



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών
«Πληροφορική»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Αυτόματη Ταξινόμηση μελωδίας σε μουσικά είδη με τη χρήση τεχνητών νευρωνικών δικτύων
Όνοματεπώνυμο Φοιτητή	Βιδάλη Ροδάνθη
Πατρώνυμο	Γεώργιος
Αριθμός Μητρώου	ΜΠΠΛ/06026
Επιβλέπων	Γεώργιος Τσιχριντζής, Καθηγητής

31 Μαρτίου

Έτος 2011

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΑΙΑ

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

(υπογραφή)

(υπογραφή)

Τσιχριντζής Γεώργιος
Καθηγητής

Αποστόλου Δημήτριος
Επίκουρος Καθηγητής

Κωνσταντόπουλος Χαράλαμπος
Λέκτορας

Ευχαριστίες

Θα θελα να ευχαριστήσω θερμά τον καθηγητή Γ. Τσιχριντζή για την εμπιστοσύνη που μου έδειξε, και τη δυνατότητα που μου έδωσε ώστε να ασχοληθώ με το αντικείμενο της Αναγνώρισης Προτύπων και των Τεχνητών Νευρωνικών Δικτύων.

Θα ήθελα επίσης να εκφράσω την ευγνωμοσύνη μου στους γονείς μου για την διαρκή τους υποστήριξη, που επέτρεψε την επιτυχή διεκπεραίωση των σπουδών μου. Τέλος θα ήθελα να ευχαριστήσω όλους τους φίλους μου.

Πίνακας Περιεχομένων

Περίληψη	8
Abstract	9
1. Εισαγωγή	10
1.1. Στόχοι και σύνοψη της Διατριβής	11
2. Αναγνώριση Προτύπων	12
2.1. Ιστορική Αναδρομή	12
2.2. Εφαρμογές της Αναγνώρισης Προτύπων	12
2.3. Μεθοδολογίες Αναγνώρισης Προτύπων	13
2.3.1. Αισθητήρες	14
2.3.2. Τμηματοποίηση και Ομαδοποίηση	14
2.3.3. Εξαγωγή Χαρακτηριστικών	15
2.3.4. Ταξινόμηση	15
3. Εισαγωγή Στα Τεχνητά Νευρωνικά Δίκτυα	16
3.1. Από τα Βιολογικά στα Τεχνητά Νευρωνικά Δίκτυα	16
3.2. Εισαγωγή στους Τεχνητούς Νευρώνες	17
3.3. Ιστορική Αναδρομή	19
3.4. Αρχιτεκτονικές των ΤΝΔ	20
3.4.1. ΤΝΔ πρόσθιας τροφοδότησης	20
3.4.2. ΤΝΔ με ανατροφοδότηση	20
3.4.3. Perceptrons	20
3.5. Πλεονεκτήματα των Νευρωνικών Δικτύων	21
3.6. Εφαρμογές των ΤΝΔ	22
3.7. Μέθοδοι εκπαίδευσης	22
4. Τεχνικές ταξινόμησης μελωδίας σε μουσικά είδη	24
4.1. Είδη μουσικής	24
4.2. Εξαγωγή Χαρακτηριστικών Γνωρισμάτων	25
4.2.1. Χροιά	26
4.2.2. Μελωδία και Αρμονία	27
4.2.3. Ρυθμός	28
4.2.4. Εξαγωγή χαρακτηριστικών γνωρισμάτων από ηχητικά τμήματα	28
4.3. Έμπειρα Συστήματα	30
4.4. Προσέγγιση χωρίς επίβλεψη	30
4.4.1. Μέτρα ομοιότητας	31
4.4.2. Αλγόριθμοι Ομαδοποίησης	31
4.5. Προσέγγιση με επίβλεψη	32
4.5.1. Ταξινομητές με επίβλεψη	32

5.	Εισαγωγή στο Neural Network Toolbox του Matlab	35
5.1.	Μοντέλο Νευρώνων	35
5.2.	Συναρτήσεις Μεταφοράς.....	36
5.3.	Νευρώνες με Είσοδο Διάνυσμα.....	38
5.4.	Αρχιτεκτονικές Δικτύου	39
5.5.	Εκπαίδευση Δικτύου Νευρώνων	42
5.5.1.	Επαυξητική εκπαίδευση	42
5.5.1.1.	Επαυξητική εκπαίδευση σε στατικά δίκτυα	42
5.5.1.2.	Επαυξητική εκπαίδευση σε δυναμικά δίκτυα	44
5.5.2.	Εκπαίδευση σε δεσμίδες.....	44
5.5.2.1.	Εκπαίδευση σε δεσμίδες σε στατικό δίκτυο	45
5.5.2.2.	Εκπαίδευση σε δεσμίδες σε δυναμικό δίκτυο	46
5.5.3.	Αλγόριθμος εκπαίδευσης με ελαστική επανατροφοδότηση (trainrp)	47
5.5.4.	Αλγόριθμος εκπαίδευσης Levenberg-Marquardt (trainlm)	48
6.	Ανάλυση μελωδίας: Θεωρία και Εργαλεία.....	49
6.1.	Εισαγωγή στη θεωρία των μουσικών οργάνων	49
6.2.	Ανάλυση Αρμονικών	52
6.3.	Χαρακτηριστικά γνωρίσματα ηχητικού σήματος.....	54
7.	Ταξινόμηση σε μουσικά είδη με τη χρήση τεχνητών νευρωνικών δικτύων	59
7.1.	Αρχιτεκτονικές ΤΝΔ.....	59
7.1.1.	Αρχιτεκτονική με Feed-Forward Neural Network	59
7.1.2.	Αρχιτεκτονική με Cascade-Forward back-propagation Neural Network.....	60
7.1.3.	Αρχιτεκτονική με Elman back-propagation Neural Network	60
7.2.	Πειράματα και Αριθμητικά αποτελέσματα.....	61
7.2.1.	Υλοποίηση με Feed-Forward Neural Network	61
7.2.2.	Υλοποίηση με Cascade-Forward Neural Network	67
7.2.3.	Υλοποίηση με Elman Back-propagation Neural Network	70
8.	Συμπεράσματα	74
	Βιβλιογραφία	76
	Παράρτημα –Κώδικας	80

Πίνακας Σχημάτων

Σχήμα 1: Βήματα Αναγνώρισης Προτύπων.....	14
Σχήμα 2: Απλουστευμένο μοντέλο ενός βιολογικού νευρώνα.....	16
Σχήμα 3 Νευρώνας.....	17
Σχήμα 4: Ένα πολυστρωματικό πλήρως διασυνδεδεμένο πρόσθιας τροφοδοσίας ΤΝΔ .	18
Σχήμα 5: Ένα πολυστρωματικό ΤΝΔ με ανατροφοδότηση.....	20
Σχήμα 6: Ένα νευρωνικό δίκτυο.....	35
Σχήμα 7: Μοντέλο Νευρώνων.....	36
Σχήμα 8: Συνάρτηση Μεταφοράς Hard-Limit.....	36
Σχήμα 9: Γραμμική συνάρτηση μεταφοράς.....	37
Σχήμα 10: Συνάρτηση Sigmoid.....	37
Σχήμα 11 Συνάρτηση tansig.....	38
Σχήμα 12: Είσοδος Διάνυσμα.....	38
Σχήμα 13: Σύντομος Συμβολισμός Νευρώνα.....	39
Σχήμα 14: Στρώμα Νευρώνων.....	40
Σχήμα 15 Περιεκτική περιγραφή στρώματος.....	41
Σχήμα 16 Δίκτυο πολλαπλών στρωμάτων.....	41
Σχήμα 17 Δίκτυο πολλαπλών στρωμάτων.....	42
Σχήμα 18: Τρόποι ταλάντωσης μιας χορδής.....	51
Σχήμα 19: Απλή ανάλυση ήχου.....	52
Σχήμα 20: Νότες πιάνου.....	53
Σχήμα 21: Ανάλυση διάφορων οργάνων.....	54
Σχήμα 22 Zerocross (πηγή MIRtoolbox Users Guide).....	55
Σχήμα 23 Centroid (πηγή MIRtoolbox Users Guide).....	55
Σχήμα 24 Spread (πηγή MIRtoolbox Users Guide).....	55
Σχήμα 25 Skewness (πηγή MIRtoolbox Users Guide).....	56
Σχήμα 26 Onset detection Curve (πηγή MIRtoolbox Users Guide).....	57
Σχήμα 27 Rollof (πηγή MIRtoolbox Users Guide).....	57
Σχήμα 28 Brightness (πηγή MIRtoolbox Users Guide).....	58
Σχήμα 29 Feed-Forward Neural Network.....	59
Σχήμα 30 Cascade-Forward Neural Network.....	60
Σχήμα 31 Elman Neural network.....	60
Σχήμα 32 Κατάληξη του αλγόριθμου εκπαίδευσης trainlm (Levenberg-Marquardt backpropagation) για δίκτυο FF με 20 νευρώνες.....	62
Σχήμα 33 Διαδικασία Εκπαίδευσης: Απόδοση του δικτύου FF με 20 νευρώνες.....	63
Σχήμα 34 Μελέτη απόδοσης δικτύου FF με 20 νευρώνες: Πραγματικές και προσομοιωμένες τιμές για ηχητικά κομμάτια.....	64

Σχήμα 35 Ακρίβεια υπολογισμού δείγματος για FF-20: Άξονας X: πλήθος τιμών που χρησιμοποιήθηκαν, Y: απόκλιση από την πραγματική τιμή	65
Σχήμα 36 Μελέτη απόδοσης δικτύου FF με 10 νευρώνες: Πραγματικές και προσομοιωμένες τιμές για ηχητικά κομμάτια.	66
Σχήμα 37 Διαδικασία εκπαίδευσης του FF με 10 νευρώνες	66
Σχήμα 38 Ακρίβεια υπολογισμού δείγματος για FF-10: Άξονας X: πλήθος τιμών που χρησιμοποιήθηκαν, Y: απόκλιση από την πραγματική τιμή	67
Σχήμα 39 Κατάληξη του αλγόριθμου εκπαίδευσης trainlm (Levenberg-Marquardt backpropagation) για δίκτυο CF με 10 νευρώνες.....	68
Σχήμα 40 Μελέτη απόδοσης δικτύου CF με 10 νευρώνες: Πραγματικές και προσομοιωμένες τιμές για ηχητικά κομμάτια.	69
Σχήμα 41 Ακρίβεια υπολογισμού δείγματος για CF-10: Άξονας X: πλήθος τιμών που χρησιμοποιήθηκαν, Y: απόκλιση από την πραγματική τιμή	70
Σχήμα 42 Κατάληξη του αλγόριθμου εκπαίδευσης trainlm (Resilient backpropagation) για δίκτυο Elman με 10 νευρώνες	71
Σχήμα 43 Μελέτη απόδοσης δικτύου Elman με 10 νευρώνες: Πραγματικές και προσομοιωμένες τιμές για ηχητικά κομμάτια.	72
Σχήμα 44 Ακρίβεια υπολογισμού δείγματος για Elman-10: Άξονας X: πλήθος τιμών που χρησιμοποιήθηκαν, Y: απόκλιση από την πραγματική τιμή	73

Περίληψη

Στα πλαίσια της εργασίας αυτής μελετάται η αυτόματη ταξινόμηση μελωδίας σε μουσικά είδη με τη χρήση τεχνητών νευρωνικών δικτύων. Τα νευρωνικά δίκτυα αποτελούνται από απλά στοιχεία που λειτουργούν παράλληλα. Τα στοιχεία ονομάζονται νευρώνες και προσομοιώνουν σε ένα βαθμό τα βιολογικά νευρωνικά συστήματα. Μπορεί κανείς να εκπαιδεύσει ένα νευρωνικό δίκτυο ώστε αυτό να προσομοιώνει μια ιδιαίτερη συνάρτηση με τη ρύθμιση των τιμών των συνδέσεων (βάρη) μεταξύ των στοιχείων. Χαρακτηριστικά, τα νευρωνικά δίκτυα ρυθμίζονται, ή εκπαιδεύονται, έτσι ώστε μια συγκεκριμένη είσοδος να οδηγεί σε μια συγκεκριμένη επιθυμητή έξοδο.

Στα πλαίσια αυτής της εργασίας, θα γίνει χρήση διάφορων περιγραφών που μπορούν να χαρακτηρίσουν μια μουσική ακολουθία, προκειμένου να εκπαιδευτεί ένα νευρωνικό δίκτυο ώστε να μπορεί να ξεχωρίζει το είδος της μουσικής που δέχεται ως είσοδο. Παράλληλα, θα δοκιμαστούν διαφορετικές αρχιτεκτονικές τεχνητών νευρωνικών δικτύων (στρώματα, πλήθος νευρώνων ανά στρώμα, συναρτήσεις μεταφοράς, διαδικασία ενημέρωσης των βαρών). Κάθε ξεχωριστό είδος δικτύου θα τροφοδοτηθεί με μερικές ώρες μουσικής από διαφορετικά μουσικά είδη, έτσι ώστε να εκπαιδευτεί και να είναι ικανό να αναγνωρίζει τα διαφορετικά αυτά είδη. Για την υλοποίηση της εργασίας θα χρησιμοποιηθεί το Neural Network Toolbox του Matlab.

Μέσα από τα αριθμητικά αποτελέσματα των πειραμάτων θα επαληθευθεί η καταλληλότητα των τεχνητών νευρωνικών δικτύων για την η αυτόματη ταξινόμηση μελωδίας σε μουσικά είδη και θα εξαχθούν επιμέρους συμπεράσματα για κάθε αρχιτεκτονική.

Abstract

In this thesis we study the automatic classification of music into genres using artificial neural networks (ANN). Artificial neural networks consist of simple elements operating in parallel. These elements are called neurons and to some extent simulate the biological neural systems. One can train a neural network so that it simulates a particular function by adjusting the connections (weights) between elements. Typically, neural networks are configured, or trained, so that a particular input leads to a specific desired output.

In this work we use various descriptors that characterize a musical sequence in order to train a neural network to be able to distinguish the type of music given input we will test different architectures of artificial neural networks (layers, number of neurons per layer, transfer functions, weight updating process). Each separate type of ANN is provisioned with a few hours of music from different genres in order to be trained to recognize these different genres. For the implementation of the work we use the Matlab Neural Network Toolbox.

Through experimental numerical results we verify the applicability of artificial neural networks for automatic classification of melody in musical genres and draw different conclusions for each tested architecture.

1. Εισαγωγή

Η δημιουργία τεράστιων βάσεων δεδομένων που προέρχονται τόσο από την επανάκτηση ήδη υπάρχοντων αναλογικών αρχείων όσο και από νέο περιεχόμενο, απαιτεί όλο και περισσότερο αξιόπιστα και γρηγορότερα εργαλεία για την ανάλυση του περιεχομένου. Τα εργαλεία αυτά μπορούν να χρησιμοποιηθούν στις αναζητήσεις/ερωτήσεις περιεχομένου. Για το λόγο αυτό, τα μουσικά είδη είναι κρίσιμοι περιγραφείς δεδομένου ότι έχουν χρησιμοποιηθεί ευρέως για χρόνια για να οργανώσουν τους καταλόγους μουσικής, τις βιβλιοθήκες και τα καταστήματα μουσικής. Παρά τη χρήση τους, τα μουσικά είδη παραμένουν μια μη καλώς ορισμένη έννοια, γεγονός που καθιστά το πρόβλημα της αυτόματης ταξινόμησης μη τετριμμένο.

Τα είδη μουσικής είναι ο πιο συνήθης τρόπος κατηγοριοποίησης που χρησιμοποιείται από τους εμπόρους και τους βιβλιοθηκάρους μουσικής για την οργάνωση των συλλογών μουσικής. Με την ηλεκτρονική διανομή μουσικής (Electronic Music Distribution), οι κατάλογοι μουσικής τείνουν να γίνουν τεράστιοι (υπάρχουν online υπηρεσίες οι οποίες παρέχουν πάνω από 1 εκατομμύριο κομμάτια όπως είναι η Amazon και η Apple - iTunes). Για το λόγο αυτό η σύνδεση ενός είδους με ένα μουσικό κομμάτι είναι σημαντική διότι βοηθάει τους χρήστες να βρουν αυτό που ψάχνουν.

Στην πραγματικότητα, το πλήθος της ψηφιακής μουσικής απαιτεί αποτελεσματικούς τρόπους αναζήτησης, οργάνωσης και δυναμικής ενημέρωσης των συλλογών μουσικής και σίγουρα απαιτεί νέα μέσα για αυτόματο σχολιασμό. Συγχρόνως, ακόμα κι αν όροι όπως τζαζ, ροκ ή ποπ χρησιμοποιούνται ευρέως, παραμένουν συχνά αόριστα καθορισμένοι έτσι ώστε το πρόβλημα της αυτόματης ταξινόμησης είδους να γίνεται ένας μη τετριμμένος στόχος. Η αυτόματη ταξινόμηση με βάση κάποια χαρακτηριστικά είναι μέρος του επιστημονικού πεδίου που ονομάζεται Αναγνώριση Προτύπων.

Ένας ορισμός για Αναγνώριση Προτύπων είναι ο ακόλουθος: Πρόκειται για την διαδικασία κατά την οποία σε πρώτο στάδιο γίνεται παρατήρηση απλών δεδομένων και στη συνέχεια γίνεται λήψη αποφάσεων με βάση την «κατηγορία» του προτύπου. Η αναγνώριση προτύπων (pattern recognition) είναι ο κλάδος εκείνος της επιστήμης ο οποίος περιγράφει και κατατάσσει τα αντικείμενα σε ένα αριθμό κατηγοριών. Τα αντικείμενα τα οποία κατατάσσονται σε κατηγορίες ονομάζονται πρότυπα (patterns). Ένα πρότυπο είναι η περιγραφή ενός αντικείμενου.

Η δημιουργία μηχανών οι οποίες αναγνωρίζουν τέτοια πρότυπα είναι απαραίτητη. Εφαρμογές όπως η αυτόματη αναγνώριση φωνής είτε η αναγνώριση αποτυπομάτων απαιτούν η αναγνώριση προτύπων από μια μηχανή να γίνεται με μεγάλη ακρίβεια και αξιοπιστία. Επίσης, η αναγνώριση προτύπων (π.χ αναγνώριση ομιλίας και η οπτική αναγνώριση) πολλές φορές εμπνέεται από τη φύση για τον σχεδιασμό των αλγορίθμων ή του υλικού που χρησιμοποιείται.

Χαρακτηριστικό παράδειγμα συστημάτων που έχουν εμπνευστεί από τη φύση και χρησιμοποιούνται εκτεταμένα στην αναγνώριση προτύπων είναι τα τεχνητά νευρωνικά δίκτυα. Το Τεχνητό Νευρωνικό Δίκτυο (ΤΝΔ) είναι μια αρχιτεκτονική δομή αποτελούμενη από ένα πλήθος διασυνδεδεμένων μονάδων, οι οποίες ονομάζονται τεχνητοί νευρώνες. Κάθε μονάδα έχει εισόδους και εξόδους και υλοποιεί τοπικά έναν απλό υπολογισμό. Είναι πιθανό να υπάρχει μιας μορφής λειτουργική ικανότητα κατά τη στιγμή της κατασκευής ενός δικτύου, αλλά σχεδόν πάντα τα νευρωνικά δίκτυα αναπτύσσουν μια συνολική λειτουργικότητα μέσω μιας μορφής εκπαίδευσης. Ένα ΤΝΔ καθορίζεται από την τοπολογία του δικτύου, τα χαρακτηριστικά των νευρώνων, τη μέθοδο εκπαίδευσης και από τα δεδομένα με τα οποία γίνεται η εκπαίδευση.

1.1. Στόχοι και σύνοψη της Διατριβής

Ο στόχος της παρούσας μεταπτυχιακής διατριβής είναι η διερεύνηση της δυνατότητας των νευρωνικών δικτύων να χρησιμοποιηθούν για την ανάλυση και κατάταξη ακολουθιών μουσικής.

Τα νευρωνικά δίκτυα αποτελούνται από απλά στοιχεία που λειτουργούν παράλληλα. Τα στοιχεία στα δίκτυα αποτελούν σε ένα βαθμό προσομοίωση των βιολογικών νευρωνικών συστημάτων. Όπως στη φύση, οι συνδέσεις μεταξύ των στοιχείων καθορίζουν κατά ένα μεγάλο μέρος τη συνάρτηση του δικτύου. Μπορεί κανείς να εκπαιδεύσει ένα νευρωνικό δίκτυο ώστε αυτό να προσομοιώνει μια ιδιαίτερη συνάρτηση με τη ρύθμιση των τιμών των συνδέσεων (βάρη) μεταξύ των στοιχείων. Χαρακτηριστικά, τα νευρωνικά δίκτυα ρυθμίζονται, ή εκπαιδεύονται, έτσι ώστε μια συγκεκριμένη είσοδος να οδηγεί σε μια συγκεκριμένη επιθυμητή έξοδο. Στα πλαίσια αυτής της εργασίας, θα γίνει χρήση διάφορων περιγραφών που μπορούν να χαρακτηρίσουν μια μουσική ακολουθία, προκειμένου να εκπαιδευτεί ένα νευρωνικό δίκτυο ώστε να μπορεί να ξεχωρίζει το είδος της μουσικής που δέχεται ως είσοδο.

Ενδεικτικά, θα δοκιμαστούν διαφορετικές αρχιτεκτονικές ΤΝΔ (στρώματα, πλήθος νευρώνων ανά στρώμα και συναρτήσεις μεταφοράς). Κάθε ξεχωριστό είδος ΤΝΔ θα τροφοδοτηθεί με μερικές ώρες κλασσικής, ελληνικής παραδοσιακής και σύγχρονης μουσικής, ώστε να εκπαιδευτεί να αναγνωρίζει τα διαφορετικά αυτά είδη. Για την υλοποίηση της εργασίας θα χρησιμοποιηθεί το Neural Network Toolbox του Matlab.

Το υπόλοιπο της διατριβής έχει την ακόλουθη δομή: στο επόμενο κεφάλαιο θα γίνει εισαγωγή στην Αναγνώριση προτύπων. Στο 3^ο κεφάλαιο θα γίνει εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα, ενώ στο Κεφάλαιο 4 θα περιγραφούν τεχνικές ταξινόμησης μελωδίας σε μουσικά είδη. Στο Κεφάλαιο 5 θα γίνει εισαγωγή στο Neural Network Toolbox του Matlab. Στο Κεφάλαιο 6 θα γίνει παρουσίαση της θεωρίας και εργαλείων για την ανάλυση της μελωδίας, ενώ στο 7^ο Κεφάλαιο προτείνονται, αναλύονται και εξετάζονται διάφορες αρχιτεκτονικές ΤΝΔ για την ταξινόμηση σε μουσικά είδη. Στο πρώτο μέρος του κεφαλαίου γίνεται η παρουσίαση διάφορων αρχιτεκτονικών ΤΝΔ, ενώ στο δεύτερο μέρος παρουσιάζονται τα πειράματα και τα αντίστοιχα αριθμητικά αποτελέσματα. Τέλος, το Κεφάλαιο 8 ολοκληρώνει την εργασία με την εξαγωγή συμπερασμάτων για την καταλληλότητα των ΤΝΔ για την αυτόματη ταξινόμηση μελωδίας σε μουσικά είδη.

2. Αναγνώριση Προτύπων

Η αναγνώριση προτύπων (*pattern recognition*) είναι το επιστημονικό πεδίο που ασχολείται με το σχεδιασμό μεθόδων και αλγορίθμων που αναλύουν την διαθέσιμη πληροφορία που προέρχεται από κάποιο (φυσικό ή τεχνητό) φαινόμενο με σκοπό να αναγνωρίζουν πρότυπα στα δεδομένα.

Η αναγνώριση προτύπων μπορεί να χωριστεί σε δύο ευρείες κατηγορίες: την αναγνώριση σαφών αντικειμένων (concrete items) και την αναγνώριση αφηρημένων αντικειμένων (abstract items). Η αναγνώριση των σαφών αντικειμένων περιλαμβάνει την αναγνώριση των χωρικών και χρονικών αντικειμένων. Παραδείγματα χωρικών αντικειμένων είναι τα δακτυλικά αποτυπώματα, οι χάρτες καιρού, οι φωτογραφίες κλπ. Παραδείγματα χρονικών αντικειμένων είναι οι κυματομορφές και οι υπογραφές. Η αναγνώριση αφηρημένων αντικειμένων περιλαμβάνει την αναγνώριση της λύσης ενός προβλήματος, μια παλιά συζήτηση ή διαφωνία, κλπ. Με άλλα λόγια, η αναγνώριση αντικειμένων χωρίς φυσική υπόσταση.

2.1. Ιστορική Αναδρομή

Οι πρώτοι που έθεσαν τις βάσεις της αναγνώρισης προτύπων ήταν ο Πλάτωνας και ο Αριστοτέλης οι οποίοι ήταν οι πρώτοι που διαχώρισαν την ουσιώδη ιδιότητα (τι μοιράζονται δηλαδή τα μέλη μιας κατηγορίας) από την επουσιώδη ιδιότητα (τις διαφορές δηλαδή ανάμεσα στα μέλη μιας κατηγορίας).

Η αναγνώριση προτύπων είναι η διαδικασία κατά την οποία ανιχνεύονται τέτοιες ουσιώδεις ιδιότητες από μια κατηγορία αντικειμένων. Ο Αριστοτέλης κατασκεύασε ένα σύστημα προκειμένου να μπορεί να γίνει ταξινόμηση των ζώων. Αυτό που έκανε σε πρώτη φάση ήταν να χωρίσει τα ζώα σε δύο κατηγορίες: σε αυτά που έχουν σπονδυλική στήλη και σε αυτά που δεν έχουν. Στη συνέχεια προχώρησε σε περαιτέρω διαχωρισμό των δυο ομάδων. Ο Θεόφραστος ταξινόμησε με παρόμοιο τρόπο τα φυτά. Στις αρχές του 18ου αιώνα ο Carolus Linnaeus χρησιμοποιώντας καινούριες γνώσεις ταξινόμησε τα ζώα, τα φυτά, τα πετρώματα αλλά και ορισμένες ασθένειες. Οι Hertzprung & Russell κατάφεραν να ταξινομήσουν τα άστρα με βάση τη λαμπρότητα και τη θερμοκρασία.

Η πρώτη προσέγγιση με τη χρήση μαθηματικής προτυποποίησης στην αναγνώριση προτύπων έγινε Ο Fisher το 1936. Η ανάπτυξη της αναγνώρισης προτύπων έχει προχωρήσει τόσο πολύ τα τελευταία 30 χρόνια ώστε πλέον θεωρείται ανεξάρτητη επιστήμη.

2.2. Εφαρμογές της Αναγνώρισης Προτύπων

Οι εφαρμογές που χρησιμοποιούν την Αναγνώριση Προτύπων είναι πολυάριθμες και χρησιμοποιούνται από πάρα πολλούς κλάδους της σύγχρονης επιστήμης. Ακολουθούν μερικά χαρακτηριστικά παραδείγματα:

- Αναγνώριση ομιλίας: Πρόκειται για συστήματα τα οποία αντιλαμβάνονται την ομιλία και την μετατρέπουν σε γραπτό κείμενο.
- Αναγνώριση οπτικών χαρακτήρων: (Optical Character Recognition - OCR) Μετατρέπει χειρόγραφο κείμενο είτε εικόνες σε κείμενο προς επεξεργασία. Υποκατηγορία της αναγνώρισης οπτικών χαρακτήρων είναι και η αναγνώριση χειρόγραφων κειμένων.

- Ταυτοποίηση ατόμων: Αφορά εφαρμογές ασφαλείας και χρησιμοποιεί διάφορες τεχνικές όπως η αναγνώριση δακτυλικών αποτυπωμάτων είτε η αναγνώριση χαρακτηριστικών του ματιού και λοιπά.
- Ιατρική διάγνωση: Αποτελεί πολύ χρήσιμο εργαλείο σε πολλούς κλάδους της ιατρικής. Μερικά παραδείγματα στα οποία χρησιμοποιείται η αναγνώριση προτύπων είναι στην ακτινογραφία, στην αξονική τομογραφία, στο υπερηχογράφημα, στα καρδιογραφήματα, στα εγκεφαλογραφήματα και σε πολλά άλλα.
- Χρηματοοικονομικές Εφαρμογές: συναλλαγές με πιστωτικές κάρτες που θεωρούνται ύποπτες, ταξινόμηση των αιτήσεων για έγκριση δανείου.
- Εξόρυξη δεδομένων (data mining): Το φιλτράρισμα και η εξαγωγή δεδομένων από τεράστιες βάσεις μπορεί να γίνει με τον έλεγχο και την ανίχνευση προτύπων στα πρωτογενή δεδομένα.

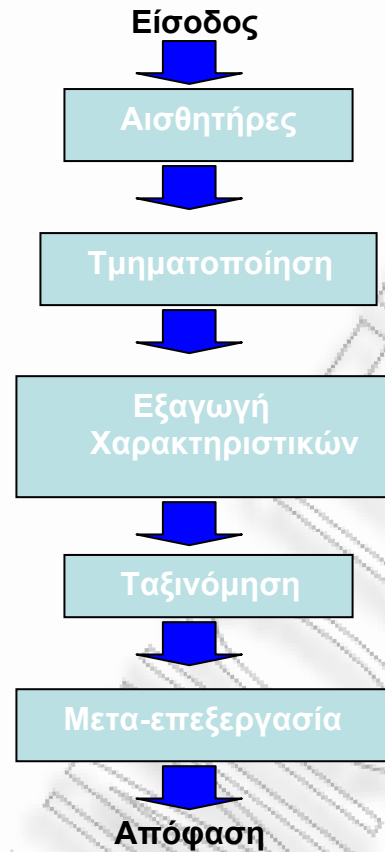
2.3. Μεθοδολογίες Αναγνώρισης Προτύπων

Υπάρχουν δύο τύποι αναγνώρισης προτύπων:

- Η στατιστική αναγνώριση προτύπων (statistical pattern recognition)
- Η συντακτική ή δομημένη αναγνώριση προτύπων (syntactic or structural pattern recognition).

Η συντακτική ή δομημένη αναγνώριση προτύπων (Syntactic pattern recognition ή structural pattern recognition) είναι ένας τύπος αναγνώρισης προτύπων στον οποίο κάθε αντικείμενο μπορεί να εκπροσωπείται από ένα μεταβλητών διαστάσεων σύνολο χαρακτηριστικών γνωρισμάτων (features). Με τον τρόπο αυτό η αναπαράσταση δομών προτύπων είναι δυνατή καθώς λαμβάνονται υπόψη πολύπλοκες σχέσεις ανάμεσα σε χαρακτηριστικά. Η συντακτική αναγνώριση προτύπων μπορεί να χρησιμοποιηθεί αντί της στατιστικής αναγνώρισης προτύπων, αν υπάρχει σαφής δομή μεταξύ των προτύπων. Ένας τρόπος για να παρουσιαστεί μια τέτοια δομή είναι με τη βοήθεια μιας σειράς από σύμβολα από κάποια τυπική γλώσσα. Σε αυτή την περίπτωση, οι διαφορές στις δομές των κλάσεων κωδικοποιούνται ως διαφορετικές γραμματικές.

Η αναγνώριση προτύπων ακολουθεί τα εξής τρία βασικά βήματα: την προεπεξεργασία, την εξαγωγή χαρακτηριστικών και την ταξινόμηση. Στο Σχήμα 1 αναλύονται τα βήματα αυτά σε περισσότερα επιμέρους βήματα. Προκειμένου να γίνει πιο κατανοητό το πρόβλημα της σχεδίασης ενός τέτοιου συστήματος πρέπει σε πρώτη φάση να γίνει επεξήγηση σχετικά με το είδος του προβλήματος που χρειάζεται να επιλυθεί από κάθε ένα από αυτά τα επιμέρους βήματα. Στη συνέχεια περιγράφεται ο τρόπος λειτουργίας για κάθε ένα από αυτά τα βήματα αλλά και τα προβλήματα που πιθανόν να παρουσιαστούν για κάθε βήμα ξεχωριστά.



Σχήμα 1: Βήματα Αναγνώρισης Προτύπων

2.3.1. Αισθητήρες

Ως είσοδος ενός συστήματος αναγνώρισης προτύπων μπορεί να θεωρηθεί οποιοσδήποτε αναμεταδότης, όπως είναι μια κάμερα ή ένα σύνολο από μικρόφωνα. Το πρόβλημα που μπορεί να προκύψει εδώ έχει να κάνει με τα χαρακτηριστικά αυτών των αισθητήρων όπως: κατά πόσο είναι καλή ή κακή η ανάλυση της εικόνας, ο ρυθμός δειγματοληψίας κ.ά.

2.3.2. Τμηματοποίηση και Ομαδοποίηση

Έστω ότι ζητείται η εύρεση της μελωδίας σε κάποιο ηχητικό απόσπασμα. Οι νότες σε κάποια μελωδία μπορεί να είναι επικαλυπτόμενες οπότε το σύστημά θα έπρεπε με κάποιο τρόπο να μπορεί να καταλάβει που τελειώνει μια νότα και που αρχίζει η επόμενη. Το κάθε πρότυπο πρέπει να τμηματοποιηθεί. Αν είχε αναγνωριστεί ήδη η κάθε νότα θα ήταν πιο εύκολο να τμηματοποιηθεί ο ήχος της. Το ζήτημα είναι πως γίνεται να τμηματοποιηθούν οι εικόνες πριν αυτές να έχουν κατηγοριοποιηθεί ή πώς να κατηγοριοποιηθούν αν αυτές δεν έχουν προηγουμένως τμηματοποιηθεί.

2.3.3. Εξαγωγή Χαρακτηριστικών

Η διαφορά ανάμεσα στην εξαγωγή χαρακτηριστικών και την ταξινόμηση δεν είναι ξεκαθαρισμένη. Από την μία, η έξοδος ενός ιδανικού εξαγωγέα χαρακτηριστικών θα πρέπει να είναι σε τέτοια μορφή ώστε να χρειάζεται ελάχιστη επεξεργασία από τον ταξινομητή προκειμένου να γίνει η ταξινόμηση. Από την άλλη, ένας ταξινομητής μπορεί να χαρακτηριστεί ως ιδανικός όταν δεν είναι απαραίτητη η βοήθεια ενός περίπλοκου εξαγωγέα.

Η αναγνώριση προτύπων βασίζεται στο γεγονός ότι χαρακτηριστικά τα οποία εξάγονται από αντικείμενα που βρίσκονται στην ίδια κατηγορία είναι παρόμοια ενώ χαρακτηριστικά τα οποία εξάγονται από αντικείμενα που βρίσκονται σε διαφορετικές κατηγορίες είναι διαφορετικά. Επίσης, τα χαρακτηριστικά γνωρίσματα που εξάγονται θα πρέπει να μένουν ανεπηρέαστα από παράγοντες που δεν αφορούν την ταξινόμηση των αντικειμένων. Στο παράδειγμά της εξαγωγής μελωδίας, ποιο όργανο έπαιξε μια συγκεκριμένη νότα είναι άσχετο με την κατηγορία στην οποία αυτή ανήκει και έτσι πρέπει και η αναπαράσταση που θα επιλέγει να είναι ανεξάρτητη από αυτά τα χαρακτηριστικά. Επειδή η καταγωγή της νότας είναι και αυτή άσχετη της ταξινόμησης, θα ήταν επίσης σωστό να είναι και τα χαρακτηριστικά ανεξάρτητα από το όργανο που τα παίζει. Τελικά, ακόμη και η ένταση της νότας μπορεί να μην είναι σημαντική. Έτσι, θα ήταν επιθυμητό τα χαρακτηριστικά της νότας που θα χρησιμοποιηθούν για την ταξινόμηση να παραμένουν αμετάβλητα όταν αλλάζουν στοιχεία όπως η διάρκεια, η ένταση ή το είδος του οργάνου.

2.3.4. Ταξινόμηση

Η ταξινόμηση δεν είναι καθόλου εύκολη διαδικασία. Οι τιμές των χαρακτηριστικών γνωρισμάτων των αντικειμένων τα οποία βρίσκονται είτε στην ίδια είτε σε διαφορετική κατηγορία μπορεί να μεταβάλλονται. Το μέγεθος δυσκολίας της ταξινόμησης έχει να κάνει με αυτή ακριβώς την αλλαγή των τιμών. Μια πιθανή αιτία της αλλαγής των τιμών των χαρακτηριστικών των αντικειμένων τα οποία βρίσκονται στην ίδια κατηγορία μπορεί να είναι είτε η πολυπλοκότητα είτε ο θόρυβος. Ο θόρυβος αποτελεί το μεγαλύτερο πρόβλημα κατά τη διαδικασία της αναγνώρισης προτύπων. Η πρόκληση που προκύπτει σε αυτό το σημείο είναι να μπορέσει να σχεδιαστεί ένας ταξινομητής ο οποίος να μπορεί να ανταποκριθεί με ακρίβεια σε αυτές τις αλλαγές των τιμών αλλά και να έχει και την καλύτερη δυνατή απόδοση.

Το πρόβλημα που προκύπτει εδώ είναι ότι δεν υπάρχει πάντα η δυνατότητα του προσδιορισμού των τιμών όλων των χαρακτηριστικών για μια δεδομένη είσοδο. Στο σύστημά για την αναγνώριση της μελωδίας, για παράδειγμα, μπορεί να μην είναι δυνατό να καθοριστεί κάποια νότα επειδή επικαλύπτεται από κάποια άλλη. Το ζήτημα είναι πώς γίνεται ο ταξινομητής να αντιμετωπίσει αυτό το πρόβλημα της απώλειας κατά την ταξινόμηση. Το να υποθέσει κανείς ότι το χαρακτηριστικό το οποίο ενδεχομένως επικαλύπτεται έχει τιμή ίση με το μηδέν ή ότι η τιμή του είναι ίση με τη μέση τιμή των άλλων τιμών δεν οδηγεί σε σωστά συμπεράσματα.

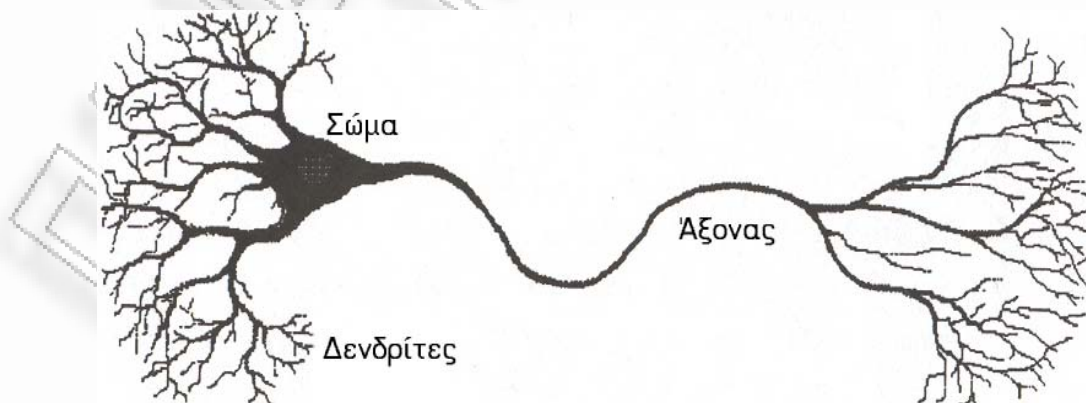
3. Εισαγωγή Στα Τεχνητά Νευρωνικά Δίκτυα

Ένα τεχνητό νευρωνικό δίκτυο (ΤΝΔ) είναι ένα μαθηματικό μοντέλο επεξεργασίας πληροφορίας του οποίου η λειτουργία είναι εμπνευσμένη από τον τρόπο με τον οποίο βιολογικοί νευρώνες, όπως αυτοί του ανθρώπινου εγκεφάλου, επεξεργάζονται πληροφορία. Το βασικό στοιχείο ενός νευρωνικού δικτύου είναι ο νευρώνας. Ένα ΤΝΔ αποτελείται από ένα μεγάλο αριθμό διασυνδεδεμένων στοιχείων επεξεργασίας (νευρώνες) τα οποία εργάζονται από κοινού για την επίλυση συγκεκριμένων προβλημάτων. Τα ΤΝΔ, όπως και οι άνθρωποι, μαθαίνουν από παραδείγματα και ρυθμίζονται προκειμένου να μπορούν να χρησιμοποιηθούν σε συγκεκριμένες εφαρμογές, όπως είναι η αναγνώριση προτύπων ή η ταξινόμηση δεδομένων, μέσα από μια διαδικασία εκπαίδευσης. Η εκπαίδευση στα βιολογικά συστήματα συνεπάγεται αναπροσαρμογές των συναπτικών συνδέσεων που υπάρχουν μεταξύ των νευρώνων. Το ίδιο ισχύει και για τα ΤΝΔ, όπου μέσω της διαδικασίας εκπαίδευσης ρυθμίζονται οι διασυνδέσεις των νευρώνων (βάρη).

3.1. Από τα Βιολογικά στα Τεχνητά Νευρωνικά Δίκτυα

Δεν είναι ακόμη πλήρως γνωστό πώς ο εγκέφαλος εκπαιδεύεται έτσι ώστε να επεξεργάζεται τις πληροφορίες. Στον ανθρώπινο εγκέφαλο, ένας τυπικός νευρώνας συλλέγει σήματα από άλλους νευρώνες μέσα από μια σειρά από εκλεπτυσμένες δομές που ονομάζονται δενδρίτες. Ο νευρώνας αποτελείται από ένα μακρύ και λεπτό άξονα ο οποίος χωρίζεται σε χιλιάδες διακλαδώσεις. Στο τέλος κάθε διακλάδωσης, μια δομή η οποία ονομάζεται σύναψη μετατρέπει την ενέργεια από τον άξονα ηλεκτρικούς παλμούς που είτε αναστέλλουν είτε διεγείρουν τους διασυνδεδεμένους νευρώνες. Όταν ένας νευρώνας δέχεται διεγερτικές εισόδους τότε στέλνει ένα παλμό ηλεκτρικής ενέργειας κατά μήκος του άξονα του, μεταδίδοντας τη διέγερση σε άλλους διασυνδεδεμένους νευρώνες. Η μάθηση προκύπτει με την αλλαγή της αποτελεσματικότητας των συνάψεων, έτσι ώστε η επιρροή ενός νευρώνα σε έναν άλλο να αλλάζει.

Συμπερασματικά λοιπόν οι νευρώνες αποτελούνται από 3 βασικά τμήματα, όπως φαίνεται και στο Σχήμα.2. Αυτά είναι: (α) το σώμα, (β) ο άξονας, και (γ) οι δενδρίτες.



Σχήμα 2: Απλουστευμένο μοντέλο ενός βιολογικού νευρώνα

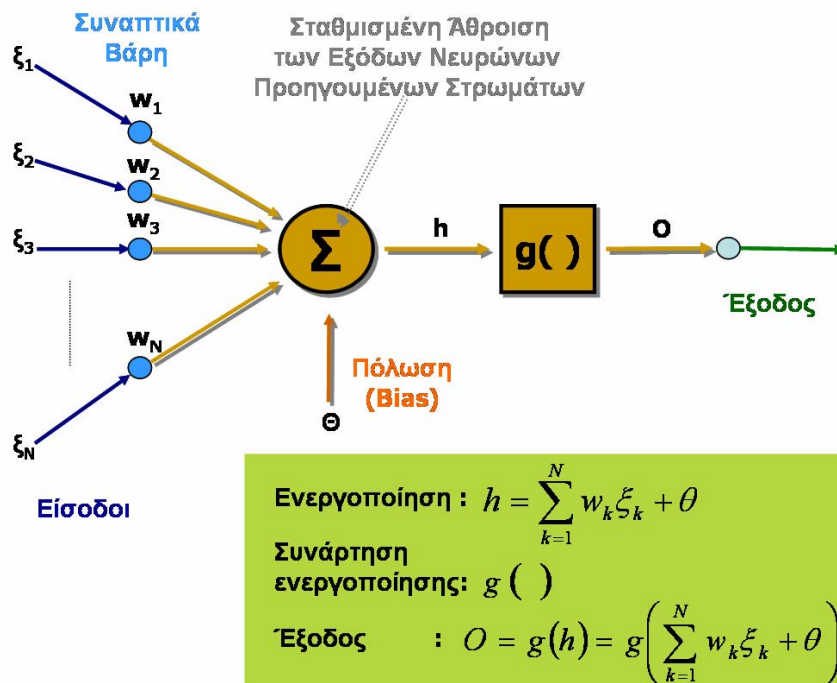
3.2. Εισαγωγή στους Τεχνητούς Νευρώνες

. Στην προηγούμενη ενότητα αναφέρθηκε πως στα βιολογικά νευρωνικά δίκτυα οι νευρώνες διασυνδέονται μεταξύ τους και μέσα από την εκπαίδευση γίνεται τροποποίηση του κατά πόσο η έξοδος του ενός επηρεάζει τον άλλο νευρώνα. Κατά αναλογία, στις συνδέσεις μεταξύ των νευρώνων σε ένα ΤΝΔ αντιστοιχεί ένας πραγματικός αριθμός, που ονομάζεται συντελεστής βάρους ή απλά βάρος και χρησιμοποιείται (όπως και οι συνάψεις μεταξύ των ανθρώπινων νευρώνων) για την τροποποίηση των εισόδων του νευρώνα.

Συνήθως, κάθε νευρώνας (εκτός από τους νευρώνες εισόδου) έχει ακόμα μια σύνδεση που ονομάζεται πόλωση (bias) και βοηθά το ΤΝΔ να έχει καλύτερη ικανότητα ταξινόμησης. Στα περισσότερα ΤΝΔ χρησιμοποιείται και η πόλωση και συνήθως όταν κανείς αναφέρεται στα βάρη υπονοεί και την πόλωση.

Η εμπειρία και η γνώση που αποκτά το ΤΝΔ αποθηκεύεται στα βάρη του. Η εκπαίδευση του ΤΝΔ είναι η επαναληπτική διαδικασία κατά την οποία γίνεται συνεχής αναπροσαρμογή των τιμών των βαρών έτσι ώστε το ΤΝΔ να έχει την επιθυμητή απόκριση και γίνεται με τη βοήθεια διαφόρων αλγορίθμων εκπαίδευσης.

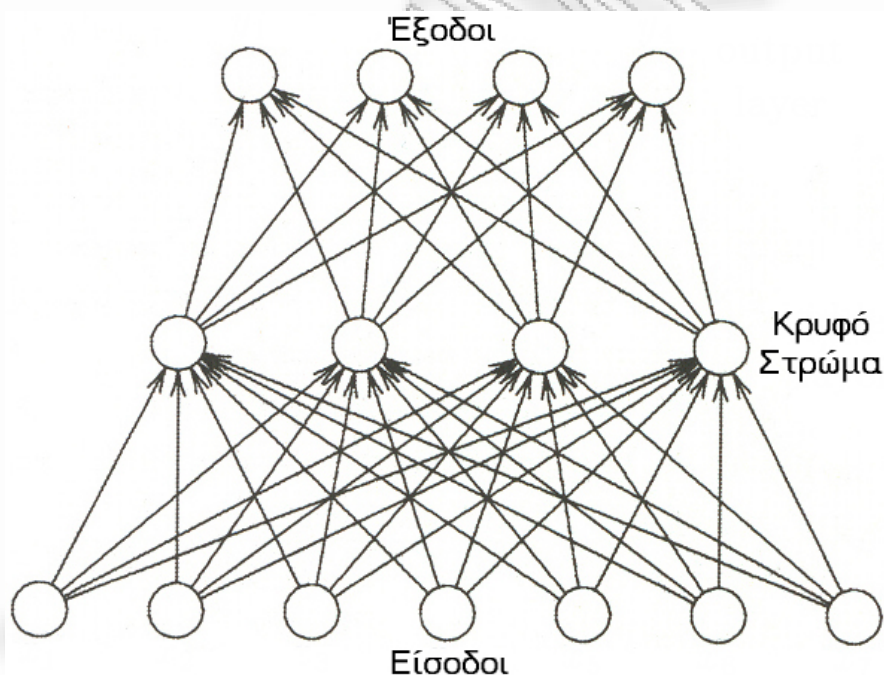
Στο σχήμα 3 φαίνονται τα δομικά στοιχεία ενός νευρώνα. Τονίζεται ότι κάθε νευρώνας στέλνει μόνο ένα σήμα κάθε φορά στους γειτονικούς του νευρώνες. Τα περισσότερα ΤΝΔ για να υπολογίσουν την ενεργοποίησή τους υπολογίζουν το γινόμενο κάθε εισόδου επί το αντίστοιχο βάρος και αθροίζουν όλα αυτά τα γινόμενα. Η είσοδος του νευρώνα είναι ένα διάνυσμα. Σε κάθε στοιχείο του διανύσματος αντιστοιχεί ένα βάρος. Η σταθμισμένη άθροιση των στοιχείων της εισόδου δίνεται ως είσοδος σε μια συνάρτηση μεταφοράς (ή συνάρτηση ενεργοποίησης). Η έξοδος της συνάρτησης είναι η έξοδος του νευρώνα.



Σχήμα 3 Νευρώνας

Ο τρόπος με τον οποίο είναι διασυνδεδεμένοι οι νευρώνες ενός δικτύου ονομάζεται τοπολογία ή αρχιτεκτονική του ΤΝΔ. Οι νευρώνες είναι λογικά χωρισμένοι σε επίπεδα, με τους νευρώνες του ίδιου επιπέδου να έχουν την ίδια συνάρτηση μεταφοράς και συνεπώς να έχουν παρόμοια συμπεριφορά. Σε κάθε ΤΝΔ αντιστοιχούν τουλάχιστον δύο στρώματα: το πρώτο στρώμα που ονομάζεται ως στρώμα εισόδου και έχει ως είσοδο την είσοδο του ΤΝΔ και το στρώμα εξόδου από το οποίο λαμβάνεται η έξοδος του δικτύου. Αν ένα ΤΝΔ δεν αποτελείται μόνο από το στρώμα εισόδου και το στρώμα εξόδου, τότε τα υπόλοιπα στρώματά του ονομάζονται κρυφά στρώματα (hidden layers) και το ΤΝΔ λέγεται πολυστρωματικό (multilayer). Όσο περισσότερα στρώματα έχει ένα νευρωνικό δίκτυο τόσο πιο δύσκολο είναι να εκπαιδευτεί κατάλληλα, αλλά από την άλλη έχει και περισσότερες δυνατότητες προσέγγισης της επιθυμητής απόκρισης.

Η αρχιτεκτονική ενός ΤΝΔ εξαρτάται και από τον τρόπο διάδοσης των πληροφοριών μεταξύ των νευρώνων. Υπάρχουν διάφορα είδη δικτύων όπως τα ΤΝΔ πρόσθιας τροφοδότησης (feedforward) τα ΤΝΔ οπίσθιας τροφοδότησης (back propagation) τα οποία θα παρουσιαστούν σε επόμενη ενότητα. Τα ΤΝΔ πρόσθιας τροφοδότησης για παράδειγμα είναι δίκτυα στα οποία το σήμα διαδίδεται έτσι ώστε να μην υπάρχει νευρώνας που η έξοδός του είναι είσοδος κάποιου νευρώνα του ίδιου ή προηγούμενου στρώματος. Επίσης, στα ΤΝΔ συνήθως κάθε νευρώνας ενός στρώματος, διασυνδέεται με όλους τους νευρώνες του επόμενου στρώματος (fully interconnected). Στο επόμενο σχήμα παρουσιάζεται πολυστρωματικό πλήρως διασυνδεδεμένο πρόσθιας τροφοδοσίας ΤΝΔ.



Σχήμα 4: Ένα πολυστρωματικό πλήρως διασυνδεδεμένο πρόσθιας τροφοδοσίας ΤΝΔ

Η συμπεριφορά ενός ΤΝΔ, δηλαδή η απόκριση σε συγκεκριμένες εισόδους εξαρτάται από: (α) την αρχιτεκτονική του, (β) τις συναρτήσεις ενεργοποίησης που χρησιμοποιεί, και (γ) τον αλγόριθμο εκπαίδευσης και τα δεδομένα εκπαίδευσης που χρησιμοποιήθηκαν για την εκπαίδευσή του.

3.3. Ιστορική Αναδρομή

Οι McCulloch και Pitts (1943) ανέπτυξαν μοντέλα νευρωνικών δικτύων βασισμένοι στην αντίληψη που είχαν για την νευρολογία. Τα μοντέλα αυτά παρήγαγαν αρκετές υποθέσεις σχετικά με το πώς λειτουργούν οι νευρώνες. Τα δίκτυά τους βασίζονταν σε απλούς νευρώνες που θεωρούνταν ότι είναι δυαδικές διατάξεις με σταθερά όρια. Τα αποτελέσματα των μοντέλων τους ήταν απλές λογικές συναρτήσεις όπως "α ή β" και "α και β". Μια άλλη προσπάθεια έγινε με τη χρήση προσομοιώσεων σε υπολογιστή από δύο ομάδες (τους Farley και Clark, 1954 και τους Rochester, Holland, Haibit και Duda, 1956). Η πρώτη ομάδα μάλιστα διατηρούσε στενή επαφή με νευροεπιστήμονες στο Πανεπιστήμιο McGill. Έτσι, κάθε φορά που τα μοντέλα τους δεν λειτουργούσαν, ζητούσαν τη γνώμη των νευροεπιστημόνων. Αυτή η αλληλεπίδραση καθιέρωσε μια τάση ανταλλαγής γνώσης και μεθοδολογιών ανάμεσα στις δυο αυτές επιστήμες η οποία συνεχίζεται μέχρι σήμερα [Sutt98].

Δε συνέβαλε μονάχα η επιστήμη της νευρολογίας στην ανάπτυξη των νευρωνικών δικτύων, αλλά και οι ψυχολόγοι και οι μηχανικοί συνέβαλαν επίσης στην εξέλιξη των προσομοιώσεων των νευρωνικών δικτύων. Ο Rosenblatt (1958) έδειξε μεγάλο ενδιαφέρον σε αυτόν τον τομέα, όταν ο ίδιος σχεδίασε και ανέπτυξε το νευρωνικό δίκτυο γνωστό και ως "Perceptron". Το "Perceptron" αποτελούνταν από τρία στρώματα όπου το μεσαίο στρώμα ήταν γνωστό ως διασυνδεδεμένο στρώμα. Αυτό το σύστημα μπορούσε να μάθει να επικοινωνεί ή να συνδέει μια δεδομένη είσοδο με μια τυχαία μονάδα εξόδου [Rose58].

Ένα άλλο σύστημα ήταν το ADALINE (ADaptive LInear Element - Προσαρμοστικό Γραμμικό Στοιχείο) [ADAL], το οποίο αναπτύχθηκε το 1960 από τους Widrow και Hoff. Η ADALINE ήταν μια αναλογική ηλεκτρονική συσκευή φτιαγμένη από απλά εξαρτήματα. Η μέθοδος που χρησιμοποιούσε για την εκμάθηση ήταν διαφορετική από εκείνη που χρησιμοποιούσε το Perceptron. Συγκεκριμένα εφαρμόζε τον κανόνα των ελάχιστων μέσων τετραγώνων (LMS) [Widr90].

Το 1969 οι Minsky και Papert [Mins69] έγραψαν ένα βιβλίο στο οποίο γενίκευαν τους περιορισμούς που παρουσίαζαν τα Perceptrons σε πολυστρωματικά συστήματα. Η δημοσίευση αυτή είχε τόσο σημαντικό αντίκτυπο στην ερευνητική κοινότητα ώστε οδήγησε στην κατάργηση της χρηματοδότησης της έρευνας σχετικής με τα νευρωνικά δίκτυα.

Το 1974 ο Paul Werbos [Werb74] [Werb90] ανέπτυξε και χρησιμοποίησε τη μέθοδο της οπίσθιας τροφοδότησης (back-propagation). Εντούτοις πέρασαν αρκετά χρόνια μέχρι να διαδοθεί αυτή η προσέγγιση. Τα δίκτυα οπίσθιας τροφοδότησης είναι ίσως η πιο γνωστή εφαρμογή των νευρωνικών δικτύων σήμερα. Στο εντωμεταξύ, ο Fukushima (F. Kunihiko) έκανε ένα βήμα παραπάνω εκπαιδύοντας πολυστρωματικά νευρωνικά δίκτυα για την ερμηνεία χειρόγραφων χαρακτήρων. Το αρχικό δίκτυο δημοσιεύθηκε το 1975 και ονομαζόταν "Cognitron" [Fuku75].

Η πρόοδος κατά τα τέλη του 1970 και στις αρχές της δεκαετίας του 1980 ήταν σημαντική για την επανεμφάνιση του ενδιαφέροντος στον τομέα των νευρωνικών δικτύων. Διάφοροι παράγοντες έπαιξαν ρόλο σε αυτό. Σημαντικό ρόλο έπαιξε και το γεγονός ότι η ανταπόκριση στα συνέδρια και στις δημοσιεύσεις που σχετίζονταν με νευρωνικά δίκτυα ήταν αρκετά θετική. Τα μέσα ενημέρωσης έδειξαν ενδιαφέρον για αυτή την αυξημένη δραστηριότητα και βοήθησαν στη διάδοση της συγκεκριμένης τεχνολογίας. Εμφανίστηκαν τα πρώτα Ακαδημαϊκά προγράμματα και τα πρώτα μαθήματα μπήκαν στα περισσότερα μεγάλα Πανεπιστήμια (σε ΗΠΑ και Ευρώπη).

Στις μέρες μας έχει σημειωθεί σημαντική πρόοδος στον τομέα των νευρωνικών δικτύων, αρκετή ώστε να προσελκύσει την μεγάλη προσοχή και τη χρηματοδότηση για περαιτέρω έρευνα. Τα νευρωνικά δίκτυα διδάσκονται πλέον στα περισσότερα πανεπιστήμια. Η πρόοδος πέρα από τις σημερινές εμπορικές εφαρμογές φαίνεται να είναι δυνατή και η έρευνα προχωράει σε πολλά μέτωπα

3.4. Αρχιτεκτονικές των ΤΝΔ

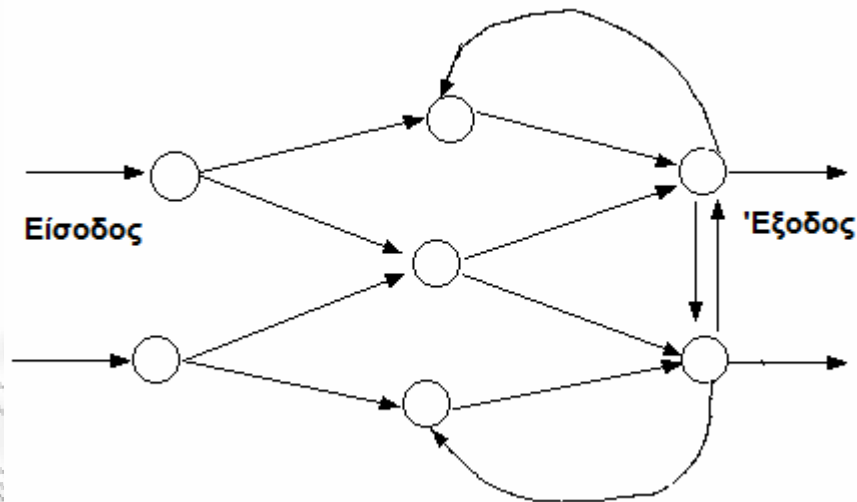
Στη συνέχεια θα γίνει παρουσίαση των διαφόρων αρχιτεκτονικών που χρησιμοποιούνται στα ΤΝΔ. Η αρχιτεκτονική ενός ΤΝΔ εξαρτάται και από τον τρόπο διάδοσης των πληροφοριών μεταξύ των νευρώνων.

3.4.1. ΤΝΔ πρόσθιας τροφοδότησης

Τα ΤΝΔ πρόσθιας τροφοδότησης για είναι δίκτυα στα οποία το σήμα διαδίδεται έτσι ώστε να μην υπάρχει νευρώνας που η έξοδός του είναι είσοδος κάποιου νευρώνα του ίδιου ή προηγούμενου στρώματος. Με άλλα λόγια, το σήμα μεταφέρεται μόνο προς μια κατεύθυνση, από την είσοδο του νευρωνικού προς την έξοδο. Δεν υπάρχουν βρόγχοι ανατροφοδότησης και η έξοδος ενός στρώματος επηρεάζει μονάχα τα επόμενα στρώματα.

3.4.2. ΤΝΔ με ανατροφοδότηση

Στα δίκτυα με ανατροφοδότηση τα σήματα ταξιδεύουν και προς τις δυο κατευθύνσεις καθώς υπάρχουν βρόγχοι στο δίκτυο. Τα δίκτυα αυτά είναι εν γένει πολύ ισχυρά και ταυτόχρονα ιδιαίτερα πολύπλοκα. Τα δίκτυα με ανατροφοδότηση είναι δυναμικά, η κατάσταση τους αλλάζει συνεχώς μέχρι να φτάσουν σε μια κατάσταση ισορροπίας, στην οποία παραμένουν μέχρι να αλλάξει η είσοδος και να βρεθεί καινούριο σημείο ισορροπίας



Σχήμα 5: Ένα πολυστρωματικό ΤΝΔ με ανατροφοδότηση

3.4.3. Perceptrons

Η έρευνα με την μεγαλύτερη επιρροή στα νευρωνικά δίκτυα την δεκαετία του 1960 συνδέθηκε με τον τίτλο «perceptrons», έναν όρος που επινοήθηκε από τον Frank Rosenblatt [rose57]. Το perceptron είναι ένα μοντέλο ΤΝΔ με νευρώνες με σταθμισμένες εισόδους και με κάποια πρόσθετη, σταθερή προεπεξεργασία. Τα Perceptrons μιμούνται τη βασική ιδέα

πίσω από το οπτικό σύστημα των θηλαστικών και χρησιμοποιήθηκαν κυρίως στην αναγνώριση προτύπων. Είναι δυαδικοί ταξινομητές οι οποίοι αντιστοιχίζουν την είσοδο x (ένα διάνυσμα με πραγματικές τιμές) σε μια τιμή εξόδου $f(x)$ (μία δυαδική τιμή) τέτοια ώστε:

$$f(x) = \begin{cases} 1, & \text{αν } w \cdot x + b > 0 \\ 0, & \text{διαφορετικά} \end{cases}$$

Το perceptron θεωρείται ως η απλούστερη μορφή δικτύου πρόσθιας τροφοδότησης. Το 1969 οι Minsky and Papert [mins69] δημοσίευσαν τους περιορισμούς ενός perceptron με ένα στρώμα. Ο αντίκτυπος που είχε η δημοσίευση αυτή ήταν τεράστιος και προκάλεσε πολλούς ερευνητές να χάσουν το ενδιαφέρον τους στα ΤΝΔ. Οι Minsky and Papert απέδειξαν μαθηματικά ότι perceptrons με ένα στρώμα δεν μπορούν να κάνουν κάποιες βασικές λειτουργίες αναγνώρισης προτύπων, όπως ο προσδιορισμός αν ένα σχήμα είναι συνδεδεμένο ή όχι. Στη συνέχεια αποδείχθηκε ότι πολυστρωματικά perceptron με την κατάλληλη εκπαίδευση μπορούν να κάνουν με επιτυχία τέτοιου είδους λειτουργίες [Gall90].

3.5. Πλεονεκτήματα των Νευρωνικών Δικτύων

Το επιστημονικό ενδιαφέρον για τα ΤΝΔ προκύπτει κυρίως από τη δυνατότητά τους να επιλύουν δύσκολα και ενδιαφέροντα υπολογιστικά προβλήματα του πραγματικού κόσμου. Η χρήση των ΤΝΔ προσφέρει τις ακόλουθες πολύ χρήσιμες ιδιότητες και δυνατότητες.

- **Αντικατάσταση μαθηματικού μοντέλου.** Η αξία των ΤΝΔ έγκειται κυρίως στο γεγονός ότι η χρήση τους δε συνεπάγεται τη δημιουργία κάποιου μοντέλου για το περιβάλλον το οποίο καλούνται να προσομοιώσουν. Όταν η περιγραφή του περιβάλλοντος είναι ιδιαίτερα πολύπλοκη για να περιγραφεί με κάποιο μαθηματικό μοντέλο, τότε η χρήση των ΤΝΔ είναι μια καλή λύση. Κατά την εκπαίδευση δίνονται στο ΤΝΔ πρότυπα εισόδου/εξόδου (δείγματα εισόδων και εξόδων της συνάρτησης που καλείται το ΤΝΔ να προσομοιώσει). Στόχος της διαδικασίας εκπαίδευσης είναι να φτάσει το ΤΝΔ σε μια τέτοια κατάσταση όπου για κάθε πρότυπο εκπαίδευσης, η έξοδός του να ταυτίζεται με την επιθυμητή έξοδο. Έτσι δημιουργείται μια συσχέτιση μεταξύ των δεδομένων εισόδου και εξόδου, χωρίς όμως τη χρήση κάποιου προκαθορισμένου στατιστικού ή άλλου μοντέλου.
- **Προσαρμογή.** Τα ΤΝΔ έχουν την ικανότητα να μεταβάλλουν την απόκρισή τους μεταβάλλοντας ελαφρώς τα βάρη των νευρώνων, ανάλογα με τα πρότυπα εισόδου. Συνεπώς, σε περιβάλλοντα τα οποία αλλάζουν διαρκώς, τα ΤΝΔ μπορούν να εφαρμοστούν εφόσον υπόκεινται περιοδικά σε εκπαίδευση με τα επικαιροποιημένα πρότυπα εισόδου/εξόδου.
- **Δυνατότητα εύκολης υλοποίησης σε Hardware.** Η απλή δομή των νευρώνων και το γεγονός ότι ένα ΤΝΔ αποτελείται από πολλά ίδια στοιχεία (νευρώνες) καθιστά σχετικά απλή την υλοποίηση των ΤΝΔ σε hardware. Με τη χρήση της τεχνολογίας ολοκλήρωσης πολύ μεγάλης κλίμακας (Very Large Scale Integration – VLSI) που είναι ιδιαίτερα διαδομένη για την κατασκευή ηλεκτρονικών πλακετών, μπορούν να δημιουργηθούν εύκολα ταχύτατα ΤΝΔ.
- **Ανεκτικότητα σε σφάλματα.** Τα ΤΝΔ που έχουν υλοποιηθεί σε υλικό (hardware) έχουν την ιδιότητα της ανεκτικότητας σε σφάλματα, γιατί η απόδοση του συστήματος μειώνεται ομαλά σε περίπτωση λάθους. Ακόμα και αν καταστραφεί ένας νευρώνας, το ΤΝΔ θα συνεχίσει να λειτουργεί με κάπως μειωμένη απόδοση.
- **Ανοχή στο θόρυβο.** Τα ΤΝΔ παρουσιάζουν ανοχή στο θόρυβο υπό την έννοια ότι αν εκπαιδευτούν για την αναγνώριση προτύπων, (για παράδειγμα σε εφαρμογές με

εικόνες) τότε είναι σε γενικές γραμμές εφικτή η ταξινόμηση από το νευρωνικό ακόμα και όταν η είσοδος έχει υποστεί κάποια αλλοίωση (θόρυβο).

3.6. Εφαρμογές των ΤΝΔ

Στη συνέχεια παρουσιάζονται μερικές από τις βασικές εφαρμογές των ΤΝΔ σε διάφορους τομείς της Επιστήμης και της Τεχνολογίας. Οι περισσότερες από αυτές τις εφαρμογές έχουν ήδη υλοποιηθεί και πολλές από αυτές αποτελούν εμπορικά προϊόντα.

- Αναγνώριση των ομιλητών στις επικοινωνίες.
- Διάγνωση της ηπατίτιδας
- Ανάκτηση των τηλεπικοινωνιών από ελαττωματικό λογισμικό.
- Ερμηνεία των κινεζικών λέξεων
- Ανίχνευση υποθαλάσσιων ναρκών
- Ανάλυση σύστασης
- Αναγνώριση τρισδιάστατων αντικειμένων
- Αναγνώριση χειρόγραφων λέξεων
- Αναγνώριση προσώπου.

Στη συνέχεια παρουσιάζονται μερικά παραδείγματα για το πώς τα ΤΝΔ μπορούν να εφαρμοστούν σε διαφορετικές επιστήμες για την Αναγνώριση Προτύπων.

- **Ιατρική:** Τα νευρωνικά δίκτυα είναι ιδανικά στην αναγνώριση ασθενειών δεδομένου ότι δεν υπάρχει ανάγκη να δημιουργηθεί κάποιο μαθηματικό μοντέλο για την εύρεση της νόσου. Τα Νευρωνικά δίκτυα εκπαιδεύονται από δείγματα που είναι αντιπροσωπευτικά του συνόλου των παραλλαγών της νόσου. Κάθε δείγμα θα πρέπει να περιέχει αποτελέσματα εξετάσεων και δείκτες που θεωρούνται ότι συσχετίζονται με τη νόσο που πρέπει να ανιχνευθεί. Τα δείγματα θα πρέπει να επιλεγούν πολύ προσεκτικά, αν θέλει κανείς να έχει ένα αποτελεσματικό και αξιόπιστο σύστημα.
- **Marketing:** Η “Airline Marketing Tactician” (εμπορικό σήμα γνωστό ως AMT) είναι ένα πληροφοριακό σύστημα που περιλαμβάνει διάφορες τεχνολογίες, όπως τα έμπειρα συστήματα (expert systems) και τα ΤΝΔ. Στο AMT έχει ενσωματωθεί ένα ΤΝΔ το οποίο χρησιμοποιείται για τον έλεγχο της χορήγησης αεροπορικών θέσεων. Το σύστημα χρησιμοποιείται για να παρακολουθεί τη διαθεσιμότητα των θέσεων και να προτείνει συμβουλές χορήγησης θέσεων για κάθε πτήση. Οι πληροφορίες αποδείχθηκε ότι έχουν άμεσο αντίκτυπο στην αποδοτικότητα της διαδικασίας δέσμευσης και χορήγησης θέσεων και μπορούν να προσφέρουν τεχνολογικό πλεονέκτημα στην αεροπορική εταιρία που χρησιμοποιεί το AMT [Hute87].

3.7. Μέθοδοι εκπαίδευσης

Όλες οι μέθοδοι εκπαίδευσης που χρησιμοποιούνται στα νευρωνικά δίκτυα μπορούν να ταξινομηθούν σε δύο μεγάλες κατηγορίες:

- Εκπαίδευση με επίβλεψη η οποία απαιτεί την εμφάνιση ενός «δασκάλου», έτσι ώστε το δίκτυο να μαθαίνει ποια είναι η επιθυμητή έξοδος για κάθε είσοδο. Κατά τη διαδικασία εκπαίδευσης μπορεί να απαιτείται γενικευμένη πληροφορία. Παραδείγματα εκπαίδευσης με επίβλεψη αποτελούν η εκπαίδευση διόρθωσης σφάλματος, η εκπαίδευση με ενίσχυση και η στοχαστική εκπαίδευση. Ένα σημαντικό ζήτημα σχετικά με την εκπαίδευση με επίβλεψη είναι το πρόβλημα της σύγκλισης λάθους, δηλαδή η ελαχιστοποίηση του σφάλματος μεταξύ της επιθυμητής και της υπολογιζόμενης τιμής. Ο στόχος είναι να καθοριστεί ένα σύνολο βαρών το οποίο ελαχιστοποιεί το σφάλμα. Μία γνωστή μέθοδος, η οποία είναι κοινή για πολλά παραδείγματα εκπαίδευσης είναι η σύγκλιση του ελάχιστου των μέσων τετραγώνων (LMS).
- Η εκπαίδευση χωρίς επίβλεψη δεν χρησιμοποιεί κάποιον «δάσκαλο» και βασίζεται μόνο σε τοπικές πληροφορίες. Αναφέρεται επίσης ως αυτο-οργάνωση, με την έννοια ότι αυτο-οργανώνει τα δεδομένα που παρουσιάζονται στο δίκτυο και εντοπίζει τις προκύπτουσες συλλογικές ιδιότητες τους. Ένα νευρωνικό δίκτυο εκπαιδεύεται off-line αν η φάση εκπαίδευσης και η φάση λειτουργίας είναι διαφορετικές. Ένα νευρωνικό δίκτυο εκπαιδεύεται on-line, αν η εκπαίδευση και η λειτουργία γίνονται ταυτόχρονα. Συνήθως, η εκπαίδευση με επίβλεψη γίνεται off-line, ενώ η εκπαίδευση χωρίς επίβλεψη γίνεται on-line.

4. Τεχνικές ταξινόμησης μελωδίας σε μουσικά είδη

Στο κεφάλαιο αυτό θα περιγραφούν υπάρχουσες τεχνικές για την αυτόματη ταξινόμηση μελωδίας σε μουσικά είδη. Τα είδη μουσικής είναι κατηγορίες που έχουν προκύψει μέσω μιας σύνθετης αλληλεπίδρασης των πολιτισμών, των καλλιτεχνών και της δυναμικής των αγορών, προκειμένου να χαρακτηριστούν οι ομοιότητες μεταξύ των μουσικών ή των συνθέσεων και να οργανωθούν οι συλλογές μουσικής. Τα όρια μεταξύ των διαφόρων ειδών παραμένουν ακόμα συγκεχυμένα, συνεπώς ο καθορισμός τους καθιστά το πρόβλημα της αυτόματης ταξινόμησης έναν μη τετριμμένο στόχο.

Το πρόβλημα ταξινόμησης σε μουσικά είδη απαιτεί μια κατηγοριοποίηση των ειδών, δηλαδή ένα ιεραρχικό σύνολο κατηγοριών χαρτογραφημένο σε μια συλλογή μουσικής. Οι Pachet και Cazaly [Pach00] μελέτησαν διάφορες μουσικές ταξινομήσεις σε μουσικά είδη που χρησιμοποιήθηκαν στη βιομηχανία και στο διαδίκτυο και έδειξαν ότι δεν είναι απλό να κατασκευαστεί μια τέτοια ιεραρχία. Δεδομένου ότι μια καλή ταξινόμηση στηρίζεται σε μια προσεκτική κατηγοριοποίηση, παρουσιάζονται στη συνέχεια μια σειρά από διάφορα θέματα που αφορούν τη διαδικασία κατηγοριοποίησης. Τα θέματα αυτά παρουσιάζονται στη δημοσίευση [Scar06].

4.1. Είδη μουσικής

Η ταξινόμηση στα διάφορα είδη μουσικής μπορεί να γίνει με βάση διάφορα κριτήρια, τα οποία παρατίθενται στη συνέχεια.

A. Ταξινόμηση με βάση τον καλλιτέχνη, το άλμπουμ ή τον τίτλο

Μια βασική ερώτηση που τίθεται είναι εάν η ταξινόμηση σε μουσικά είδη πρέπει να εφαρμόζεται στον καλλιτέχνη, το άλμπουμ ή τον τίτλο. Εάν υποθέσει κανείς ότι ένα τραγούδι μπορεί να ταξινομηθεί σε ένα μόνο είδος (γεγονός που είναι ήδη αμφισβητήσιμο), δεν είναι το ίδιο απλό πλέον για ένα άλμπουμ, το οποίο μπορεί να περιέχει ετερογενές υλικό. Το ίδιο ισχύει και για τους καλλιτέχνες, μερικοί από τους οποίους έχουν καλύψει ένα ευρύ φάσμα διαφόρων ειδών κατά τη διάρκεια της καριέρας τους με αποτέλεσμα να μην έχει πάρα πολύ νόημα να προσπαθήσει κανείς να εντάξει αυτά τα κομμάτια σε μια συγκεκριμένη κατηγορία.

B. Ασυμφωνία στις ταξινομήσεις

Οι Pachet και Cazaly [Pach00] έδειξαν ότι δεν υπάρχει μια γενικά αποδεκτή διαδικασία στις ταξινομήσεις σε μουσικά είδη. Παίρνοντας για παράδειγμα γνωστούς ιστοχώρους όπως το Allmusic (531 είδη), το Amazon (719 είδη), και το Mp3 (430 είδη), βρέθηκαν μόνο 70 κοινοί όροι στις 3 ταξινομήσεις. Παρατηρήθηκε ότι μερικοί ευρέως χρησιμοποιημένοι όροι όπως το ροκ ή το ποπ αναφέρονται σε διαφορετικά σύνολα τραγουδιών ενώ παράλληλα αυτές οι ιεραρχίες των ειδών είναι διαφορετικά δομημένες από τη μια ταξινόμηση στην άλλη.

Γ. Καθορισμένες ετικέτες είδους

Εξετάζοντας περισσότερο μερικά συγκεκριμένα και ευρέως χρησιμοποιημένα μουσικά είδη, παρατηρεί κανείς πόσο διαφορετικά μπορεί να είναι τα κριτήρια που καθορίζουν το κάθε είδος. Στη συνέχεια παρουσιάζονται ορισμένα τέτοια παραδείγματα:

- Η «ινδική μουσική» είναι γεωγραφικός προσδιορισμός.

- Η «μουσική μπαρόκ» συσχετίζεται με μια συγκεκριμένη εποχή στην ιστορία (καλύπτοντας παράλληλα ένα ευρύ φάσμα από μουσικά στυλ και μια ευρεία γεωγραφική περιοχή)
- Το Post-rock είναι ένας όρος που επινοήθηκε από τον κριτικό μουσικής Simon Reynolds

Οι Pachet και Cazaly [Pach00] υποστηρίζουν ότι η σημασιολογική σύγχυση σε μια ενιαία ταξινόμηση μπορεί να οδηγήσει σε πλεονασμούς που μπορεί να μην προκαλούν σύγχυση στους χρήστες αλλά δύσκολα μπορούν να λυθούν από αυτόματα συστήματα. Επιπλέον, η ταξινόμηση είδους μπορεί να εξαρτάται από πολιτιστικές αναφορές. Για παράδειγμα, ένα τραγούδι από κάποιο Γάλλο τραγουδιστή θα μπορούσε να ταξινομηθεί στην κατηγορία “pop” στη Γαλλία, αλλά στην Αμερική μπορεί να αρχειοθετούταν ως παγκόσμια μουσική.

Δ. Εξελιξιμότητα της ταξινόμησης σε μουσικά είδη

Οι ιεραρχίες των ειδών πρέπει επίσης να εξετάζουν τη δυνατότητα προσθήκης νέων ειδών ώστε να λαμβάνουν υπόψη την εξέλιξη της μουσικής. Συχνά εμφανίζονται νέα είδη μουσικής τα οποία είναι αποτέλεσμα κάποιας συγχώνευσης διαφορετικών ειδών μουσικής (το Psychobilly μπορεί να θεωρηθεί ως συγχώνευση του Rockabilly και του πανκ) ή του διαχωρισμού ενός είδους σε υποείδος (το Hip-Hop έχει οδηγήσει σε διαφορετικά υποείδη όπως το Gangsta Rap, το Turntablism, και το Conscious Rap). Αυτό είναι ένα σημαντικό θέμα για τα αυτόματα συστήματα. Η προσθήκη νέων ειδών και υποειδών σε μια ταξινόμηση είναι εύκολη αλλά η δημιουργία ενός αυτόματου συστήματος που απαιτεί συνεχή εκπαίδευση ικανή ώστε να προσαρμόζεται μόνο του στις αλλαγές, είναι δύσκολη.

4.2. Εξαγωγή Χαρακτηριστικών Γνωρισμάτων

Στον κόσμο των ψηφιακών μέσων, οι ηχητικές πληροφορίες αντιπροσωπεύονται συνήθως από bits επιτρέποντας μια άμεση αναδημιουργία της αναλογικής κυματομορφής από την οποία προέρχονται. Μπορεί όμως να δεχθεί κανείς ότι η μουσική πληροφορία (π.χ. στην κοινή δυτική μουσική) μπορεί να περιγραφεί με περισσότερη ή λιγότερη ακρίβεια από μερικά υψηλότερου επιπέδου format βασισμένα σε συγκεκριμένα μοντέλα – όπως το MIDI ή το MusicXML.

Εντούτοις, στις εφαρμογές του πραγματικού κόσμου μια ακριβής συμβολική αντιπροσώπευση ενός (νέου) τραγουδιού είναι σπάνια διαθέσιμη και κάποιος πρέπει να χειριστεί την απλούστερη αυτή μορφή, όπως για παράδειγμα ένα ηχητικό δείγμα του κομματιού. Τα ηχητικά δείγματα, που λαμβάνονται με τη δειγματοληψία της ακριβούς κυματομορφής, δεν μπορούν να χρησιμοποιηθούν άμεσα από τα αυτόματα συστήματα ανάλυσης λόγω του χαμηλού επιπέδου και της χαμηλής «πυκνότητας» των πληροφοριών που περιέχουν. Από την άλλη, η αποθήκευση της ακριβούς κυματομορφής κοστίζει πολύ καθώς ο όγκος δεδομένων είναι τεράστιος. Συνεπώς, το πρώτο βήμα των συστημάτων ανάλυσης είναι να εξαχθούν μερικά χαρακτηριστικά γνωρίσματα από τα ηχητικά δεδομένα ώστε να είναι δυνατός ο χειρισμός των σημαντικότερων πληροφοριών και η μείωση της περαιτέρω επεξεργασίας.

Η εξαγωγή χαρακτηριστικών γνωρισμάτων είναι το πρώτο βήμα στα περισσότερα συστήματα αναγνώρισης προτύπων. Εφόσον εξαχθούν τα σημαντικά χαρακτηριστικά γνωρίσματα, μπορούν να χρησιμοποιηθούν διάφορες τεχνικές ταξινόμησης. Στην περίπτωση των ηχητικών σημάτων, τα χαρακτηριστικά γνωρίσματα μπορούν να αφορούν τις κύριες διαστάσεις της μουσικής όπως για παράδειγμα η μελωδία, η αρμονία, ο ρυθμός, η χροιά και άλλα.

4.2.1. Χροιά

Η χροιά ορίζεται στη βιβλιογραφία ως το γνώρισμα που κάνει δύο ήχους με τον ίδιο τόνο (pitch) και διαφορετική ένταση να ακούγονται διαφορετικά. Τα χαρακτηριστικά γνωρίσματα που περιγράφουν τη χροιά αφορούν κυρίως τη φασματική κατανομή του σήματος αν και μερικά απ' αυτά υπολογίζονται στο πεδίο του χρόνου. Αυτά τα χαρακτηριστικά γνωρίσματα είναι σφαιρικά υπό την έννοια ότι ενσωματώνουν πληροφορίες όλων των πηγών και των μουσικών οργάνων συγχρόνως.

Στη δημοσίευση [Peet04] υπάρχει ένας λεπτομερής κατάλογος χαρακτηριστικών γνωρισμάτων που χρησιμοποιούνται για να χαρακτηρίσουν τη χροιά των μουσικών οργάνων. Οι περισσότεροι από αυτούς τους περιγραφείς έχουν χρησιμοποιηθεί στα πλαίσια της αναγνώρισης του είδους μουσικής, αν και μερικά χαρακτηριστικά γνωρίσματα είναι περισσότερο προσαρμοσμένα για το χαρακτηρισμό μονοφωνικών μουσικών οργάνων παρά για πολυφωνικές μείξεις. Αυτοί οι περιγραφείς αναφέρονται συνήθως ως οντότητες χαμηλού επιπέδου και τυπικά υπολογίζονται για δείγμα σήματος μεταξύ 10 και 60 δευτερολέπτων. Μερικοί από αυτούς τους περιγραφείς έχουν κανονικοποιηθεί με βάση το πρότυπο preg-7 και ο αλγόριθμος εξαγωγής τους είναι κανονικοποιημένος.

Παρακάτω συνοψίζονται τα κύρια χαρακτηριστικά γνωρίσματα χαμηλού επιπέδου που χρησιμοποιούνται στις εφαρμογές χαρακτηρισμού μουσικού είδους:

- Χρονικά γνωρίσματα (Temporal features): χαρακτηριστικά γνωρίσματα που υπολογίζονται από το πλαίσιο των ηχητικών σημάτων (zero-crossing rate, γραμμικοί συντελεστές πρόβλεψης, κ.λπ.).
- Ενεργειακά χαρακτηριστικά γνωρίσματα (Energy features): χαρακτηριστικά γνωρίσματα που αναφέρονται στο ενεργειακό περιεχόμενο του σήματος (ενέργεια της μέσης τετραγωνικής τιμής του πλαισίου του σήματος, ενέργεια της αρμονικής συνιστώσας του φάσματος, ενέργεια του μέρους του φάσματος που αντιστοιχεί σε θόρυβο, και άλλα).
- Χαρακτηριστικά γνωρίσματα μορφής φάσματος (Spectral shape features): χαρακτηριστικά γνωρίσματα που περιγράφουν τη μορφή του φάσματος ενός πλαισίου σήματος: κέντρο βάρους (centroid), διάδοση (spread), καμπυλότητα (skewness), κύρτωση (kurtosis), κλίση, διακύμανση και άλλα.
- Χαρακτηριστικά γνωρίσματα που σχετίζονται με την ανθρώπινη αντίληψη (Perceptual features): τα χαρακτηριστικά γνωρίσματα υπολογίζονται χρησιμοποιώντας ένα πρότυπο της διαδικασίας ακοής του ανθρώπου (ηχηρότητα, οξύτητα, διάδοση).

Οι μετασχηματισμοί των χαρακτηριστικών γνωρισμάτων όπως οι πρώτες και δεύτερες παράγωγοι χρησιμοποιούνται συνήθως για να δημιουργήσουν νέα χαρακτηριστικά γνωρίσματα ή για να αυξήσουν τις διαστάσεις των διανυσμάτων των χαρακτηριστικών γνωρισμάτων.

Οι περισσότεροι από αυτούς τους περιγραφείς υπολογίζονται ανά τακτά χρονικά διαστήματα, σε σύντομα χρονικά παράθυρα τυπικού μήκους μεταξύ 10 έως 60 ms. Στα πλαίσια της ταξινόμησης, οι περιγραφείς χροιάς στη συνέχεια συχνά συνοψίζονται υπολογίζοντας στατιστικά μετρικά της κατανομής των χαρακτηριστικών γνωρισμάτων σε μεγαλύτερα παράθυρα γνωστά ως παράθυρα δομής (texture windows) [Tzan02] [Meng02]. Η μοντελοποίηση της χροιάς σε μεγαλύτερη χρονική κλίμακα όχι μόνο μειώνει το υπολογιστικό κόστος που απαιτείται αλλά είναι επίσης πιο κοντά στην ανθρώπινη αντίληψη εξαιτίας των

μικρών πλαισίων σήματος που χρησιμοποιούνται για τον υπολογισμό των χαρακτηριστικών γνωρισμάτων.

Ο αντίκτυπος του μεγέθους του παραθύρου υψής στην ακρίβεια της ταξινόμησης έχει μελετηθεί στο [Tzan02]. Αποδεικνύεται ότι πράγματι η χρήση ενός παραθύρου αυξάνει σημαντικά την ακρίβεια ταξινόμησης σε σύγκριση με την άμεση ανάλυση των πλαισίων. Το συμπέρασμα είναι ότι τα παράθυρα υψής ενός δευτερολέπτου είναι ένας καλός συμβιβασμός δεδομένου ότι κανένα σημαντικό κέρδος στην ακρίβεια ταξινόμησης δεν λαμβάνεται με τη λήψη μεγαλύτερων παραθύρων ενώ η ακρίβεια μειώνεται (σχεδόν γραμμικά) καθώς το παράθυρο μικραίνει.

Αντί για τη χρήση παραθύρων υψής με σταθερό μέγεθος και αυθαίρετες θέσεις, μερικοί ερευνητές προσπαθούν να συνδέσουν τα παράθυρα με πραγματικά μουσικά γεγονότα. Οι West και Cox [West05] διαιρούν το ηχητικό σήμα με έναν ανιχνευτή έναρξης (onset detector) ενώ οι Scaringella και Zoia [Scar05] χρησιμοποιούν ένα ανιχνευτή ρυθμού (beat tracker). Τα εξαγόμενα τμήματα χρησιμοποιούνται έπειτα ως συνηθισμένα παράθυρα υψής με τις πληροφορίες χροιάς να έχουν μεγαλύτερη συνοχή.

4.2.2. *Μελωδία και Αρμονία*

Η αρμονία μπορεί να οριστεί ως η χρήση και η μελέτη του τόνου (pitch) και των συγχορδιών. Αντίθετα, η μελωδία είναι μια διαδοχή των τονικών γεγονότων θεωρημένα ως μια ενιαία οντότητα. Η αρμονία αναφέρεται μερικές φορές ως κάθετο στοιχείο της μουσικής με τη μελωδία να αποτελεί το οριζόντιο στοιχείο. Η ανάλυση της μελωδίας και της αρμονίας που έχει χρησιμοποιηθεί για πολύ καιρό από μουσικολόγους στη μελέτη μουσικών δομών, προσπαθεί να ενσωματώσει μια τέτοια ανάλυση κατά την μοντελοποίηση του μουσικού είδους.

Στη δημοσίευση [Gome03] υπάρχει μια καλή επισκόπηση για την περιγραφή και την εξαγωγή της μελωδίας στα πλαίσια της επεξεργασίας ηχητικού περιεχομένου. Για την εκτίμηση των πολλαπλών θεμελιωδών συχνοτήτων ταυτόχρονων μουσικών ήχων, υπάρχει μια αναφορά στο [Klap03] ενώ η εξαγωγή συγχορδιών εξετάζεται στο [Zoia04].

Σε κάθε περίπτωση, το μελωδικό και αρμονικό περιεχόμενο περιγράφεται πιο αυτοδύναμα από τις ιδιότητες χαμηλότερων επιπέδων από ότι οι νότες ή οι συγχορδίες. Είναι γνωστό ότι έχει υπάρξει μόνο μια προσπάθεια να χρησιμοποιηθούν τέτοια χαρακτηριστικά γνωρίσματα κατά τη μοντελοποίηση των ειδών των ηχητικών σημάτων [Tzan02] ενώ έχουν χρησιμοποιηθεί πιο εντατικά στα πλαίσια της σημασιολογικής κατάτμησης και της σύνοψης της μουσικής [Chai06]. Η βασική ιδέα είναι να χρησιμοποιηθεί μια συνάρτηση που να χαρακτηρίζει τη κατανομή των τόνων (pitch) ενός μικρού τμήματος όπως στις περισσότερες συσκευές ανάλυσης μελωδίας/αρμονίας. Η διαφορά είναι ότι δεν λαμβάνεται υπ' όψη καμία απόφαση σχετικά με τη θεμελιώδη συχνότητα, τη συγχορδία, το μουσικό κλειδί ή άλλα χαρακτηριστικά γνωρίσματα υψηλού επιπέδου. Αντίθετα, ένα σύνολο περιγραφών υπολογίζεται από αυτήν την συνάρτηση συμπεριλαμβανομένου του πλάτους και των θέσεων των κύριων κορυφών της, το διάστημα μεταξύ των κορυφών, το άθροισμα της συνάρτησης ανιχνεύσης και ενδεχομένως οποιοδήποτε είδος στατιστικού περιγραφέα της κατανομής της συνάρτησης. Τυπικά χρησιμοποιούνται δύο εκδόσεις της συνάρτησης τόνων: μια αναδιπλωμένη έκδοση που περιέχει τις πληροφορίες για το εύρος των τόνων του κομματιού και τη «διπλωμένη» έκδοση, στην οποία χαρτογραφούνται όλες οι θεμελιώδεις συχνότητες σε μια ενιαία οκτάβα προσφέροντας μια καλή περιγραφή του αρμονικού περιεχομένου του κομματιού.

4.2.3. Ρυθμός

Δεν υπάρχει ένας ακριβής ορισμός για το ρυθμό. Οι περισσότεροι συγγραφείς αναφέρονται στην ιδέα της χρονικής κανονικότητας (temporal regularity). Στην πραγματικότητα, η κανονικότητα όπως την αντιλαμβάνεται ο άνθρωπος είναι διαφορετική από το ρυθμό. Γενικά, η λέξη ρυθμός μπορεί να χρησιμοποιηθεί για να αναφερθεί σε όλες τις χρονικές πλευρές ενός μουσικού έργου.

Διαισθητικά, είναι σαφές ότι το ρυθμικό περιεχόμενο μπορεί να αποτελεί μια διάσταση της μουσικής όταν γίνεται διάκριση μεταξύ της καθαρής μουσικής ροκ, από τη ρυθμική και πιο σύνθετη μουσική λάτιν, ή κατά την ανίχνευση κλασικής μουσικής.

Στη δημοσίευση [Gouy05] μπορεί να βρει κανείς μια αναθεώρηση των αυτόματων συστημάτων περιγραφής ρυθμού. Αυτά τα αυτόματα συστήματα μπορούν να χρησιμοποιηθούν σε διαφορετικές εφαρμογές: στην επαγωγή ρυθμού (tempo), στην ανίχνευση ρυθμού (beat tracking), στην επαγωγή του μέτρου, στην κβαντοποίηση του προς εκτέλεση ρυθμού, ή στο χαρακτηρισμό των σκόπιμων αποκλίσεων συγχρονισμού. Ακόμα, δεδομένου ότι τα συστήματα περιγραφής του ρυθμού που είναι υψηλότερου επιπέδου έχουν ακόμα ορισμένες αδυναμίες, για την αναγνώριση είδους μπορεί να χρησιμοποιηθεί μια προσέγγιση χαμηλότερου επιπέδου. Παραδείγματος χάριν, οι αλγόριθμοι για το ρυθμό (tempo) και την ανίχνευση ρυθμού εμφανίζουν λάθη με αποτέλεσμα να δίνουν αναξιόπιστες πληροφορίες για τους αλγόριθμους εκμάθησης μηχανών.

Μετά από την ίδια προσέγγιση όπως αυτή που ακολουθήθηκε στις χαμηλού επιπέδου ιδιότητες του τόνου, οι περιγραφείς μπορούν να εξαχθούν από μια συνάρτηση μέτρησης της σημασίας της περιοδικότητας στο εύρος των διακριτών ρυθμών (tempi) (χαρακτηριστικά μεταξύ 40 και 200 bpm στις εφαρμογές ταξινόμησης είδους). Τέτοια συνάρτηση μπορεί να ληφθεί εξετάζοντας τη διαφοροποίηση της αυτοσυσχέτισης (autocorrelation-like transform) των χαρακτηριστικών γνωρισμάτων με την πάροδο του χρόνου (ενδιαφέρον χαρακτηριστικό γνώρισμα είναι συνήθως η ενέργεια σε διαφορετικές ζώνες συχνότητας). Είναι επίσης δυνατό να χρησιμοποιηθεί μετασχηματισμός Fourier (Fast Fourier Transformation – FFT) για τον υπολογισμό των εναλλαγών χαρακτηριστικών γνωρισμάτων (τυπικά για παράθυρα άνω των 6 δευτερολέπτων). Οι ερευνητές στη δημοσίευση [Gouy04] παρουσιάζουν μια μελέτη σε βάθος για χαμηλού επιπέδου ρυθμικούς περιγραφείς που εξάγονται από μια συνάρτηση περιοδικότητας.

4.2.4. Εξαγωγή χαρακτηριστικών γνωρισμάτων από ηχητικά τμήματα

Οι περιγραφείς που παρουσιάστηκαν προηγουμένως μπορούν να εξαχθούν για ολόκληρα ηχητικά σήματα. Όμως, σε πολλές ταξινομήσεις, χρησιμοποιείται ένα μικρό τμήμα του ηχητικού σήματος, δεδομένου ότι μπορεί να περιέχει ικανοποιητικές πληροφορίες ώστε να χαρακτηρίσει το περιεχόμενο ολόκληρου του τραγουδιού, επειδή σε πολλά μουσικά είδη οι επαναλήψεις είναι έμφυτες στη δομή της μουσικής. Αυτή η ιδέα είναι ακόμη πιο σημαντική από τη στιγμή που εξετάζοντας μόνο ένα μικρό μέρος του σήματος, ο απαραίτητος υπολογισμός μπορεί να μειωθεί πολύ.

Οι περισσότεροι από τους προτεινόμενους αλγόριθμους για την ταξινόμηση είδους μουσικής χρησιμοποιούν πράγματι ένα μικρό τμήμα του ηχητικού σήματος ανά τίτλο: τυπικά ένα τμήμα 30 δευτερολέπτων που αρχίζει 30 δευτερόλεπτα μετά από την αρχή του κομματιού

για να αποφύγει την εισαγωγή που μπορεί να μην είναι αντιπροσωπευτική ολόκληρου του κομματιού.

Στα πλαίσια του προσδιορισμού των καλλιτεχνών, στη δημοσίευση [Bere02] προτείνεται η αυτόματη ανίχνευση τμημάτων ενός τραγουδιού. Τα αποτελέσματα που προέκυψαν με την ανάλυση μόνο του μέρους του τραγουδιού που περιέχει φωνή ήταν βελτιωμένα: καθώς μπορεί να είναι ευκολότερο να προσδιοριστεί ο καλλιτέχνης με το άκουσμα της φωνής του παρά με τη μουσική του.

Στη συνέχεια παρατίθενται επιμέρους συμπεράσματα από την έρευνα μέχρι στιγμής. Ο πίνακας I συνοψίζει τους τύπους των χαρακτηριστικών γνωρισμάτων που χρησιμοποιούνται στις εφαρμογές ανάκτησης πληροφοριών μουσικής [Scar06].

ΠΙΝΑΚΑΣ 1		
ΤΥΠΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΤΟΝ ΧΑΡΑΚΤΗΡΙΣΜΟ ΤΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ ΜΟΥΣΙΚΗΣ		
Χροιά	Μελωδία/Αρμονία	Ρυθμός
Μοντέλο δομής: μοντέλο χαρακτηριστικών γνωρισμάτων στο παράθυρο δομής:	Συνάρτηση τόνου: μέτρηση της ενέργειας στη συνάρτηση των μουσικών νότων:	Συνάρτηση περιοδικότητας: μετράει την περιοδικότητα των χαρακτηριστικών γνωρισμάτων:
1. Απλή μοντελοποίηση με στατιστικά χαμηλής προτεραιότητας	1. «αναδιπλωμένη συνάρτηση» : περιγράφει το περιεχόμενο και το εύρος των τόνων	1. Τεμπο: περιοδικότητες που κυμαίνονται μεταξύ 0.3 και 1.5s (δηλαδή 200 με 40 bpm)
2. Μοντελοποίηση με αλγορίθμους εκτίμησης κατανομής	2. «διπλωμένη συνάρτηση» : περιγράφει το περιεχόμενο της αρμονίας	2. Μουσικό πρότυπο: περιοδικότητας μεταξύ 2 και 6 δευτερολέπτων

Η εξαγωγή περιγραφών υψηλού επιπέδου από γενικά πολυφωνικά ηχητικά σήματα δεν είναι ακόμα ώριμη τεχνολογία. Οι περισσότερες προσεγγίσεις εστιάζουν στη μοντελοποίηση της χροιάς (timbre) βασισμένη στον συνδυασμό περιγραφών χαμηλού επιπέδου. Η χροιά μπορεί να περιέχει ικανοποιητικές πληροφορίες για να χαρακτηρίσει κατά προσέγγιση τα μουσικά είδη, δεδομένου ότι η έρευνα κατέδειξε πως άνθρωποι με λίγες ή καθόλου μουσικές γνώσεις ήταν σε θέση να κάνουν μια σωστή ταξινόμηση μουσικής (μεταξύ 10 ειδών) στο 53% των περιπτώσεων ύστερα από το άκουσμα μόλις 250 χιλιοστών του δευτερολέπτου και στο 72% των περιπτώσεων βασισμένων μόνο σε 3 δευτερόλεπτα του ηχητικού σήματος [Perr99]. Αυτό αποδεικνύει ότι δεν απαιτείται υψηλού επιπέδου κατανόηση της μουσικής για να χαρακτηριστεί το είδος, καθώς τα 250 χιλιοστά του δευτερολέπτου αλλά ακόμα και τα 3 δευτερόλεπτα είναι επίσης λίγος χρόνος προκειμένου να αναγνωριστεί μια μουσική δομή.

Οι Aucouturier και Pachet [Aucu03] έχουν μια πιο απαισιόδοξη άποψη. Έχουν μελετήσει το συσχετισμό μεταξύ της ομοιότητας της χροιάς και του είδους. Χρησιμοποίησαν ένα τελευταίας τεχνολογίας μέτρο ομοιότητας της χροιάς [Aucu02] και μια βάση δεδομένων 20.000 τίτλων με πάνω από 18 είδη. Τα αποτελέσματά τους δείχνουν ότι υπάρχει μόνο μια μικρή συσχέτιση μεταξύ της χροιάς και των ειδών προτείνοντας ότι τα συστήματα ταξινόμησης που βασίζονται απλώς στη χροιά είναι πραγματικά περιορισμένα. Επίσης προτείνουν ότι αυτά τα συστήματα ταξινόμησης μπορούν μετά βίας να αποδώσουν όσο αυξάνεται το πλήθος τόσο στον αριθμό των τίτλων όσο και στον αριθμό κατηγοριών είδους.

Υποστηρίζοντας ότι μπορεί να μην υπάρχουν αρκετές πληροφορίες στα ηχητικά σήματα για να χαρακτηρίσουν το μουσικό είδος ενός τίτλου, προσπάθησαν να λάβουν μερικά πολιτιστικά χαρακτηριστικά γνωρίσματα υπόψη [Aucu03] αντλώντας πληροφορία από το διαδίκτυο για να εξάγουν σχετικές λέξεις κλειδιά που σχετίζονται με τους τίτλους μουσικής.

4.3. Έμπειρα Συστήματα

Τα έμπειρα συστήματα εφαρμόζουν ρητά ορισμένα σύνολα κανόνων. Ο στόχος της ταξινόμησης είδους, θα μπορούσε να είναι ισοδύναμος με την απαρίθμηση διαφόρων κανόνων οι οποίοι θα χαρακτήριζαν ακριβώς και μοναδικά το κάθε είδος. Από όσο είναι γνωστό, κανένα πρότυπο βασισμένο στα έμπειρα συστήματα δεν έχει προταθεί για τον χαρακτηρισμό μουσικών ειδών. Η δουλειά των Pachet και Cazaly [Pach00] για την ταξινόμηση των μουσικών ειδών μπορεί να συγκριθεί με μια προσέγγιση έμπειρων συστημάτων παρόλο που δεν οδήγησε σε μια πραγματική εφαρμογή. Η συγκεκριμένη δουλειά αξίζει επίσης να αναφερθεί, δεδομένου ότι επιτρέπει την βαθύτερη κατανόηση των δυσκολιών της ταξινόμησης ειδών μουσικής.

Οι Pachet και Cazaly έχουν προσπαθήσει να καθορίσουν τα χαρακτηριστικά των ειδών και των μεταξύ τους σχέσεων. Έχουν δηλώσει τυπικά τις διαφορές μεταξύ των ειδών με μια γλώσσα βασισμένη στους περιγραφείς όπως είναι η ενορχήστρωση, ο τύπος της φωνής, ο τύπος του ρυθμού και ο ρυθμός του τραγουδιού. Αυτό υπονοεί ότι αυτοί οι περιγραφείς πρέπει να αναλυθούν λεπτομερώς για να χαρακτηρίσουν τις διαφορές μεταξύ των επιμέρους ειδών.

Αυτή η προσέγγιση, πιθανότατα, δεν είναι κατάλληλη για την ταξινόμηση του είδους, δεδομένης της πολυπλοκότητας του στόχου και της δυσκολίας να περιγράψει αντικειμενικά τα πολύ συγκεκριμένα επιμέρους είδη. Επιπλέον απαιτεί έναν (αυτόματο) τρόπο για τη λήψη αξιόπιστων περιγραφών υψηλού επιπέδου από το ακουστικό σήμα, το οποίο είναι ένα θέμα που δεν έχει λυθεί ακόμα, όπως περιγράφηκε στο προηγούμενο τμήμα.

Τα έμπειρα συστήματα, αν και ενσωματώνουν βαθιά γνώση του θέματος, είναι μια «ακριβή λύση» επειδή χρειάζονται συντήρηση. Καθώς ο αριθμός των χειροκίνητων κανόνων αυξάνεται, παράγονται απροσδόκητες αλληλεπιδράσεις και παρενέργειες, έτσι ώστε τα ζητήματα τεχνολογίας λογισμικού να γίνονται όλο και περισσότερο σημαντικά. Στα τελευταία έτη, η προσέγγιση της εκμάθησης μηχανών έχει συγκεντρώσει αυξανόμενο ενδιαφέρον. Γενικότερα, η προσέγγιση της εκπαίδευσης της μηχανής (machine learning) τείνει να επικρατήσει ανάμεσα σε παρόμοιους τομείς επεξεργασίας φυσικής γλώσσας και αναγνώρισης προτύπων όπως η αυτόματη αναγνώριση ομιλίας (speech recognition) η αναγνώριση προσώπου (face recognition).

4.4. Προσέγγιση χωρίς επίβλεψη

Ενώ μερικές προσεγγίσεις τείνουν να ταξινομήσουν τη μουσική με βάση μια αυθαίρετη ταξινόμηση μουσικών ειδών, μια άλλη άποψη είναι η συγκέντρωση των στοιχείων με έναν μη επιβλεπόμενο τρόπο έτσι ώστε η ταξινόμηση να προκύπτει από τα ίδια τα δεδομένα και να βασίζεται σε αντικειμενικά μέτρα ομοιότητας. Το πλεονέκτημα αυτής της προσέγγισης είναι η αποφυγή του περιορισμού μιας σταθερής και καθορισμένης ταξινόμησης, η οποία μπορεί να “πάσχει” από ασάφειες και ασυνέπειες όπως αναφέρθηκε προηγουμένως. Ένα άλλο πρόβλημα το οποίο μπορεί να λύσει αυτή η προσέγγιση είναι ότι μερικοί τίτλοι μπορεί απλά να μην ταιριάζουν σε καμία από τις ήδη υπάρχουσες ταξινόμησεις.

Στην χωρίς επίβλεψη προσέγγιση, ένας τίτλος αντιπροσωπεύεται από ένα σύνολο χαρακτηριστικών γνωρισμάτων όπως αναφέρθηκε σε προηγούμενη παράγραφο. Επίσης σε

αυτή την προσέγγιση για τη σύγκριση μεταξύ των τίτλων χρησιμοποιείται ένα μέτρο ομοιότητας. Στη συνέχεια οι αλγόριθμοι ομαδοποίησης χωρίς επίβλεψη αξιοποιούν το μέτρο ομοιότητας για να οργανώσουν τις συλλογές μουσικής δημιουργώντας ομάδες με παρόμοιους τίτλους.

4.4.1. Μέτρα ομοιότητας

Η απλούστερη επιλογή για να μετρήσει κανείς την απόσταση μεταξύ δύο διανυσμάτων χαρακτηριστικών γνωρισμάτων είναι, για παράδειγμα, να χρησιμοποιηθεί η Ευκλείδεια απόσταση. Εντούτοις αυτές οι αποστάσεις έχουν νόημα μόνο εάν τα διανύσματα χαρακτηριστικών γνωρισμάτων είναι αμετάβλητα στο χρόνο. Διαφορετικά δύο παρόμοιοι τίτλοι μπορεί να απέχουν με βάση το μέτρο, εάν παρόμοια χαρακτηριστικά γνωρίσματα είναι μεταβλητά στο χρόνο. Για να μετατραπεί μια χρονική σειρά σε μια αμετάβλητη στο χρόνο απεικόνιση του ηχητικού σήματος, συνήθως κατασκευάζεται ένα στατιστικό μοντέλο κατανομής των χαρακτηριστικών γνωρισμάτων και έπειτα χρησιμοποιείται η απόσταση για τη σύγκριση αυτών των μοντέλων.

Τυπικά μοντέλα περιλαμβάνουν γκαουσιανά μίγματα (Gaussian Mixtures - GMM) (τα GMMs έχουν χρησιμοποιηθεί για τη μοντελοποίηση της χροιάς ενός κομματιού στις δημοσιεύσεις [Aucu02], [Pamp05] και [Mand05]). Η απόκλιση kullback-Leibler ή σχετική εντροπία είναι ο φυσικός τρόπος να υπολογιστεί η απόσταση μεταξύ των κατανομών πιθανότητας αλλά δεν είναι κατάλληλη για τα GMM. Στο [Aucu02] παρουσιάζονται κάποια εναλλακτικά μέτρα συμπεριλαμβανομένης της δειγματοληψίας, της απόστασης Earth's Mover και της ασυμπτωτικής προσέγγισης πιθανότητας (Asymptotic Likelihood Approximation).

Εξετάζοντας το γεγονός ότι, αντίθετα με τα περισσότερα κλασικά προβλήματα αναγνώρισης προτύπων, τα στοιχεία που ταξινομούνται είναι χρονικές σειρές, ο Shao και οι λοιποί στο [Shao04] χρησιμοποιούν μοντέλα Hidden Markov (Hidden Markov Models - HMM) για να μοντελοποιήσουν τη σχέση μεταξύ των χαρακτηριστικών γνωρισμάτων και του χρόνου. Ένα ενδιαφέρον των μοντέλων HMM είναι ότι παρέχουν μια κατάλληλη μετρική για την απόσταση έτσι ώστε μόλις χαρακτηριστεί κάθε κομμάτι από το δικό του HMM, μπορεί να υπολογιστεί η απόσταση μεταξύ οποιωνδήποτε κομματιών στη βάση δεδομένων.

4.4.2. Αλγόριθμοι Ομαδοποίησης

Ο αλγόριθμος K-means είναι ίσως ο απλούστερος και δημοφιλέστερος αλγόριθμος ομαδοποίησης. Επιτρέπει το διαχωρισμό ενός συνόλου διανυσμάτων σε K ξεχωριστά υποσύνολα. Μια από τις αδυναμίες του αλγορίθμου είναι ότι απαιτεί εκ των προτέρων τον αριθμό συστάδων (K). Ο Shao και οι λοιποί στο [Shao04] ομαδοποίησαν τη συλλογή της μουσικής τους με έναν αλγόριθμο συσσωρευτικής ιεραρχικής ομαδοποίησης (Agglomerative Hierarchical Clustering). Πρόκειται για έναν αλγόριθμο ομαδοποίησης που ξεκινάει με N μονήρεις (singleton) συστάδες (όπου N είναι ο αριθμός των κομματιών στη βάση δεδομένων) και έτσι διαμορφώνει μια ακολουθία συστάδων με διαδοχική συγχώνευση.

Ο αυτό-οργανούμενος χάρτης (Self-Organizing Map - SOM) και ο αυξανόμενος ιεραρχικός αυτό-οργανούμενος χάρτης (Growing Hierarchical Self-Organizing Map - GHSOM) χρησιμοποιούνται για την ομαδοποίηση των δεδομένων και την οργάνωσή τους σε ένα δισδιάστατο χώρο κατά τέτοιο τρόπο ώστε τα διανύσματα με παρόμοια χαρακτηριστικά γνωρίσματα να συγκεντρώνονται κοντά. Τα SOM είναι τεχνητά νευρωνικά δίκτυα χωρίς επίβλεψη που χαρτογραφούν τα πολυδιάστατα δεδομένα εισόδου σε λιγότερες διαστάσεις διατηρώντας τις τοπολογικές σχέσεις μεταξύ των δεδομένων εισόδου όσο το δυνατό με μεγαλύτερη ακρίβεια. Τα GHSOM είναι μια ειδική περίπτωση των SOM που χρησιμοποιούν

μια ιεραρχική δομή με πολλαπλά στρώματα όπου κάθε στρώμα αποτελείται από διάφορα ανεξάρτητα SOM. Ο Rauber και οι λοιποί στο [Raub02] χαρτογράφησαν τα δεδομένα εισόδου σε 2 διαστάσεις προκειμένου να επιτρέψουν την οπτική αναπαράσταση μιας συλλογής μουσικής με ένα GHSOM.

Το σημαντικότερο μειονέκτημα των τεχνικών χωρίς επίβλεψη μπορεί να είναι ότι οι συστάδες που δημιουργούνται δεν αποκτούν κάποια ετικέτα. Εν πάση περιπτώσει, οι συστάδες αυτές δεν απεικονίζουν πάντα ιεραρχίες ειδούς, ενώ οι ομοιότητες εξαρτώνται από τον τύπο των χαρακτηριστικών γνωρισμάτων (ομοιότητες ρυθμού, ομοιότητες μελωδίας, κ.λπ.).

4.5. Προσέγγιση με επίβλεψη

Η προσέγγιση με επίβλεψη στην ταξινόμηση είδους μουσικής έχει μελετηθεί εκτενέστερα. Οι μέθοδοι αυτής της ομάδας θεωρούν ότι δίνεται μια ταξινόμηση ειδών και στη συνέχεια γίνεται προσπάθεια χαρτογράφησης της βάσης δεδομένων των τραγουδιών σε αυτήν, με τη βοήθεια των αλγορίθμων εκμάθησης μηχανών. Σαν πρώτο βήμα, το σύστημα εκπαιδεύεται με ορισμένα δεδομένα τα οποία έχουν λάβει ετικέτες χειροκίνητα και έπειτα χρησιμοποιείται για να ταξινομήσει τα στοιχεία που δεν έχουν ετικέτες. Το σημαντικότερο ενδιαφέρον της ταξινόμησης με επίβλεψη σε αντίθεση με την προσέγγιση των έμπειρων συστημάτων είναι ότι δεν χρειάζεται να περιγράψει κανείς με σαφήνεια τα μουσικά είδη: ο ταξινομητής προσπαθεί να διαμορφώσει αυτόματα τις σχέσεις μεταξύ των χαρακτηριστικών γνωρισμάτων του συνόλου προς εκπαίδευση και των σχετικών κατηγοριών.

Στη συνέχεια περιγράφονται ορισμένοι από τους πιο γνωστούς αλγορίθμους εκμάθησης μηχανών με επίβλεψη. Κατόπιν παρουσιάζονται τα αποτελέσματα που επιτυγχάνονται με αυτούς τους αλγορίθμους στη βιβλιογραφία.

4.5.1. Ταξινομητές με επίβλεψη

- K-κοντινότερος γείτονας (KNN): πρόκειται για έναν μη παραμετρικό ταξινομητή ο οποίος βασίζεται στην ιδέα ότι ένας μικρός αριθμός γειτόνων επηρεάζει την απόφαση σε ένα σημείο. Πιο συγκεκριμένα, για ένα δεδομένο διάνυσμα χαρακτηριστικών γνωρισμάτων στο σύνολο που μελετάται, επιλέγονται τα K πιο κοντινά διανύσματα (σύμφωνα με ορισμένα μέτρα απόστασης). Το διάνυσμα αυτό των χαρακτηριστικών γνωρισμάτων ορίζεται ως ετικέτα της πιο αντιπροσωπευτικής κλάσης στη “γειτονιά” του K. Ο KNN αξιολογείται στα πλαίσια της ταξινόμησης είδους στο [Tzan02] και στο [Pamp05].
- Gaussian Mixture Models (GMM): Τα GMM μοντελοποιούν την κατανομή των διανυσμάτων χαρακτηριστικών γνωρισμάτων. Για κάθε κλάση, θεωρείται ότι υπάρχει μια συνάρτηση πυκνότητας πιθανότητας η οποία εκφράζεται ως μίγμα διάφορων πολυδιάστατων γκαουσιανών κατανομών. Για τον υπολογισμό των παραμέτρων σε κάθε γκαουσιανό συστατικό καθώς και τον υπολογισμό των βαρών του μίγματος χρησιμοποιείται συνήθως ο επαναληπτικός αλγόριθμος μεγιστοποίησης προσδοκίας (Expectation Maximization - EM). Τα GMM χρησιμοποιούνται ευρέως στην κοινότητα ανάκτησης πληροφοριών μουσικής, ειδικότερα για την κατασκευή μοντέλων χροιάς όπως περιγράφηκε νωρίτερα. Τα μοντέλα GMM μπορούν να χρησιμοποιηθούν ως ταξινομητές οι οποίοι μπορούν, με τη χρήση ενός κριτηρίου μέγιστης πιθανότητας, να βρουν το καταλληλότερο μοντέλο για ένα συγκεκριμένο τραγούδι. Τέτοιοι ταξινομητές έχουν χρησιμοποιηθεί για τη μοντελοποίηση μουσικών ειδών στο [Tzan02]. Στη

δημοσίευση Burge03] χρησιμοποιείται μια δεντρική δομή GMM για τη μοντελοποίηση της ταξινόμησης είδους. Με μια στρατηγική “Διαίρει και βασίλευε” ταξινομούνται πρώτα τα στοιχεία σε ένα πρώτο επίπεδο και έπειτα σε διαδοχικά επίπεδα με μεγαλύτερη ακρίβεια. Η απόφαση ταξινόμησης αποσυντίθεται έτσι σε έναν αριθμό από τοπικές αποφάσεις δρομολόγησης. Επιπλέον, η επιλογή χαρακτηριστικών γνωρισμάτων σε κάθε επίπεδο επιτρέπει τη βελτιστοποίηση των αποτελεσμάτων ταξινόμησης. Οι West και COX στο [West05] χρησιμοποιούν ένα δυαδικό δέντρο (Maximal Classification Binary Tree) το οποίο κατασκευάζεται λαμβάνοντας αρχικά έναν κόμβο ρίζα ο οποίος περιέχει όλα τα δεδομένα και στη συνέχεια το στοιχείο αυτό διαχωρίζεται σε δύο κόμβους παιδιά με τη χρήση του ενιαίου γκαουσιανού ταξινομητή και βάση μετρήσεων απόστασης Mahalanobis. Προκειμένου να χωριστεί ένας κόμβος, δημιουργούνται όλοι οι πιθανοί συνδυασμοί κλάσεων και στη συνέχεια επιλέγεται εκείνος ο συνδυασμός κλάσεων που παράγει τον καλύτερο διαχωρισμό (σημειώνεται ότι η δημιουργία του δέντρου γίνεται χωρίς επίβλεψη ενώ οι ταξινομητές που χρησιμοποιούνται για το διαχωρισμό σε κάθε κόμβο εκπαιδεύονται με επίβλεψη).

- Hidden Markov Mode (HMM): Τα κρυμμένα μοντέλα Markov μπορούν να χρησιμοποιηθούν για ταξινόμηση. Έχουν χρησιμοποιηθεί εκτενώς στην αναγνώριση φωνής λόγω της ικανότητάς τους να χειρίζονται χρονοσειρές δεδομένων. Τα HMM μπορούν να θεωρηθούν ως μια διπλή ενσωματωμένη στοχαστική διαδικασία: μια διαδικασία η οποία δεν είναι άμεσα αισθητή (κρυμμένη) και μπορεί μόνο να παρατηρηθεί μέσω μιας άλλης στοχαστικής διαδικασίας (παρατηρήσιμη) που παράγει το σύνολο των παρατηρήσεων. Αν και μπορούν να είναι κατάλληλα για τη μοντελοποίηση της μουσικής, από όσο είναι γνωστό, τα HMM έχουν χρησιμοποιηθεί μόνο στο [Scar03] και στο [Solt98] για την ταξινόμηση είδους ακουστικού περιεχομένου (έχουν χρησιμοποιηθεί και στο [Shao04] επίσης αλλά στην περίπτωση της οργάνωσης μιας συλλογής μουσικής χωρίς επίβλεψη).
- Γραμμική διακρίνουσα ανάλυση (LDA): Η βασική ιδέα του LDA είναι η εύρεση ενός γραμμικού μετασχηματισμού ο οποίος να επιτυγχάνει τον καλύτερο διαχωρισμό μεταξύ των κλάσεων καθώς και η εκτέλεση της ταξινόμησης στο μετασχηματισμένο χώρο - βασισμένη σε μερικές μετρικές όπως είναι η Ευκλείδεια απόσταση. Στο [Casa05] χρησιμοποιήθηκε ο αλγόριθμος LDA σε συνδυασμό με τον Adaboost. Ο αλγόριθμος Adaboost, για την προσαρμοστική ώθηση, χρησιμοποιείται από κοινού με άλλους αλγορίθμους εκμάθησης (σε αυτήν την περίπτωση με τον LDA) για να βελτιώσει τις αποδόσεις των ταξινομήσεων και τις γενικεύσεις. Ο Adaboost είναι ένας προσαρμοστικός αλγόριθμος με την έννοια ότι οι επόμενοι ταξινομητές που κατασκευάζονται είναι βελτιωμένες ταξινομήσεις των περιπτώσεων που ταξινομήθηκαν λάθος από προηγούμενους ταξινομητές. Στο [West05], χρησιμοποιείται το κριτήριο των ψαράδων Multi-class LDA προκειμένου να μειωθεί η διαστατικότητα του προβλήματος της ταξινόμησης πριν τη μοντελοποίηση με μια γκαουσιανή διανομή.
- Διανυσματικές μηχανές υποστήριξης (SVM): Ο αλγόριθμος SVM βασίζεται σε δύο ιδιότητες: τη μεγιστοποίηση περιθωρίου (που επιτρέπει μια καλή γενίκευση του ταξινομητή) και το μη γραμμικό μετασχηματισμό του χώρου χαρακτηριστικών γνωρισμάτων με πυρήνες (καθώς ένα σύνολο στοιχείων είναι πιο εύκολα διαχωρίσιμο σε ένα χώρο περισσότερων διαστάσεων - χαρακτηριστικών γνωρισμάτων). Ο SVM έχει χρησιμοποιηθεί στα πλαίσια της ταξινόμησης είδους στο [Scar05] και στο [Lidy05]. Στο [Mand05], ο SVM χρησιμοποιείται για την ταξινόμηση είδους και βασίζεται σε έναν αλγόριθμο για την μέτρηση της απόστασης μεταξύ δυο τραγουδιών που λέγεται Kullback

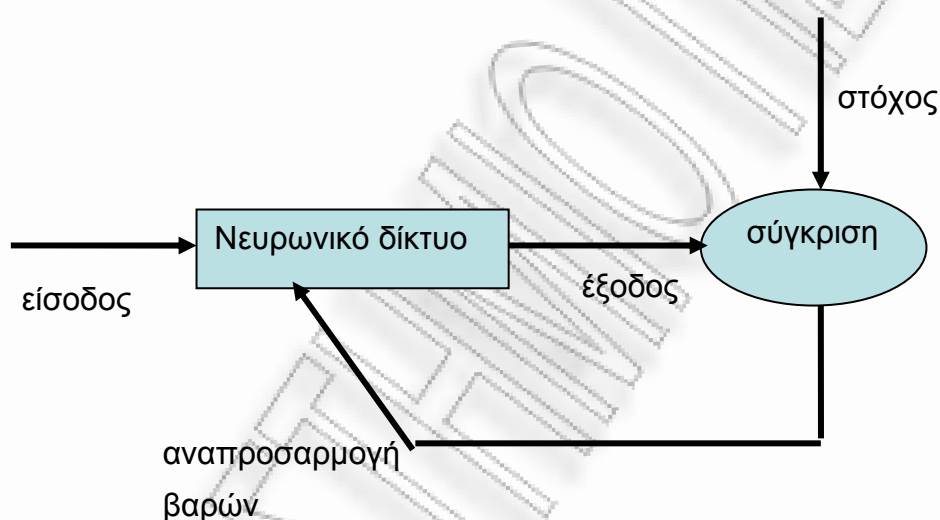
Lleiber divergence-based kernel. Στο [Scari05], η ταξινόμηση είδους γίνεται με ένα μίγμα εμπειρογνομόνων SVM. Ένα μίγμα εμπειρογνομόνων μπορεί να λύσει ένα πρόβλημα ταξινόμησης με τη χρήση διάφορων ταξινομητών με σκοπό να το αποσυνθέσει σε μια σειρά από υποπροβλήματα. Έτσι όχι μόνο μειώνεται η πολυπλοκότητα κάθε ενιαίου στόχου αλλά βελτιώνεται επίσης η σφαιρική ακρίβεια με το συνδυασμό των αποτελεσμάτων από διαφορετικούς ταξινομητές (εμπειρογνώμονες). Φυσικά, ο αριθμός ταξινομητών που απαιτούνται αυξάνεται, όμως, έχοντας κάποιον από αυτούς να χειρίζεται ένα απλούστερο πρόβλημα, η συνολική υπολογιστική δύναμη που απαιτείται μειώνεται.

- Τεχνητό νευρωνικό δίκτυο (ANN): Τα πιο γνωστά τεχνητά νευρωνικά δίκτυα με επίβλεψη που χρησιμοποιούνται για την αναγνώριση προτύπων είναι τα πολυστρωματικά Perceptron (MLP). Πρόκειται για ένα πολύ γενικό μοντέλο που μπορεί σε γενικές γραμμές να προσεγγίσει οποιαδήποτε μη γραμμική συνάρτηση. Τα MLP έχουν χρησιμοποιηθεί στο [Bere02] στα πλαίσια του προσδιορισμού του καλλιτέχνη. Τα νευρωνικά δίκτυα καθώς επίσης και οι άλλες αναθεωρημένες αρχιτεκτονικές (εκτός από το HMM), μπορούν να χειριστούν μόνο τα στατικά πρότυπα. Αυτή η αδυναμία αντιμετωπίζεται εν μέρει στο [Bere02] με την εισαγωγή διάφορων συνεχόμενων διανυσμάτων χαρακτηριστικών γνωρισμάτων στο δίκτυο έτσι ώστε να λαμβάνονται υπόψη οι πληροφορίες που βασίζονται στα συμφοραζόμενα: αυτή η στρατηγική αντιστοιχεί στο αποκαλούμενο Feedforward Time-Delay Neural Network (Νευρωνικό δίκτυο χρονικής καθυστέρησης (TDNN)). Έχουν προταθεί και άλλα παραδείγματα που προσανατολίζονται προς την επεξεργασία των χρονικών ακολουθιών (επαναλαμβανόμενα δίκτυα όπως το δίκτυο elman) αλλά δεν έχουν χρησιμοποιηθεί ακόμα στα πλαίσια της ταξινόμησης είδους μουσικής. Ο Soltau και λοιποί στο [Solt98] έχουν εισάγει στα πλαίσια της αναγνώρισης ειδών μουσικής μια αρχική μέθοδο για τη σαφή μοντελοποίηση του χρόνου της προσωρινής δομής της μουσικής (etm-NN): ένα MLP εκπαιδεύεται ώστε να αναγνωρίζει τα είδη μουσικής, αλλά αντί να εξετάζεται η έξοδος του, εξετάζεται η ενεργοποίηση των κρυμμένων νευρώνων του, ως μια συμπαγής αντιπροσώπευση του διανύσματος χαρακτηριστικών γνωρισμάτων που λαμβάνεται ως είσοδος (είναι γνωστό πράγματι ότι το πρώτο μισό ενός feed-forward δικτύου εκτελεί έναν συγκεκριμένο μη γραμμικό μετασχηματισμό των δεδομένων εισόδου σε ένα χώρο στον οποίο η διάκριση πρέπει να είναι απλούστερη). Κάθε κρυμμένος νευρώνας μπορεί να θεωρηθεί ως μια σύνοψη μουσικού γεγονότος - όχι απαραίτητα σχετική με μια πραγματική μουσική έκφραση. Η ακολουθία της σύνοψης των γεγονότων με την πάροδο του χρόνου αναλύεται έπειτα για να δημιουργηθεί ένα ενιαίο διάλυμα χαρακτηριστικών γνωρισμάτων με το οποίο τροφοδοτείται ένα δεύτερο δίκτυο που εφαρμόζει τη τελική απόφαση για το είδος του μουσικού κομματιού. Στο [Scar05] αξιολογείται η αρχιτεκτονική etm-NN σε σύγκριση με άλλους ταξινομητές.

5. Εισαγωγή στο Neural Network Toolbox του Matlab

Τα νευρωνικά δίκτυα αποτελούνται από απλά στοιχεία που λειτουργούν παράλληλα. Τα στοιχεία στα δίκτυα αυτά προσομοιώνουν σε ένα βαθμό τα βιολογικά νευρωνικά συστήματα. Όπως στη φύση, οι συνδέσεις μεταξύ των στοιχείων καθορίζουν κατά ένα μεγάλο μέρος τη συνάρτηση του δικτύου. Μπορεί κανείς να εκπαιδεύσει ένα νευρωνικό δίκτυο έτσι ώστε αυτό να εκτελεί μια ιδιαίτερη συνάρτηση με τη ρύθμιση των τιμών των συνδέσεων (βάρη) μεταξύ των στοιχείων.

Χαρακτηριστικά, τα νευρωνικά δίκτυα ρυθμίζονται, ή εκπαιδεύονται, έτσι ώστε μια συγκεκριμένη είσοδος να οδηγεί σε μια συγκεκριμένη επιθυμητή έξοδο. Το σχήμα που ακολουθεί δείχνει ένα τέτοιο παράδειγμα. Στο σχήμα αυτό το δίκτυο ρυθμίζεται έτσι ώστε η έξοδος τελικά να ταιριάζει με τον αρχικό στόχο που έχει τεθεί. Αυτό επιτυγχάνεται με μια σύγκριση που γίνεται μεταξύ της εξόδου και του επιθυμητού στόχου. Χαρακτηριστικά, πολλά τέτοια ζευγάρια εισόδων/στόχων απαιτούνται για να εκπαιδεύσουν ένα δίκτυο.



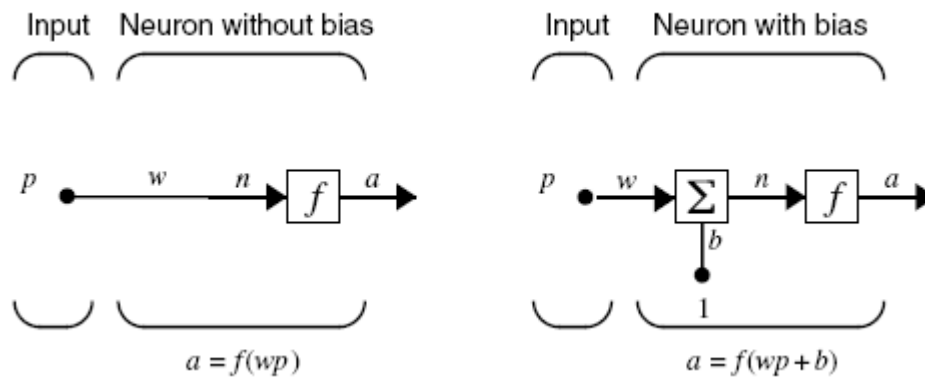
Σχήμα 6: Ένα νευρωνικό δίκτυο

Τα νευρωνικά δίκτυα έχουν χρησιμοποιηθεί για την εκτέλεση σύνθετων λειτουργιών σε διάφορους τομείς, όπως η αναγνώριση προτύπων, η επεξεργασία και αναγνώριση ομιλίας, εικόνας, και τα συστήματα ελέγχου.

Τα νευρωνικά δίκτυα μπορούν επίσης να εκπαιδευθούν για να λύσουν τα προβλήματα που είναι δύσκολα για τους συμβατικούς υπολογιστές ή τους ανθρώπους. Οι επόμενες ενότητες περιγράφουν τα βασικά χαρακτηριστικά της εργαλειοθήκης του Matlab στα νευρωνικά δίκτυα. Πιο συγκεκριμένα, παρουσιάζονται οι βασικές αρχές του προτύπου νευρώνων, οι αρχιτεκτονικές των νευρωνικών δικτύων και τα βασικά στοιχεία καθώς και η ονοματολογία γύρω από αυτά τα δίκτυα. Στη συνέχεια, γίνεται παρουσίαση της διαδικασίας δημιουργίας και εκπαίδευσης απλών δικτύων με τη χρήση της εργαλειοθήκης του Matlab. Παρουσιάζεται επίσης ένα γραφικό περιβάλλον για τον χρήστη (GUI) που μπορεί να χρησιμοποιήσει κανείς για να λύσει προβλήματα χωρίς ιδιαίτερη χρήση κώδικα.

5.1. Μοντέλο Νευρώνων

Ένας νευρώνας με μια μοναδική αριθμητική είσοδο και καμία πόλωση εμφανίζεται στο αριστερό μέρος του παρακάτω σχήματος.



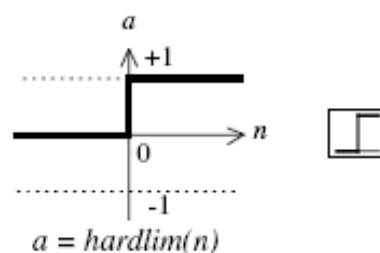
Σχήμα 7: Μοντέλο Νευρώνων

Η αριθμητική είσοδος P διαβιβάζεται μέσω μιας σύνδεσης που πολλαπλασιάζει τη δύναμή της με το αριθμητικό βάρος (weight) W για να διαμορφώσει το προϊόν wp , που είναι επίσης αριθμός. Εδώ η σταθμισμένη εισαγωγή wp είναι το μόνο όρισμα της συνάρτησης μεταφοράς f , η οποία παράγει την αριθμητική έξοδο a . Ο νευρώνας στα δεξιά του σχήματος έχει μια αριθμητική πόλωση (bias), b . Κανείς μπορεί να αντιληφθεί τη πόλωση σαν ένα προσθετέο που απλά προστίθεται στο προϊόν wp όπως παρουσιάζεται από τον κόμβο άθροιση Σ ή ως μετατόπιση της συνάρτησης f προς τα αριστερά κατά ένα ποσό b . Η πόλωση είναι σαν ένα βάρος με τη διαφορά ότι έχει σταθερή είσοδο 1 .

Η συνάρτηση μεταφοράς n είναι επίσης ένας αριθμός και ισοδυναμεί με το άθροισμα της σταθμισμένης εισαγωγής wp με την πόλωση b . Αυτό το ποσό είναι το όρισμα της συνάρτησης μεταφοράς f . Στην περίπτωση αυτή η f είναι μια συνάρτηση μεταφοράς, για παράδειγμα μια συνάρτηση βημάτων ή μια σιγμοειδής συνάρτηση, η οποία παίρνει το όρισμα n και παράγει την έξοδο a . Αξίζει να σημειωθεί ότι το w και το b είναι και οι δύο διευθετήσιμες αριθμητικές παράμετροι του νευρώνα. Η κεντρική ιδέα των νευρικών δικτύων είναι ότι τέτοιες παράμετροι μπορούν να ρυθμιστούν έτσι ώστε το δίκτυο να παρουσιάζει κάποια επιθυμητή ή ενδιαφέρουσα συμπεριφορά. Κατά συνέπεια, μπορεί κανείς να εκπαιδεύσει το δίκτυο να εκτελεί μια ιδιαίτερη εργασία με τη ρύθμιση του βάρους ή της πόλωσης, ή ίσως το ίδιο το δίκτυο θα ρυθμίσει αυτές τις παραμέτρους για να επιτύχει κάποιο επιθυμητό αποτέλεσμα.

5.2. Συναρτήσεις Μεταφοράς

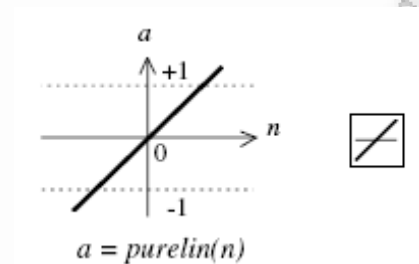
Πολλές συναρτήσεις μεταφοράς συμπεριλαμβάνονται στο Neural Network Toolbox. Στη συνέχεια παρουσιάζονται τέσσερις από τις πιο συχνά χρησιμοποιούμενες συναρτήσεις.



Σχήμα 8: Συνάρτηση Μεταφοράς Hard-Limit

Η συνάρτηση μεταφοράς αυστηρού ορίου (hard-limit) που εμφανίζεται στο παραπάνω σχήμα περιορίζει την έξοδο του νευρώνα είτε σε 0, εάν το όρισμα n είναι λιγότερο από 0, είτε 1, εάν το n είναι μεγαλύτερο ή ίσο με 0. Αυτή η λειτουργία χρησιμοποιείται στα «Perceptrons,» για να δημιουργήσει τους νευρώνες που λαμβάνουν αποφάσεις ταξινόμησης. Η υλοποίηση της παραπάνω μαθηματικής έννοιας λέγεται `hardlim` στο Neural Network Toolbox.

Μια ακόμα συνηθισμένη συνάρτηση μεταφοράς είναι η γραμμική συνάρτηση μεταφοράς. Η αντίστοιχη συνάρτηση του Neural Network Toolbox (`purelin`) και το σύμβολό της παρουσιάζονται παρακάτω.

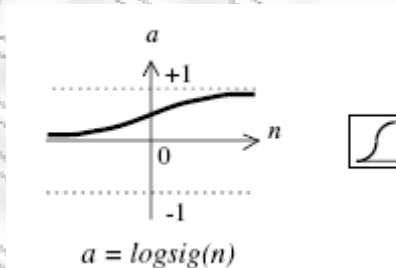


Σχήμα 9: Γραμμική συνάρτηση μεταφοράς

Η `purelin` δίνεται από τον παρακάτω τύπο:

$$\text{purelin}(n) = n$$

Νευρώνες αυτού του τύπου χρησιμοποιούνται ως γραμμικές προσεγγίσεις στα γραμμικά φίλτρα. Μια ακόμα συνήθης συνάρτηση μεταφοράς στα νευρωνικά δίκτυα είναι η σιγμοειδής (sigmoid). Η συνάρτηση αυτή παίρνει σαν όρισμα μια τιμή ανάμεσα στο πλην άπειρο και τον συν άπειρο και περιορίζει την έξοδο στο εύρος 0 με 1.

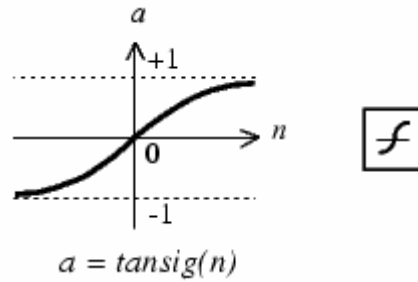


Σχήμα 10: Συνάρτηση Sigmoid

Η `logsig` δίνεται από τον παρακάτω τύπο:

$$\text{logsig}(n) = \frac{1}{1 + e^{-n}}$$

Παραλλαγή της `logsig` είναι η `tansig`. Η `tansig` (Hyperbolic tangent sigmoid transfer function) εικονίζεται στο παρακάτω σχήμα:



Σχήμα 11 Συνάρτηση tansig

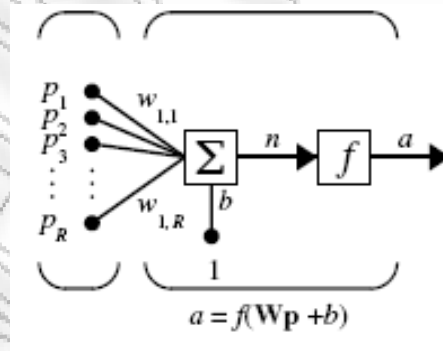
Η tansig δίνεται από τον παρακάτω τύπο:

$$\text{tansig}(n) = \frac{2}{1 + e^{-2n}} - 1$$

Το σύμβολο στο τετράγωνο δεξιά κάθε γραφικής παράστασης των συναρτήσεων μεταφοράς που παρουσιάστηκαν στα προηγούμενα σχήματα αντιπροσωπεύει τη σχετική συνάρτηση μεταφοράς. Αυτά τα εικονίδια αντικαθιστούν το γενικό f στα διαγράμματα δικτύων για να υποδηλώσουν τη χρήση της συγκεκριμένης συνάρτησης μεταφοράς.

5.3. Νευρώνες με Είσοδο Διάνυσμα

Ένας νευρώνας εκτός από αριθμητική τιμή μπορεί να έχει ως είσοδο και ένα διάνυσμα.



Σχήμα 12: Είσοδος Διάνυσμα

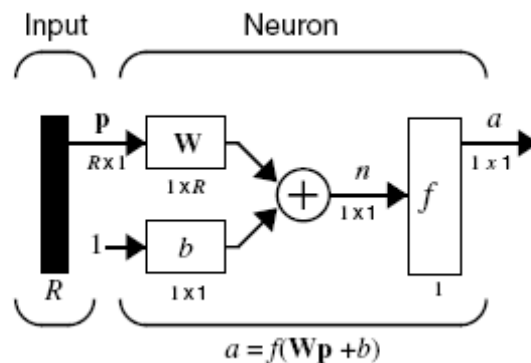
Αν η είσοδος είναι ένα διάνυσμα \mathbf{p} με R ξεχωριστά στοιχεία: $\{p_1, p_2, \dots, p_R\}$ τότε αυτό πολλαπλασιάζεται με το αντίστοιχο διάνυσμα βαρών $\mathbf{W} = \{w_1, w_2, \dots, w_R\}$. Αν ο νευρώνας έχει και πόλωση b , τότε το όρισμα της συνάρτησης μεταφοράς του νευρώνα είναι:

$$n = w_1 p_1 + w_2 p_2 + \dots + w_R p_R + b.$$

Το διάγραμμα ενός μόνο νευρώνα που παρουσιάζεται στο προηγούμενο σχήμα περιέχει πολλή λεπτομέρεια. Όταν εξετάζει κανείς δίκτυα με πολλούς νευρώνες, και ίσως στρώματα

πολλών νευρώνων χρειάζεται ένα πιο λιτό συμβολισμό. Για το λόγο αυτό, χρησιμοποιείται ένας πιο λιτός συμβολισμός ανά νευρώνα για τη σχεδίαση κυκλωμάτων πολλαπλών νευρώνων, ο οποίος παρουσιάζεται στο επόμενο σχήμα. Εδώ το διάνυσμα είσοδος \mathbf{p} αντιπροσωπεύεται από την κατακόρυφη μπάρα στα αριστερά.

Οι διαστάσεις του \mathbf{p} παρουσιάζονται κάτω από το σύμβολο \mathbf{p} στο σχήμα ως $R \times 1$. (Σημειώνετε ότι ένα κεφαλαίο γράμμα, όπως το R στην προηγούμενη πρόταση, χρησιμοποιείται κατά αναφορά στο μέγεθος ενός διανύσματος.) Κατά συνέπεια, το \mathbf{p} είναι ένα διάνυσμα R στοιχείων εισόδου.

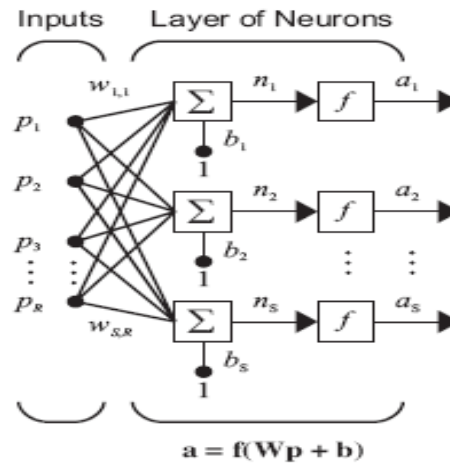


Σχήμα 13: Σύντομος Συμβολισμός Νευρώνα

Το παραπάνω σχήμα ορίζει ένα στρώμα του δικτύου. Ένα στρώμα περιέχει το συνδυασμό βαρών και το πολλαπλασιασμό (\mathbf{Wp}), τη πρόσθεση της πόλωσης b και τη συνάρτηση μεταφοράς f . Το διάνυσμα εισόδων \mathbf{p} δεν συμπεριλαμβάνεται στο στρώμα. Κάθε φορά που χρησιμοποιείται αυτός ο σύντομος συμβολισμός, τα μεγέθη των πινάκων φαίνονται κάτω ακριβώς από τα ονόματα των μεταβλητών που τους αναπαριστούν. Στο παραπάνω σχήμα, όταν χρησιμοποιείται κάποια συγκεκριμένη συνάρτηση μεταφοράς, όπως αυτές που περιγράφηκαν προηγουμένως, ο ειδικός συμβολισμός τους αντικαθιστά το κουτί με το f .

5.4. Αρχιτεκτονικές Δικτύου

Δύο ή περισσότεροι από τους νευρώνες που παρουσιάστηκαν νωρίτερα μπορούν να συνδυαστούν σε ένα στρώμα, και ένα δίκτυο θα μπορούσε να περιέχει ένα ή περισσότερα τέτοια στρώματα.



Σχήμα 14: Στρώμα Νευρώνων

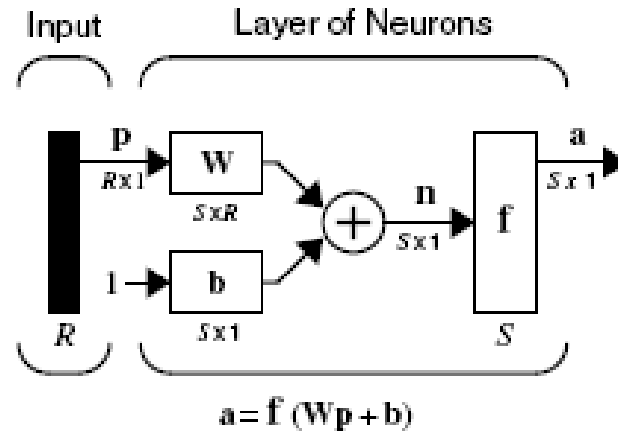
Στο παραπάνω σχήμα παρουσιάζεται ένα δίκτυο από S νευρώνες με R εισόδους. Σε αυτό το δίκτυο, κάθε στοιχείο του διανύσματος εισόδου \mathbf{p} συνδέεται με κάθε νευρώνα που εισάγεται μέσω του πίνακα βάρους \mathbf{W} . Ο i -στός νευρώνας έχει έναν αθροιστή που συλλέγει τις σταθμισμένες εισαγωγές και την πόλωση του για να διαμορφώσει την αριθμητική έξοδο $n(i)$. Τα διάφορα $n(i)$ όταν συλλέγονται δημιουργούν ένα δίκτυο αποτελούμενο από S στοιχεία. Τέλος, οι έξοδοι του στρώματος διαμορφώνουν μια στήλη-διάνυσμα \mathbf{a} . Η έκφραση για το \mathbf{a} παρουσιάζεται στο κατώτατο σημείο του σχήματος. Στο παραπάνω σχήμα R είναι ο αριθμός στοιχείων στο διάνυσμα εισόδου, ενώ S είναι ο αριθμός νευρώνων στο στρώμα. Είναι σύνηθες ο αριθμός εισόδων να είναι διαφορετικός από τον αριθμό νευρώνων.

Μπορεί να οριστεί ένα σύνθετο στρώμα νευρώνων που να έχει διαφορετικές συναρτήσεις μεταφοράς βάζοντας δυο απλά δίκτυα παράλληλα. Και τα 2 δίκτυα θα πρέπει να έχουν τις ίδιες εισόδους, και κάθε δίκτυο θα παράγει κάποιες από τις εξόδους. Η είσοδος γίνεται με τη βοήθεια ενός πίνακα βαρών \mathbf{W} :

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,R} \\ w_{2,1} & w_{2,2} & \dots & w_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ w_{S,1} & w_{S,2} & \dots & w_{S,R} \end{bmatrix}$$

Ο δείκτης γραμμών σε κάθε στοιχείο καθορίζει το νευρώνα για τον οποίο προορίζεται το συγκεκριμένο στοιχείο-βάρος ενώ ο δείκτης στήλης ποια είσοδος ταιριάζει με το συγκεκριμένο βάρος. Συνεπώς, οι δείκτες στο $w_{1,2}$ υποδηλώνουν ότι η ισχύς του σήματος από την δεύτερη είσοδο στον πρώτο νευρώνα είναι $w_{1,2}$.

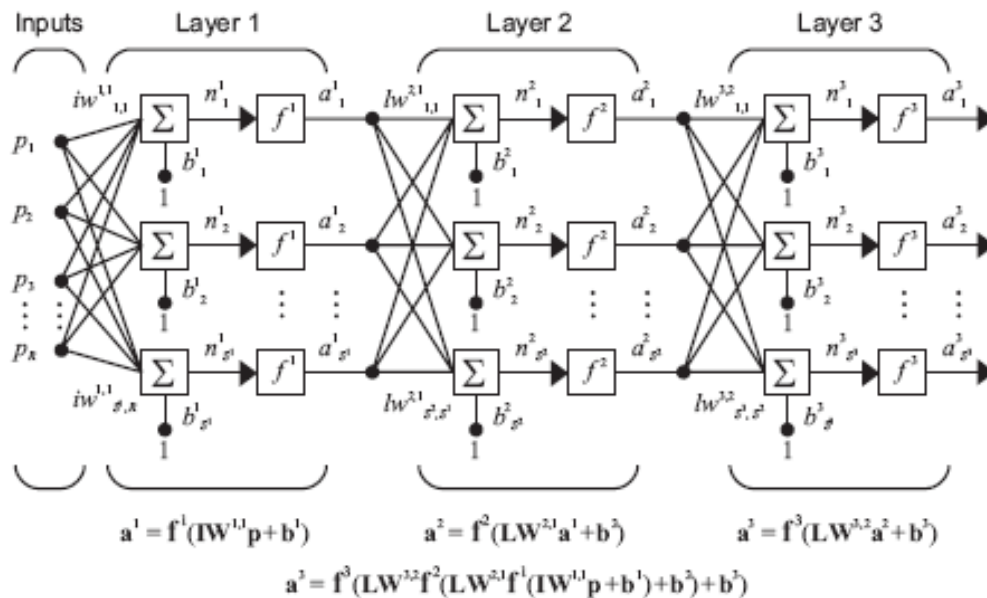
Το δίκτυο που αποτελείται από S νευρώνες και R εισόδους μπορεί επίσης να σχεδιαστεί με συντομία ως εξής:



Σχήμα 15 Περιεκτική περιγραφή στρώματος

Εδώ το p είναι ένα διάνυσμα εισόδου με μήκος R και W είναι ένας πίνακας $S \times R$, και τα a και b είναι διανύσματα μήκους S . Το δίκτυο αυτό περιλαμβάνει το πίνακα βαρών, τις πράξεις πολλαπλασιασμού, το διάνυσμα πόλωσης b , τον αθροιστή και τα κουτιά των συναρτήσεων μεταφοράς.

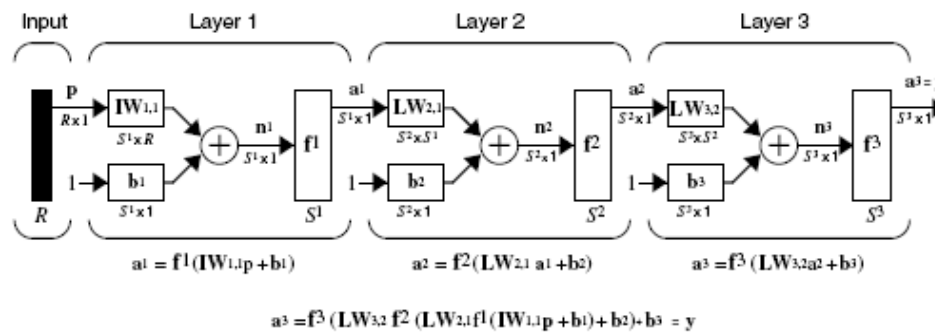
Ένα δίκτυο μπορεί να έχει πολλαπλά στρώματα νευρώνων. Κάθε στρώμα έχει ένα πίνακα βαρών W , το διάνυσμα πόλωσης b και το διάνυσμα εξόδου a . Για να ξεχωρίσει κανείς τους πίνακες βαρών, τα διανύσματα πόλωσης κτλ του κάθε στρώματος προστίθεται ο αριθμός του στρώματος σαν άνω κείμενο σε κάθε μεταβλητή. Ο συμβολισμός αυτός φαίνεται στο παρακάτω σχήμα:



Σχήμα 16 Δίκτυο πολλαπλών στρωμάτων

Το παραπάνω δίκτυο έχει R^1 εισόδους, S^1 νευρώνες στο πρώτο στρώμα, S^2 νευρώνες στο δεύτερο στρώμα κτλ. Είναι σύνηθες κάθε στρώμα να έχει διαφορετικό αριθμό νευρώνων. Οι εξοδοί κάθε ενδιάμεσου στρώματος είναι οι εισοδοί του επόμενου.

Τα στρώματα σε ένα δίκτυο νευρώνων παίζουν διαφορετικό ρόλο. Το στρώμα που παράγει την έξοδο του δικτύου ονομάζεται στρώμα εξόδου. Όλα τα υπόλοιπα στρώματα ονομάζονται κρυφά.



Σχήμα 17 Δίκτυο πολλαπλών στρώματων

Τα δίκτυα που έχουν πολλαπλά στρώματα έχουν σημαντικές δυνατότητες. Για παράδειγμα, ένα δίκτυο με δυο στρώματα, όπου το πρώτο στρώμα έχει συνάρτηση μεταφοράς sigmoid και το δεύτερο γραμμική, μπορεί να εκπαιδευτεί ώστε να προσεγγίζει οποιαδήποτε συνάρτηση (με πεπερασμένο αριθμό ασυνεχειών) αρκετά καλά. Στο Neural Network Toolbox η έξοδος του δικτύου συμβολίζεται με y .

5.5. Εκπαίδευση Δικτύου Νευρώνων

Το τμήμα αυτό περιγράφει δύο διαφορετικές μορφές εκπαίδευσης. Την επαυξητική εκπαίδευση και την εκπαίδευση σε δεσμίδες. Στην επαυξητική εκπαίδευση (Incremental training) τα βάρη και οι πολώσεις του δικτύου ενημερώνονται κάθε φορά που εμφανίζεται μια νέα είσοδος στο δίκτυο. Στην εκπαίδευση που γίνεται σε δεσμίδες (batch training) τα βάρη και οι πολώσεις κάθε νευρώνα ενημερώνονται μόνο αφότου παρουσιάζονται όλες οι εισόδοι.

5.5.1. Επαυξητική εκπαίδευση

Η επαυξητική εκπαίδευση μπορεί να εφαρμοστεί τόσο στα στατικά όσο και στα δυναμικά δίκτυα, αν και συνήθως χρησιμοποιείται στα δυναμικά δίκτυα, όπως είναι τα προσαρμοστικά φίλτρα. Σε αυτή την ενότητα παρουσιάζεται η απόδοση της επαυξητικής εκπαίδευσης τόσο στα στατικά όσο και στα δυναμικά δίκτυα.

5.5.1.1. Επαυξητική εκπαίδευση σε στατικά δίκτυα

Έστω ένα στατικό δίκτυο το οποίο πρέπει να εκπαιδευτεί επαυξητικά, έτσι ώστε μετά από κάθε είσοδο που θα προκύπτει, τα βάρη και οι πολώσεις του δικτύου να ενημερώνονται. Σε αυτή την περίπτωση χρησιμοποιείται η συνάρτηση “adapt”, και οι εισόδοι και οι έξοδοι εμφανίζονται σαν ακολουθίες.

Έστω ότι πρέπει να εκπαιδευτεί το δίκτυο έτσι ώστε να παραχθεί η παρακάτω γραμμική συνάρτηση

$$t = 2p_1 + p_2$$

Τότε για τις εισόδους:

$$p_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, p_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, p_3 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, p_4 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Οι αντίστοιχες εξόδοι είναι

$$t_1 = [4], t_2 = [5], t_3 = [7], t_4 = [7]$$

Στην επαυξητική εκπαίδευση, οι είσοδοι εμφανίζονται σαν ακολουθίες:

$$P = \{[1;2] [2;1] [2;3] [3;1]\};$$

$$T = \{4 \ 5 \ 7 \ 7\};$$

Πρώτον, αρχικοποιείται το δίκτυο με μηδενικά βάρη και πολώσεις. Επίσης, ο αρχικός ρυθμός εκπαίδευσης τίθεται ίσος με το μηδέν ώστε να φανεί η επίδραση της επαυξητικής εκπαίδευσης.

```
net = newlin(P,T,0,0);
net.IW{1,1} = [0 0];
net.b{1} = 0;
```

Υπάρχουν δυο βασικοί τύποι διανυσμάτων εισόδου: αυτά που εμφανίζονται ταυτόχρονα (την ίδια χρονική στιγμή ή σε όχι συγκεκριμένες χρονικές ακολουθίες) και σε αυτά που εμφανίζονται ακολουθιακά στο χρόνο. Για τα πρώτα, η σειρά δεν είναι σημαντική, και στην περίπτωση που υπάρχει ένα πλήθος δικτύων που τρέχουν παράλληλα, μπορεί να εμφανίζεται ένα διάνυσμα εισόδου για κάθε ένα από αυτά τα δίκτυα. Για τα δεύτερα, η σειρά με την οποία τα διανύσματα εμφανίζονται είναι σημαντική.

Για ένα στατικό δίκτυο, η προσομοίωση του δικτύου παράγει τις ίδιες εξόδους είτε οι είσοδοι εμφανίζονται σαν ένας πίνακας “ταυτόχρονων” διανυσμάτων είτε σαν πίνακας κελιών διαδοχικών διανυσμάτων. Ωστόσο, αυτό δεν επαληθεύεται κατά την εκπαίδευση ενός δικτύου. Όταν χρησιμοποιείται η συνάρτηση “adapt”, εάν οι είσοδοι εμφανίζονται σαν ένας πίνακας κελιών διαδοχικών διανυσμάτων, τότε τα βάρη ενημερώνονται κάθε φορά που προκύπτει μια νέα είσοδος (επαυξητική μέθοδος). Όπως φαίνεται και στην επόμενη ενότητα, εάν οι είσοδοι εμφανίζονται σαν ένας πίνακας “ταυτόχρονων” διανυσμάτων, τότε τα βάρη ενημερώνονται μόνο όταν παρουσιαστούν όλες οι είσοδοι (μέθοδος δεσμίδων).

Τώρα μπορεί πλέον να αρχίσει η εκπαίδευση του δικτύου επαυξητικά.

```
[net,a,e,pf] = adapt(net,P,T);
```

Οι εξόδοι του δικτύου παραμένουν μηδέν, διότι ο ρυθμός εκπαίδευσης και τα βάρη δεν άλλαξαν. Τα λάθη είναι ίδια με τις εξόδους:

```
a = [0] [0] [0] [0]
e = [4] [5] [7] [7]
```

Εάν τώρα τεθεί ο ρυθμός εκπαίδευσης ίσος με 0,1 τότε μπορεί κανείς να δει πως προσαρμόζεται το δίκτυο μετά από κάθε νέα είσοδο που προκύπτει:

```

net.inputWeights{1,1}.learnParam.lr=0.1;
net.biases{1,1}.learnParam.lr=0.1;
[net,a,e,pf] = adapt(net,P,T);
a = [0] [2] [6] [5.8]
e = [4] [3] [1] [1.2]

```

Η πρώτη έξοδος είναι η ίδια με την έξοδο που είχε προκύψει όταν ο ρυθμός εκπαίδευσης ήταν μηδέν, διότι δεν προέκυψε καμία ενημέρωση μέχρι που να εμφανιστεί η πρώτη είσοδος. Η δεύτερη έξοδος είναι διαφορετική, διότι τα βάρη έχουν ενημερωθεί. Τα βάρη συνεχίζουν να μεταβάλλονται κάθε φορά που υπολογίζεται ένα λάθος. Εάν το δίκτυο είναι ικανό και ο ρυθμός εκπαίδευσης τίθενται σωστά, το λάθος τελικά τείνει στο μηδέν.

5.5.1.2. Επαναληπτική εκπαίδευση σε δυναμικά δίκτυα

Μπορεί επίσης να εκπαιδευτεί επαναληπτικά και ένα δυναμικό δίκτυο. Ουσιαστικά, αυτή είναι και η πιο συνηθισμένη περίπτωση.

Παρακάτω παρουσιάζονται οι αρχικές εισοδοί P_i , οι εισοδοί P και οι έξοδοι T σαν στοιχεία του πίνακα κελιών.

```

Pi = {1};
P = {2 3 4};
T = {3 5 7};

```

Έστω ότι δημιουργείται ένα γραμμικό δίκτυο με μια καθυστέρηση στην είσοδο, όπως χρησιμοποιήθηκε και στο προηγούμενο παράδειγμα. Τα βάρη αρχικοποιούνται στο μηδέν και ο ρυθμός εκπαίδευσης στο 0,1.

```

net = newlin(P,T,[0 1],0.1);
net.IW{1,1} = [0 0];
net.biasConnect = 0;

```

Έστω ότι το δίκτυο πρέπει να εκπαιδευτεί έτσι ώστε η τρέχουσα έξοδος να προκύπτει από το άθροισμα της τρέχουσας και της προηγούμενης εισόδου. Πρόκειται για την ίδια ακολουθία εισόδου που χρησιμοποιήθηκε στο προηγούμενο παράδειγμα (με τη χρήση του `sim`) με την διαφορά ότι ως αρχικός περιορισμός για την καθυστέρηση (`delay`) ορίζεται ο πρώτος όρος της ακολουθίας. Στη συνέχεια το δίκτυο μπορεί να εκπαιδευτεί διαδοχικά με τη χρήση της “`adapt`”.

```

[net,a,e,pf] = adapt(net,P,T,Pi);
a = [0] [2.4] [7.98]
e = [3] [2.6] [-0.98]

```

Η πρώτη έξοδος είναι το μηδέν, διότι τα βάρη δεν έχουν ενημερωθεί ακόμη. Τα βάρη αλλάζουν σε κάθε επόμενο χρονικά βήμα.

5.5.2 Εκπαίδευση σε δεσμίδες

Η εκπαίδευση σε δεσμίδες, στην οποία τα βάρη και οι πολώσεις ενημερώνονται μόνο εφόσον έχουν παρουσιαστεί όλες οι εισοδοί και οι έξοδοι, μπορεί να εφαρμοστεί τόσο στα στατικά όσο και στα δυναμικά δίκτυα. Στη συνέχεια μελετώνται οι δυο αυτές περιπτώσεις.

5.5.2.1 Εκπαίδευση σε δεσμίδες σε στατικό δίκτυο

Η εκπαίδευση σε δεσμίδες μπορεί να γίνει με τη χρήση είτε της “adapt” είτε της “train”, αν και η “train” είναι γενικά η καλύτερη επιλογή, διότι τυπικά έχει πρόσβαση σε περισσότερους αποτελεσματικούς αλγόριθμους εκπαίδευσης. Η επαυξητική εκπαίδευση μπορεί να πραγματοποιηθεί μόνο με τη χρήση της “adapt” και η εκπαίδευση σε δεσμίδες μόνο με τη χρήση της “train”.

Για την εκπαίδευση σε δεσμίδες ενός στατικού δικτύου με τη χρήση της “adapt”, τα διανύσματα εισόδου πρέπει να είναι σε μορφή ενός πίνακα συντρεχουσών διανυσμάτων.

```
P = [1 2 2 3; 2 1 3 1];
T = [4 5 7 7];
```

Έστω το στατικό δίκτυο που χρησιμοποιήθηκε στα προηγούμενα παραδείγματα. Ο ρυθμός εκπαίδευσης τίθενται ίσος με 0,1.

```
net = newlin(P,T,0,0.1);
net.IW{1,1} = [0 0];
net.b{1} = 0;
```

Όταν καλείται η “adapt”, καλείται η “trains” (η εξορισμού συνάρτηση προσαρμογής για τα γραμμικά δίκτυα) και η “learnwh” (η εξορισμού συνάρτηση εκπαίδευσης για τα βάρη και τις πολώσεις). Η “trains” χρησιμοποιεί την εκπαίδευση “Widrow-Hoff” [Widr60]

```
[net,a,e,pf] = adapt(net,P,T);
a = 0 0 0 0
e = 4 5 7 7
```

Να σημειωθεί ότι οι όλες οι εξοδοί του δικτύου είναι μηδέν, διότι τα βάρη δεν έχουν ενημερωθεί έως ότου παρουσιαστούν όλα τα σύνολα εκπαίδευσης. Για τα βάρη λοιπόν προκύπτει ότι:

```
»net.IW{1,1}
ans = 4.9000    4.1000
»net.b{1}
ans =
    2.3000
```

Τα αποτελέσματα αυτά είναι διαφορετικά από αυτά που προέκυψαν μετά από το πρώτο πέρασμα της “adapt” στην επαυξητική εκπαίδευση.

Τώρα παρουσιάζεται η ίδια εκπαίδευση σε δεσμίδες με τη χρήση όμως την “train”. Επειδή ο κανόνας του “Widrow-Hoff” μπορεί να εφαρμοστεί είτε στην επαυξητική μέθοδο είτε στη μέθοδο σε δεσμίδες, μπορεί να καλεστεί τόσο από την “adapt” όσο και από την “train”. (Υπάρχουν αρκετοί αλγόριθμοι οι οποίοι μπορούν να χρησιμοποιηθούν μόνο στην μέθοδο σε δεσμίδες, επομένως αυτοί οι αλγόριθμοι μπορούν να καλεστούν μόνο από την “train”).

Σε αυτή την περίπτωση, τα διανύσματα εισόδου πρέπει να είναι με τη μορφή “ταυτόχρονων” διανυσμάτων ή με τη μορφή πίνακα κελιών διαδοχικών διανυσμάτων. Επειδή το δίκτυο είναι στατικό και επειδή η “train” πάντα λειτουργεί στη μέθοδο σε δεσμίδες, η “train” μετατρέπει οποιοδήποτε πίνακα κελιών διαδοχικών διανυσμάτων σε πίνακα “ταυτόχρονων” διανυσμάτων. Η κατάσταση ταυτόχρονης λειτουργίας χρησιμοποιείται όπου είναι δυνατόν επειδή έχει μια πιο αποτελεσματική εφαρμογή στον κώδικα MATLAB:

```
P = [1 2 2 3; 2 1 3 1];
T = [4 5 7 7];
```

Το δίκτυο στήνεται με τον ίδιο τρόπο.

```
net = newlin(P,T,0,0.1);
net.IW{1,1} = [0 0];
net.b{1} = 0;
```

Τώρα το δίκτυο είναι πλέον έτοιμο να εκπαιδευτεί. Η εκπαίδευση γίνεται μόνο για μια επανάληψη, διότι χρησιμοποιείται μόνο ένα πέρασμα της “adapt”. Η εξορισμού συνάρτηση εκπαίδευσης για τα γραμμικά δίκτυα είναι η trainb, και η εξορισμού συνάρτηση εκπαίδευσης για τα βάρη και τις πολώσεις είναι η learnwh, έτσι ώστε να προκύψουν τα ίδια αποτελέσματα με αυτά που προέκυψαν με τη χρήση της “adapt” στο προηγούμενο παράδειγμα, όπου η εξορισμού συνάρτηση προσαρμογής ήταν η “trains”.

```
net.inputWeights{1,1}.learnParam.lr = 0.1;
net.biases{1}.learnParam.lr = 0.1;
net.trainParam.epochs = 1;
net = train(net,P,T);
```

Μετά από μια επανάληψη προκύπτει:

```
»net.IW{1,1}
ans = 4.9000    4.1000
»net.b{1}
ans =
    2.3000
```

Πρόκειται για τα ίδια αποτελέσματα με αυτά που προέκυψαν από την εκπαίδευση με τη μέθοδο των δεσμίδων με τη χρήση της “adapt”. Στα στατικά δίκτυα, η συνάρτηση “adapt” μπορεί να υλοποιήσει επαυξητική εκπαίδευση ή εκπαίδευση με δεσμίδες, ανάλογα με το είδος των δεδομένων εισόδου. Εάν τα δεδομένα είναι σε μορφή πίνακα “ταυτόχρονων” διανυσμάτων, τότε εφαρμόζεται η εκπαίδευση σε δεσμίδες. Εάν τα δεδομένα είναι σε μορφή διανύσματος τότε εφαρμόζεται η επαυξητική εκπαίδευση. Αυτό δεν ισχύει για την train, η οποία εκτελεί πάντα εκπαίδευση σε δεσμίδες, ανεξάρτητα από τη μορφή των δεδομένων εισόδου.

5.5.2.2 Εκπαίδευση σε δεσμίδες σε δυναμικό δίκτυο

Η εκπαίδευση των στατικών δικτύων, είναι σχετικά απλή. Εάν χρησιμοποιηθεί η “train” το δίκτυο εκπαιδεύεται με τη μέθοδο των δεσμίδων και οι εισοδοί μετατρέπονται σε “ταυτόχρονα” διανύσματα (στήλες του πίνακα), ακόμη και στην περίπτωση που αρχικά είχαν τη μορφή ακολουθίας (στοιχεία ενός πίνακα κελιών). Εάν χρησιμοποιηθεί η “adapt”, τότε η μορφή των δεδομένων εισόδου καθορίζει την μέθοδο εκπαίδευσης. Εάν τα δεδομένα εισόδου έχουν τη μορφή ακολουθίας, τότε το δίκτυο εκπαιδεύεται με τη επαυξητική μέθοδο. Εάν τα δεδομένα εισόδου έχουν τη μορφή συντρεχουσών διανυσμάτων, τότε το δίκτυο εκπαιδεύεται με τη μέθοδο των δεσμίδων.

Με τα δυναμικά δίκτυα, η εκπαίδευση με τη μέθοδο των δεσμίδων επιτυγχάνεται κυρίως με τη χρήση της “train” μόνο, ειδικά εάν υπάρχει μόνο μια ακολουθία για εκπαίδευση. Για να φανεί αυτό, ας θεωρηθεί το γραμμικό δίκτυο με καθυστέρηση και ας ληφθεί ένας ρυθμός εκπαίδευσης ίσος με 0.02 για την εκπαίδευση.

```

Pi = {1};
P = {2 3 4};
T = {3 5 6};
net = newlin(P,T,[0 1],0.02);
net.IW{1,1} = [0 0];
net.biasConnect = 0;
net.trainParam.epochs = 1;

```

Έστω ότι το δίκτυο εκπαιδεύεται με την ίδια ακολουθία που χρησιμοποιήθηκε για την επαυξητική μέθοδο προηγούμεως, με τη διαφορά ότι τώρα τα βάρη ενημερώνονται μόνο όταν όλες οι εισοδοί εμφανιστούν (μέθοδος σε δεσμίδες). Το δίκτυο προσομοιώνεται σε λειτουργία ακολουθίας, επειδή η είσοδος είναι μια ακολουθία, αλλά τα βάρη ενημερώνονται με τη μέθοδο των δεσμίδων.

```
net = train(net,P,T,Pi);
```

Τα βάρη που προκύπτουν μετά από μια επανάληψη είναι

```

»net.IW{1,1}
ans = 0.9000    0.6200

```

Τα βάρη αυτά είναι διαφορετικά από αυτά που θα προέκυπταν εάν είχε χρησιμοποιηθεί η επαυξητική μέθοδος, όπου τα βάρη θα ενημερώνονταν τρεις φορές κατά τη διάρκεια ενός περάσματος στο σύνολο προς εκπαίδευση. Για την εκπαίδευση σε δεσμίδες τα βάρη ενημερώνονται μόνο μια φορά σε κάθε επανάληψη.

5.5.3 Αλγόριθμος εκπαίδευσης με ελαστική επανατροφοδότηση (*trainrp*)

Τα πολυστρωματικά δίκτυα συνήθως χρησιμοποιούν σιγμοειδής συναρτήσεις μεταφοράς στα κρυφά στρώματα. Αυτές οι συναρτήσεις συχνά ονομάζονται συναρτήσεις συμπίεσης (“squashing”), επειδή συμπιέζουν την είσοδο με άπειρο διάστημα σε έξοδο με πεπερασμένο διάστημα. Οι σιγμοειδείς συναρτήσεις χαρακτηρίζονται από το γεγονός ότι οι κλίσεις τους πρέπει να τείνουν στο μηδέν όσο η είσοδος μεγαλώνει. Αυτό προκαλεί πρόβλημα όταν χρησιμοποιείται απότομη αρνητική κλίση (descent) για την εκπαίδευση ενός πολυστρωματικού δικτύου με σιγμοειδής συναρτήσεις, διότι η κλίση μπορεί να είναι πολύ μικρή με αποτέλεσμα να προκύπτουν μικρές αλλαγές στα βάρη και στις πολώσεις, ακόμη και όταν τα βάρη και οι πολώσεις απέχουν αρκετά από τις βέλτιστες τιμές τους.

Ο σκοπός του αλγόριθμου εκπαίδευσης με ελαστική επανατροφοδότηση (*trainrp*) είναι η εξάλειψη αυτών των προβλημάτων από τις μερικές παραγώγους. Μόνο το πρόσημο των παραγώγων μπορεί να καθορίσει την κατεύθυνση της ενημέρωσης βάρους. Το μέγεθος της παραγώγου δεν έχει καμία επιρροή στην ενημέρωση των βαρών. Το μέγεθος της αλλαγής του βάρους προσδιορίζεται από μια ξεχωριστή ενημερωμένη τιμή. Η ενημερωμένη τιμή για κάθε βάρους και πόλωση αυξάνεται κατά έναν παράγοντα *delt_inc* κάθε φορά που η παράγωγος της συνάρτησης απόδοσης (performance) σε σχέση με το βάρους έχει το ίδιο πρόσημο για δυο διαδοχικές επαναλήψεις. Η ενημερωμένη τιμή μειώνεται κατά έναν παράγοντα *delt_dec* κάθε φορά που η παράγωγος σε σχέση με το συγκεκριμένο βάρους αλλάζει πρόσημο από την προηγούμενη επανάληψη. Εάν η παράγωγος είναι μηδέν, η ενημερωμένη τιμή παραμένει ίδια. Όταν τα βάρη αυξομειώνονται, η αλλαγή βάρους μειώνεται. Εάν τα βάρη εξακολουθούν

να αλλάζουν προς την ίδια κατεύθυνση για αρκετές επαναλήψεις το μέγεθος των αλλαγών βάρους αυξάνεται.

5.5.4 Αλγόριθμος εκπαίδευσης *Levenberg-Marquardt (trainlm)*

Ο αλγόριθμος Levenberg-Marquardt κατασκευάστηκε με σκοπό να επιτυγχάνει μεγάλες ταχύτητες εκπαίδευσης χωρίς να χρειάζεται ο υπολογισμός του πίνακα Hessian. Όταν η συνάρτηση απόδοσης έχει τη μορφή αθροίσματος τετραγώνων, τότε για τον πίνακα Hessian μπορεί να γίνει η παρακάτω προσέγγιση

$$\mathbf{H} = \mathbf{J}^T \mathbf{J}$$

και η κλίση μπορεί να υπολογιστεί σύμφωνα με τον παρακάτω τύπο

$$\mathbf{g} = \mathbf{J}^T \mathbf{e}$$

όπου \mathbf{J} είναι ο Ιακωβιανός πίνακας ο οποίος περιέχει τις πρώτες παραγώγους των λαθών του δικτύου σε σχέση με τα βάρη και τις πολώσεις, και \mathbf{e} είναι ένα διάνυσμα των λαθών του δικτύου. Ο Ιακωβιανός πίνακας μπορεί να υπολογιστεί με τη βοήθεια μιας τυπικής “backpropagation” τεχνικής η οποία είναι λιγότερο πολύπλοκη από τον υπολογισμό του πίνακα Hessian.

Ο αλγόριθμος Levenberg-Marquardt χρησιμοποιεί αυτή την προσέγγιση στον πίνακα Hessian στον παρακάτω τύπο του Newton:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{J}^T \mathbf{J} + \mu \mathbf{I}]^{-1} \mathbf{J}^T \mathbf{e}$$

όταν η μεταβλητή-αριθμός μ είναι μηδέν τότε πρόκειται για τη μέθοδο Newton με τη χρήση της προσέγγισης του πίνακα Hessian. Η μέθοδος του Newton είναι πιο γρήγορη και πιο ακριβής κοντά σε ένα ελάχιστο λάθος, οπότε ο στόχος είναι η χρήση αυτής της μεθόδου. Έτσι, το μ μειώνεται μετά από κάθε επιτυχημένο βήμα (μείωση στη συνάρτηση απόδοσης) και αυξάνεται μόνο όταν ένα βήμα θα μπορούσε να αυξήσει την συνάρτηση απόδοσης. Με αυτόν τον τρόπο, η συνάρτηση απόδοσης πάντα μειώνεται σε κάθε επανάληψη του αλγορίθμου.

6. Ανάλυση μελωδίας: Θεωρία και Εργαλεία

6.1. Εισαγωγή στη θεωρία των μουσικών οργάνων

Όλα τα ελαστικά σώματα που δονούνται παράγουν ήχο: μια μεταλλική χορδή, ένα έλασμα, μια ελαστική μεμβράνη, κλπ. Πολλά αντικείμενα της καθημερινής μας ζωής έχουν αυτή την ιδιότητα και ο ήχος που παράγουν περιέχει μερικές φορές ίχνη από μουσικές νότες. Ο τύπος του ήχου που παράγουν αυτές οι ταλαντώσεις εξαρτάται από δύο παράγοντες: την ταχύτητα της ίδιας της ταλάντωσης και τον τρόπο με τον οποίο το σώμα ταλαντώνεται. Αν μπορεί να ελεγχθεί ο τρόπος με τον οποίο πάλλεται ένα αντικείμενο τότε μπορεί να ελεγχθεί και ο ήχος που παράγει. Απλό παράδειγμα αποτελούν όσοι παίζουν μουσική χρησιμοποιώντας κρυστάλλινα ποτήρια με νερό. Αφού γεμίσουν κάθε ποτήρι με διαφορετική ποσότητα νερού, το κάνουν να δονείται διαφορετικά, παράγοντας κατ' αυτό τον τρόπο μια διαφορετική νότα. Η ίδια αρχή εφαρμόζεται και στην κατασκευή των ηχείων στα στερεοφωνικά συγκροτήματα, στο εσωτερικό των οποίων τοποθετούνται δύο πλαστικοί δίσκοι (κόνιοι) που πάλλονται με διάφορους τρόπους. Το ζήτημα είναι σε τι διαφέρουν οι ήχοι που παράγονται από ποτήρια με διαφορετικές ποσότητες νερού. Ουσιαστικά αυτό που αλλάζει είναι η συχνότητα του ήχου, δηλαδή ο αριθμός των κύκλων από ηχητικές ταλαντώσεις μέσα σ' ένα δευτερόλεπτο - αυτό που ονομάζεται Χερτς (Hz). Κάθε μουσική νότα χαρακτηρίζεται από μια συγκεκριμένη συχνότητα, έτσι όπως τα χρώματα χαρακτηρίζονται από μια συγκεκριμένη συχνότητα φωτός. Η συχνότητα του "λα", της νότας πάνω στην οποία κουρδίζονται τα μουσικά όργανα στη Δύση, είναι 440 Hz. Από ακουστικής άποψης όσο πιο ψηλή είναι η συχνότητα τόσο πιο οξύς είναι ο ήχος.

Μιλώντας για ήχους και συχνότητες, μέχρι στιγμής εξετάστηκαν περιπτώσεις που θα μπορούσαν να χαρακτηριστούν ως ιδανικές. Πράγματι, υπάρχει διάφορα ανάμεσα στον καθαρό, θεωρητικό ήχο και στον πραγματικό. Έστω για παράδειγμα η νότα "λα" των 440 Hz. Ακούγεται διαφορετικά από μια κιθάρα και διαφορετικά από ένα πιάνο, μια τρομπέτα ή από τη φωνή ενός τραγουδιστή. Αυτό συμβαίνει επειδή κανένα μουσικό όργανο δεν παράγει αποκλειστικά μια καθαρή δόνηση 440 κύκλων το δευτερόλεπτο. Το κύμα που παράγει είναι πολύ πιο σύνθετο και περιέχει μεταξύ άλλων μια σειρά από ήχους ασθενέστερης έντασης, η συχνότητα των οποίων έχει κάποια αριθμητική σχέση με τη συχνότητα του βασικού ήχου (διπλάσια, τριπλάσια κ.ο.κ.). Αυτές οι επιπλέον συχνότητες ονομάζονται αρμονικές και είναι εκείνες που κάνουν κάθε ήχο ιδιαίτερο και μοναδικό. Αν λείπουν οι άρτιες αρμονικές, εκείνες με συχνότητα διπλάσια, τετραπλάσια κ.λπ., ο ήχος θα έχει τη χαρακτηριστική χροιά που ονομάζουμε ένρινη.

Ο πρώτος που μελέτησε τις αρμονικές των ήχων ήταν ο Ζαν-Μπατίστ Ζοζέφ Φουριέ, το 19ο αιώνα. Ανακάλυψε ότι ο κάθε ήχος στη φύση μπορεί να αναλυθεί σ' ένα συγκεκριμένο αριθμό παλμικών κινήσεων που αντιστοιχούν σε μια θεμελιώδη συχνότητα και επιπλέον, σ' έναν αριθμό από δευτερεύοντες ήχους, τις αρμονικές, που είναι πολλαπλάσια της θεμελιώδους συχνότητας. Αυτή η αρχή έχει αποφασιστική σημασία για τη μουσική. Κι αυτό γιατί το πόσες και ποιας ποιότητας αρμονικές εμπεριέχει ένας ήχος, αυτό που ονομάζεται χροιά, είναι εκείνο που καθορίζει τη διαφορά ανάμεσα στις "φωνές" των διαφόρων μουσικών οργάνων.

Σε αυτό το τμήμα της εργασίας παρουσιάζονται μουσικά παραδείγματα και εργαλεία του Matlab που μπορούν να χρησιμοποιηθούν για τη δημιουργία και ανάλυση μουσικής. Η μελέτη του μετασχηματισμού Fourier από μια μουσική προοπτική προσφέρει μεγάλη διορατικότητα στις βασικές μαθηματικές έννοιες και τη φυσική των μουσικών οργάνων. Το Matlab διαθέτει πολλά εργαλεία τα οποία επιτρέπουν την ανάλυση μιας μελωδίας στις

κυματομορφές και τις αρμονικές των καταγραμμένων ήχων καθώς και τη σύνθεση νέων ήχων.

Το κύριο πρόβλημα είναι η εξίσωση κύματος μιας χορδής με συνοριακές συνθήκες Dirichlet:

$$\begin{cases} u_{tt} = c^2 u_{xx}, & x \in (0, L), t \geq 0 \\ u(0, t) = u(L, t) = 0, & t \geq 0 \\ u(x, 0) = \alpha(x), u_t(x, 0) = \beta(x), & x \in (0, L). \end{cases}$$

Από τη πλευρά της φυσικής, μπορεί να σκεφτεί κανείς το $u(x, t)$ ως τη μετατόπιση μιας χορδής κιθάρας με αρχική μετατόπιση $\alpha(x)$ και αρχική ταχύτητα $\beta(x)$. Επιπλέον περιορισμοί μπορούν να προστεθούν στην εξίσωση κύματος για την ακαμψία της χορδής και την εσωτερική απόσβεση. Η λύση, που λαμβάνεται από το χωρισμό των μεταβλητών, είναι

$$u(x, t) = \sum_{n=1}^{\infty} \sin \frac{n\pi}{L} x \left(a_n \cos \frac{n\pi}{L} ct + b_n \sin \frac{n\pi}{L} ct \right),$$

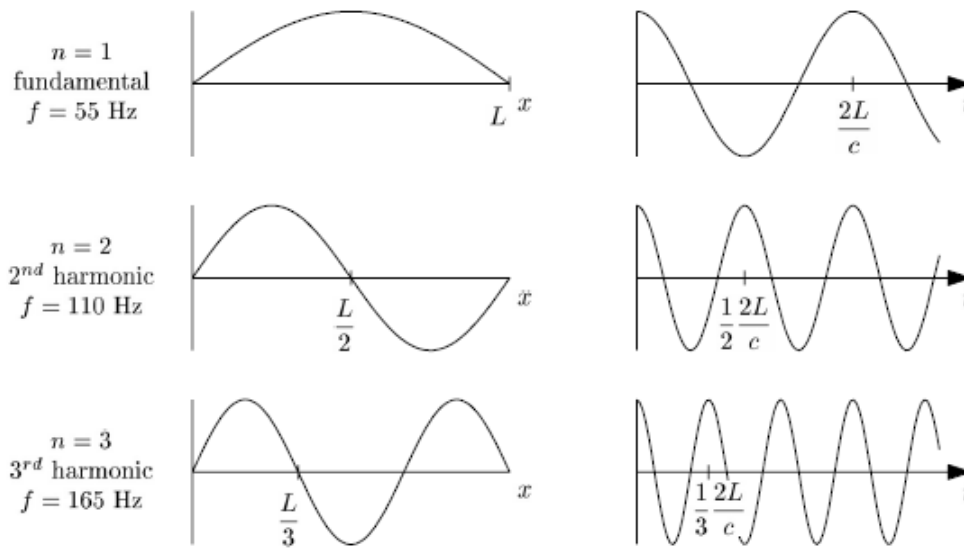
όπου

$$a_n = \frac{2}{\pi} \int_0^L \alpha(x) \sin \frac{n\pi}{L} x dx$$

$$b_n = \frac{2L}{cn\pi^2} \int_0^L \beta(x) \sin \frac{n\pi}{L} x dx$$

είναι οι συντελεστές Fourier των $\alpha(x)$ και $\beta(x)$. Χρησιμοποιώντας την ταυτότητα $\cos(X - Y) = \cos(X) \cos(Y) + \sin(X) \sin(Y)$ ο παραπάνω τύπος γίνεται:

$$u(x, t) = \sum_{n=1}^{\infty} p_n \sin \frac{n\pi}{L} x \cos \frac{n\pi}{L} c (t - \gamma_n).$$



Σχήμα 18: Τρόποι ταλάντωσης μιας χορδής

Η λύση είναι η επαλληλία πολλών τρόπων ταλάντωσης, με την κάθε ταλάντωση να βρίσκεται σε διαφορετική συχνότητα, όπως φαίνεται στο Σχήμα 18. Ο πρώτος παράγοντας στο άθροισμα, $\sin \frac{n\pi}{L}x$, είναι ένα στιγμιότυπο της δονούμενης χορδής. Ο δεύτερος παράγοντας, $\cos \frac{n\pi}{L}C(t - \gamma_n)$, δείχνει πως λειτουργεί η ταλάντωση με το χρόνο. Η n-οστή επαλληλία έχει συχνότητα $f_n = \frac{c}{2L}$, διαφορά φάσης γ_n και πλάτος $p_n = \sqrt{a_n^2 + b_n^2}$.

Ένας μουσικός θα ονόμαζε την f_1 ως θεμελιώδη συχνότητα (pitch) και f_2, f_3, \dots τις αρμονικές. Ο χαρακτηριστικός ήχος κάθε νότας συσχετίζεται με μια συγκεκριμένη θεμελιώδη συχνότητα. Για παράδειγμα, μια πολύ χαμηλή Ντο έχει θεμελιώδη συχνότητα 55 Hz και οι αρμονικές της έχουν 110, 165, 220, 275... Οι αρμονικές συχνότητες δημιουργούν μια αριθμητική ακολουθία, $f_n = nf_1$. Μέρος της ακολουθίας αυτής εμφανίζεται στο Σχήμα 18. Η ταχύτητα του κύματος c μπορεί να οριστεί ως η τετραγωνική ρίζα της τάσης της χορδής προς την γραμμική πυκνότητά [Dem98]. Συνεπώς ο χαρακτηριστικός ήχος μιας χορδής περιγράφεται από τρεις παραμέτρους:

$$f_1 = \frac{1}{2\text{μηνκος}} \sqrt{\frac{\text{τάση}}{\gamma\text{.πυκνότητα}}}$$

Τεντώνοντας μια χορδή αυξάνεται η θεμελιώδης συχνότητα, ο ήχος γίνεται πιο υψίσυχνος, ενώ επιλέγοντας πιο φαρδιές χορδές χαμηλώνει η θεμελιώδης συχνότητα. Παράδειγμα είναι ο κιθαρίστας ο οποίος πατώντας στα τάστα της κιθάρας μειώνει το μήκος, αλλάζει τον χαρακτηριστικό ήχο της χορδής και παίζει διαφορετική νότα.

Όταν μια χορδή ταλαντώνεται θέτει τον αέρα που την περικλείει σε κίνηση παράγοντας διαφορές στην πίεση. Αυτά είναι ηχητικά κύματα που ανιχνεύονται από το ανθρώπινο αυτί. Συνεπώς, ένα μικρόφωνο κοντά στη χορδή ανιχνεύει το κύμα σύμφωνα με τη σχέση:

$$y(t) = \sum_{n=1}^{\infty} p_n \cos 2\pi f_n (t - \gamma_n).$$

Από φυσική άποψη, τα πλάτη του κύματος p_n περιγράφουν την ενέργεια που συσχετίζεται με κάθε αρμονική. Το φάσμα ισχύος είναι η κατανομή της ενέργειας ανάμεσα στις αρμονικές, στην περίπτωση αυτή δηλαδή το διάνυσμα $[p_1, p_2, p_3, \dots]$. Για πραγματικούς ήχους, η ισχύς συγκεντρώνεται στις πρώτες 10 με 20 αρμονικές και το φάσμα ισχύος είναι μια συνεχής συνάρτηση της συχνότητας.

6.2. Ανάλυση Αρμονικών

Στο τμήμα αυτό θα μελετηθεί και θα γίνουν κάποια αρχικά πειράματα για το πώς αλλάζει το φάσμα των αρμονικών για διάφορες νότες (και όργανα) και πως η όλη αυτή διαδικασία μπορεί να γίνει μέσα από το Matlab.

```
file='test';

[y, Fs] = wavread(file);      % y:δεδομένα ήχου, Fs: συχνότητα δειγματοληψίας.
t = (1:length(y))/Fs;        % χρόνος

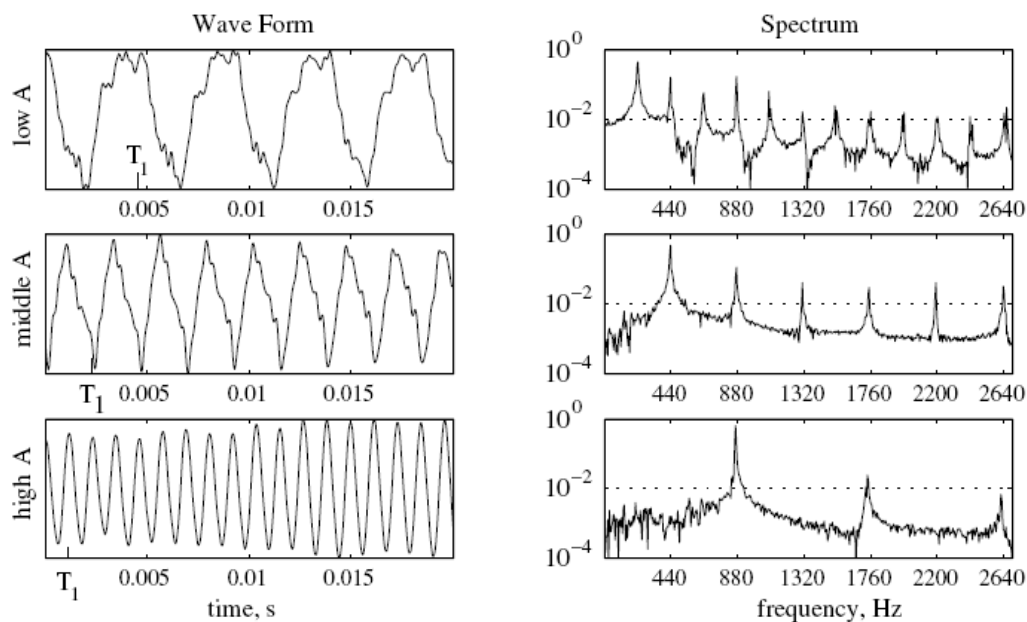
ind = find(t>0.1 & t<0.2);   % χρονικά όρια κυματομορφής
figure; subplot(2,1,1)
plot(t(ind), y(ind))
axis tight
title(['Waveform of ' file])

N = 2^12;                     % αριθμός σημείων προς ανάλυση
c = fft(y(1:N))/N;            % μετασχηματισμός Fourier
p = 2*abs(c(2:N/2));          % υπολογισμός ισχύος σε κάθε συχνότητα
f = (1:N/2-1)*Fs/N;          % συχνότητα που αντιστοιχεί σε κάθε ισχύ

subplot(2,1,2)                % εκτύπωση αποτελεσμάτων
semilogy(f,p)
axis([0 4000 10^-4 1])
title(['Power Spectrum of ' file])
```

Σχήμα 19: Απλή ανάλυση ήχου

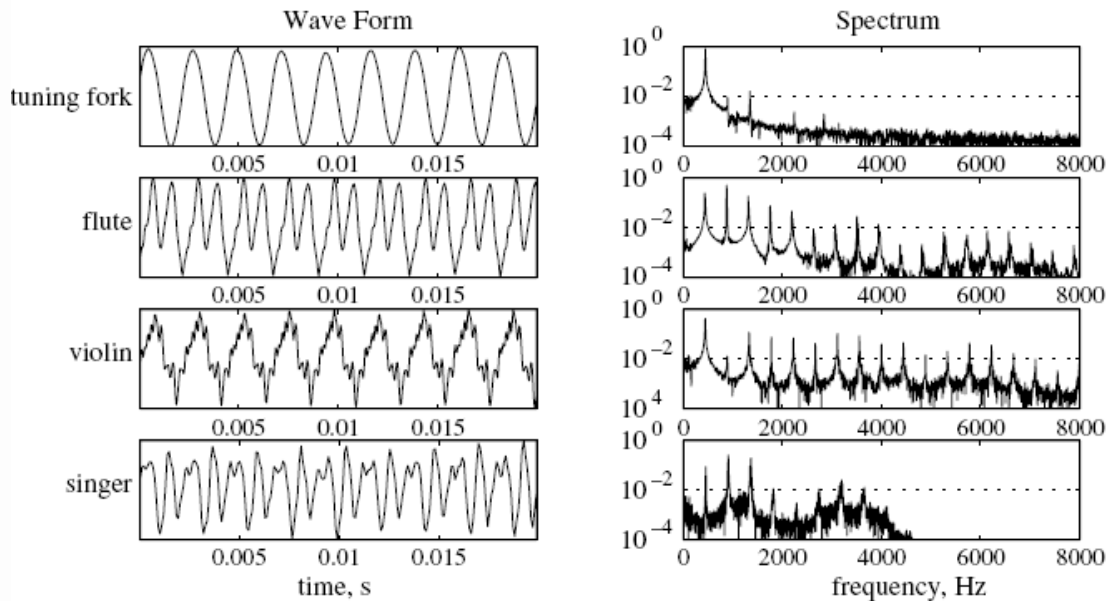
Στον παραπάνω κώδικα χρησιμοποιούνται οι συναρτήσεις `wavread` και `fft` για τον υπολογισμό του φάσματος ισχύος ενός μουσικού αρχείου (.wav). Οι καμπύλες που παράγονται από τον κώδικα αυτό χρησιμοποιούνται για τη ανίχνευση της θεμελιώδους συχνότητας και της έντασης του ήχου. Στο Σχήμα 20 φαίνονται τα αποτελέσματα για μια χαμηλή Ντο, μια μέση και μια ψηλή Ντο που παίζεται σε ένα πιάνο.



Σχήμα 20: Νότες πιάνου

Οι κυματομορφές στα αριστερά είναι οι διακυμάνσεις στην πίεση ως προς το χρόνο, όπως αυτές ανιχνεύονται από το μικρόφωνο. Το πλάτος του κύματος είναι ένα μέτρο της έντασης του ήχου. Η ένταση μετριέται σε decibels, που είναι ο λογάριθμος της πίεσης. Από τις κυματομορφές στο παραπάνω σχήμα μπορεί κανείς να εντοπίσει εύκολα τη περιοδική φύση κάθε ήχου και να επιλέξει την θεμελιώδη περίοδο T . Οι αντίστοιχες θεμελιώδεις συχνότητες $f_1=1/T_1$ είναι 220, 440 και 880 Hz αντίστοιχα, για τις τρεις νότες και φαίνονται ως οι πρώτες κορυφές στα διαγράμματα του φάσματος ισχύος. Η θεμελιώδης συχνότητα διπλασιάζεται σε κάθε μουσική οκτάβα και οι αρμονικές διατάσσονται σε αντιστοιχία με την f_1 . Το φάσμα ανάμεσα σε δυο αρμονικές είναι μη μηδενικό επειδή τα κύματα δεν είναι τελείως περιοδικά.

Ο ανθρώπινος εγκέφαλος μπορεί να αναγνωρίσει ουσιαστικά ένα απεριόριστο αριθμό οργάνων από τον ήχο τους και μόνο, ακόμα και αν παίζουν την ίδια θεμελιώδη συχνότητα στην ίδια ένταση. Στη συνέχεια θα εξεταστεί ποιο είναι αυτό το στοιχείο που τα διαχωρίζει. Το Σχήμα 21 παρουσιάζει την κυματομορφή και το φάσμα ισχύος διάφορων οργάνων που παίζουν την ίδια νότα. Η κυματομορφή φαίνεται τελείως διαφορετική σε κάθε περίπτωση, αλλά όλες έχουν την ίδια θεμελιώδη συχνότητα. Ο ήχος που παράγει το όργανο οφείλεται στην σχετική ισχύ των αρμονικών. Ένας καθαρός τόνος, όπως ο ήχος που κάνει το όργανο που χρησιμοποιείται για το κούρδισμα των οργάνων (tuning fork στο σχήμα) έχει αρμονικές με σχετικά χαμηλή ισχύ. Αντίθετα, όταν η ισχύς κατανέμεται στις ψηλότερες αρμονικές προσδίδει περισσότερο «χρώμα» στον ήχο. Η αντίθεση αυτή παρατηρείται στο Σχήμα 21, όπου για παράδειγμα το φλάουτο έχει ισχυρότερες τις πρώτες αρμονικές και παράγει σχετικά απλό ήχο που μοιάζει με σφύριγμα, ενώ ο ήχος από βιολί που είναι πιο πολύπλοκος και «ζεστός» παρουσιάζει πολύ ισχυρές ανώτερες αρμονικές.



Σχήμα 21: Ανάλυση διάφορων οργάνων

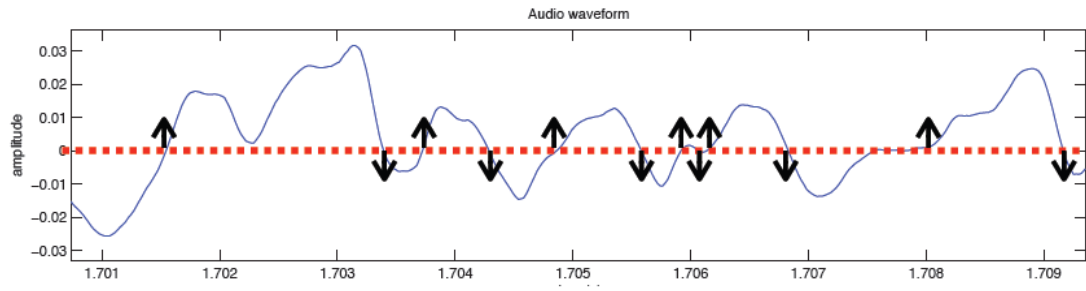
6.3. Χαρακτηριστικά γνωρίσματα ηχητικού σήματος

Στη συνέχεια θα γίνει παρουσίαση πιο σύνθετων μετρικών που μπορούν να χρησιμοποιηθούν για τη ταξινόμηση σε μουσικά είδη. Κάποια από τα χαρακτηριστικά γνωρίσματα αυτά αναφέρθηκαν κατά την παρουσίαση της βιβλιογραφίας σε προηγούμενο κεφάλαιο και έχουν χρησιμοποιηθεί σε διάφορες τεχνικές ταξινόμησης μουσικών ειδών.

- **Ρίζα των μέσων τετραγώνων της ενέργειας (Root Mean Square –RMS).** Η συνολική ενέργεια ενός σήματος x μπορεί να υπολογιστεί παίρνοντας το μέσο όρο της ρίζας των τετραγώνων του πλάτους, γνωστή και ως ρίζα των μέσων τετραγώνων της ενέργειας. Το μετρικό αυτό χρησιμοποιείται ουσιαστικά για να μετρήσει την ενέργεια του σήματος.

$$x_{\text{rms}} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$$

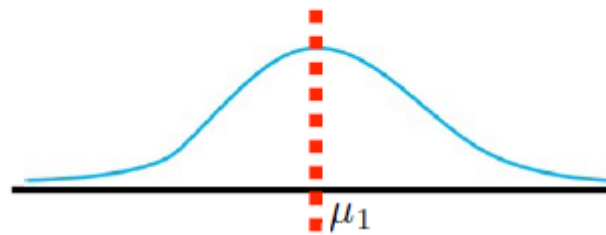
- **Zero-cross - Κυματομορφή ρυθμού αλλαγής πρόσημου.** Ένας απλός δείκτης θορύβου ο οποίος προκύπτει από τον αριθμό των φορών που το σήμα διασταυρώνεται με τον άξονα των x (ή με άλλα λόγια, πόσες φορές αλλάζει το πρόσημο).



Σχήμα 22 Zerocross (πηγή MIRtoolbox Users Guide)

- **Centroid (Κέντρο βάρους).** Το χαρακτηριστικό αυτό γνώρισμα επιστρέφει το κέντρο βάρους των δεδομένων. Μια σημαντική και χρήσιμη περιγραφή του σχήματος μιας κατανομής μπορεί να προκύψει με τη χρήση των ροπών. Η πρώτη ροπή, γνωστή ως mean, είναι το γεωμετρικό κέντρο (κέντρο βάρους) της κατανομής και αποτελεί ένα μέτρο της κύριας τάσης για την τυχαία μεταβλητή.

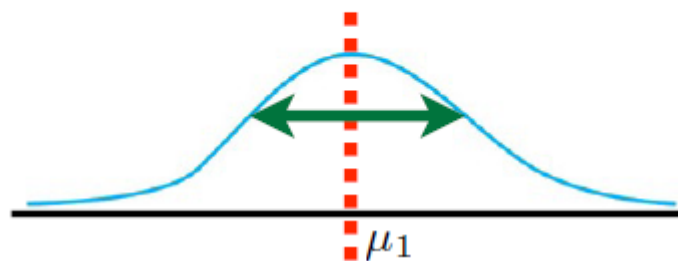
$$\mu_1 = \int x f(x) dx$$



Σχήμα 23 Centroid (πηγή MIRtoolbox Users Guide)

- **Spread (εξάπλωση):** επιστρέφει την διακύμανση (variance) των δεδομένων. Η δεύτερη κεντρική ροπή, γνωστή ως variance, συνήθως περιγράφεται από το σύμβολο σ^2 και ορίζεται ως εξής:

$$\sigma^2 = \mu_2 = \int (x - \mu_1)^2 f(x) dx$$



Σχήμα 24 Spread (πηγή MIRtoolbox Users Guide)

Το spread είναι η τετραγωνική απόκλιση της τυχαίας μεταβλητής από τη μέση τιμή, είναι πάντα θετική και αποτελεί ένα μέτρο για τη διασπορά ή το «άνοιγμα» της κατανομής. Η ρίζα της διακύμανσης λέγεται τυπική απόκλιση και είναι πιο χρήσιμη

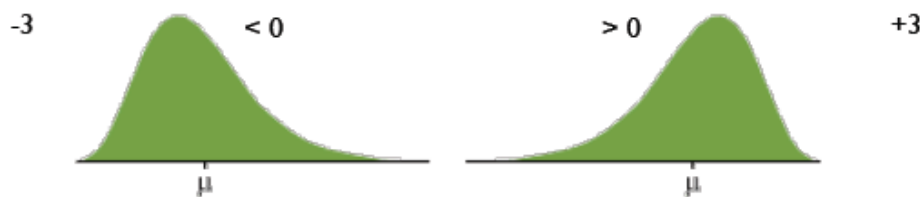
στην περιγραφή της φύσης της κατανομής καθώς έχει τις ίδιες μονάδες με την τυχαία μεταβλητή.

- **Skewness:** επιστρέφει τον συντελεστή λοξότητας των δεδομένων. Η τρίτη κύρια ροπή, ονομάζεται skewness και αποτελεί ένα μέτρο συμμετρίας της κατανομής. Η skewness έχει θετική τιμή στην περίπτωση που η κατανομή έχει λίγες τιμές μεγαλύτερες από το μέσο όρο οπότε και εμφανίζεται μια “ουρά” στα αριστερά. Μια αρνητικά skewed κατανομή εμφανίζει την “ουρά” στα δεξιά. Μια συμμετρική κατανομή εμφανίζει skewness στο μηδέν.

$$\mu_3 = \int (x - \mu_1)^3 f(x) dx$$

Ο συντελεστής skewness είναι ο λόγος του skewness προς την κανονική απόκλιση υψωμένα στην τρίτη δύναμη.

$$\frac{\mu_3}{\sigma^3}$$



Σχήμα 25 Skewness (πηγή MIRtoolbox Users Guide)

- **Regularity:** Η ασυμμετρία ενός φάσματος είναι ο βαθμός μεταβολής των διαδοχικών κορυφών του φάσματος. Υπάρχουν 2 διαφορετικοί ορισμοί της ασυμμετρίας. Με βάση την έρευνα που έγινε στο [Jens99] είναι το άθροισμα των τετραγώνων της διαφοράς πλάτους ανάμεσα σε παρακείμενες κορυφές:

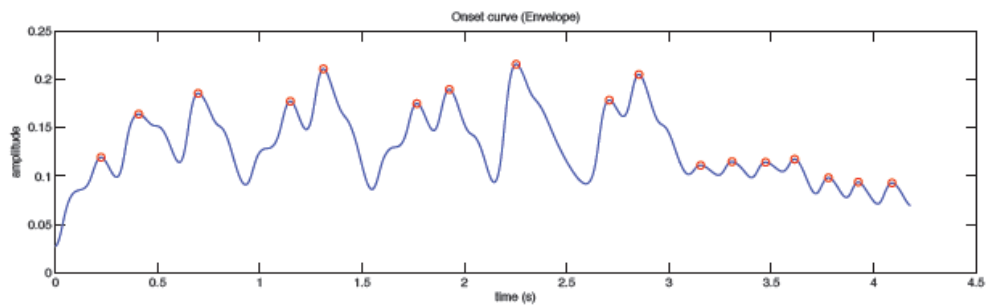
$$\left(\sum_{k=1}^N (a_k - a_{k+1})^2 \right) / \sum_{k=1}^N a_k^2$$

Με βάση όμως την έρευνα που έγινε στο [Krim94] είναι το άθροισμα του πλάτους της κάθε κορυφής μείον το μέσο της προηγούμενης, επόμενης και τρέχουσας κορυφής:

$$\sum_{k=2}^{N-1} \left| a_k - \frac{a_{k-1} + a_k + a_{k+1}}{3} \right|$$

- **Tempo:** Υπολογίζει το ρυθμό – τέμπο αναζητώντας περιοδικότητες από την αρχή της περιοδικής καμπύλης. Το tempo μπορεί να υπολογιστεί με διάφορους τρόπους. Βασίζεται γενικά στον υπολογισμό των διαδοχικών εκρήξεων ενέργειας που αντιστοιχούν σε διαδοχικούς παλμούς. Η καμπύλη που περιγράφει τους παλμούς

αυτούς λέγεται καμπύλη ανίχνευσης ενάρξεων (onset detection curve). Το Tempo μπορεί να προκύψει από τη συνάρτηση αυτοσυσχέτισης αυτής της καμπύλης.

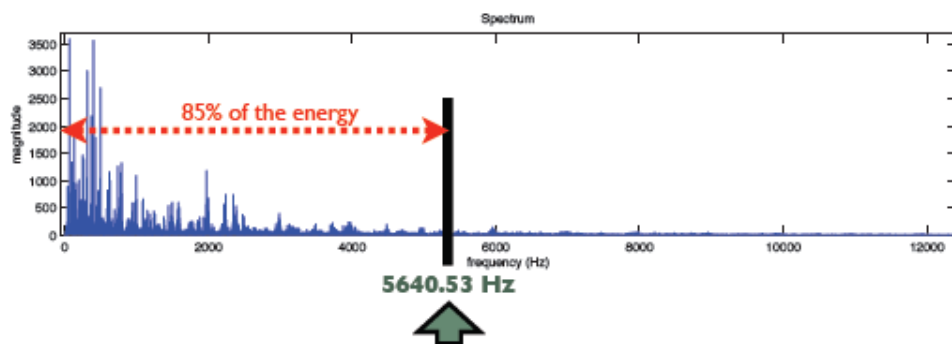


Σχήμα 26 Onset detection Curve (πηγή MIRtoolbox Users Guide)

- **Pitch – Θεμελιώδης συχνότητα**

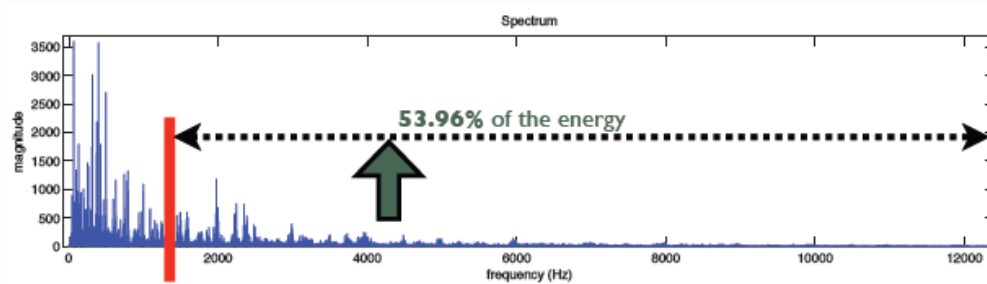
Για τη Θεμελιώδη συχνότητα έγινε λεπτομερής αναφορά στην ενότητα 6.1.

- **Rollof:** ενέργεια υψηλών συχνοτήτων. Ένας τρόπος υπολογισμού της ποσότητας υψηλών συχνοτήτων του σήματος συνίσταται στην εύρεση μιας συχνότητας τέτοιας ώστε ένα προκαθορισμένο κλάσμα της συνολικής ενέργειας να βρίσκεται κάτω από αυτή τη συχνότητα. Μια τυπική τιμή για το λόγο που χρησιμοποιείται συνήθως στο rollof είναι 85%.



Σχήμα 27 Rollof (πηγή MIRtoolbox Users Guide)

- **Brightness:** Το γνώρισμα αυτό είναι μια εναλλακτική μέθοδος για τον υπολογισμό της ενέργειας των υψηλών συχνοτήτων. Συνίσταται στον υπολογισμό της cut-off συχνότητας και στον υπολογισμό του ποσού της ενέργειας που βρίσκεται πάνω από αυτή τη συχνότητα. Η συχνότητα cut-off είναι ένα λογικό όριο στο φάσμα του σήματος μετά από το οποίο η ενέργεια τείνει να μειώνεται. [Jus100]. Το αποτέλεσμα είναι ένας αριθμός ανάμεσα στο 0 και το 1.



Σχήμα 28 Brightness (πηγή MIRtoolbox Users Guide)

7. Ταξινόμηση σε μουσικά είδη με τη χρήση τεχνητών νευρωνικών δικτύων

7.1. Αρχιτεκτονικές ΤΝΔ

Με τη βοήθεια του Neural Network Toolbox του Matlab θα εξεταστούν 3 διαφορετικές αρχιτεκτονικές νευρωνικών δικτύων. Πιο συγκεκριμένα, θα δοκιμαστούν οι παρακάτω τύποι νευρωνικών δικτύων:

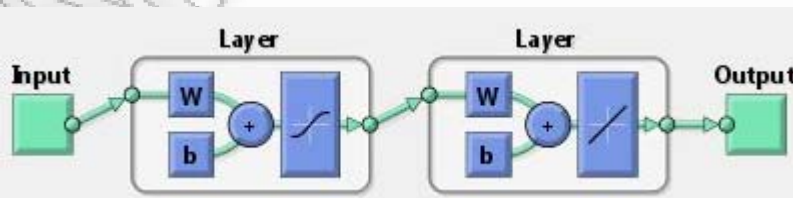
- Feed-Forward Neural Network
- Cascade-Forward Neural Network
- Elman back-propagation Neural Network

Η αρχιτεκτονική του κάθε είδους από τα παραπάνω νευρωνικά δίκτυα παρουσιάζεται στη συνέχεια.

7.1.1. Αρχιτεκτονική με Feed-Forward Neural Network

Τα δίκτυα τύπου « Feed-Forward » είναι πολυστρωματικά δίκτυα. Το πρώτο στρώμα έχει βάρη που καθορίζονται από την είσοδο. Κάθε ένα από τα ενδιάμεσα στρώματα έχει πολώσεις και βάρη. Τα βάρη καθορίζονται αποκλειστικά από την έξοδο του αμέσως προηγούμενου στρώματος. Το τελευταίο στρώμα είναι η έξοδος του δικτύου.

Πιο συγκεκριμένα, χρησιμοποιήθηκαν 2 στρώματα. Η συνάρτηση μεταφοράς που χρησιμοποιήθηκε στο κρυφό στρώμα είναι η *tansig*, ενώ στο στρώμα εξόδου είναι η *purelin*. Το διάγραμμα του δικτύου φαίνεται στο παρακάτω σχήμα:

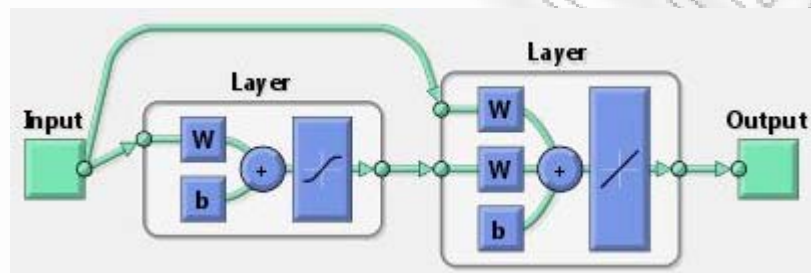


Σχήμα 29 Feed-Forward Neural Network

7.1.2. Αρχιτεκτονική με *Cascade-Forward back-propagation Neural Network*

Τα δίκτυα τύπου «Cascade-Forward» είναι πολυστρωματικά δίκτυα. Το πρώτο στρώμα έχει βάρη που καθορίζονται από την είσοδο. Κάθε ένα από τα ενδιάμεσα στρώματα έχει πολώσεις και βάρη. Τα βάρη στα ενδιάμεσα στρώματα καθορίζονται από την έξοδο του αμέσως προηγούμενου στρώματος αλλά και από την είσοδο του νευρωνικού δικτύου. Το τελευταίο στρώμα είναι η έξοδος του δικτύου.

Πιο συγκεκριμένα, χρησιμοποιήθηκαν 2 στρώματα. Η συνάρτηση μεταφοράς που χρησιμοποιήθηκε στο κρυφό στρώμα είναι η *tansig*, ενώ στο στρώμα εξόδου είναι η *purelin*. Το διάγραμμα του δικτύου φαίνεται στο παρακάτω σχήμα:

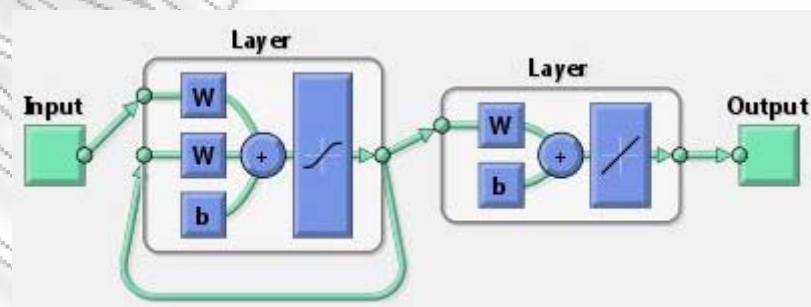


Σχήμα 30 Cascade-Forward Neural Network

7.1.3. Αρχιτεκτονική με *Elman back-propagation Neural Network*

Τα νευρωνικά δίκτυα τύπου «Elman» είναι πολυστρωματικά δίκτυα. Το πρώτο στρώμα καθώς και τα ενδιάμεσα έχουν βάρη τα οποία καθορίζονται από την έξοδο του αμέσως προηγούμενου στρώματος αλλά και βάρη τα οποία καθορίζονται από την έξοδο του ίδιου στρώματος. Η ιδιότητα αυτή χαρακτηρίζεται ως μνήμη του δικτύου. Το τελευταίο στρώμα είναι η έξοδος του δικτύου.

Πιο συγκεκριμένα, χρησιμοποιήθηκαν 2 στρώματα. Η συνάρτηση μεταφοράς που χρησιμοποιήθηκε στο κρυφό στρώμα είναι η *tansig*, ενώ στο στρώμα εξόδου είναι η *purelin*. Το διάγραμμα του δικτύου φαίνεται στο παρακάτω σχήμα:



Σχήμα 31 Elman Neural network

7.2. Πειράματα και Αριθμητικά αποτελέσματα

Στην ενότητα αυτή παρουσιάζονται τα πειράματα για την αξιολόγηση της καταλληλότητας της χρήσης νευρωνικών δικτύων για την κατάταξη μελωδίας σε μουσικά είδη.

Πιο συγκεκριμένα, για την εκπαίδευση του νευρωνικού δικτύου χρησιμοποιήθηκαν πολλά λεπτά μουσικής από κομμάτια από διαφορετικές κατηγορίες: Κλασική μουσική, ελληνική παραδοσιακή μουσική (νησιώτικα), ελληνική παραδοσιακή δημοτική μουσική (Ηπειρώτικα), Ελληνική παραδοσιακή μουσική (Κρητικά) και μουσική ντίσκο. Οι κατηγορίες αυτές επιλέχθηκαν έτσι ώστε να υπάρχουν κατηγορίες με ιδιαίτερα συγγενές ήχο (οι υποκατηγορίες της ελληνικής παραδοσιακής μουσικής) αλλά και κατηγορίες με τελείως διαφορετικό ήχο, τουλάχιστον σε ότι αφορά την αντίληψη του ήχου από τον άνθρωπο.

Κάθε μουσικό κομμάτι ψηφιοποιήθηκε σε μορφή wav, και με τη χρήση του Mir ToolBox, εξάχθηκαν τα χαρακτηριστικά γνωρίσματα που περιγράφηκαν στην ενότητα 6.3. Ύστερα από διάφορες δοκιμές, επιλέχθηκαν τελικά τα παρακάτω έξι χαρακτηριστικά γνωρίσματα ως είσοδοι στο νευρωνικό δίκτυο:

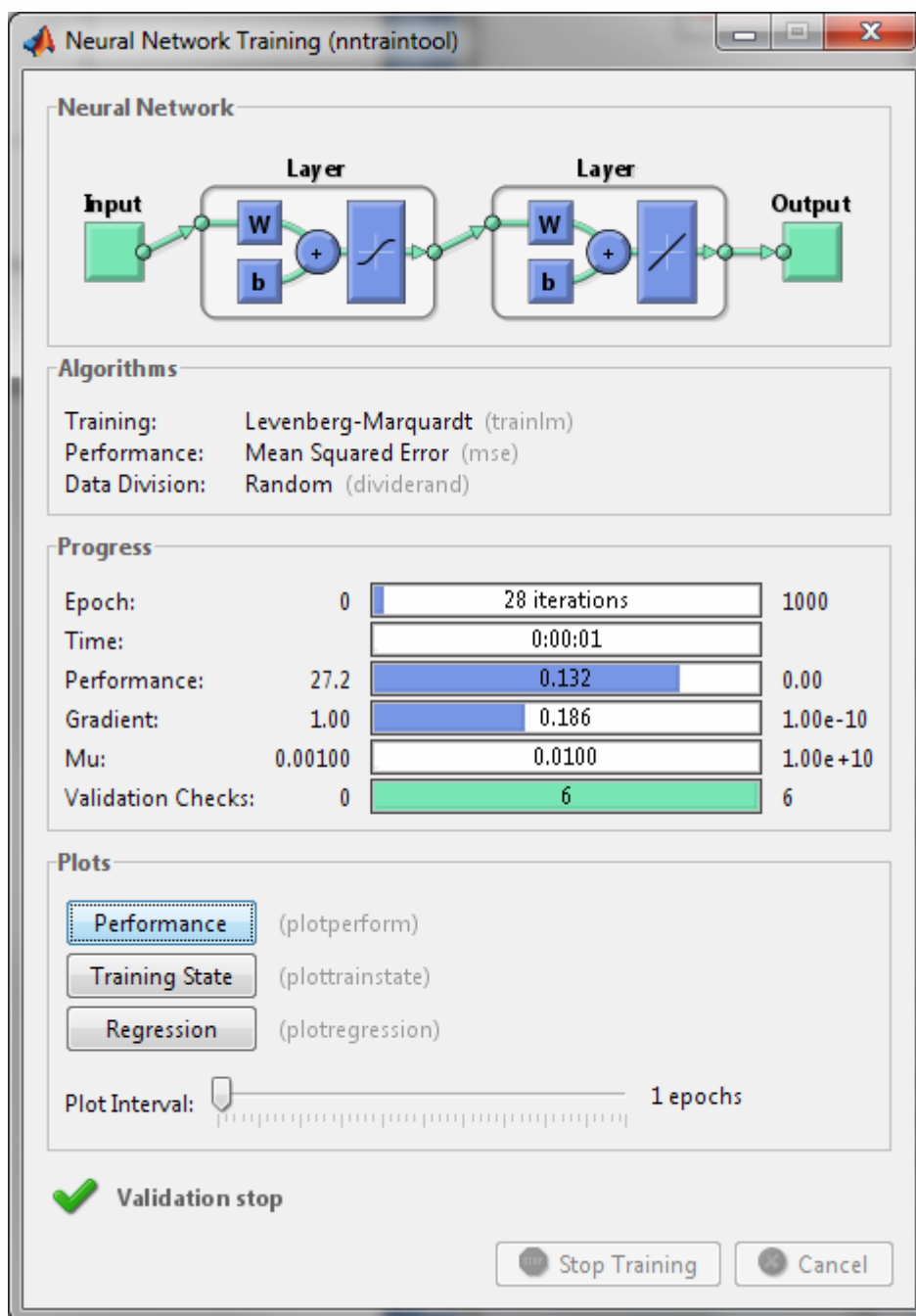
- RMS
- ZeroCross
- Centroid
- Spread
- Skewness
- Regularity

Συνεπώς, αν υπάρχει ένα ηχητικό κομμάτι το οποίο αποτελείται από 500 δείγματα, τότε δημιουργείται ως είσοδος για την εκπαίδευση του νευρωνικού δικτύου, ένας πίνακας 500x6. Για την κατηγοριοποίηση της μουσικής, ορίζεται ως επιθυμητή έξοδος μια ακέραια τιμή >0. Στα αριθμητικά αποτελέσματα που θα ακολουθήσουν, με 1 ήταν τα κρητικά, με 2 ήταν έξοδος για νησιώτικα, 3 για δημοτικά, 4 για ντίσκο, 5 η κλασική μουσική. Στη συνέχεια και για λόγους ευκολίας, θα αναφερόμαστε στα παραπάνω ως μουσικό είδος 1,2,3,4,5, με κάθε είδος να έχει επιθυμητή τιμή εξόδου στο νευρωνικό δίκτυο ίση με την διάταξη αυτή.

7.2.1. Υλοποίηση με Feed-Forward Neural Network

Ακολουθούν τα αριθμητικά αποτελέσματα με την χρήση Feed-Forward Neural Network - FF. Η απόδοση του δικτύου αυτού δοκιμάστηκε με 2 διαφορετικά πλήθη κρυμμένων νευρώνων: 10 και 20 νευρώνες.

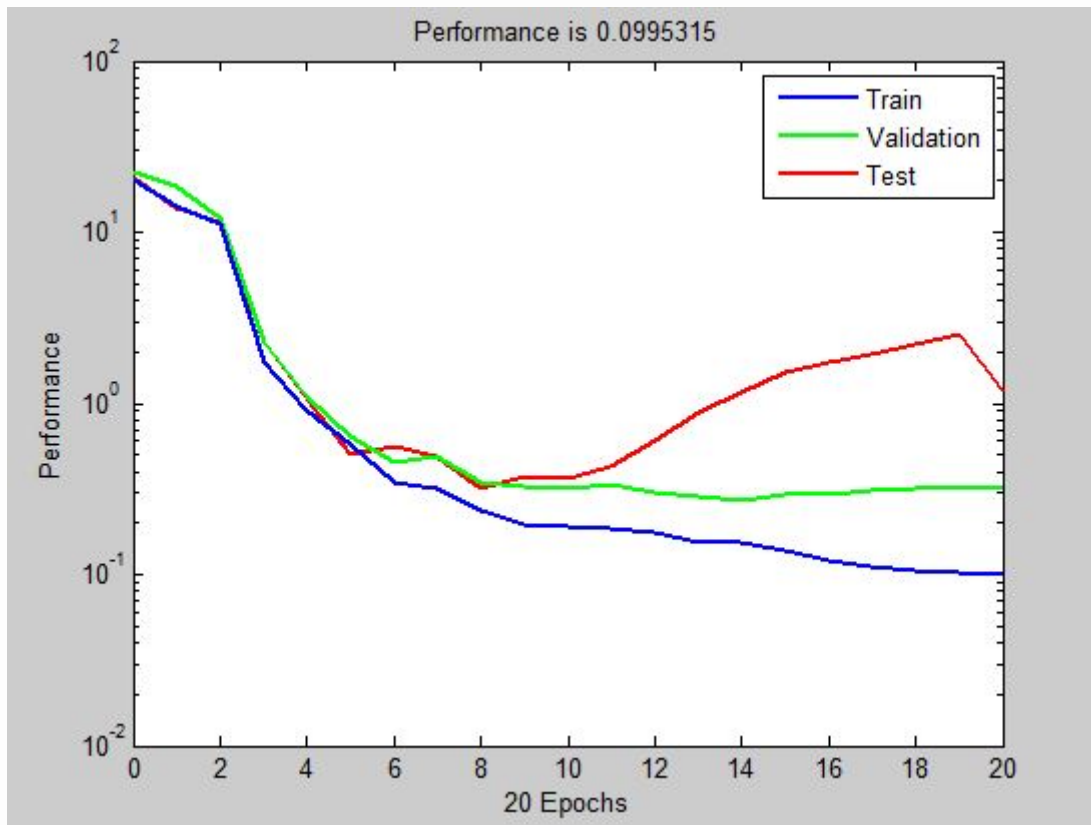
Για την εκπαίδευση του δικτύου έγινε χρήση της μεθόδου trainlm. Ακολουθεί στιγμιότυπο από την διαδικασία εκπαίδευσης του δικτύου. Παρατηρούμε ότι το δίκτυο εκπαιδεύεται πολύ γρήγορα, αρκούν μονάχα 28 επαναλήψεις προκειμένου να επιτευχθεί σταθερή απόδοση του δικτύου.



Σχήμα 32 Κατάληξη του αλγόριθμου εκπαίδευσης trainlm (Levenberg-Marquardt backpropagation) για δίκτυο FF με 20 νευρώνες

Όπως περιγράφηκε και σε προηγούμενη ενότητα, η διαδικασία της εκπαίδευσης είναι μια επαναληπτική διαδικασία, η οποία τερματίζει είτε όταν η απόδοση του δικτύου πιάσει το επιθυμητό όριο είτε περάσει ο μέγιστος αριθμός επαναλήψεων. Κάποιες από τις τιμές του διανύσματος εισόδου επιλέγονται τυχαία για την εκπαίδευση του δικτύου και κάποιες για τον έλεγχο της απόδοσης. Η απόδοση ορίζεται ως το μέσο τετραγωνικό λάθος (Mean Square Error).

Οι καμπύλες της απόδοσης σε συνάρτηση με τον αριθμό επαναλήψεων εικονίζονται στο επόμενο σχήμα. Όταν επιτευχθεί συγκεκριμένος προκαθορισμένος αριθμός επιτυχών επαληθεύσεων (Validation checks=6), η εκπαίδευση τερματίζει.



Σχήμα 33 Διαδικασία Εκπαίδευσης: Απόδοση του δικτύου FF με 20 νευρώνες

Στο επόμενο σχήμα παρουσιάζονται καμπύλες για την αξιολόγηση του νευρωνικού δικτύου. Πιο συγκεκριμένα, δίνονται ως είσοδοι ηχητικά κομμάτια από κάθε είδος μουσικής και τυπώνονται οι πραγματικές τιμές εξόδου που είναι το σύνολο $\{1,2,3,4\}$ και οι τιμές που επιστρέφει το δίκτυο.

Για την κατάταξη σε μουσικό είδος χρησιμοποιείται η μέση τιμή του διανύσματος εξόδου, ενώ μέτρο της ακρίβειας της κατάταξης αποτελεί και η τυπική απόκλιση των τιμών του διανύσματος.

Για το δίκτυο FF-20 έχουμε:

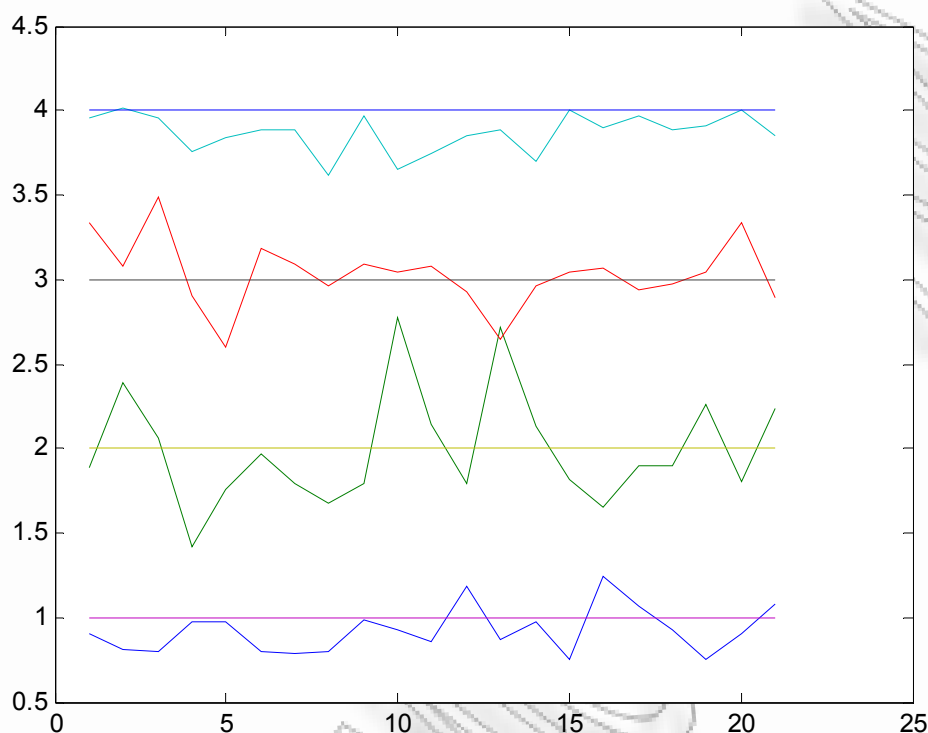
$\text{mean}(1) = 0.9224$, $\text{std}(1) = 0.1368$

$\text{mean}(2) = 1.9930$, $\text{std}(2) = 0.3376$

$\text{mean}(3) = 3.0327$, $\text{std}(3) = 0.2045$

$\text{mean}(4) = 3.8695$, $\text{std}(4) = 0.1142$

Στο σχήμα που ακολουθεί, εμφανίζονται τα διαγράμματα για τις παραπάνω τιμές. Πιο συγκεκριμένα, οι σταθερές καμπύλες ισοδυναμούν στις πραγματικές τιμές κάθε κομματιού $\{1,2,3,4\}$, ενώ οι άλλες καμπύλες αντιστοιχούν στις προσομοιωμένες τιμές. Παρατηρούμε ότι με τη χρήση του μέσου μπορεί να επιτευχθεί ακριβής ανίχνευση και κατάταξη της μελωδίας σε μουσικό είδος.

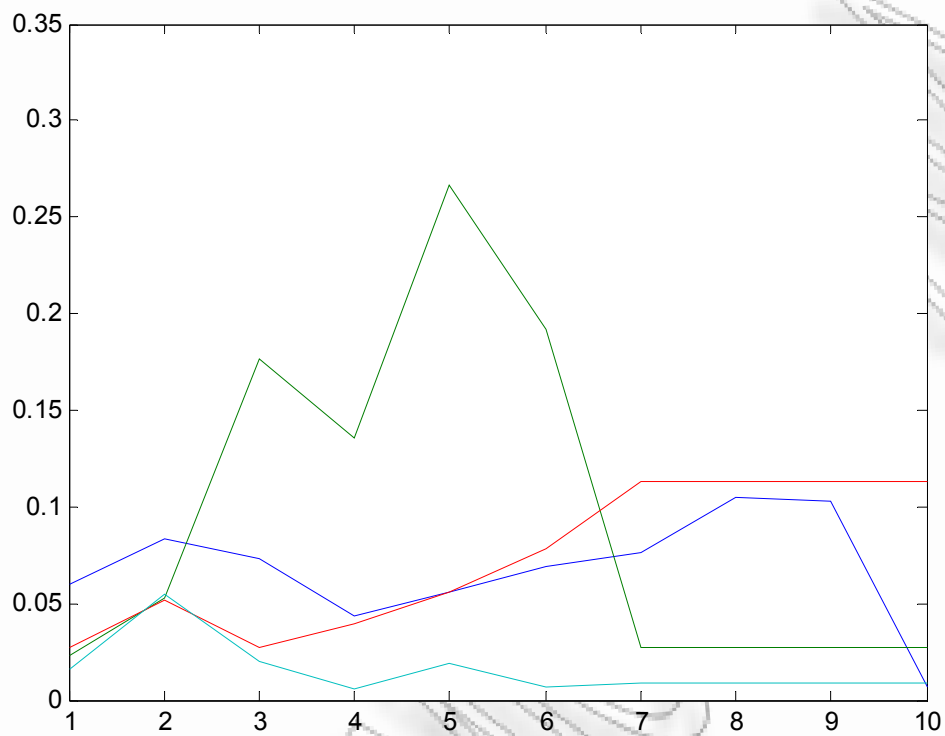


Σχήμα 34 Μελέτη απόδοσης δικτύου FF με 20 νευρώνες: Πραγματικές και προσομοιωμένες τιμές για ηχητικά κομμάτια.

Στη συνέχεια, θα γίνει μια προσπάθεια υπολογισμού του αριθμού τιμών που απαιτούνται από το δείγμα προκειμένου να υπολογιστεί με ακρίβεια το μουσικό είδος. Το επόμενο σχήμα παρουσιάζει τη μέση απόκλιση του μέσου όρου του διανύσματος εξόδου από τη πραγματική τιμή, ως συνάρτηση του αριθμού δειγμάτων που χρησιμοποιήθηκαν για την κατάταξη.

Για παράδειγμα, αν χρησιμοποιήθηκαν 3 τιμές, για ένα δείγμα που ανήκει στο μουσικό είδος 2, υπολογίζουμε την απόκλιση του μέσου όρου των τριών αυτών τιμών από το 3. Το ίδιο κάνουμε για εύρος τιμών από 1 ως 10, από το μουσικό κομμάτι.

Παρατηρούμε, ότι για εύρος δείγματος μεγαλύτερο του 7, η απόκλιση σταθεροποιείται.



Σχήμα 35 Ακρίβεια υπολογισμού δείγματος για FF-20: Άξονας X: πλήθος τιμών που χρησιμοποιήθηκαν, Y: απόκλιση από την πραγματική τιμή

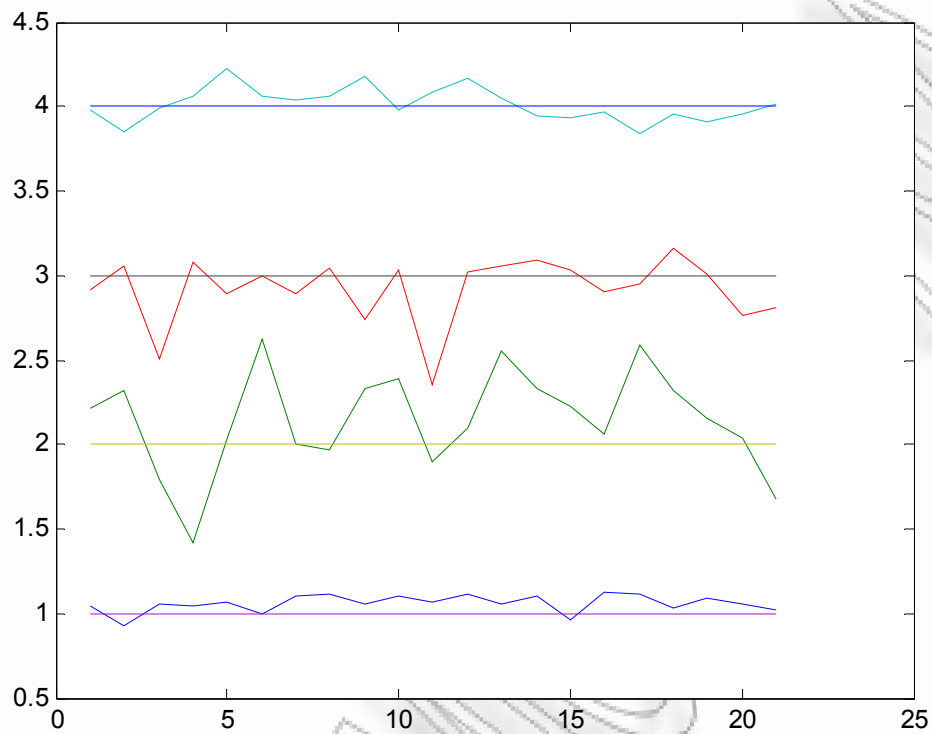
Ακολουθούν τα αντίστοιχα αριθμητικά αποτελέσματα και καμπύλες για το FF-10, δηλαδή δίκτυο Feed Forward με 10 κρυμμένους νευρώνες.

$\text{mean}(1) = 1.0598, \text{std}(1) = 0.0522$

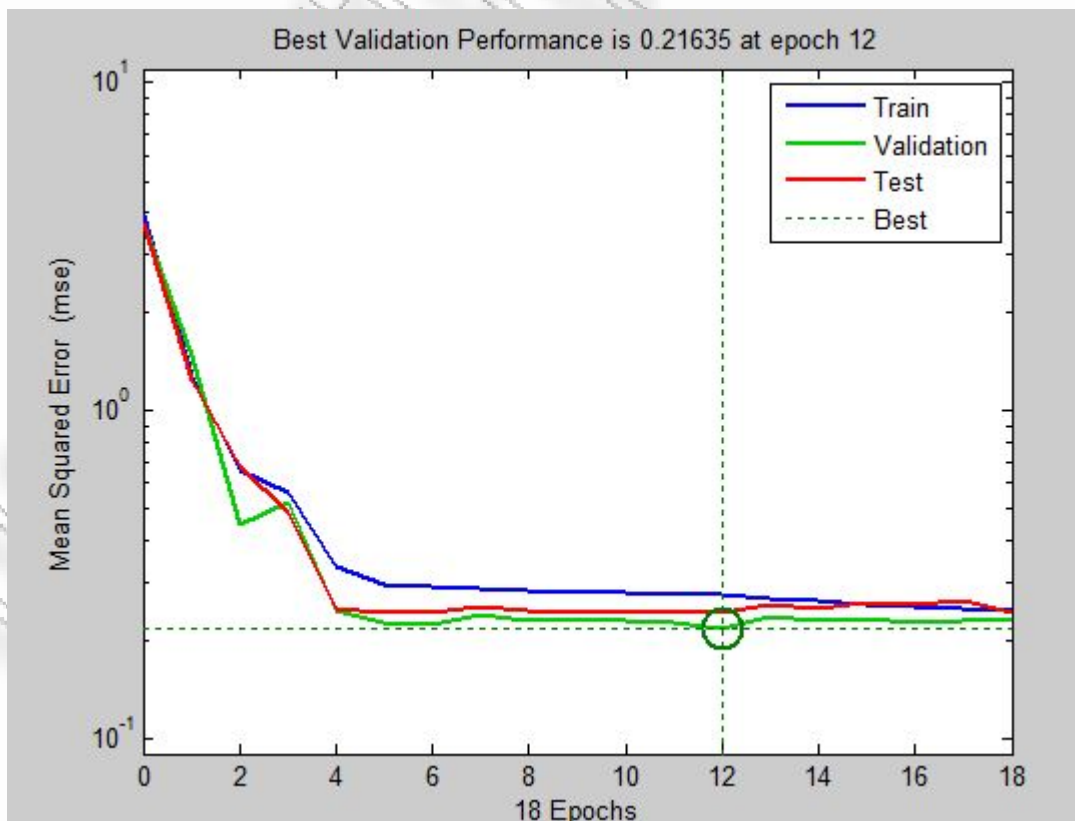
$\text{mean}(2) = 2.1431, \text{std}(2) = 0.3020$

$\text{mean}(3) = 2.9182, \text{std}(3) = 0.1954$

$\text{mean}(4) = 4.0114, \text{std}(4) = 0.1002$

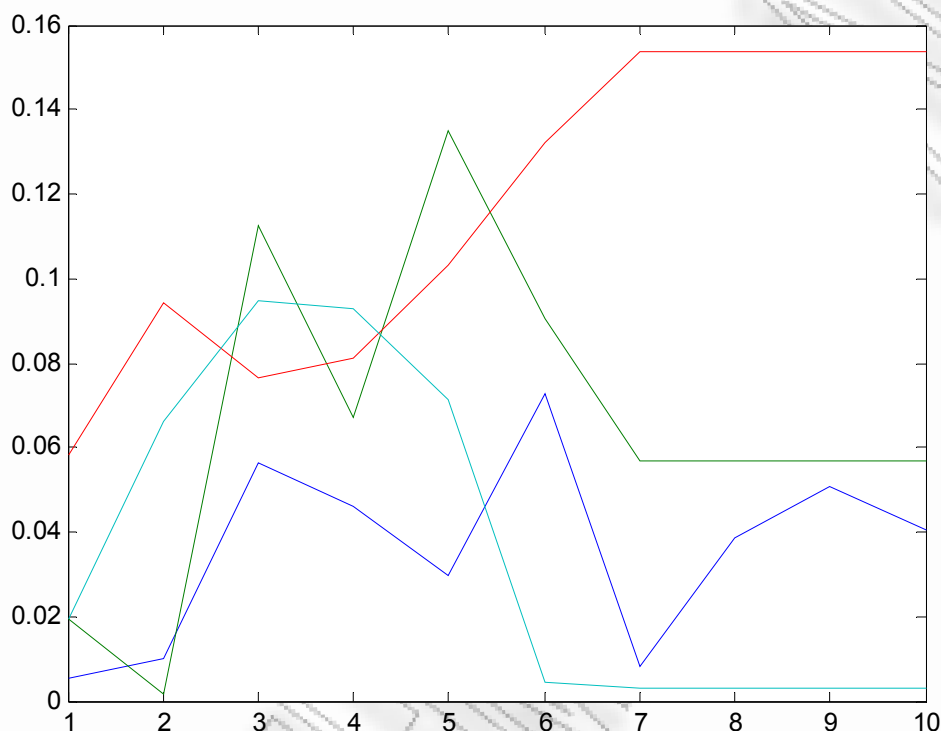


Σχήμα 36 Μελέτη απόδοσης δικτύου FF με 10 νευρώνες: Πραγματικές και προσομοιωμένες τιμές για ηχητικά κομμάτια.



Σχήμα 37 Διαδικασία εκπαίδευσης του FF με 10 νευρώνες

Σε αντιστοιχία με το FF-20, και στο FF-10 για τα περισσότερα μουσικά είδη παρατηρείται σταθεροποίηση της απόκλισης για αριθμό δειγμάτων >7 .



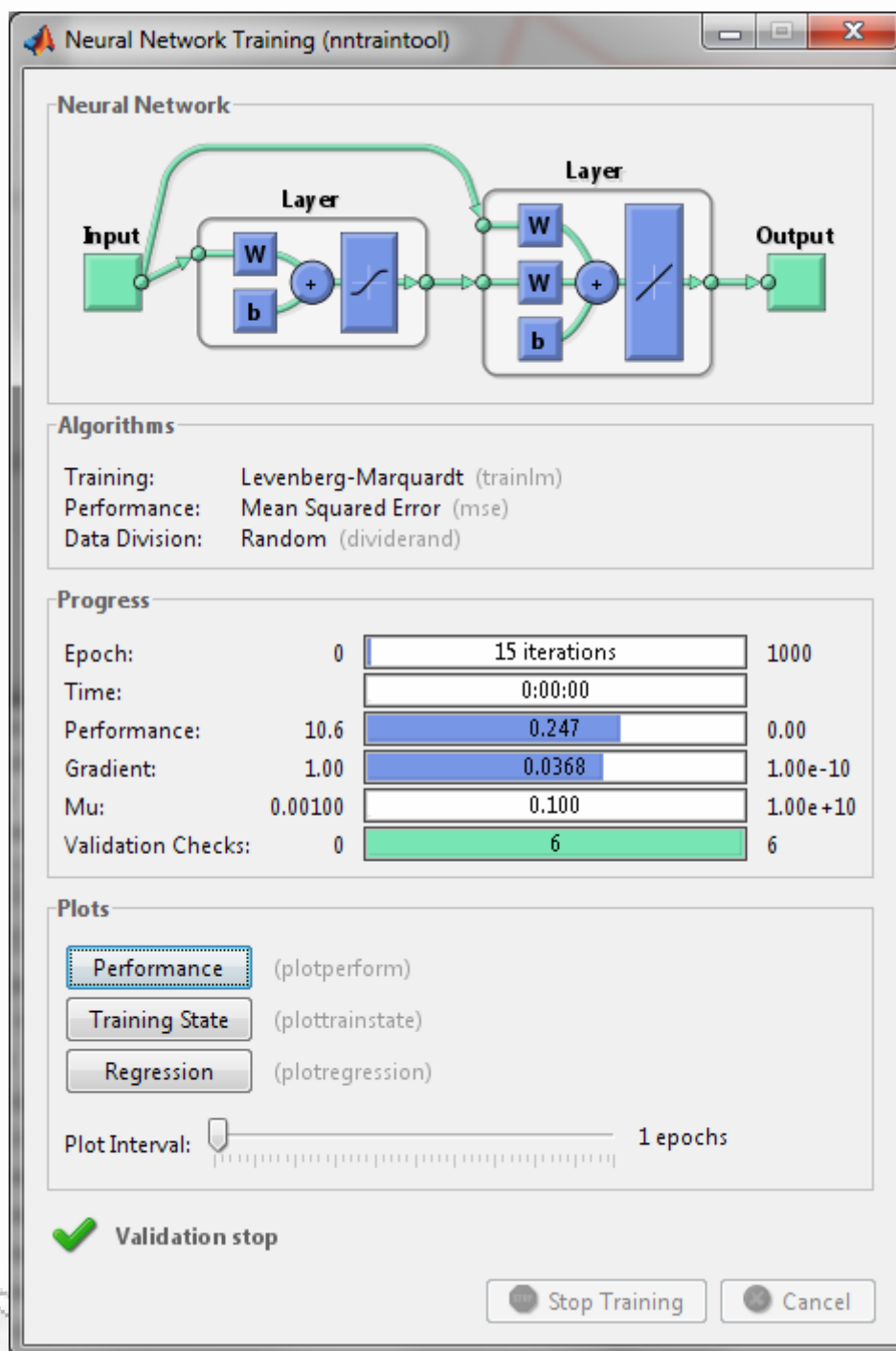
Σχήμα 38 Ακρίβεια υπολογισμού δείγματος για FF-10: Αξονας X: πλήθος τιμών που χρησιμοποιήθηκαν, Y: απόκλιση από την πραγματική τιμή

Παρατηρούμε ότι το δίκτυο έχει παρόμοια απόδοση είτε με τη χρήση 10 είτε με την χρήση 20 κρυμμένων νευρώνων. Στα επόμενα δίκτυα, θα παρουσιαστούν μονάχα τα αριθμητικά αποτελέσματα με αριθμό κρυμμένων νευρώνων ίσο με 10.

7.2.2. Υλοποίηση με Cascade-Forward Neural Network

Ακολουθούν τα αριθμητικά αποτελέσματα με την χρήση Cascade-Forward Neural Network - CF. Η απόδοση του δικτύου αυτού δοκιμάστηκε με αριθμό κρυμμένων νευρώνων: 10.

Για την εκπαίδευση του δικτύου έγινε χρήση της μεθόδου `trainlm`. Ακολουθεί στιγμότυπο από την διαδικασία εκπαίδευσης του δικτύου. Παρατηρούμε ότι το δίκτυο εκπαιδεύεται πολύ γρήγορα, αρκούν μονάχα 15 επαναλήψεις προκειμένου να επιτευχθεί σταθερή απόδοση του δικτύου.



Σχήμα 39 Κατάληξη του αλγόριθμου εκπαίδευσης trainlm (Levenberg-Marquardt backpropagation) για δίκτυο CF με 10 νευρώνες

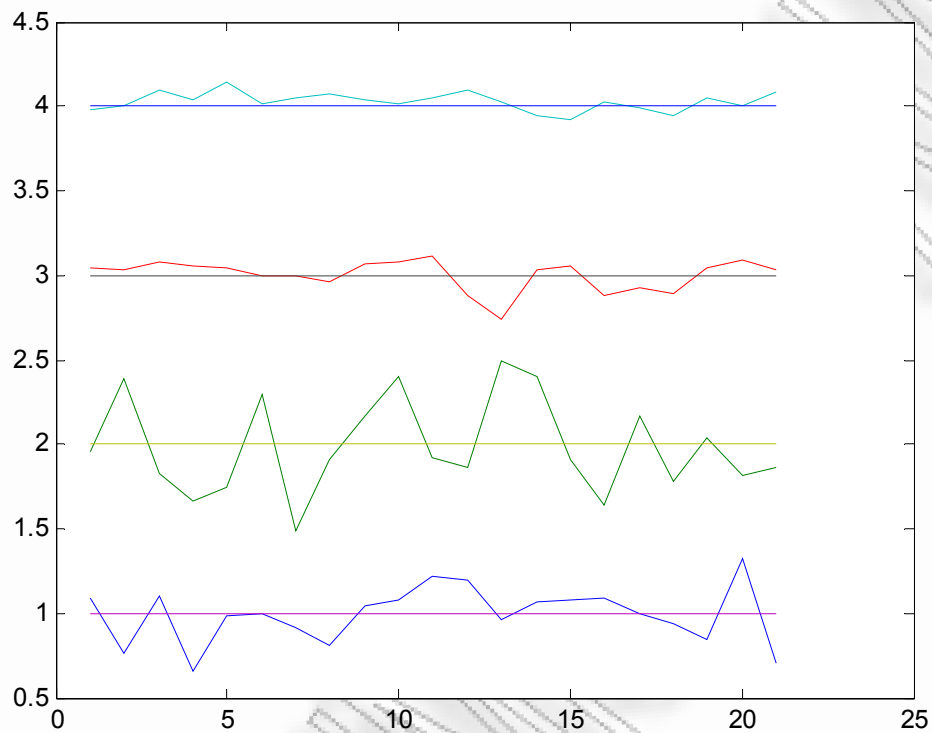
Ακολουθούν τα αντίστοιχα αριθμητικά αποτελέσματα και οι καμπύλες για το CF-10, δηλαδή δίκτυο Cascade Forward με 10 κρυμμένους νευρώνες.

mean(1) = 1.0395, std(1)= 0.1105

mean(2) = 1.9038, std(2)= 0.3890

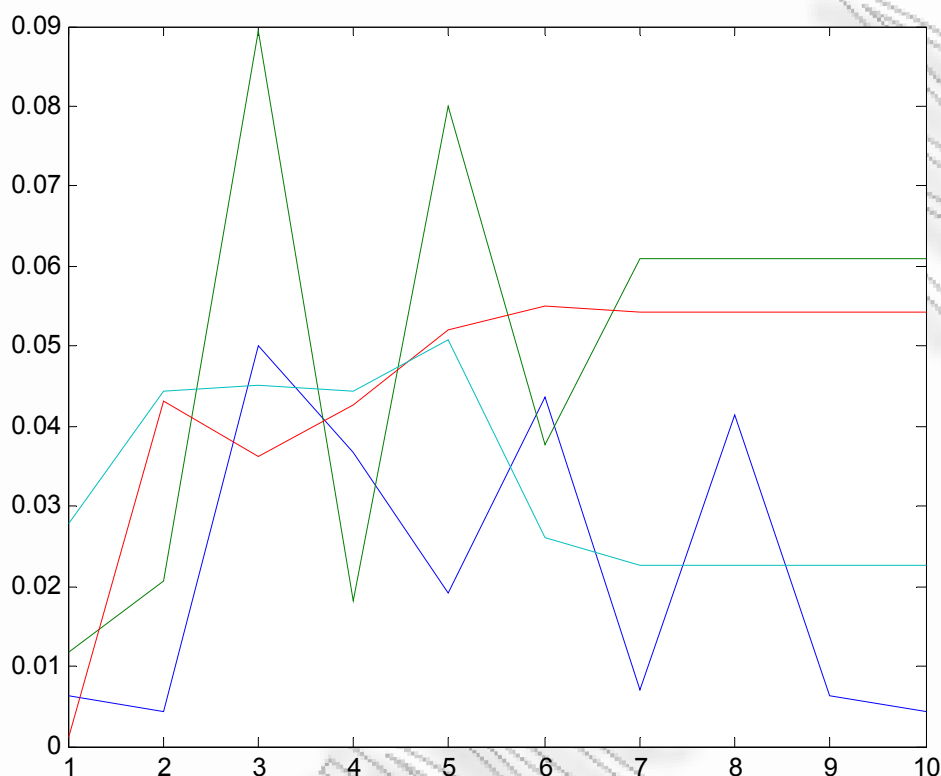
mean(3) = 3.0941, std(3)= 0.2441

mean(4) = 4.2014, std(4)= 0.1691



Σχήμα 40 Μελέτη απόδοσης δικτύου CF με 10 νευρώνες: Πραγματικές και προσομοιωμένες τιμές για ηχητικά κομμάτια.

Στο CF-10 για τα περισσότερα μουσικά είδη παρατηρείται σταθεροποίηση της απόκλισης για αριθμό δειγμάτων >7 .



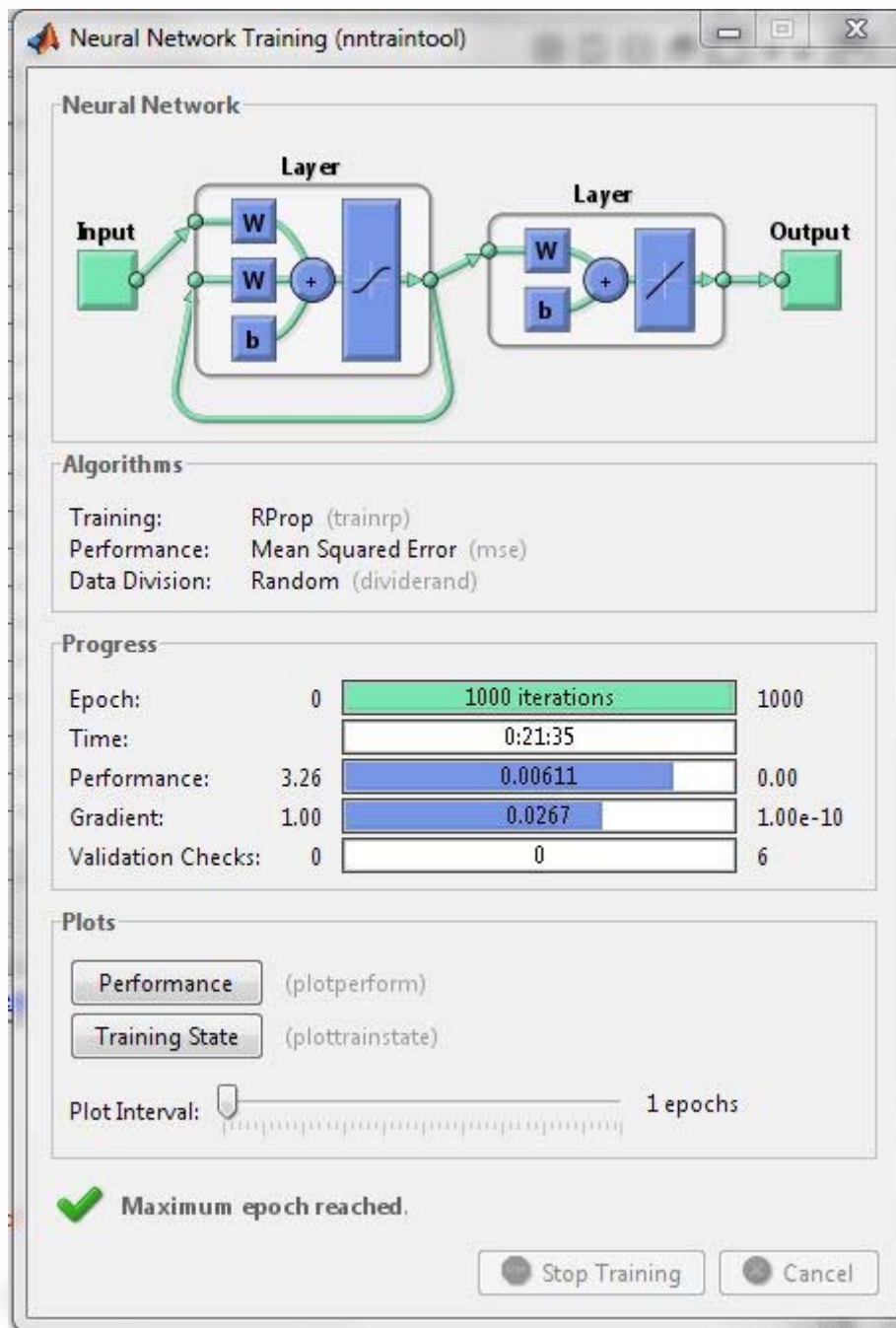
Σχήμα 41 Ακρίβεια υπολογισμού δείγματος για CF-10: Άξονας X: πλήθος τιμών που χρησιμοποιήθηκαν, Y: απόκλιση από την πραγματική τιμή

Από τη μελέτη των παραπάνω αριθμητικών αποτελεσμάτων και διαγραμμάτων προκύπτει ότι και με τη χρήση του δικτύου CF-10, μπορεί να γίνει με επιτυχία κατάταξη μελωδίας σε μουσικά είδη.

7.2.3. Υλοποίηση με *Elman Back-propagation Neural Network*

Ακολουθούν τα αριθμητικά αποτελέσματα με την χρήση *Elman Back-propagation Neural Network*. Η απόδοση του δικτύου αυτού δοκιμάστηκε με αριθμό κρυμμένων νευρώνων: 10.

Για την εκπαίδευση του δικτύου έγινε χρήση της μεθόδου `trainrp`, γιατί η μέθοδος `trainlm` δε προτείνεται (είναι πάρα πολύ αργή και χρειάζεται περισσότερη μνήμη) σε αυτό το τύπο δικτύου. Ακολουθεί στιγμιότυπο από την διαδικασία εκπαίδευσης του δικτύου. Παρατηρούμε ότι το δίκτυο εκπαιδεύεται πολύ πιο αργά σε σχέση με τις προηγούμενες περιπτώσεις.



Σχήμα 42 Κατάληξη του αλγόριθμου εκπαίδευσης trainlm (Resilient backpropagation) για δίκτυο Elman με 10 νευρώνες

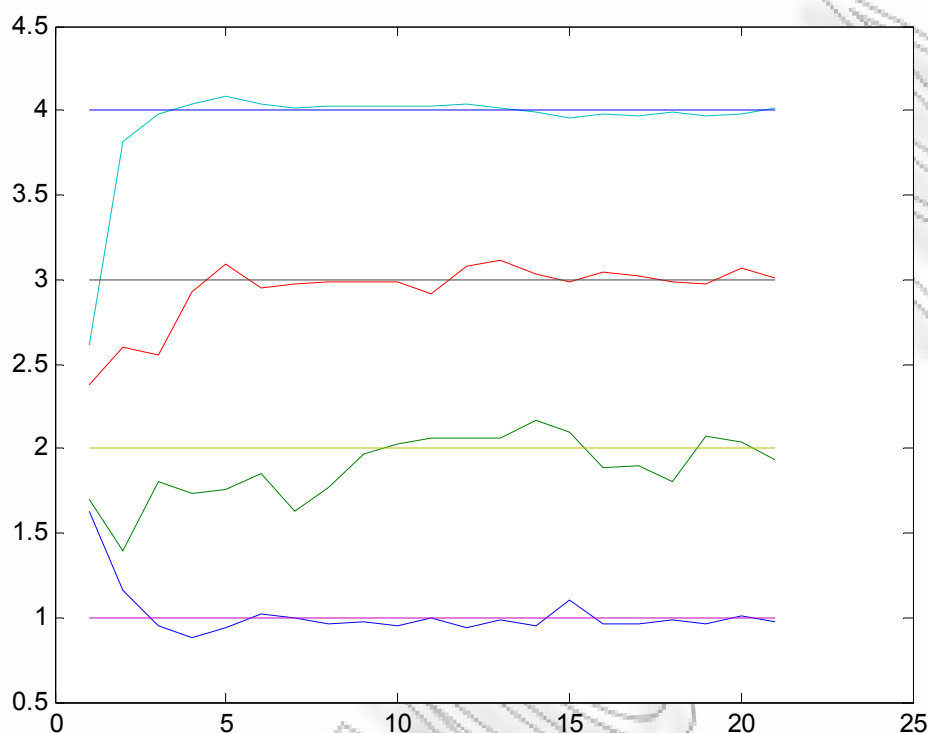
Ακολουθούν τα αντίστοιχα αριθμητικά αποτελέσματα και καμπύλες για το Elman-10, δηλαδή δίκτυο Elman με 10 κρυμμένους νευρώνες.

mean(1) = 1.0138, std(1)= 0.1520

mean(2) = 1.8896, std(2)= 0.1901

mean(3) = 2.9351, std(3)= 0.1898

mean(4) = 3.9324, std(4)= 0.3066



Σχήμα 43 Μελέτη απόδοσης δικτύου Elman με 10 νευρώνες: Πραγματικές και προσομοιωμένες τιμές για ηχητικά κομμάτια.

Στη περίπτωση του δικτύου αυτού παρατηρούμε ότι υπάρχει μεγάλη απόκλιση στις αρχικές τιμές του μουσικού κομματιού. Αν για τον υπολογισμό του μέσου όρου δε χρησιμοποιηθούν οι 5 πρώτες τιμές (δε χρησιμοποιηθεί δηλαδή η εισαγωγή του κομματιού) λαμβάνουμε τις εξής τιμές:

$$\text{mean}(1) = 0.9824, \text{std}(1) = 0.0388$$

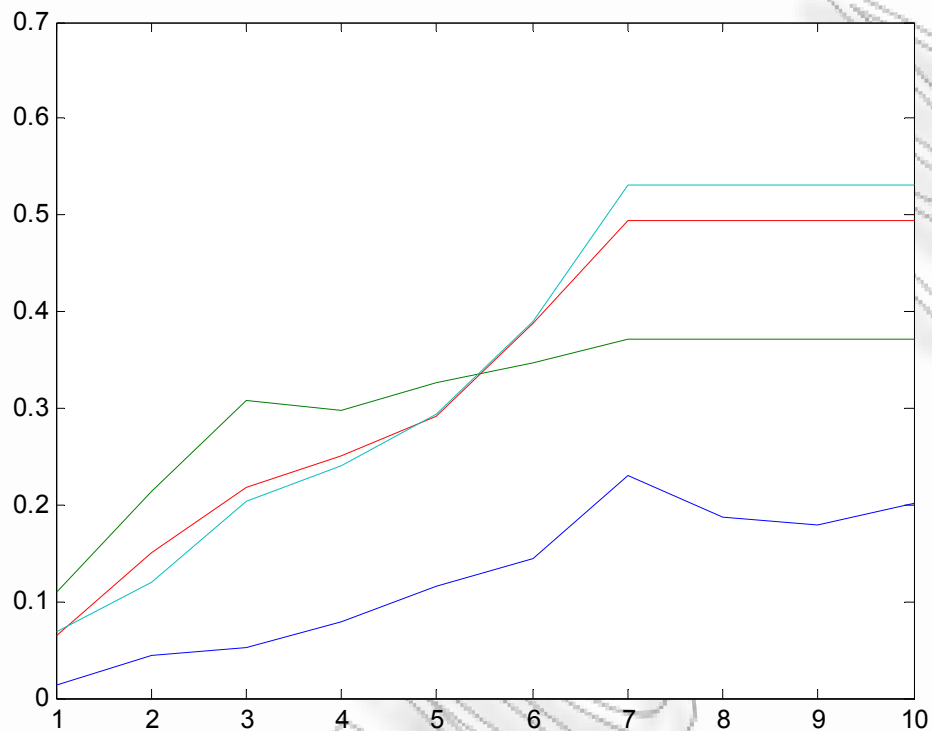
$$\text{mean}(2) = 1.9570, \text{std}(2) = 0.1440$$

$$\text{mean}(3) = 3.0057, \text{std}(3) = 0.0503$$

$$\text{mean}(4) = 4.0033, \text{std}(4) = 0.0262$$

Με την αποκοπή των 5 πρώτων τιμών, η ακρίβεια στην κατάταξη αυξάνεται περίπου κατά μια τάξη μεγέθους.

Στο Elman-10 για τα περισσότερα μουσικά είδη παρατηρείται σταθεροποίηση της απόκλισης για αριθμό δειγμάτων >7 .



Σχήμα 44 Ακρίβεια υπολογισμού δείγματος για Elman-10; Άξονας X: πλήθος τιμών που χρησιμοποιήθηκαν, Y: απόκλιση από την πραγματική τιμή

Από τη μελέτη των παραπάνω αριθμητικών αποτελεσμάτων και διαγραμμάτων προκύπτει ότι και με τη χρήση του δικτύου Elman-10, μπορεί να γίνει με επιτυχία κατάταξη μελωδίας σε μουσικά είδη. Μάλιστα, αν αποκοπεί μέρος της αρχής του μουσικού κομματιού, η κατάταξη με χρήση του δικτύου Elman έχει τη μεγαλύτερη ακρίβεια από όλα τα δίκτυα. Η εκπαίδευση όμως του δικτύου είναι πιο απαιτητική σε χρόνο και υπολογιστικούς πόρους.

8. Συμπεράσματα

Η δημιουργία τεράστιων βάσεων δεδομένων που προέρχονται τόσο από την επανάκτηση ήδη υπάρχοντων αναλογικών αρχείων όσο και από νέο περιεχόμενο, απαιτεί όλο και περισσότερο αξιόπιστα και γρηγορότερα εργαλεία για την ανάλυση του περιεχομένου. Τα εργαλεία αυτά μπορούν να χρησιμοποιηθούν στις αναζητήσεις/ερωτήσεις περιεχομένου. Για το λόγο αυτό, τα μουσικά είδη είναι κρίσιμοι περιγραφείς δεδομένου ότι έχουν χρησιμοποιηθεί ευρέως για χρόνια για να οργανώσουν τους καταλόγους μουσικής, τις βιβλιοθήκες και τα καταστήματα μουσικής. Παρά τη χρήση τους, τα μουσικά είδη παραμένουν μια μη καλώς ορισμένη έννοια, γεγονός που καθιστά το πρόβλημα της αυτόματης ταξινόμησης μη τετριμμένο.

Στην πραγματικότητα, το πλήθος της ψηφιακής μουσικής απαιτεί αποτελεσματικούς τρόπους αναζήτησης, οργάνωσης και δυναμικής ενημέρωσης των συλλογών μουσικής και σίγουρα απαιτεί νέα μέσα για αυτόματο σχολιασμό. Συγχρόνως, ακόμα κι αν όροι όπως τζαζ, ροκ ή ποπ χρησιμοποιούνται ευρέως, παραμένουν συχνά αόριστα καθορισμένοι έτσι ώστε το πρόβλημα της αυτόματης ταξινόμησης είδους να γίνεται ένας μη τετριμμένος στόχος.

Στα πλαίσια της εργασίας αυτής μελετήθηκε η αυτόματη ταξινόμηση μελωδίας σε μουσικά είδη με τη χρήση τεχνητών νευρωνικών δικτύων. Τα νευρωνικά δίκτυα αποτελούνται από απλά στοιχεία που λειτουργούν παράλληλα. Τα στοιχεία στα δίκτυα αυτά προσομοιώνουν σε ένα βαθμό τα βιολογικά νευρωνικά συστήματα. Όπως στη φύση, οι συνδέσεις μεταξύ των στοιχείων καθορίζουν κατά ένα μεγάλο μέρος τη συνάρτηση του δικτύου. Μπορεί κανείς να εκπαιδεύσει ένα νευρωνικό δίκτυο ώστε αυτό να προσομοιώνει μια ιδιαίτερη συνάρτηση με τη ρύθμιση των τιμών των συνδέσεων (βάρη) μεταξύ των στοιχείων. Χαρακτηριστικά, τα νευρωνικά δίκτυα ρυθμίζονται, ή εκπαιδεύονται, έτσι ώστε μια συγκεκριμένη είσοδος να οδηγεί σε μια συγκεκριμένη επιθυμητή έξοδο.

Στα πλαίσια αυτής της εργασίας, έγινε χρήση διάφορων περιγραφών που μπορούν να χαρακτηρίσουν μια μουσική ακολουθία, προκειμένου να εκπαιδευτεί ένα νευρωνικό δίκτυο ώστε να μπορεί να ξεχωρίζει το είδος της μουσικής που δέχεται ως είσοδο. Ύστερα από διάφορες δοκιμές, επιλέχθηκαν τελικά τα παρακάτω έξι χαρακτηριστικά γνωρίσματα ως είσοδοι στο νευρωνικό δίκτυο:

- RMS: μετράει ουσιαστικά την ενέργεια του σήματος.
- ZeroCross: ένας απλός δείκτης που μετρά πόσες φορές αλλάζει πρόσημο η κυματομορφή.
- Centroid: επιστρέφει το κέντρο βάρους των δεδομένων (1^η κύρια ροπή).
- Spread: επιστρέφει την διακύμανση (variance) των δεδομένων (2^η ροπή).
- Skewness: επιστρέφει τον συντελεστή λοξότητας των δεδομένων. Πρόκειται για την 3^η κύρια ροπή, ονομάζεται skewness και αποτελεί ένα μέτρο συμμετρίας της κατανομής.
- Regularity: Η ασυμμετρία ενός φάσματος είναι ο βαθμός μεταβολής των διαδοχικών κορυφών του φάσματος.

Παράλληλα, έγιναν δοκιμές σε διαφορετικές αρχιτεκτονικές ΤΝΔ. Κάθε ξεχωριστό είδος ΤΝΔ τροφοδοτήθηκε με μερικές ώρες μουσικής από διαφορετικά μουσικά είδη, έτσι ώστε να εκπαιδευτεί και να είναι ικανό να αναγνωρίζει τα διαφορετικά αυτά είδη. Τα είδη δικτύων που δοκιμάστηκαν ήταν τα παρακάτω:

- Feed-Forward Neural Network (FF): Πολυστρωματικά δίκτυα με βάρη που καθορίζονται αποκλειστικά από την έξοδο του αμέσως προηγούμενου στρώματος και το τελευταίο στρώμα είναι η έξοδος του δικτύου. Χρησιμοποιήθηκαν 2 στρώματα με συναρτήσεις μεταφοράς tansig και purelin σε κρυφό στρώμα και στρώμα εξόδου αντίστοιχα.
- Cascade-Forward Neural Network (CF): Πολυστρωματικά δίκτυα με βάρη που καθορίζονται από την έξοδο του αμέσως προηγούμενου στρώματος αλλά και από την είσοδο του νευρωνικού δικτύου. Χρησιμοποιήθηκαν 2 στρώματα με συναρτήσεις μεταφοράς tansig και purelin σε κρυφό στρώμα και στρώμα εξόδου αντίστοιχα.
- Elman back-propagation Neural Network (Elman): Πολυστρωματικά δίκτυα με βάρη που καθορίζονται από την έξοδο του αμέσως προηγούμενου στρώματος και από την έξοδο του ίδιου στρώματος. Χρησιμοποιήθηκαν 2 στρώματα με συναρτήσεις μεταφοράς tansig και purelin σε κρυφό στρώμα και στρώμα εξόδου αντίστοιχα.

Από τη μελέτη των αριθμητικών αποτελεσμάτων και διαγραμμάτων προκύπτει ότι μπορεί να γίνει αυτόματη ταξινόμηση μελωδίας με επιτυχία με τη χρήση και των τριών τύπων δικτύων. Πέρα από την επαλήθευση της καταλληλότητας των ΤΝΔ για την ταξινόμηση μελωδίας σε μουσικά είδη, μπορούν να εξαχθούν και τα ακόλουθα επιμέρους συμπεράσματα:

- Με τη χρήση των έξι περιγραφέων που παραθέσαμε προηγουμένως, δε χρειάζεται μεγάλος αριθμός νευρώνων για την επιτυχή ταξινόμηση. Πιο συγκεκριμένα, παρατηρήθηκε ότι τα ΤΝΔ έχουν παρόμοια απόδοση είτε με τη χρήση 10 είτε με την χρήση 20 κρυμμένων νευρώνων.
- Τα δίκτυα FF και CF έχουν παρόμοια συμπεριφορά (απόδοση και ταχύτητα εκπαίδευσης). Συγκεκριμένα, παρατηρήθηκε ότι στα συγκεκριμένα πειράματα και είδη μουσικής, τα δίκτυα αυτά εκπαιδεύονται πολύ γρήγορα καθώς αρκούν μονάχα 15 με 30 επαναλήψεις προκειμένου να επιτευχθεί σταθερή απόδοση.
- Με τη χρήση δικτύου Elman, μπορεί να γίνει με επιτυχία κατάταξη μελωδίας σε μουσικά είδη. Μάλιστα, αν αποκοπεί μέρος της αρχής του μουσικού κομματιού, η κατάταξη με χρήση του δικτύου Elman έχει αρκετά μεγαλύτερη ακρίβεια από τα υπόλοιπα δίκτυα. Η εκπαίδευση όμως αυτού του τύπου δικτύου είναι πιο απαιτητική σε χρόνο και υπολογιστικούς πόρους. Συνεπώς, αν η εκπαίδευση του δικτύου είναι μια διαδικασία που δε γίνεται συχνά, τότε η χρήση ΤΝΔ τύπου Elman είναι προτιμότερη από τα δίκτυα τύπου FF και CF.
- Έγιναν επίσης μετρήσεις για τον υπολογισμό των δειγμάτων που απαιτούνται για την ακριβή κατάταξη της μελωδίας. Στα πειράματά μας, με περίοδο δειγματοληψίας 10 sec, για τα περισσότερα μουσικά είδη παρατηρείται πρακτικά σταθεροποίηση της απόκλισης για αριθμό δειγμάτων > 7.

Βιβλιογραφία

- [Acco82] J. Acoust. Eng. Soc., Vol. 30, No. 6, 396-406. Pollard H. F. & Jansson, E.V. (1982). A tristimulus method for the specification of musical timbre, *Acustica*, vol. 51, 162-171.
- [Aucu02] J.J. Aucouturier, F. Pachet, “Music similarity measures: what’s the use?”, in Proc. of the 3rd Int. Symposium on Music Information Retrieval, 2002.
- [Aucu03] J.J. Aucouturier, F. Pachet, “Representing musical genre: a state of the art”, in *Journal of New Music Research*, vol. 32, no. 1, pp. 83-93, 2003.
- [Bell00] Bello, J. P., G. Monti and M. L. R. Sandler, (2000), An Implementation of Automatic Transcription of Monophonic Music with a Blackboard System. In Proceedings of the Irish Signals and Systems conference (ISSC 2000), Dublin, Ireland, June.
- [Bere02] A. Berenzweig, D. Ellis, S. Lawrence, “Using voice segments to improve artist classification of music”, in Proc. of the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, 2002.
- [Burre03] J.J. Burred, A. Lerch, “A hierarchical approach to automatic musical genre classification”, in Proc. Of the 6th Int. Conf. on Digital Audio Effects (DAFx), London, UK, 2003.
- [Casa05] N. Casagrande, D. Eck, B. K., “Geometry in sound: a speech/music audio classifier inspired by an image classifier”, in Proc. Of the Int. Computer Music Conference (ICMC), 2005.
- [Chai06] W. Chai, “Semantic segmentation and summarization of music”, in *IEEE Signal Processing Magazine: Special Issue on Semantic Retrieval of Multimedia*, March 2006.
- [Dem98] H. Demuth, M. Beale, M. Hagan, *Neural Network Toolbox™ 6, User’s Guide*, 2008
- [Dep98] Depalle, (1998), Feature Extraction and Temporal Segmentation of Acoustic Signals. Proceedings of the ICMC.
- [Flet08] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (2nd ed.), Springer, 1998.
- [Fuku75] Fukushima, Kunihiko, *Cognitron: A self-organizing multilayered neural network*, *Biological Cybernetics*, Springer Berlin, pp: 121-136, Volume: 20, Issue: 3, 1975
- [Gall90] Gallant, S. I. (1990). Perceptron-based learning algorithms. *IEEE Transactions on Neural Networks*, vol. 1, no. 2, pp. 179–191.
- [Gome03] E. Gomez, A. Klapuri, B. Meudic, “Melody description and extraction in the context of music content processing”, in *Journal of New Music Research*, Vol. 32(1), 2003.
- [Gouy04] F. Gouyon, S. Dixon, E. Pampalk, G. Widmer, “Evaluating rhythmic descriptors for musical genre classification”, in AES 25th International Conference, London, England, 2004.
- [Gouy05] F. Gouyon, S. Dixon, “A review of automatic rhythm description system”, in *Computer Music Journal*, vol. 29, pp. 34-54, 2005

- [Hutc87] Hutchison, W.R. & Stephens, K.R. The Airline Marketing Tactician (AMT): a commercial application of adaptive networking. Proceedings of the first International Conference on Neural Networks, 4: 753-756. IEEE Press, 1987.
- [Jens99] Jensen, K. & Arnspang, J. (1999). Binary Tree Classification of Musical Instruments, Proceedings of the ICMC, Beijing, China, 414-417.
- [Jens99] Jensen, Timbre Models of Musical Sounds, Rapport 99/7, University of Copenhagen, 1999.
- [Jusl00] Juslin, P. N. Cue utilization in communication of emotion in music performance: relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 1797-813, 2000
- [Ked86] Kedem. B., (1986) Spectral analysis and discrimination by zero-crossings. *Proc. IEEE*, 74(11), 1477-1493
- [Klap03] A. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness", in *IEEE Trans. Speech and Audio Proc.*, 11(6), 804-816, 2003.
- [Krim94] Krimphoff, J., McAdams, S. & Winsberg, S, (Caractérisation du timbre des sons complexes. II : Analyses acoustiques et quantification psychophysique. *Journal de Physique*, 4(C5), 625-628, 1994
- [Kris01] Kristoffer Jensen and Declan Murphy, Segmenting Melodies into Notes, In Proceedings of the DSAGM, Copenhagen, Denmark, 2001
- [Lidy05] T. Lidy, A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification", in Proc. of the 6th Int. Symposium on Music Information Retrieval, London, UK, 2005.
- [Mand05] M. Mandel, D. Ellis, "Song-level features and support vector machines for music classification", in Proc. of the 6th Int. Symposium on Music Information Retrieval, London, UK, 2005.
- [Mark07] Mark R. Petersen, "Musical Analysis and Synthesis in Matlab", Applied Mathematics Department, University of Colorado, 2007
- [Masr96] Masri, P. and Bateman, A. (1996). Improved modelling of attack transient in music analysis-resynthesis. In Proceedings of the International Computer Music Conference, Hong Kong, 100-104.
- [Meng05] A. Meng, P. Ahrendt, J. Larsen, "Improving music genre classification by short-time feature integration", in Proceedings of the 6th Int. Symposium on Music Information Retrieval, London, UK, 2005.
- [Mins69] Minsky M. L. and Papert S. A. 1969. Perceptrons. Cambridge, MA: MIT Press.
- [McCu43] McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115-133.
- [Pach00] F.Pachet, D. Cazaly, "A taxonomy of musical genres", in Proc. Content-Based Multimedia Information Access (RIAO), Paris, France, 2000.
- [Pamp05] E. Pampalk, A. Flexer, G. Widmer, "Improvements of audio based music similarity and genre classification?", in Proc. of the 6th Int. Symposium on Music Information Retrieval, London, UK, 2005.
- [Peet04] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project", CUIDADO I.S.T. Project Report, 2004.

- [Perr99] D. Perrott, R. O. Gjerdingen, "Scanning the dial: an exploration of factors in the identification of musical style", Research Notes, Department of music, Northwestern University, Illinois, 1999.
- [Rabi77] Rabiner, L. R., (1977) On the use of autocorrelation analysis for pitch detection. IEEE Trans. ASSP, Vol. ASSP-25, No. 1, February, 23-33.
- [Raub02] A. Rauber, E. Pampalk, D. Merkl, "Using psycho-acoustic models and self-organizing maps to create ahierarchical structuring of music by sound similarity", in Proc. of the 3rd Int. Conf. on Music Information Retrieval, Paris, France, 2002.
- [Rose57] Rosenblatt, Frank (1957), The Perceptron--a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory.
- [Ross00] Rossignol, S., X. Rodet, J: Soumagne, J.-L. Colette and P. Depalle (2000), Automatic characterisation of musical signals: feature extraction and temporal segmentation. Journal of New Musical Research.
- [Scar05] N. Scaringella, G. Zoia, "On the modeling of time information for automatic genre recognition systems in audio signals", in Proceedings of the 6th Int. Symposium on Music Information Retrieval, London, UK, 2005.
- [Scar06] N. Scaringella, G. Zoia, D. Mlynek, "Automatic genre classification of music content: a survey", IEEE Signal Processing Magazine, Vol. 23, Issue 2, 133-141, 2006
- [Scari05] N. Scaringella, D. Mlynek, "A mixture of support vector machines for audio classification", Music Information Retrieval Evaluation exchange (MIREX) website, http://www.music-ir.org/evaluation/mirexresults/articles/audio_genre/scaringella.pdf, 2005.
- [Sche97] Scheirer, E. and M. Slaney (1997), Construction and evaluation of a robust multi features speech/music discriminator, IEEE Transactions on Acoustics, Speech, and Signal Processing (ICASSP'97), 1331-1334.
- [Shao04] X. Shao, C. Xu, M. Kankanhalli, "Unsupervised classification of musical genre using hidden Markov model", in IEEE Int. Conf. of Multimedia Explore (ICME), Taipei, Taiwan, China, 2004.
- [Solt98] H. Soltau, T. Schultz, M. Westphal, A. Waibel, "Recognition of music types", in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Seattle, USA, 1998.
- [Sutt98] Richard S. Sutton, Andrew G. Barto, "Reinforcement learning: an introduction", MIT Press, 1998
- [Tzan02] G. Tzanetakis, P. Cook, "Musical genre classification of audio signals", in IEEE Transactions on Speech and Audio Processing, Vol. 10, No. 5, July 2002.
- [West05] K. West, S. Cox, "Finding an optimal segmentation for audio genre classification", in Proceedings of the 6th Int. Symposium on Music Information Retrieval, London, UK, 2005.
- [Zoia04] G. Zoia, R. Zhou, D. Mlynek, "A multi-timbre chord/harmony analyzer based on signal processing and neural networks", in Proc. IEEE Int. Workshop on Multimedia Signal Processing, Siena, Italy, 2004.
- [Zwi57] Zwicker, E., G. Flottorp, S. S. Stevens, (1957), Critical band width in loudness summation. J. Acoust. Soc. Am. Vol. 29, No. 5, May, 548-557.

- [Werb74] Paul J. Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University, 1974
- [Werb90] Paul J. Werbos. Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, Volume 78, Issue 10, 1550 - 1560, Oct 1990
- [Widr90] Widrow, B., Lehr, M.A., "30 years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation," *Proc. IEEE*, vol 78, no 9, pp. 1415–1442, (1990).

Παράρτημα –Κώδικας

Αρχικοποίηση Δεδομένων

```
%This is not a program, this is a list of commands used sequentially to collect
music features.

%parameters
frame_length=10;

%%get frames for all data
cd 'paradosiaka';
paradosiaka = mirframe(miraudio('Folder'),'Length',frame_length);
cd '../';
cd 'classic';
classic = mirframe(miraudio('Folder'),'Length',frame_length);
cd '../';
cd 'disco';
disco = mirframe(miraudio('Folder'),'Length',frame_length);
cd '../';

%paradosiaka features

% timbre energy : root mean square energy
p_rms = mirgetdata(mirrms(paradosiaka));
p_rms = [ p_rms(1:46,1)' p_rms(1:28,2)' p_rms(1:42,3)' ]'

%timbre temporal
p_zc = mirgetdata(mirzerocross(paradosiaka));
p_zc = [ p_zc(1:46,1)' p_zc(1:28,2)' p_zc(1:42,3)' ]'

%spectral shape features
p_centroid = mirgetdata(mircentroid(paradosiaka));
```



```
p_centroid = [ p_centroid(1:46,1)' p_centroid(1:28,2)' p_centroid(1:42,3)' ]'

p_spread = mirgetdata(mirspread(paradosiaka));
p_spread = [ p_spread(1:46,1)' p_spread(1:28,2)' p_spread(1:42,3)' ]'

p_skewness = mirgetdata(mirskewness(paradosiaka));
p_skewness = [ p_skewness(1:46,1)' p_skewness(1:28,2)' p_skewness(1:42,3)' ]'

%p_mfcc = mirgetdata(mirmfcc(paradosiaka));

%regularity
p_regularity = mirgetdata(mirregularity(paradosiaka));
p_regularity = [ p_regularity(1:46,1)' p_regularity(1:28,2)' p_regularity(1:42,3)'
]'

%pitch
%p_pitch = mirgetdata(mirpitch(paradosiaka));

%classic features
c_rms = mirgetdata(mirrms(classic));
c_rms = [ c_rms(1:83,1)' c_rms(:,2)' ]';

c_zc = mirgetdata(mirzerocross(classic));
c_zc = [ c_zc(1:83,1)' c_zc(:,2)' ]';

c_centroid = mirgetdata(mircentroid(classic));
c_centroid = [ c_centroid(1:83,1)' c_centroid(:,2)' ]';

c_spread = mirgetdata(mirspread(classic));
c_spread = [ c_spread(1:83,1)' c_spread(:,2)' ]';

c_skewness = mirgetdata(mirskewness(classic));
```

```

c_skewness = [ c_skewness(1:83,1)' c_skewness(:,2)' ]';

c_regularity = mirgetdata(mirregularity(classic));
c_regularity = [ c_regularity(1:83,1)' c_regularity(:,2)' ]';

%disco features
d_rms = mirgetdata(mirrms(disco));
d_rms = [ d_rms(1:45,1)' d_rms(1:41,2)' d_rms(1:36,3)' ]'

d_zc = mirgetdata(mirzerocross(disco));
d_zc = [ d_zc(1:45,1)' d_zc(1:41,2)' d_zc(1:36,3)' ]'

d_centroid = mirgetdata(mircentroid(disco));
d_centroid = [ d_centroid(1:45,1)' d_centroid(1:41,2)' d_centroid(1:36,3)' ]'

d_spread = mirgetdata(mirspread(disco));
d_spread = [ d_spread(1:45,1)' d_spread(1:41,2)' d_spread(1:36,3)' ]'

d_skewness = mirgetdata(mirskewness(disco));
d_skewness = [ d_skewness(1:45,1)' d_skewness(1:41,2)' d_skewness(1:36,3)' ]'

d_regularity = mirgetdata(mirregularity(disco));
d_regularity = [ d_regularity(1:45,1)' d_regularity(1:41,2)' d_regularity(1:36,3)'
]'

%dimotika
dim = mirframe( miraudio('dim1') + miraudio('dim2')+ miraudio('dim3')
,'length',frame_length) ;

dim_rms = mirgetdata(mirrms(dim));
dim_zc = mirgetdata(mirzerocross(dim));

```

```
dim_centroid = mirgetdata(mircentroid(dim));
dim_spread = mirgetdata(mirspread(dim));
dim_skewness = mirgetdata(mirskewness(dim));
dim_regularity = mirgetdata(mirregularity(dim));

%kritika
krit = mirframe( miraudio('krit1') , 'Length', frame_length) ;
krit_rms = mirgetdata(mirrms(krit));
krit_zc = mirgetdata(mirzerocross(krit));
krit_centroid = mirgetdata(mircentroid(krit));
krit_spread = mirgetdata(mirspread(krit));
krit_skewness = mirgetdata(mirskewness(krit));
krit_regularity = mirgetdata(mirregularity(krit));
```

Κώδικας για τη δημιουργία και εκπαίδευση δικτύου cf

```
clear
load c_features
load d_features
load p_features
load dim_features
load k_features

par_input = [ p_rms p_zc p_centroid p_spread p_skewness p_regularity ]' ;
disco_input = [ d_rms d_zc d_centroid d_spread d_skewness d_regularity ]' ;
classic_input = [ c_rms c_zc c_centroid c_spread c_skewness c_regularity ]' ;
```

```
dim_input = [ dim_rms dim_zc dim_centroid dim_spread dim_skewness dim_regularity ]'  
;  
  
krit_input = [ krit_rms krit_zc krit_centroid krit_spread krit_skewness  
krit_regularity ]' ;  
  
Tc=ones(1,211);  
Tp=2*ones(1,116);  
Tdim=3*ones(1,45);  
Td=4*ones(1,122);  
  
P=[ classic_input par_input dim_input disco_input ];  
T=[ Tc Tp Tdim Td ];  
  
ctest = classic_input(:,5:25);  
ptest = par_input(:,5:25);  
dtest = disco_input(:,5:25);  
dimtest = dim_input(:,5:25);  
ktest = krit_input(:,5:25);  
  
net = create_cf_net(P,T,10);  
  
c = sim(net,ctest)';  
d = sim(net,dtest)';  
dim = sim(net,dimtest)';  
p = sim(net,ptest)';  
  
Y= [c p dim d ones(1,21)' 2*ones(1,21)' 3*ones(1,21)' 4*ones(1,21)'];  
  
figure;  
plot(Y)
```

```
mean(c)
std(c)
mean(p)
std(p)
mean(dim)
std(dim)
mean(d)
std(d)

figure
plot_accuracy(c,p,dim,d)
```

Κώδικας για τη δημιουργία και εκπαίδευση δικτύου elman

```
clear
load c_features
load d_features
load p_features
load dim_features
load k_features

par_input = [ p_rms p_zc p_centroid p_spread p_skewness p_regularity ]' ;

disco_input = [ d_rms d_zc d_centroid d_spread d_skewness d_regularity ]' ;

classic_input = [ c_rms c_zc c_centroid c_spread c_skewness c_regularity ]' ;

dim_input = [ dim_rms dim_zc dim_centroid dim_spread dim_skewness dim_regularity ]'
;

krit_input = [ krit_rms krit_zc krit_centroid krit_spread krit_skewness
krit_regularity ]' ;
```

```
Tc=ones(1,211);
Tp=2*ones(1,116);
Tdim=3*ones(1,45);
Td=4*ones(1,122);

P=[ classic_input par_input dim_input disco_input ];
T=[ Tc Tp Tdim Td ];

ctest = classic_input(:,5:25);
ctestSeq = con2seq(ctest);
ptest = par_input(:,5:25);
ptestSeq = con2seq(ptest);
dtest = disco_input(:,5:25);
dtestSeq = con2seq(dtest);
dimtest = dim_input(:,5:25);
dimtestSeq = con2seq(dimtest);

ktest = krit_input(:,5:25);
ktestSeq = con2seq(ktest);

net = create_elman_net(P,T,10);

c = sim(net,ctestSeq)';
d = sim(net,dtestSeq)';
dim = sim(net,dimtestSeq)';
p = sim(net,ptestSeq)';
```

```
Y= [cell2mat(c) cell2mat(p) cell2mat(dim) cell2mat(d) ones(1,21)' 2*ones(1,21)'
3*ones(1,21)' 4*ones(1,21)'];

figure;
plot(Y)
mean(cell2mat(c))
mean(cell2mat(p))
mean(cell2mat(dim))
mean(cell2mat(d))

std(cell2mat(c))
std(cell2mat(p))
std(cell2mat(dim))
std(cell2mat(d))

figure
plot_accuracy(cell2mat(c), cell2mat(p),cell2mat(dim) , cell2mat(d))
```

Κώδικας για τη δημιουργία και εκπαίδευση δικτύου ff

```
clear
load c_features
load d_features
load p_features
load dim_features
load k_features

par_input = [ p_rms p_zc p_centroid p_spread p_skewness p_regularity ]' ;

disco_input = [ d_rms d_zc d_centroid d_spread d_skewness d_regularity ]' ;
```

```
classic_input = [ c_rms c_zc c_centroid c_spread c_skewness c_regularity ]' ;

dim_input = [ dim_rms dim_zc dim_centroid dim_spread dim_skewness dim_regularity ]'
;

krit_input = [ krit_rms krit_zc krit_centroid krit_spread krit_skewness
krit_regularity ]' ;

Tc=ones(1,211);
Tp=2*ones(1,116);
Tdim=3*ones(1,45);
Td=4*ones(1,122);

P=[ classic_input par_input dim_input disco_input ];
T=[ Tc Tp Tdim Td ];

ctest = classic_input(:,5:25);
ptest = par_input(:,5:25);
dtest = disco_input(:,5:25);
dimtest = dim_input(:,5:25);

ktest = krit_input(:,5:25);

net = create_ff_net(P,T,10);

c = sim(net,ctest)';
d = sim(net,dtest)';
dim = sim(net,dimtest)';
p = sim(net,ptest)';
```



```
Y= [c p dim d ones(1,21)' 2*ones(1,21)' 3*ones(1,21)' 4*ones(1,21)'];

figure;
plot(Y)
mean(c)
std(c)
mean(p)
std(p)
mean(dim)
std(dim)
mean(d)
std(d)

figure
plot_accuracy(c,p,dim,d)
```

Βοηθητικές Συναρτήσεις

```
function net = create_cf_net(inputs,targets,numHiddenNeurons)
%CREATE_FIT_NET Creates and trains a fitting neural network.
%
% NET = CREATE_FIT_NET(INPUTS,TARGETS) takes these arguments:
%   INPUTS - R x Q matrix of Q R-element input samples
%   TARGETS - S x Q matrix of Q S-element associated target samples
% arranged as columns, and returns these results:
%   NET - The trained neural network
```

```
% Create Network
net = newcf(inputs,targets,numHiddenNeurons);

net.divideParam.trainRatio = 70/100;           % Adjust as desired
net.divideParam.valRatio = 15/100;           % Adjust as desired
net.divideParam.testRatio = 15/100;         % Adjust as desired

% Train and Apply Network
[net,tr] = train(net,inputs,targets);
outputs = sim(net,inputs);

% Plot
figure
plotperf(tr)
```

```
function net = create_elman_net(inputs,targets,numHiddenNeurons)
%CREATE_FIT_NET Creates and trains a fitting neural network.
%
% NET = CREATE_FIT_NET(INPUTS,TARGETS) takes these arguments:
%   INPUTS - RxQ matrix of Q R-element input samples
%   TARGETS - SxQ matrix of Q S-element associated target samples
% arranged as columns, and returns these results:
%   NET - The trained neural network

% Create Network
net = newelm(inputs,targets,numHiddenNeurons);
net.trainFcn='trainrp';

net.divideParam.trainRatio = 70/100;           % Adjust as desired
```

```
net.divideParam.valRatio = 15/100;           % Adjust as desired
net.divideParam.testRatio = 15/100;        % Adjust as desired

%arrange these as sequences
Pseq = con2seq(inputs);
Tseq = con2seq(targets);

% Train and Apply Network
[net,tr] = train(net,Pseq,Tseq);
outputs = sim(net,Pseq);

% Plot
figure
plotperf(tr)
```

```
function net = create_ff_net(inputs,targets,numHiddenNeurons)
%CREATE_FIT_NET Creates and trains a fitting neural network.
%
% NET = CREATE_FIT_NET(INPUTS,TARGETS) takes these arguments:
%   INPUTS - RxQ matrix of Q R-element input samples
%   TARGETS - SxQ matrix of Q S-element associated target samples
% arranged as columns, and returns these results:
%   NET - The trained neural network

% Create Network
net = newff(inputs,targets,numHiddenNeurons);

net.divideParam.trainRatio = 70/100;       % Adjust as desired
```

```
net.divideParam.valRatio = 15/100;           % Adjust as desired
net.divideParam.testRatio = 15/100;        % Adjust as desired

% Train and Apply Network
[net,tr] = train(net,inputs,targets);
outputs = sim(net,inputs);

% Plot
figure
plotperf(tr)
```

```
function plot_accuracy(c,p,dim,d)

for step=1:10
    j=1;
    cc=[];
    pp=[];
    dimdim=[];
    dd=[];
    for i=1:step:numel(c)
        cc(j)=c(i);
        dd(j)=d(j);
        pp(j)=p(j);
        dimdim(j)=dim(j);
        j=j+1;
    end
    mean_c(step)=abs(mean(cc)-1);
    mean_p(step)=abs(mean(pp)-2);
    mean_dim(step)=abs(mean(dimdim)-3);
```

```
mean_d(step)=abs(mean(dd)-4);  
end  
  
yy=[ mean_c' mean_p' mean_dim' mean_d' ] ;  
plot(yy);
```