbol{x})$ ένα αρνητικό. Δηλαδή ο περιορισμός ομοιότητας (a, b, c) μπορεί να μετασχηματιστεί σε δύο παραδείγματα $(\boldsymbol{x}, +1)$ και $(-\boldsymbol{x}, -1)$ για μάθηση δυαδικής ταξινόμησης. Στην εικόνα που ακολουθεί παρουσιάζεται αυτός ο μετασχηματισμός του προβλήματος. Επιπλέον ,το διάνυσμα $w = (w_1, \dots, w_d)$ που ορίζει τη συνάρτηση απόστασης (3.3.4) με μοναδικό τρόπο, ορίζει και την υπερεπιφάνεια (hyperplane) του σχετικού προβλήματος ταξινόμησης.



Εικόνα 9 Η Μετατροπή του περιορισμού σχετικής απόστασης σε δύο πρότυπα εκπαίδευσης του αντίστοιχου δυαδικού προβλήματος ταξινόμησης όπως περιγράφεται στο [Cheng et al., 2008]

Παρόλη την απλότητα του το γραμμικό μοντέλο (3.3.4) παρουσιάζει αρκετά προτερήματα. Είναι εύκολα ερμηνεύσιμο, καθώς τα βάρη w_i βρίσκονται σε άμεση αντιστοιχία με την σημαντικότητα του τοπικού μέτρου. Επιπλέον επιτρέπει να ενσωματωθούν πρόσθετες γνώσεις με εύκολο τρόπο (π.χ. η γνώση ότι το χαρακτηριστικό f_i είναι τουλάχιστον όσο σημαντικό είναι το χαρακτηριστικό f_j μπορεί να προστεθεί σαν $w_i \geq w_j$). Επιπλέον, από την οπτική της μηχανικής μάθησης το γραμμικό μοντέλο είναι αρκετά ελκυστικό, καθώς μπορεί να επιδεχθεί επίλυση με αποδοτικούς αλγορίθμους ή και με μη – γραμμικές προεκτάσεις μέσω χρήσης συναρτήσεων πυρήνα.

Στην παρούσα εργασία για την επίλυση του δυαδικού προβλήματος ταξινόμησης χρησιμοποιήσαμε ένα Νευρωνικό Δίκτυο ενός Επιπέδου Πρόσθιας Τροφοδότησης, Γραμμικό SVM ένα μη γραμμικό SVM

ΚΕΦΑΛΑΙΟ 4:

Τεχνητά Νευρωνικά Δίκτυα

4.1 Ιστορική Αναδρομή

Ο ανθρώπινος εγκέφαλος αποτελεί ένα σύνθετο, μη γραμμικό σύστημα επεξεργασίας δεδομένων με δυνατότητες παράλληλης επεξεργασίας και συνεχούς μάθησης μέσω της αλληλεπίδρασης με το περιβάλλον, για την εκτέλεση πολύπλοκων υπολογισμών και λειτουργιών (π.χ. αναγνώριση προτύπων κ.α.). Η επιστημονική κοινότητα που ασχολείται με τον τομέα της Τεχνητής Νοημοσύνης, προσπαθεί να προσεγγίσει τη λειτουργία του ανθρώπινου εγκέφαλου, προκειμένου είτε να κατανοήσει πλήρως την λειτουργία του είτε μέσω της μίμησης της λειτουργίας του να κατασκευάσει αποδοτικότερους αλγορίθμους.

Η κατανόηση της λειτουργίας του εγκεφάλου αν και έχει τις ρίζες της στην αρχαιότητα και στις έρευνες του Αλκμέωνα και του Ιπποκράτη πάνω σε ανθρώπινους και μη εγκεφάλους, ευνοήθηκε από την ανακάλυψη από τον Camilo Golgi της μεθόδου του εμποτισμού του νευρικού ιστού με χρωμιούχο άργυρο, η οποία πυροδότησε την υπόθεση του Ramon Y Cajal το 1911, ότι ο εγκέφαλος απαρτίζεται από βασικές αυτόνομες δομικές μονάδες, που ονομάζονται *νευρώνες*.

Αργότερα (το 1943) με την συνεισφορά του νευροφυσιολόγου Warren McCulloch και του μαθηματικού Walter Pitts κατασκευάστηκε το πρώτο μοντέλο ενός νευρώνα. Το μοντέλο αυτό, εμπνευσμένο από τα πραγματικά νευρωνικά κύτταρα, είχε έναν αριθμό δυαδικών εισόδων οι οποίες ήταν διεγερτικές και κάποιες που ήταν ανασταλτικές. Η τιμή της εξόδου καθοριζόταν από την σύγκριση του αθροίσματος των εισόδων με μια προκαθορισμένη τιμή, γνωστή και ως κατώφλι. Η έξοδος του μοντέλου αυτού ,ήταν 1 όταν το προαναφερθέν άθροισμα ήταν μεγαλύτερο από την τιμή του κατωφλίου και 0 αν ήταν μικρότερη.

Η επόμενη μεγάλη συνεισφορά μπορεί να αποδοθεί στον ψυχολόγο Donald Hebb ο οποίος υποστηρίζοντας το μοντέλο των McCulloch – Pitts και την λειτουργία του, περιέγραψε τον τρόπο με τον οποίο οι νευρικές οδοί ενισχύονται κάθε φορά που χρησιμοποιούνται, μια έννοια θεμελιώδης στον τρόπο που μαθαίνει ο ανθρώπινος νους.

Μερικά χρόνια αργότερα, το 1958, ο Rosenblatt ανέπτυξε το μοντέλο του perceptron, το οποίο αποτελεί το πρώτο πρακτικά τεχνητό νευρωνικό δίκτυο. Το perceptron αποτελεί ουσιαστικά ένα δίκτυο δύο επιπέδων, το οποίο ήταν ικανό να μάθει συγκεκριμένες κατηγοριοποιήσεις (classifications) προσαρμόζοντας τα συναπτικά βάρη.

Το 1959 οι Widrow και Hoff ανέπτυξαν το μοντέλο ADALINE (από τα αρχικά των λέξεων Adaptive Linear Element), το οποίο αποτελεί ένα απλό σύστημα, το οποίο ομοιάζει με το perceptron, και πραγματοποιεί κατηγοριοποιήσεις προσαρμόζοντας τα συναπτικά βάρη, με

τέτοιο τρόπο ώστε να ελαχιστοποιεί το Μέσο Τετραγωνικό Σφάλμα (Mean Square Error) σε κάθε επανάληψη.
Παρόλο που το μοντέλο του perceptron αυτό ήταν αρκετά επιτυχές στην κατηγοριοποίηση συγκεκριμένων προτύπων, εμπεριείχε όμως έναν αριθμό περιορισμών, όπως το γεγονός ότι επίλυε μόνο γραμμικά διαχωρίσιμα προβλήματα, οι οποίοι περιγράφηκαν από το βιβλίο των Minsky και Papert «Perceptrons» το 1969, οδηγώντας έτσι τις ακμάζουσες μέχρι τότε ερευνητικές εργασίες πάνω στα perceptrons, σε απότομη διακοπή.
Χρειάστηκε να περάσουν μερικά χρόνια, μέχρι την απόδειξη του John Hoppfield το 1982 πως ένα νευρωνικό δίκτυο μπορεί να χρησιμοποιηθεί σαν αποθηκευτικός χώρος, και μπορεί να ανακτήσει όλη τη πληροφορία ενός συστήματος αν δοθούν μερικά τμήματα του συστήματος, καθώς και την πρόταση μιας νέας διαδικασίας εκπαίδευσης από McClelland και Rumelhart (1986), την μέθοδο οπισθοδρόμησης (back-propagation), για να επανέλθουν τα νευρωνικά δίκτυα στο επίκεντρο του επιστημονικού ενδιαφέροντος.

4.2 Βιολογικό Πρότυπο

Για να γίνει ευκολότερα κατανοητή η λειτουργία των τεχνητών νευρωνικών_δικτύων θα ήταν χρήσιμη μια αναφορά στα δομή των βιολογικών νευρώνων και στη λειτουργία τους.

Ο νευρώνας αποτελεί τη βασική λειτουργική μονάδα του ανθρώπινου εγκεφάλου. Ο ανθρώπινος εγκέφαλο αποτελείται από νευρώνες, οι όποιοι ανέρχονται περίπου στον αριθμό των 10^{11}, και διαφοροποιούνται αρκετά μεταξύ τους. Υπάρχουν αρκετές κατηγοριοποιήσεις νευρώνων (σύμφωνα με το σχήμα τους, τη λειτουργία τους κ.α.) αλλά όλοι διέπονται από κάποιες βασικές αρχές δομής και λειτουργίας τις οποίες θα αναφέρουμε.
Ένα νευρωνικό κύτταρο αποτελείται από τα εξής βασικά στοιχεία: το σώμα του κυττάρου(cell body), μέσα στον οποίο βρίσκεται και ο πυρήνας του κυττάρου, τον άξονα του κυττάρου (axon), ο οποίος μοιάζει με μια λεπτή ίνα και τους δενδρίτες (dendrites), οι οποίοι αποτελούν μια διακλαδωτική διάρθρωση εισροών. Οι δενδρίτες και ο άξονας ενός κυττάρου συνδέονται μέσω «επαφών» οι οποίες ονομάζονται συνάψεις με γειτονικά κύτταρα. Κάθε νευρώνας έχει έναν μόνο άξονα ο οποίος συνδέεται μέσω συνάψεων με τους δενδρίτες των γειτονικών νευρώνων. Η ροή πληροφοριών μέσα σε ένα νευρώνα έχει κατεύθυνση από τους δενδρίτες προς το σώμα και αν πρόκειται να μεταδοθεί στον άξονα, όπου μέσω της σύναψης με έκκριση χημικών ουσιών , τους νευροδιαβιβαστές, μεταφέρεται προς τους δενδρίτες γειτονικών νευρώνων. Το προσυναπτικό νευρωνικό κύτταρο (αυτό που απελευθερώνει το νευροδιαβιβαστή) μπορεί να επάγει στο μετασυναπτικό κύτταρο (το οποίο προσλαμβάνει το νευροδιαβιβαστή) μια ηλεκτρική διέγερση που θα διαβιβαστεί στο αξονικό λοφίδιο ώστε να δημιουργηθεί ένα δυναμικό ενέργειας (action potential) το οποίο μετά θα διαβιβαστεί ως ηλεκτρική διέγερση κατά μήκος του νευράξονα. Κατά την άφιξη στην απόληξη του νευράξονα, προκαλείται απελευθέρωση του νευροδιαβιβαστή στο συναπτικό κενό. Οι νευροδιαβιβαστές γενικά μπορεί είτε να προκαλέσουν διέγερση είτε να εμποδίσουν τη διέγερση του κυττάρου-στόχου. Το δυναμικό ενέργειας θα παραχθεί στο κύτταρο-στόχο αν τα μόρια του νευροδιαβιβαστή που δρουν στους μετα-συναπτικούς υποδοχείς οδηγήσουν το κύτταρο-στόχο στο να φτάσει τον ουδό ή αλλιώς κατώφλι πυροδότησής του.

4.3 Δομή και Λειτουργία Τεχνητού Νευρώνα

Σύμφωνα με τον Simon Haykin [Haykin, 1994]:

"*Ένα νευρωνικό δίκτυο αποτελεί έναν μαζικά κατανεμημένο παράλληλο επεξεργαστή, ο οποίος διαθέτει φυσική κλίση στην αποθήκευση εμπειρικής γνώσης και στη διάθεση της προς χρήση. Ομοιάζει στον ανθρώπινο εγκέφαλο υπό δύο έννοιες:*

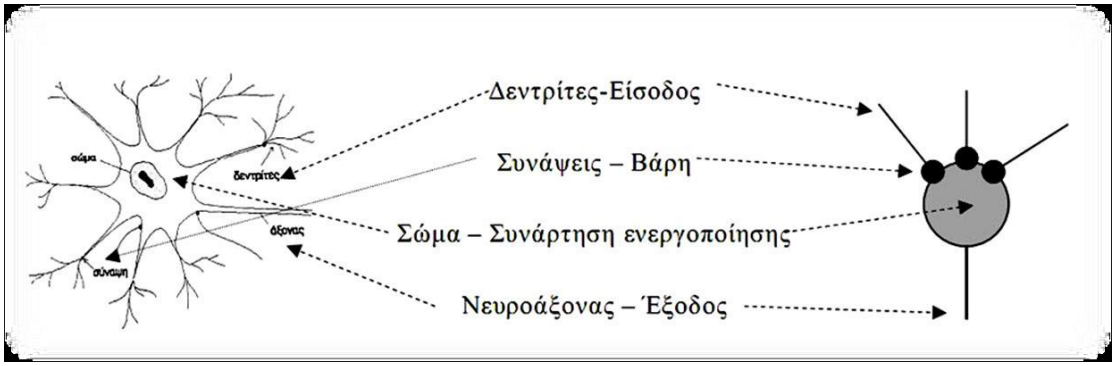
Η γνώση αποκτάται από το δίκτυο μέσω μιας διαδικασίας μάθησης

Η γνώση αποθηκεύεται στα συναπτικά βάρη μεταξύ των νευρώνων."

Οι αρχές λειτουργίας των τεχνητών νευρωνικών δικτύων είναι οι ακόλουθες:

- Η επεξεργασία της πληροφορίας πραγματοποιείται στους νευρώνες, οι οποίοι είναι μη γραμμικοί επεξεργαστές.
- Τα σήματα μεταβιβάζονται μεταξύ των νευρώνων με την βοήθεια των συνάψεων, δηλαδή των συνδετικών κλάδων μεταξύ των νευρώνων.
- Σε κάθε σύναψη αντιστοιχεί ένας συντελεστής, τα αποκαλούμενα συναπτικά βάρη, το οποίο αποτελεί έναν πολλαπλασιαστή σήματος.
- Τα σήματα εισόδου, πολλαπλασιασμένα με τα βάρη και αφού υποστούν άθροιση, εισάγονται στην συνάρτηση ενεργοποίησης το αποτέλεσμα της οποίας λαμβάνεται ως έξοδος.

Ο αναγνώστης μπορεί εύκολα να συμπεράνει την αντιστοιχία μεταξύ των μερών του Τεχνητού Νευρωνικού Δικτύου και του αντίστοιχου βιολογικού η οποία παρουσιάζεται και στο ακόλουθο σχήμα



Εικόνα ΙΙ Αντιστοιχία Βιολογικού και Τεχνητού Νευρώνα

Σε πλήρη αντιστοιχία με τον ανθρώπινο εγκέφαλο και το βιολογικό νευρώνα , ένας τεχνητός νευρώνας, αποτελώντας τη βασική δομική μονάδα των τεχνητών νευρωνικών δικτύων, δέχεται σαν είσοδο πληροφορίες, τις επεξεργάζεται και αποδίδει μια τιμή εξόδου.

Ένας νευρώνας αποτελεί τη στοιχειώδη μονάδα διαχείρισης πληροφοριών των τεχνητών νευρωνικών δικτύων και αποτελείται από τα ακόλουθα στοιχεία:

- Ένα σύνολο **συνδέσεων / συνάψεων** προς άλλους νευρώνες του δικτύου. Κάθε μια από αυτές τις συνδέσεις φέρει σαν χαρακτηριστικό, το **συναπτικό βάρος** w_{kj} που αναφέρεται στην σύναψη j του νευρώνα k .

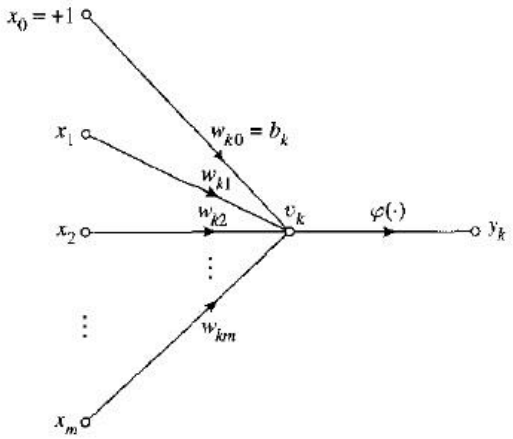
- Έναν **αθροιστή** που επιτελεί την λειτουργία της άθροισης των εισερχόμενων σημάτων πολλαπλασιασμένων με τα αντίστοιχα συναπτικά βάρη
- Μια **συνάρτηση ενεργοποίησης** η οποία περιορίζει το πλάτος της εξόδου του νευρώνα, συνήθως στο πεδίο τιμών [-1,1] ή [0,1].
- Συνήθως υπάρχει και η πόλωση του νευρώνα (bias) ή αλλιώς κατώφλι (threshold).

Οι είσοδοι του νευρώνα πολλαπλασιάζονται με τα αντίστοιχα συναπτικά βάρη και μαζί με την πόλωση εισέρχονται στον αθροιστή και από την έξοδο αυτού προκύπτει το δυναμικό ενεργοποίησης (activation potential). Μέχρι αυτό το σημείο ο νευρώνας επιτελεί μόνο τη λειτουργία του γραμμικού συνδυασμού. Στη συνέχεια στην έξοδο του αθροιστή εφαρμόζεται η συνάρτηση ενεργοποίησης (activation function), η οποία τελικά θα καθορίσει την τιμή της εξόδου του νευρώνα.

Χρησιμοποιώντας μαθηματικούς όρους μπορούμε να περιγράψουμε τον νευρώνα με τις ακόλουθες σχέσεις:

$$u_k = \sum_{j=1}^{p} w_{kj} x_j \qquad \text{και} \qquad y_k = \varphi(u_k - \theta_k)$$

, όπου $x_1, x_2, \dots, x_p$ είναι τα σήματα εισόδου του νευρώνα, $w_{k1}, w_{k2}, \dots, w_{kp}$ είναι τα συναπτικά βάρη του k νευρώνα, u_k είναι το δυναμικό ενεργοποίησης, θ_k το κατώφλι, $\varphi(\cdot)$ η συνάρτηση ενεργοποίησης και y_k η έξοδος του νευρώνα. Η χρήση του κατωφλιού θ_k έχει το αποτέλεσμα της εφαρμογής ενός Αφινικού Μετασχηματισμού (affine transformation) στην έξοδο u_k του αθροιστή.



Εικόνα 11 Διάγραμμα Ροής Σημάτων ενός τυπικού Νευρώνα

Όπως φαίνεται και στο σχήμα είναι : $v_k = u_k - \theta_k$

Για λόγους απλούστευσης των υπολογισμών η πόλωση/ το κατώφλι θεωρείται σαν ένα επιπλέον συναπτικό βάρος του οποίου η τιμή εισόδου είναι πάντα ίση με -1 και συνεπώς προκύπτουν οι ακόλουθες σχέσεις:

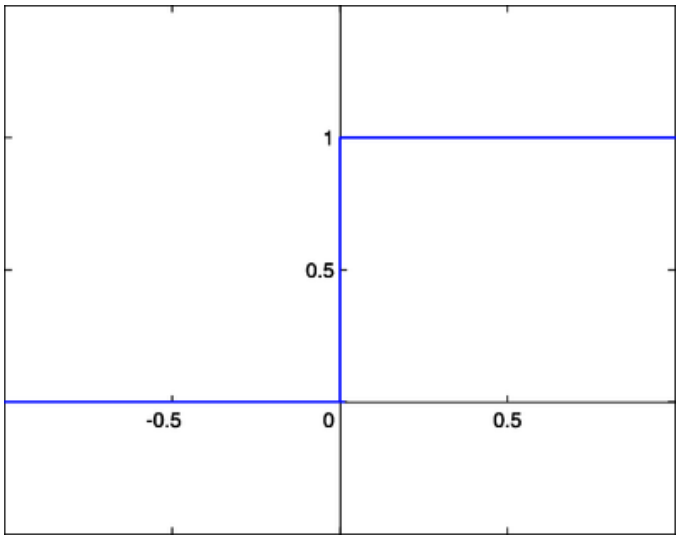
$$u_k = \sum_{j=0}^{p} w_{kj} x_j \qquad \text{και} \qquad y_k = \varphi(u_k)$$

Για λόγους πληρότητας αναφέρουμε ότι αν στο μοντέλο χρησιμοποιηθεί η έννοια της πόλωσης αντί για αυτή του κατωφλίου τότε θεωρείται πάλι ένα συναπτικό βάρος w_0 ίσο με την τιμή της πόλωσης αλλά σε αυτή τη περίπτωση η αντίστοιχη τιμή της εισόδου ισούται πάντα με 1. Τα δύο αυτά μαθηματικά μοντέλα είναι μαθηματικά ισοδύναμα.

4.4 Τύποι Συναρτήσεων Ενεργοποίησης:

Σε έναν βιολογικό νευρώνα η συνάρτηση ενεργοποίησης πιστεύεται ότι ομοιάζει με αυτήν που αποκαλούμε συνάρτηση κατωφλίου (threshold function), η οποία ορίζεται από την ακόλουθη σχέση

$$f(v) = \begin{cases} 1\,, v \geq 0 \\ 0\,, v \leq 0 \end{cases}$$



Εικόνα 12 Συνάρτηση Κατωφλιού Τεχνητού Νευρώνα

Στα τεχνητά νευρωνικά δίκτυα όμως, επειδή επιθυμούμε να έχουμε συνεχείς και διαφορίσιμες συναρτήσεις ενεργοποίησης, για λόγους που θα εξηγηθούν στη συνέχεια, χρησιμοποιούνται συνήθως είτε *γραμμικές ή τμηματικά γραμμικές* είτε *σιγμοειδής* συναρτήσεις, οι οποίες προσεγγίζουν την λειτουργία της βηματικής αλλά επιπλέον προσφέρουν δυνατότητες παραγωγισιμότητας.

Η γραμμική συνάρτηση αναπαρίσταται από την εξίσωση:

$$f(x) = k \cdot x$$

Μια τμηματικά γραμμική συνάρτηση περιγράφεται από την εξίσωση:

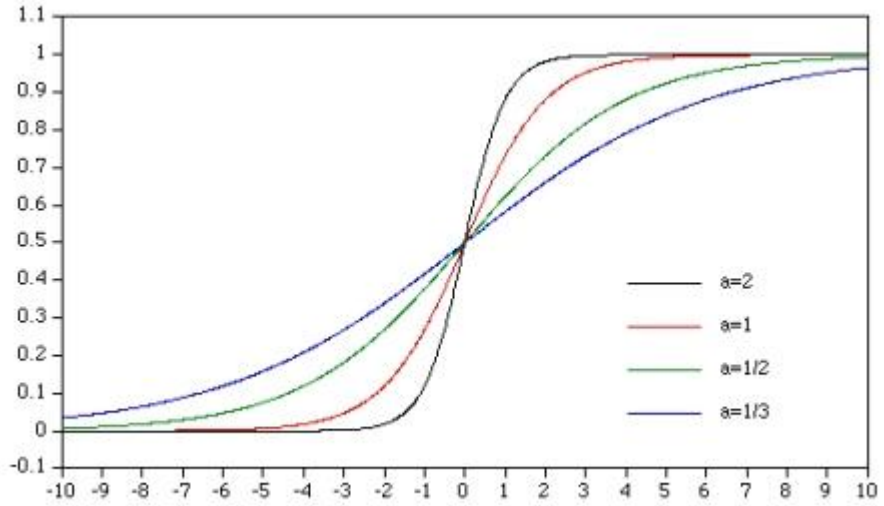
$$\varphi(v) = \begin{cases} 1\,, & v \geq \frac{1}{2} \\ v\,, & \frac{1}{2} > v > -\frac{1}{2} \\ 0\,, & -\frac{1}{2} \geq v \end{cases}$$

Συνηθέστερα χρησιμοποιείται η *λογιστική συνάρτηση (logistic function)* η οποία ανήκει στην κατηγορία των σιγμοειδών και ορίζεται από την ακόλουθη σχέση:

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

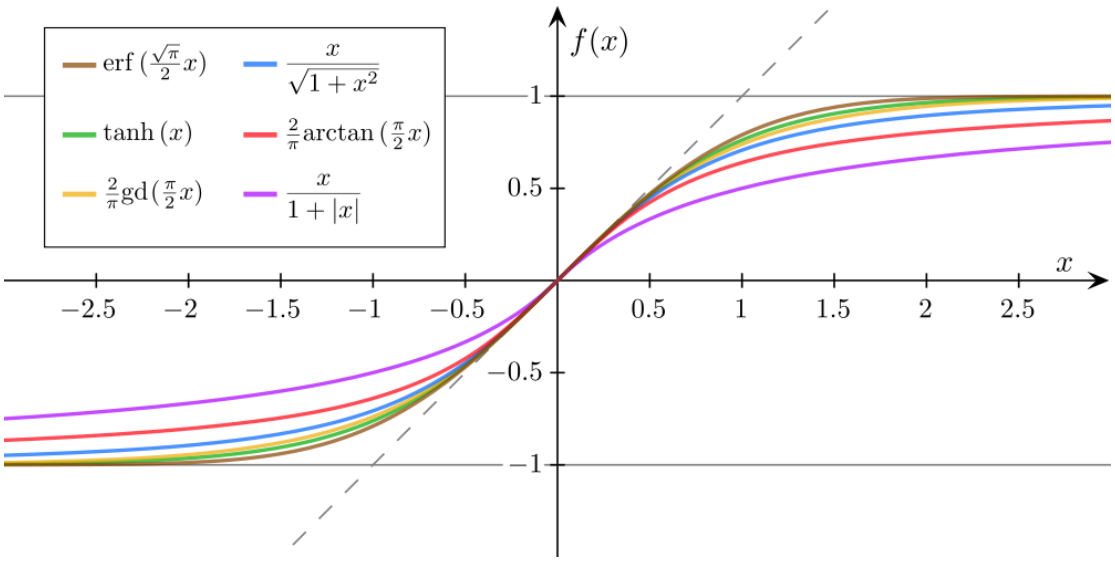
όπου με την παράμετρο α μπορούμε να μεταβάλλουμε την κλίση της σιγμοειδούς.

Στην εικόνα που ακολουθεί γίνεται εμφανής η σχέση της κλίσης της λογιστικής συνάρτησης συναρτήσει της παραμέτρου α.



Εικόνα 13 Λογιστική Συνάρτηση: Απεικόνιση σχέσης κλίσης και τιμής της παραμέτρου

Στην επόμενη εικόνα παρουσιάζονται περεταίρω σιγμοειδείς συναρτήσεις που μπορούν να χρησιμοποιηθούν σαν συναρτήσεις ενεργοποίησης.



Εικόνα 14 Σιγμοειδείς συναρτήσεις ενεργοποίησης

Αξίζει να σημειωθεί σε αυτό το σημείο ότι η πόλωση μας επιτρέπει να καθορίσουμε τη μετακίνηση της συνάρτησης ενεργοποίησης δεξιά – αριστερά ώστε να επιτευχθεί σωστά η ταξινόμηση.

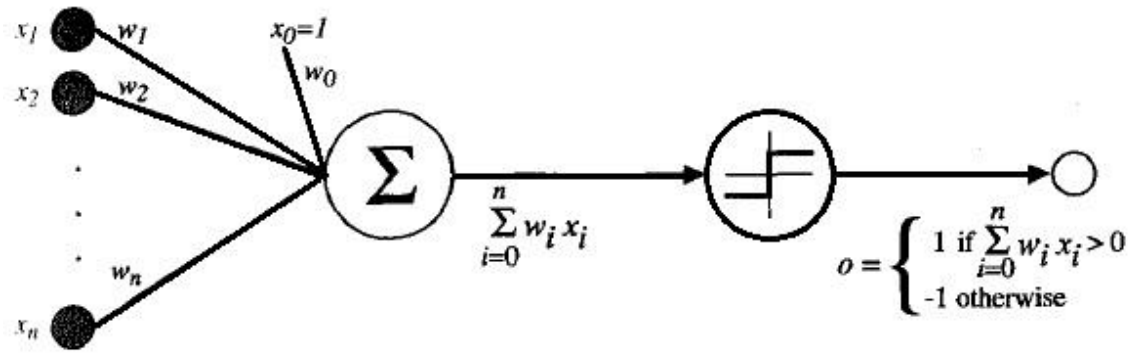
4.5 Perceptron:

Αποτελεί την πιο απλή μορφή νευρωνικού δικτύου αποτελούμενος από δύο επίπεδα νευρώνων, το επίπεδο εισόδου και το επίπεδο εξόδου. Χρησιμοποιείται για ταξινόμηση γραμμικά διαχωρίσιμων δεδομένων. Ο perceptron δέχεται σαν είσοδο ένα διάνυσμα πραγματικών τιμών, υπολογίζει έναν γραμμικό συνδυασμό από τις εισόδους αυτές στο επίπεδο εξόδου και επιστρέφει 1 αν το αποτέλεσμα προέκυψε πάνω από ένα συγκεκριμένο κατώφλι, διαφορετικά επιστρέφει -1. Συνεπώς ένας perceptron έχει τη δυνατότητα να ταξινομήσει πρότυπα που ανήκουν σε δύο κλάσεις.

Πιο συγκεκριμένα, δεδομένης της εισόδου $x_1, x_2, \dots, x_n$ η έξοδος του perceptron $o(x_1, x_2, \dots, x_n)$ υπολογίζεται ως εξής:

$$o(x_1, x_2, \dots, x_n) = \begin{cases} 1, & \alpha\nu\ w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n > 0 \\ -1, & \alpha\lambda\lambda\iota\acute{\omega}\varsigma \end{cases}$$

, όπου κάθε w_i ονομάζεται βάρος και αποτελεί μια σταθερά πραγματικής τιμής, η οποία καθορίζει την συνεισφορά της εισόδου x_i στο αποτέλεσμα της εξόδου. Σημειώνουμε ότι το κατώφλι που αναφέρθηκε πιο πριν είναι η τιμή $(-w_0)$ και είναι η τιμή που πρέπει να υπερβεί το άθροισμα $w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$ για να προκύψει η έξοδος του perceptron ίση με 1.



Εικόνα III Απεικόνιση Perceptron

Συνεπώς η έξοδος του perceptron δίνεται από τη σχέση: $\upsilon = \sum_{i=1}^{n} w_i x_i - w_0$

Ο σκοπός του perceptron είναι να ταξινομεί το σύνολο των εξωτερικών ερεθισμάτων $x_1, x_2, \dots, x_n$ σε μια από τις δύο κλάσσεις, έστω $\mathcal{C}_1$ και $\mathcal{C}_2$. Ο κανόνας διαχωρισμού για την ταξινόμηση είναι η ανάθεση του σημείου που αναπαριστάται με τις εισόδους $x_1, x_2, \dots, x_n$ στη κλάση $\mathcal{C}_1$ αν η έξοδος του perceptron είναι 1 και η ανάθεσή του στη κλάση $\mathcal{C}_2$ αν η έξοδος είναι -1.

Για να απλοποιήσουμε την σημειογραφία, υποθέτουμε ότι υπάρχει μια επιπλέον σταθερή είσοδος $x_0 = 1$, το οποίο μας επιτρέπει να γράφουμε τις ανωτέρω σχέσεις σε διανυσματική μορφή:

$$\vec{w} \cdot \vec{x} > 0 \text{ και } o(\vec{x}) = sgn(\vec{w} \cdot \vec{x})$$

, όπου sgn η συνάρτηση προσήμου.

Ενώ το θεώρημα σύγκλισης του perceptron εγγυάται ότι η ταξινόμηση των γραμμικά διαχωριζομένων δεδομένων θα γίνει σωστά, στη πραγματικότητα σπάνια έχουμε να αντιμετωπίσουμε τέτοια προβλήματα. Στη περίπτωση μη γραμμικά διαχωρίσιμων δεδομένων ενδείκνυται η χρήση ενός πολυεπίπεδου perceptron. Σε ένα πολυεπίπεδο perceptron η έξοδος των perceptron του πρώτου επιπέδου γίνεται είσοδος του επόμενου επιπέδου. όμως το μοντέλο αυτό αντιμετωπίζει πρόβλημα μάθησης καθώς το Θεώρημα Σύγκλισης δεν επεκτείνεται στα πολυεπίπεδα perceptron.

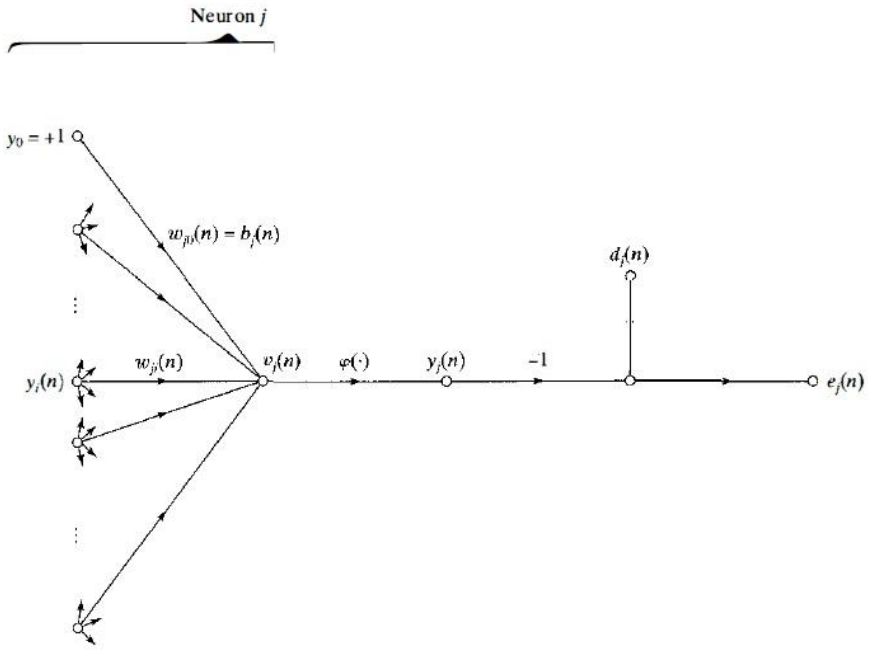
4.6 Multilayer Neural Networks

Η ικανότητα εκπαίδευσης πολυεπίπεδων δικτύων αποτελεί σημαντικό βήμα στη κατασκευή ευφυών μηχανών αποτελούμενες από νευρώνες. Στη συνέχεια θα παρουσιάσουμε μια υποκλάση των Πολυεπίπεδων Νευρωνικών Δικτύων και συγκεκριμένα αυτή των πλήρως διασυνδεδεμένων , Πρόσθιας Τροφοδότησης Δικτύων(Fully Connected, Feedforword). Η κομβική ενεργοποίηση έχει ροή από το επίπεδο εισόδου στο επίπεδο εξόδου μέσω ενός κρυφού επιπέδου.

Ένας τυπικός νευρώνας αυτού του δικτύου παρουσιάζεται στην εικόνα που ακολουθεί και περιγράφεται από τη σχέση

$$y_j = \sum_{i=1}^{N} w_{ij} x_i - \theta_j$$

,όπου με x_i συμβολίζουμε κάθε μια από τις N εισόδους στον κόμβο j, με w_{ij} το συναπτικό βάρος μεταξύ του κόμβου i και του κόμβου j,με θ_j την πόλωση (bias) του κόμβου j,και με y_j την έξοδο από τον κόμβο j.

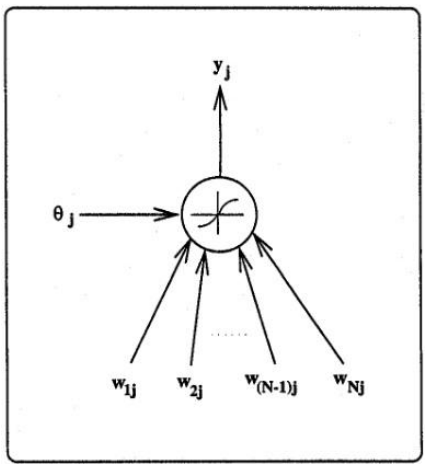


Εικόνα 16 Απεικόνιση της ροής των σημάτων στον νευρώνα j του στρώματος εξόδου

Κάθε νευρώνας ενός επιπέδου ενώνεται προς την πρόσθια κατεύθυνση με κάθε νευρώνα του επόμενου επιπέδου. Η γνώση του δικτύου αποθηκεύεται στα συναπτικά βάρη μεταξύ των νευρώνων. Κάθε επίπεδο ενός πολυεπίπεδου δικτύου μπορεί να εκφραστεί σε μορφή πίνακα ως:

$$y = f(w \cdot x + \theta)$$

,όπου y είναι ένα διάνυσμα στήλη μήκους ίσου με τον αριθμό των νευρώνων (M) στο παρόν στρώμα, x είναι ένα διάνυσμα στήλη με μήκος ίσο με τον αριθμό των εισόδων από το προηγούμενο επίπεδο, w ένας $N \times M$ πίνακας με τα συναπτικά βάρη και θ ένα διάνυσμα στήλη μήκους M.



Εικόνα 17 Τυπικός Νευρώνας Πολυεπίπεδου Νευρωνικού Δικτύου

Η ύπαρξη κρυφών επιπέδων επιτρέπει στο δίκτυο να ανιχνεύσει σύνθετα χαρακτηριστικά. Αλλά κυρίως η συμπεριφορά των κρυφών αυτών επιπέδων μαθαίνεται αυτόματα, δεν προγραμματίζεται από πριν. Όταν ένα νευρωνικό δίκτυο χρησιμοποιείται για την επίλυση ενός προβλήματος, το δίκτυο μαθαίνει την σχέση εισόδου – εξόδου και είναι σε θέση να γενικεύσει την έξοδο όταν εκτίθεται σε εισόδους τις οποίες δεν έχει δει ακόμα. Σύμφωνα με τους [Mendel et al., 1970]. , με την κατάλληλη τοπολογία του δικτύου και την κατάλληλη συνάρτηση ενεργοποίησης, ένα τεχνητό νευρωνικό δίκτυο μπορεί να προσεγγίσει κάθε συνάρτηση αρκετά καλά.

4.7 Εκπαίδευση Τεχνητών Νευρωνικών Δικτύων:

Τα τεχνητά νευρωνικά δίκτυα έχουν την ικανότητα να μαθαίνουν από το περιβάλλον και να βελτιώνουν την απόδοσή τους μέσω μάθησης. Με τον όρο *μάθηση* αναφερόμαστε στη επαναληπτική διαδικασία μεταβολής των συναπτικών βαρών του δικτύου με σκοπό την επίτευξη μια συγκεκριμένης επιθυμητής συμπεριφοράς. Ανάλογα με την λειτουργία που έχει να επιτελέσει το νευρωνικό δίκτυο επιλέγουμε και με πια συνάρτηση μάθησης επιθυμούμε να λειτουργήσει, αν και συνηθίζεται να δοκιμάζουμε την επίλυση ενός προβλήματος με περισσότερους αλγορίθμους μάθησης προκειμένου να καταλήξουμε σε αυτόν που μας δίνει καλύτερα αποτελέσματα.

Σύμφωνα με τους [Mendel et al., 1970] « Μάθηση είναι μια διαδικασία με την οποία προσαρμόζονται οι ελεύθερες παράμετροι ενός νευρωνικού δικτύου μέσω μιας συνεχούς διαδικασίας διέγερσης από το περιβάλλον στο οποίο βρίσκεται το δίκτυο. Το είδος της μάθησης καθορίζεται από τον τρόπο με τον οποίο πραγματοποιούνται οι αλλαγές των παραμέτρων.

Οι αλγόριθμοι μάθησης χωρίζονται σε τρείς βασικές κατηγορίες:

- Επιβλεπόμενη Μάθηση (Supervised Learning)
- Μη επιβλεπόμενη ή Αυτό-οργανούμενη Μάθηση (Self-organised /Unsupervised Learning)
- Ενισχυτική Μάθηση (Reinforcement Learning)

Κατά την επιβλεπόμενη μάθηση ο αλγόριθμος προσπαθεί να απεικονίσει τις δοσμένες εισόδους (σύνολο εκπαίδευσης - training set) στις επιθυμητές τιμές εξόδους μεταβάλλοντας τα συναπτικά βάρη του δικτύου, με απώτερο σκοπό τη γενίκευση της συνάρτησης αυτής και για εισόδους για τις οποίες δεν γνωρίζει την έξοδο. Η διαδικασία αυτή επαναλαμβάνεται για όλο το μέγεθος της εισόδου μέχρι την ελαχιστοποίηση του σφάλματος της εξόδου του δικτύου προς την επιθυμητή έξοδο. Για την τροποποίηση των συναπτικών βαρών χρησιμοποιούνται συνήθως οι ακόλουθες μέθοδοι:

- Μέθοδος Κατιούσας Κλίσης (Gradient Descend)
- Μέθοδος Newton
- Μέθοδος Μεγίστης Κλίσης (Steepest Descend)
- Κανόνας Polak – Ribiere
- Αλγόριθμος Ανάστροφης Μετάδοσης Λάθους (back propagation)

Στην μη επιβλεπόμενη μάθηση κατασκευάζεται ένα μοντέλο για κάποιο σύνολο εισόδων χωρίς όμως να γνωρίζουμε τις επιθυμητές εξόδους. Το δίκτυο αυτοδιοργανώνεται μόνο του, υπό την προϋπόθεση όμως ότι παρέχεται αρκετά μεγάλος αριθμός δεδομένων εισόδου, τα οποία έχουν επιλεγεί κατάλληλα έτσι ώστε να περιέχουν χαρακτηριστικά που να περιγράφουν τις ξεχωριστές ιδιότητες των αντικειμένων τα οποία επιθυμούμε να ταξινομήσουμε σε κλάσεις. Οι μέθοδοι που χρησιμοποιούνται για αυτό το είδος μάθησης είναι:

- Κανόνας του Hebb
- Ανταγωνιστικός Νόμος
- Διαφορικός νόμος του Hebb
- Διαφορικός Ανταγωνιστικός Νόμος

Τέλος στην ενισχυτική μάθηση, ο αλγόριθμος ακολουθεί μια συγκεκριμένη στρατηγική ενεργειών για μία συγκεκριμένη παρατήρηση.

4.7.1 Backpropagation

Ένας τρόπος εκπαίδευσης ενός πολυεπίπεδου δικτύου πρόσθιας τροφοδότησης είναι η χρησιμοποίηση του αλγορίθμου εκπαίδευσης backpropagation. Ο αλγόριθμος αυτός προτάθηκε από τους Rumelhart et al. και θεωρείται σήμερα ο πιο ευρέως χρησιμοποιούμενος αλγόριθμος για επιβλεπόμενη μάθηση σε πολυεπίπεδα νευρωνικά δίκτυα. Ο στόχος ενός τέτοιου αλγορίθμου είναι να εκπαιδεύσει το δίκτυο να αντιστοιχίζει συγκεκριμένα πρότυπα εισόδου στα αντίστοιχα πρότυπα εξόδου ή σε αντίστοιχες τιμές εξόδου, μέσω της μεταβολής των συναπτικών βαρών προκειμένου να επιτευχθεί η ελαχιστοποίηση του σφάλματος μεταξύ της επιθυμητής εξόδου και της εξόδου που δίνει το δίκτυο. Για την βελτιστοποίηση αυτή χρησιμοποιείται συνήθως ένας αλγόριθμος κατιούσας κλίσης (gradient descend). Οι κόμβοι σε ένα δίκτυο ανάστροφης τροφοδότησης απαιτείται να έχουν μια μονότονα αύξουσα παραγωγίσημη συνάρτηση ενεργοποίησης. Συνήθως χρησιμοποιείται μια σιγμοείδης συνάρτηση για αυτόν τον σκοπό, η οποία εξασφαλίζει μια συνεχή έξοδο με τιμές μεταξύ 0 και 1.

Στη λειτουργία της μάθησης με τον αλγόριθμο backpropagation μπορούμε να διακρίνουμε δύο φάσεις: την πρόσθια φάση και την οπίσθια φάση. κατά την πρόσθια φάση τα σήματα εισόδου διαδίδονται μέσα στο δίκτυο περνώντας από το ένα επίπεδο στο επόμενο, παράγοντας έτσι μια έξοδο, η οποία συγκρίνεται με την επιθυμητή έξοδο. Τα σήματα σφάλματος που προκύπτουν από την διαφορά της επιθυμητής από την πραγματική έξοδο, μεταδίδεται στο δίκτυο προς τα πίσω, και αυτή η φάση αποτελεί την οπίσθια φάση του αλγορίθμου.

Με μαθηματικούς όρους, αν συμβολίσουμε την επιθυμητή έξοδο του νευρώνα εξόδου j $d_j(n)$, την πραγματική του έξοδο $y_j(n)$ και το σήμα σφάλματος $e_j(n)$ κατά την παρουσίαση του n προτύπου εκπαίδευσης, τότε έχουμε:

$$e_j(n) = d_j(n) - y_j(n) \qquad (1)$$

Ορίζουμε την στιγμιαία τιμή του τετραγωνικού σφάλματος του νευρώνα j ως $\frac{1}{2}e_j^2(n)$.
Αντίστοιχα, η στιγμιαία τιμή $\mathcal{E}(n)$ του αθροίσματος των τετραγωνικών σφαλμάτων εξασφαλίζεται αθροίζοντας τα $\frac{1}{2}e_j^2(n)$ όλων των νευρώνων του στρώματος εξόδου του δικτύου:

$$\mathcal{E}(n) = \frac{1}{2}\sum_{j\in C}\frac{1}{2}e_j^2(n) \quad (2)$$

,όπου C το σύνολο που περιλαμβάνει όλους τους νευρώνες του στρώματος εξόδου. Αν N είναι ο συνολικός αριθμός των προτύπων εισόδου που περιέχονται στο σύνολο εκπαίδευσης, το Μέσο Τετραγωνικό Σφάλμα βρίσκεται αθροίζοντας όλα τα $\mathcal{E}(n)$ για όλα τα n κανονικοποιώντας τελικά με το μέγεθος του συνόλου C:

$$\mathcal{E}_{av} = \frac{1}{N}\sum_{n=1}^{N}\mathcal{E}(n) \quad (3)$$

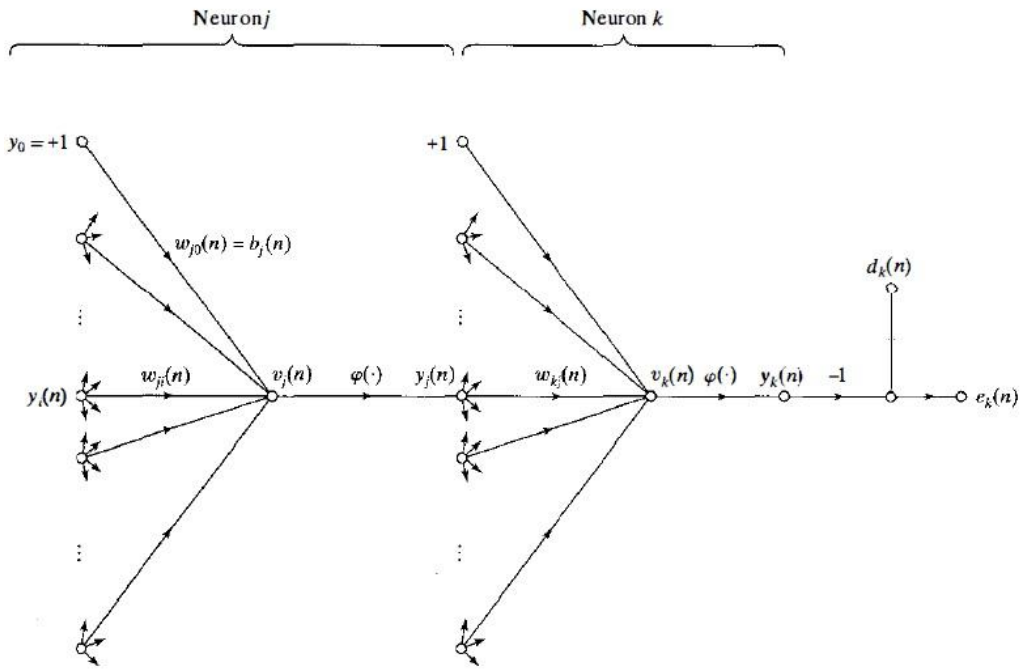
Το $\mathcal{E}_{av}$ είναι συνάρτηση όλων των ελεύθερων παραμέτρων του δικτύου (συναπτικά βάρη και κατώφλια) και για ένα δεδομένο σύνολο εκπαίδευσης αποτελεί την Συνάρτηση Κόστους σαν μέτρο της επίδοσης μάθησης του συνόλου. Ο σκοπός της διαδικασίας μάθησης είναι η προσαρμογή των ελεύθερων παραμέτρων έτσι ώστε να ελαχιστοποιηθεί το $\mathcal{E}_{av}$.

Στη συνέχεια της ανάλυσης της διαδικασίας εκπαίδευσης υποθέτουμε ότι τα συναπτικά βάρη ανανεώνονται ανάλογα με το σχετικό σφάλμα που υπολογίζεται μετά την παρουσίαση κάθε προτύπου. Συνεπώς ο αριθμητικός μέσος αυτών των μεταβολών των βαρών για όλο το σύνολο εκπαίδευσης αποτελεί μια εκτίμηση της πραγματικής αλλαγής που θα προέκυπτε αν μεταβάλλαμε τα συναπτικά βάρη βασιζόμενοι στην ελαχιστοποίηση της συνάρτησης κόστους $\mathcal{E}_{av}$ πάνω σε όλο το σύνολο εκπαίδευσης.
Στην εικόνα που ακολουθεί αναπαρίσταται ένας νευρώνας j ο οποίος λαμβάνει σήματα από το επίπεδο νευρώνων που βρίσκεται στα αριστερά του. Η net internal activity $v_j(n)$ που εμφανίζεται στην είσοδο της συνάρτησης ενεργοποίησης εκφράζεται ως:

$$v_j(n) = \sum_{i=0}^{N} w_{ji}(n)y_i(n) \quad (4)$$

,όπου N ο συνολικός αριθμός των εισόδων που εφαρμόζονται στον νευρώνα j. Το συναπτικό βάρος w_{j0} που αντιστοιχεί στη σταθερή είσοδο $y_0 = -1$ ισούται με την τιμή του κατωφλιού θ_j που εφαρμόζεται στον νευρώνα αυτόν. Συνεπώς, η έξοδος του νευρώνα j στην επανάληψη n θα δίνεται από την ακόλουθη σχέση:

$$y_j(n) = \varphi_j\left(v_j(n)\right) \quad (5)$$



Εικόνα 18 Απεικόνιση της ροής των σημάτων της σύνδεσης του νευρώνα k του στρώματος εξόδου με τον νευρώνα j του κρυφού επιπέδου

Ο αλγόριθμος της όπισθεν διάδοσης επιβάλει μια διόρθωση $\Delta w_{ji}(n)$ στο συναπτικό βάρος $w_{ji}(n)$, η οποία είναι ανάλογη με τη στιγμιαία κλίση $\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)}$, η οποία εφαρμόζοντας τον κανόνα της αλυσίδας εκφράζεται ως:

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \qquad (6)$$

Η κλίση αυτή αποτελεί έναν παράγοντα ευαισθησίας, καθορίζοντας την κατεύθυνση της αναζήτησης στον χώρο των βαρών για το συναπτικό βάρος w_{ji}.
Παραγωγίζοντας και τις δύο πλευρές τις εξίσωσης (2) ως προς $e_j(n)$, έχουμε:

$$\frac{\partial \mathcal{E}(n)}{\partial e_j(n)} = e_j(n) \qquad (7)$$

Παραγωγίζοντας τις δύο πλευρές της εξίσωσης (1) ως προς $y_j(n)$ παίρνουμε:

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \qquad (8)$$

Διαφορίζοντας την εξίσωση (5) ως προς $v_j(n)$, έχουμε:

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi_j'(v_j(n)) \qquad (9)$$

, όπου ο τονισμός της συνάρτησης φ ερμηνεύεται ως διαφορισμός ως προς τη μεταβλητή της.
Τέλος παραγωγίζοντας την (4) ως προς $w_{ji}(n)$ παίρνουμε:

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n) \quad (10)$$

Έτσι χρησιμοποιώντας τις σχέσεις (7) και (10) στην εξίσωση (6) προκύπτει:

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} = -e_j(n)\varphi_j'\big(v_j(n)\big)y_i(n) \quad (11)$$

Η διόρθωση $\Delta w_{ji}(n)$ που θα εφαρμοστεί στο βάρος $w_{ji}(n)$ δίνεται από τον *Κανόνα Δέλτα*:

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{ji}(n)} \quad (12)$$

, όπου η είναι η παράμετρος του ρυθμού μάθησης του αλγορίθμου.
Η χρήση του πρόσημου μείον στην ανωτέρω σχέση ευθύνεται για την *Κάθοδο Κλίσης* στο χώρο των βαρών.
Η χρήση της σχέσης (11) στην (12) οδηγεί στην:

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \quad (13)$$

,όπου η *τοπική κλίση* $\delta_j(n)$ ορίζεται από την σχέση:

$$\delta_j(n) = -\frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \quad (14)$$

$$= e_j(n)\, \varphi_j'\big(v_j(n)\big)$$

Η τοπική κλίση δείχνει τις απαιτούμενες αλλαγές των συναπτικών βαρών. Στη σχέση (14) η τοπική κλίση για τον νευρώνα εξόδου j ισούται με το γινόμενο του σήματος σφάλματος $e_j(n)$ με την παράγωγο $\varphi_j'\big(v_j(n)\big)$.

Αξίζει να σημειωθεί μελετώντας τις εξισώσεις (13) και (14) ότι ένας βασικός παράγοντας για τον υπολογισμό της προσαρμογής του βάρους $\Delta w_{ji}(n)$ είναι το σήμα σφάλματος $e_j(n)$ στην έξοδο του νευρώνα j. Σε αυτό το σημείο μπορούμε να διαχωρίσουμε δύο περιπτώσεις, ανάλογα με τη θέση του νευρώνα j μέσα στο δίκτυο:

- <u>Ο νευρώνας j βρίσκεται στο επίπεδο εξόδου</u> :

  Για τον νευρώνα αυτόν γνωρίζουμε την επιθυμητή του απόκριση, συνεπώς μπορούμε να χρησιμοποιήσουμε την εξίσωση (1) για να υπολογίσουμε το σήμα σφάλματος και στη συνέχεια μέσω της σχέσης (14) να υπολογίσουμε την τοπική κλίση.
- <u>Ο νευρώνας j βρίσκεται σε κρυφό επίπεδο</u> :

  Για τον νευρώνα αυτόν δεν υπάρχει συγκεκριμένη επιθυμητή απόκριση. Συνεπώς το σήμα σφάλματος του νευρώνα αυτού θα πρέπει να υπολογιστεί αναδρομικά σε σχέση με τα σήματα σφάλματος όλων των νευρώνων με τους οποίους συνδέεται ο νευρώνας αυτός.

  Στην εικόνα που ακολουθεί απεικονίζεται ο νευρώνας j ως κρυφός και ένας νευρώνας k ως νευρώνας εξόδου.

Από την εξίσωση (14) μπορούμε να επαναορίσουμε την τοπική κλίση ως:

$$\delta_j(n) = -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)}\frac{\partial y_j(n)}{\partial v_j(n)}$$

$$= -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)}\,\varphi_j'\big(v_j(n)\big) \quad \text{, o j είναι} \qquad (15)$$

κρυφός

, όπου στην δεύτερη γραμμή χρησιμοποιήθηκε η εξίσωση (9).

Για να υπολογίσουμε τη μερική παράγωγο $\frac{\partial \mathcal{E}(n)}{\partial y_j(n)}$, παραγωγίζουμε την εξίσωση (2) - εκφρασμένη για τον νευρώνα k , τον νευρώνα εξόδου – ως προς $y_j(n)$ και έχουμε:

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial v_k(n)}\frac{\partial v_k(n)}{\partial y_j(n)} \qquad (16)$$

Στην την προηγούμενη εικόνα παρατηρούμε:

$$e_k(n) = d_k(n) - y_k(n) = d_k(n) - \varphi_k(v_k(n)) \qquad (17)$$

Και

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -\varphi_k'(v_k(n)) \qquad (18)$$

Παρατηρούμε ακόμα στην εικόνα ότι η net internal activity level του νευρώνα k είναι

$$v_k(n) = \sum_{i=0}^{P} w_{kj}(n) y_j(n) \qquad (19)$$

,όπου P ο συνολικός αριθμός των εισόδων που εφαρμόζονται στον νευρώνα k. Διαφορίζοντας την (19) ως προς $y_j(n)$ έχουμε:

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n) \qquad (20)$$

Έτσι χρησιμοποιώντας τις σχέσεις (18) και (20) στην (16) καταλήγουμε στην επιθυμητή μερική παράγωγο:

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = -\sum_k e_k(n)\,\varphi_k'(v_k(n))\,w_{kj}(n) \qquad (21)$$

$$= -\sum_k \delta_k(n)\,w_{kj}(n)$$

Τέλος χρησιμοποιώντας τις εξισώσεις (15) και (21) υπολογίζουμε την τοπική κλίση $\delta_j(n)$ του κρυφού νευρώνα j:

$$\delta_j(n) = \varphi_j'\big(v_j(n)\big)\sum_k \delta_k(n)\,w_{kj}(n) \qquad (22)$$

Συνεπώς για την διόρθωση $\Delta w_{ji}(n)$ που εφαρμόζεται στο συναπτικό βάρος που ενώνει τον νευρώνα i με τον νευρώνα j χρησιμοποιούμε τον κανόνα δέλτα:

$$\begin{pmatrix} Διώρθωση \\ βάρους \\ \Delta w_{ji}(n) \end{pmatrix} = \begin{pmatrix} Ρυθμός \\ Μάθησης \\ \eta \end{pmatrix} \cdot \begin{pmatrix} Τόπική \\ Κλίση \\ \delta_j(n) \end{pmatrix} \cdot \begin{pmatrix} Είσοδος \\ νευρώνα\ j \\ y_i(n) \end{pmatrix} \quad (23)$$

Επιπλέον η τοπική κλίση $\delta_j(n)$ εξαρτάται από τη θέση του νευρώνα μέσα στο δίκτυο:

- Αν ο j ανήκει στο στρώμα εξόδου, τότε η τοπική κλίση δίνεται από την σχέση (14)
- Αν ο j ανήκει σε κρυφό επίπεδο, τότε η τοπική κλίση εξαρτάται από την $\varphi_j'(v_j(n))$ και το σταθμισμένο άθροισμα των τοπικών κλίσεων που έχουν υπολογιστεί για τους νευρώνες που βρίσκονται στο επόμενο επίπεδο και συνδέονται με τον j (σχέση (22)).

Συνεπώς κατά την εφαρμογή του αλγορίθμου της όπισθεν διάδοσης έχουμε δύο περάσματα υπολογισμών στο δίκτυο:

- Το *εμπρός πέρασμα* όπου τα συναπτικά βάρη μένουν αμετάβλητα μέσα στο δίκτυο και υπολογίζονται τα σήματα εξόδου του κάθε νευρώνα ανά επίπεδο, μέχρι ο υπολογισμός να φτάσει στο επίπεδο εξόδου, όπου βάση της επιθυμητής εξόδου υπολογίζονται τα σήματα σφάλματος.
- Το *πίσω πέρασμα* όπου τα σήματα σφάλματος διαδίδονται από το στρώμα εξόδου προς τα προηγούμενα στρώματα προκειμένου να υπολογιστούν αναδρομικά οι τοπικές κλίσεις για κάθε νευρώνα. Η αναδρομική αυτή διαδικασία επιτρέπει στα συναπτικά βάρη του δικτύου να υποστούν αλλαγές βάση του κανόνα δέλτα όπως αυτός περιγράφεται στη σχέση (23)

4.8 Κανόνες Μάθησης

Μάθηση Με Διόρθωση Σφάλματος (Error – Correction Learning):

Σύμφωνα με αυτόν τον τύπο μάθησης κάθε φορά που παρουσιάζεται στο δίκτυο μια είσοδος κατά τη διάρκεια της εκπαίδευσης, γίνεται σύγκριση της τιμής που επιστρέφει το δίκτυο ως έξοδο με την επιθυμητή έξοδο και στη συνέχεια τα συναπτικά βάρη μεταβάλλονται προς την κατεύθυνση εκείνη που θα οδηγήσει σε μείωση του σφάλματος μεταξύ της πραγματικής και της επιθυμητής εξόδου.

Αναλυτικότερα, αν $y_k(n)$ η πραγματική έξοδος ενός νευρώνα k και $d_k(n)$ η επιθυμητή έξοδος, τότε το σφάλμα ορίζεται από την σχέση:

$$e_k(n) = d_k(n) - y_k(n)$$

Ο τελικός στόχος της μάθησης με διόρθωση σφάλματος είναι η ελαχιστοποίηση μιας συνάρτησης κόστους βασισμένη πάνω στο προαναφερθέν σφάλμα. Ένα κριτήριο που χρησιμοποιείται αρκετά συχνά είναι το κριτήριο Ελαχίστων Τετραγώνων (Mean – Squared – Error Criterion):

$$J = E\left[\frac{1}{2}\sum_{k} e_k^2(n)\right]$$

,όπου η άθροιση γίνεται πάνω σε όλους τους νευρώνες του επιπέδου εξόδου του δικτύου και Ε είναι ο τελεστής της στατιστικής πιθανότητας. Ελαχιστοποιώντας την ανωτέρω συνάρτηση κόστους J σε σχέση με τις παραμέτρους του δικτύου οδηγεί στη γνωστή ***Μέθοδο Κατιούσας Κλίσης (Gradient Descend).***

Επειδή όμως δεν είμαστε σε θέση να γνωρίζουμε τα στατιστικά χαρακτηριστικά των υποκείμενων διαδικασιών της J ,χρησιμοποιούμε την ακόλουθη προσεγγιστική λύση στο πρόβλημα βελτιστοποίησης, υπολογίζοντας το στιγμιαίο κριτήριο ελαχίστων τετραγώνων:

$$\mathcal{E}(n) = \frac{1}{2}\sum_{k} e_k^2(n)$$

Συνεπώς το δίκτυο βελτιστοποιείται ελαχιστοποιώντας την $\mathcal{E}(n)$ σε σχέση με τα συναπτικά βάρη του δικτύου. Έτσι σύμφωνα με τον κανόνα διόρθωσης σφάλματος (ή κανόνας δέλτα (delta rule)), η προσαρμογή $\Delta w_{kj}(n)$ που γίνεται στο συναπτικό βάρος w_{kj} τη χρονική στιγμή n δίνεται από τη σχέση:

$$\Delta w_{kj} = \eta\, e_k(n) x_j(n)$$

,όπου *η* είναι μια θετική σταθερά που ονομάζεται *ρυθμός μάθησης (rate of learning).*
Συνεπώς η νέα τιμή του συνοπτικού βάρους τη χρονική στιγμή (n+1) γράφεται:

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n)$$

Με την μέθοδο αυτή το δίκτυο λειτουργεί σαν σύστημα κλειστού βρόχου και θα πρέπει να δοθεί αρκετή βαρύτητα στη σωστή επιλογή του ρυθμού μάθησης η, αφού ο ρυθμός μάθησης έχει σημαντική επίδραση στην επίδοση του συστήματος. Αν το η είναι μικρό, τότε η διαδικασία μάθησης εξελίσσεται ομαλά, αλλά μπορεί να χρειαστεί αρκετός χρόνος για το σύστημα για να συγκλίνει σε μια ευσταθή λύση. Από την άλλη, όταν το η είναι μεγάλο, ο ρυθμός μάθησης επιταχύνεται αλλά υποβόσκει ο κίνδυνος ότι η διαδικασία μάθησης δεν θα συγκλίνει και το σύστημα θα είναι ασταθές.

<u>Χεμπιανή Μάθηση (Hebbian Learning):</u>

Το αξίωμα του Hebb αποτελεί τον γνωστότερο κανόνα μάθησης:

Όταν ένας άξονας του κυττάρου Α είναι αρκετά κοντα έτσι ώστε να διεγείρει ένα κύτταρο Β και το πυροδοτεί επανειλημμένα ή σταθερά, κάποια διαδικασία ανάπτυξης ή μεταβολική αλλάγη συμβαίνει σε ένα ή και στα δύο κύτταρα, έτσι ώστε επιτυγχάνεται η αύξηση της αποτελεσματικότητας του Α στην πυροδότητση του κυττάρου Β.

Η Χεμπιανή εκπαίδευση υιοθετεί την παραπάνω υπόθεση και συνεπώς όταν δύο νευρώνες ενεργοποιούνται ταυτόχρονα, το βάρος της μεταξύ τους σύναψης αυξάνει, ενώ αντίθετα όταν ενεργοποιούνται ασύγχρονα το βάρος τους μειώνεται ή εξαλείφεται τελείως.
Μια τέτοια σύναψη αποκαλείται Χεμπιανή Σύναψη και ορίζεται ως η σύναψη εκείνη η οποία χρησιμοποιεί έναν χρονοεξαρτώμενο, τοπικό και ισχυρά αλληλεπιδραστικό μηχανισμό για να αυξήσει την συναπτική αποτελεσματικότητα σαν συνάρτηση της συσχέτισης μεταξύ της πρόσυναπτικής και της μετα-συναπτικής δραστηριότητας.

Με μαθηματικούς όρους η μεταβολή του βάρους w_{kj} με ροποσυναπτική και μετασυναπτική δραστηριότητα x_j και y_k αντίστοιχα εκφράζεται ως:

$$\Delta w_{kj}(n) = F\big(y_k(n), x_k(n)\big)$$

, όπου $F(\cdot,\cdot)$ είναι μια συνάρτηση της προσυναπτικής και της μετασυναπτικής δραστηριότητας. Η πιο απλή εκδοχή της ανωτέρω εξίσωσης είναι η : $\Delta w_{kj}(n) = \eta y_k(n) x_j(n)$
,όπου η είναι μια θετική σταθερά που ονομάζεται ρυθμός μάθησης. Ο ανωτέρω κανόνας μεταβολής αναφέρεται συχνά ως Κανόνας Γινομένου Δραστηριότητας (Activity Product Rule).

Ανταγωνιστική Μάθηση (Competitive Learning)

Στην ανταγωνιστική μάθηση, οι νευρώνες του στρώματος εξόδου ενός δικτύου ανταγωνίζονται μεταξύ τους για το ποιος θα είναι ο μοναδικός που θα ενεργοποιηθεί. Σε ένα ανταγωνιστικό δίκτυο μόνο ένας νευρώνας εξόδου επιτρέπεται να είναι ενεργός σε κάθε χρονική στιγμή, ευνοώντας έτσι την χρησιμοποίηση του δικτύου για ανεύρεση εξεχόντων στατιστικών χαρακτηριστικών που μπορεί να χρησιμοποιηθούν για ταξινόμηση ενός συνόλου προτύπων.
Υπάρχουν τρία βασικά στοιχεία σε έναν κανόνα ανταγωνιστικής μάθησης:

- Ένα σύνολο από νευρώνες που είναι όλοι ίδιοι εκτός από κάποια τυχαία κατανεμημένα συναπτικά βάρη, που έχουν σαν αποτέλεσμα τη διαφορετική απόκριση σε ένα δοσμένο σύνολο από πρότυπα εισόδου
- Ένα όριο που καθορίζει τη «δύναμη» του κάθε νευρώνα
- Έναν μηχανισμό που επιτρέπει στους νευρώνες να ανταγωνίζονται για το δικαίωμα να αποκριθούν για ένα δεδομένο υποσύνολο εισόδων, έτσι ώστε μόνο ένας νευρώνας εξόδου να είναι ενεργός τη κάθε χρονική στιγμή.

Ο απλούστερος κανόνας ανταγωνιστικής εκπαίδευσης είναι:

$$\Delta w_{ji} = \begin{cases} \eta\,(x_i - w_{ji}) & ,\text{αν ο νευρώνας } j \text{ κερδίσει τον ανταγωνισμό} \\ 0 & ,\text{αν ο νευρώνας } j \text{ χάσει} \end{cases}$$

Μάθηση Boltzmann

Ο κανόνας Μάθησης Boltzmann αποτελεί έναν στοχαστικό αλγόριθμο μάθησης που προέκυψε από θερμοδυναμικές θεωρήσεις. Σε μια μηχανή Boltzmann, οι νευρώνες συνθέτουν μια αναδρομική δομή και λειτουργούν δυαδικά, είναι δηλαδή είτε ενεργοί είτε όχι. Η μηχανή χαρακτηρίζεται από μια συνάρτηση ενέργειας Ε, η τιμή της οποίας υπολογίζεται από τις επιμέρους καταστάσεις των νευρώνων και εκφράζεται ως:

$$E = -\frac{1}{2}\sum_{i}\sum_{\substack{j \\ i \neq j}} w_{ji} s_j s_i$$

,όπου s_iη κατάσταση του νευρώνα i, w_{ji} το συναπτικό βάρος μεταξύ των νευρώνων i και j.

Η μηχανή επιλέγει έναν νευρώνα τυχαία, έστω τον j, σε κάποιο βήμα της διαδικασίας μάθησης, και αλλάζει τον νευρώνα αυτόν στη δυαδική κατάσταση του σε κάποια θερμοκρασία Τ με πιθανότητα

$$W(s_j \to -s_j) = \frac{1}{1 + e^{-\frac{\Delta E_j}{T}}}$$

,όπου ΔΕ η μεταβολή της ενέργειας που προκύπτει από την αλλαγή αυτή.

ΚΕΦΑΛΑΙΟ 5:

Μηχανές Διανυσμάτων Υποστήριξης – Support Vector Machines (SVM)

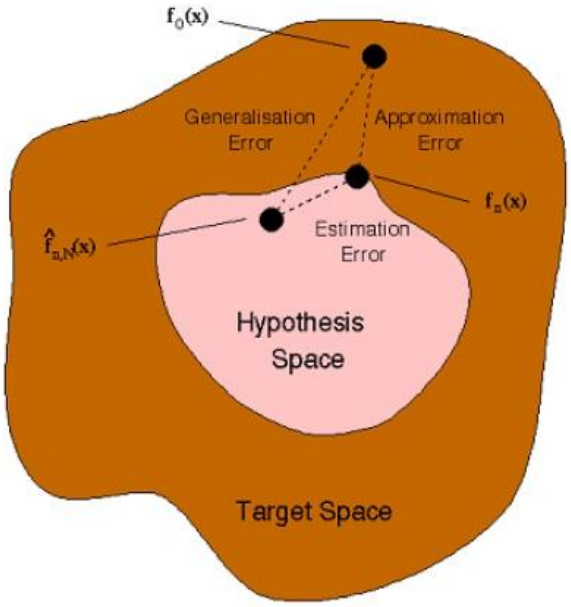
5.1 Εισαγωγή

Τα θεμέλια των Μηχανών Διανυσμάτων Υποστήριξης (SVM) ετέθησαν από τον Vladimir Vapnik το 1995 και κερδίζουν δημοτικότητα λόγω των ελκυστικών χαρακτηριστικών και επιδόσεων που παρουσιάζουν. Η διατύπωση τους εμπεριέχει την Αρχή Ελαχιστοποίησης του Δομικού Κινδύνου (Structural Risk Minimization – SRM), η οποία έχει αποδειχθεί ως ανώτερη σε σχέση με την Αρχή Ελαχιστοποίησης του Εμπειρικού Κινδύνου (Empirical Risk Minimization – ERM), που χρησιμοποιείται από τα συμβατικά νευρωνικά δίκτυα. Η SRM ελαχιστοποιεί ένα άνω όριο του αναμενόμενου κινδύνου, σε αντίθεση με την ERM που ελαχιστοποιεί το σφάλμα στα δεδομένα εκπαίδευσης. Αυτή η διαφορά καθιστά τα SVM πιο ικανά στη γενίκευση.

5.2 Θεωρία Στατιστικής Μάθησης (Statistical Learning Theory)

Στο κεφάλαιο αυτό επιχειρούμε μια σύντομη εισαγωγή στις αρχές της στατιστικής μάθησης, προκειμένου να γίνει στη συνέχεια κατανοητή η θεωρία των SVM.
Ο στόχος μιας μοντελοποίησης είναι η επιλογή ενός μοντέλου από τον χώρο υπόθεσης, το οποίο να βρίσκεται κοντύτερα (αναφορικά με κάποιο μέτρο σφάλματος) στην υποκείμενη συνάρτηση στον χώρο στόχου. Τα σφάλματα σε αυτή τη διαδικασία προκύπτουν από δύο περιπτώσεις:

- ***Σφάλμα Προσέγγισης (Approximation Error)***: αποτελεί συνέπεια του μικρότερου μεγέθους του χώρου υπόθεσης σε σχέση με το χώρο στόχο και ως εκ τούτου η υποκείμενη συνάρτηση μπορεί να κείται έξω από τον χώρο υπόθεσης. Μια κακή επιλογή του χώρου μοντελοποίησης μπορεί να οδηγήσει σε μεγάλο σφάλμα προσέγγισης και αποκαλείται ασυμφωνία μοντέλου.
- ***Σφάλμα Εκτίμησης (Estimation Error)***: αποτελεί το σφάλμα που οφείλεται στην διαδικασία μάθησης, η οποία οδηγεί σε μια τεχνική επιλογής του μη βέλτιστου μοντέλου από τον χώρο υπόθεσης.



Εικόνα 18 Σφάλμα Προσέγγισης και Σφάλμα Εκτίμησης

Σε συνδυασμό αυτά τα δύο σφάλματα συγκροτούν το σφάλμα γενίκευσης.
Κατά την κατασκευή του μοντέλου επιθυμούμε να βρούμε τη συνάρτηση f, η οποία θα ελαχιστοποιεί τον κίνδυνο:

$$R[f] = \int_{X\times Y} L\big(y, f(x)\big)\, P(x,y)\, dx\, dy \tag{5.1}$$

Όμως το $P(x,y)$ είναι άγνωστο και είναι δυνατόν να βρεθεί μια προσέγγιση σύμφωνα με την αρχή ERM:

$$R_{emp}[f] = \frac{1}{l}\sum_{i=1}^{l} L\big(y^i, f(x^i)\big) \tag{5.2}$$

η οποία ελαχιστοποιεί τον εμπειρικό κίνδυνο

$$\widehat{f_{n,l}}(x) = \arg\min_{f\in H_n} R_{emp}[f] \tag{5.3}$$

Η ελαχιστοποίηση του εμπειρικού κινδύνου έχει νόημα μόνο αν

$$\lim_{l\to\infty} R_{emp}[f] = R[f] \tag{5.4}$$

,το οποίο ισχύει σύμφωνα με τον νόμο των μεγάλων αριθμών. Ωστόσο, ο εμπειρικός κίνδυνος θα πρέπει να ικανοποιεί και την ακόλουθη σχέση:

$$\lim_{l\to\infty} \min_{f\in H_n} R_{emp}[f] = \min_{f\in H_n} R[f] \tag{5.5}$$

η οποία ισχύει όταν το σύνολο H_n είναι αρκετά μικρό.
Το ακόλουθο όριο ισχύει με πιθανότητα $1-\delta$:

$$R[f] \le R_{emp}[f] + \sqrt{\frac{h\, ln\left(\frac{2l}{h}+1\right) - ln\left(\frac{\delta}{4}\right)}{l}} \tag{5.6}$$

5.3 Η Διάσταση VC (VC Dimension)

Η διάσταση VC αποτελεί ένα βαθμωτό μέγεθος που μετράει την χωρητικότητα ενός συνόλου συναρτήσεων.

Σύμφωνα με τον ορισμό των Vapnik – Chervonenkis "η VC διάσταση ενός συνόλου συναρτήσεων ισούται με p αν και μόνο αν υπάρχει ένα σύνολο σημείων $\{x^i\}_{i=1}^{p}$ τέτοιο ώστε τα σημεία αυτά να μπορούν να χωριστούν σε 2^p πιθανές διατάξεις, και δεν υπάρχει κανένα σύνολο $\{x^i\}_{i=1}^{q}$ με $q > p$ το οποίο να ικανοποιεί αυτήν την ιδιότητα.

Η εικόνα που ακολουθεί δείχνει πως 3 σημεία πάνω σε μια επιφάνεια μπορούν να χωριστούν από ένα σύνολο γραμμικών συναρτήσεων, ενώ τέσσερα σημεία δεν μπορούν. Το σύνολο των γραμμικών συναρτήσεων στον n-διάστατο χώρο έχει VC διάσταση ίση με n+1.

5.4 Αρχή Ελαχιστοποίησης του Δομικού Κινδύνου (Structural Risk Minimization – SRM)

Αν κατασκευαστεί μια δομή ένθετων ομάδων συναρτήσεων $F_1 \subset F_2 \subset \cdots \subset F_\infty$, τέτοια ώστε F_h να είναι ο χώρος υπόθεσης με VC διάστασης ίση με h, τότε η αρχή προχωράει θεωρώντας τις $f_1, f_2, \dots, f_h$ ως λύσεις του ελαχιστοποιημένου εμπειρικού κινδύνου/σφάλματος στις ομάδες των συναρτήσεων F_i. Η αρχή SRM επιλέγει την ομάδα συναρτήσεων F_i η οποία ικανοποιεί το ακόλουθο πρόβλημα:

$$\min_{S_h} R_{emp}[f] + \sqrt{\frac{h\, ln\left(\frac{2l}{h}+1\right) - ln\left(\frac{\delta}{4}\right)}{l}} \qquad (5.7)$$

5.5 Ταξινομητές Υπερεπιπέδου

Έστω ότι έχουμε l παραδείγματα εκπαίδευσης $\{x_i, y_i\}, i = 1, \dots, l$, όπου κάθε παράδειγμα αποτελεί ένα διάνυσμα d διαστάσεων ($x_i \in \mathbb{R}^d$) καθώς και μια ετικέτα κλάσης με τιμές $y_i \in \{-1,1\}$. Όλα τα υπερεπίπεδα στο $\mathbb{R}^d$ παραμετροποιούνται με ένα διάνυσμα w και μια σταθερά b, και περιγράφονται από την εξίσωση:

$$\boldsymbol{w^T} \cdot \boldsymbol{x} + b = 0 \qquad (5.8)$$

Το διάνυσμα w αποτελεί το κάθετο διάνυσμα στο υπερεπίπεδο. Έχοντας το υπερεπίπεδο $(\boldsymbol{w}, b)$ το οποίο διαχωρίζει τα δεδομένα, οδηγούμαστε στην εξίσωση:

$$f(\boldsymbol{x}) = sign(\boldsymbol{w^T} \cdot \boldsymbol{x} + b) \qquad (5.9)$$

η οποία ταξινομεί σωστά τα δεδομένα εκπαίδευσης. Βέβαια, ένα δοσμένο υπερεπίπεδο που περιγράφεται από τα $(\boldsymbol{w}, b)$, μπορεί να περιγραφεί εξίσου από όλα τα ζεύγη $\{\lambda \boldsymbol{w}, \lambda b\}$ για $\lambda \in \mathbb{R}^+$. Συνεπώς ορίζουμε το K*ανονικό Υπερεπίπεδο (Canonical Hyperplane)* ως εκείνο το οποίο διαχωρίζει τα δεδομένα από το υπερεπίπεδο κατά μια απόσταση 1. Συνεπώς αναφερόμαστε σε υπερεπιπεδα που ικανοποιούν τις ακόλουθες ανισότητες

$$\boldsymbol{w^T} \cdot \boldsymbol{x_i} + b \geq +1 \text{ , όταν } y_i = +1 \qquad (5.10)$$

$$\boldsymbol{w^T} \cdot \boldsymbol{x_i} + b \leq -1 \text{ , όταν } y_i = -1 \qquad (5.11)$$

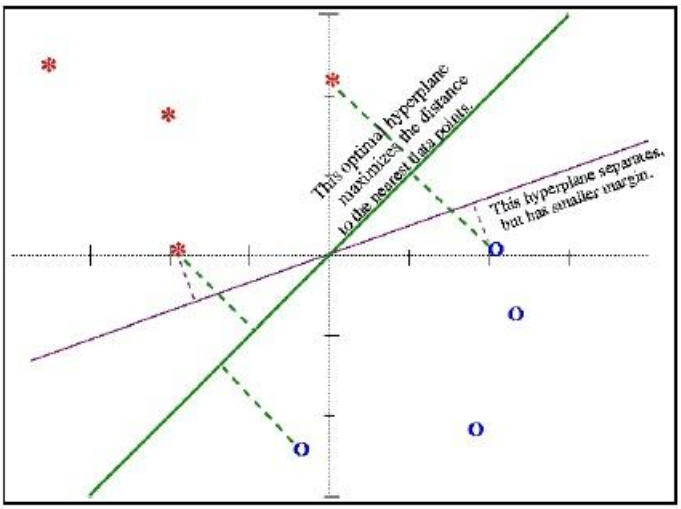
ή σε πιο συμπαγή μορφή:

$$y_i(\boldsymbol{w^T} \cdot \boldsymbol{x_i} + b) \geq +1 \quad \forall i \qquad (5.12)$$

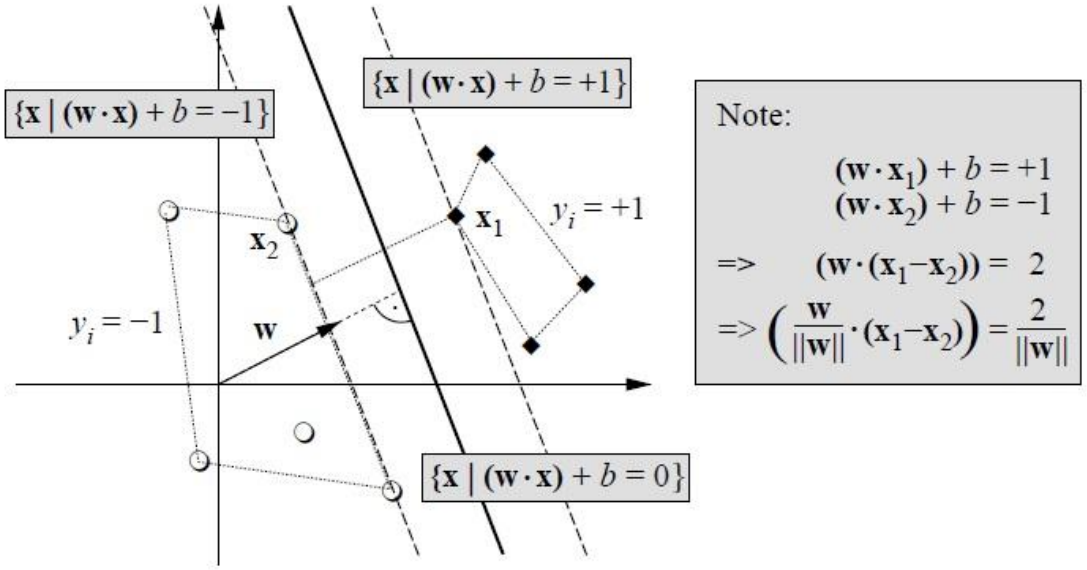
Όλα τα υπερεπίπεδα αυτά έχουν μια «λειτουργική απόσταση» ≥ 1. Δεν θα πρέπει όμως να συγχέεται με την γεωμετρική ή την Ευκλείδια απόσταση. Για ένα δοσμένο υπερεπίπεδο $(\boldsymbol{w}, b)$, όλα τα ζεύγη $\{\lambda \boldsymbol{w}, \lambda b\}$ ορίζουν ακριβώς την ίδια υπερεπιφάνεια, αλλά η κάθε μια έχει διαφορετική λειτουργική απόσταση από τα δεδομένα. Για να προσδιορίσουμε την γεωμετρική απόσταση από την υπερεπιφάνεια προς ένα σημείο δεδομένων έχουμε:

$$d\big((w, b), x_i\big) = \frac{y_i(\boldsymbol{w^T} \cdot \boldsymbol{x_i} + b)}{\|w\|} \qquad (5.13)$$

Διαισθητικά επιθυμούμε το υπερεπίπεδο εκείνο το οποίο μεγιστοποιεί τη γεωμετρική απόσταση των πιο κοντινών σε αυτό σημείων, όπως φαίνεται και στη ακόλουθη εικόνα.



Εικόνα 4 Το βέλτιστο υπερεπίπεδο που μεγιστοποιεί τις αποστάσεις των κοντινών σε αυτό σημείων (πράσινο) και ένα μη βέλτιστο υπερεπίπεδο (κόκκινο)



Εικόνα 20 Υπερεπίπεδο που διαχωρίζει τις δύο κλάσεις

Το γεωμετρικό εύρος του ταξινομητή ισούται με το μέγιστο μήκος της ζώνης που μπορεί να διαχωρίσει τις δύο κλάσεις, το οποίο ισούται με το διπλάσιο της ελάχιστης απόστασης (5.13) για όλα τα δεδομένα, όπως φαίνεται και στην Εικόνα . Συνεπώς το γεωμετρικό εύρος του ταξινομητή ισούται με:

$$margin = \frac{\mathbf{2}}{\|\boldsymbol{w}\|} \qquad (5.14)$$

Για να κατασκευάσουμε το βέλτιστο υπερεπίπεδο (εικόνα 19), θα πρέπει να μεγιστοποιήσουμε το γεωμετρικό εύρος, δηλαδή να λύσουμε το ακόλουθο πρόβλημα βελτιστοποίησης:

$$minimize \qquad \tau(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|^2 \qquad (5.15)$$

$$subject\ to\ \ y_i \cdot (\boldsymbol{w^T} \cdot \boldsymbol{x_i} + b)\ \geq 1\ , i = 1, \dots, l \qquad (5.16)$$

Παρατηρούμε ότι το γεωμετρικό εύρος του ταξινομητή είναι ανεξάρτητο της κλιμάκωσης των παραμέτρων w και b, αφού κανονικοποιείται με την νόρμα του w. Συνεπώς μπορούμε να προβούμε σε οποιαδήποτε κλιμάκωση του w χωρίς να επηρεάσουμε το γεωμετρικό εύρος. Έτσι απαιτούμε $\|w\| = 1$ και απαιτώντας όλα τα σημεία δεδομένων να έχουν απόσταση τουλάχιστον 1 καταλήγουμε στον περιορισμό (5.16) του προβλήματος που διατυπώσαμε ανωτέρω.

Το πρόβλημα αυτό της βελτιστοποίησης υπό περιορισμούς επιλύεται χρησιμοποιώντας πολλαπλασιαστές Lagrange $a_i \geq 0$ και την Λαγκραντζιανή:

$$L(\boldsymbol{w}, b, a) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{l} a_i\big(y_i\big((w^T \cdot x_i) + b\big) - 1\big) \qquad (5.17)$$

Η Λανγκρατζιανή L πρέπει να ελαχιστοποιηθεί ως προς τις πρωτογενείς μεταβλητές w και b και να μεγιστοποιηθεί ως προς τις δυαδικές μεταβλητές a_i (δηλαδή πρέπει να βρεθεί ένα σαγματικό σημείο).

Διαισθητικά βλέπουμε ότι αν παραβιαστεί η συνθήκη (5.16) , δηλαδή αν $y_i\big((x_i \cdot w) + b\big) - 1 < 0$, η L μπορεί να μεγιστοποιηθεί μεγιστοποιώντας τα αντίστοιχα a_i. Για να εμποδίσουμε το $-a_i\big(y_i\big((x_i \cdot w) + b\big) - 1\big)$ να γίνει αυθέρετα μεγάλο, η αλλάγη των w και b θα διασφαλίσει, ότι δεδομένου ότι το πρόβλημα είναι διαχωρίσιμο, ο περιορισμός τελικά θα ικανοποιηθεί. Παρόμοια μπορούμε να κατανοήσουμε ότι για όλους τους περιορισμούς οι οποίοι δεν ικανοποιούνται ως ισότητα αλλά ως ανισότητα, δηλαδή ισχύει $y_i\big((x_i \cdot w) + b\big) - 1 > 0$, τα αντίστοιχα a_i πρέπει να είναι 0, προκειμένου να μεγιστοποιηθεί η L. Το τελευταίο αποτελεί διατύπωση των συμπληρωματικών συνθηκών Karush-Kuhn_Tucker της θεωρίας βελτιστοποίησης.

Ο περιορισμός ότι στο σαγματικό σημείο οι παράγωγοι της L ως προς τις πρωτογενείς μεταβλητές πρέπει να μηδενίζεται, δηλαδή

$$\frac{\partial}{\partial b}L(\boldsymbol{w}, b, a) = 0 \ , \frac{\partial}{\partial \boldsymbol{w}}L(\boldsymbol{w}, b, a) = 0 \qquad (5.18)$$

οδηγεί στις ακόλουθες σχέσεις:

$$\sum_{i=1}^{l} a_i y_i = 0, \qquad (5.19)$$

$$w = \sum_{i=1}^{l} a_i y_i x_i \ , \qquad (5.20)$$

Το διάνυσμα λύσης έχει συνεπώς μια επέκταση όσον αφορά το υποσύνολο των προτύπων εκπαίδευσης, ήτοι τα πρότυπα των οποίων οι συντελεστές a_i είναι μη μηδενικοί, τα οποία αποκαλούνται Διανύσματα Υποστήριξης.

Από τις συμπληρωματικές συνθήκες Karush-Kuhn-Tucker

$$a_i\big(y_i\big((x_i{}^T \cdot w) + b\big) - 1\big) = 0 \ , \quad i = 1, \dots, l \qquad (5.21)$$

τα Διανύσματα Υποστήριξης κείτονται πάνω στο εύρος (margin) . Όλα τα υπόλοιπα παραδείγματα του συνόλου εκπαίδευσης είναι άνευ σημασίας, ο περιορισμός (5.16) που σχετίζεται με αυτά δεν συμβάλει καθόλου στην βελτιστοποίηση και συνεπώς δεν συμβάλουν στο **w** (5.20). Συνεπώς αυτό συνάδει με την διαίσθησή μας για το πρόβλημα: αφού το υπερεπίπεδο καθορίζεται πλήρως από τα πρότυπα που βρίσκονται πιο κοντά σε αυτό, η λύση δεν θα πρέπει να εξαρτάται από τα υπόλοιπα παραδείγματα.

Αντικαθιστώντας τις σχέσεις (5.19) και (5.20) στην L, εξαλείφονται οι πρωτογενείς μεταβλητές και καταλήγουμε στο Δυαδικό Πρόβλημα του Προβλήματος Βελτιστοποίησης, όπου ψάχνουμε τους πολλαπλασιαστές a_i οι οποίοι:

$$maximize \quad W(a) = \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i,j=1}^{l} a_i a_j y_i y_j \left(x_i{}^T \cdot x_j\right) \qquad (5.22)$$

$$subject\ to\ \ 0 \le a_i\ , i = 1, \dots, l\ \kappa\alpha\iota \sum_{i=1}^{l} a_i y_i = 0 \qquad (5.23)$$

Ορίζοντας τον πίνακα $(H)_{ij} = y_i y_j \left(x_i \cdot x_j\right)$ καταλήγουμε σε μια πιο συμπαγή έκφραση:

$$maximize \quad W(a) = \alpha^T - \frac{1}{2}\alpha^T H\alpha \qquad (5.24)$$

$$subject\ to\ \ \boldsymbol{0 \le \alpha}\ , i = 1, \dots, l\ \kappa\alpha\iota\ \alpha^T y = 0 \qquad (5.25)$$

Έστω ότι έχουμε υπολογίσει τα βέλτιστα a (από τα οποία κατασκευάζουμε τα w), απομένει να προσδιορίσουμε τα b για να καθορίσουμε πλήρως το υπερεπίπεδο. Για το σκοπό αυτό παίρνουμε ένα θετικό και ένα αρνητικό διάνυσμα υποστήριξης, $\boldsymbol{x^+}$ και $\boldsymbol{x^-}$ για τα οποία ξέρουμε ότι:

$$(\boldsymbol{w^T} \cdot \boldsymbol{x^+} + b) = +1 \qquad (5.26)$$

$$(\boldsymbol{w^T} \cdot \boldsymbol{x^-} + b) = -1 \qquad (5.27)$$

Από τις προηγούμενες εξισώσεις καταλήγουμε στο:

$$b = -\frac{1}{2}(\boldsymbol{w^T} \cdot \boldsymbol{x^+} + \boldsymbol{w^T} \cdot \boldsymbol{x^-}) \qquad (5.28)$$

Η συνάρτηση απόφασης του υπερεπιπέδου μπορεί συνεπώς να γραφεί:

$$f(\boldsymbol{x}) = sign\left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{l}} \boldsymbol{y_i a_i} \cdot (\boldsymbol{x^T} \cdot \boldsymbol{x_i}) + b\right) \qquad (5.29)$$

Αν $f(x) > 0 \Longrightarrow y = +1$ και αν $f(x) < 0 \Longrightarrow y = -1$.

5.6 Η Ικανότητα γενίκευσης του τέλεια εκπαιδευμένου SVM

Έστω ότι διαθέτουμε το βέλτιστο υπερεπίπεδο που διαχωρίζει τα δεδομένα και από τα l παραδείγματα εκπαίδευσης τα N_s αποτελούν διανύσματα υποστήριξης. Μπορεί να δειχθεί ότι το αναμενόμενο σφάλμα Π για τα δεδομένα που δεν έχουν παρουσιαστεί στο σύστημα φράζεται ως εξής:

$$\Pi \le \frac{N_s}{l-1} \qquad (5.30)$$

Το αποτέλεσμα αυτό αποδεικνύεται ότι όσο λιγότερα διανύσματα υποστήριξης έχουμε και συνεπώς απλούστερο και πιο συμπαγές σύστημα, τόσο καλύτερη θα είναι και επίδοση του συστήματος.

5.7 Soft Margin Classification

Σε περιπτώσεις που τα δεδομένα του προβλήματος δεν μπορούν να διαχωριστούν γραμμικά, αλλά και γενικότερα όταν επιθυμούμε το σύστημα να είναι εύρωστο στον θόρυβο, επιτρέπουμε στον ταξινομητή να κάνει κάποια λάθη κατά την διάρκεια της εκπαίδευσης, προκειμένου να αποφύγουμε την υπερεκπαίδευση του συστήματος. Έτσι επιτρέπουμε σε μερικά σημεία δεδομένων να βρίσκονται μέσα στο περιθώριο/ εύρος.
Για την υλοποίηση αυτού του μοντέλου εισάγουμε κάποιες *μεταβλητές χαλαρότητας* ξ_i *(slack variables)* στο πρόβλημα (5.15) και επαναδιατυπώνουμε το πρόβλημα ως εξής:

$$minimize \qquad \tau(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_i \xi_i \qquad (5.31)$$

$$subject\ to\ \ y_i \cdot (\boldsymbol{w^T} \cdot \boldsymbol{x_i} + b)\ \geq 1 - \xi_i\,, \quad i = 1, \dots, l \qquad (5.32)$$

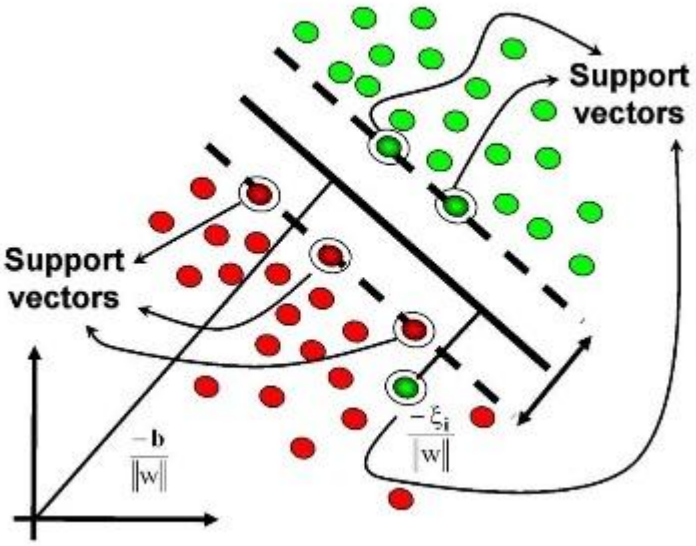
Το δυαδικό πρόβλημα των (5.31) – (5.32) εκφράζεται ως εξής:

$$maximize \qquad W(a) = \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i,j=1}^{l} a_i a_j y_i y_j \left(x_i{}^T \cdot x_j\right) \qquad (5.33)$$

$$subject\ to\ \ 0 \leq a_i \leq C\ \ , i = 1, \dots, l\ \kappa\alpha\iota\ \sum_{i=1}^{l} a_i y_i = 0 \qquad (5.34)$$

Παρατηρούμε ότι στο δυαδικό πρόβλημα δεν εμφανίζονται ούτε οι *μεταβλητές χαλαρότητας* ξ_i ούτε οι πολλαπλασιαστές Lagrange που σχετίζονται με αυτές.

Ο περιορισμός $a_i \leq C$ συμβάλλει στον έλεγχο του περιθωρίου. Όταν $C \to \infty$ το βέλτιστο υπερεπίπεδο θα είναι αυτό το οποίο διαχωρίζει τελείως τα δεδομένα (εάν υπάρχει). Για πεπερασμένο C, το πρόβλημα μετατρέπεται στην εύρεση ενός ταξινομητή «μαλακού» περιθωρίου (soft margin classifier), που επιτρέπει σε μερικά από τα δεδομένα να ταξινομηθούν λανθασμένα. Μπορούμε συνεπώς να θεωρήσουμε την σταθερά C σαν μια ρυθμιστική παράμετρο: μεγάλη τιμή του C αντιστοιχεί σε μεγαλύτερη βαρύτητα στην σωστή ταξινόμηση των δεδομένων εκπαίδευσης, ενώ μια μικρή τιμή του C οδηγεί σε ένα πιο «ευέλικτο» υπερεπίπεδο το οποίο προσπαθεί να ελαχιστοποιήσει το σφάλμα περιθωρίου για κάθε παράδειγμα. Οι πεπερασμένες τιμές του C αποδεικνύονται ως αρκετά χρήσιμες σε περιπτώσεις που τα δεδομένα δεν διαχωρίζονται εύκολα, ίσως λόγω παρουσίας θορύβου στα δεδομένα εισόδου, επιτρέποντας να βρεθεί μια λύση για το πρόβλημα.



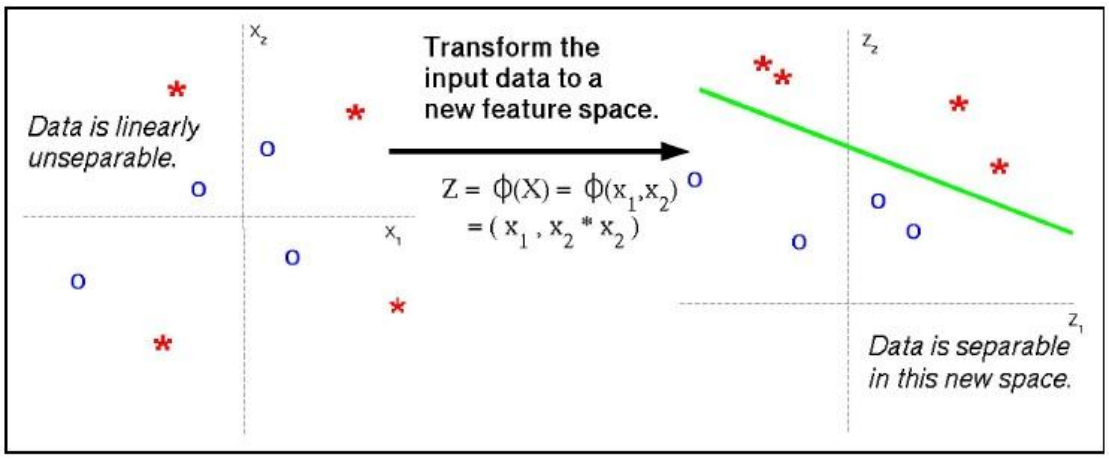
Εικόνα 21 Διανύσματα Υποστήριξης

Το πρόβλημα (5.33) υπακούει στην Αρχή Ελαχιστοποίησης Δομικού Κινδύνου, στην οποία αναφερθήκαμε σε προηγούμενη παράγραφο, και αποτελεί μια εξισορρόπηση μεταξύ του εμπειρικού σφάλματος -τα λάθη εκπαίδευσης που επιτρέπονται και εκφράζονται με τον δεύτερο όρο - και της πολυπλοκότητας του μοντέλου, που εκφράζεται με τον πρώτο όρο.

5.8 Χρήση Συναρτήσεων Πυρήνα

Σε περιπτώσεις που τα δεδομένα δεν είναι γραμμικώς διαχωρίσιμα, ακόμα και με τη χρήση μαλακού περιθωρίου χρησιμοποιείται μια «προεπεξεργασία» των δεδομένων ([Scholkopf et al., 2001]) προκειμένου το πρόβλημα να αναχθεί σε γραμμικώς διαχωρίσιμο και να ανευρεθεί μια λύση με ανάλογο τρόπο όπως προηγουμένως.

Για τον σκοπό αυτό ορίζουμε μια αντιστοίχηση $\boldsymbol{z} = \boldsymbol{\varphi}(\boldsymbol{x})$, η οποία μετατρέπει το διάνυσμα εισόδου $\boldsymbol{x}$ διάστασης $\boldsymbol{d}$, σε ένα διάνυσμα $\boldsymbol{z}$ διάστασης $\boldsymbol{d'}$ (συνήθως $\boldsymbol{d} < \boldsymbol{d'}$). Επιθυμούμε να επιλέξουμε μια συνάρτηση $\boldsymbol{\varphi}(\cdot)$, η οποία θα καταστήσει τα νέα δεδομένα εισόδου $\{\boldsymbol{\varphi}(\boldsymbol{x_i}), \boldsymbol{y_i}\}$ γραμμικώς διαχωρίσιμα.



Εικόνα 22 Χρήση Συνάρτησης Πυρήνα για διαχωρισμό των δεδομένων στο νέο χώρο χαρακτηριστικών

Ιδιαίτερη προσοχή πρέπει να δοθεί στην επιλογή της συνάρτησης πυρήνα $\varphi(\cdot)$: Αν η $\varphi(\cdot)$ αντιστοιχίζει τα δεδομένα εισόδου σε ένα χώρο πολύ υψηλής διάστασης ($d \ll d'$) , τα δεδομένα τελικά θα είναι διαχωρίσιμα, όμως θα εισήχθη πολύ μεγάλη υπολογιστική πολυπλοκότητα στο σύστημα, εξαιτίας του υπολογισμού των γινομένων $(x_i \cdot x_j)$ που προϋποθέτει η κατασκευή του πίνακα H. Αν το d' είναι εκθετικά μεγαλύτερο από το d, ο υπολογισμός του H καθίσταται απαγορευτικός τόσο ως προς το υπολογιστικό όσο και ως προς το χωρικό κόστος. Επιπλέον καθώς αυξάνουμε την πολυπλοκότητα του συστήματος με αυτόν τον τρόπο αυξάνουμε και την πιθανότητα *υπερεκπαίδευσης* (overfitting) στα δεδομένα εκπαίδευσης.

Δοσμένης μιας απεικόνισης $\boldsymbol{z} = \varphi(x)$, για να κατασκευάσουμε το καινούργιο πρόβλημα βελτιστοποίησης, αντικαθιστούμε όλες τις εμφανίσεις του x με το $\varphi(x)$. Η νέα διατύπωση του προβλήματος είναι

$$maximize \quad W(a) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j=1}^{l} a_i a_j y_i y_j \left(\varphi(x_i) \cdot \varphi(x_j)\right) \qquad (5.35)$$

$$subject\ to\ \ 0 \le a_i \le C\ , i = 1, \dots, l\ \kappa\alpha\iota \sum_{i=1}^{l} a_i y_i = 0 \qquad (5.36)$$

και

$$w = \sum_{i=1}^{l} a_i y_i \varphi(x_i) \qquad (5.37)$$

και η εξίσωση (21) γίνεται:

$$\begin{aligned} f(\boldsymbol{x}) = \ & sign(\boldsymbol{w} \cdot \varphi(x) + b) \\ & = sign\left(\left[\sum_i a_i y_i \varphi(\boldsymbol{x}_i)\right] \cdot \varphi(\boldsymbol{x}) + b\right) \\ & = sign\left(\sum_i a_i y_i \left(\varphi(\boldsymbol{x_i}) \cdot \varphi(\boldsymbol{x})\right) + b\right) \end{aligned} \qquad (5.38)$$

Παρατηρούμε ότι η συνάρτηση $\boldsymbol{\varphi(\cdot)}$ εμφανίζεται μόνο σε γινόμενα της μορφής $\boldsymbol{\varphi(x_i) \cdot \varphi(x_j)}$. Συνεπώς αντί να υπολογίσουμε την απεικόνιση $\boldsymbol{x \mid \rightarrow \varphi(x)}$ για κάθε σημείο και στη συνέχεια να υπολογήσουμε το εσωτερικό γινόμενο, μπορούμέ να χρησιμοποιήσουμε το αποκαλούμενο Kernel Trick και να χρησιμοποιήσουμε μια συνάρτηση πυρήνα $\boldsymbol{K(u,v) = \varphi(u) \cdot \varphi(v)}$.

Ουσιαστικά η συνάρτηση πυρήνα μας επιτρέπει να υπολογίσουμε το εσωτερικό γινόμενο στον χώρο μεγαλύτερης διάστασης, χωρίς να είναι απαραίτητο να ορίσουμε την συνάρτηση φ(), αλλά χρησιμοποιώντας μόνο τα δεδομένα όπως αυτά ορίζονται στον αρχικό χώρο χαρακτηριστικών. Σαν συναρτήσεις πυρήνα μπορούν να χρησιμοποιηθούν οι συναρτήσεις που υπακούουν στο θεώρημα του Mercer (Mercer's theorem).

Το πρόβλημα βελτιστοποίησης με χρήση συνάρτησης πυρήνα εκφράζεται ωε εξής:

$$maximize \quad W(a) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j=1}^{l} a_i a_j y_i y_j K(x_i, x_j) \qquad (5.40)$$

$$subject\ to\ 0 \le a_i \le C, i = 1, \dots, l \ και \sum_{i=1}^{l} a_i y_i = 0 \qquad (5.41)$$

Συναρτήσεις που χρησιμοποιούνται συνηθέστερα ως συναρτήσεις πυρήνα είναι οι εξής:

- *Γραμμική:* $K(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v}$
- *Πολυωνυμική:* $K(\boldsymbol{u}, \boldsymbol{v}) = (\gamma \boldsymbol{u} \cdot \boldsymbol{v} + r)^p, \gamma > 0$
- *Ακτινωτή Συνάρτηση Βάσης (Radial Basis Function –RBF):* $K(\boldsymbol{u}, \boldsymbol{v}) = e^{-\gamma |\boldsymbol{u} - \boldsymbol{v}|^2}$
- *Σιγμοειδής:* $K(\boldsymbol{u}, \boldsymbol{v}) = tanh(\gamma \boldsymbol{u} \cdot \boldsymbol{v} + r)$

Συνηθέστερα χρησιμοποιείται ο πυρήνας RBF, ο οποίος απεικονίζει τα δεδομένα σε μεγαλύτερη διάσταση με μη γραμμικό τρόπο. Όπως έχουν αποδείξει οι Keerthi and Lin ένα SVM με γραμμική συνάρτηση πυρήνα με παράμετρο τιμωρίας $\tilde{C}$ παρουσιάζει την ίδια επίδοση με ένα SVM με RBF πυρήνα με παραμέτρους κάποια (C, γ), συνεπώς η γραμμική περιπτωση μπορεί να θεωρηθεί σαν υποπερίπτωση του RBF.

Ένας επιπλέον λόγος αυτής της επιλογής έγκειται στο γεγονός ότι ο αριθμός των παραμέτρων του μοντέλου που χρησιμοποιεί RBF είναι μικρότερος σε σχέση με τον πολυωνυμικό πυρήνα, καθώς και το γεγονός ότι εμπεριέχει λιγότερες αριθμητικές δυσκολίες. Η συνάρτηση πυρήνα RBF έχει πεδίο τιμών το $(0,1]$, σε αντίθεση με την πολυωνυμική που μπορεί να πάρει τιμές κοντά στο άπειρο για μεγάλους εκθέτες. Επιπλέον αξίζει να σημειωθεί ότι η σιγμοειδής συνάρτηση δεν αποτελεί έγκυρη συνάρτηση πυρήνα για κάποιες παραμέτρους (δεν ικανοποιεί το θεώρημα του Mercer).

Για τους λόγους που προαναφέρθηκαν, στη παρούσα εργασία χρησιμοποιήσαμε έναν γραμμικό ταξινομητή μαλακού περιθωρίου (όπως στο [Wolf et al., 2012]), έναν μη γραμμικό ταξινομητή με πυρήνα RBF, καθώς και μια παραλλαγή αυτού, το wSVM, που αναλύεται σε επόμενη παράγραφο.

5.9 Εκπαίδευση SVM

Ένα πολύ σημαντικό κομμάτι της εκπαίδευσης των SVM, και ίσως το πιο απαιτητικό από άποψη χρόνου, είναι η αναζήτηση των παραμέτρων που υπεισέρχονται σε κάθε μοντέλο. Σκοπός μας είναι να εντοπίσουμε τις τιμές των παραμέτρων για τις οποίες ο ταξινομητής μπορεί να κάνει προβλέψεις με ακρίβεια την κλάση των άγνωστων δεδομένων.
Για την αναζήτηση των παραμέτρων εκτελούμε k-fold crossvalidation, επειδή αυτή η διαδικασία μπορεί να εμποδίσει το φαινόμενο της υπερεκπαίδευσης. Διαιρούμε το σύνολο εκπαίδευσης σε k υποσύνολα ίσου μεγέθους και διαδοχικά εκπαιδεύουμε τον ταξινομητή με τα k-1 υποσύνολα και κάνουμε επαλήθευση πάνω στο ένα υποσύνολο, το οποίο δεν χρησιμοποιήθηκε στην εκπαίδευση. Συνήθως χρησιμοποιείται 5 ή 10 fold cross validation και επιθυμούμε, ανάλογα με τους χρονικούς περιορισμούς που ανακύπτουν κάθε φορά, να εκτελούμε τη διαδικασία αυτή με όσο το δυνατόν μεγαλύτερο k γίνεται. Στην ιδανική περίπτωση "leave-one-out", η οποία βέβαια είναι υπολογιστικά χρονοβόρα, αν έχουμε N παραδείγματα εισόδου, εκπαιδεύουμε τον ταξινομητή για τα N-1 παραδείγματα και ελέγχουμε πάνω στο 1 που έχει απομείνει, εκτελώντας αυτή τη διαδικασία N φορές.

Για το SVM με πυρήνα RBF εκτελέσαμε το λεγόμενο "grid search" cross validation. ([Burges, 1998]) Κατά τη διαδικασία αυτή δοκιμάζονται διάφορα ζευγάρια παραμέτρων (C, γ) με τιμές επιλεγμένες πάνω από ένα «πλέγμα» και το ζευγάρι με την καλύτερη cross validation accuracy επιλέγεται ως ζευγάρι παραμέτρων για τον τελικό ταξινομητή. Συνήθως χρησιμοποιούνται εκθετικά αυξανόμενες ακολουθίες για τις τιμές των παραμέτρων ($\{2^{-15}, 2^{-14}, \dots, 2^{14}, 2^{15}\}$), και εκτελούνται διαδοχικά τρία grid search: στο πρώτο εντοπίζεται χοντρικά η περιοχή γύρω από τις βέλτιστες παραμέτρους και στα επόμενα δύο grid search γίνεται πιο λεπτομερής αναζήτηση γύρω από αυτή τη βέλτιστη περιοχή, μικραίνοντας το βήμα αναζήτησης.

5.10 SVM με βάρη (wSVM)

Ορισμένες φορές καλούμαστε να επιλύσουμε ένα πρόβλημα ταξινόμησης, στο οποίο κάποια συγκεκριμένα σημεία έχουν περισσότερη βαρύτητα από τα υπόλοιπα και απαιτείται να ταξινομηθούν σωστά. Στις περιπτώσεις αυτές χρησιμοποιούνται διαφορετικές παράμετροι τιμωρίας για κάθε πρότυπο, έτσι ώστε να επιβάλλεται μεγάλη τιμωρία αν ταξινομηθεί λάθος ένα από τα σημεία που μας ενδιαφέρουν.
Αν διαθέτουμε έναν πίνακα P,μεγέθους 1x N (N ο αριθμός προτύπων εισόδου) ο οποίος περιέχει ένα βάρος p_iγια κάθε πρότυπο εκπαίδευσης, όπου καθώς αυξάνεται η τιμή του βάρους, αυξάνεται και το ενδιαφέρον μας προκυμένου να μην ταξινομηθεί λάθος το σημείο αυτό, τότε μπορούμε να χρησιμοποιήσουμε την ακόλουθη αντικειμενική συνάρτηση:

$$minimize \qquad \tau(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C \cdot p_i \sum_i \xi_i \qquad (5.42)$$

$$subject\ to\ \ y_i \cdot (\boldsymbol{w^T} \cdot \boldsymbol{x_i} + b)\ \geq 1 - \xi_i\,, \quad i = 1, \dots, l \qquad (5.43)$$

Το μοντέλο αυτό το χρησιμοποιούμε σε κάποια από τα πειράματά μας, χρησιμοποιώντας τα βάρη των ακμών του γράφου στη θέση των p_i.

ΚΕΦΑΛΑΙΟ 6:

Support Vector Neural Network – (SVNN)

6.1 Εισαγωγή

Για την αποτροπή του φαινομένου της υπερεκπαίδευσης στα πρότυπα εισόδου, χρησιμοποιείται αρκετά συχνά η τεχνική του regularization. Για τα Νευρωνικά Δίκτυα τέτοιες τεχνικές είναι το Early Stopping, η Εξομάλυνση βάση Καμπυλότητας (Curvature Driven Smoothing) και η Μείωση των Βαρών (weight decay).

Στο κεφάλαιο αυτό θα παρουσιάσουμε έναν αλγόριθμο εκπαίδευσης Νευρωνικού Δικτύου με ανοχή στο θόρυβο των δεδομένων, ο οποίος υιοθετεί μια συνάρτηση βελτιστοποίησης παρόμοια με αυτή που χρησιμοποιείται στα SVM και για τον λόγω αυτόν αποκαλείται SVNN. Ο αλγόριθμος αυτός αποτελεί μέρος της διδακτορικής διατριβής του Oswaldo Ludwig [Ludwig, 2012].

Ο αλγόριθμος αυτός βασίζεται σε μια καινούργια τεχνική regularization, την αποκαλούμενη Μείωση των Ιδιοτιμών (eigenvalue decay), η οποία περιλαμβάνει έναν επιπλέον όρο στη συνάρτηση κόστους του αλγορίθμου μάθησης, ο οποίος τιμωρεί τις υπερβολικά μεγάλες και μικρές ιδιοτιμές του πίνακα $W_1 W_1^T$, όπου W_1 είναι ο πίνακας των συναπτικών βαρών του πρώτου επιπέδου του δικτύου, στοχεύοντας στην βελτίωση του περιθωρίου του ταξινομητή.

6.2 Eigenvalue Decay

Η πιο συχνά χρησιμοποιούμενη αντικειμενική συνάρτηση στα Νευρωνικά Δίκτυα είναι το Μέσο Τετραγωνικό Σφάλμα:

$$e = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad (6.1)$$

, όπου N το μέγεθος του συνόλου εκπαίδευσης, y_i η επιθυμητή έξοδος και $\hat{y}_i$ η έξοδος του MLP για την είσοδο x_i.

Στην περίπτωση της συνήθους weight decay μεθόδου, στην πιο πάνω εξίσωση προστίθενται περεταίρω όροι οι οποίοι τιμωρούν υπερβολικά υψηλές τιμές των βαρών και των biases και η αντικειμενική συνάρτηση παίρνει την ακόλουθη μορφή:

$$e^* = e + \kappa_1 \sum_{w_i \in W_1} w_i^2 + \kappa_2 \sum_{w_j \in w_2} w_j^2 + \kappa_3 \sum_{b_{(1,k)} \in b_1} b_{(1,k)}^2 + \kappa_4 b_2^2 \qquad (6.2)$$

, όπου $W_1, w_2, b_1 \kappa\alpha\iota\ b_2$ οι παράμετροι του MLP και κ_i οι υπερπαράμετροι του συστήματος.

Στη περίπτωση της eigenvalue decay η αντικειμενική συνάρτηση ορίζεται ως:

$$e^{**} = e + \kappa(\lambda_{min} + \lambda_{max}) \qquad (6.3)$$

,όπου λ_{min} και λ_{max} οι μικρότερη και η μεγαλύτερη ιδιοτιμή αντίστοιχα του $W_1 W_1^T$.

6.3 Ανάλυση της Eigenvalue Decay

Έστω K το πεδίο των πραγματικών αριθμών, $K^{n \times n}$ ένας χώρος διανυσμάτων που περιέχει όλους τους πίνακες με n σειρές και n στήλες με περιεχόμενα από το K, $A \in K^{n \times n}$ ένας συμμετρικός θετικά ημιορισμένος πίνακας, λ_{min} και λ_{max} οι μικρότερη και η μεγαλύτερη ιδιοτιμή αντίστοιχα του πίνακα A. Για κάθε $x \in K^n$ ισχύει η ακόλουθη ανισότητα:

$$\lambda_{min} x^T x \leq x^T A x \leq \lambda_{max} x^T x \qquad (6.4)$$

[Απόδειξη]

Έστω $V = [v_1 \dots v_n]$ ο τετραγωνικός πίνακας μεγέθους $n \times n$ του οποίου η i-στη στήλη αποτελεί το ιδιοδιάνυσμα v_i του A, και Λ ο διαγώνιος πίνακας του οποίου το i-στο διαγώνιο στοιχείο είναι η ιδιοτιμή λ_i του A. Οι ακόλουθες σχέσεις ισχύουν:

$$x^T A x = x^T V V^{-1} A V V^{-1} x = x^T V \Lambda V^T x \qquad (6.5)$$

Λαμβάνοντας υπόψη ότι ο πίνακας A είναι θετικά ημιορισμένος, δηλαδή δεν έχει αρνητικές ιδιοτιμές έχουμε:

$$x^T V(\lambda_{min} I) V^T x \leq x^T V \Lambda V^T x \leq x^T V(\lambda_{max} I) V^T x \qquad (6.6)$$
$$\Longrightarrow \lambda_{min} x^T x \leq x^T V \Lambda V^T x \leq \lambda_{max} x^T x$$

Χρησιμοποιώντας την εξίσωση (6.5) στην (6.6) αποδεικνύεται η ανισότητα (6.4).

Το επόμενο θεώρημα δίνει ένα άνω και κάτω φράγμα για το εύρος ταξινόμησης:

Θεώρημα:

Έστω m_i το περιθώριο του παραδείγματος εκπαίδευσης i, δηλαδή η μικρότερη κάθετη απόσταση μεταξύ του υπερεπιπέδου απόφασης του ταξινομητή και του πρώτου ,που i, λ_{min} και λ_{max} οι μικρότερη και η μεγαλύτερη ιδιοτιμή αντίστοιχα του $W_1 W_1^T$, τότε για $m_i > 0$, δηλαδή για ένα σωστά ταξινομημένο παράδειγμα, η ακόλουθη ανισότητα ισχύει.

$$\frac{1}{\sqrt{\lambda_{max}}}\mu \leq m_i \leq \frac{1}{\sqrt{\lambda_{min}}}\mu \tag{6.7}$$

όπου

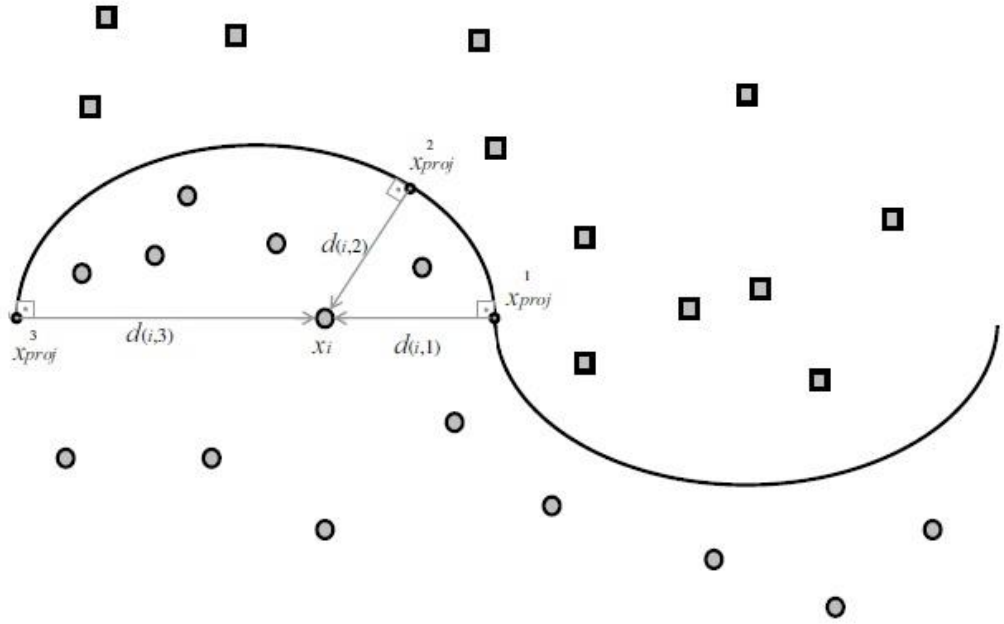
$$\mu = \min_j \left(y_i \frac{w_2^T \Gamma_j W_1 (x_i - x_{proj}^j)}{\sqrt{w_2^T \Gamma_j \Gamma_j^T w_2}} \right) \tag{6.8}$$

$$\Gamma_j = \begin{bmatrix} \varphi'(v_1) & 0 & \cdots & 0 \\ 0 & \varphi'(v_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \varphi'(v_n) \end{bmatrix} \tag{6.9}$$

$$[v_1, v_2, \dots, v_n]^T = W_1 \cdot x_i + b_1 \tag{6.10}$$

$$\varphi'(v_n) = \frac{\partial \varphi}{\partial v_n}\Big|_{x_{proj}^j} \tag{6.11}$$

,όπου x_{proj}^j η j-στη προβολή του i παραδείγματος πάνω στο υπερεπίπεδο., όπως φαίνεται στην ακόλουθη εικόνα.



Εικόνα 23 Απεικόνιση Χώρου Χαρακτηριστικών με μη γραμμική διαχωριστική υπερεπιφάνεια με τις προβολές x_{proj}^{j} του i-στου παραδείγματος εκπαίδευσης x_i και οι κάθετες αποστάσεις $d_{(i,j)}$

[Απόδειξη]

Αρχικά υπολογίζουμε την κλίση της εξόδου του Νευρωνικού Δικτύου σε σχέση με το διάνυσμα εισόδου x. Ο υπολογισμός γίνεται στο σημείο x_{proj}^{j}.

Το δίκτυο που χρησιμοποιούμε αποτελείται από ένα κρυφό επίπεδο με σιγμοειδή συναρτήσεις ενεργοποίησης και από ένα επίπεδο εξόδου με γραμμικές συναρτήσεις ενεργοποίησης. Συνεπώς το δίκτυο περιγράφεται από τις σχέσεις:

$$y_h = \varphi(W_1 \cdot x + b_1) \tag{6.12}$$

$$\hat{y} = w_2^T y_h + b_2$$

Άρα για την κλίση έχουμε :

$$\nabla\hat{y}_{(i,j)} = \frac{\partial\hat{y}}{\partial x}\Big|_{x_{proj}^{j}} = w_2^T \Gamma_j W_1 \tag{6.13}$$

Το μοναδιαίο διάνυσμα

$$\overrightarrow{\boldsymbol{p_j}} = \frac{\nabla\hat{y}_{(i,j)}}{\left\|\nabla\hat{y}_{(i,j)}\right\|} \tag{6.14}$$

είναι κάθετο στην διαχωριστική επιφάνεια, δίνοντας την κατεύθυνση από το x_i στο x_{proj}^{j}

Έτσι έχουμε

$$x_i - x_{proj}^{j} = d_{(i,j)}\,\overrightarrow{\boldsymbol{p_j}} \tag{6.15}$$

, όπου $d_{(i,j)}$ η βαθμωτή απόσταση μεταξύ x_i και x_{proj}^{j}

Από την προηγούμενη σχέση έχουμε

$$\nabla\hat{y}_{(i,j)}\left(x_i - x_{proj}^{j}\right) = d_{(i,j)}\,\nabla\hat{y}_{(i,j)}\,\overrightarrow{\boldsymbol{p_J}} \qquad (6.16)$$

Αντικαθιστώντας την (6.13) στην (6.12)και επιλύοντας ως προς $d_{(i,j)}$, έχουμε

$$d_{(i,j)} = \frac{\nabla\hat{y}_{(i,j)}\left(x_i - x_{proj}^{j}\right)}{\left\|\nabla\hat{y}_{(i,j)}\right\|} \qquad (6.17)$$

Η απόλυτη τιμή του εύρους του ταξινομητή, m_i, είναι η μικρότερη τιμή του $d_{(i,j)}$ ως προς j:

$$|m_i| = \min_j d_{(i,j)} \qquad (6.18)$$

Το πρόσημο του m_i εξαρτάται από την επιθυμητή κλάση y_i και για αυτό:

$$m_i = \min_j\left(y_i \frac{\nabla\hat{y}_{(i,j)}\left(x_i - x_{proj}^{j}\right)}{\left\|\nabla\hat{y}_{(i,j)}\right\|}\right) \qquad (6.19)$$

Αντικαθιστώντας την (6.11) στην (6.17) προκύπτει:

$$m_i = \min_j\left(y_i \frac{w_2^T\Gamma_j W_1\left(x_i - x_{proj}^{j}\right)}{\sqrt{w_2^T\Gamma_j W_1 W_1^T\Gamma_j^T w_2}}\right) \qquad (6.20)$$

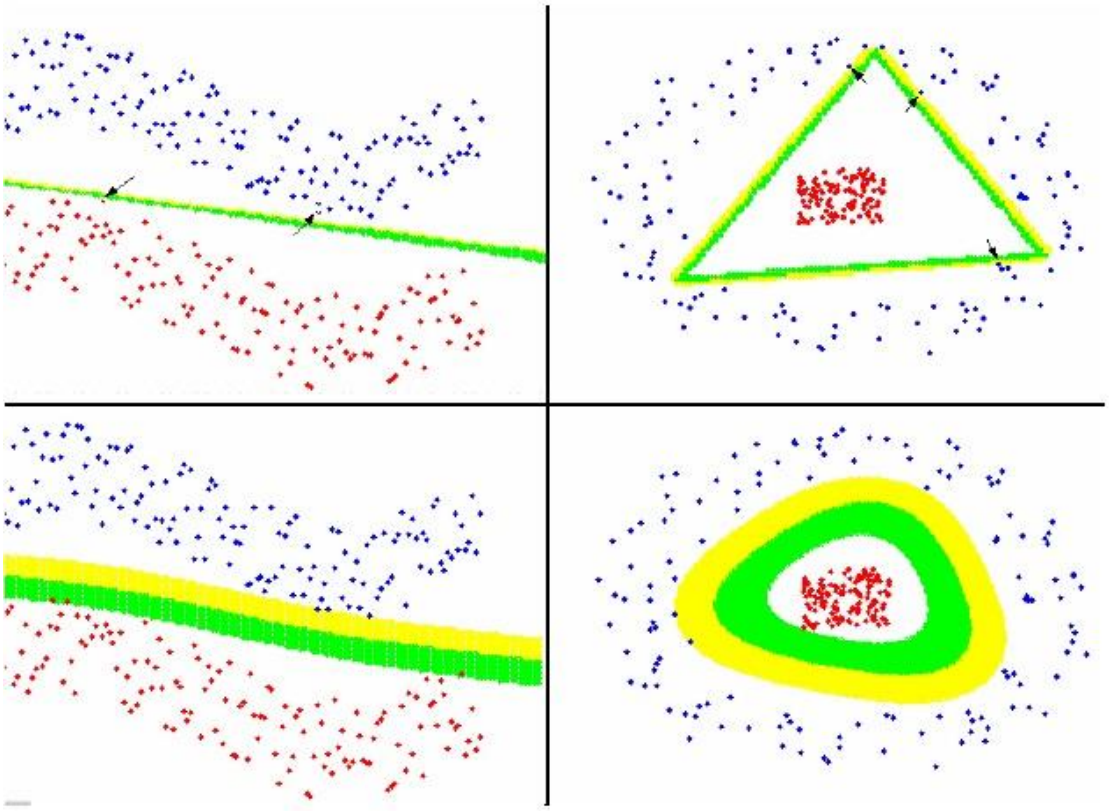
Σημειώνουμε ότι ο πίνακας $W_1W_1^T$ είναισυμμετρικός θετικά ημιορισμένος, συνεπώς από την (6.4) έχουμε για όλα τα Γ_j και w_2

$$\lambda_{min} w_2^T\Gamma_j\Gamma_j^T w_2 \le w_2^T\Gamma_j W_1 W_1^T\Gamma_j^T w_2 \le \lambda_{max} w_2^T\Gamma_j\Gamma_j^T w_2 \qquad (6.21)$$

Από τις (6.20) και (6.21) προκύπτει η (6.7).

Παρατηρώντας ότι τα λ_{min} και λ_{max} αποτελούν τους παρανομαστές του ανωτέρω φράγματος, η μέθοδος εκπαίδευσης που ακολουθεί την eigenvalue decay προσπαθεί να μειώσει τις ιδιοτιμές αυτές προκειμένου να αυξήσει το άνω και το κάτω φράγμα του περιθωρίου ταξινόμησης.

6.4 Ταξινόμηση με το SVNN



Εικόνα 24 Η επίδραση του Eigenvalue Decay στο εύρος ταξινόμησης.

Στην εικόνα 24 απεικονίζονται δύο παραδείγματα ταξινόμησης με και χωρίς την χρήση του κριτηρίου Eigenvalue Decay για την εκπαίδευση ενός Νευρωνικού Δικτύου. Στην πάνω σειρά εικόνων παρουσιάζεται ο χώρος χαρακτηριστικών με την διαχωριστική υπερεπιφάνεια που προκύπτει από την εκπαίδευση του Νευρωνικού χωρίς την χρήση του Eigenvalue Decay και στην κάτω σειρά παρουσιάζονται τα ίδια παραδείγματα με την χρήση του προαναφερθέντος κριτηρίου. Παρατηρούμε ότι ο διαχωρισμός πλέον ομοιάζει πολύ με αυτόν που θα επιτύγχανε ένας SVM ταξινομητής. Για τα δεδομένα που βρίσκονται μέσα στην κίτρινη περιοχή το νευρωνικό δίκτυο επιστρέφει $0 \leq \hat{y}_i < 1$, ενώ για αυτά που βρίσκονται στην πράσινη περιοχή επιστρέφει $0 > \hat{y}_i > -1$. Το όριο μεταξύ των δύο χρωματιστών περιοχών αντιπροσωπεύει την διαχωριστική υπερεπιφάνια.

Η τεχνική αυτή παραπέμπει στη λειτουργία των SVM ταξινομητών, καθώς τα δεδομένα εισόδου που κείτονται μέσα στο εύρος ταξινόμησης ή σε λανθασμένη πλευρά αυξάνουν τον όρο τιμωρίας. Ο αλγόριθμος ελαχιστοποιεί τον όρο τιμωρίας με τέτοιο τρόπο, έτσι ώστε να μετακινήσει την «χρωματιστή περιοχή» , δηλαδή την διαχωριστική υπερεπιφάνια μακριά από τα δεδομένα εκπαίδευσης. Έτσι όσο μεγαλύτερο προκύπτει το εύρος της «χρωματισμένης περιοχής», δηλαδή όσο μικρότερες είναι οι ιδιοτιμές του πίνακα $W_1 W_1^T$,

τόσο μεγαλύτερη θα είναι και η απόσταση μεταξύ των δεδομένων και της διαχωριστικής επιφάνειας.

Σύμφωνα με τη μέθοδο του SVNN το πρόβλημα που έχει να βελτιστοποιήσει το νευρωνικό δίκτυο, σε αναλογία με το πρόβλημα βελτιστοποίησης του SVM (24), είναι το ακόλουθο

$$\min_{W_1,\xi_i}\left(\lambda_{min}+\lambda_{max}+\frac{C}{N}\sum_{i=1}^{N}\xi_i\right) \quad (6.22)$$

$$subject\ to\ \ 0\le\xi_i\ i=1,\dots,N\ \kappa\alpha\iota\ y_i\hat{y}_i\ge 1-\xi_i \quad (6.23)$$

, όπου C η υπερπαράμετρος κανονικοποίησης και ξ_i οι μεταβλητές χαλάρωσης, που σχετίζονται με τα σφάλματα ταξινόμησης στην εκπαίδευση.

Το πρόβλημα βελτιστοποίησης υπό περιορισμούς αντικαθίσταται από το ακόλουθο πρόβλημα βελτιστοποίησης χωρίς περιορισμούς:

$$\min_{W_1,w_2,b_1,b_2}(\Phi) \quad (6.24)$$

, όπου

$$\Phi=\lambda_{min}+\lambda_{max}+\frac{C}{N}\sum_{i=1}^{N}H\ \ (y_i\hat{y}_\iota) \quad (6.25)$$

και $H(t)=\max(0,1-t)$ το Hinge loss.

Λόγω των ασυνεχειών που παρουσιάζει η Φ, δεν είναι δυνατή η εφαρμογή της Καθόδου Κλίσης (Gradient Descend) και συνεπώς για την βελτιστοποίηση χρησιμοποιείται ένας γενετικός αλγόριθμος, που χρησιμοποιεί την Φ σαν συνάρτηση καταλληλότητας. (fitness function).

Στην εξίσωση (6.25) παρατηρούμε ότι ο τελευταίος όρος αποτελεί όρο τιμωρίας για τις εκτιμώμενες εξόδους του ταξινομητή οι οποίες δεν υπακούουν στον περιορισμό $y_i\cdot\hat{y}_i\ge 1$, με τέτοιο τρόπο ώστε να διασφαλιστεί ένα ελάχιστο εύρος ταξινόμησης. Στη ίδια εξίσωση οι δύο πρώτοι όροι στοχεύουν στη μεγιστοποίηση αυτού του ελαχίστου εύρους, σύμφωνα με το κριτήριο της Eigenvalue Decay, όπως αναφέρθηκε στο προηγούμενο θεώρημα.

ΚΕΦΑΛΑΙΟ 7:

Μέτρα Αξιολόγησης - Evaluation Metrics:

7.1 Εισαγωγή

Στις εφαρμογές ταξινόμησης με διακριτές καταστάσεις η επίδοση του ταξινομητή περιγράφεται συνήθως από έναν confusion matrix. Τα στοιχεία του confusion matrix καταδεικνύουν τον αριθμό των παραδειγμάτων που ταξινομήθηκαν σωστά ή λάθος. Τα μέτρα αξιολόγησης συνοψίζουν τον πίνακα αυτόν σε μια τιμή που μπορεί να χρησιμοποιηθεί για την σύγκριση διαφορετικών ταξινομητών.

Η επιλογή του κατάλληλου μετρικής αποτελεί κρίσιμο κομμάτι της κατασκευής του συστήματος ταξινόμησης και εξαρτάται από την εκάστοτε εφαρμογή. Ένα μη επαρκώς καθορισμένο μέτρο μπορεί να οδηγήσει σε επιλογή μη ιδανικού μοντέλου ή να προκαλέσει λανθασμένα συμπεράσματα κατά τη σύγκριση των επιδόσεων δύο ταξινομητών.

Πληθώρα μέτρων αξιολόγησης έχουν χρησιμοποιηθεί για την εκτίμηση των ταξινομητών. Για προβλήματα με ισορροπημένο σύνολο δεδομένων, δηλαδή όταν οι δύο κλάσεις είναι ισοπίθανες, η συνολική accuracy χρησιμοποιείται συχνότερα.

Θα εξετάσουμε αρχικά την κατασκευή του confusion matrix και στη συνέχεια θα αναλύσουμε τα μέτρα που προκύπτουν από αυτόν.

7.2 Confusion Matrix

Ο confusion matrix αποτελεί έναν πίνακα συγκεκριμένης διάταξης, ο οποίος διευκολύνει την απεικόνιση της επίδοσης του αλγορίθμου ταξινόμησης. Το όνομα του προέρχεται από το γεγονός ότι διευκολύνει να συμπεράνουμε αν το σύστημα ταξινόμησης συγχέει δύο κλάσεις (δηλαδή ταξινομεί τα στοιχεία μιας κλάσης σε άλλη).

Για την κατασκευή του πίνακα χρησιμοποιούνται είτε μετρήσεις του απόλυτου αριθμού των φορών που κάθε προβλεπόμενη ετικέτα (predicted label) συσχετίζεται με τις αληθινές κλάσεις (συνήθως συμβολίζονται με κεφαλαία γράμματα), είτε πιθανοτικές τιμές (χρησιμοποιούνται συνήθως πεζά σύμβολα).

Δοθέντος ενός ταξινομητή και ενός παραδείγματος μπορούμε να λάβουμε τέσσερα πιθανά αποτελέσματα. Αν το παράδειγμα είναι θετικό και ταξινομηθεί ως θετικό, τότε προσμετράται ως *πραγματικά θετικό (true positive - tp)* , ενώ αν ταξινομηθεί ως αρνητικό προσμετράται ως *εσφαλμένα αρνητικό (false negative - fn)*. Αν το παράδειγμα ανήκει στην αρνητική κλάση και ταξινομηθεί ως αρνητικό, τότε υπολογίζεται σαν *πραγματικά αρνητικό (true negative - tn)*, ενώ αν ταξινομηθεί ως θετικό θεωρείται *εσφαλμένα θετικό (false positive – fp)*.

Κάθε γραμμή του confusion matrix αντιπροσωπεύει τα παραδείγματα που ανήκουν σε μία προβλεπόμενη κλάση, σύμφωνα με το αποτέλεσμα του ταξινομητή, ενώ κάθε στήλη αντιπροσωπεύει τα παραδείγματα που ανήκουν σε μια πραγματική κλάση. Στην επόμενη εικόνα παρουσιάζεται η συνήθης μορφή ενός confusion matrix.

Εικόνα 25 Confusion Matrix

Τα σύμβολα $+R$ και $-R$ συμβολίζουν τις πραγματικές κλάσσεις στις οποίες ανήκουν τα παραδείγματα, ενώ με +P και –P συμβολίζονται οι κλάσεις τις οποίες επιστρέφει ο ταξινομητής. Με tp, fp, fn, tn συμβολίζονται τα ποσοστά των πραγματικά θετικών, εσφαλμένα αρνητικών , πραγματικά αρνητικών και εσφαλμένα θετικών που προέκυψαν κατά τη διάρκεια της επαλήθευσης, ενώ με pp και pn συμβολίζονται τα ποσοστά των παραδειγμάτων που προβλέφθηκαν ως θετικά και αρνητικά αντίστοιχα. Η πράσινη διαγώνιος συνεπώς συμβολίζει τις σωστές προβλέψεις και η κόκκινη τις λανθασμένες.

Τα πιο συχνά χρησιμοποιούμενα μέτρα είναι:

- ***Ανάκληση (Recall) ή Ευαισθησία (Sensitivity) ή (True Positive Rate)*** : Αποτελεί το ποσοστό των πραγματικά θετικών παραδειγμάτων που προβλέφθηκαν ως θετικά. Δηλαδή

$$Recall = \frac{\#\ Θετικών\ παραδειγμάτων\ που\ ταξινομηθηκαν\ σωστα}{\#\ Πραγματικά\ Θετικών\ Παραδειγμάτων} = \frac{tp}{rp} = \frac{tp}{tp + fn}$$

- ***Ακρίβεια (Precision) ή Εμπιστοσύνη (Confidence)***: Εκφράζει το ποσοστό των προβλεφθέντων θετικών παραδειγμάτων , τα οποία είναι πραγματικά θετικά. Δηλαδή

$$Precision = \frac{\#\ Θετικών\ παραδειγμάτων\ που\ ταξινομηθηκαν\ σωστα}{\#\ Παραδειγμάτων\ που\ προβλεύθηκαν\ ως\ θετικά} = \frac{tp}{pp} = \frac{tp}{tp + fp}$$

- ***Accuracy*** : Ορίζεται ως ο αριθμός των σωστά ταξινομημένων παραδειγμάτων προς το σύνολο των παραδειγμάτων, δηλαδή:

$$Accuracy = \frac{\#\ παραδειγμάτων\ που\ ταξινομηθηκαν\ σωστα}{Συνολικός\ \#\ παραδειγμάτων} = \frac{tp + tn}{tp + fp + tn + fn}$$

- ***F1 – Score***: Αποτελεί τον αρμονικό μέσο των Precision και Recall

$$F1 - Score = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision}$$

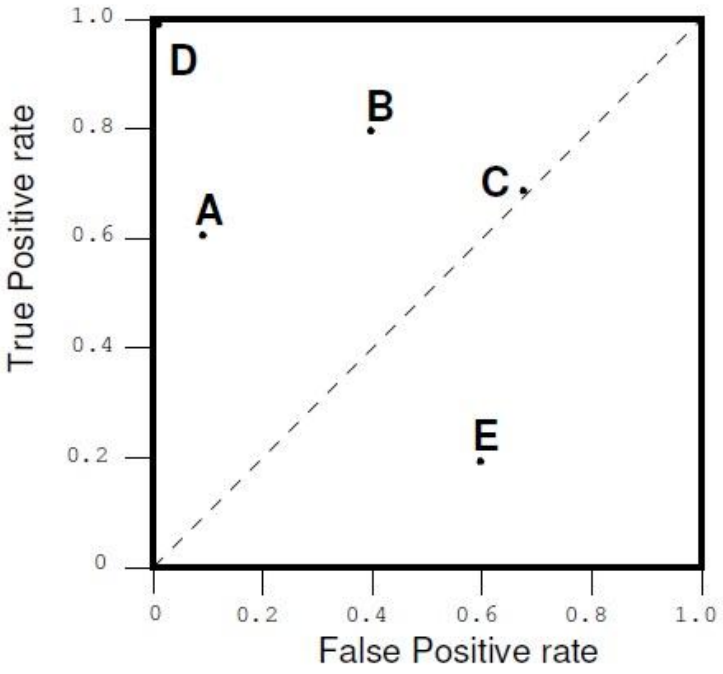
- ***Ποσοστό Ψευδών Θετικών (False Positive Rate)***:

$$False\ Positive\ Rate = \frac{\#\ Αρνητικών\ παραδειγμάτων\ που\ ταξινομηθηκαν\ λάθως}{\#\ Πραγματικά\ Αρνητικών\ Παραδειγμάτων} =$$

Εκτός από αυτά τα μέτρα χρησιμοποιούνται σπανιότερα: η Αντίστροφη Ανάκληση (Inverse Recall) και η Αντίστροφη Ακρίβεια (Inverse Precision), οι οποίες ορίζονται με παρόμοιο τρόπο όπως ορίσαμε πιο πάνω την Ανάκληση και την Ακρίβεια, εκφράζοντας τα αρνητικά παραδείγματα, καθώς και ο Ρυθμός Αποτυχίας (Miss Rate) και Fallout.

7.3 ROC (Receiver Operating Characteristic)

Αρκετά συχνά για την αξιολόγηση ταξινομητών χρησιμοποιούνται τα γραφήματα ROC (Receiver Operating Characteristic), τα οποία αποτελούν δισδιάστατα γραφήματα που απεικονίζουν τη σχέση του True Positive rate με το False Positive Rate. Στην εικόνα που ακολουθεί παρουσιάζεται ένα ROC γράφημα πέντε ταξινομητών , επισημασμένοι με γράμματα από A μέχρι E.



Εικόνα 26 Γραφήματα ROC πέντε ταξινομητών

Ένας διακριτός ταξινομητής επιστρέφει μια μόνο τιμή ετικέτας κλάσης και περιγράφει ένα ζευγάρι (fp rate , tp rate), το οποίο αντιστοιχεί σε ένα σημείο στον ROC χώρο. Στην προηγούμενη εικόνα και οι πέντε ταξινομητές είναι διακριτοί.

Στον χώρο ROC, το σημείο (0,0) αντιστοιχεί στον ταξινομητή που δεν επιστρέφει ποτέ θετική πρόβλεψη, δηλαδή δεν ταξινομεί κανένα παράδειγμα στη θετική κλάση, αποφεύγοντας έτσι τα false positive λάθη και χάνοντας και τα true positive. Το σημείο (1,1) εκφράζει την αντίθετη στρατηγική, δηλαδή τη ταξινόμηση όλων των παραδειγμάτων ως θετικών. Το σημείο (0,1) αντιπροσωπεύει την ιδανική ταξινόμηση και συνεπώς ο ταξινομητής D θεωρείται ιδανικός.

Οι ταξινομητές που απεικονίζονται στην αριστερή πλευρά του ROC γραφήματος, κοντά στον X άξονα, θεωρούνται "συντηρητικοί" : επιστρέφουν θετικές ταξινομήσεις μόνο με ισχυρές ενδείξεις, επιτυγχάνοντας έτσι χαμηλό αριθμό false positive λαθών, αλλά παρουσιάζοντας συγχρόνως και χαμηλό αριθμό true positive. Οι ταξινομητές στην επάνω δεξιά γωνία του ROC γραφήματος μπορούν να θεωρηθούν "γενναιόδωροι" : ταξινομούν στην θετική κλάση με ασθενείς ενδείξεις, με αποτέλεσμα να ταξινομούν όλα τα θετικά παραδείγματα σωστά αλλά συγχρόνως να εμφανίζουν υψηλά false positive rates. Στην εικόνα ο ταξινομητής A είναι πιο συντηρητικός από τον B.

Η διαγώνιος x=y αντιπροσωπεύει το αποτέλεσμα ενός ταξινομητή, όπως ο C που θα επέστρεφε τυχαία αποτελέσματα. Οι ταξινομητές που εμφανίζονται στο κάτω δεξί τριγωνικό κομμάτι του ROC γραφήματος, παρουσιάζουν επιδόσεις χειρότερες από την τυχαία εικασία. Συνεπώς επιθυμούμε οι ταξινομητές μας να εμφανίζονται στο άνω τριγωνικό αριστερό κομμάτι του γραφήματος.

Αξίζει να σημειωθεί ότι, λόγω του γεγονότος ότι ο χώρος απόφασης είναι συμμετρικός ως προς την διαγώνιο, ένας ταξινομητής με επίδοση χειρότερη από την τυχαία εικασία μπορεί να μετατραπεί σε ταξινομητή με επίδοση καλύτερη, προσθέτοντας μια άρνηση στην έξοδό του, δηλαδή αν εναλλάξουμε τις αποφάσεις ταξινόμησης που αυτός επιστρέφει. Στην προηγούμενη εικόνα, ο ταξινομητής E, που παρουσιάζει πολύ χειρότερη επίδοση από την τυχαία, με μια άρνηση στην έξοδό του μετατρέπεται στον ταξινομητή B, που έχει αποδεκτή απόδοση. Συνεπώς ενώ ένας ταξινομητής που εμφανίζεται πάνω στη διαγώνιο μπορούμε να πούμε ότι δεν διαθέτει καμία πληροφορία για τις κλάσεις, ένας ταξινομητής κάτω από την διαγώνιο κατέχει πληροφορίες για τις κλάσεις αλλά τις χρησιμοποιεί εσφαλμένα.

Συχνά χρησιμοποιείται σαν κριτήριο σύγκρισης ταξινομητών το εμβαδόν κάτω από την ROC καμπύλη (Area Under Curve - AUC), η οποία για διακριτούς ταξινομητές ξεκινάει από το σημείο (0,0), περνάει από το σημείο του ταξινομητή στον ROC χώρο και καταλήγει στο σημείο (1,1). Θεωρείται καλύτερος ένας ταξινομητής για τον οποίο μεγιστοποιείται το AUC.

ΚΕΦΑΛΑΙΟ 8:

Visualization

8.1 Εισαγωγή

Μια υπόθεση που συναντάται πολύ συχνά, πάνω στην οποία μάλιστα βασίζονται πολλές μέθοδοι κατηγοριοποίησης, υποστηρίζει ότι η κατηγοριοποίηση αντικειμένων συνδέεται πολύ στενά με την έννοια της ομοιότητας. Δηλαδή ότι όμοιες οντότητες τείνουν να ομαδοποιούνται μαζί στις ίδιες κατηγορίες.

Βέβαια, υπάρχουν διαφορετικές απόψεις σχετικά με την σχέση των δύο αυτών εννοιών, της ομοιότητας και της κατηγοριοποίησης.. Από τη μία, η ομοιότητα θεωρείται πολύ ευέλικτη έννοια για να αποτελέσει μια βάση για κατηγοριοποίηση. Οποιεσδήποτε δύο οντότητες μπορεί να ειδωθούν σαν όμοιες υπό ένα ορισμένο πρίσμα. Από την άλλη, η έννοια της ομοιότητας μπορεί να θεωρηθεί πολύ αυστηρή και περιοριστική για να λογισθεί η ποικιλία των κατηγοριοποιήσεων σύμφωνα με τον ανθρώπινο νου.

Και εδώ ανακύπτει ένα πολύ ενδιαφέρον ζήτημα_ πως αντιπροσωπεύονται οι οντότητες και οι ιδιότητές τους. Αν τα αντικείμενα περιγράφονται μόνο στα πλαίσια των αντιληπτικών ιδιοτήτων τους (οπτικές, ακουστικές κτλ.), τότε προφανώς η ομοιότητα αποδεικνύεται ανεπαρκής για πολλές περιπτώσεις κατηγοριοποίησης. Βέβαια αν ληφθεί υπόψη οποιοδήποτε είδος ιδιοτήτων, αφηρημένες ή σχεσιακές επιπλέον των αντιληπτικών, τότε η έννοια της ομοιότητας καθίσταται πιο ευέλικτη.

Είναι έκδηλο το γεγονός ότι οι έννοιες της κατηγοριοποίησης, της ομοιότητας και της αναπαράστασης (representation) των οντοτήτων και των ιδιοτήτων τους είναι πολύ στενά συνδεδεμένες. Φαίνεται πολύ απλό το μοντέλο που θέλει τον άνθρωπο να ξεκινά από μια ακριβή περιγραφή των οντοτήτων και των ιδιοτήτων τους, να ανευρίσκει ομοιότητες μεταξύ αυτών και τελικά να κατηγοριοποιεί μαζί τις πιο όμοιες από αυτές. Αναμφισβήτητα είναι πιο εύλογη η σκέψη ότι καθώς ο άνθρωπος οργανώνει τη γνώση του για τον κόσμο, αλλάζει συνεχώς τις αναπαραστάσεις των οντοτήτων ταυτόχρονα με τις αναδυόμενες κατηγοριοποιήσεις και τις αποφάσεις περι ομοιότητας.

8.2 Multidimensional Scaling (MDS)

Το Multidimensional scaling (MDS) αποτελεί ένα σύνολο μεθόδων για την ανάλυση της (αν)ομοιότητας των δεδομένων. Χρησιμοποιείται συνήθως για την γεωμετρική αναπαράσταση ενός συνόλου αντικειμένων ως σύνολο σημείων πάνω σε ένα γράφημα (συνήθως 2 ή 3 διαστάσεων) με βάση τις μεταξύ τους αποστάσεις/(αν)ομοιότητες.

Το MDS έχει τις ρίζες του στην ψυχομετρία, όπου προτάθηκε ως μέθοδος που θα βοηθούσε στη κατανόηση της ανθρώπινης κρίσης σχετικά με την ομοιότητα των μελών ενός συνόλου αντικειμένων. Ο Torgeson το 1952 πρότεινε για πρώτη φορά την πρώτη MDS μέθοδο και καθιέρωσε τον όρο. Από τότε το MDS χρησιμοποιείται σαν τεχνική ανάλυσης δεδομένων σε ευρύ φάσμα εφαρμογών, όπως κοινωνιολογία, φυσική, πολιτικές επιστήμες, βιολογία κ.α.

Ξεκινώντας από έναν πινάκα με τις αποστάσεις μεταξύ όλων των ζευγών των αντικειμένων, ο αλγόριθμος MDS προσδιορίζει μια θέση για κάθε αντικείμενο σε έναν m-διάστατο χώρο, χαμηλότερης διάστασης από τον αρχικό χώρο αναπαράστασης των δεδομένων, όπου θα γίνει η προβολή.

Τα δεδομένα που θέλουμε να προβάλλουμε είναι ένα σύνολο I που απαρτίζεται από n αντικείμενα (π.χ. μουσικά κομμάτια) για τα οποία ορίζεται μια συνάρτηση απόστασης.

Συμβολίζουμε με $\delta_{i,j}$ την απόσταση μεταξύ του αντικειμένου i από το j.

Αυτές οι αποστάσεις αποτελούν τα στοιχεία του πίνακα ανομοιομοτήτων Δ (dissimilarity matrix):

$$\Delta := \begin{bmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,n} \\ \delta_{2,1} & \delta_{2,2} & \cdots & \delta_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n,1} & \delta_{n,2} & \cdots & \delta_{n,n} \end{bmatrix}$$

Ο στόχος του MDS αλγόριθμου είναι, δεδομένου του πίνακα Δ και της διάστασης m του χώρου στον οποίο θα γίνει η προβολή, να υπολογίσει n διανύσματα $x_1, x_2, \dots, x_n \in \mathbb{R}^m$ τέτοια ώστε $\|x_i - x_j\| \approx \delta_{i,j} \ \forall i,j \in I$, όπου με $\|\cdot\|$ συμβολίζουμε κάποια διανυσματική νόρμα. Προσπαθεί δηλαδή να βρει το βέλτιστο σύστημα συντεταγμένων στον $\mathbb{R}^m$ χώρο και να ενσωματώσει τα n αντικείμενα στον χώρο αυτό, με τέτοιο τρόπο ώστε οι αποστάσεις μεταξύ των αντικειμένων να διατηρούνται.

Τα διανύσματα x_i που επιστρέφει ο αλγόριθμος *δεν είναι μοναδικά :* Αν χρησιμοποιηθεί η Ευκλείδεια απόσταση (νόρμα L_2) , όπως γίνεται στον classical MDS, τα διανύσματα αυτά μπορεί να είναι αυθαίρετα μετατοπισμένα, περιστρεμμένα και ανακλασμένα, δεδομένου ότι αυτοί οι μετασχηματισμοί δεν επηρεάζουν τις αποστάσεις των ζευγών $\|x_i - x_j\|$.

Συνήθως ο MDS διατυπώνεται σαν πρόβλημα βελτιστοποίησης, όπου τα x_i υπολογίζονται ως ελαχιστοποιητές κάποιας συνάρτησης κόστους. Η λύση στη συνέχεια προκύπτει από τεχνικές αριθμητικής βελτιστοποίησης.

Ανάλογα με την συνάρτηση κόστους και τον πίνακα (αν)ομοιότητας οι μέθοδοι MDS χωρίζονται σε δύο προσεγγίσεις: τη μετρική και την μη-μετρική. Αν ο πίνακας Δ είναι βασισμένος σε *μετρική απόσταση*, τότε μιλάμε για μετρική MDS, διαφορετικά αναφερόμαστε σε μη-μετρική MDS. Επιπλέον ορίζεται περεταίρω διαχωρισμός στις μεθόδους MDS με βάση

τον αριθμό των πινάκων που δέχεται σαν είσοδο ο αλγόριθμος και ορίζονται η κλασσική MDS, η αναδιπλούμενη MDS και η σταθμισμένη MDS ([Zhang et al. 2010]).

8.2.1 Μετρική Κλασσική MDS (CMDS) :

Είναι γνωστή και σαν Ανάλυση Κυρίων Συντεταγμένων (Principal Coordinates Analysis)ή Κλιμάκωση κατά Torgeson (Torgerson Scaling). Το χαρακτηριστικό της κλασσικής MDS μεθόδου είναι ότι διαχειρίζεται μόνο έναν πίνακα ανομοιότητας. Χρησιμοποιεί την Ευκλείδεια απόσταση για να αναπαραστήσει τις ανομοιότητες/αποστάσεις μεταξύ των αντικειμένων. Δηλαδή η απόσταση μεταξύ δύο σημείων i και j είναι

$d_{ij} = \sqrt{\sum_a (x_{ia} - x_{ja})^2}$, όπου x_{ia} η συντεταγμένη του σημείου i στη διάσταση α.

Το προτέρημα της κλασικής MDS είναι ότι παρέχει μια αναλυτική λύση χωρίς τη χρήση επαναλήψεων. Δεδομένου ενός πίνακα ανομοιοτήτων μεταξύ ζευγών αντικειμένων ο αλγόριθμος επιστρέφει έναν πίνακα συντεταγμένων των οποίων η διάταξη ελαχιστοποιεί μια συνάρτηση κόστους.

Ο κλασσικός MDS αλγόριθμος βασίζεται στο γεγονός ότι μπορούμε να εξάγουμε τον πίνακα συντεταγμένων X από την ανάλυση ιδιοτιμών (eigenvalue decomposition) του πίνακα $B = XX'$.

Αρχικά από τον πίνακα απόστασης Δ υπολογίζουμε τις τετραγωνισμένες αποστάσεις $\Delta^{(2)} = [\delta^2]$ και στη συνέχεια εφαρμόζουμε το διπλό κεντράρισμα(double centering) $B = -\frac{1}{2} H \Delta^{(2)} H$, χρησιμοποιώντας τον πίνακα $H = I - n^{-1}11'$, όπου n είναι το σύνολο των αντικειμένων.

Στη συνέχεια εξάγουμε τις m μεγαλύτερες θετικές ιδιοτιμές $\lambda_1, \dots, \lambda_m$ του B και τα m αντίστοιχα ιδιοδιανύσματα $e_1, \dots, e_m$.

Μια m-διάστατη χωρική διάταξη n αντικειμένων προκύπτει τελικά από τον πίνακα συντεταγμένων $X = E_m \Lambda_m^{1/2}$, όπου E_m είναι ο πίνακας των m ιδιοδιανυσμάτων και Λ_m ο διαγώνιος πίνακας των m ιδιοτιμών του B.

Η συνάρτηση κόστους που χρησιμοποιείται από την κλασική MDS είναι

$f(\Delta) = \left\| H\Delta^{(2)}H - H\Delta^* H \right\|$

8.2.2 Μη μετρική CMDS (Non-metric MDS)

Η υπόθεση ότι οι ανομοιότητες μεταξύ των αντικειμένων προσεγγίζονται με αποστάσεις τις αποστάσεις είναι ίσως αρκετά περιοριστική όταν η μέθοδος MDS χρησιμοποιείται για την εξερεύνηση του αντιληπτικού χώρου των ανθρώπινων υποκειμένων. Για την αντιμετώπιση του προβλήματος αυτού ο Shepard (1962) και ο Kruskal (1964) ανέπτυξαν την μη μετρική MDS.

Στη μη μετρική MDS ([Kruskal et al., 1986])χρησιμοποιείται μόνο η πληροφορία που είναι σχετική με την τακτική θέση στις ανομοιότητες για την κατασκευή της χωρικής διάταξης. Υπολογίζεται ένας μονοτονικός μη γραμμικός μετασχηματισμός, ο οποίος οδηγεί σε διαβαθμισμένες ανομοιότητες. Εξαιτίας της μονοτονίας του μετασχηματισμού οι μεγάλες ή μικρές ανομοιότητες θα απεικονίζονται ως μεγάλες ή μικρές αποστάσεις στη διάταση που θα επιστρέφει ο αλγόριθμος. Οι βέλτιστα κλιμακωμένες ανομοιότητες συχνά αποκαλούνται αποκλίσεις(disparities) $\hat{d} = f(p)$.

Το πρόβλημα της μη μετρικής MDS είναι η αναζήτηση της διάταξης σημείων η οποία ελαχιστοποιεί της τετραγωνισμένες διαφορές μεταξύ των βέλτιστα κλιμακωμένων ανομοιοτήτων και των αποστάσεων μεταξύ των σημείων. Δηλαδή αν συμβολίσουμε με d τις ευκλείδειες αποστάσεις μεταξύ των σημείων που καθορίζει ο MDS, με p το διάνυσμα των ανομοιοτήτων (το άνω τριγωνικό κομμάτι του πίνακα ανομοιοτήτων Δ) και έστω $f(p)$ ένας μονοτονικός μετασχηματισμός του p, τότε οι συντεταγμένες που θα επιστρέψει ο αλγόριθμος θα πρέπει να ελαχιστοποιούν το αποκαλούμενο κριτήριο Stress1:

$$STRESS1 = \sqrt{\frac{\sum_{i<j}\left(f(p_{ij}) - d_{ij}\right)^2}{\sum_{i<j} d_{ij}^2}}$$

Ο κορμός του μη μετρικού MDS αλγορίθμου είναι μια διπλή διαδικασία βελτιστοποίησης. Αρχικά πρέπει να βρεθεί ένας βέλτιστος μονοτονικός μετασχηματισμός των ανομοιοτήτων και στη συνέχεια τα σημεία της διάταξης πρέπει να κατανεμηθούν βέλτιστα, έτσι ώστε οι αποστάσεις τους να αντιστοιχούν στις κλιμακούμενες ανομοιότητας όσο γίνεται καλύτερα.

Τα βασικά βήματα του αλγορίθμου ακολουθούν:

1]. Υπολογίζεται μια τυχαία κατανομή σημείων
2]. Υπολογίζονται τις αποστάσεις d μεταξύ αυτών των σημείων.
3]. Ανευρίσκει έναν βέλτιστο μονοτονικό μετασχηματισμό των ανομοιοτήτων προκειμένου να εξασφαλίσει βέλτιστα κλιμακωμένα δεδομένα $f(p)$.
4]. Ελαχιστοποιεί το stress μεταξύ των βέλτιστα κλιμακούμενων δεδομένων και των αποστάσεων που έχει υπολογίσει στο βήμα 2, βρίσκοντας μια καινούργια κατανομή σημείων.
5]. Συγκρίνοντας το stress με κάποιο κατώφλι, αν το stress είναι αρκετά χαμηλό ο αλγόριθμος τερματίζει επιστρέφοντας την τελευταία κατανομή σημείων που υπολογιστικέ, διαφορετικά επιστρέφει στο βήμα 2.

Το κριτήριο stress μπορεί να χρησιμοποιηθεί επίσης και για την αξιολόγηση του πόσο καλά έγινε η προσαρμογή των σημείων(goodness of fit) με τη χρήση του MDS. Μια μικρή τιμή stress καταδεικνύει μια λύση καλά προσαρμοσμένων σημείων.

Πρέπει όμως πάντα να λαμβάνεται υπόψη ότι η τιμή του stress μειώνεται καθώς ο αριθμός των διαστάσεων της λύσης αυξάνει. Δηλαδή μια λύση σε δύο διαστάσεις έχει πάντα μεγαλύτερο stress από ότι μια λύση σε τρεις διαστάσεις.

Για την περίπτωση όπου χρησιμοποιείται το κριτήριο stress του Kruskal μια καλή τιμή του stress είναι το 0.05 ή και μικρότερο, ενώ οι τιμές stress πάνω από 0.20 καταδεικνύουν μη καλά προσαρμοσμένες λύσεις της MDS.

Επειδή η τιμή του stress από μόνη της αποτελεί μια ακαθόριστη ένδειξη της καλής προσαρμογής έχουν καθιερωθεί δύο επιπλέον τεχνικές για την αξιολόγηση της καταλληλότητας μιας λύσης της MDS: το διάγραμμα Shepard και το scree διάγραμμα.

Στο διάγραμμα scree η τιμή του stress απεικονίζεται σε σχέση με το μέγεθος της διάστασης της λύσης. Αφού το stress φθίνει μονοτονικά καθώς αυξάνουν οι διαστάσεις, ψάχνουμε για τον μικρότερο αριθμό διαστάσεων για τον οποίο προκύπτει αποδεκτή τιμή stress. Ένα σημείο γονάτου στο διάγραμμα υποδεικνύει ότι χρησιμοποιώντας περισσότερες διαστάσεις στη λύση θα οδηγούμασταν σε ελάχιστη βελτίωση όσον αφορά το stress. Συνεπώς η καλύτερα προσαρμοσμένη λύση MDS έχει τόσες διαστάσεις όσες ο αριθμός των διαστάσεων στο σημείο γονάτου στο scree διάγραμμα.

Το διάγραμμα Shepard απεικονίζει την σχέση μεταξύ των ανομοιοτήτων και των αποστάσεων μεταξύ των σημείων της διάταξης. Μικρότερη εξάπλωση των σημείων στο χώρο φανερώνει μια καλή προσαρμογή των σημείων. Στη μη μετρική MDS, η ιδανική θέση των σημείων στο διάγραμμα Shepard είναι μια μονοτονοειδώς αύξουσα γραμμή που περιγράφει τις αποκαλούμενες αποκλίσεις(disparities).

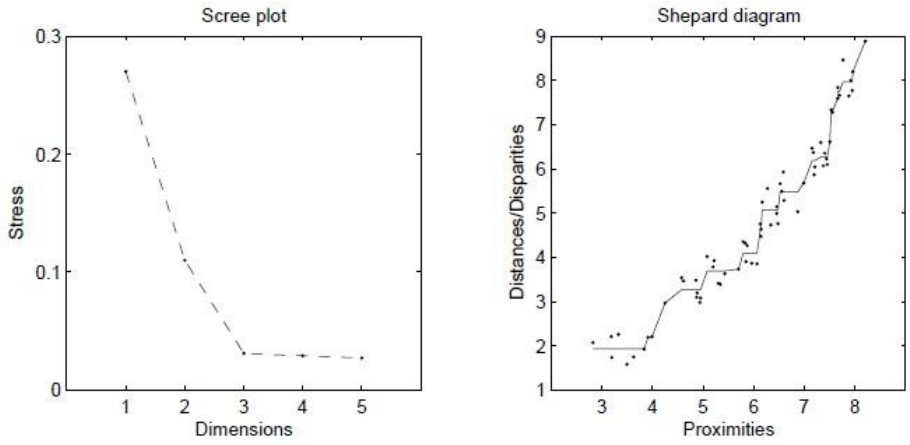


Figure 3: Left panel: Scree plot displaying an elbow at three dimensions. Right panel: Shepard diagram with the optimally scaled proximities.

ΚΕΦΑΛΑΙΟ 9:

Πειραματικό Μέρος

9.1 Επιλογή Dataset

Για το πειραματικό κομμάτι της εργασίας χρησιμοποιήθηκε το dataset του Magnatagatune και η υλοποίηση έγινε με τη βοήθεια του προγραμματιστικού περιβάλλοντος MATLAB (versions: 2010b και 2012a) και τα διάφορα toolboxes που παρέχονται με αυτό στα οποία θα αναφερθούμε εκτενέστερα στη συνέχεια.

Ύστερα από έρευνα σχετικά με το ποιο μουσικό dataset θα χρησιμοποιούσαμε για την υλοποίηση της εργασίας καταλήξαμε στο Magnatagatune για τους εξής λόγους:

- Πρώτον το dataset περιέχει μεγάλο πλήθος μουσικών κομματιών (5405), επιλεγμένα από αρκετά μουσικά είδη και από 230 διαφορετικούς καλλιτέχνες.
- Το dataset περιλαμβάνει σχολιασμούς κομματιών από χρήστες με την μορφή tags (188 μοναδικά tags) αλλά και με τη μορφή ψήφων ανομοιότητας μεταξύ τριάδων κομματιών (7650 ψήφοι για 533 τριάδες κομματιών). Τόσο τα tags όσο και οι ψήφοι ανομοιότητας συλλέχτηκαν από το διαδικτυακό παιχνίδι TagATune που κατασκευάστηκε από τον Edith Law.
- Επιπλέον περιλαμβάνεται μια λεπτομερής ανάλυση της δομής και του μουσικού περιεχομένου των κομματιών υπό τη μορφή ενός xml αρχείου για κάθε μουσικό απόσπασμα, η οποία προέρχεται από την υπηρεσία Echo Nest.
- Τέλος, με το dataset παρέχονται και μουσικά αποσπάσματα 29 δευτερολέπτων σε μορφή mp3 (mono, 16kHz, 32kbps) για κάθε κομμάτι προερχόμενα από τα δημοσιευμένα κομμάτια της υπηρεσίας Magnatune.com, γεγονός που αποτέλεσε πολύ ελκυστικό χαρακτηριστικό, δεδομένου ότι μας επιτρέπει να κάνουμε μόνοι μας εξαγωγή επιπλέον χαρακτηριστικών πέραν αυτών που παρέχονται από την ανάλυση του Echo Nest. Συνολικά περιέχονται 25863 αποσπάσματα των 29 δευτερολέπτων προερχόμενα από 5405 κομμάτια. (Για κάθε κομμάτι ανάλογα με τη διάρκειά του παρέχεται και ο ανάλογος αριθμός αποσπασμάτων.)

Το Magnatune είναι μια διαδικτυακή δισκογραφική εταιρεία που ιδρύθηκε το 2003 από τον John Buckman στη Καλιφόρνια με σκοπό να βοηθήσει τους καλλιτέχνες να εκθέσουν το έργο τους στο κοινό παρακάμπτοντας τις οικονομικές και νομικές δυσκολίες που προκύπτουν κατά

την έκδοση ενός μουσικού δίσκου σε μια κανονική δισκογραφική εταιρία και κατά τη διανομή της από μουσικούς εμπόρους. Παρατηρώντας ότι πολύ συχνά οι καλλιτέχνες κερδίζουν ελάχιστα από την πώληση των δίσκων τους και επιπλέον εγκλωβίζονται νομικά από τις δισκογραφικές εταιρίες, δημιούργησε την Magnatune η οποία λειτουργεί με Άδεια Creative Commons (αντί για το σύστημα "all rights reserved" που ακολουθείται από τις κλασσικές δισκογραφικές) και υποστηρίζει ότι αποδίδει στους καλλιτέχνες το 50% της τιμής πώλησης των έργων τους καθώς και το 50% του ποσού που προκύπτει από τις συνδρομές των χρηστών[8].

Ο Buckman σε συνεργασία με τους E. Law και P.Lamere κατασκεύασαν τη βάση δεδομένων που χρησιμοποιούμε στην εργασία αυτή.

Το διαδικτυακό παιχνίδι TagATune εντάσσεται στα παιχνίδια που ονομάζονται «*παιχνίδια με σκοπό*» ("games with a purpose" - GWAPs) . Τα παιχνίδια αυτού του είδους έχουν κατασκευαστεί με σκοπό τη συλλογή δεδομένων που θα χρησιμοποιηθούν σε προβλήματα μηχανικής μάθησης, εκμεταλλευόμενα τη διάθεση του χρήστη για ψυχαγωγία.

Έτσι στο παιχνίδι TagATune δύο παίκτες ακούνε ταυτόχρονα το ίδιο ή διαφορετικό μουσικό απόσπασμα 29 δευτερολέπτων[9] (χωρίς να ξέρει κανένας αν ο συμπαίκτης του ακούει το ίδιο) και στη συνέχεια ερωτούνται να περιγράψουν το κομμάτι που μόλις άκουσαν. Βασιζόμενοι στις περιγραφές (tags) που έδωσαν οι δύο παίκτες καλούνται να αποφασίσουν αν τελικά άκουσαν το ίδιο μουσικό κομμάτι ή όχι. Επιπλέον στον γύρο μπόνους που περιλαμβάνει το παιχνίδι, οι δύο χρήστες προσπαθούν να συμφωνήσουν για το πιο από τα τρία μουσικά αποσπάσματα που τους παρουσιάστηκαν ήταν το πιο διαφορετικό σε σχέση με τα υπόλοιπα δύο.

Με το παιχνίδι αυτό συλλέχθηκαν τα 188 tags που παρέχονται με το dataset και από τον γύρο bonus προέκυψαν οι 7650 ψήφοι για 346 διαφορετικές τριπλέτες μουσικών αποσπασμάτων (οι οποίες αναφέρονται σε 1019 μουσικά αποσπάσματα).

9.2 Κατασκευή Διανύσματος Μουσικών Χαρακτηριστικών

9.2.1 Echo Nest Analyzer

Το EchoNest αποτελεί μια εταιρία παροχής μουσικών πληροφοριών σε προγραμματιστές και εταιρίες media, που ιδρύθηκε το 2005 από τους διδακτορικούς αποφοίτους του MIT Brian Whitman και Tristan Jehan. Η εταιρία με χρήση τεχνικών εξόρυξης πληροφορίας και ψηφιακής ανάλυσης σήματος παρέχει υπηρεσίες όπως μουσική πρόταση (music recommendation) , παραγωγή playlist, ακουστική ανάλυση και μουσική αναγνώριση. Επιπλέον η εταιρία παρέχει τα μουσικά της δεδομένα μέσω ενός API σε προγραμματιστές προκειμένου να κατασκευάσουν τις δικές τους μουσικές εφαρμογές.

[8] Αναφέρεται οτί 25% των χρημάτων από κάθε συνδρομή αποδίδεται στους καλλιτέχνες που ακούει on-line ενας χρήστης και 25% μοιράζρται στους καλλιτέχνες των οποίων τα κομμάτια κατεβάζει ο χρήστης.

[9] Ο λόγος που χρησιμοποιούνται μικρά μουσικά αποσπάσματα πιστεύεται από τους δημιουργούς του παιχνιδιού ότι διασφαλίζει την άμεση σχέση μεταξύ του περιεχομένου της μουσικής και των περιγραφών που παρέχονται από τον χρήστη.

Στο dataset του Magnatagatune περιέχεται από ένα xml για κάθε μουσικό κομμάτι, το οποίο περιλαμβάνει περιγραφές μουσικού περιεχομένου που έχουν προκύψει από την ανάλυση του Echonest. Το Echonest Analyzer παρέχει για κάθε κομμάτι ένα αρχείο μορφοποιημένο σύμφωνα με το πρότυπο JSON (JavaScript Object Notation).

Σε κάθε τέτοιο αρχείο περιέχονται οι ακόλουθες πληροφορίες:

- **μεταδεδομένα (metadata)** : πληροφορίες σχετικές με τη διαδικασία της ανάλυσης και πληροφορίες για το κομμάτι
- **δεδομένα κομματιού (track data):**
  - **ένδειξη ρυθμού(time signature)**: μια συνολικά εκτιμώμενη ένδειξη ρυθμού για το κομμάτι. Αποτελεί μια παγκόσμια σύμβαση για τον καθορισμό του αριθμού των beat που περιέχονται σε κάθε μέτρο (bar)
  - **Κλίμακα (key)**: η συνολικά εκτιμώμενη κλίμακα του κομματιού.
  - **mode** : Προσδιορίζει το modality (μείζον ή ελάσσων) του κομματιού, το είδος της κλίμακας από την οποία προήλθε το μελωδικό περιεχόμενο
  - **Ρυθμός (tempo)** : μια συνολική εκτίμηση του ρυθμού του κομματιού σε beats ανά λεπτό (bpm) Αποτελεί τη ταχύτητα ενός κομματιού και προκύπτει κατευθείαν από τη μέση διάρκεια του beat
  - **Ένταση (loudness)** : την συνολική ένταση του κομματιού σε dB
  - **Διάρκεια (duration)** : η ακριβής διάρκεια του κομματιού όπως καταγράφεται από τον αποκωδικοποιητή ήχου
  - **Τέλος του fade in (end of fade in)** : σε δευτερόλεπτα
  - **Αρχή του fade out (start of fade out)** : σε δευτερόλεπτα
- **Διαδοχικά Δεδομένα (sequenced data)**: ο αναλυτής σπάει το ακουστικό κομμάτι σε μουσικά στοιχεία που εμφανίζονται σε αλληλουχίες στον χρόνο. Από τη μικρότερη στη μεγαλύτερη αυτές είναι:
  - **segments** : Ένα σύνολο από μουσικές οντότητες (συνήθως κάτω από ένα δευτερόλεπτο) η κάθε μια σχετικά ομοιόμορφη σε ηχόχρωμα και αρμονία.
    - **loudness_start** : υποδεικνύει το επίπεδο έντασης κατά την έναρξη του segment
    - **loudness_max_time** : το αρχικό σημείο μέσα στο segment όπου εμφανίζεται η μέγιστη ένταση
    - **loudness_max** : η τιμή της μέγιστης έντασης
    - **Συντελεστές Τονικότητας (pitch coefficients)** : διάνυσμα 12 διαστάσεων με κάθε διάσταση να αντιστοιχεί σε μια από τις 12 τονικές κλάσεις με τιμές [0,1]
    - **Συντελεστές Ηχοχρώματος (timbral coefficients)** : διάνυσμα 12 διαστάσεων. (Αναλύεται περισσότερο στη συνέχεια)
    - **Αντιληπτική Εμφάνιση (perceptual onset)**
    - **Διάρκεια (duration)** : σε δευτερόλεπτα
  - **tatums** : μια λίστα από δείκτες tatum σε δευτερόλεπτα. Τα tatums αναπαριστούν τη χαμηλότερη ακολουθία παλμών που ο ακροατής

συμπεραίνει ενστικτωδώς από τον χρονισμό των αντιλαμβανόμενων μουσικών γεγονότων.

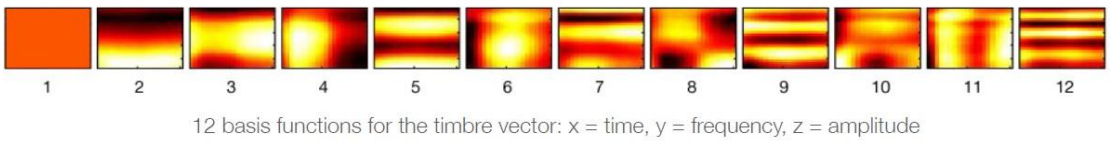
- **beats** : μια λίστα από δείκτες beat σε δευτερόλεπτα. Ένα beat αποτελεί τη βασική μονάδα χρόνου σε ένα μουσικό κομμάτι. Τα beat είναι συνήθως πολλαπλάσια των tatums
- **Μέτρα (bars)** : μια λίστα από δείκτες μέτρων σε δευτερόλεπτα. Ένα μέτρο αποτελεί ένα κομμάτι του χρόνου ορισμένα σαν ένα δεδομένο αριθμό beat
- **Ενότητες (sections)** : ένα σύνολο από δείκτες ενοτήτων σε δευτερόλεπτα. Οι ενότητες ορίζονται από μεγάλες μεταβολές στο ρυθμό ή το ηχόχρωμα.

9.2.2 Echo Nest Timbral (ENT) Χαρακτηριστικά

Στο σημείο αυτό αξίζει να αναφερθούμε σους συντελεστές ηχοχρώματος που παρέχονται από την EchoNest ανάλυση ([Jehan et al., 2011]), καθώς διαφέρουν από τα MFCC που αναφέραμε προηγουμένως. Ονομάζονται ENT Features και προέρχονται από την ερευνητική εργασία του Jehan, ο οποίος όταν κατασκεύασε μια αναπαράσταση του ηχοχρώματος με 12 συντελεστές μέσω PCA, αποφάσισε να μην την δημοσιεύσει και να την κρατήσει κλειστή, χρησιμοποιώντας την στην ανάλυση του Echonest. Τα ENT αποτελούν ουσιαστικά μια πιο συμπαγή περιγραφή των κλασσικών MFCC και συνθέτουν ένα διάνυσμα με 12 τιμές χωρίς άνω ή κάτω φράγμα, κεντραρισμένες γύρω από το 0. Οι τιμές αυτές αποτελούν υψηλού επιπέδου γενίκευση της φασματικής επιφάνειας, ταξινομημένες σύμφωνα με το βαθμό σημαντικότητας τους και ομοιάζουν με το αποτέλεσμα μιας δισδιάστατης συνέλιξης με 12 διαφορετικά φίλτρα, τα οποία αντιστοιχούν σε χαρακτηριστικά φασματικά σχήματα. Η πρώτη διάσταση αναπαριστά τη μέση ένταση του segment, το δεύτερο τονίζει την φωτεινότητα, η τρίτη σχετίζεται με την ομαλότητα του ήχου και το τέταρτο σχετίζεται με ήχους με πιο δυνατό attack. Για τις υπόλοιπες διαστάσεις του ENT διανύσματος δεν διατίθενται περισσότερες πληροφορίες. Στην εικόνα που ακολουθεί παρουσιάζονται οι 12 συναρτήσεις βάσης. Το πραγματικό ηχόχρωμα ενός segment περιγράφεται σαν γραμμικός συνδυασμός αυτών των 12 συναρτήσεων βάσης σταθμισμένων με τις τιμές των συντελεστών ENT:

$$timbre = c_1 \cdot b_1 + c_2 \cdot b_2 + \cdots + c_{12} \cdot b_{12}$$

,όπου με c_i αναπαρίστανται οι 12 συντελεστές και b_i οι 12 συναρτήσεις βάσης.



Εικόνα V Οι 12 συναρτήσεις βάσης που χρησιμοποιούνται στον υπολογισμό του διανύσματος ηχοχρώματος

Για την υλοποίηση των πειραμάτων επιλέξαμε να χρησιμοποιήσουμε τα 12 ENT features (όπως στα [Wolf et al., 2012],[Stober et al., 2011]),τα οποία παρέχονται από το

Echonest ως μια πιο συμπαγή μορφή των MFCC που χρησιμοποιούνται ως επί το πλείστον για εφαρμογές μουσικής ομοιότητας. Δεδομένου ότι ο αριθμός των συντελεστών MFCC που συναντάται στη βιβλιογραφία ανέρχεται στους 13- 25 ([Bozzon et al., 2008],[Schnitzer et al., 2011], [Logan et al., 2001]), καθιστά αρκετά ελκυστική την αναπαράσταση του ηχοχρώματος που επινόησε ο Jehan.

9.2.3 Αναπαράσταση ENT Χαρακτηριστικών

Το πρόβλημα που ανέκυψε στη συνέχεια αφορούσε τον τρόπο που θα αναπαριστούσαμε το σύνολο των ENT χαρακτηριστικών για κάθε κομμάτι, καθώς οι συντελεστές αυτοί είναι υπολογισμένοι για κάθε segment ξεχωριστά. Στη βιβλιογραφία συναντάται αρκετά συχνά η χρήση πολυμεταβλητών Γκαουσσιανών ([Schnitzer et al., 2011]) ή Μοντέλων Μίξης Γκαουσσιανών (Gaussian Mixture Models - GMM)([McFee., 2009]), συνήθως τριών Γκαουσσιανών ([Aucouturier et al., 2002]), για την αναπαράσταση του συνόλου των MFCC για κάθε κομμάτι. Στο [Aucouturier et al., 2004] , όπου μελετάται η επίδραση του αριθμού των Γκαουσσιανών που θα επιλεγούν για αναπαράσταση των MFCC ενός κομματιού, αναφέρεται ότι ο ιδανικός αριθμός είναι 50 Γκαουσσιανές για εφαρμογές μουσικής ομοιότητας. Με τη χρήση ενός τέτοιου μοντέλου αναπαράστασης των MFCC για να υπολογίσουμε την απόσταση μεταξύ δύο κομματιών θα έπρεπε να υπολογίσουμε την απόσταση μεταξύ δύο GMM και συνεπώς να χρησιμοποιήσουμε την απόκλιση Kullback – Leibler (Kullback – Leibler Divergence), η οποία όμως δεν ορίζεται σε κλειστή μορφή για δύο GMM. Επιπλέον λαμβάνοντας υπόψη και τη δαπανηρή, από άποψη χρόνου, εκπαίδευση που επιβάλει ένα τέτοιο μοντέλο, καθώς βασίζεται τον Expectation Minimization Αλγόριθμο, αναζητήσαμε μια διαφορετικού είδους αναπαράσταση.

Στο [Stober et al., 2011] αναφέρεται η αναπαράσταση των ENT features μέσω της μέσης τιμής και την τυπικής απόκλισης που εμφανίζει κάθε διάσταση του διανύσματος μέσα σε ένα μουσικό απόσπασμα. Προτείνεται δηλαδή αναπαράσταση του ηχοχρώματος ενός μουσικού αποσπάσματος από ένα διάνυσμα 24 διαστάσεων, αποτελούμενο από 12 διαστάσεις προερχόμενες από τη μέση τιμή ανά διάσταση των ENT χαρακτηριστικών για όλα τα segments ενός μουσικού αποσπάσματος και άλλες 12 προερχόμενες από την τυπική απόκλιση αυτών.

Στη παρούσα εργασία χρησιμοποιήσαμε στην **πρώτη εκτέλεση** (στα αποτελέσματα αναφέρεται ως **24D**) του πειράματός μας σαν διάνυσμα χαρακτηριστικών το διάνυσμα αυτών των **24 διαστάσεων των ENT (12 τιμές για μέση τιμή και 12 για τυπική απόκλιση)**.

9.2.4 Επιλογή Επιπλέον Χαρακτηριστικών

Στη **δεύτερη εκτέλεση** (στα αποτελέσματα αναφέρεται ως **29D**)του πειράματος επιλέξαμε να διευρύνουμε το διάνυσμα αυτό με κάποια επιπλέον χαρακτηριστικά προκειμένου να μελετήσουμε την επίδραση αυτών στην απόδοση των ταξινομητών. Επιλέχθηκαν γενικά χαρακτηριστικά: η **Συνολική Εκτίμηση του Ρυθμού (tempo)** του κομματιού και τη **Συνολική Ένταση (loudness)** όπως αυτά παρέχονται από την ανάλυση του EchoNest, καθώς και την **Ενέργεια Βραχέως Χρόνου (RMS)**, το **Ρυθμό Μηδενισμών (Zero Crossing Rate - ZCR)** (όπως στο [Tzanetakis et al., 2001]) καθώς και το **Σημείο**

Φασματικού Roll off (όπως στα [Bozzon et al., 2008],[Wang et al., 2011]) . Τα τελευταία τρία χαρακτηριστικά τα εξάγαμε με την βοήθεια του MIR Toolbox 1.4.

Το MIR Toolbox ([Lartillot, 2011]) αποτελεί ένα σύνολο συναρτήσεων γραμμένες να λειτουργούν στο περιβάλλον Matlab, με κύριο στόχο την εξαγωγή μουσικών χαρακτηριστικών από ηχητικά αρχεία. Κατασκευάστηκε στα πλαίσια του προγράμματος Brain Tuning, το οποίο είχε σαν έναν από τους κυρίους στόχους του την ανάλυση της σχέσης των μουσικών χαρακτηριστικών με τα προκαλούμενα από τη μουσική συναισθήματα και την σχετιζόμενη νευρωνική δραστηριότητα.

Με την βοήθεια του εργαλείου αυτού παρέχεται η δυνατότητα μιας βασικής ανάλυσης σήματος για μουσικά ηχητικά αρχεία. Οι συναρτήσεις που χρησιμοποιήθηκαν για την εξαγωγή των χαρακτηριστικών είναι οι:mirtempo() , mirrms() και η mirzerocross().

Η εξαγωγή των χαρακτηριστικών δεν έγινε για το σύνολο της μουσικής βάσης αρχικά, αλλά μόνο για τα 1019 κομμάτια εκείνα για τα οποία διατίθενται ψήφοι ανομοιότητας από το παιχνίδι TagATune [Law et al., 2009].

9.3 Γράφος Ανομοιότητας

Οι ψήφοι ανομοιότητας παρέχονται στο αρχείο τύπου csv με όνομα "comparisons_final" όπου κάθε γραμμή του αρχείου αυτού περιέχει για κάθε κομμάτι: τον αναγνωριστικό αριθμό του στη βάση, τον αριθμό ψήφων ανομοιότητας που έχει λάβει καθώς και τη διαδρομή στο φάκελο με τα audio και τα xml αρχεία, όπου μπορεί να βρεθεί. Μεταφέροντας τα δεδομένα αυτού του αρχείου σε μορφή πίνακα στο Matlab.

Στη συνέχεια ανέκυψε το ζήτημα του τρόπου ερμηνείας του πίνακα αυτού και προέκυψαν τρεις τρόποι.

- Σύμφωνα με τον απλούστερο τρόπο ερμηνείας του πίνακα, για κάθε τριάδα κομματιών προκύπτει ένας μοναδικός «νικητής» ανομοιότητας, το κομμάτι που έχει λάβει τους περισσότερους ψήφους. Για να προκύψουν πιο αξιόπιστα αποτελέσματα έπρεπε να αγνοηθούν οι τριάδες εκείνες για τις οποίες δεν προέκυπτε κανένας νικητής (ισοπαλία) ή η διαφορά των ψήφων μεταξύ του νικητή και κάποιου άλλου από τα εναπομείναντα δύο κομμάτια ήταν μοναδιαία. Για τον λόγο αυτόν αφαιρέθηκαν 87 τριάδες.

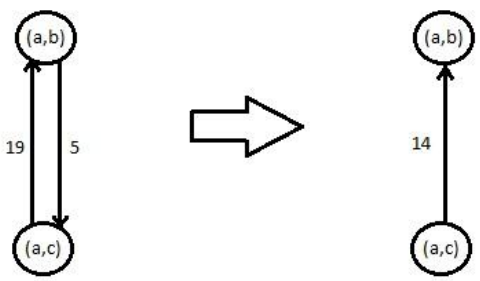
  Για τον νικητή κάθε τριάδας, δηλαδή το πιο ανόμοιο κομμάτι μεταξύ των τριών, κατασκευάζουμε περιορισμούς σχετικών αποστάσεων με βάση τη σχέση (I). Για κάθε τριάδα συνεπώς προκύπτουν δύο περιορισμοί. Στη συνέχεια κατασκευάζουμε έναν κατευθυνόμενο γράφο όπως αναλύθηκε σε προηγούμενο κεφάλαιο όπου τοποθετούμε μία ακμή για κάθε περιορισμό. Συνεπώς στο γράφο προκύπτουν 892 ακμές , 718 από τις οποίες είναι μοναδικές.

  Στη συνέχεια προκειμένου να καταλήξουμε σε κατευθυνόμενο ακυκλικό γράφημα (DAG) αφαιρούμε τους κύκλους μήκους 2 που εμφανίζονται στον γράφο, δεδομένου ότι οι κύκλοι αυτοί προκαλούν ασυνέπειες στο πρόβλημά μας, και καταλήγουμε σε ένα γράφο 818 ακμών (644 από αυτές είναι μοναδικές). Εξετάζοντας αυτόν τον γράφο για

κύκλους, διαπιστώνουμε ότι αποτελεί πλέον ένα DAG. Στη συνέχεια μπορούμε να κατασκευάσουμε το διάνυσμα εισόδου του ταξινομητή, δημιουργώντας ένα θετικό και ένα αρνητικό παράδειγμα για κάθε ακμή του DAG.

- Σύμφωνα με τον δεύτερο τρόπο ερμηνείας του πίνακα ψήφων ανομοιοτήτων, εκμεταλλευόμενοι τους ψήφους και των τριών κομματιών κάθε τριάδας, κατασκευάζουμε περιορισμούς σχετικών αποστάσεων βασιζόμενοι στη σχέση (I), για κάθε κομμάτι που έχει λάβει τουλάχιστον έναν ψήφο ανομοιότητας. Κατασκευάζουμε συνεπώς, με παρόμοιο τρόπο όπως πριν, έναν κατευθυνόμενο γράφο μοναδιαίων ακμών, εισάγοντας 6 το πολύ περιορισμούς για κάθε τριάδα, προσθέτοντας μια ακμή από τον κόμβο (a,b) προς τους κόμβους (a,c) και (b,c) αν τουλάχιστον ένας χρήστης έχει ψηφίσει το κομμάτι c σαν το πιο ανόμοιο μεταξύ των τριών και αντίστοιχα για τα a και b. Με τον τρόπο αυτό δημιουργείται ένας γράφος, ο οποίος αποτελείται από 1598 μοναδιαίες ακμές, όπου αφαιρώντας τους κύκλους μήκους 2 που εμφανίζονται στο γράφο (δηλαδή τους αντιφατικούς περιορισμούς) προκύπτουν 860 ακμές.
  Στη συνέχεια ελέγχοντας τον γράφο αυτό για πιθανούς κύκλους, διαπιστώνουμε ότι αυτός αποτελεί κατευθυνόμενο ακυκλικό γράφημα και υπολογίζουμε το transitive reduction του γραφήματος αυτού. Με τον υπογράφο που προκύπτει (αποτελούμενο από 674 ακμές) κατασκευάζουμε το σύνολο εκπαίδευσης δημιουργώντας ένα θετικό και ένα αρνητικό παράδειγμα για κάθε μια από τις προκύπτουσες ακμές.

- Ο τρίτος τρόπος ερμηνείας του πίνακα, ομοιάζει πολύ με τον δεύτερο, με την διαφορά όμως, ότι κατασκευάζουμε βεβαρυμένο κατευθυνόμενο γράφο, όπου το βάρος κάθε ακμής αντιπροσωπεύει το πλήθος των ψήφων που αντιστοιχεί σε κάθε σύγκριση. Έτσι αν το κομμάτι c της τριάδας (a,b,c) έχει λάβει w ψήφους ανομοιότητας, προσθέτουμε στον γράφο ακμές βάρους w από τον κόμβο (a,b) στους (a,c) και (b,c). Το αποτέλεσμα αυτής της διαδικασίας είναι 1598 βεβαρυμένες ακμές, ή αν αντικαταστήσουμε κάθε βεβαρυμένη ακμή με τόσες μοναδιαίες ακμές όσες το βάρος αυτής, προκύπτουν 15300 μοναδιαίες ακμές. Εξαλείφοντας τους κύκλους μήκους 2 όπως παρουσιάζεται στην επόμενη εικόνα, προκύπτουν 860 βεβαρυμένες ακμές (ή 6898 μοναδιαίες), οι οποίες αποτελούν τη βάση για τη κατασκευή του συνόλου εκπαίδευσης.



Δεδομένου ότι ο τρίτος τρόπος ερμηνείας εκμεταλλεύεται και αξιοποιεί μεγαλύτερο μέρος του συνόλου της πληροφορίας που προέρχεται από τους ψήφους ανομοιότητας επιλέξαμε να υιοθετήσουμε αυτόν, λαμβάνοντας υπόψη ότι αποτελεί την ερμηνεία που έχουν αποδώσει και οι [Wolf et al., 2012]

9.4 Data Partitioning

Προκειμένου να αξιολογήσουμε την επίδοση των ταξινομητών που περιγράφηκαν ανωτέρω, συγκρίναμε δύο παραλλαγές διασταυρούμενης επαλήθευσης (cross validation) για την κατασκευή ανεξάρτητων συνόλων εκπαίδευσης και ελέγχου.

Χρησιμοποιήσαμε την άμεση μέθοδο τυχαίας επιλογής περιορισμών για τη κατασκευή μη επικαλυπτόμενων συνόλων εκπαίδευσης και ελέγχου για 10-fold cross validation, όπως έχει χρησιμοποιηθεί και από τους [Wolf et al., 2011]. Στο υπόλοιπο της εργασίας αυτός ο τρόπος χωρισμού δεδομένων θα αναφέρεται ως **Sampling A.**

Για τον δεύτερο τρόπο, θεωρήσαμε ότι δεδομένου ότι από κάθε ψήφο κατασκευάζονται δύο περιορισμοί (βλέπε 3.I), εκχωρώντας τον έναν από αυτούς τους περιορισμούς στο σύνολο εκπαίδευσης και τον άλλον στο σύνολο ελέγχου, εισάγεται κάποια μεροληψία στο αποτέλεσμα του ταξινομητή, αφού ουσιαστικά και οι δύο περιορισμοί αναφέρονται στην ίδια πληροφορία, όπως στο [Wolf et al., 2012]. Οι 860 ακμές του βεβαρυμένου γράφου ομοιότητας συνδέουν ουσιαστικά 337 συνιστώσες από τρεις κόμβους η κάθε μια, οι οποίες αντιστοιχούν στις τριάδες του πίνακα ανομοιοτήτων. Συνεπώς στην δεύτερη παραλλαγή η επιλογή δεδομένων για έλεγχο γίνεται ανά τριάδα, αντί να γίνεται ανά περιορισμό. Ο τρόπος αυτός κατασκευής του συνόλου επαλήθευσης θα αναφέρεται ως **Sampling B** στη παρουσίαση των πειραμάτων.

Για την αξιολόγηση της γενίκευσης και της γενικότερης επίδοσης των ταξινομητών εκτελούμε τη διαδικασία της εκπαίδευσης και αξιολόγησης για αυξανόμενο μέγεθος του συνόλου εκπαίδευσης, ξεκινώντας με 86 περιορισμούς για το Sampling A και 17 τριάδες για το Sampling B.

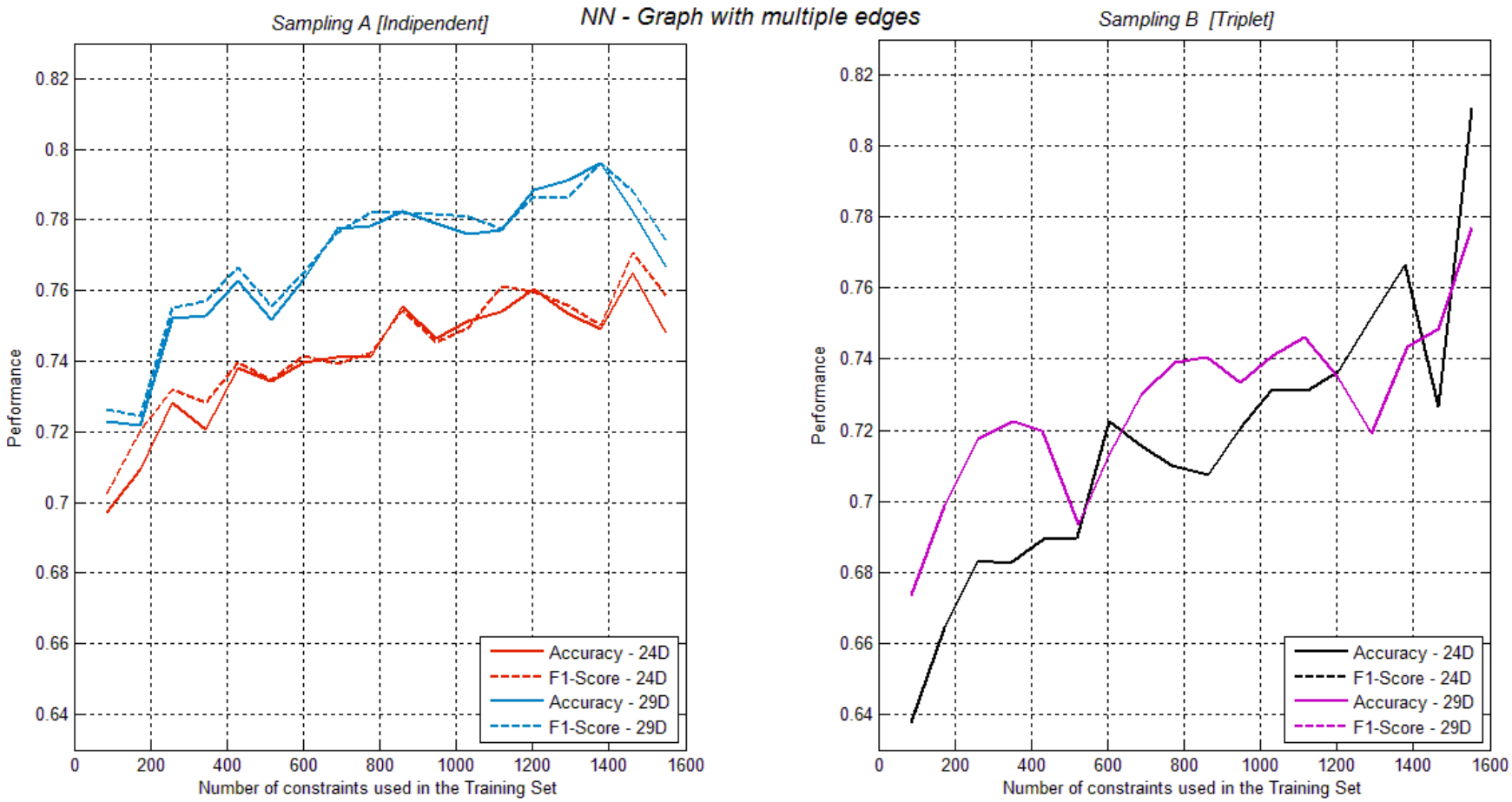
9.5 Ταξινομητές

9.5.1 Neural Network:

Για την επίλυση του προβλήματος με Νευρωνικό Δίκτυο, χρησιμοποιήθηκε το Neural Network toolbox που παρέχεται μαζί με το Matlab.
Κατασκευάστηκε ένα νευρωνικό δίκτυο πρόσθιας τροφοδότησης ενός επιπέδου (Single Layer Feed Forward Neural Network), το οποίο εκπαιδεύτηκε με αλγόριθμο εκπαίδευσης BFGS quasi-Newton backpropagation.

Για να λύσουμε όμως με συνέπεια το πρόβλημα, χρειάστηκε να προσθέσουμε στην συνάρτηση εκπαίδευσης του toolbox τον περιορισμό των μη αρνητικών συναπτικών βαρών προσθέτοντας έναν επιπλέον περιορισμό μηδενισμού των βαρών εκείνων, που μετά την ανανέωσή τους κατά την διαδικασία της εκπαίδευσης, αποκτούν τιμή μικρότερη του μηδενός.

Στα επόμενα διαγράμματα παρουσιάζονται οι επιδόσεις του ταξινομητή με βάση το Accuracy και το F1-Score συναρτήσει του μεγέθους του διανύσματος εκπαίδευσης για τα δύο διανύσματα χαρακτηριστικών και τις δύο παραλλαγές επιλογής του συνόλου ελέγχου. Τα αποτελέσματα που παρουσιάζονται και εδώ αλλά και τις ακόλουθες υλοποιήσεις αποτελούν αποτελέσματα 10-fold cross validation.


NN - Graph with multiple edges
Sampling A [Indipendent]
Sampling B [Triplet]
Performance
Number of constraints used in the Training Set
Accuracy - 24D
F1-Score - 24D
Accuracy - 29D
F1-Score - 29D
0.64
0.66
0.68
0.7
0.72
0.74
0.76
0.78
0.8
0.82
0
200
400
600
800
1000
1200
1400
1600

9.5.2 SVM:

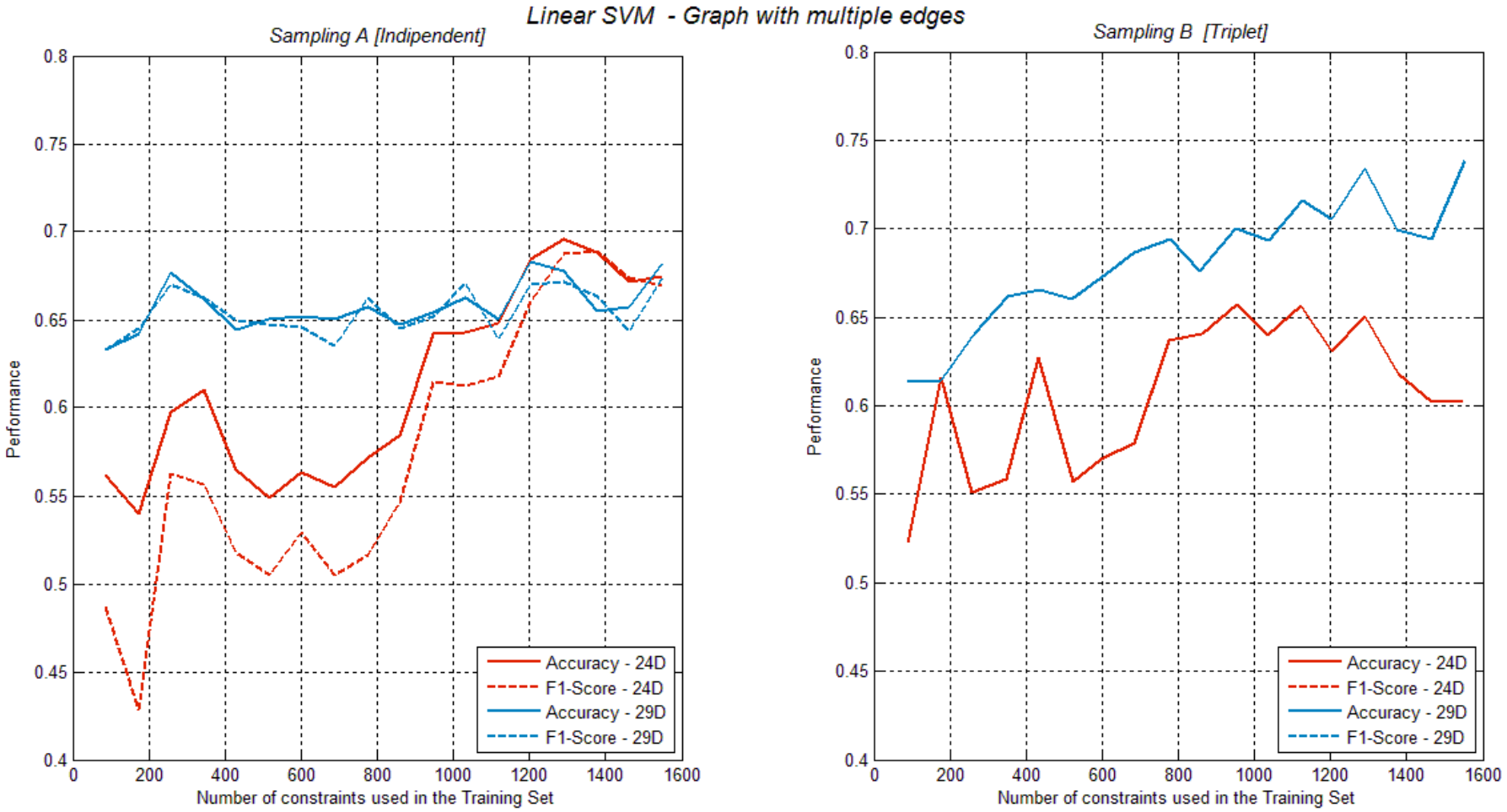
Για την κατασκευή του SVM ταξινομητή χρησιμοποιήθηκε η βιβλιοθήκη συναρτήσεων LIBSVM (Version 3.16 και 3.17) [Chang et al., 2011].

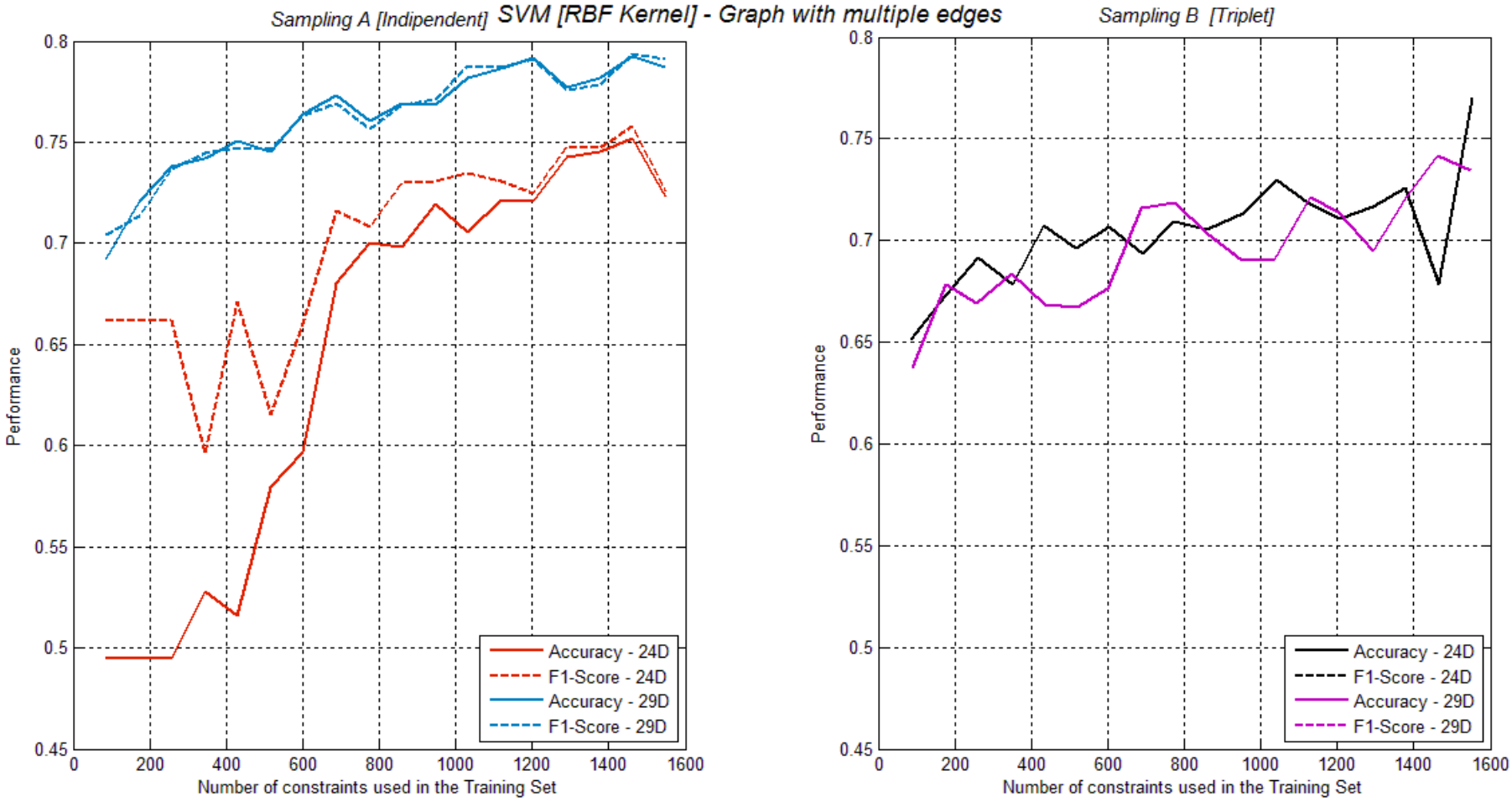
Χρησιμοποιήσαμε Γραμμικό SVM, SVM με πυρήνα RBF και την προσέγγιση του wSVM με RBF πυρήνα. Ο λόγος που δεν χρησιμοποιήθηκε γραμμικό wSVM έγκειται στο γεγονός ότι τα πειράματα με το γραμμικό SVM δεν παρουσίασαν καθόλου καλές επιδόσεις, οπότε κρίναμε ότι δεν θα ήταν σκόπιμο να δοκιμάσουμε και αυτό το μοντέλο.

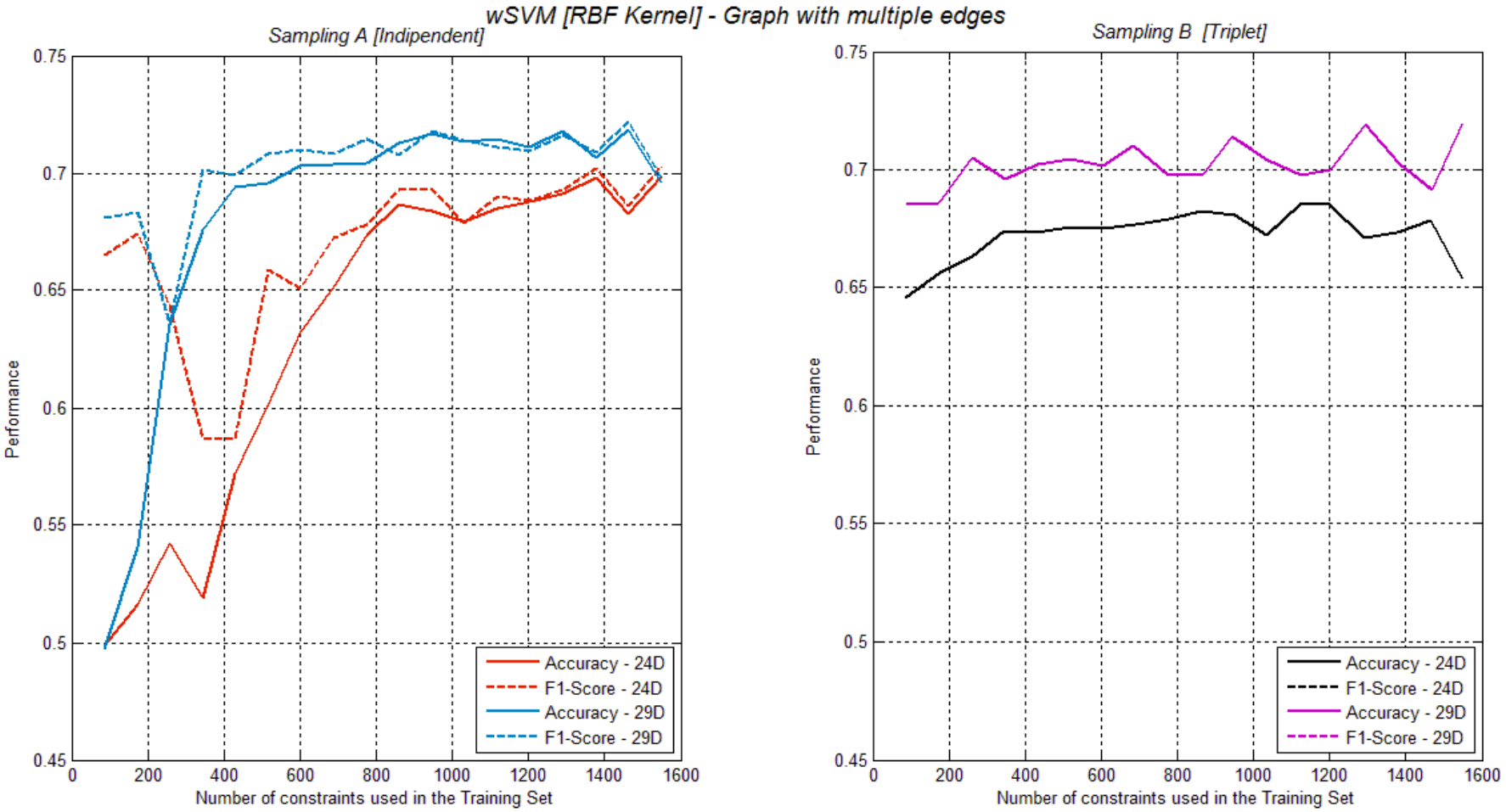
Στις επόμενες σελίδες ακολουθούν τα διαγράμματα για τις τρείς μορφές SVM που χρησιμοποιήθηκαν και για τις δυο παραλλαγές επιλογής συνόλου ελέγχου.

9.5.3 SVNN:

Για τον SVNN ταξινομητή χρησιμοποιήσαμε τον κώδικα που διαθέτει ο Oswaldo Ludwig [Ludwig, 2012] στην file Exchange σελίδα του Matlab:


Linear SVM - Graph with multiple edges
Sampling A [Indipendent]
Sampling B [Triplet]
Performance
Number of constraints used in the Training Set
Accuracy - 24D
F1-Score - 24D
Accuracy - 29D
F1-Score - 29D
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0
200
400
600
800
1000
1200
1400
1600


Sampling A [Indipendent] SVM [RBF Kernel] - Graph with multiple edges
Sampling B [Triplet]
Performance
Number of constraints used in the Training Set
Accuracy - 24D
F1-Score - 24D
Accuracy - 29D
F1-Score - 29D
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0
200
400
600
800
1000
1200
1400
1600

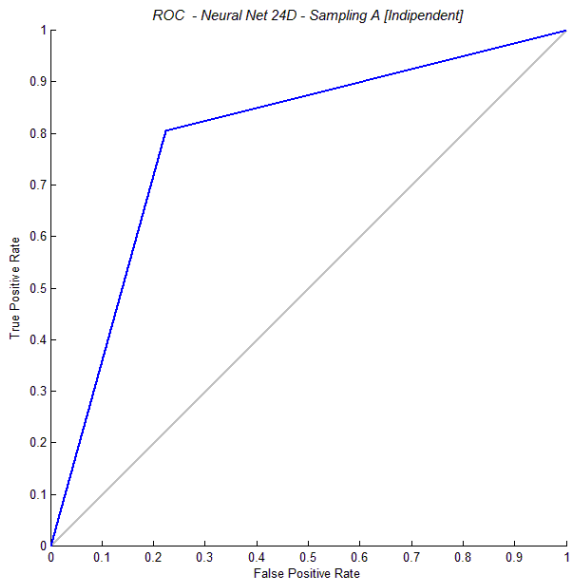

wSVM [RBF Kernel] - Graph with multiple edges
Sampling A [Indipendent]
Sampling B [Triplet]
Performance
Number of constraints used in the Training Set
Accuracy - 24D
F1-Score - 24D
Accuracy - 29D
F1-Score - 29D
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0
200
400
600
800
1000
1200
1400
1600

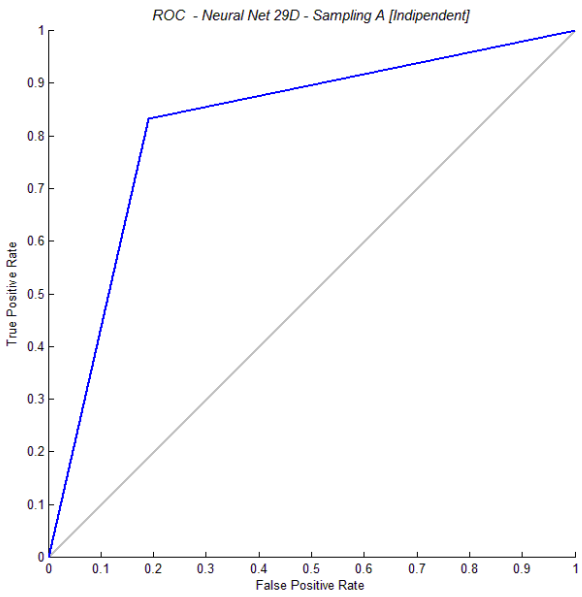

SVNN - Graph with multiple edges
Sampling A [Indipendent]
Sampling B [Triplet]
Performance
Number of constraints used in the Training Set
Accuracy - 24D
F1-Score - 24D
Accuracy - 29D
F1-Score - 29D
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0
200
400
600
800
1000
1200
1400
1600

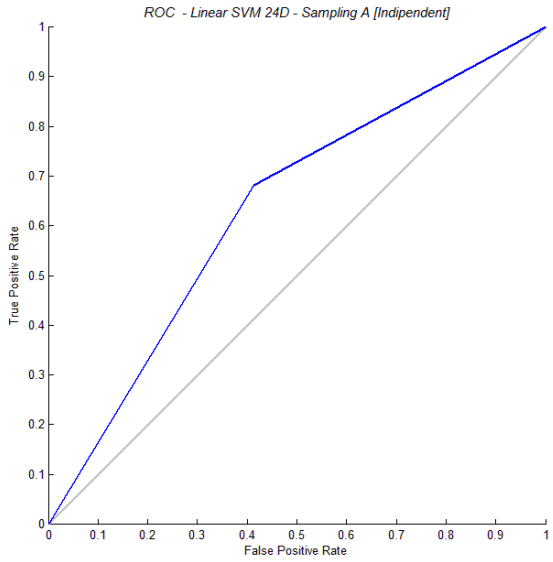
Στη συνέχεια παρουσιάζονται τα αποτελέσματα για μέγεθος συνόλου εκπαίδευσης ίσο με το 80% του συνόλου των δεδομένων (1376) και σύνολο ελέγχου ίσο με το 20% του συνόλου των δεδομένων (344). Το 5-fold cross validation αυτό αποτελεί την πιο συχνά χρησιμοποιούμενη μέθοδο για την εκπαίδευση ενός συστήματος. Τα αποτελέσματα των μέτρων αξιολόγησης των μοντέλων που περιγράφηκαν πιο πριν είναι τα ακόλουθα.

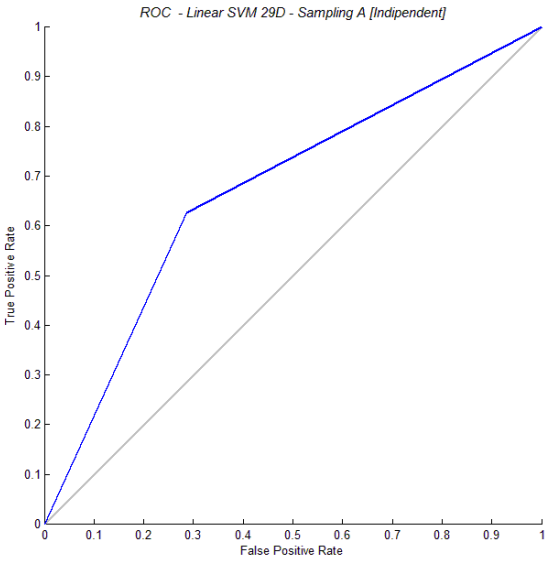
	Accuracy	**Recall**	**Precision**	**F1 - Score**
24D				
NN	0.7922	0.8056	0.7831	0.7994
Linear SVM	0.6339	0.6806	0.6227	0.6503
RBF SVM	0.7668	0.8064	0.7474	0.7757
wSVM	0.7151	0.7193	0.7110	0.7151
SVNN	0.7740	0.8038	0.7849	0.7942
29D				
NN	0.8211	0.8333	0.8136	0.8233
Linear SVM	0.6699	0.6879	0.7484	0.7168
RBF SVM	0.7889	0.7829	0.7921	0.7874
wSVM	0.7267	0.6879	0.7484	0.7168
SVNN	0.8011	0.8082	0.7971	0.8026

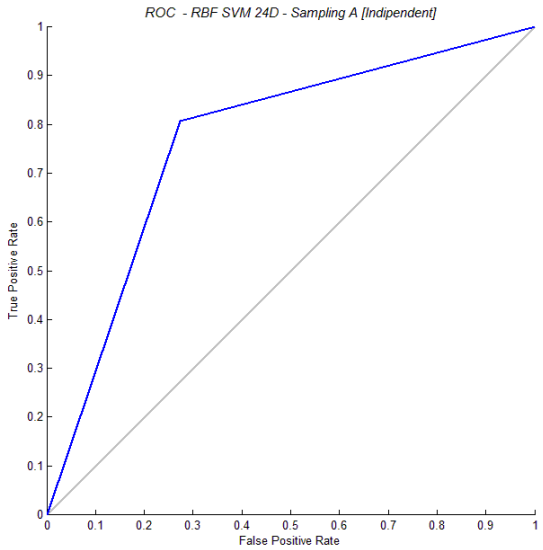
Τα διαγράμματα ROC για τα πιο πάνω συστήματα παρουσιάζονται στις επόμενες σελίδες:

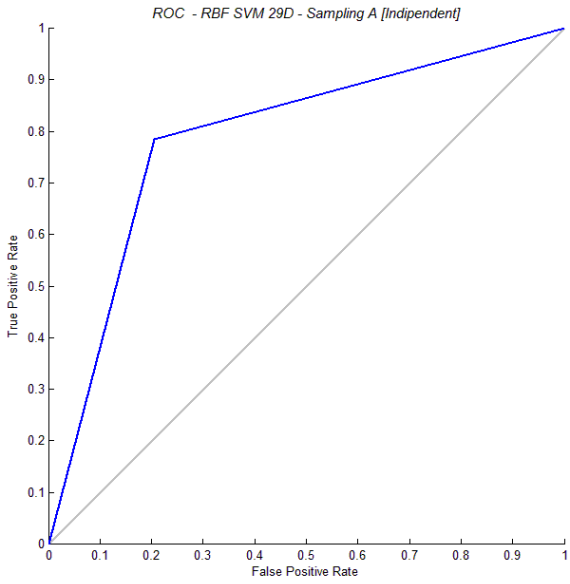


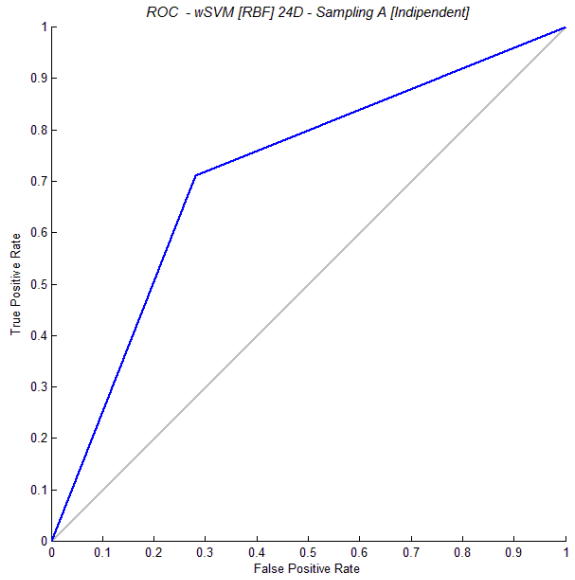


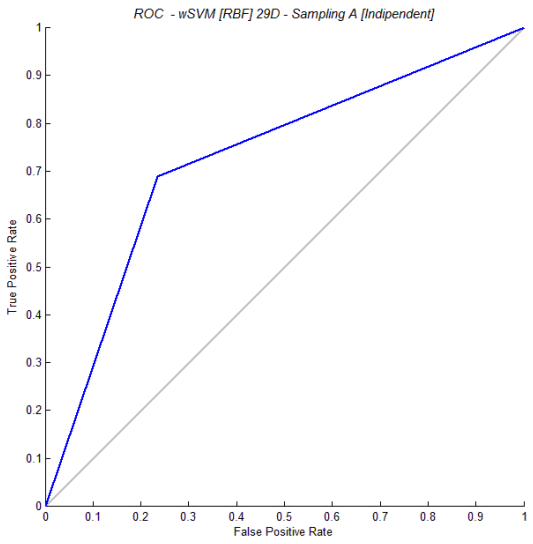

ROC - Linear SVM 24D - Sampling A [Indipendent]
True Positive Rate
False Positive Rate

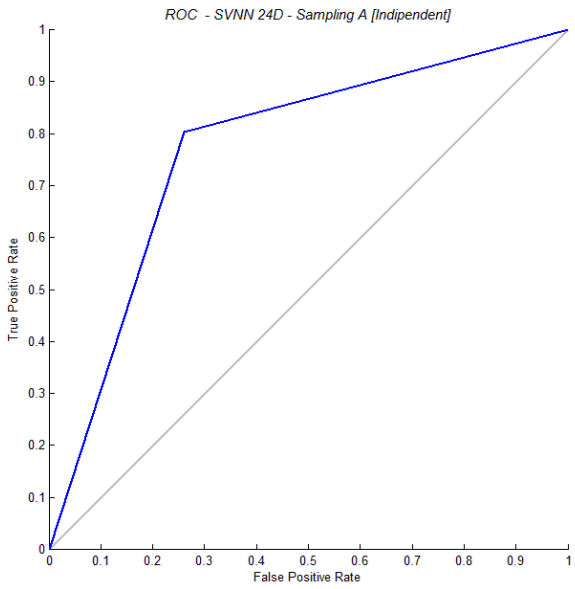

ROC - Linear SVM 29D - Sampling A [Indipendent]
True Positive Rate
False Positive Rate

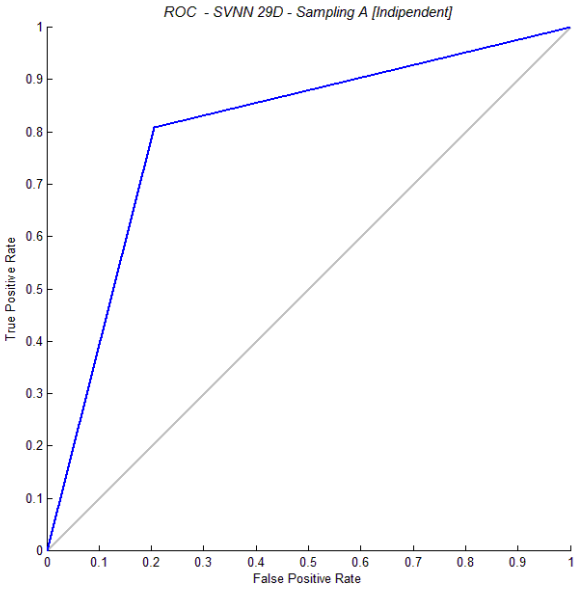

ROC - RBF SVM 24D - Sampling A [Indipendent]
True Positive Rate
False Positive Rate


ROC - RBF SVM 29D - Sampling A [Indipendent]
True Positive Rate
False Positive Rate









9.6 Αξιολόγηση Αποτελεσμάτων

9.6.1 Διανύσματα Χαρακτηριστικών

Αρχικά μπορούμε να παρατηρήσουμε ότι στα περισσότερα από τα μοντέλα που χρησιμοποιήθηκαν για την επίλυση του προβλήματός μας, η χρήση του επαυξημένου διανύσματος χαρακτηριστικών (των 29 διαστάσεων) αποδείχτηκε ότι συνείσφερε στην βελτίωση της απόδοσης των εκάστοτε συστημάτων. Αυτό βέβαια ήταν εν μέρει διαισθητικά αναμενόμενο. Το πρώτο διάνυσμα χαρακτηριστικών που εφαρμόστηκε (τα ENT features),

βασιζόταν μόνο σε πληροφορία ηχοχρώματος και συνεπώς δεν θα ήταν δυνατό να συνοψίσει την έννοια της μουσικής ομοιότητας.

Η προσθήκη των επιπλέον γενικών χαρακτηριστικών έχει συμβάλλει στην αύξηση της απόδοσης των συστημάτων κατά ένα 2 – 3%., το οποίο αν λάβουμε υπόψη μας ότι αφορά την προσθήκη μόλις 5 επιπλέον διαστάσεων στο πρόβλημα μας, μπορεί να θεωρηθεί ικανοποιητικό αποτέλεσμα.

Από τις καμπύλες ROC μπορούμε να παρατηρήσουμε για κάθε σύστημα τη μεταβολή του ρυθμού των Πραγματικά Θετικών και των Λανθασμένα Θετικών παραδειγμάτων. Είναι φανερό ότι η χρήση του διανύσματος χαρακτηριστικών 29 διαστάσεων, μετατόπισε τα σημεία ROC σε θέσεις που αντιστοιχούν σε μικρότερο ρυθμό Λανθασμένων Θετικών (FP) και σε αρκετές περιπτώσεις και σε μεγαλύτερο αριθμό Πραγματικά Θετικών (TP) , δηλαδή οδήγησε στην κατασκευή ταξινομητών με μεγαλύτερη ικανότητα γενίκευσης.

9.6.2 Μοντέλο επιλογής Συνόλων Ελέγχου και Εκπαίδευσης

Όσον αφορά την επιλογή του μοντέλου κατασκευής του συνόλου ελέγχου διαπιστώνουμε ότι η μέθοδος Sampling B, δηλαδή η επιλογή των συνόλων εκπαίδευσης και ελέγχου με βάση τις τριάδες παρουσιάζει λίγο χειρότερη επίδοση από τη μέθοδο Sampling A. Είναι αρκετά εύλογο να συμπεράνουμε ότι η υπόθεση που είχαμε διατυπώσει στην αρχή, περί επανάληψης πληροφορίας που προκύπτει από την ίδια τριάδα κομματιών στο σύνολο εκπαίδευσης και ελέγχου προσθέτει μεροληψία (bias) στο σύστημα. Συνεπώς μπορούμε να υποθέσουμε ότι με τη μέθοδο Sampling A ένα μικρό κομμάτι του συνόλου ελέγχου αφορά πληροφορία βάση της οποίας εκπαιδεύτηκε ο ταξινομητής και ήδη γνωρίζει την σωστή κατηγοριοποίηση κατά το στάδιο του ελέγχου.

Επιπροσθέτως μπορεί να παρατηρηθεί στα προηγούμενα διαγράμματα ότι για τα περισσότερα πειράματα όπου η επιλογή του συνόλου ελέγχου έγινε βάση του Sampling B, η καμπύλη της accuracy συμπίπτει με αυτή του f1-Score. Αυτό έγκειται στο γεγονός ότι η επειδή η εκπαίδευση γίνεται με βάση τις τριάδες, θα απονεμηθούν συγχρόνως στο σύνολο εκπαίδευσης ή ελέγχου και το θετικό και το αρνητικό παράδειγμα που αφορά σε έναν περιορισμό. Συνεπώς αν ένα θετικό παράδειγμα από το σύνολο ελέγχου ταξινομηθεί σωστά, λόγω της φύσης του προβλήματος, θα ταξινομηθεί σωστά και το αντίστοιχο αρνητικό και συνεπώς στην αξιολόγηση θα προκύψει ίσος αριθμός Πραγματικά Θετικών (TP) και Πραγματικά Αρνητικών (TN) παραδειγμάτων. Από τις σχέσεις που έχουν δοθεί σε προηγούμενη ενότητα για το f1-score και το accuracy μπορεί εύκολα να υπολογιστεί ότι για την περίπτωση που TN=TP τα δύο αυτά μέτρα έχουν ίδια τιμή.

9.6.3 Μέγεθος Συνόλου Εκπαίδευσης

Όσον αφορά το κατάλληλο μέγεθος εκπαίδευσης του κάθε ταξινομητή παρατηρούμε ότι ενώ και το Νευρωνικό και το SVM με πυρήνα RBF και το SVNN για το επαυξημένο διάνυσμα χαρακτηριστικών πετυχαίνουν accuracy περίπου 80% για το Sampling A, εντούτοις το

Νευρωνικό Δίκτυο χρειάζεται περισσότερα παραδείγματα εκπαίδευσης για να επιτύχει το αποτέλεσμα αυτό. Παρατηρούμε ότι ενώ το SVNN χρειάζεται 1100 περίπου παραδείγματα εκπαίδευσης για να επιτύχει αυτό το ποσοστό, το SVM χρειάζεται 1200, ενώ το Νευρωνικό 1400 παραδείγματα.

Στον πίνακα με τις επιδόσεις των αλγορίθμων μπορούμε να παρατηρήσουμε ότι το Νευρωνικό Δίκτυο παρουσιάζει καλύτερη απόδοση για αυτό το μέγεθος συνόλου εκπαίδευσης σε σχέση με το SVM με RBF πυρήνα και το SVNN. Αυτό οφείλεται στο γεγονός που παρατηρήθηκε πιο πάνω βάση των διαγραμμάτων ότι τα δύο αυτά συστήματα πετυχαίνουν τη μέγιστη απόδοση τους για μικρότερο σύνολο εκπαίδευσης και κάθε περεταίρω αύξηση του συνόλου εκπαίδευσης είναι ικανή να τα οδηγήσει σε υπερεκπαίδευση και να μειώσει την ικανότητα τους για γενίκευση.

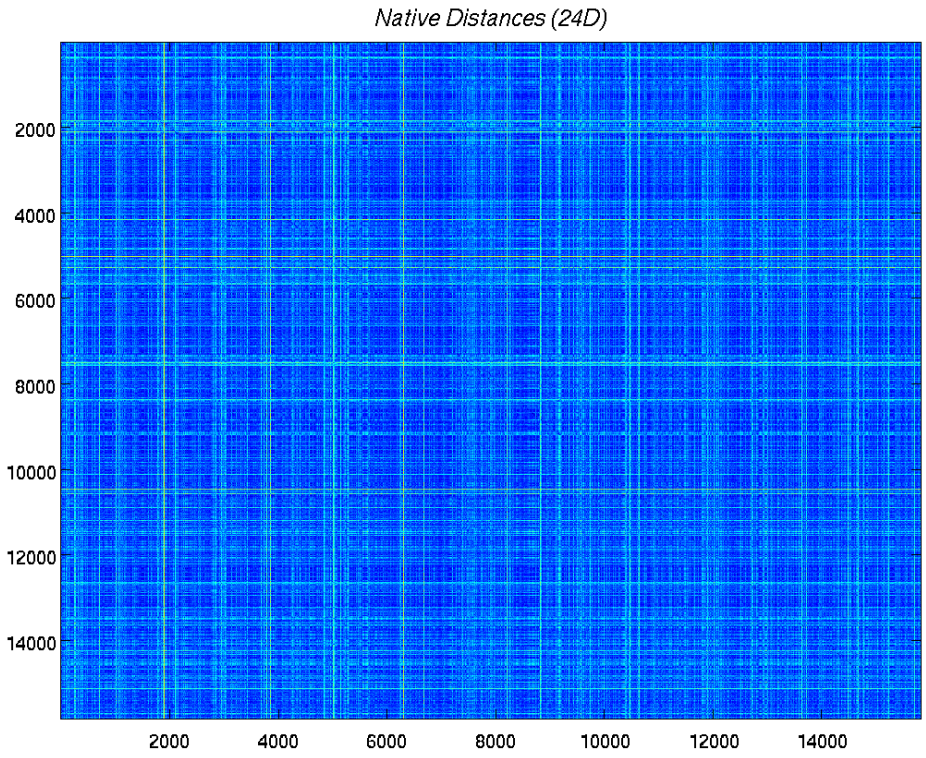
9.7 Visualization

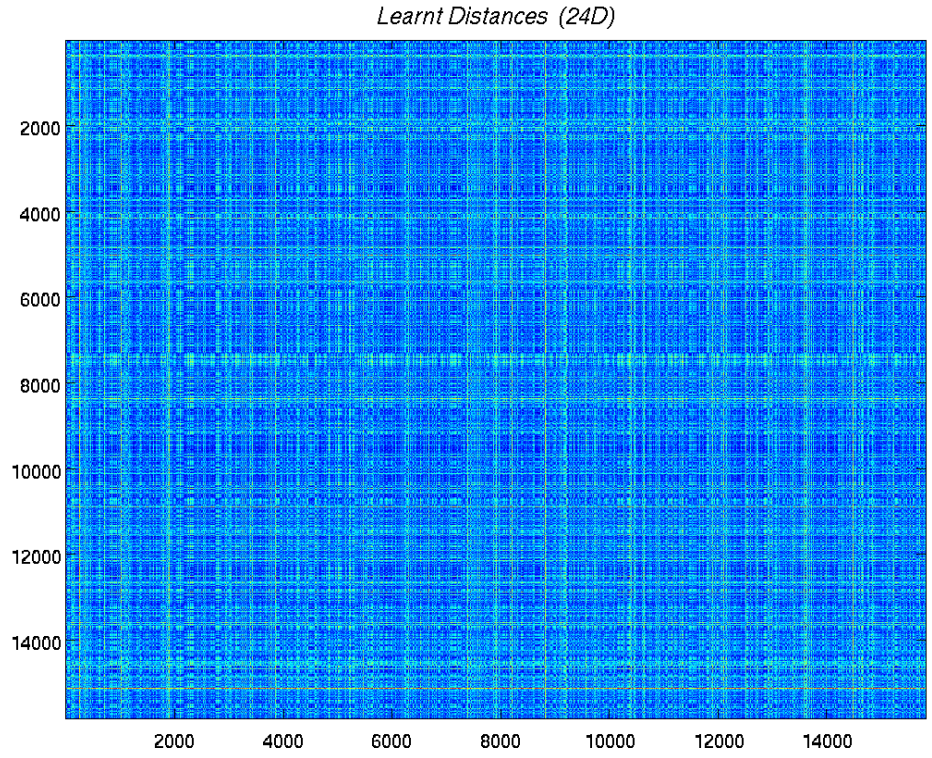
Για τη διαδικασία της οπτικοποίησης των αποστάσεων ομοιότητας/ανομοιότητας , χρειαζόμαστε το διάνυσμα των βαρών που προκύπτει για κάθε μοντέλο ταξινομητή, το οποίο θα χρησιμοποιηθεί στην εξίσωση (3.4).
Στην περίπτωση του Νευρωνικού Δικτύου ενός επιπέδου, τα βάρη διατίθενται από το ίδιο το toolbox του Matlab ως μέρος του μοντέλου του ταξινομητή που εκπαιδεύσαμε.
Για τον SVM ταξινομητή τα βάρη υπολογίζονται μέσω των σχέσεων (5.20) και (5.37) για το γραμμικό και το μοντέλο με RBF πυρήνα αντίστοιχα. Από τη σχέση (5.37) παρατηρούμε ότι ο υπολογισμός του διανύσματος βαρών δεν μπορεί να πραγματοποιηθεί επειδή δεν γνωρίζουμε την μορφή της απεικονιστικής συνάρτησης φ(). Συνεπώς αποκλείεται το ενδεχόμενο κατασκευής οπτικοποίησης των αποτελεσμάτων βασισμένης στο RBF SVM και στο wSVM.

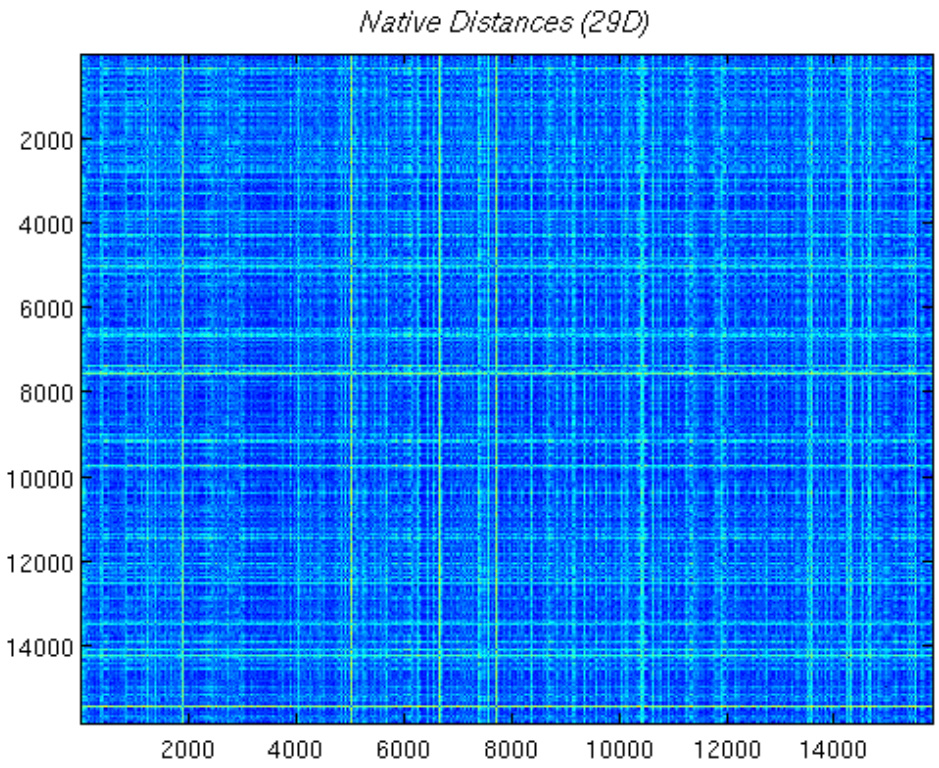
Όσον αφορά το γραμμικό μοντέλο θεωρήσαμε ότι λόγω της χαμηλής επίδοσής του θα ήταν άσκοπο να βασίσουμε την οπτικοποίησης πάνω στο διάνυσμα βαρών που θα προέκυπτε από αυτό, δεδομένου ότι το αποτέλεσμα που θα προέκυπτε θα λάνθανε αξιοπιστίας.

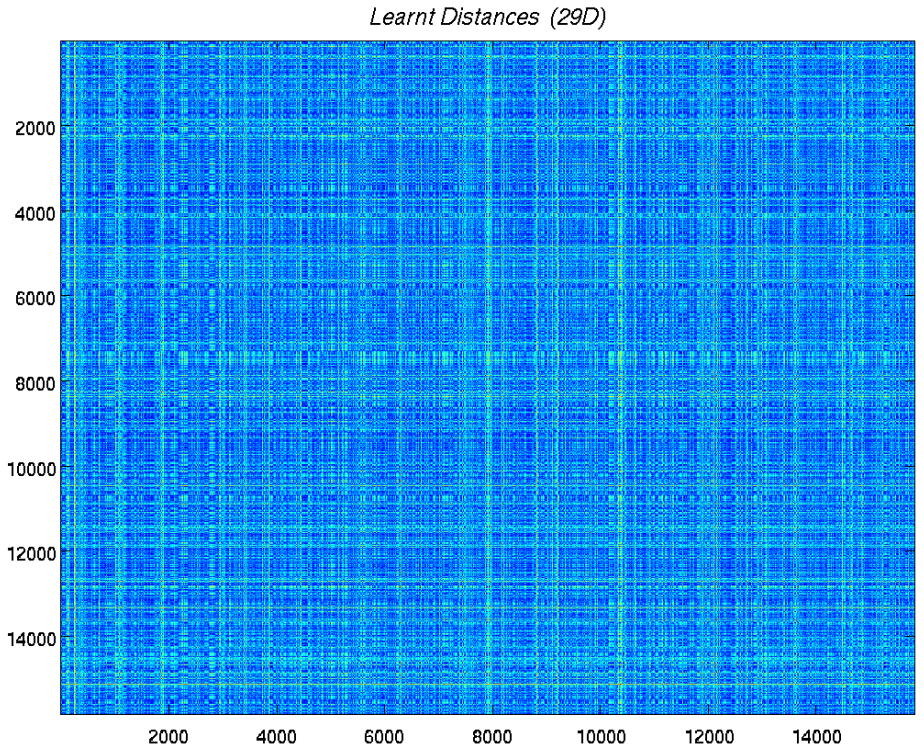
Συνεπώς χρησιμοποιήσαμε τα βάρη που προέκυψαν από το Νευρωνικό Δίκτυο με μέγεθος συνόλου εκπαίδευσης 80% του συνολικού συνόλου δεδομένων (όπου όπως αναλύθηκε προηγουμένως είναι το σημείο όπου το Νευρωνικό επιτυγχάνει τη μεγαλύτερη επίδοση του) και υπολογίσαμε τις αποστάσεις μεταξύ των κομματιών όλης της βάσης του Magnatagatune.

Αρχικά θεωρούμε ότι θα ήταν χρήσιμο να παρουσιάσουμε τις τιμές των αποστάσεων των κομματιών όπως αυτές υπολογίζονται βάση της Ευκλείδειας απόστασης μεταξύ των χαρακτηριστικών που περιγράφουν τα κομμάτια (native distance) και όπως προκύπτουν ύστερα από την χρήση των συναπτικών βαρών του Νευρωνικού πάνω στις Ευκλείδειες αποστάσεις των κομματιών σύμφωνα με την εξίσωση (3.4).


Native Distances (24D)
2000
4000
6000
8000
10000
12000
14000
2000
4000
6000
8000
10000
12000
14000


Learnt Distances (24D)
2000
4000
6000
8000
10000
12000
14000
2000
4000
6000
8000
10000
12000
14000


Native Distances (29D)
2000
4000
6000
8000
10000
12000
14000
2000
4000
6000
8000
10000
12000
14000


Learnt Distances (29D)
2000
4000
6000
8000
10000
12000
14000
2000
4000
6000
8000
10000
12000
14000

Στις προηγούμενες εικόνες απεικονίζονται οι τιμές των αποστάσεων μεταξύ των διαφορετικών κομματιών της βάσης. Με σκούρο μπλε αναπαριστώνται οι τιμές που είναι μηδενικές ή βρίσκονται πολύ κοντά στο 0 και καθώς αυξάνει η τιμή της απόστασης το χρώμα της αναπαράστασης τείνει προς το ανοιχτό γαλάζιο και το κίτρινο.

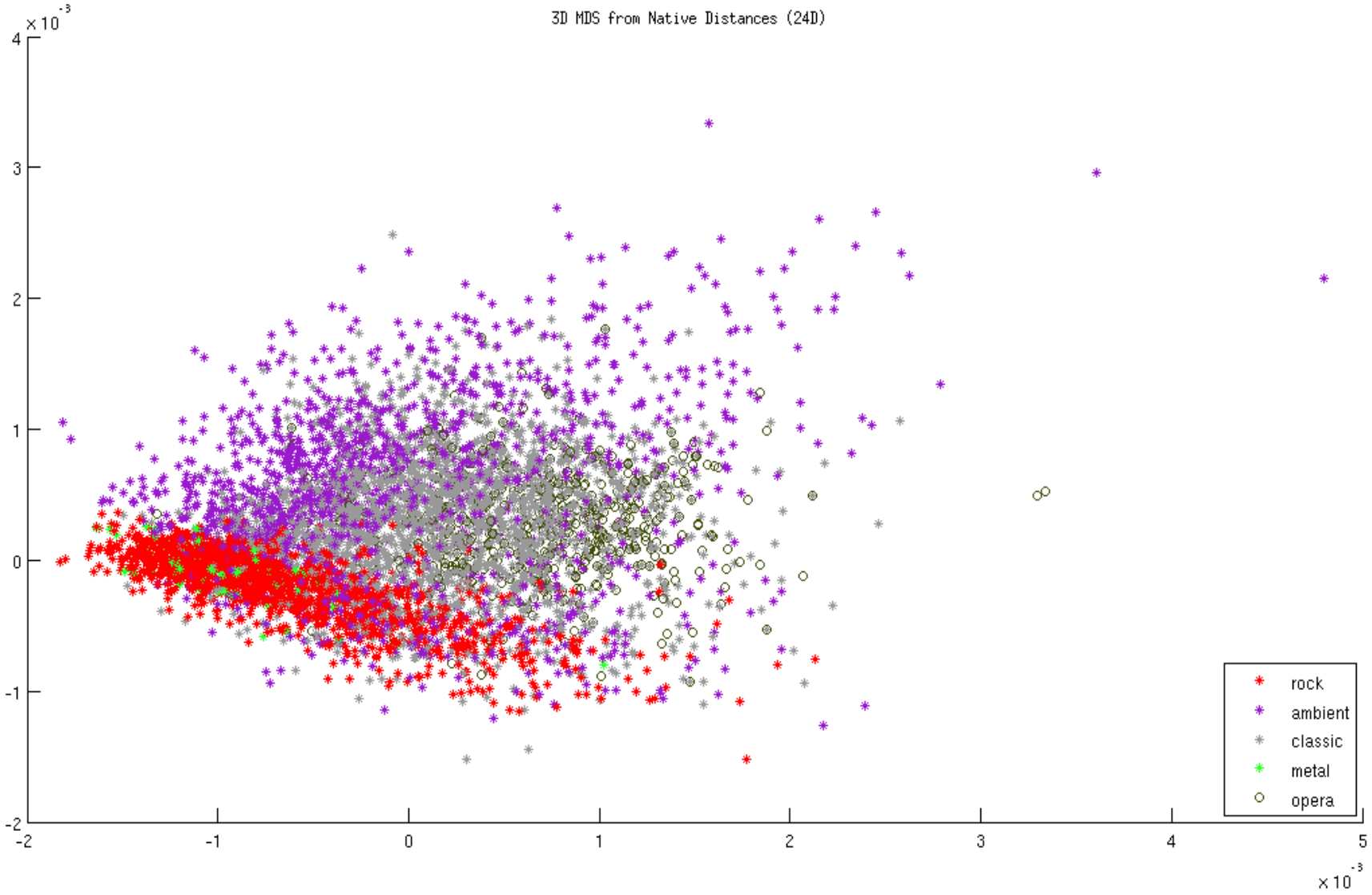
Στην πρώτη εικόνα, όπου παρουσιάζονται οι native/έμφυτες αποστάσεις των κομματιών υπολογισμένες με χρήση του διανύσματος των 24 διαστάσεων, παρατηρούμε ότι υπάρχει μεγάλη ομοιογένεια μεταξύ των τιμών με τις περισσότερες να βρίσκονται πολύ κοντά στο 0 (σκούρο μπλε). Αντιθέτως στη δεύτερη εικόνα, όπου αποτυπώνονται οι αποστάσεις που προκύπτουν ύστερα από τον πολλαπλασιασμό με τον πίνακα των συναπτικών βαρών παρατηρούμε ότι οι τιμές έχουν διαφοροποιηθεί αρκετά.

Παρατηρώντας την τρίτη εικόνα, όπου αναπαρίστανται οι τιμές των native αποστάσεων των κομματιών υπολογισμένες βάση του διανύσματος των 29 διαστάσεων, διαπιστώνουμε ότι εμφανίζονται μεγαλύτερες διακυμάνσεις στις τιμές σε σχέση με την αντίστοιχη απεικόνιση βάση του διανύσματος των 24 διαστάσεων. Συμπεραίνουμε συνεπώς ότι ακόμα και χωρίς την χρήση του ταξινομητή, το επαυξημένο διάνυσμα χαρακτηριστικών προσθέτει σημαντική πληροφορία στο πρόβλημά μας.

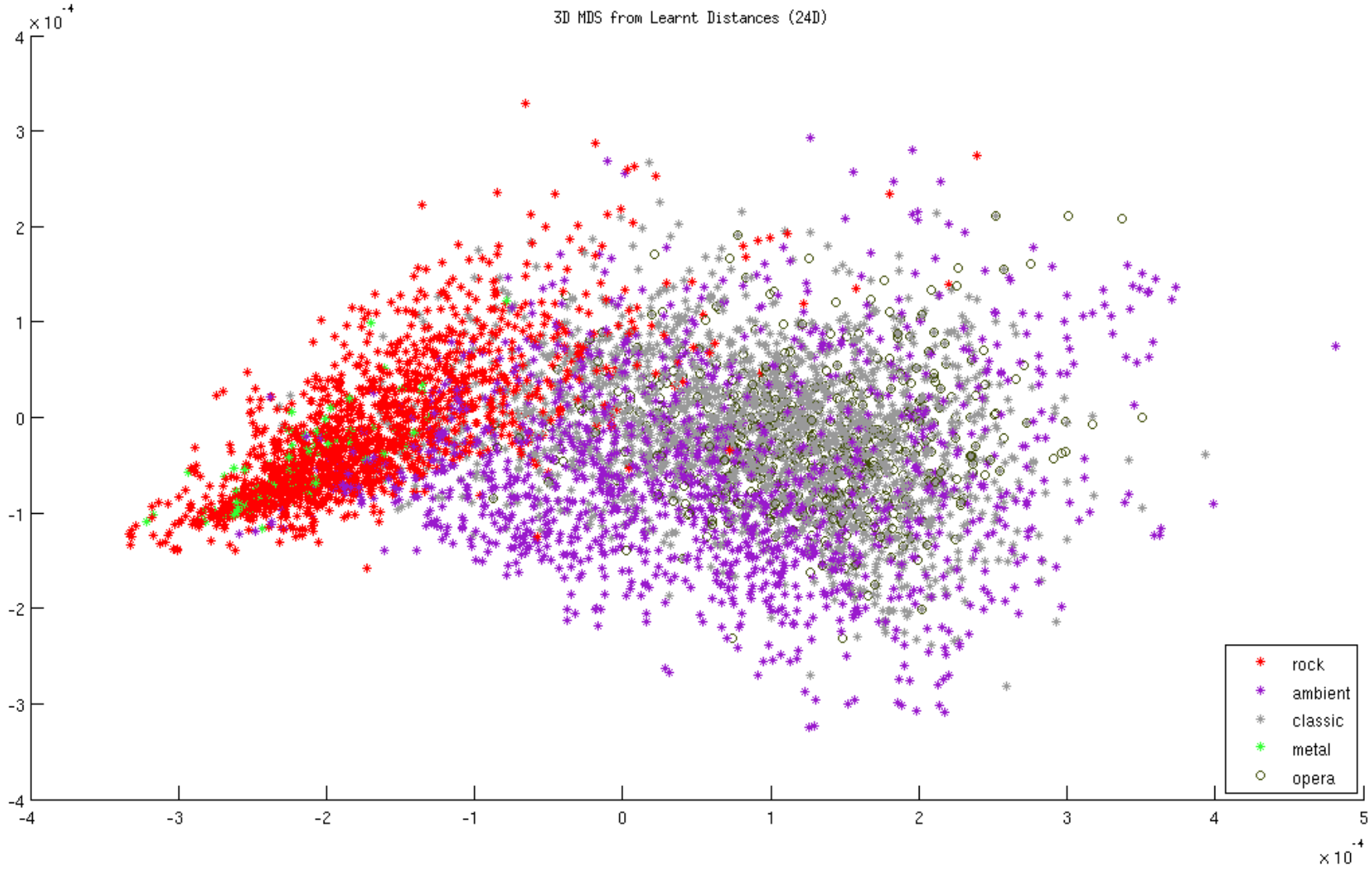
Τέλος παρατηρούμε ότι η τέταρτη εικόνα, στην οποία απεικονίζονται οι τιμές των αποστάσεων βάση του διανύσματος των 29 διαστάσεων ύστερα από τη χρήση του διανύσματος βαρών, ομοιάζει αρκετά με την δεύτερη, με διάφορες τροποποιήσεις βέβαια σε ορισμένα σημεία. Αν ανατρέξουμε στα αποτελέσματα των επιδόσεων που παρουσιάστηκαν ανωτέρω, διαπιστώνουμε ότι το accuracy του Νευρωνικού με χρήση του διανύσματος 29 διαστάσεων είναι κατά περίπου 3% καλύτερη σε σχέση με τον αντίστοιχο ταξινομητή εκπαιδευμένο με διάνυσμα 24 διαστάσεων. Συνεπώς είναι αναμενόμενο οι αποστάσεις που προκύπτουν ύστερα από την χρήση των συναπτικών βαρών των δύο αυτών ταξινομητών να ομοιάζουν αρκετά.

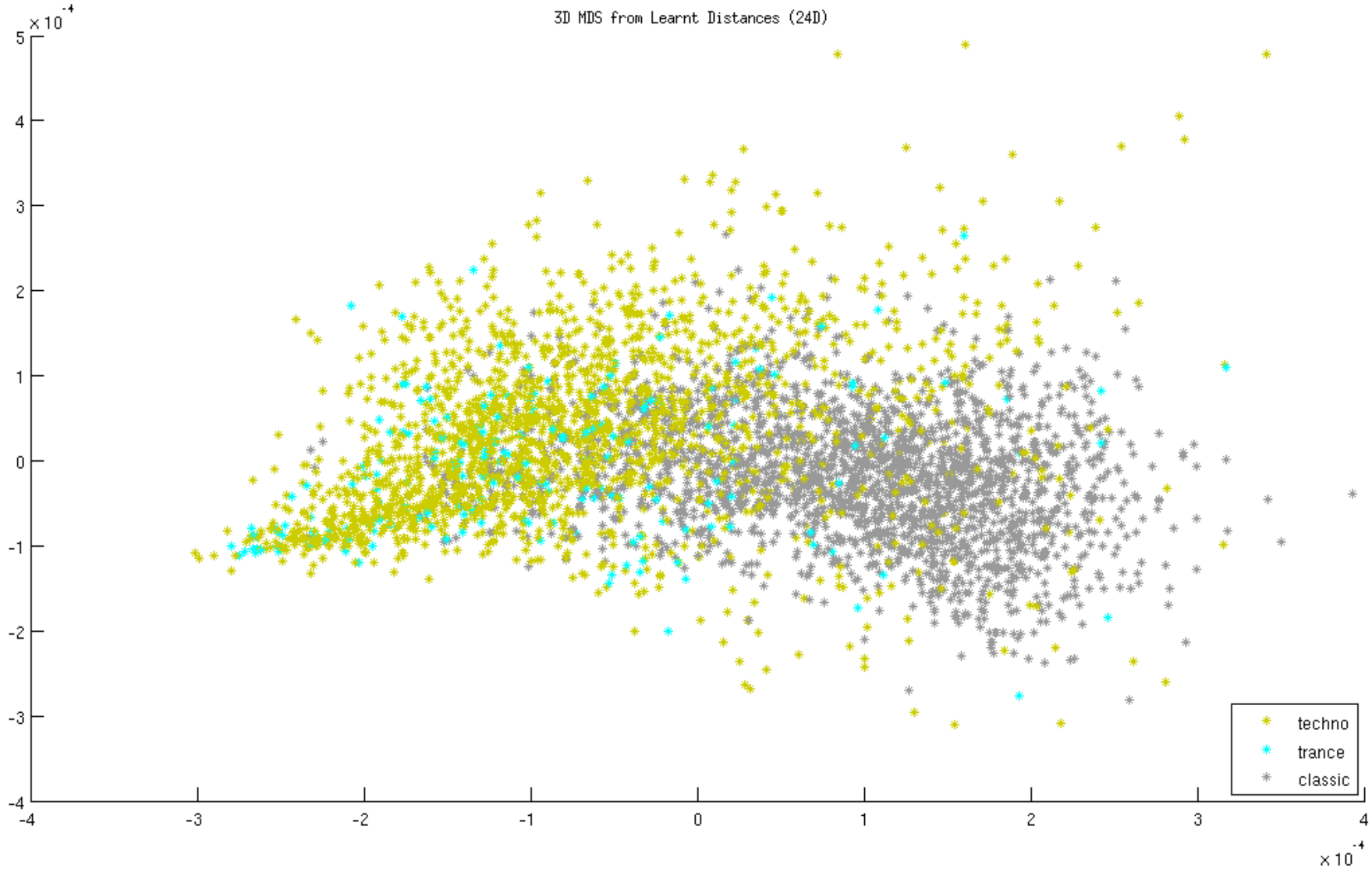
9.7.1 MDS Αναπαράσταση

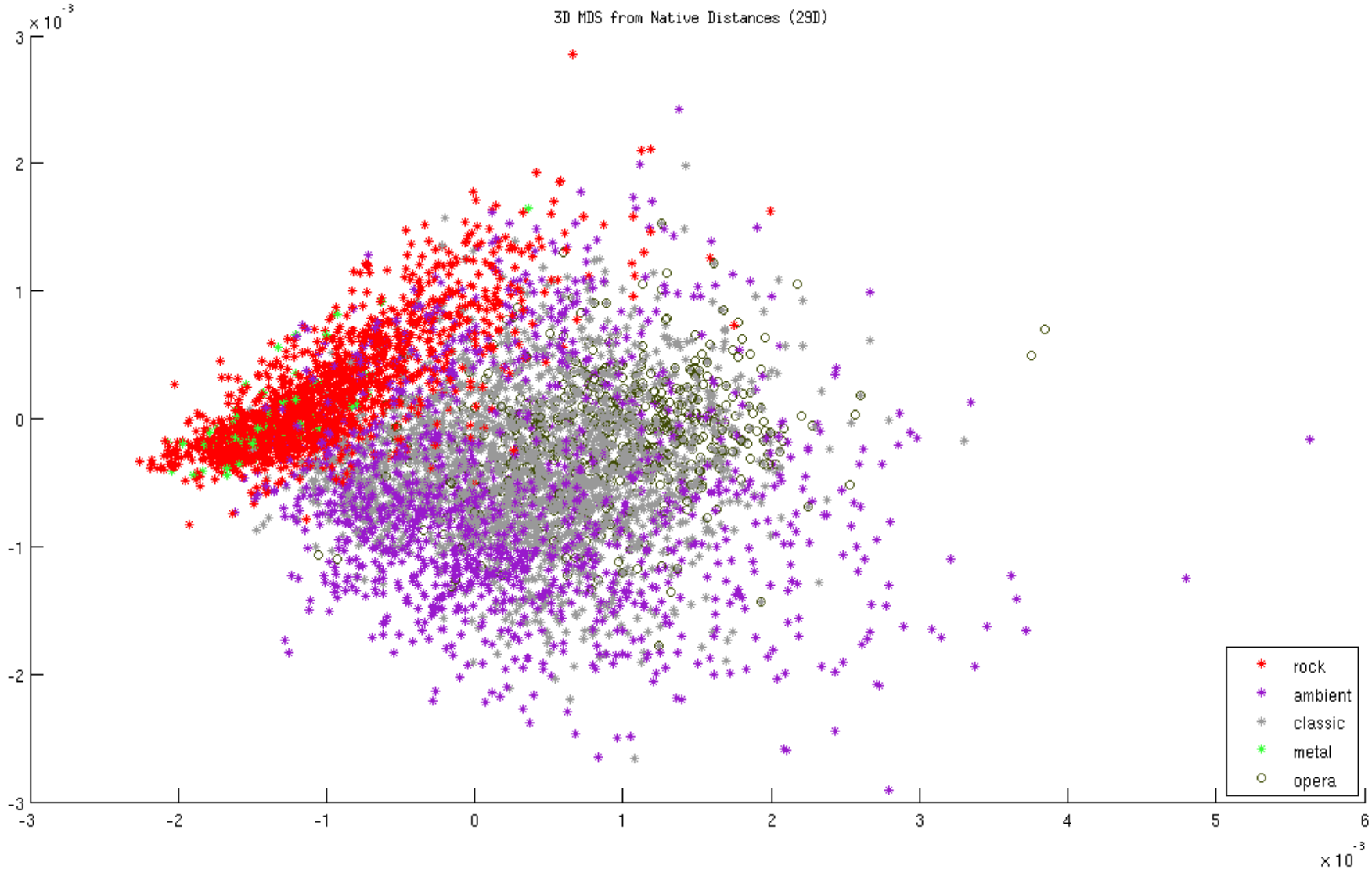
Χρησιμοποιώντας τον μετρικό αλγόριθμο MDS όπως αυτός περιγράφηκε σε προηγούμενη ενότητα, επιχειρήσαμε να απεικονίσουμε τα μουσικά κομμάτια σε έναν 3-διάστατο χώρο βασιζόμενοι στην μουσική ομοιότητα που υπολογίστηκε από τον ταξινομητή Νευρωνικού Δικτύου. Συγχρόνως κάνοντας χρήση των tags που παρέχονται με τη βάση του Magnatagatune και κυρίως των tags που αφορούν σε μουσικό είδος, επιχειρήσαμε να εξετάσουμε κατά πόσο η μουσική ομοιότητα που έχει υπολογιστεί με το σύστημά μας σχετίζεται με το μουσικό είδος στο οποίο ανήκει ένα μουσικό κομμάτι. Τα αποτελέσματα παρουσιάζονται στη συνέχεια:


3D MDS from Native Distances (24D)
x 10^-3
4
3
2
1
0
-1
-2
-2
-1
0
1
2
3
4
5
x 10^-3
rock
ambient
classic
metal
opera

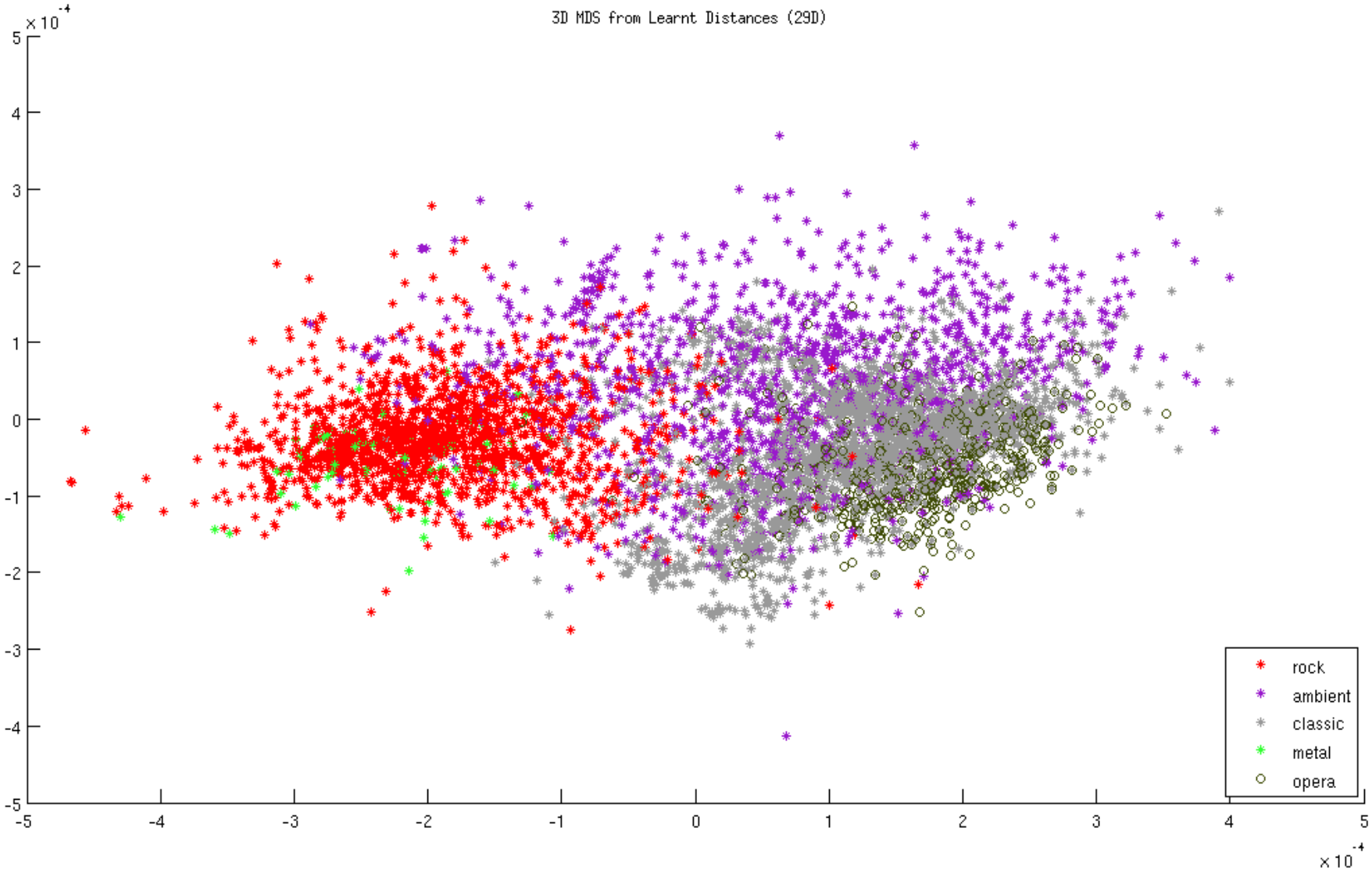

3D MDS from Native Distances (24D)
x 10^-3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
x 10^-3
techno
trance
classic

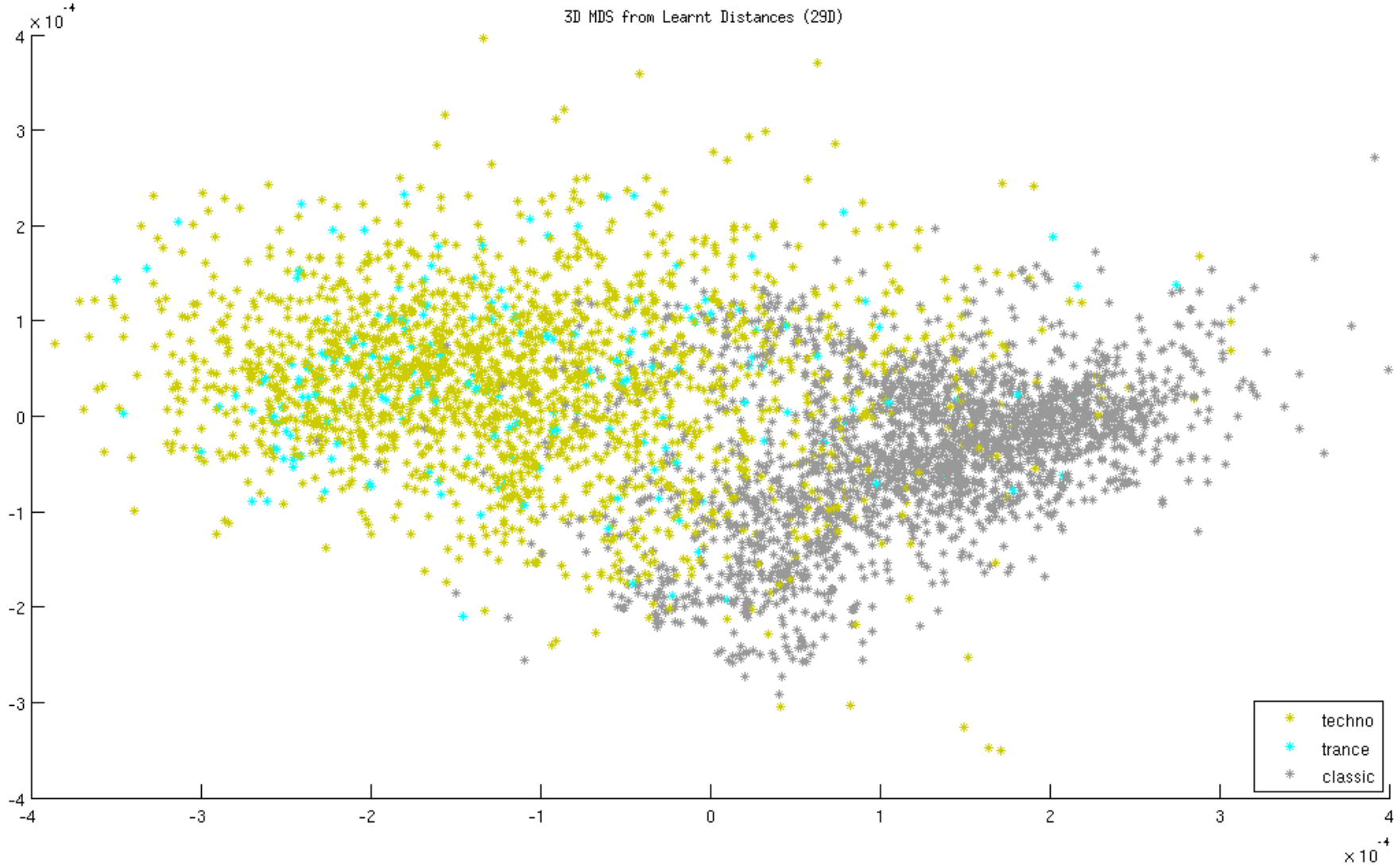

3D MDS from Learnt Distances (24D)
× 10^-4
4
3
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
5
× 10^-4
rock
ambient
classic
metal
opera


3D MDS from Learnt Distances (24D)
× 10^-4
5
4
3
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
× 10^-4
techno
trance
classic


3D MDS from Native Distances (29D)
x 10^-3
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
4
5
6
x 10^-3
rock
ambient
classic
metal
opera


3D MDS from Native Distances (29D)
× 10^-3
4
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
4
× 10^-3
techno
trance
classic


3D MDS from Learnt Distances (29D)
× 10^-4
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5
× 10^-4
rock
ambient
classic
metal
opera


3D MDS from Learnt Distances (29D)
× 10^-4
4
3
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
× 10^-4
techno
trance
classic

Στις προηγούμενες εικόνες απεικονίζεται μόνο ένα μέρος των κομματιών της βάσης, το οποίο αξιολογείται ως αρκετά ικανοποιητικό με βάση την θέση που αποδοθεί σε κάθε μουσικό είδος, αν υποθέσουμε ότι η κατηγοριοποίηση κατά είδος αποτελεί μια βασική παράμετρο της μουσικής ομοιότητας.

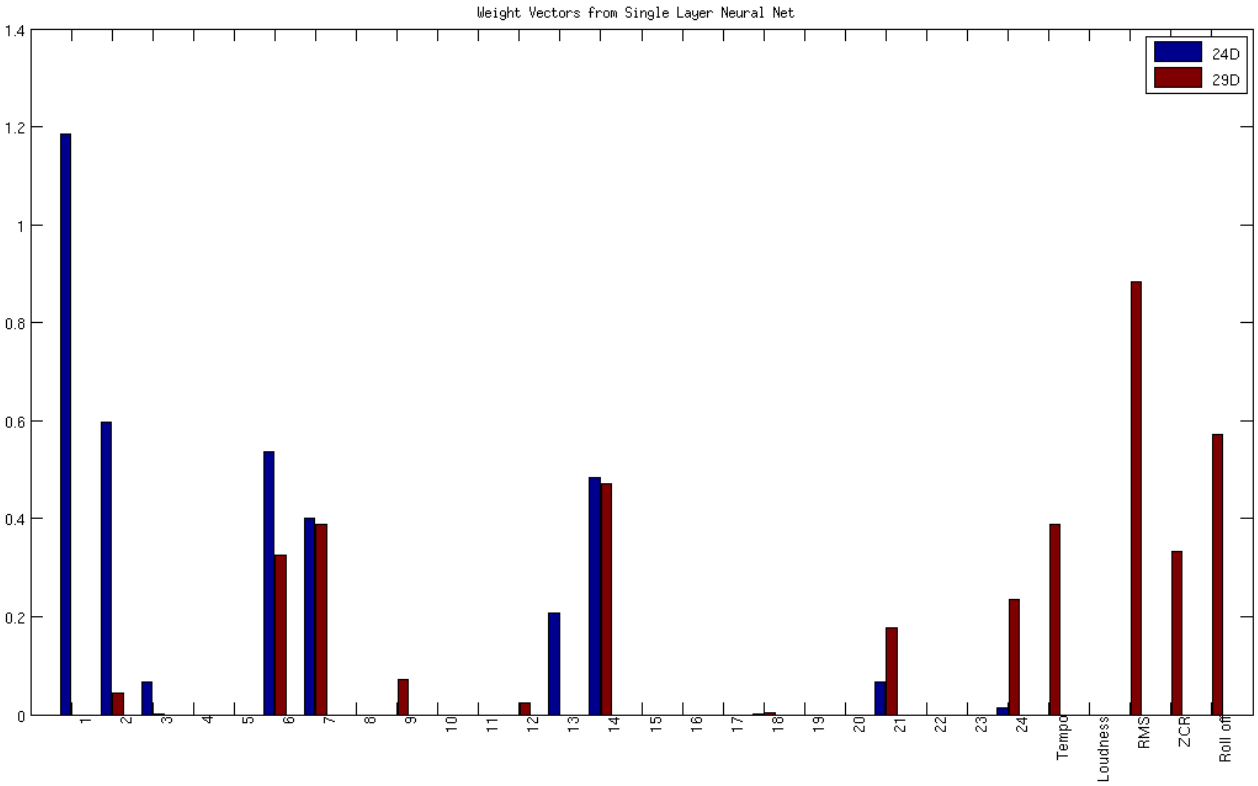
Αρχικά θα πρέπει να αναφέρουμε ότι είναι εμφανές ότι τα χαρακτηριστικά ENT αποτελούν μια αρκετά καλή επιλογή αναπαράστασης των μουσικών κομματιών, δεδομένης της πρώτης και της δεύτερης απεικόνισης που προκύπτει οι οποίες έχουν πραγματοποιηθεί με την χρήση των native αποστάσεων για διάνυσμα 24 διαστάσεων. Παρατηρούμε ότι τα μουσικά είδη που απεικονίζονται, αν και δεν εμφανίζονται τελείως ξένα μεταξύ τους, τείνουν προς τον σχηματισμό ενός cluster ανά είδος. Βέβαια παρατηρούμε ότι οι αποστάσεις μεταξύ των διαφορετικών ειδών που απεικονίζονται είναι πάρα πολύ μικρές, γεγονός που πιθανώς δεν συνάδει με την αναπαράσταση της μουσικής ομοιότητας που θέλουμε να κατασκευάσουμε. Παρατηρούμε δηλαδή ότι τα κλασσικά κομμάτια εμφανίζονται πάρα πολύ κοντά στα rock κάτι που σύμφωνα με την ανθρώπινη διαίσθηση δεν ανταποκρίνεται πιστά στο κριτήριο της ομοιότητας.

Στην απεικόνιση για τις αποστάσεις που έχουν προκύψει μέσω του Νευρωνικού Δικτύου για το διάνυσμα των 24 διαστάσεων παρατηρείται ένας διαχωρισμός μεταξύ των κλασσικών και των rock κομματιών, ο οποίος γίνεται αρκετά πιο εμφανής στην αντίστοιχη απεικόνιση για το διάνυσμα των 29 διαστάσεων.

Σε όλες τις απεικονίσεις, αλλά κυρίως στην τελευταία, είναι εμφανής η σχέση των rock – metal και classical – opera κομματιών , αφού τα metal εμφανίζονται να καταλαμβάνουν μια γωνία από τα rock και αντίστοιχα τα οπερετικά για τα κλασσικά κομμάτια, καθώς και ο διαχωρισμός των rock/metal και ambient/classical. Στην αντίστοιχη απεικόνιση όπου παρουσιάζεται η σχέση μεταξύ των techno/trance και classical, παρατηρούμε επίσης τον διαχωρισμό της trance από την κλασσική μουσική.

9.8 Ερμηνεία και Ανάλυση Πίνακα Συναπτικών Βαρών

Αρκετό ενδιαφέρον παρουσιάζει όμως και ο ίδιος ο πίνακας των συναπτικών βαρών που έχει προκύψει από την εκπαίδευση του Νευρωνικού, καθώς μπορεί να αναδείξει ποια χαρακτηριστικά καθορίζουν το αποτέλεσμα της ομοιότητας/ ανομοιότητας δύο μουσικών κομματιών. Στο επόμενο ιστόγραμμα παρουσιάζονται οι τιμές που έχουν λάβει τα βάρη ανά διάσταση κατά την εκπαίδευση με τα διανύσματα των 24 και 29 διαστάσεων.



Παρατηρούμε ότι ενώ στο πείραμα με τις 24 διαστάσεις, το πρώτο ENT χαρακτηριστικό, το οποίο όπως έχει αναφερθεί προηγουμένως, σχετίζεται με την συνολική ένταση του κομματιού, παίζει πολύ καθοριστικό ρόλο, καθώς το βάρος που αντιστοιχεί σε αυτό κατέχει την υψηλότερη τιμή, εντούτοις όταν χρησιμοποιούμε το διάνυσμα των 29 διαστάσεων, όχι μόνο το βάρος που αντιστοιχεί στο χαρακτηριστικό αυτό μηδενίζεται, αλλά φαίνεται πλέον να μην επηρεάζει σημαντικά η ένταση το αποτέλεσμα, αφού και το συναπτικό βάρος που αντιστοιχεί στο χαρακτηριστικό Loudness έχει επίσης μηδενική τιμή. Συνεπώς κάποιο από τα χαρακτηριστικά που προσθέσαμε εμπεριέχει την έννοια της έντασης. Πράγματι το χαρακτηριστικό αυτό είναι η ενέργεια βραχέως χρόνου (RMS), αν και δεν μπορούν να θεωρηθούν ταυτόσημα σαν έννοιες τα δύο αυτά χαρακτηριστικά. Ήχοι με την ίδια RMS συχνότητα μπορεί να μην γίνονται αντιληπτοί σαν ήχοι ίσης έντασης, λόγω του διαφορετικού συχνοτικού περιεχομένου τους. Συγκεκριμένα το ανθρώπινο αυτί αντιλαμβάνεται τους πολύ χαμηλόσυχνους και τους πολύ υψηλόσυχνους ήχους με χαμηλότερη ένταση από αυτήν που αντιλαμβάνεται έναν ήχο ίδιας RMS που ανήκει στις μεσαίες συχνότητες. Συνεπώς τα δύο αυτά χαρακτηριστικά δεν θεωρούνται ταυτόσημα, και για αυτό το λόγο άλλωστε χρησιμοποιήθηκαν στο διάνυσμα των 29 διαστάσεων. Μπορούμε να συμπεράνουμε όμως ότι για το πείραμα των 24 διαστάσεων, ελλείψει περιγραφής της ενέργειας RMS, το ENT χαρακτηριστικό που αντιπροσώπευε την ένταση αναδείχθηκε ως το πιο σημαντικό.

Επιπλέον παρατηρούμε ότι η δέκατη τέταρτη διάσταση του διανύσματος, η οποία αφορά την τυπική απόκλιση του πρώτου ENT feature παίζει σημαντικό ρόλο στην απόφαση και των δύο εκδοχών του ταξινομητή.

Ακόμα είναι φανερό ότι τα επιπλέον 5 χαρακτηριστικά που προσθέσαμε στο πείραμά των 29 διαστάσεων (εκτός από το Loudness) φέρουν καίριο ρόλο στον καθορισμό της μουσικής ομοιότητας, καθώς τα συναπτικά βάρη που αντιστοιχούν σε αυτά διαθέτουν τις υψηλότερες τιμές.

ΚΕΦΑΛΑΙΟ 10:

10.1 Σύνοψη και Συμπεράσματα

Στην εργασία αυτή εξετάστηκε η έννοια της μουσικής ομοιότητας, η οποία αποτελεί μια πολυσύνθετη και ασθενώς ορισμένη έννοια. Έγινε μια εισαγωγή στο αντικείμενο του κλάδου έρευνας Εξόρυξης Μουσικής Πληροφορίας (MIR) και αναλύθηκαν τα είδη πληροφορίας που χρησιμοποιούνται για την υλοποίηση των εφαρμογών MIR, δίνοντας ιδιαίτερη έμφαση στην πληροφορία που προέρχεται από το ηχητικό σήμα.

Στη συνέχεια περιγράφηκαν οι τρόποι με τους οποίους μπορεί να συλλεχθούν πληροφορίες που σχετίζονται με την ανθρώπινη αντίληψη της ομοιότητας, μέσω αντιληπτικών πειραμάτων, και αναλύθηκε ο τρόπος με τον οποίο μπορεί να πραγματοποιηθεί η αναπαράσταση των αποτελεσμάτων ενός πειράματος τριαδικής σύγκρισης σε κατευθυνόμενο γράφο ανομοιοτήτων. Στην συνέχεια διατυπώθηκε το πρόβλημα μάθησης της απόστασης ή ανομοιότητας μεταξύ των μουσικών κομματιών και εξηγήθηκε ο τρόπος με τον οποίο αυτό μπορεί να μετασχηματιστεί αυτό σε πρόβλημα δυαδικής ταξινόμησης.

Εν συνεχεία, έγινε χρήση πέντε διαφορετικών ταξινομητών προκειμένου να διερευνηθεί η ικανότητά τους να επιλύσουν το πρόβλημα της ταξινόμησης και μελετήθηκε η επίδοση τους σε σχέση με το μέγεθος του συνόλου εκπαίδευσης, σε σχέση με τον τρόπο κατασκευής των συνόλων εκπαίδευσης και ελέγχου, καθώς και σε σχέση με το διάνυσμα χαρακτηριστικών που χρησιμοποιείται σαν περιγραφέας για τα μουσικά αποσπάσματα. Παρατηρήθηκε ότι το Τεχνητό Νευρωνικό Δίκτυο, το SVM με πυρήνα RBF και το SVNN πέτυχαν τιμές του f1-score περίπου 79% για το διάνυσμα των ENT χαρακτηριστικών και 80 – 82 % για το επαυξημένο διάνυσμα χαρακτηριστικών που χρησιμοποιήθηκε. Διαπιστώσαμε ότι το Νευρωνικό Δίκτυο εκμεταλλεύτηκε καλύτερα την πληροφορία των επιπλέον χαρακτηριστικών που προσθέσαμε, αγγίζοντας όμως την καλύτερη επίδοση του για μεγαλύτερο μέγεθος συνόλου εκπαίδευσης σε σχέση με τα SVM και SVNN. Αξίζει να σημειωθεί ότι παρόλο που τα διανύσματα των μουσικών χαρακτηριστικών που χρησιμοποιήθηκαν αποτελούνταν από πολύ λίγες διαστάσεις σε σχέση με τους περιγραφείς που χρησιμοποιούνται σε εμπορικές MIR εφαρμογές, τα αποτελέσματα που προέκυψαν ήταν αξιόλογα.

Στη συνέχεια επιχειρήθηκε μια αναπαράσταση της μουσική ομοιότητας με πληροφορία χωρικής εγγύτητας σε μια τρισδιάστατη απεικόνιση μέσω της μεθόδου MDS για το σύνολο των κομματιών της βάσης του MagnaTagATune με βάση τα συναπτικά βάρη που προέκυψαν από την εκπαίδευση του Νευρωνικού Δικτύου και χρησιμοποιήσαμε tags που σχετίζονται με το μουσικό είδος, προκειμένου να διαπιστώσουμε τη σχέση της μουσικής ομοιότητας με την κατάταξη ανά μουσικό είδος. Μελετήθηκαν οι πίνακες αποστάσεων μεταξύ των κομματιών, πριν γίνει η χρήση των συναπτικών βαρών του νευρωνικού και μετά, και διαπιστώθηκε σημαντική μεταβολή στις αποστάσεις. Από την

απεικόνιση μέσω της MDS παρατηρήθηκε ότι η μουσική ομοιότητα που «έμαθε» το σύστημά μας σχετίζεται αρκετά με την κατηγοριοποίηση σε μουσικά είδη, αφού διαφάνηκε ένας διαχωρισμός μεταξύ των κομματιών που είναι χαρακτηρισμένα ως rock και metal από τα κλασικά και ambient, καθώς και διαχωρισμός μεταξύ των techno και των κλασσικών κομματιών.

Τέλος επιχειρήσαμε να ερμηνεύσουμε τον πίνακα των συναπτικών βαρών που προέκυψαν από την εκπαίδευση του Νευρωνικού Δικτύου για να διαπιστώσουμε την σημαντικότητα του κάθε χαρακτηριστικού στον καθορισμό της ομοιότητας. Παρατηρήθηκε μια αρκετά διαφορετική ιεραρχία σημαντικότητας μεταξύ των δύο διανυσμάτων χαρακτηριστικών και έγινε μια προσπάθεια ερμηνείας αυτών. Συμπεράναμε ότι ίσως ο πιο σημαντικός παράγοντας (από αυτούς που δοκιμάστηκαν) στην απόφαση περί ομοιότητας-ανομοιότητας είναι η Ενέργεια Βραχέως Χρόνου, και όταν δεν παρέχεται αυτή, σημαντικό παράγοντα αποτελεί η Συνολική Ένταση του κομματιού.

10.2 Μελλοντικές Επεκτάσεις

Δυνατότητες μελλοντικής εργασίας πάνω στο ίδιο αντικείμενο αναδεικνύονται πάρα πολλές. Αυτές περιλαμβάνουν την επανάληψη των πειραμάτων με χρήση περισσότερων μουσικών χαρακτηριστικών, έτσι ώστε να μπορέσει να καλυφθεί το πολύπλευρό φάσμα της έννοιας της μουσικής ομοιότητας, καθώς και η μελέτη της επίδρασης των χαρακτηριστικών που θα χρησιμοποιηθούν στο αποτέλεσμα των πειραμάτων. Μια απαραίτητη προσθήκη είναι αυτή των χαρακτηριστικών σημασιολογικού περιεχομένου, όπως τα tags, η συνύπαρξη κομματιών στην ίδια playlist κ.α.

Ενδεχομένως να είναι χρήσιμο να δοκιμαστεί ένα Ensemble σύστημα ταξινομητών, όπου κάθε ταξινομητής θα ταξινομεί τις εισόδους βασιζόμενος σε διαφορετικούς περιγραφείς του μουσικού κομματιού, αποφασίζοντας έτσι για την ομοιότητα σύμφωνα με μια διαφορετική ηχητική παράμετρο και στο τέλος να πραγματοποιείται ο συνδυασμός των επιμέρους αποτελεσμάτων μέσω voting ή κάποιας άλλης τεχνικής. Με την μέθοδο αυτή διευκολύνεται η χρήση διαφορετικών μουσικών χαρακτηριστικών, χωρίς να είναι απαραίτητο να διαθέτουν όλα μια κοινή αναπαράσταση.

Ακόμα ίσως θα ήταν σκόπιμη η δοκιμή της διαφορετικής αναπαράστασης των ENT χαρακτηριστικών ηχοχρώματος μέσω Gaussian Mixture Models ή πολυμεταβλητών Γκαουσσιανών. Αν και αυτά τα μοντέλα έχουν χρησιμοποιηθεί κατά κόρον στο παρελθόν για τα συνήθη MFCC θα ήταν χρήσιμη η εκπόνηση μιας συγκριτικής μελέτης ανάλογης της [Aucouturier et al., 2004] που κατέληγε στο βέλτιστο μοντέλο αναπαράστασης των ENT features για κάθε κομμάτι.

Μια ακόμα εναλλακτική που θα ήταν άξιο να διερευνηθεί θα ήταν η υιοθέτηση των υποθέσεων του Tversky ([Tversky, 1977]), περί των ιδιοτήτων της ασυμμετρίας και της κατευθυντικότητας της έννοιας της ομοιότητας, και κατ' επέκταση της μουσικής ομοιότητας, και η κατασκευή ενός μοντέλου που θα υλοποιεί αυτήν την αναπαράσταση με χρήση ασαφούς λογικής, με παρόμοιο τρόπο όπως στα [Santini et al.,1999], [Eckhardt et al., 2009].

Η σύγκριση της μοντελοποίησης αυτής κατά Tversky και της συνήθους μοντελοποίησης μέσω μετρικών θα οδηγούσε σε χρήσιμα αποτελέσματα, τόσο για την θεωρητική όσο και για την πρακτική προσέγγιση του ζητήματος της ομοιότητας.

Οι ταξινομητές που κατασκευάσαμε θα μπορούσαν να αποτελέσουν βάση ενός recommender συστήματος, το οποίο εκμεταλλευόμενο το μέτρο απόστασης που έχει υπολογιστεί μέσω των συναπτικών βαρών του ταξινομητή, και πιθανών εκμεταλλευόμενο και πληροφορία μεταδεδομένων προκειμένου να καλυφθεί το σημασιολογικό κενό μεταξύ ηχητικού σήματος και ανθρώπινης αντίληψης, θα προβαίνει σε προτάσεις προς ακρόαση.
Εναλλακτικά θα μπορούσε να κατασκευαστεί ένα σύστημα αυτόματης παραγωγής playlist βασισμένης σε έναν αλγόριθμο βραχύτερου μονοπατιού (shortest path) πάνω στην MDS απεικόνιση.

Βιβλιογραφία:

[Ashby et al., 1988] Ashby, F. G., and N. A. Perrin. (1988). Toward a unified theory of similarity and recognition. Psychological Review 95: 124–150.

[Aucouturier et al., 2002] J.-J. Aucouturier and F. Pachet. (2002). Music similarity measures: What's the use?. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR).

[Aucouturier et al., 2002b] J.-J. Aucouturier and F. Pachet. (2002). Finding songs that sound the same. In Proceedings of the IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002), (Leuven, Belgium).

[Aucouturier et al., 2004] J. Aucouturier, F. Pachet. (2004). Improving Timbre Similarity: How high's the sky ? . *Journal of Negative Results in Speech and Audio Sciences*, Vol. 1, No. 1.

[Berenzweig, 2003] Berenzweig, A., Logan, B., Ellis, D. P. W. Whitman, B. (2003). A large-scale evaluation of acoustic and subjective music similarity measures. *Computer Music Journal* 28, 63–76.

[Bertin-Mahieux et al., 2010] T. Bertin-Mahieux, D. Eck, M. Mandel. (2010). Automatic tagging of audio: The state-of-the-art. In *Machine Audition: Principles, Algorithms and Systems*. IGI Publishing.

[Bogert et al., 1963] B. P. Bogert, M. J. R. Healy, and J. W. Tukey. (1963). The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking. *Proceedings of the Symposium on Time Series Analysis* (M. Rosenblatt, Ed) Chapter 15, 209-243.

[Bozzon et al., 2008] A. Bozzon, G. Prandi, G. Valenzise, M. Tagliasacchi. (2008). A music recommendation system based on semantic audio segments similarity, EuroIMSA2008 Internet and Multimedia Systems and Applications) , March 17 – 19, 2008 Innsbruck, Austria.

[Burges, 1998] C. J. C. Burges. (1998). A tutorial on support vector machines for pattern recognition. In *Data Mining and Knowledge Discovery*, 2(2):121–167.

[Burton et al., 1976] Burton, M. L. , Nerlove, S. B. (1976). Balanced designs for triads tests: Two examples from English. *Social Science Research* 5, 247–267.

[Cambouropoulos, 2009] E. Cambouropoulos (2009). How similar is similar?. In *Musicae Scientiae*

[Casey et al., 2008] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney (2008). Content-Based Music Information Retrieval: Current Directions and Future Challenges. Proceedings of the IEEE 96(4), 668–695.

[Cheng et al., 2008] W. Cheng, E. Hüllermeier. (2008). Learning similarity functions from qualitative feedback. In *Proc. of ECCBR 2008*.

[Chang et al., 2011] C.-C. Chang, C.-J. Lin. (2011). LIBSVM: alibrary for support vector machines. ACM *Transactions on Intelligent Systems and Technology.*

[Downie et al., 2005] Downie, J. Stephen, Kris West, Andreas Ehmann , Emmanuel Vincent (2005). The 2005 Music Information Retrieval Evaluation eXchange (MIREX 2005): Preliminary

Overview. In *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR 2005), London, UK, 11-15 September 2005*. Queen Mary, UK: University of London, pp. 320-323.

[EchoNest, 2009] The Echo Nest, 2009. http://www.echonest.com/.

[Eckhardt et al., 2009] A. Eckhardt, T. Skopal, P. Vojtás. (2009) *On Fuzzy vs. Metric Similarity Search in Complex Databases. In Flexible Query Answering Systems Lecture Notes in Computer Science* Volume 5822, 2009, pp 64-75 2009: 64-75

[Freed, 2006] A. Freed (2006). Music metadata quality: A multiyear case study using the music of Skip James. In *Proc. AES 121st Conv.,* San Francisco, CA.

[Goodman, 1972] N. Goodman (1972). Seven strictures on similarity. In *Project and Problems*, by N. Goodman, The bobbs-merrill Company, inc., New York.

[Haykin, *1994]* S. Haykin. (1994). *Neural Networks: A Comprehensive Foundation.* MacMillan Publishing Company.

[Herre et al., 2001] Herre, J. & Allamanche, E. & Hellmuth, O. (2001). Robust Matching of Audio Signals Using Spectral Flatness Features. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, (pp. 127:130)*, New Paltz, NY.

[Herre et al., 2003] J. Herre, E. Allamanche, C. Ertel (2003). How similar do songs sound? towards modeling human perception of musical similarities. *In Proceedings of IEEE Wokshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Mohonk, NY.

[Hofmann, 2002] Hofmann L. (2002). Rhythmic Similarity: A theoretical and empirical approach. In *Proceedings of the 7th International Conference on Music Perception and Cognition*, *Sydney 2002*. Ed. C. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick. Adelaide, Causal Productions, 2002

[Jehan et al., 2011] T. Jehan, D. DesRoches. (2011). Analyzer documentation (analyzer version 3.08). Website.

[Lamont et al., 2001]Lamont, A. and Dibben, N. (2001). Motivic structure and the perception of similarity.*Music Perception* 18, 245–274.

[Lartillot, 2011] O. Lartillot. (2011). MIRtoolbox 1.3.4, for MatLab. User's Manual. b. http:// www.mathworks.com/matlabcentral/fileexchange/24583-mirtoolbox (accessed 2 July 2013).

[Law et al., 2009] E. Law, K. West, M. Mandel, M. Bay, J. S. Downie.(2009). Evaluation of algorithms using games: the case of music annotation. In *Proc. of the 10th Int. Conf. on Music Information Retrieval (ISMIR'09)*, pages 387–392.

[Logan et al., 2001] B. Logan and A. Salomon. (2001). A music similarity function based on signal analysis. In *Proc. IEEE Int. Conf. Multimedia Expo*, pp. 745–748.

[Ludwig, 2012] O. Ludwig. (2012). Study on Non-parametric Methods for Fast Pattern Recognition with Emphasis on Neural Networks and Cascade Classifiers. PhD Thesis, University of Coimbra, Coimbra, 2012.

[Kruskal et al., 1986] J. Kruskal, M. Wish. (1986). Multidimensional Scaling. Sage.

[Magnatune]http://magnatune.com/.

[Mandel et al., 2008] M.I. Mandel, D. P. W. Ellis. (2008). A web-based game for collecting music metadata. In *Journal of New Music Research*,37(2):151–165.

[McAdams et al., 2004] McAdams, S., Vieillard, S., Houix, O., Reynolds, R. (2004). Perception of musical similarity among contemporary thematic materials in two instrumentations. Music Perception 22, 207–237.

[McFee., 2009] B. McFee, G. Lanckriet. (2009). Heterogeneous embedding for subjective artist similarity. In *Proc. of the 10th Int. Conf. on Music Information Retrieval (ISMIR'09)*.

[McFee et al., 2010]B. Mcfee and G. Lanckriet. (2010). Metric learning to rank. In *Proceedings of the 27th annual International Conference on Machine Learning (ICML)*.

[Mckay, 2005] C. McKay (2005). Automatic music classification and similarity analysis. *Course Paper.* Université de Montreal, Canada.

[Mendel et al., 1970] Mendel, J. M., McLaren, R. W. (1970). Reinforcement learning control and pattern recognition systems. In Mendel, J. M. and Fu, K. S., editors, *Adaptive, Learning and Pattern Recognition Systems: Theory and Applications*, pages 287-318. Academic Press, New York.

[MIREX, 2005] http://www.music-ir.org/mirex/wiki/2005:Main_Page

[MIREX, 2013] http://www.music-ir.org/mirex/wiki/2013:Main_Page

[Mitrovic et al., 2010] D. Mitrovic, M. Zeppelzauer, C. Breiteneder. (2010). Features for Content-Based Audio Retrieval. In *Advances in Computers Volume 78 Improving the Web*, Elsevier :71 - 150.

[Müllensiefen et al., 2004] Daniel Müllensiefen, Klaus Frieler. (2004). Melodic Similarity: Approaches and Applications. In *Proc. of the 8th International Conference on Music Perception and Cognition (ICMPC8),* Evanston, IL.

[Novello et al., 2006] A. Novello, M. F. Mckinney A. Kohlrausch. (2006). Perceptual evaluation of music similarity. In ISMIR 2006.

[Orpen et al., 1992] K. S. Orpen, Huron, D. (1992). Measurements of similarity in music: A quantitative approach for non parametric representations. Computers in Music Research 4, 1–44.

[Rabiner et al., 1993] L. Rabiner , B. Juang.(1993). Fundamentals of speech recognition. Prentice-Hall.

[Santini et al.,1999] S. Santini, Ramesh J. (1999). *Similarity measures*. IEEE Pattern Analysis and Machine Intelligence, 21(9):871–883.

[Schnitzer et al., 2011] Schnitzer D., Flexer A., Schedl M., Widmer G. (2011). Using Mutual Proximity to Improve Content-Based Audio Similarity. InProceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR'11), Miami, FL, USA.

[Schwartz, 2004] B. Schwartz. (2004). The tyranny of choice. In *Scientific American Mind*.

[Schwartz, 2010] P. Schwartz. (2010). Music Information Retrieval. *Topics in Sound and Music – Volume 0*

[Scholkopf et al., 2001] B. Scholkopf , A. J. Smola. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[Shepard, 1964] R. N. Shepard. (1964). Circularity in judgements of relative pitch. The Journal of the Acoustical Society of America, 36: 2346-2353.

[Stober et al., 2011] S. Stober, A. Nürnberger. (2011). An Experimental Comparison of Similarity Adaptation Approaches. *Adaptive Multimedia Retrieval 2011*: 96-113

[Stober, 2011b] S. Stober. (2011). Adaptive Distance Measures for Exploration and Structuring of Music Collections. In *Proc. of the AES 42ND International Conference*, Ilmenau, Germany.

[Tolos et al., 2005] M. Tolos, R. Tato, T. Kemp. (2005). Mood-based navigation through large collections of musical data. 2nd IEEE Consumer Communications and Networking Conference, pages 71--75.

[Tversky, 1977] A. Tversky. (1977). Features of similarity. *Psychological Review*, 84:327–352.

[Tzanetakis et al., 2001] G. Tzanetakis, G. Essl, P.Cook. (2001). Automatic Musical Genre Classification of Audio Signals. In Proc. Int. Symposium on Music Information Retrieval (ISMIR2001), Bloomington.

[Wang et al., 2011] J.-C. Wang, H.-S. Lee, H.-M. Wang, S.-K. Jeng. (2011). Learning the similarity of audio music in bag-of-frames representation from tagged music data. In *Proc. of 12th Int. Society for Music Information RetrievalConference, (ISMIR 2011),* pp. 85–90.

[Wolf et al., 2011] D. Wolff, T. Weyde. (2011). Adapting metrics for music similarity using comparative judgements. In *Proc. International Symposium on Music Information Retrieval (ISMIR* 2011).

[Wolf et al., 2012] D. Wolf, T. Weyde, S. Stober, A. Nuernberger. (2012). A Systematic Comparison of Music Similarity Adaptation Approaches. In *Proc. of the 13th Int. Conf. on Music Information Retrieval (ISMIR 2012).*

[Wolf et al., 2012b] D. Wolf, T. Weyde. (2012). Adapting similarity on the MagnaTagATune database: effects of model and feature choices. In A. Mille, F. L. Gandon, J. Misselis, M. Rabinovich, and S. Staab (Eds.), *WWW Companion Volume,* 931-936, ACM.

[Zhang et al., 2006] Zhang, X. , Ras, Zbigniew W. (2006). Sound Isolation by Harmonic Peak Partition For Music Instrument Recognition. In *Fundamenta Informaticae Journal Special issue on Tilings and Cellular Automata*, IOS Press, pp. 612-628.

[Zhang et al., 2010] Z. Zhang, Y. Takane. (2010). Statistics: Multidimensional Scaling, In *E. Baker, B. McGaw & P. Peterson (Eds.), International Encyclopedia of Education* (3rd Edition).Oxford, UK: Elsevier.