



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

## **Προσαρμοζόμενες τεχνικές αναγνώρισης συναισθημάτων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Θεόδωρος Ε. Σκάρας

**Επιβλέπων Καθηγητής :** Στέφανος Κόλλιας, Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2010





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

## Προσαρμοζόμενες τεχνικές αναγνώρισης συναισθημάτων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Θεόδωρος Ε. Σκάρας

Επιβλέπων Καθηγητής : Στέφανος Κόλλιας, Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 3<sup>η</sup> Νοεμβρίου 2010

.....  
Σ. Κόλλιας  
Καθηγητής Ε.Μ.Π.

.....  
Α. Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π.

.....  
Γ. Στάμου  
Λέκτορας Ε.Μ.Π.

Θεόδωρος Ε. Σκάρας

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

Copyright © Θεόδωρος Ε. Σκάρας

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου

# Περιεχόμενα

Περιεχόμενα .....	5
Περίληψη .....	8
Abstract .....	9
Εισαγωγή .....	10
Σκοπός- Κίνητρα και Διάρθρωση της διπλωματικής .....	11
<b>Κεφάλαιο 1 : Τεχνητά Νευρωνικά Δίκτυα .....</b>	<b>12</b>
1.1 Εισαγωγή.....	12
1.2 Το μοντέλο McCulloch-Pitts .....	12
1.3 Το δίκτυο Perceptron .....	14
1.4 Μάθηση στο δίκτυο Perceptron .....	15
1.5 Το μοντέλο Multi Layer Perceptron(MLP) .....	16
1.6 Μάθηση των πολυστρωματικών δικτύων Perceptron.....	17
1.6.1 Επιβλεπόμενη μάθηση.....	17
1.6.2 Ενισχυτική μάθηση.....	19
1.6.3 Μη επιβλεπόμενη μάθηση (αυτο-οργανούμενη μάθηση) .....	20
1.7 Τεχνητά αναδρομικά νευρωνικά δίκτυα.....	20
1.7.1 Γενικά για τα τεχνητά αναδρομικά νευρωνικά δίκτυα.....	20
1.7.2 Λειτουργία του αναδρομικού νευρωνικού δικτύου Elman.....	21
<b>Κεφάλαιο 2: Τεχνικές καταγραφής ανθρώπινης έκφρασης.....</b>	<b>24</b>
2.1 Εισαγωγή.....	24
2.2 Το πρότυπο MPEG – 4 .....	24

2.3 Τεχνικές καταγραφής παραμόρφωσης χαρακτηριστικών ανθρώπινου προσώπου.....	26
2.3.1 MPEG – 4 και πρόσωπο – FDPs και FAPs.....	26
2.3.2 Σύνολο Παραμέτρων Προσδιορισμού Προσώπου (FDPs).....	26
2.3.3 Σύνολο Παραμέτρων Κίνησης Προσώπου (FAPs).....	28
2.3.4 Πίνακες παρεμβολής FAPs.....	33
2.3.5 Καθορισμός του μοντέλου του προσώπου.....	34
<b>Κεφάλαιο 3: Αναγνώριση ανθρώπινου συναισθήματος με MLP.....</b>	<b>35</b>
3.1 Εισαγωγή.....	35
3.2 Δεδομένα εισόδου του προβλήματος.....	35
3.3 Υλοποίηση με feed – forward multilayer perceptron.....	38
3.3.1 Γενικά.....	38
3.3.2 Χωρίς προεπεξεργασία δεδομένων.....	39
3.3.3 Αναζήτηση βέλτιστης αρχιτεκτονικής.....	40
3.3.4 Υλοποίηση με είσοδο μόνο των φωνητικών χαρακτηριστικώ.....	41
<b>Κεφάλαιο 4 : Στατιστική ανάλυση των faps.....</b>	<b>43</b>
4.1 Εισαγωγή.....	43
4.2 Συσχέτιση δύο μεταβλητών.....	43
4.3 Συσχέτιση των faps.....	46
4.4 Ομαδοποίηση των faps με βάση τη συσχέτισή τους.....	47
4.5 Εκπαίδευση και έλεγχος απόδοσης mlr με είσοδο που έχει μεταβλητές επιλεγμένες κάθε φορά, με βάση την συσχέτιση τους.....	49
4.5.1 Πείραμα 1.....	50
4.5.2 Πείραμα 2.....	51
4.5.3 Πείραμα 3.....	51

4.6 Αναζήτηση βέλτιστης αρχιτεκτονικής με είσοδο τα χαρακτηριστικά φωνής και τα ασυσχέτιστα faps. ....	53
--	----

## **Κεφάλαιο 5: Ανάλυση ευαισθησίας σε νευρωνικό δίκτυο εκπαιδευμένο για αναγνώριση συναισθήματος. ....**

5.1 Εισαγωγή .....	54
5.2 Ανάλυση ευαισθησίας (Θεωρητικό υπόβαθρο).....	54
5.3 Αλγόριθμος μείωσης δεδομένων εισόδου.....	56
5.4 Υλοποίηση αλγορίθμου μείωσης δεδομένων εισόδου.....	57
5.5 Αναζήτηση βέλτιστης αρχιτεκτονικής.....	64
5.6 Εκπαίδευση και έλεγχος νευρωνικού δικτύου με εισόδους επιλεγμένες με βάση την ανάλυση ευαισθησίας σε συνδυασμό με την στατιστική συσχέτιση των faps....	65
5.6.1 Εκπαίδευση και έλεγχος νευρωνικού δικτύου με μεταβλητές εισόδου όλα τα χαρακτηριστικά φωνής, τα τρία σημαντικότερα ασυσχέτιστα χαρακτηριστικά προσώπου και τα δύο, μη συσχετισμένα μεταξύ τους, σημαντικότερα χαρακτηριστικά προσώπου. ....	68

## **Κεφάλαιο 6: Επίλογος.....**

6.1 Συμπεράσματα.....	69
6.2 Μελλοντικές επεκτάσεις.....	72
<b>Βιβλιογραφία.....</b>	<b>74</b>

## Περίληψη

*Στην παρούσα εργασία εξετάζεται η αναγνώριση της συναισθηματικής κατάστασης ενός ανθρώπου από τα φυσικά χαρακτηριστικά του προσώπου του και την ομιλία του. Η διαδικασία της αναγνώρισης πραγματοποιείται με την βοήθεια ενός πολυστρωματικού νευρωνικού δικτύου προσοτροφodότησης . Επιπλέον, μελετάται η λειτουργία και η εκπαίδευση τέτοιου τύπου νευρωνικών δικτύων, όπως επίσης ο τρόπος με τον οποίο μπορούν να αξιολογηθούν και στη συνέχεια να αφαιρεθούν τα χαρακτηριστικά εισόδου τους με χρήση στατιστικών μεθόδων, όπως η ανάλυση ευαισθησίας και η ανάλυση συσχέτισης . Τα παραγόμενα αποτελέσματα από την χρήση αυτών των μεθόδων, οδηγούν σε σημαντικές πληροφορίες για την επίδραση κάθε εισόδου στο σύστημα καθώς και για τη συσχέτιση μεταξύ των εισόδων. Παράλληλα, επιτυγχάνουν να μειώσουν την διάσταση του προβλήματος, άρα και την πολυπλοκότητά του, καθώς και να βελτιώσουν την ταχύτητα και την απόδοση του δικτύου.*



## Abstract

The present work examines the recognition of the sentimental situation of a person from the natural characteristics of his face and his speech. The process of recognition is taking place with the help of a multi-layer, feed- forward neural network. In addition, the operation and the education of such type of networks are studied, their characteristics of entry are evaluated with use of statistical methods, such as sensitivity analysis and correlation analysis and afterwards are partly removed. The produced results from the use of these methods, lead to important information on the effect of each entry to the system and the correlation between the entries. At the same time, they achieve to decrease the dimension and the complexity of the problem, and also improve the speed and attribution of the network.

## Εισαγωγή

Τα τελευταία χρόνια γινόμαστε μάρτυρες ολοένα και περισσότερο της εφαρμογής των τεχνητών νευρωνικών δικτύων στην επίλυση προβλημάτων. Η ανάπτυξη της τεχνολογίας των τεχνητών νευρωνικών δικτύων αποτελεί πλέον ένα ισχυρό εργαλείο στη μοντελοποίηση προβλημάτων που εμπλέκουν όλες εκείνες τις δυσκολίες που παρουσιάζει ο πραγματικός κόσμος. Πλεονάζουσες πληροφορίες, αντιφατικές, μη ακριβείς, έλλειψη δυνατότητας ενσωμάτωσης γνώσης που έχει ήδη αποκτηθεί κ.α. Όλα αυτά αποτελούν συνηθισμένα –και αρκετά περιοριστικά– χαρακτηριστικά που συναντώνται κατά την αντιμετώπιση σύνθετων πραγματικών προβλημάτων. Η δυνατότητα χειρισμού των πληροφοριών σε ένα τέτοιο περιβάλλον, που η “μηχανή” δεν αντιλαμβάνεται με παραδοσιακές μεθόδους, σε συνδυασμό με την ικανότητα εισαγωγής προηγούμενης εξειδικευμένης εμπειρίας αλλά και με τη δυνατότητα εκπαίδευσης και προσαρμογής σε δυναμικά μεταβαλλόμενα περιβάλλοντα, καθιστούν τα νευρωνικά δίκτυα τον ενδεδειγμένο τρόπο για να ξεπεραστούν όλα εκείνα τα εμπόδια που αναδεικνύονται κατά την προσέγγιση πραγματικών προβλημάτων και προοιωνίζουν τον αυξημένο ρόλο που πρόκειται να παίξουν στο μέλλον.

Η θεωρία των τεχνητών νευρωνικών δικτύων δίνει τη δυνατότητα στις υλοποιήσεις μας να εκπαιδεύονται και να προσαρμόζονται στα νέα δεδομένα που λαμβάνουν, καθιστώντας τα έτσι ικανά να αντεπεξέρχονται σε μεταβαλλόμενες απαιτήσεις με όσο το δυνατόν λιγότερη εξάρτηση από τον ανθρώπινο παράγοντα, παρ’ όλο που αποτελεί ακόμα σημαντικό περιοριστικό παράγοντα για την ικανοποιητική λειτουργία τους ( βλ. επιβλεπόμενη μάθηση) . Σε αυτήν την εργασία το πραγματικό πρόβλημα είναι η αναγνώριση του ανθρώπινου συναισθήματος. Με την χρήση τεχνητών νευρωνικών δικτύων επιχειρούμε να καταστήσουμε δυνατή από τον υπολογιστή μια τέτοια ‘ ανθρώπινη ‘ λειτουργία με όσο το δυνατό μεγαλύτερη επιτυχία και ακρίβεια.

## Σκοπός- Κίνητρα και Διάρθρωση της διπλωματικής

Όπως προκύπτει και από την εισαγωγή, τα τεχνητά νευρωνικά δίκτυα αποτελούν ένα πεδίο έρευνας στο οποίο έχουν επενδυθεί πολλές ώρες εργασίας και αναμένονται υλοποιήσεις υψηλών επιδόσεων που θα ωθήσουν την τεχνολογία σε ακόμα ταχύτερους και αποδοτικότερους τρόπους αντιμετώπισης των σύγχρονων προβλημάτων. Αρχικός στόχος αυτής τη εργασίας είναι να υλοποιηθεί ένα τεχνητό νευρωνικό δίκτυο , το οποίο θα αναγνωρίζει με καλά ποσοστά επιτυχίας τις εναλλαγές του ανθρώπινου συναισθήματος ενός ομιλούντος ανθρώπου. Τα δεδομένα που θα επεξεργάζεται θα βασίζονται στη φωνή και το πρόσωπο του. Στη συνέχεια , σκοπός είναι να μπορέσουμε, σύμφωνα με τις επιταγές της εποχής, να καταφέρουμε να κάνουμε το δίκτυο λιγότερο πολύπλοκο και επομένως ταχύτερο και αποδοτικότερο.

Η παρούσα διπλωματική εργασία ακολουθεί την παρακάτω διάρθρωση :

Στο 1° κεφάλαιο γίνεται μια συνοπτική παρουσίαση των βασικών εννοιών των τεχνητών νευρωνικών δικτύων που ενσωματώνονται στην υλοποίησή μας.

Στο 2° κεφάλαιο γίνεται μια παρουσίαση των τεχνικών καταγραφής ανθρώπινης έκφρασης που χρησιμοποιούνται για την εξαγωγή των χαρακτηριστικών εισόδου του δικτύου.

Στο 3° κεφάλαιο υλοποιείται το δίκτυο αναγνώρισης συναισθήματος.

Στο 4° κεφάλαιο υπολογίζεται η στατιστική συσχέτιση των δεδομένων εισόδου και βάσει των αποτελεσμάτων μειώνουμε την πολυπλοκότητα του δικτύου.

Στο 5° κεφάλαιο πραγματοποιείται ανάλυση ευαισθησίας των δεδομένων εισόδου στην έξοδο και με βάση τα αποτελέσματα της ανάλυσης αυτής και τα αποτελέσματα του προηγούμενου κεφαλαίου μειώνουμε την πολυπλοκότητα του δικτύου.

Στο 6° και τελευταίο κεφάλαιο συγκρίνουμε , ανακεφαλαιωτικά , τα αποτελέσματα όλης της εργασίας και παραθέτουμε τα συμπεράσματά μας.

Τέλος, ακολουθεί η βιβλιογραφία στην οποία βασίστηκε η συγγραφή της παρούσας διπλωματικής.

# Κεφάλαιο 1 : Τεχνητά Νευρωνικά Δίκτυα

Όπως έχει ήδη αναφερθεί, τα τεχνητά νευρωνικά δίκτυα αποτελούν θεμελιώδες κομμάτι στο οποίο στηρίχτηκε το αντικείμενο αυτής της διπλωματικής. Γι' αυτό, στο παρόν κεφάλαιο γίνεται μια συνοπτική παρουσίαση, χωρίς να μπαίνουμε σε μεγάλη μαθηματική λεπτομέρεια, των κεντρικών σημείων της θεωρίας στην οποία έχουν βασιστεί.

## 1.1 Εισαγωγή

Πηγή έμπνευσης της θεωρίας των τεχνητών νευρωνικών δικτύων είναι ο τρόπος λειτουργίας και η δομή του ανθρώπινου εγκεφάλου. Οι δομικοί λίθοι του ανθρώπινου εγκεφάλου είναι οι νευρώνες, οι οποίοι, συνδεόμενοι μεταξύ τους με πολυάριθμες συνδέσεις σχηματίζουν ένα εξαιρετικά πολύπλοκο βιολογικό νευρωνικό δίκτυο. Στην προσπάθειά μας λοιπόν να κατασκευάσουμε ένα υπολογιστικό μοντέλο που να μπορεί να παρουσιάσει ευφυή συμπεριφορά, στραφήκαμε στη δημιουργία μιας δικτυακής δομής όμοιας με αυτή του εγκεφάλου. Τα κλασσικά τεχνητά νευρωνικά αποτελούν μία πολύ απλοποιημένη προσέγγιση του μοντέλου των βιολογικών νευρώνων και σε καμία περίπτωση δεν παρουσιάζουν την πολυπλοκότητα της δομής των βιολογικών τους 'συγγενών'. Παρ' όλα αυτά, μπορούν να αποτελέσουν τη βάση για τη δημιουργία δικτύων που παρουσιάζουν ενδιαφέροντα χαρακτηριστικά, αρκεί να πληρούν δύο προϋποθέσεις:

- οι νευρώνες να έχουν ρυθμιζόμενες παραμέτρους, ώστε να διευκολύνεται η διαδικασία της μάθησης – ιδιότητα γνωστή ως πλαστικότητα των νευρώνων.
- το δίκτυο να αποτελείται από ικανό πλήθος νευρώνων ώστε να επιτυγχάνεται παραλληλισμός της επεξεργασίας και κατανομή της πληροφορίας.

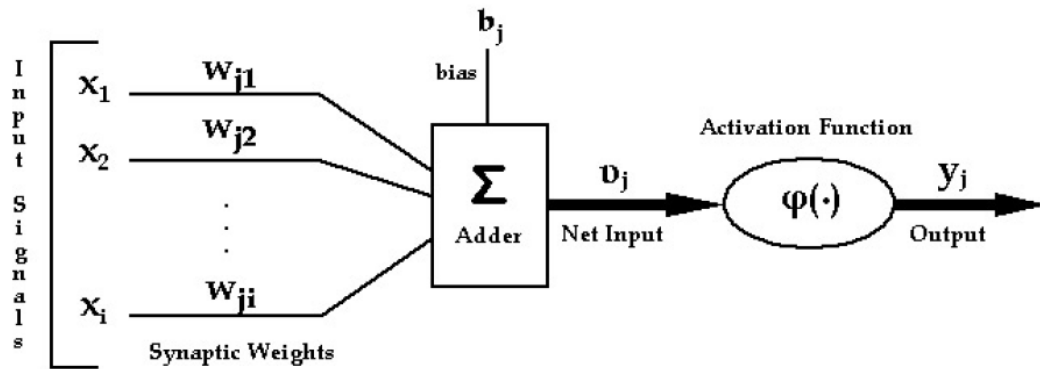
Παράλληλα, πρέπει να αναπτυχθούν οι κατάλληλοι αλγόριθμοι εκπαίδευσης που θα επιτρέψουν στα δημιουργηθέντα δίκτυα να παρουσιάζουν ευφυή συμπεριφορά όπως αυτές που μόλις αναφέρθηκαν. Δηλαδή, θα εκπαιδεύονται σε ένα επιβλεπόμενο περιβάλλον, όπου κάποιος οδηγός θα ελέγχει κάθε φορά τα αποτελέσματα ή το δίκτυο θα αυτοοργανώνεται χωρίς καμία εξωτερική επίβλεψη;

## 1.2 Το μοντέλο McCulloch-Pitts

Το 1943, οι McCulloch και Pitts περιέγραψαν ένα απλό μοντέλο της δραστηριότητας του νευρώνα. Η κατάσταση του νευρώνα προσδιορίζεται από ένα δυαδικό αριθμό  $y_j$  (για τον τυχαίο  $j$ -οστό νευρώνα).

Αν  $y = 0$  , τότε ο νευρώνας είναι ανενεργός

Αν  $y = 1$  , τότε ο νευρώνας είναι ενεργός



Σχήμα 1.1 : Το μοντέλο νευρώνα των McCulloch και Pitts

Τα συναπτικά βάρη είναι πραγματικοί αριθμοί, θετικοί για τις ενισχυτικές συμβολές των συνάψεων και αρνητικοί για τις αναιρετικές συμβολές. Αν  $x_1, x_2, \dots, x_n$  είναι οι εισόδου του νευρώνα τότε το σταθμισμένο άθροισμα των εισόδων που δέχεται ο νευρώνας είναι :

$$u = \sum_{i=1}^n w_i x_i$$

Αν το άθροισμα  $u$  είναι μεγαλύτερο από το κατώφλι  $b_j$  τότε ο νευρώνας ενεργοποιείται, αλλιώς παραμένει ανενεργός. Πιο συγκεκριμένα,

$$y = \varphi(u - b)$$

όπου  $\varphi(\cdot)$ , είναι η βηματική συνάρτηση, δηλαδή

$$\varphi(u) = \begin{cases} 0, & u \leq 0 \\ 1, & u > 0 \end{cases}$$

Το κατώφλι  $b$  και τα συναπτικά βάρη  $w_1, w_2, \dots, w_n$  είναι πραγματικοί αριθμοί. Υπάρχουν και διάφορες παραλλαγές του μοντέλου αυτού, που έγκεινται στη χρησιμοποίηση διαφορετικών μη-γραμμικών συναρτήσεων  $\varphi()$  (που ονομάζονται συναρτήσεις ενεργοποίησης του νευρώνα) που καθορίζουν την έξοδο του νευρώνα. Ορισμένες από αυτές τις παραλλαγές είναι :

Συνάρτηση προσήμου :

$$\varphi(u) = \begin{cases} -1, & u \leq 0 \\ 1, & u > 0 \end{cases}$$

Σιγμοειδής συνάρτηση :

$$\varphi(u) = 1/(1 + e^{-u})$$

Συνάρτηση υπερβολικής εφαπτομένης :

$$\varphi(u) = (1 - e^{-u})/(1 + e^{-u})$$

Συνάρτηση κατωφλιού :

$$\varphi(u) = \begin{cases} 0, & u \leq 0 \\ u, & 0 < u \leq 1 \\ 1, & u > 1 \end{cases}$$

Συνάρτηση κλίσης :

$$\varphi(u) = \begin{cases} 0, & u \leq 0 \\ u, & u > 0 \end{cases}$$

Γραμμική συνάρτηση :

$$\varphi(u) = u$$

### 1.3 Το δίκτυο Perceptron

Η πιο απλή μορφή που μπορεί να λάβει ένα νευρωνικό δίκτυο, είναι αυτή με τη συμμετοχή ενός μόνο νευρώνα. Η μορφή αυτή (που ουσιαστικά ταυτίζεται με τη μορφή που παρουσιάστηκε στο σχήμα 1.1, ονομάζεται μοντέλο Perceptron και η συνάρτηση μεταφοράς του είναι :

$$y = \varphi(u)$$

όπου

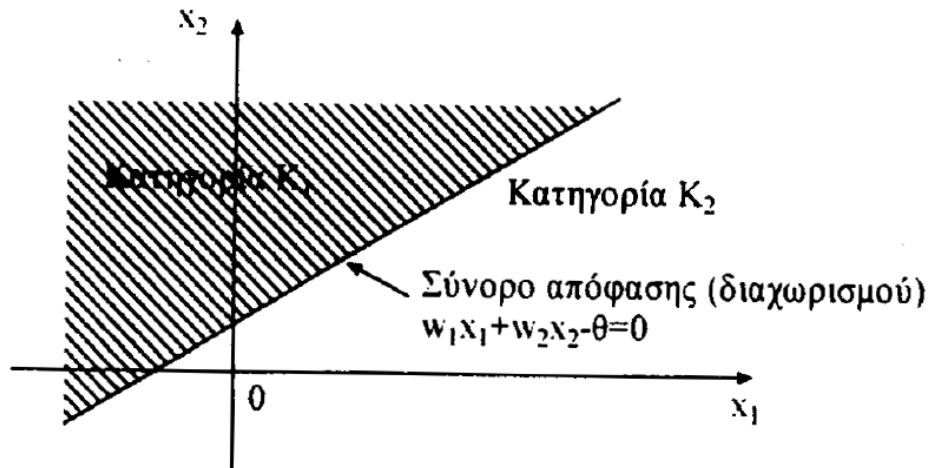
$$u = \sum_{i=1}^n w_i x_i - b$$

Η συνάρτηση μεταφοράς πραγματοποιεί την απεικόνιση του διανύσματος εισόδου  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$  (αποκαλούμενο στο εξής και με το όνομα πρότυπο) στην έξοδο  $y$ . Οι παράμετροι  $w_1, w_2, \dots, w_n$  ονομάζονται, όπως έχουμε αναφέρει, συναπτικά βάρη και το  $b$  ονομάζεται κατώφλι ενεργοποίησης και οφείλει το όνομά του στο ότι, ο νευρώνας ενεργοποιείται αν το άθροισμα  $\sum_{i=1}^n w_i x_i$  είναι μεγαλύτερο του  $b$ .

Η συνάρτηση ενεργοποίησης  $\varphi(\cdot)$ , που δέχεται ως μεταβλητή τη διέγερση  $u$  και καθορίζει την έξοδό μας, στο δίκτυο Perceptron μπορεί να είναι είτε η βηματική, είτε η συνάρτηση προσήμου. Έστω ότι χρησιμοποιούμε τη βηματική συνάρτηση. Τότε η εξίσωση  $u = \sum_{i=1}^n w_i x_i - b$  αντιστοιχεί σε ένα υπερεπίπεδο στο χώρο των  $n$  διαστάσεων. Τα σημεία  $\mathbf{x}$  που αντιστοιχούν σε θετικές τιμές,  $u > 0$ , βρίσκονται από τη μία πλευρά του επιπέδου, ενώ τα σημεία που αντιστοιχούν σε αρνητικές τιμές,  $u < 0$ , βρίσκονται στην απέναντι πλευρά του υπερεπίπεδου. Τα σημεία που αντιστοιχούν σε  $u = 0$  βρίσκονται πάνω στο υπερεπίπεδο.

Μια καλύτερη εξήγηση των παραπάνω προκύπτει από την εφαρμογή τους στις δύο διαστάσεις. Στο επίπεδο λοιπόν, η εξίσωση  $u = w_1 x_1 + w_2 x_2 - b = 0$ , ορίζει μια ευθεία που είναι κάθετη στο διάνυσμα των συναπτικών βαρών. Αυτή η ευθεία χωρίζει λοιπόν το επίπεδο σε δύο ημιεπίπεδα, όπου, στο ένα ανήκουν τα  $\mathbf{x}$  για τα οποία  $y = 1$  και στο άλλο ανήκουν τα  $\mathbf{x}$  για τα οποία  $y = 0$ .

Έστω ότι υπάρχει μία τέτοια ευθεία η οποία να μπορεί να χωρίσει το επίπεδο (ή υπερεπίπεδο) σε δύο ημιεπίπεδα, έτσι ώστε όλα τα πρότυπα εισόδου που ανήκουν στο ένα ημιεπίπεδο να ανήκουν στην ίδια κατηγορία (έστω ότι 16 αντιπροσωπεύουν άνδρες) και όλα τα υπόλοιπα να ανήκουν επίσης σε μία, μοναδική, διαφορετική κατηγορία (έστω γυναίκες). Τότε λέμε ότι οι κατηγορίες (κλάσεις) είναι γραμμικά διαχωρίσιμες. Όταν ισχύουν τα παραπάνω τότε υπάρχει και ένα δίκτυο Perceptron το οποίο αντιστοιχεί στην κατάλληλη ευθεία που επιτελεί τον παραπάνω διαχωρισμό[1].



Σχήμα 1.2 : Παράδειγμα διαχωρίσιμων κλάσεων στο επίπεδο

#### 1.4 Μάθηση στο δίκτυο Perceptron

Κύριο μέλημά μας σε ένα νευρωνικό δίκτυο είναι η εύρεση ενός τρόπου, ούτως ώστε να μπορεί, με κάποιο μηχανισμό, το ίδιο να ανανεώνει τις παραμέτρους του, για να επιτύχει την επιθυμητή απόκριση. Το δίκτυο Perceptron εκπαιδεύεται με επιβλεπόμενη μάθηση. Δηλαδή είναι πάντα γνωστή η επιθυμητή τιμή εξόδου  $d_p$  για κάθε πρότυπο εκπαίδευσης  $p$ . Το δίκτυο τώρα, για κάθε πρότυπο  $p$  ανανεώνει τα βάρη του με βάση κάποιον κανόνα, έτσι ώστε να πλησιάζει περισσότερο στην επιθυμητή έξοδο[2].

Ένας τέτοιος επαναληπτικός κανόνας εκπαίδευσης είναι ο κανόνας σταθερής αύξησης. Τα πρότυπα εισάγονται στο δίκτυο ένα ένα και όταν εισαχθεί και το τελευταίο, η διαδικασία ξεκινάει από την αρχή. Κάθε ένας ολοκληρωμένος κύκλος εισαγωγής όλων των προτύπων ονομάζεται εποχή.

Ο κανόνας ανανεώνει το διάνυσμα των συναπτικών βαρών μόνο όταν υπάρχει διαφορά μεταξύ του προτύπου που εισήχθη και της καθορισμένης επιθυμητής τιμής. Όταν παρατηρείται λοιπόν μια τέτοια διαφορά, η ανανέωση των βαρών γίνεται προσθέτοντας ή αφαιρώντας ένα ποσοστό του προτύπου εισόδου. Συγκεκριμένα, αν κατά την επανάληψη  $k$ , εισάγεται το πρότυπο  $p$  τότε ο κανόνας ανανέωσης είναι :

$$w(k) = w(k-1) + \eta(d_p - y)x_p$$

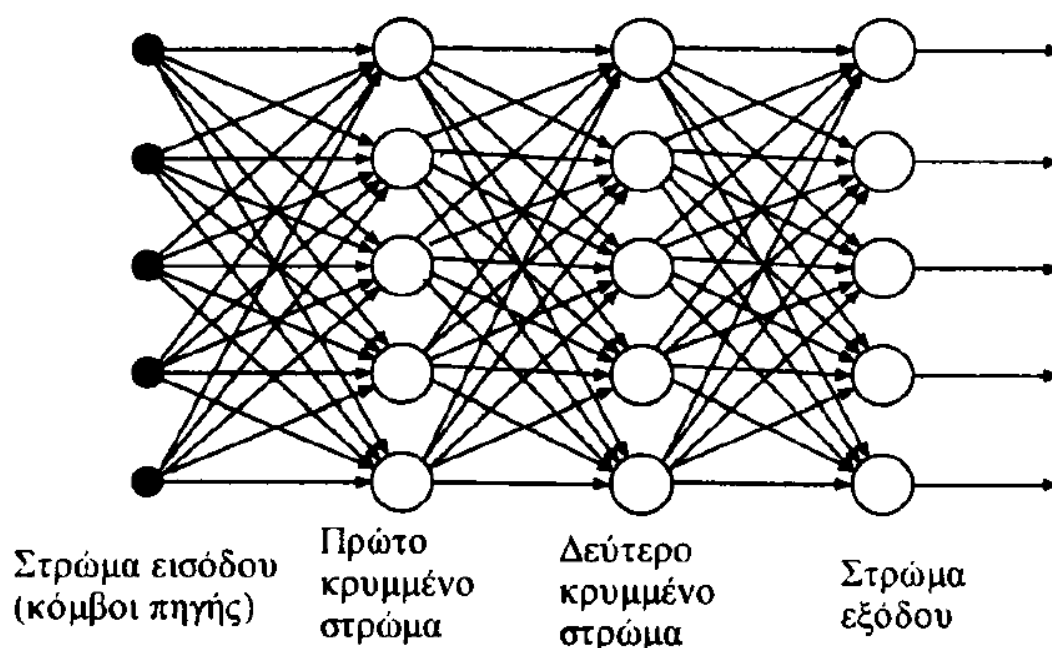
όπου  $w(k)$  είναι το διάνυσμα των συναπτικών βαρών μετά την επανάληψη  $k$ . Η παράμετρος  $\eta$  καθορίζει το μέγεθος της διόρθωσης και καλείται ρυθμός μάθησης ή βήμα μάθησης και είναι ένας μικρός θετικός αριθμός.

Αποδεικνύεται ότι η εκπαίδευση του  $w$  γίνεται με τέτοιο τρόπο ώστε να εξασφαλίζεται πάντα, ότι ένα πρότυπο που στην τρέχουσα εποχή ταξινομήθηκε λάθος, στην επόμενη εποχή, είτε θα ταξινομηθεί σωστά είτε η τιμή του θα πλησιάζει περισσότερο στην επιθυμητή.

Αποδεικνύεται επίσης ότι το μοντέλο Perceptron συγκλίνει πάντα σε μία λύση, αρκεί αυτή να υπάρχει, δηλαδή το πρόβλημα που αντιμετωπίζουμε να είναι γραμμικά διαχωρίσιμο.

### 1.5 Το μοντέλο Multi Layer Perceptron (MLP)

Η ικανότητα ταξινόμησης προτύπων του perceptron περιορίζεται σημαντικά από το γεγονός ότι οι κατηγορίες των προτύπων πρέπει να είναι γραμμικά διαχωρίσιμες. Ωστόσο, υπάρχουν προβλήματα, ακόμα και πολύ απλά, τα οποία δεν ικανοποιούν αυτό τον περιορισμό (χαρακτηριστικότερο παράδειγμα αυτής της κατηγορίας είναι το πρόβλημα XOR). Ο περιορισμός αυτός αίρεται με την προσθήκη κρυμμένων στρωμάτων.



Σχήμα 1.3 : Αρχιτεκτονική ενός MLP με δύο κρυμμένα στρώματα

Δίκτυα της μορφής αυτής, όπως του σχήματος 1.3, καλούνται πολυστρωματικά δίκτυα Perceptron (multi-layer). Το χαρακτηριστικό των δικτύων αυτών είναι ότι οι νευρώνες οποιουδήποτε στρώματος (στη γενική περίπτωση, ένα MLP έχει  $n$  στρώματα) τροφοδοτούνται αποκλειστικά από τους νευρώνες του προηγούμενου επιπέδου και τροφοδοτούν αποκλειστικά τους νευρώνες του επόμενου επιπέδου, γι' αυτό ονομάζονται και δίκτυα προσοτροφόδοτης (feed – forward) [3].

Στα δίκτυα Perceptron πολλών στρωμάτων, για την ενεργοποίηση (ή μη) των νευρώνων αποφεύγεται η χρήση της βηματικής συνάρτησης ή της συνάρτησης προσήμου, διότι δεν είναι διαφορίσιμες, ενώ οι περισσότεροι κανόνες εκπαίδευσης κάνουν χρήση παραγώγων για τη βελτιστοποίηση των παραμέτρων των MLP. Γι' αυτό χρησιμοποιείται η σιγμοειδής συνάρτηση ή κάποια άλλη, παραγωγίσιμη.

Τα πολυστρωματικά δίκτυα Perceptron που κάνουν χρήση της σιγμοειδούς συνάρτησης έχουν πολύ ισχυρές δυνατότητες για την αναπαράσταση συναρτήσεων. Η πιο σημαντική τους δυνατότητα, είναι ότι μπορούν να προσεγγίσουν οποιαδήποτε ομαλή συνάρτηση όσο κοντά επιθυμούμε [1]. Λόγω αυτής της ιδιότητας τα MLP καλούνται και universal approximators (καθολικοί προσεγγιστές).



## 1.6 Μάθηση των πολυστρωματικών δικτύων Perceptron

Η εκπαίδευση ενός MLP θέτει ως στόχο, τον κοινό στόχο που έχουν όλες οι διαδικασίες εκπαίδευσης των τεχνητών νευρωνικών δικτύων. Την κατάλληλη ρύθμιση των παραμέτρων του (συναπτικά βάρη), έτσι ώστε να ικανοποιείται κάποιο κριτήριο που έχουμε εμείς θέσει εκ των προτέρων ως επαρκές. Για την εκπαίδευση των MLP υπάρχουν διάφορες κατηγορίες εκπαίδευσης, οι οποίες θα παρουσιασθούν εδώ. Αναλυτικότερα θα παρουσιασθεί η μέθοδος κατάβασης δυναμικού (αλγόριθμος Back-Propagation), η οποία χρησιμοποιείται σε αυτήν την εργασία, ενώ θα αναφερθούν συνοπτικά και η ενισχυτική μάθηση, καθώς και η αυτοοργανούμενη μάθηση.

### 1.6.1 Επιβλεπόμενη μάθηση

#### α) Ο αλγόριθμος εκπαίδευσης Back-Propagation (ανάστροφης διάδοσης)

Γενικά οι αλγόριθμοι ανάστροφης διάδοσης έχουν το χαρακτηριστικό ότι υπολογίζουν ανάστροφα σήματα σφάλματος και μεταβάλλουν τα βάρη του κάθε νευρώνα ανάλογα με το πόσο συνέβαλλε στο σφάλμα. Για συνάρτηση σφάλματος συνήθως χρησιμοποιείται το μέσο τετραγωνικό σφάλμα του διανύσματος εξόδου του δικτύου από την επιθυμητή έξοδο. Αν θεωρήσουμε δεδομένα τα παραδείγματα εκπαίδευσης, τότε είναι σαφές ότι η συνάρτηση σφάλματος εξαρτάται αποκλειστικά από τα βάρη του δικτύου. Στόχος της εκπαιδευτικής διαδικασίας είναι η ελαχιστοποίηση αυτής της συνάρτησης ως προς τις παραμέτρους του δικτύου. Κάτι τέτοιο είναι δυνατό να επιτευχθεί με την κατάβαση δυναμικού (gradient descent).

Ο αλγόριθμος ανάστροφης διάδοσης λειτουργεί ως εξής:

Αρχικά παρουσιάζουμε κάποιο παράδειγμα στην είσοδο του δικτύου και από την έξοδό του υπολογίζουμε την τιμή της συνάρτησης σφάλματος. Στη συνέχεια υπολογίζουμε με φορά από την έξοδο προς την είσοδο (ανάστροφη) τα σήματα σφάλματος  $\delta$ . Τέλος με βάση τα σήματα αυτά καθώς και τις εισόδους των νευρώνων ανανεώνουμε τα βάρη. Κάθε σύνολο δεδομένων εκπαίδευσης χρειάζεται να περάσει πολλές φορές από ένα δίκτυο ώστε να συγκλίνουν τα βάρη σε τέτοιες τιμές που να εξασφαλίζουν ότι το νευρωνικό δίκτυο έμαθε τα παραδείγματα. Πολλοί αλγόριθμοι εκπαίδευσης αντί να αλλάζουν τα βάρη με κάθε παράδειγμα που τους παρουσιάζεται κάνουν συνολική αλλαγή των βαρών μετά το πέρας όλου του συνόλου εκπαίδευσης, δηλαδή στο τέλος κάθε εποχής.

Αν  $f(\cdot)$  είναι η συνάρτηση ενεργοποίησης ενός νευρώνα,  $y_i^k$  είναι η έξοδος (που είναι ίση με την είσοδο για το πρώτο επίπεδο  $k=1$ ),  $\delta_i^k$  είναι το σήμα σφάλματος του  $i$  νευρώνα του  $K$ ου επιπέδου. Το βάρος  $w_{ji}^k$  συνδέει τον  $j$  νευρώνα του  $K-1$  επιπέδου με τον  $i$  του  $K$ . Η έξοδος σε κάθε επίπεδο για κάθε νευρώνα είναι:

$$y_i^k = f\left(\sum_j y_j^{k-1} \cdot w_{ji}^k + b_i^k\right)$$

και η συνάρτηση σφάλματος είναι:

$$E_p^k = \frac{1}{2} \sum_j (e_j^k)^2$$

$$\text{με } e_j^k = d_j^k - y_j^k$$

όπου  $d_j^k$  είναι η επιθυμητή έξοδος για την έξοδο  $j$  του επιπέδου  $k$ .

Με βάσει τον κανόνα της κατάβασης δυναμικού, έχουμε τον παρακάτω τύπο που μας δίνει έναν τρόπο για την αλλαγή των βαρών ώστε να μειωθεί το σφάλμα του δικτύου:

$$\Delta w_{ij}^k = -\eta \frac{dE_p^k}{dw_{ij}^k}$$

όπου  $\eta$  είναι μια θετική παράμετρος μικρότερη της μονάδας και είναι ο γνωστός ρυθμός μάθησης.

Συμβολίζοντας τώρα με  $h_i^k$  το άθροισμα πριν περάσει από τη συνάρτηση ενεργοποίησης  $f$ , έχουμε:

$$h_i^k = \sum_j y_j^{k-1} \cdot w_{ji}^k + b_i^k$$

Η παράγωγος του σφάλματος ως προς τα βάρη είναι:

$$\frac{dE_p^k}{dw_{ij}^k} = \frac{dE_p^k}{de_j^k} \frac{de_j^k}{dy_j^k} \frac{dy_j^k}{dh_j^k} \frac{dh_j^k}{dw_{ij}^k} = e_j^k (-1) \frac{df(h_j^k)}{dh_j^k} y_i^k$$

ενώ

$$\delta_j^k = e_j^k \frac{df(h_j^k)}{dh_j^k}$$

Χωρίς να μπορούμε σε περαιτέρω λεπτομέρειες για τον υπολογισμό των μερικών παραγώγων παρατίθεται ο αλγόριθμος back-propagation :

1. Αρχικοποιούμε τα βάρη σε μικρές τυχαίες τιμές.
2. Παρουσιάζουμε στο νευρωνικό δίκτυο ένα διάνυσμα εκπαίδευσης με τις επιθυμητές εξόδους.
3. Υπολογίζουμε τις εξόδους κάθε νευρώνα του δικτύου από την είσοδο προς την έξοδο, σε ένα επίπεδο κάθε φορά.
4. Υπολογίζουμε τα σήματα σφάλματος  $\delta$  με βάση τις επιθυμητές εξόδους από το επίπεδο εξόδου προς την είσοδο.
5. Ανανεώνουμε τα βάρη στην κατεύθυνση της ελαχιστοποίησης της συναρτήσεως σφάλματος.
6. Συνεχίζουμε από το βήμα 2 παρουσιάζοντας στο δίκτυο ένα άλλο παράδειγμα από τα δεδομένα εκπαίδευσης μέχρι να παρουσιασθεί και το τελευταίο.
7. Επαναλαμβάνουμε την διαδικασία από το βήμα 2 παρουσιάζοντας ξανά όλα τα δεδομένα έως ότου επιτευχθεί η επιθυμητή απόδοση του δικτύου, δηλαδή έως ότου πέσει η τιμή του σφάλματος κάτω από μια τιμή ή μέχρι να ολοκληρωθεί ο μέγιστος αριθμός εποχών που έχουμε θέσει.

Πολύ σημαντική για τη λειτουργία του δικτύου είναι η κανονικοποίηση των δεδομένων που παρουσιάζονται σε αυτό ως προς το εύρος. Αν και η κανονικοποίηση δεν είναι απαραίτητη για την λειτουργία του αλγόριθμου εκπαίδευσης ωστόσο εξασφαλίζει καλύτερη συμπεριφορά και γρηγορότερη σύγκλιση.

Συχνά στον τύπο για την ανανέωση των βαρών προσθέτουμε μία παράμετρο ορμής η οποία μας εξασφαλίζει πιο γρήγορη κατάβαση. Ο όρος αυτός ορμής λαμβάνει υπόψη του την προηγούμενη μεταβολή του βάρους. Πιο συγκεκριμένα:

$$w(t+1) = w(t) + \eta \delta y + \alpha [w(t) - w(t-1)]$$

όπου  $w(t)$  είναι το βάρος τη χρονική στιγμή  $t$ , και  $\alpha$  ένας αριθμός όπου  $0 < \alpha < 1$  που ονομάζεται όρος ορμής. Τέλος οι όροι  $\delta$ ,  $y$  είναι αυτοί που υπολογίστηκαν παραπάνω για το κάθε βάρος.

Ο αλγόριθμος αυτός εκπαίδευσης δεν εγγυάται τη σύγκλιση όταν τα βάρη ανανεώνονται με το πέρασμα κάθε παραδείγματος του συνόλου εκπαίδευσης[1].

Αυτό συμβαίνει γιατί η κλίση που υπολογίζεται για την ανανέωση των βαρών υπολογίζεται μόνο για ένα παράδειγμα και όχι για όλο το σύνολο εκπαίδευσης. Για να εξασφαλισθεί η σύγκλιση ο αλγόριθμος πρέπει να βασίζεται σε όλο το σετ εκπαίδευσης. Ο βελτιωμένος αυτός αλγόριθμος ανανεώνει τα βάρη μια φορά ομαδικά στο τέλος κάθε εποχής (batch training). Για τη συνάρτηση σφάλματος χρησιμοποιούμε το συνολικό κριτήριο σφάλματος για όλα τα δεδομένα εκπαίδευσης:

$$\bar{E} = \frac{1}{N} \sum_{i=1}^N E_i$$

Ο αλγόριθμος αυτός αν και πιο βελτιωμένος ως προς τον στόχο που θέλουμε να πετύχουμε, ωστόσο έχει πολύ μεγαλύτερες απαιτήσεις σε μνήμη.

### 1.6.2 Ενισχυτική μάθηση

Κατά την ενισχυτική μάθηση, το τεχνητό νευρωνικό δίκτυο τροφοδοτείται με πρότυπα εισόδου αλλά δεν του δίνονται και οι επιθυμητές αποκρίσεις στο κάθε πρότυπο. Αντ' αυτού, όταν το δίκτυο παράγει μία έξοδο, του δίνεται ως απάντηση ένας δείκτης που αποτιμά την συμπεριφορά του.

Η βασική ιδέα στην οποία βασίζεται η ενισχυτική μάθηση είναι ότι αν μία ενέργεια του συστήματος ακολουθείται από κάποιας μορφής επιβράβευση, η τάση του συστήματος να παράγει αυτή την ενέργεια ενισχύεται. Αντίθετα, αν κάποια ενέργειά του ακολουθείται από κάποιας μορφής τιμωρία, η τάση του να παράγει την ενέργεια αυτή εξασθενεί.

Σύμφωνα με τα παραπάνω, η ενισχυτική μάθηση λειτουργεί ως εξής :

- ☐ Το δίκτυο δέχεται ως ερέθισμα ένα διάνυσμα εισόδου και παράγει ως αποτέλεσμα μία έξοδο, όπως αυτή προκύπτει με τις τρέχουσες τιμές των βαρών.
- ☐ Το σύστημα αξιολογεί την έξοδο και παράγει ως μέτρο της αξιολόγησης αυτής ένα ενισχυτικό σήμα, το οποίο τροφοδοτείται στο δίκτυο.

□ Με βάση το ενισχυτικό σήμα, το σύστημα τείνει να αυξάνει τα βάρη εκείνα που συμβάλλουν στην καλή συμπεριφορά και να μειώνει εκείνα που προκαλούν κακή συμπεριφορά.

□ Το δίκτυο αναζητά ένα σύνολο βαρών τα οποία τείνουν να αποφεύγουν να λαμβάνουν αρνητικά ενισχυτικά σήματα.

Η ενισχυτική μάθηση χωρίζεται σε συσχετιστική και μη-συσχετιστική. Στην πρώτη περίπτωση, το σύστημα, εκτός από το ενισχυτικό σήμα, τροφοδοτείται και με άλλες πληροφορίες από τις οποίες πρέπει να μάθει μια απεικόνιση με τη μορφή αιτίου-αποτελέσματος. Αντίθετα, στη δεύτερη περίπτωση, το δίκτυο τροφοδοτείται αποκλειστικά με το ενισχυτικό σήμα και σκοπός του δικτύου είναι να μάθει αποκλειστικά μία μοναδική βέλτιστη ενέργεια και όχι να συσχετίσει διάφορες ενέργειες με διαφορετικά ερεθίσματα[2].

### 1.6.3 Μη επιβλεπόμενη μάθηση (αυτο-οργανούμενη μάθηση)

Σε αυτό τον τύπο μάθησης δεν χρησιμοποιείται εξωτερικός δάσκαλος για να επιβλέψει την εκπαίδευση του δικτύου. Το δίκτυο, αντί να μάθει συγκεκριμένα ζευγάρια εισόδου – εξόδου, μαθαίνει ένα μέτρο της ποιότητας της παράστασης. Οι ελεύθεροι παράμετροι του δικτύου, προσαρμόζονται έτσι ώστε να μεγιστοποιηθεί το μέτρο αυτό.

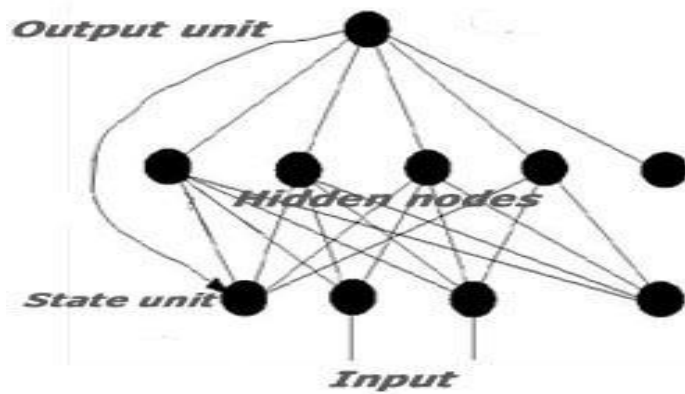
Η μη επιβλεπόμενη μάθηση μπορεί να συνδυαστεί με την επιβλεπόμενη μάθηση σε ένα πολυστρωματικό δίκτυο προσοτροφodότησης εκπαιδευόμενο με τον αλγόριθμο ανάστροφης διάδοσης για να επιταχύνει τη διαδικασία μάθησης. [4]

## 1.7 Τεχνητά αναδρομικά νευρωνικά δίκτυα

Τα τεχνητά αναδρομικά (ή ανατροφοδοτούμενα) νευρωνικά δίκτυα [3] αποτελούν την εξέλιξη των τεχνητών νευρωνικών δικτύων [2] τα οποία, με την σειρά τους, επιχειρούν να προσομοιώσουν την λειτουργία των φυσικών νευρωνικών δικτύων.

### 1.7.1 Γενικά για τα τεχνητά αναδρομικά νευρωνικά δίκτυα

Εάν το νευρωνικό δίκτυο περιέχει τουλάχιστον ένα βρόχο ανατροφοδότησης ο οποίος ανακυκλώνει πληροφορία μέσω του ίδιου ή προηγούμενων στρωμάτων, τότε ονομάζεται 'αναδρομικό νευρωνικό δίκτυο' ή νευρωνικό δίκτυο ανατροφοδότησης. Το αποτέλεσμα της ανατροφοδότησης είναι ότι, όταν ένα διάνυσμα (πρότυπο) εισόδου εισέρχεται στο αναδρομικό ΝΔ, δεν παράγει ένα πρότυπο εξόδου σε πεπερασμένο αριθμό χρονικών βημάτων, αλλά δρα με έναν κυκλικό τρόπο, όπου τα ίδια στρώματα ενεργοποιούνται επαναληπτικά. Εάν το ΝΔ είναι αφ' εαυτού ευσταθές πιθανά να ταλαντωθεί για κάποιο χρονικό διάστημα προτού φτάσει σε μια σταθερή κατάσταση στην οποία οι νευρωνικές ενεργοποιήσεις θα σταματήσουν να αλλάζουν με αποτέλεσμα να παραχθεί μια σταθερή έξοδος. Διαφορετικά, εάν το ΝΔ δε είναι ευσταθές, οι ταλαντώσεις θα συνεχίσουν αδιάκοπα. Συνεπώς όταν εκπαιδεύουμε ένα αναδρομικό ΝΔ είναι σημαντικό είναι σημαντικό να βρούμε το σύνολο των συναπτικών βαρών που του επιτρέπουν να σταθεροποιηθεί στις επιθυμητές τιμές εξόδου.



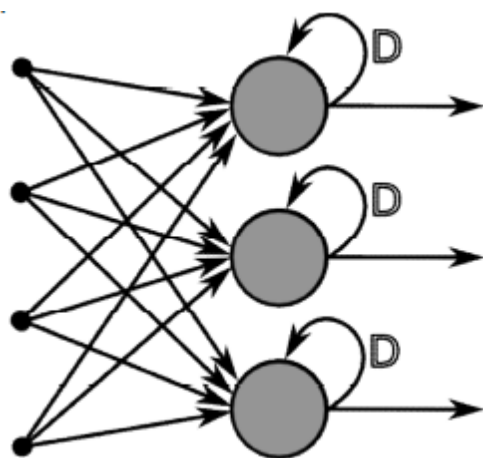
Σχήμα 1.4 : Αναδρομικό νευρωνικό δίκτυο με ένα βρόχο ανατροφοδότησης από τον νευρώνα εξόδου σε ένα νευρώνα της εισόδου.

Στην συνέχεια θα ασχοληθούμε με την αναλυτική λειτουργία του αναδρομικού νευρωνικού δικτύου Elman μιας και αυτό το δίκτυο χρησιμοποιούμε για της ανάγκες της παρούσης εργασίας.

### 1.7.2 Λειτουργία αναδρομικού νευρωνικού δικτύου Elman

Το δίκτυο Elman έχει την ακόλουθη αρχιτεκτονική (Σχ.1.5):

Χρησιμοποιείται ένα δίκτυο τριών επιπέδων στο οποίο έχει προστεθεί ένα σύνολο από μονάδες περιεχομένου (context units) στο επίπεδο εισόδου. Υπάρχουν συνδέσεις από το μεσαίο (κρυφό) επίπεδο σε αυτές τις μονάδες περιεχομένου οι οποίες έχουν σταθερό βάρος ίσο με 1. Σε κάθε βήμα, η είσοδος προωθείται προς το επόμενο επίπεδο (feed-forward) κι έπειτα εφαρμόζεται ένας κανόνας εκμάθησης. Οι σταθερές συνδέσεις προς τα πίσω καταλήγουν στις μονάδες περιεχομένου κρατώντας ένα αντίγραφο των προηγούμενων τιμών από τις κρυφές μονάδες. Έτσι, το δίκτυο μπορεί να διατηρήσει μία κατάσταση, έτσι ώστε να έχει την ικανότητα να φέρνει εις πέρας εργασίες όπως πρόβλεψη ακολουθίας που είναι πέρα από τις δυνατότητες ενός απλού πολυεπίπεδου νευρωνικού δικτύου. [5].



Σχήμα 1.5 : Αναδρομικό νευρωνικό δίκτυο Elman

Θα εξετάσουμε την λειτουργία ανατροφοδοτούμενου νευρωνικού δικτύου ενός επιπέδου αλλά η λογική επαληθεύεται και για δίκτυα περισσότερων κρυφών επιπέδων. Θεωρούμε:

- $x_i(t)$  την  $i$ -οστή είσοδο του δικτύου
- $y_j(t)$  την έξοδο του  $j$ -κρυφού νευρώνα
- $y_h(t-1)$  την προηγούμενη χρονικά έξοδο του  $h$ -κρυφού νευρώνα
- $y_k(t)$  την  $k$ -οστή έξοδο του δικτύου

Κατά την εκπαίδευσή του, το δίκτυο κατασκευάζει τους πίνακες βαρών  $V$ ,  $U$  και  $W$  τα στοιχεία των οποίων είναι αντίστοιχα:

- $v_{ji}$  το βάρος ανάμεσα στην  $i$ -είσοδο και τον  $j$ -κρυφό νευρώνα
- $u_{jh}$  το βάρος ανάμεσα στην προηγούμενη χρονικά έξοδο του  $h$ -κρυφού νευρώνα και την παρούσα είσοδο του  $j$ -κρυφού νευρώνα
- $w_{kj}$  το βάρος ανάμεσα στην έξοδο του  $j$ -κρυφού νευρώνα και την  $k$ -οστή έξοδο του δικτύου.

Για λόγους προφύλαξης του δικτύου κάθε νευρώνας των δύο τελευταίων επιπέδων, λαμβάνει και μια σταθερή είσοδο που ονομάζεται bias. Στο δίκτυο που μελετάμε έχουμε:

- $\theta_j$  το bias του  $j$ -κρυφού νευρώνα
- $\theta_k$  το bias της  $k$ -εξόδου.

Για να δημιουργηθεί η συνολική είσοδος κάθε νευρώνα αθροίζονται τα γινόμενα της εκάστοτε εισόδου με το αντίστοιχο βάρος και προστίθεται και το bias. Το αποτέλεσμα αυτό μπαίνει ως όρισμα στην συνάρτηση ενεργοποίησης του επιπέδου. Οι συναρτήσεις ενεργοποίησης των δύο τελευταίων επιπέδων μπορεί να είναι ίδιες ή διαφορετικές. Ορίζουμε :

- $f$  την συνάρτηση ενεργοποίησης του κρυφού επιπέδου
- $g$  την συνάρτηση ενεργοποίησης της εξόδου.

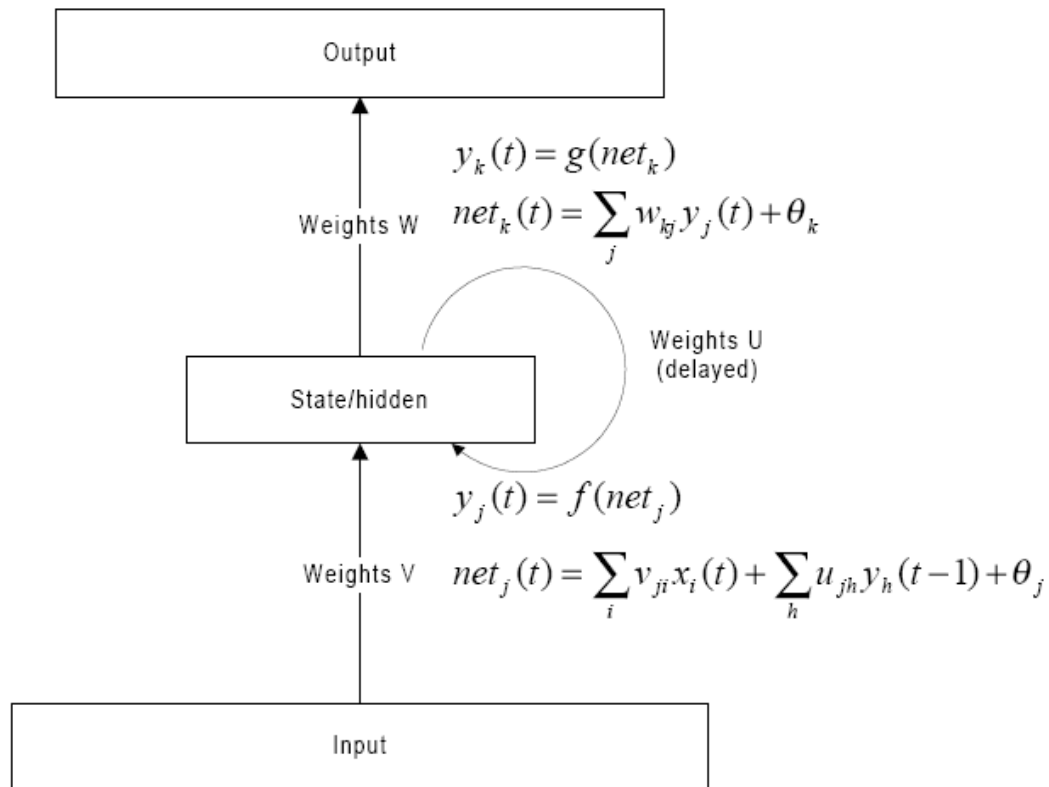
Οι εξισώσεις που δίνουν την έξοδο του κρυφού επιπέδου και την έξοδο του δικτύου είναι αντίστοιχα:

$$y_j(t) = f\left(\sum_i v_{ji} x_i(t) + \sum_h u_{jh} y_h(t-1) + \theta_j\right)$$

$$y_k(t) = g\left(\sum_j w_{kj} y_j(t) + \theta_k\right)$$

Συγκεντρωτικά, όλα τα παραπάνω παρουσιάζονται στο Σχήμα 2.3.

Όπως αντιλαμβανόμαστε η εκπαίδευση ενός τέτοιου δικτύου μπορεί να πραγματοποιηθεί με μια από τις μεθόδους που αναφέραμε και εξηγήσαμε στη παράγραφο 1.6 για τα MLP, μόνο που τώρα πρέπει φυσικά να λάβουμε υπ' όψιν μας και τα βάρη των ανατροφοδοτήσεων, τα οποία και ανανεώνονται ανάλογα με τον κανόνα μάθησης που έχει επιλεγεί. [5].



Σχήμα 1.6 : Λειτουργία αναδρομικού νευρωνικού δικτύου Elman .

## Κεφάλαιο 2 : Τεχνικές καταγραφής ανθρώπινης έκφρασης

### 2.1 Εισαγωγή

Στο κεφάλαιο αυτό θα δούμε πως με τη χρήση του προτύπου MPEG-4 μπορούμε να καταγράψουμε διαστάσεις μεταξύ των χαρακτηριστικών του ανθρώπινου προσώπου, επομένως να μελετήσουμε μικρές ή μεγάλες μεταβολές κίνησης των χαρακτηριστικών και καταληκτικά να κατηγοριοποιήσουμε ανθρώπινα συναισθήματα βάσει αυτών των μεταβολών.

### 2.2 Το πρότυπο MPEG - 4

*Το πρότυπο MPEG-4 στοχεύει και στο συνδυασμό πραγματικού και συνθετικού οπτικοακουστικού υλικού, έτσι ώστε να γίνει δυνατή η μετάδοση και απόδοσή του από συστήματα μέσω δικτύων μικρού εύρους ζώνης, αλλά και η διαλογική συμμετοχή του χρήστη – θεατή στη διαμόρφωση του περιεχομένου. Για αυτό το σκοπό, η αναπαράσταση και επεξεργασία συνθετικών αντικειμένων αποτελεί ένα κυρίαρχο τμήμα του προτύπου, με έμφαση στην αφαιρετική αναπαράσταση αντικειμένων όπως το ανθρώπινο πρόσωπο και σώμα, αλλά και χαρακτηριστικών κινήσεων ή ιδιοτήτων τους, όπως η υφή ή η έκφραση.*

Μια από τις βασικές διαφορές του προτύπου MPEG-4 σε σχέση με τα προηγούμενα αφορά τη θεώρηση της βασικής μονάδας της πληροφορίας που επεξεργάζεται. Στο πρότυπο MPEG-1, τόσο η κινούμενη εικόνα, όσο και το ηχητικό σήμα μπορούν να αναπαραχθούν με την ίδια λογική με την οποία αναπαράγεται μια αναλογική εικόνα VHS σε μια κοινή συσκευή VCR : αυτό πρακτικά σημαίνει ότι δεν υπάρχει η δυνατότητα τυχαίας προσπέλασης στο υλικό, παρά μόνο γρήγορης μετακίνησης προς τα εμπρός ή προς τα πίσω. Στο πρότυπο MPEG-2, η συμπίεση και κωδικοποίηση της κινούμενης εικόνας κάνει δυνατή την αποθήκευση ολόκληρων κινηματογραφικών ταινιών σε ψηφιακούς δίσκους. Παρόλα αυτά, το θέμα της αφαιρετικής αναπαράστασης του υλικού και της εξαγωγής συγκεκριμένων χαρακτηριστικών ή οντοτήτων από αυτό δε θίγεται καθόλου.

Με την πρόοδο των δικτύων μεταφοράς και των οπτικοακουστικών μέσων, τα ζητήματα της ανεξαρτησίας από το φυσικό μέσο, της εύκολης και πιο ορθολογικής μετάδοσης, αλλά και της αυξημένης διαλογικότητας φαίνονται να γίνονται περισσότερο σημαντικά. Έτσι, το πρότυπο θα πρέπει να ανταποκρίνεται αποτελεσματικά σε τομείς όπως:

- η αποτελεσματική αναπαράσταση βίντεο σε ποιότητα που μπορεί να κυμαίνεται από πολύ υψηλή μέχρι σχετικά χαμηλή, δεδομένων μουσικής ή ομιλίας, συνθετικών τρισδιάστατων αντικειμένων γενικής ή εξειδικευμένης μορφής, συνθετικής ομιλίας και κειμένου
- η αντιμετώπιση σφαλμάτων κατά την κωδικοποίηση, ειδικά σε συνθήκες δικτύων χωρίς μεγάλο εύρος ζώνης
- η αναπαράσταση των αντικειμένων με ανεξάρτητο τρόπο, έτσι ώστε να είναι δυνατή η διακριτή επεξεργασία και επαναχρησιμοποίησή τους



- η σύνθεση αντικειμένων ήχου, πραγματικής και συνθετικής εικόνας σε μια συνολική ολοκληρωμένη σκηνή
- η υποστήριξη περιγραφής και διαλογικότητας στα αντικείμενα, καθώς και ενσωμάτωση συνδέσμων περιήγησης στη σκηνή
- η διαχείριση και προστασία των πνευματικών δικαιωμάτων στη σκηνή και τα περιεχόμενά της, έτσι ώστε να έχουν πρόσβαση μόνο οι διαπιστευμένοι χρήστες και να διαπιστώνεται με ασφάλεια η αντιγραφή
- ο καθορισμός μιας ενιαίας μορφοποίησης που να είναι διαφανής ως προς το μέσο, το περιεχόμενο και την υπολογιστική πλατφόρμα [10].

Η βάση για να καταστεί δυνατή η ικανοποίηση όλων αυτών των αναγκών ήταν η αναπαράσταση της οπτικοακουστικής πληροφορίας με αντικείμενα, τα οποία συνδέονται μεταξύ τους με σχέσεις στο χώρο και το χρόνο. Ένα σημαντικό πλεονέκτημα της αντικειμενοστρεφούς αναπαράστασης είναι ότι διαφορετικοί τύποι αντικειμένων μπορεί να αντιστοιχούν σε διαφορετικές βέλτιστες μεθόδους απόδοσης, χωρίς όμως να εμποδίζεται η αρμονική συνύπαρξή τους σε μια ενιαία σκηνή. Για παράδειγμα, ένας συνθετικός ανθρωπόμορφος χαρακτήρας, του οποίου οι παραμορφώσεις περιγράφονται αφαιρετικά από λεκτικές παραμέτρους μπορεί να συντεθεί με ένα αντικείμενο ψηφιακού βίντεο, το οποίο αναπαριστάται από μια ορθογώνια δομή από pixels και μια μάσκα δύο διαστάσεων για την απόκρυψη ορισμένων περιοχών του. Σε μια τέτοια οργάνωση, η διαλογική επεξεργασία και πλοήγηση μέσα στο περιεχόμενο γίνονται με σαφώς απλούστερο τρόπο και διευκολύνονται εργασίες όπως η επαναχρησιμοποίηση του υλικού, η μεταφορά σε εναλλακτικά σχήματα κωδικοποίησης και η προοδευτική μετάδοση μέσα από δίκτυα χαμηλού εύρους ζώνης, όπως το Internet ή τα ασύρματα δίκτυα. Επειδή οι ιδέες που ενσωματώνονται στο MPEG-4 μπορούν να ωφελήσουν μια σειρά από διαφορετικές εφαρμογές και περιβάλλοντα, το πρότυπο είναι σχεδιασμένο με τη νοοτροπία της "εργαλειοθήκης", παρά της αυστηρής περιγραφής και προτυποποίησης κάθε ιδέας. Αυτό σημαίνει ότι, ανάλογα με την εφαρμογή και το περιβάλλον στο οποίο θα εφαρμοστεί, μπορούμε να καθορίσουμε μια σειρά από profiles, με κατάλληλες συμβάσεις και ρυθμίσεις για το υλικό, τη μετάδοσή του και την απόδοση – εμφάνιση στον τελικό χρήστη. Το περιεχόμενο αυτών των ρυθμίσεων αφορά περισσότερο τον ίδιο το σχεδιαστή της εφαρμογής, αφού συνήθως είναι αρκετό να εντοπίσει και να χρησιμοποιήσει ένα μικρό υποσύνολο όλων των ιδεών που περιγράφονται στο πρότυπο. Σε αντίθεση με το πρότυπο MPEG-2, το MPEG-4 μπορεί να ωφελήσει μια σειρά από διαφοροποιημένες χρήσεις, τόσο στον τομέα του ίδιου του υλικού και της κατασκευής του, όσο και αναφορικά με τη μετάδοση και την αναπαράστασή του. Αν και συνήθως ένα πρότυπο θέτει μια σειρά από περιορισμούς με σκοπό να επιβάλλει την τελική συμβατότητα, το MPEG-4 φαίνεται να ακολουθεί μια φιλοσοφία ορισμού του ελάχιστου δυνατού για να επιβάλλει τη μέγιστη χρηστικότητα. Αυτό πρακτικά σημαίνει ότι η υλοποίηση πολλών από τις ιδέες που ενσωματώνονται στο πρότυπο αφήνεται στην ευχέρεια του σχεδιαστή της εφαρμογής ή του χειριστή του υλικού. Ένα ενδιαφέρον σημείο αυτής της φιλοσοφίας είναι ότι εναλλακτικοί αλγόριθμοι μπορεί να αναπτυχθούν παράλληλα με το πρότυπο και να ξεχωρίσουν μέσα από διαδικασίες υποστήριξης και ανταγωνισμού.

Στην πρώτη του έκδοση, το πρότυπο MPEG-4 περιλαμβάνει έξι συγκεκριμένα τμήματα:

- **Systems:** περιλαμβάνει την περιγραφή της σκηνής, την πολυπλεξία, το συγχρονισμό, τη διαχείριση και την προστασία της πνευματικής ιδιοκτησίας
- **Visual:** περιγράφει την κωδικοποιημένη αναπαράσταση της πραγματικής και συνθετικής οπτικής πληροφορίας
- **Audio:** περιγράφει την κωδικοποιημένη αναπαράσταση της πραγματικής και συνθετικής ακουστικής πληροφορίας
- **Conformance testing:** καθορίζει τις συνθήκες συμβατότητας για τις συσκευές και το κωδικοποιημένο περιεχόμενο
- **Reference software:** περιλαμβάνει λογισμικό και εργαλεία για τα περισσότερα από τα τμήματα του MPEG-4 και το οποίο μπορεί να χρησιμοποιηθεί για την ανάπτυξη συμβατών εργαλείων
- **Delivery Multimedia Integration Framework (DMIF):** περιγράφει πρωτόκολλα για τη διαχείριση της προοδευτικής μετάδοσης (streaming) της πληροφορίας πολλαπλών μέσων μέσα από τεχνολογίες γενικής μορφής.

Τα τμήματα 1 ως 3, καθώς και το τμήμα 6, αποτελούν τον πυρήνα (core) της τεχνολογίας του MPEG-4, ενώ τα τμήματα 4 και 5 είναι συμπληρωματικά. Ακόμα, τα τρία πρώτα τμήματα είναι ανεξάρτητα από την εμφάνιση της πληροφορίας, αφήνοντας στο τμήμα 6 να διαχειριστεί τις ιδιαιτερότητες του στρώματος αυτού. Γενικά, αν και καθένα από τα τμήματα του προτύπου μπορεί να χρησιμοποιηθεί ξεχωριστά από τα υπόλοιπα, ακόμα και σε συνδυασμό με διαφορετικές τεχνολογίες, το πνεύμα τους είναι τέτοιο που να αποδίδουν καλύτερα όταν υλοποιούνται συνδυαστικά [6].

## 2.3 Τεχνικές καταγραφής παραμόρφωσης χαρακτηριστικών ανθρώπινου προσώπου

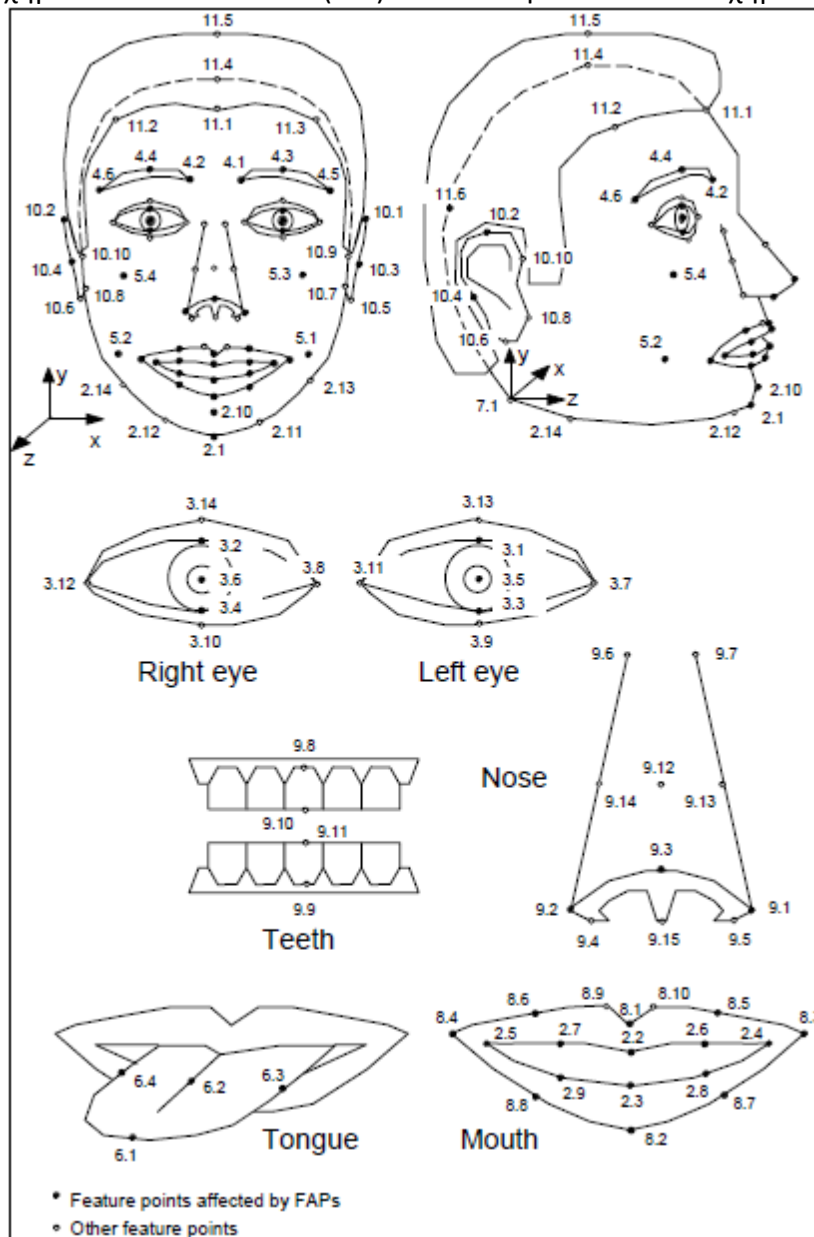
### 2.3.1. MPEG-4 και Πρόσωπο - FDPs και FAPs

Χρησιμοποιούμε ένα σύνολο παραμέτρων για τον προσδιορισμό του σχήματος, του μεγέθους και της υφής του προσώπου (FDPs – Facial Definition Parameters), σε συνδυασμό με ένα άλλο σύνολο παραμέτρων που χρησιμοποιείται για τον προσδιορισμό των «κινήσεων» του προσώπου είτε αυτές προέρχονται από έκφραση συναισθημάτων είτε προκύπτουν κατά την ομιλία (FAPs – Facial Animation Parameters) . Το σύνολο των FDPs μας παρέχει τη δυνατότητα ακριβούς προσδιορισμού των κινήσεων συγκεκριμένων χαρακτηριστικών του προσώπου, ενώ από τα FAPs, με σωστή ερμηνεία, μπορούμε να προσδιορίσουμε εκφράσεις και εκφορά λόγου σε διάφορα μοντέλα προσώπων χωρίς να είναι απαραίτητες η αρχικοποίηση και η βαθμονόμηση (calibration) της κάμερας. Η μετατροπή των FAPs σε φωνήματα μπορεί να μας οδηγήσει σε ένα μοντέλο κεφαλιού που θα έχει τη δυνατότητα να μιλάει σε όλες τις γλώσσες.

### 2.3.2 Σύνολο Παραμέτρων Προσδιορισμού Προσώπου (FDPs)

Με τα FDPs μπορούμε να προσαρμόσουμε ένα μοντέλο προσώπου σε ένα δεδομένο πρόσωπο. Το σύνολο των FDPs μπορεί να περιλαμβάνει πλέγμα 3D με πληροφορίες για την υφή, τρισδιάστατα χαρακτηριστικά σημεία και, προαιρετικά, κάποια εικόνα της υφής και άλλα χαρακτηριστικά όπως μαλλιά, γυαλιά οράσεως, ηλικία ή φύλο. Το πλέγμα 3D χρησιμοποιείται για τον προσδιορισμό του σχήματος του προσώπου, ενώ τα τρισδιάστατα χαρακτηριστικά σημεία χρησιμοποιούνται για τον

προσδιορισμό των χαρακτηριστικών του προσώπου στο συγκεκριμένο τρισδιάστατο σχήμα. Τα Feature Points (FPs) των FDPs φαίνονται στο Σχήμα 2.1.



Σχήμα 2.1 : Τα feature points (FPs)

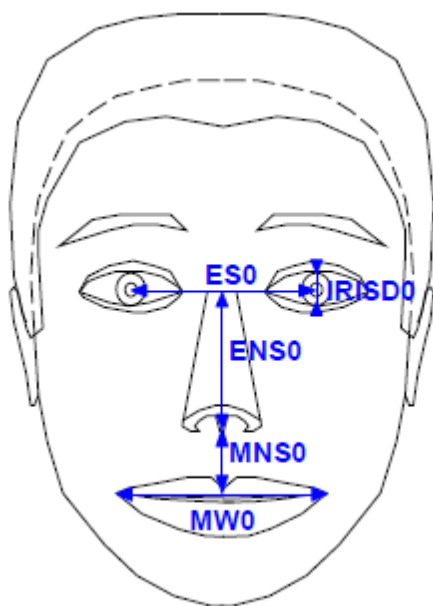
Τα FDPs περιέχουν τα εξής πεδία:

- **FeaturePointsCoord** – τα τρισδιάστατα χαρακτηριστικά σημεία που χρησιμοποιούνται για τη ζυγοστάθμιση του μοντέλου του προσώπου.
- **TextureCoords** – συντεταγμένες της υφής για τα χαρακτηριστικά σημεία
- **TextureType** – πληροφορία για τον αποκωδικοποιητή σχετική με τον τύπο της εικόνας της υφής.
- **FaceDefTables** – περιγραφή της συμπεριφορά των FAPs.
- **FaceSceneGraph** - περιέχει την εικόνα της υφής ή πληροφορίες για την ιεραρχία του μοντέλου [7].

### 2.3.3 Σύνολο Παραμέτρων Κίνησης Προσώπου (FAPs)

Τα FAPs βασίζονται στη μελέτη των ελαχίστων κινήσεων του προσώπου και συνδέονται στενά με τις κινήσεις των μυών. Αντιπροσωπεύουν ένα ολοκληρωμένο σύνολο βασικών ενεργειών του προσώπου, επιτρέποντας την απεικόνιση της πλειοψηφίας των φυσιολογικών ανθρωπίνων εκφράσεων, ενώ οι υπερβολικές τιμές μας επιτρέπουν να ορίσουμε ενέργειες αδύνατες για έναν άνθρωπο, αλλά απαραίτητες, για παράδειγμα, για τους χαρακτήρες cartoon. Όλες οι παράμετροι που εμπεριέχουν μεταφορική κίνηση εκφράζονται με τους όρους των μονάδων κίνησης των χαρακτηριστικών του προσώπου (FAPU – Facial Animation Parameter Units) [9].

Προέκυψε η ανάγκη ορισμού των μονάδων αυτών ώστε να είναι δυνατή η εφαρμογή των FAPs σε οποιοδήποτε μοντέλο προσώπου με σταθερό τρόπο, καθώς και η παραγωγή λογικών και αναμενόμενων αποτελεσμάτων, όσον αφορά στις εκφράσεις και στην εκφορά του λόγου. Τα FAPUs απεικονίζονται στην Εικόνα 2.2 και αντιστοιχούν, στην ουσία, σε κλάσματα της απόστασης ορισμένων σημείων – κλειδιών του προσώπου, επιλεγμένα με τέτοιο τρόπο ώστε να δίνουν την απαιτούμενη ακρίβεια.



Σχήμα 2.2 : Οι παράμετροι FAPUs

Οι παράμετροι FAPUs που απεικονίζονται στο Σχήμα 2.2, δίνουν τις σχέσεις του Πίνακα 2.1 για τις μονάδες μέτρησης του MPEG-4:

Ονομασία	Περιγραφή	Σχέσεις ορισμού μονάδων μέτρησης MPEG-4
IRISD0	Διάμετρος της ίριδας (εξ ορισμού ισούται με την απόσταση άνω και κάτω βλεφάρου) στο ουδέτερο πρόσωπο	$IRISD = \frac{IRISD0}{1024}$
ES0	Απόσταση ματιών	$ES = \frac{ES0}{1024}$
ENS0	Απόσταση μύτης-ματιών	$ENS = \frac{ENS0}{1024}$
MNS0	Απόσταση στόματος-μύτης	$MNS = \frac{MNS0}{1024}$
MW0	Πλάτος στόματος	$MW = \frac{MW0}{1024}$
AU	Γωνιακή μονάδα	$AU = 10E - 5rad$

*Πίνακας 2.1 : Σχέσεις ορισμού μονάδων μέτρησης στο MPEG-4 για κάθε παράμετρο FAPUs*

Για κάθε FAP ορίζουμε αν έχει μόνο θετικές (unidirectional) ή και αρνητικές τιμές (bidirectional), προς ποια κατεύθυνση είναι οι θετικές τιμές και ποιες ακριβώς είναι οι μονάδες μέτρησης (μοίρες ή κάποιο από τα FAPUs), ενώ μόνο για την περίπτωση του σαγονιού χρησιμοποιούμε μονάδες που μετρούν την «ένταση» πίεσης του σαγονιού με κλίμακα από το 1 έως το 10. Υπάρχουν FAPs που προσδιορίζουν την κίνηση του βολβού του ματιού, του βλεφάρου, το μέγεθος της κόρης του ματιού, την κίνηση των φρυδιών (με τρία σημεία για κάθε φρύδι), την κίνηση της μύτης και των ρουθουνιών ειδικότερα, όσο μικρή κι αν είναι αυτή. Άλλα FAPs προσδιορίζουν τις κινήσεις από τα μάγουλα και την πίεση που πιθανώς ασκείται στο σαγόκι καθώς και την κίνηση του στόματος (χρησιμοποιώντας τρία σημεία και προσδιορίζοντας συγχρόνως και την πιθανή κατακόρυφη κίνηση του άνω και του κάτω χείλους ή το τράβηγμα των άκρων του στόματος, εσωτερικά αλλά και εξωτερικά). FAPs υπάρχουν ακόμη και για την κίνηση της άκρης της γλώσσας ή και για πιθανή κίνηση των αυτιών. Υπάρχουν κάποια «ανώτερα FAPs» που περιγράφουν, χωρίς να διευκρινίζουν λεπτομέρειες, κάποια από τις έξι γνωστές εκφράσεις για το πρόσωπο που θα αποκωδικοποιηθεί. Τα FAPs αυτά (έξι, ένα για κάθε βασική έκφραση) έχουν προτεραιότητα έναντι των υπολοίπων, δηλαδή ακόμα και αν, σύμφωνα με τα υπόλοιπα FAPs, ορίζεται κάτι διαφορετικό, ο αποκωδικοποιητής θα λάβει υπόψη του μόνο τα ανώτερα FAPs.

Επίσης, έχουν οριστεί και άλλες παράμετροι που προσδιορίζουν τη στάση του κεφαλιού και την ακριβή έκφραση του προσώπου.

Το MPEG-4 προσδιορίζει 84 σημεία χαρακτηριστικών στο ουδέτερο πρόσωπο, τα οποία παρέχουν χωρική αναφορά για τον ορισμό των FAPs. Ο ορισμός του Facial Animation στο πλαίσιο του προτύπου ISO MPEG-4 βασίζεται στο Σύστημα Κωδικοποίησης Ενεργειών του Προσώπου (FACS – Facial Action Coding System). Συγκεκριμένα, τα σύνολα παραμέτρων ορισμού προσώπου FDP (Facial Definition Parameter) και οι παράμετροι κίνησης προσώπου FAP (Facial Animation Parameter)

σχεδιάστηκαν στο πρότυπο MPEG-4 για να επιτρέψουν τον ορισμό του σχήματος και της υφής του προσώπου, μειώνοντας την ανάγκη προσδιορισμού της τοπολογίας της υπάρχουσας γεωμετρίας μέσω των FDPs και την κίνηση των χαρακτηριστικών του προσώπου, αναπαράγοντας εκφράσεις, συναισθήματα και εκφορά λόγου μέσω των FAPs.

Ο ορισμός του viseme έχει συμπεριληφθεί στο πρότυπο για τον συγχρονισμό των κινήσεων του στόματος ως προς τα φωνήματα με τη συνολική κίνηση των χαρακτηριστικών του προσώπου (facial animation). Παρακολουθώντας τα νεύματα (gestures) του προσώπου που αντιστοιχούν σε FDP και/ή FAP κινήσεις ως προς το χρόνο, είναι δυνατή η απόδοση στοιχείων (cues) που αφορούν τις εκφράσεις και τα συναισθήματα του χρήστη. Ποικίλα αποτελέσματα έχουν παρουσιαστεί που αφορούν στην κατάταξη βασικών εκφράσεων του προσώπου, βασιζόμενα κυρίως σε χαρακτηριστικά ή σημεία που εξάγονται από περιοχές του προσώπου γύρω από το στόμα ή τα μάτια. Αυτά τα αποτελέσματα υποδηλώνουν ότι οι εκφράσεις του προσώπου σε συνδυασμό με κάποιες χειρονομίες και την φωνή, όταν αυτή είναι διαθέσιμη, παρέχουν στοιχεία που μπορούν να χρησιμοποιηθούν για την κατανόηση της συναισθηματικής κατάστασης του ατόμου. Τα FAPs ομαδοποιούνται ανάλογα με την λογική ορισμού τους, όπως φαίνεται στον Πίνακα 2.2 [7].

Ομάδες FAP		Πλήθος FAP
1	Εκφράσεις	2
2	Σαγόνι, πηγούνι, εσωτερικό του κάτω χείλους, άκρες χειλιών, μέσο χειλιών	16
3	Βολβός οφθαλμού, κόρη οφθαλμού, βλέφαρο	12
4	Φρύδι	8
5	Μάγουλο	4
6	Γλώσσα	5
7	Περιστροφή κεφαλιού	3
8	Θέση εξωτερικού χείλος	10
9	Μύτη	4
10	Αυτιά	4

Πίνακας 2.2 : Ομάδες FAP

α/α	Όνομα FAP	Περιγραφή FAP	Μονάδα μέτρησης	Συνήθης / απλή / κατεύθυνση	Κατεύθυνση Κίνησης	Ομάδα
1	Viseme	Καθορίζει την ακολουθία των visemes (κίνηση στόματος για προφορά συγκεκριμένων φωνημάτων) για το συγκεκριμένο καρέ.	-	-	-	1
2	Expression	Ένα σύνολο τιμών που καθορίζουν την έκφραση του προσώπου.	-	-	-	1
3	open_jaw	Κατακόρυφη μετατόπιση σαγονιών (δεν επηρεάζεται από το άνοιγμα του στόματος)	MNS	M	Κάτω	2
4	lower_t_midlip	Κατακόρυφη μετατόπιση του μέσου του εσωτερικού του άνω χείλους	MNS	Δ	Κάτω	2
5	raise_b_midlip	Κατακόρυφη μετατόπιση του μέσου του εσωτερικού του κάτω χείλους	MNS	Δ	Πάνω	2
6	stretch_l_cornerlip	Οριζόντια μετατόπιση της αριστερής εσωτερικής γωνίας των χειλιών	MW	Δ	Αριστερά	2
7	stretch_r_cornerlip	Οριζόντια μετατόπιση της δεξιάς εσωτερικής γωνίας των χειλιών	MW	Δ	Δεξιά	2
8	lower_t_lip_lm	Κατακόρυφη μετατόπιση του μέσου μεταξύ της αριστερής γωνίας και του μέσου του εσωτερικού του άνω χείλους	MNS	Δ	Κάτω	2
9	lower_t_lip_rm	Κατακόρυφη μετατόπιση του μέσου μεταξύ της δεξιάς γωνίας και του μέσου του εσωτερικού του άνω χείλους	MNS	Δ	Κάτω	2
10	raise_b_lip_lm	Κατακόρυφη μετατόπιση του μέσου μεταξύ της αριστερής γωνίας και του μέσου του εσωτερικού του κάτω χείλους	MNS	Δ	Πάνω	2
11	raise_b_lip_rm	Κατακόρυφη μετατόπιση του μέσου μεταξύ της δεξιάς γωνίας και του μέσου του εσωτερικού του κάτω χείλους	MNS	Δ	Πάνω	2
12	raise_l_cornerlip	Κατακόρυφη μετατόπιση της εσωτερικής αριστερής γωνίας των χειλιών	MNS	Δ	Πάνω	2
13	raise_r_cornerlip	Κατακόρυφη μετατόπιση της εσωτερικής δεξιάς γωνίας των χειλιών	MNS	Δ	Πάνω	2
14	thrust_jaw	Προώθηση του σαγονιού	MNS	M	Μπροστά	2
15	shift_jaw	Οριζόντια μετατόπιση του σαγονιού	MW	Δ	Δεξιά	2
16	push_b_lip	Προώθηση της μέσης του κάτω χείλους	MNS	Δ	Μπροστά	2
17	push_t_lip	Προώθηση της μέσης του άνω χείλους	MNS	Δ	Μπροστά	2
18	depress_chin	Ανύψωση και συμπίεση του σαγονιού (όπως στην έκφραση της λύπης).	MNS	Δ	Πάνω	2
19	close_t_l_eyelid	Κατακόρυφη μετατόπιση του άνω αριστερού	IRISD	Δ	Κάτω	3



		βλέφαρου				
20	close_t_r_eyelid	Κατακρόρυφη μετατόπιση του άνω δεξιού βλέφαρου	IRISD	Δ	Κάτω	3
21	close_b_l_eyelid	Κατακρόρυφη μετατόπιση του κάτω αριστερού βλέφαρου	IRISD	Δ	Πάνω	3
22	close_b_r_eyelid	Κατακρόρυφη μετατόπιση του κάτω δεξιού βλέφαρου	IRISD	Δ	Πάνω	3
23	yaw_l_eyeball	Οριζόντιος προσανατολισμός του βολβού του αριστερού οφθαλμού	AU	Δ	Αριστερά	3
24	yaw_r_eyeball	Οριζόντιος προσανατολισμός του βολβού του δεξιού οφθαλμού	AU	Δ	Αριστερά	3
25	pitch_l_eyeball	Κατακρόρυφος προσανατολισμός του βολβού του αριστερού οφθαλμού	AU	Δ	Κάτω	3
26	pitch_r_eyeball	Κατακρόρυφος προσανατολισμός του βολβού του δεξιού οφθαλμού	AU	Δ	Κάτω	3
27	thrust_l_eyeball	Προώθηση του βολβού του αριστερού οφθαλμού	ES	Δ	Μπροστά	3
28	thrust_r_eyeball	Προώθηση του βολβού του δεξιού οφθαλμού	ES	Δ	Μπροστά	3
29	dilate_l_pupil	Διαστολή κόρης αριστερού οφθαλμού	IRISD	Δ	Αυξητικά	3
30	dilate_r_pupil	Διαστολή κόρης δεξιού οφθαλμού	IRISD	Δ	Αυξητικά	3
31	raise_l_i_eyebrow	Κατακρόρυφη μετατόπιση του αριστερού εσωτερικού φρυδιού	ENS	Δ	Πάνω	4
32	raise_r_i_eyebrow	Κατακρόρυφη μετατόπιση του δεξιού εσωτερικού φρυδιού	ENS	Δ	Πάνω	4
33	raise_l_m_eyebrow	Κατακρόρυφη μετατόπιση του αριστερού μέσου φρυδιού	ENS	Δ	Πάνω	4
34	raise_r_m_eyebrow	Κατακρόρυφη μετατόπιση του δεξιού μέσου φρυδιού	ENS	Δ	Πάνω	4
35	raise_l_o_eyebrow	Κατακρόρυφη μετατόπιση του αριστερού εξωτερικού φρυδιού	ENS	Δ	Πάνω	4
36	raise_r_o_eyebrow	Κατακρόρυφη μετατόπιση του δεξιού εξωτερικού φρυδιού	ENS	Δ	Πάνω	4
37	squeeze_l_eyebrow	Οριζόντια μετατόπιση του αριστερού φρυδιού	ES	Δ	Δεξιά	4
38	squeeze_r_eyebrow	Οριζόντια μετατόπιση του δεξιού φρυδιού	ES	Δ	Αριστερά	4
39	puff_l_cheek	Οριζόντια μετατόπιση του αριστερού μάγουλου	ES	Δ	Αριστερά	5
40	puff_r_cheek	Οριζόντια μετατόπιση του δεξιού μάγουλου	ES	Δ	Δεξιά	5
41	lift_l_cheek	Κατακρόρυφη μετατόπιση του αριστερού μάγουλου	ENS	M	Πάνω	5
42	lift_r_cheek	Κατακρόρυφη μετατόπιση του δεξιού μάγουλου	ENS	M	Πάνω	5
43	shift_tongue_tip	Οριζόντια μετατόπιση της άκρης της γλώσσας	MW	Δ	Δεξιά	6
44	raise_tongue_tip	Κατακρόρυφη μετατόπιση της άκρης της γλώσσας	MNS	Δ	Πάνω	6
45	thrust_tongue_tip	Προώθηση της άκρης της γλώσσας	MW	Δ	Μπροστά	6
46	raise_tongue	Κατακρόρυφη μετατόπιση της γλώσσας	MNS	Δ	Πάνω	6
47	tongue_roll	Δίπλωμα της γλώσσας σε σχήμα U	AU	M	Κοίλη προς τα πάνω	6
48	head_pitch	Γωνία κεφαλής από την κορυφή της σπονδυλικής στήλης	AU	Δ	Κάτω	7
49	head_yaw	Εκτροπή κεφαλής από την κορυφή της σπονδυλικής στήλης	AU	Δ	Αριστερά	7



50	head_roll	Στροφή κεφαλής από την κορυφή της σπονδυλικής στήλης	AU	Δ	Δεξιά	7
51	lower_t_midlip_o	Κατακόρυφη μετατόπιση του μέσου του εξωτερικού του άνω χείλους	MNS	Δ	Κάτω	8
52	raise_b_midlip_o	Κατακόρυφη μετατόπιση του μέσου του εξωτερικού του κάτω χείλους	MNS	Δ	Πάνω	8
53	stretch_l_cornerlip_o	Οριζόντια μετατόπιση της αριστερής εξωτερικής γωνίας των χειλιών.	MW	Δ	Αριστερά	8
54	stretch_r_cornerlip_o	Οριζόντια μετατόπιση της δεξιάς εξωτερικής γωνίας των χειλιών.	MW	Δ	Δεξιά	8
55	lower_t_lip_lm_o	Κατακόρυφη μετατόπιση του μέσου μεταξύ της αριστερής γωνίας και του μέσου του εξωτερικού του άνω χείλους	MNS	Δ	Κάτω	8
56	lower_t_lip_rm_o	Κατακόρυφη μετατόπιση του μέσου μεταξύ της δεξιάς γωνίας και του μέσου του εξωτερικού του άνω χείλους	MNS	Δ	Κάτω	8
57	raise_b_lip_lm_o	Κατακόρυφη μετατόπιση του μέσου μεταξύ της αριστερής γωνίας και του μέσου του εξωτερικού του κάτω χείλους	MNS	Δ	Πάνω	8
58	raise_b_lip_rm_o	Κατακόρυφη μετατόπιση του μέσου μεταξύ της δεξιάς γωνίας και του μέσου του εξωτερικού του κάτω χείλους	MNS	Δ	Πάνω	8
59	raise_l_cornerlip_o	Κατακόρυφη μετατόπιση της αριστερής εξωτερικής γωνίας των χειλιών	MNS	Δ	Πάνω	8
60	raise_r_cornerlip_o	Κατακόρυφη μετατόπιση της δεξιάς εξωτερικής γωνίας των χειλιών	MNS	Δ	Πάνω	8
61	stretch_r_nose	Οριζόντια μετατόπιση της αριστερής πλευράς της μύτης	ENS	Δ	Αριστερά	9
62	stretch_r_nose	Οριζόντια μετατόπιση της δεξιάς πλευράς της μύτης	ENS	Δ	Δεξιά	9
63	raise_nose	Κατακόρυφη μετατόπιση της άκρης της μύτης	ENS	Δ	Πάνω	9
64	bend_nose	Οριζόντια μετατόπιση της άκρης της μύτης	ENS	Δ	Δεξιά	9
65	raise_l_ear	Κατακόρυφη μετατόπιση του αριστερού αφτιού	ENS	Δ	Πάνω	10
66	raise_r_ear	Κατακόρυφη μετατόπιση του δεξιού αφτιού	ENS	Δ	Πάνω	10
67	pull_l_ear	Οριζόντια μετατόπιση του αριστερού αφτιού	ENS	Δ	Αριστερά	10
68	pull_r_ear	Οριζόντια μετατόπιση του δεξιού αφτιού	ENS	Δ	Δεξιά	10

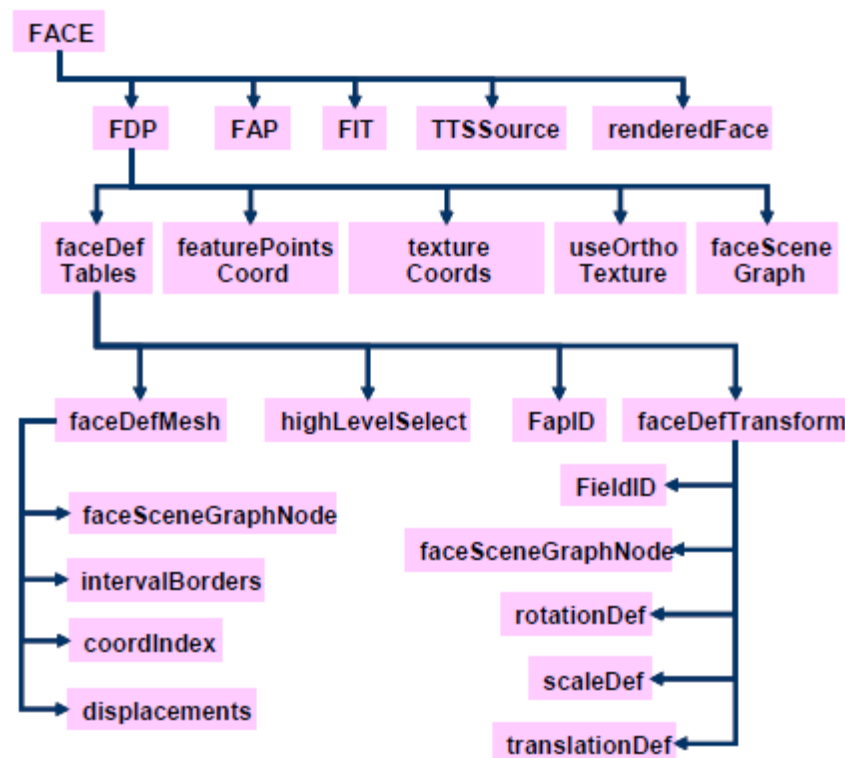
Πίνακας 2.3 : Περιγραφές FAPs

#### 2.3.4 Πίνακες Παρεμβολής FAPs

Όπως αναφέρθηκε προηγουμένως, ο κωδικοποιητής μπορεί να επιτρέψει στον αποκωδικοποιητή να προσεγγίσει τις τιμές κάποιων FAPs με βάση τα μεταδιδόμενα FAPs.

Εναλλακτικά, ο αποκωδικοποιητής μπορεί να προσδιορίσει τους κανόνες παρεμβολής με τη χρήση πινάκων παρεμβολής FAPs (FIT – FAPs Interpolation Tables). Ένας FIT επιτρέπει

ένα μικρότερο σύνολο από FAPs να σταλεί για προσδιορισμό κίνησης προσώπου. Αυτό το μικρό σύνολο μπορεί αργότερα να χρησιμοποιηθεί για τον καθορισμό των τιμών άλλων FAPs, με χρήση λογικής πολυωνυμικής συσχέτισης (rational polynomial mapping) μεταξύ των παραμέτρων. Για παράδειγμα, το FAP του εσωτερικού του πάνω χείλους μπορεί να σταλεί και στη συνέχεια να χρησιμοποιηθεί για τον προσδιορισμό του FAP του εξωτερικού του πάνω χείλους, με χρήση συνάρτησης λογικού πολυωνύμου που προσδιορίζεται στο FIT. Σχηματικά, ο ολοκληρωμένος τρόπος κωδικοποίησης της κίνησης του προσώπου, σύμφωνα με το πρότυπο MPEG-4 δίνεται στο Σχήμα 2.3:



Σχήμα 2.3 : Κωδικοποίηση της κίνησης του προσώπου με το πρότυπο MPEG-4

### 2.3.5 Καθορισμός του μοντέλου του προσώπου

Κάθε αποκωδικοποιητής MPEG-4 που λαμβάνει τα FAPs πρέπει να παράσχει ένα πρότυπο μοντέλο προσώπου MPEG-4 που να περιέχει και κίνηση. Συνήθως, αυτό είναι ένα πρότυπο το οποίο γνωρίζει ο αποκωδικοποιητής. Ο κωδικοποιητής δεν γνωρίζει τη μορφή του πρότυπου προσώπου. Χρησιμοποιώντας παραμέτρους ορισμού του προσώπου (FDP- Face Definition Parameter), το MPEG-4 επιτρέπει στον κωδικοποιητή να διευκρινίσει με μεγάλη ακρίβεια το μοντέλο του προσώπου, του οποίου τα χαρακτηριστικά θα κινηθούν.

Για να υλοποιηθεί αυτό πρέπει να έχει καθοριστεί η στατική γεωμετρία του προτύπου προσώπου στη φυσική του κατάσταση με σκοπό να καθοριστούν οι ιδιότητες της επιφάνειας του προσώπου καθώς και οι κανόνες κίνησης, όπως αυτοί καταγράφονται στους πίνακες κίνησης προσώπου (FAT- Facial Animation Tables), οι οποίοι διευκρινίζουν πώς αυτό το μοντέλο μετασχηματίζεται από τα FAPs [7].

## Κεφάλαιο 3 : Αναγνώριση ανθρώπινου συναισθήματος με MLP

### 3.1 Εισαγωγή

Πόσο καλά πρέπει να αναγνωρίζει ένας υπολογιστής το ανθρώπινο συναίσθημα για να θεωρηθεί ευφυής ; Πρέπει να σημειώσουμε ότι κανένας άνθρωπος δεν μπορεί να κατανοεί απόλυτα όλα τα συναισθήματα καθώς και ότι κάποιες φορές οι άνθρωποι δεν μπορούν να κατανοήσουν ούτε καν τα δικά τους συναισθήματα. Κανένας τρόπος αναγνώρισης συναισθήματος δεν είναι αλάθητος, όμως μερικές πτυχές των εσωτερικών συναισθηματικών διαβαθμίσεων παραμένουν κρυφές, ειδικά αν το άτομο το επιδιώκει ή αν επιλέγει να τα «μεταμφιέσει». Αυτό που μπορεί να δει κάποιος είναι αυτό που μπορεί να παρατηρηθεί και να εκλογικευθεί, και αυτό συνοδευόμενο πάντα με κάποια αμφιβολία. Παρόλα αυτά, οι άνθρωποι αναγνωρίζουν τα συναισθήματα ο ένας του άλλου, αρκετά καλά ώστε να μπορούν να προσαρμόζονται ανάλογα. Ο απώτερος στόχος, επομένως, είναι να αποκτήσουν οι υπολογιστές ικανότητες αναγνώρισης συναισθημάτων ανάλογες με αυτές των ανθρώπων. Σε αυτό το κεφάλαιο θα προσομοιώσουμε την ικανότητα ενός νευρωνικού δικτύου να αναγνωρίζει το ανθρώπινο συναίσθημα. Η εκπαίδευση θα γίνει με επιβλεπόμενη μάθηση μέσω, ενός πολυστρωματικού δικτύου προσοτροφodότησης, πάνω σε δεδομένα που προκύπτουν από μια μαγνητοσκοπημένη ομιλία ενός ανθρώπου.

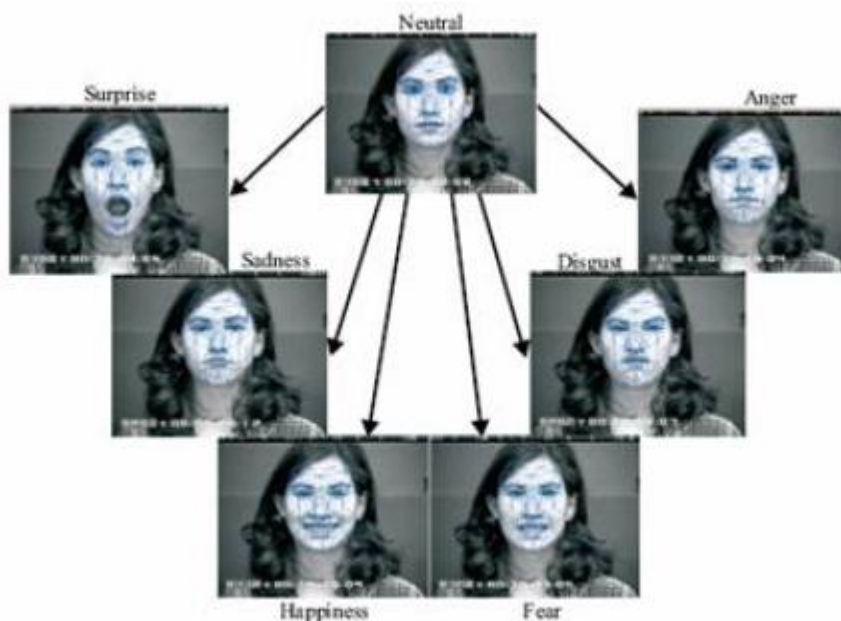
### 3.2 Δεδομένα εισόδου του προβλήματος

Τα δεδομένα, που εξήχθησαν από την μαγνητοσκοπημένη ομιλία, χωρίζονται σε οπτικά και ακουστικά. Τα οπτικά δίνουν πληροφορίες για τη μορφή των χαρακτηριστικών του προσώπου ανάλογα με την έκφραση (Σχήμα 2.1) που προκαλεί το εκάστοτε συναίσθημα στον ομιλητή ενώ τα ακουστικά προκύπτουν από την επεξεργασία των ακουστικών σημάτων που παράγει . Η βιντεοσκόπηση αυτή, που αποτελεί το σύνολο των δεδομένων μας έχει χωριστεί σε διαδοχικά (ως προς τον χρόνο) τμήματα τα οποία ονομάζονται tunes και το καθένα από αυτά αντιστοιχεί σε μία κλάση. Τα frames που ανήκουν στο ίδιο tune ανήκουν προφανώς στην ίδια κλάση και αυτή είναι και η κλάση του tune. Έχουμε, λοιπόν, 477 tunes (το κάθε tune έχει έναν αριθμό από frames) στα οποία υπάρχουν 49 είσοδοι και 5 έξοδοι. Από τις 49 εισόδους, οι 17 αντιστοιχούν στα χαρακτηριστικά του προσώπου και οι 32 στην ομιλία. Συνοπτικά, τα στοιχεία δίνονται στον πίνακα 3.1:

Tunes	είσοδοι	πρόσωπο	φωνήτικά	έξοδοι
477	49	17	32	5

Πίνακας 3.1 : Στοιχεία συνόλου δεδομένων

Στην πειραματική αυτή διαδικασία ασχολούμαστε με 17 παραμέτρους (Πίνακας 3.2) οι οποίες αντιπροσωπεύουν “καταστάσεις” στο πρόσωπο και είναι κωδικοποιημένες με Facial Animation Parameters–FAPs (Παράμετροι Κίνησης Προσώπου), όπως είδαμε στο προηγούμενο κεφάλαιο.



*Σχήμα 3.1 : Εκφράσεις προσώπου που αντιπροσωπεύουν συναίσθημα*

Το μέγεθος της μετακίνησης που περιγράφεται από ένα FAP εκφράζεται μέσω των μονάδων μέτρησης Facial Animation Parameter Units (FAPU), που αντιπροσωπεύουν τμήματα συγκεκριμένων αποστάσεων στο πρόσωπο. Οι περιστροφές περιγράφονται σαν τμήματα ενός ακτινίου.

α/α	Περιγραφή	Όνομα
1	Άνοιγμα σαγονιού	F3_open_jaw
2	Κατέβασμα μεσαίου σημείου πάνω χείλους	F4_lower_top_midlip
3	Ανέβασμα μεσαίου σημείου κάτω χείλους	F5_raise_bottom_midlip inv
4	Διαπλάτυνση στόματος	Fx_widening_mouth
5	Κλείσιμο αριστερού ματιού	F19_21_close_left_eye inv
6	Κλείσιμο δεξιού ματιού	F20_22_close_right_eye inv
7	Σήκωμα εσωτερικού σημείου αριστερού φρυδιού	F31_raise_left_inner_eyebrow
8	Σήκωμα εσωτερικού σημείου δεξιού φρυδιού	F32_raise_right_inner_eyebrow
9	Σήκωμα μεσαίου σημείου αριστερού	F33_raise_left_medium_eyebrow

	φρυδιού	
10	Σήκωμα μεσαίου σημείου δεξιού φρυδιού	F34_raise_right_medium_eyebrow
11	Σήκωμα εξωτερικού σημείου αριστερού φρυδιού	F35_raise_left_outer_eyebrow
12	Σήκωμα εξωτερικού σημείου δεξιού φρυδιού	F36_raise_right_outer_eyebrow
13	Συμπίεση αριστερού φρυδιού	F37_squeeze_left_eyebrow
14	Συμπίεση δεξιού φρυδιού	F38_squeeze_right_eyebrow
15	Πτυχές μεταξύ φρυδιών	Fx_wrinkles_between_eyebrows inv
16	Σήκωμα αριστερής γωνίας χελιών	F59_raise_left_outer_cornerlip inv
17	Σήκωμα δεξιάς γωνίας χελιών	F60_raise_right_outer_cornerlip inv

*Πίνακας 3.2 : Faps που χρησιμοποιούνται σαν δεδομένα εισόδου*

Τα ακουστικά χαρακτηριστικά βασίζονται αποκλειστικά στην προσωδία και σχετίζονται με τον τόνο και τον ρυθμό ενώ το σύνολο των φωνητικών δεδομένων είναι χωρισμένο σε τμήματα. Κάθε tune αναλύθηκε με μία μέθοδο που χρησιμοποιεί την προσωδική αναπαράσταση βασιζόμενη στη μέθοδο που ονομάζεται 'Prosogram'. Η Prosogram [5] βασίζεται στην επεξεργασία των χαρακτηριστικών της βασικής συχνότητας των φωνητικών πυρήνων (το ελάχιστο από τα τμήματα στα οποία μπορεί να "διασπαστεί" η ομιλία). Δίνει δηλαδή σε κάθε φωνητικό πυρήνα έναν τόνο και ένα μήκος. Στη μελέτη αυτή, το μεγάλο πλήθος συναισθημάτων που μπορεί να νιώσει ο άνθρωπος οργανώνεται σε 5 γενικές κατηγορίες συναισθηματικής κατάστασης που ονομάζονται κλάσεις. Αυτή είναι μια κωδικοποίηση για τις εξής γενικές κατηγορίες συναισθηματικής κατάστασης:

- Θετική ενεργητική (κλάση 1)
- Θετική παθητική (κλάση 2)
- Αρνητική ενεργητική (κλάση 3)
- Αρνητική παθητική (κλάση 4)
- Ουδέτερη (κλάση 5)

Σε ένα σύστημα σαν το δικό μας, με τόσο μεγάλο πλήθος εισόδων, θα περίμενε κανείς πολύ μεγάλο πλήθος κρυφών κόμβων οι οποίοι είναι εκθετικά ανάλογοι των εισόδων. Αυτό, βέβαια, θα είχε ως αποτέλεσμα και την εκθετική αύξηση της πολυπλοκότητάς του. Στόχος μας είναι η αποφυγή αυτού του φαινομένου διατηρώντας, όμως, την καλή απόδοση του δικτύου.

### 3.3 Υλοποίηση με ένα feed – forward multilayer perceptron

#### 3.3.1 Γενικά

Η δημιουργία τεχνητού νευρωνικού δικτύου προϋποθέτει τον προσδιορισμό ορισμένων χαρακτηριστικών:

- Την αρχιτεκτονική (δομή και πλήθος κόμβων)
- Τις συναρτήσεις ενεργοποίησης κάθε νευρώνα.

Στόχος της εκπαίδευσης του νευρωνικού δικτύου είναι η οριστικοποίηση της συνάρτησης μεταφοράς του μέσω του καθορισμού των πινάκων βαρών. Για να γίνει αυτό, θα πρέπει να ορίσουμε τον πίνακα εισόδων και τον πίνακα επιθυμητών εξόδων. Άλλες, απαραίτητες για την εκπαίδευση, παράμετροι προς προσδιορισμό είναι:

- Το πλήθος των εποχών (φάσεις εκπαίδευσης )
- Ο ρυθμός μάθησης
- Η ορμή (momentum)

Οι ονομασίες των παραμέτρων έχουν ήδη εξηγηθεί στο κεφάλαιο 1. Η ορμή, όπως είδαμε προσθέτει ένα τμήμα της προηγούμενης ανανέωσης βάρους στην τωρινή. Η μέθοδος αυτή ονομάζεται εκπαίδευση με προσαρμοσμένο ρυθμό μάθησης. Αυτό βοηθάει το σύστημα να συνεχίσει την πορεία του προς την ελαχιστοποίηση του σφάλματος, στην περίπτωση που έχει “φυλακιστεί” σε κάποιο τοπικό ελάχιστο. Σε αυτήν την εργασία σε όλα τα δίκτυα χρησιμοποιείται εκπαίδευση με προσαρμοσμένο ρυθμό μάθησης , με ορμή ίση με 0,9.

Η μεθοδολογία που ακολουθείται συνήθως για την εκπαίδευση ενός νευρωνικού δικτύου είναι η εξής:

Ένα μέρος του συνόλου δεδομένων μας ( data set) χρησιμοποιείται για την εκπαίδευση του δικτύου (train set) ενώ το υπόλοιπο χρησιμοποιείται για να ελεγχθεί η αποδοτικότητα του δικτύου σε άγνωστα δεδομένα (test set). Στην συνέχεια, εφ’ όσον έχει αποφασιστεί η αρχιτεκτονική του δικτύου, γίνονται δοκιμές αλλάζοντας τις παραμέτρους εκπαίδευσης έως ότου να βρεθεί ο συνδυασμός που οδηγεί στην καλύτερη απόδοση. Στόχος είναι το νευρωνικό να εκπαιδευτεί καλά, χωρίς όμως να φτάσει στην υπερεκπαίδευση. Η τελευταία αφορά συνήθως την κλάση εξόδου που εμφανίζεται σε πολύ μεγάλο ποσοστό σε σχέση με τις άλλες στο σύνολο εκπαίδευσης. Τότε έχουμε το φαινόμενο κατά το οποίο το σύστημα αναγνωρίζει παντού την έξοδο αυτή. Αυτό, σε συνδυασμό με το γεγονός ότι η συγκεκριμένη έξοδος συναντάται πολύ συχνά στο σύνολο εκπαίδευσης, έχει ως αποτέλεσμα ένα αρκετά καλό ποσοστό απόδοσης το οποίο όμως είναι πλασματικό. Αυτή ήταν και μία από τις δυσκολίες που συναντήσαμε στην πειραματική μας μελέτη.

Αφού εκπαιδεύσουμε το δίκτυο, μας στην συνέχεια ελέγχουμε την απόδοσή του με βάση του πόσο μικρό είναι το μέσο τετραγωνικό σφάλμα κατά το πέρας της εκπαίδευσης και με βάση την ακρίβεια και την ανάκληση που έχει πάνω στο test set. Ορίζουμε ως εξής:

- Ακρίβεια (Precision) είναι ο λόγος των σωστά ταξινομημένων προτύπων στην κατηγορία προς το σύνολο των ταξινομημένων προτύπων στην κατηγορία επί τοις εκατό (%).

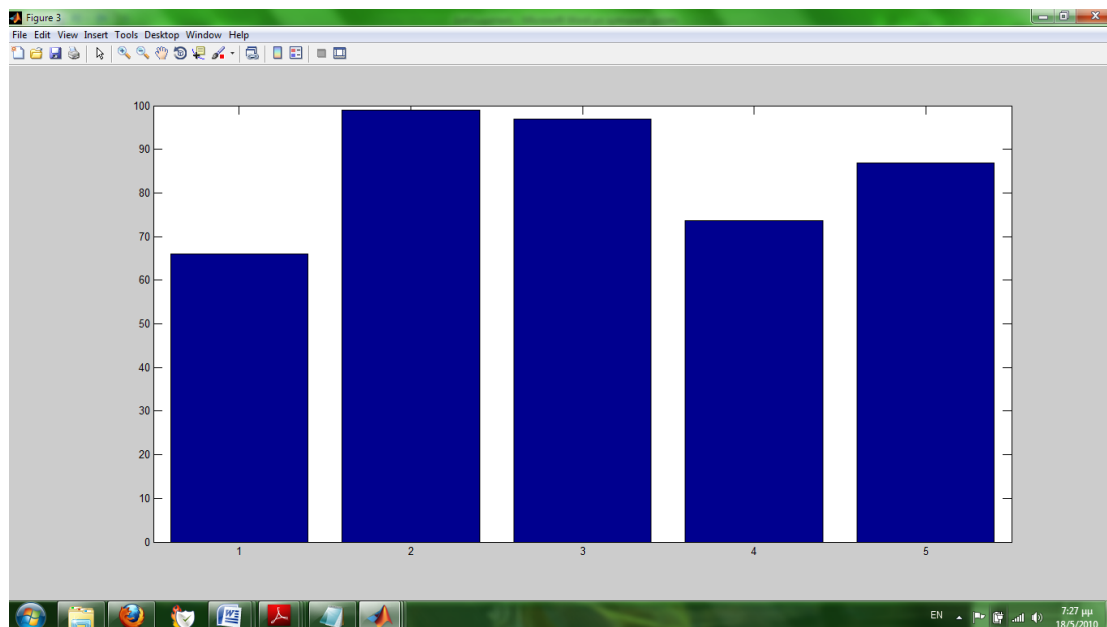


- Ανάκληση (Recall) είναι ο λόγος των σωστά ταξινομημένων προτύπων στην κατηγορία προς το σύνολο των προτύπων στην κατηγορία επί τοις εκατό (%).
- Λόγος επιτυχίας ή επιτυχία (Hit rate) είναι ο λόγος του συνόλου των σωστά ταξινομημένων προτύπων προς το σύνολο των προτύπων του συνόλου ελέγχου.

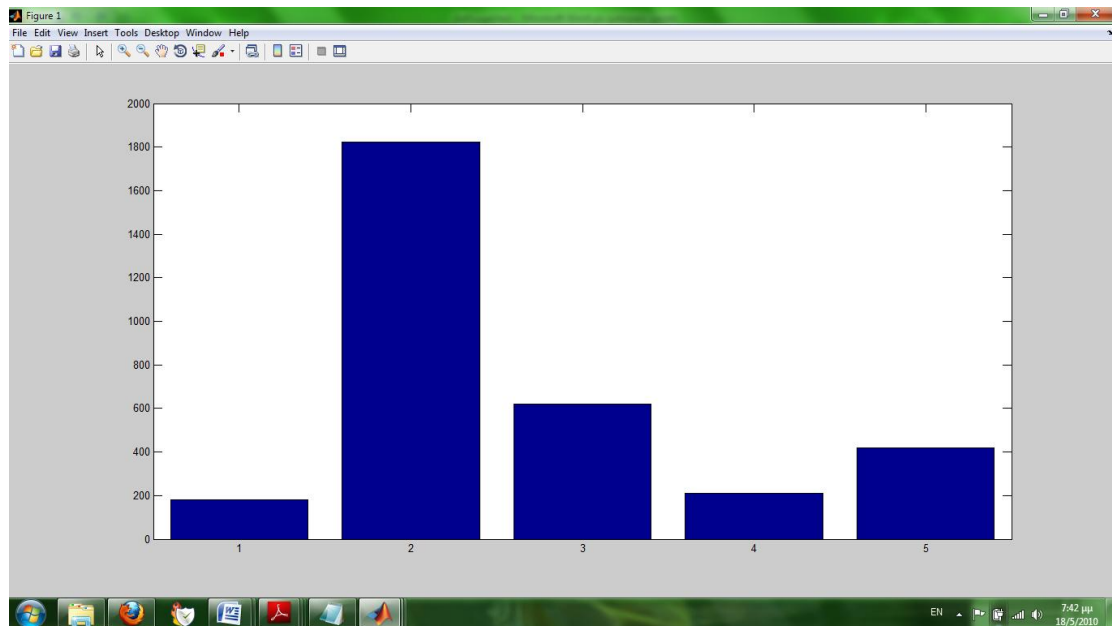
### 3.3.2 Χωρίς προεπεξεργασία δεδομένων

Το δικό μας σύνολο δεδομένων αποτελείται από 477 tunes και κάθε tune από frames. Όλα τα frames συνολικά είναι 13011. Επομένως το data set μας αποτελείται από 13011 πρότυπα 49 μεταβλητών το καθένα.

Δημιουργούμε ένα multilayer perceptron με 49 νευρώνες εισόδου (όσα και τα χαρακτηριστικά εισόδου προφανώς), 1 κρυφό επίπεδο με 20 νευρώνες και 5 νευρώνες εξόδου (όσες και οι κατηγορίες συναισθήματος). Τα 75 % του συνόλου των προτύπων θα χρησιμοποιηθεί για την εκπαίδευση και το υπόλοιπο 25% για τον έλεγχο. Ο αλγόριθμος μάθησης που ακολουθείται είναι back – propagation. Μετά τον έλεγχο προέκυψε ότι οι κλάσεις 2 και 3 αναγνωρίζονται πολύ καλά, αλλά αντίθετα οι υπόλοιπες υστερούν. (Σχ.3.2). Ο λόγος επιτυχίας είναι 92,5 % και το μέσο τετραγωνικό σφάλμα 0,0358. Τα παραπάνω μεγέθη φαίνονται πολύ καλά αλλά είναι πλασματικά μιας και το δίκτυο μας δεν αποδίδει. Γι' αυτό ευθύνεται το γεγονός ότι έχει υπερεκπαιδευτεί όπως εξηγήσαμε και στη προηγούμενη παράγραφο, αφού η κλάση 2 πιο πολύ, αλλά και η 3 εμφανίζονται πολλές περισσότερες φορές (Σχ. 3.3) με αποτέλεσμα να αναγνωρίζονται συνέχεια.



Σχήμα 3.2 : Recall mlp δικτύου χωρίς προεπεξεργασία δεδομένων



Σχήμα 3.3 : Σύνολο εμφανιζόμενων προτύπων από κάθε κατηγορία στο test set

### 3.3.3 Αναζήτηση βέλτιστης αρχιτεκτονικής

Προκειμένου να αποφύγουμε το φαινόμενο της υπερεκπαίδευσης θα προεπεξεργαστούμε τα δεδομένα εισόδου, ώστε να εμφανίζεται περίπου ο ίδιος αριθμός δεδομένων για κάθε κλάση στην είσοδο. Σε αυτήν την εργασία από δω και στο εξής, η παρακάτω διαδικασία θα εφαρμόζεται σε κάθε πείραμα. Στις κλάσεις που αντιπροσωπεύονται περισσότερο αφαιρούμε πρότυπα και στις κλάσεις με μικρότερη αντιπροσώπευση εμφανίζουμε κάποια πρότυπα περισσότερες από μια φορές. Τελικά κάθε κλάση αντιπροσωπεύεται από 4375 πρότυπα. Στη συνέχεια πειραματιζόμαστε με τις διάφορες παραμέτρους του δικτύου για να πετύχουμε τον καλύτερο λόγο επιτυχίας.

Τα σύνολα ελέγχου και εκπαίδευσης χωρίζονται όπως παραπάνω. ( 75% train set ,25% test set.)

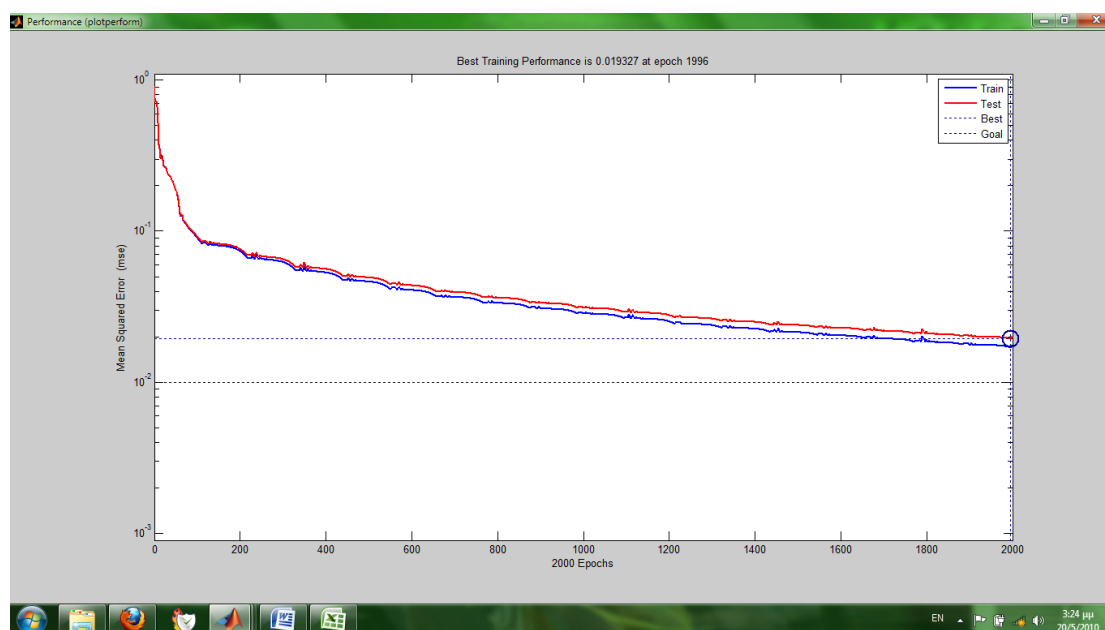
Στον πίνακα που παρατίθεται παρακάτω αποτυπώνονται τα στοιχεία του κάθε δικτύου. (Πίνακας 3.3)

Παρατηρούμε ότι το τρίτο δίκτυο έχει εντυπωσιακή απόδοση και καταλήγουμε σε αυτό ως το καλύτερο. Στα παρακάτω διαγράμματα ( Σχήματα 3.4 και 3.5 ) απεικονίζονται η ακρίβεια καθώς και το μέσο τετραγωνικό σφάλμα του σε συνάρτηση με τις εποχές. Παρότι το μέσο τετραγωνικό σφάλμα είναι απόλυτο μέγεθος, καθώς πάντα συγκλίνει σε μια συγκεκριμένη τιμή, ενώ η επιτυχία εξαρτάται από τις αρχικοποιήσεις των βαρών και δεν είναι αξιόπιστο μέτρο σύγκρισης για μικρές διαφορές, εδώ επιλέξαμε σαν καλύτερο δίκτυο, αυτό με την μεγαλύτερη επιτυχία, επειδή η διαφορά του μέσου τετραγωνικού του σφάλματος ( 0,0177) είναι πολύ μικρή σε σχέση με το 0,172 το οποίο επιτεύχθηκε με 60 νευρώνες στο κρυφό επίπεδο, δηλαδή 10 νευρώνες παραπάνω. Στη συνέχεια, όταν θα συγκρίνουμε αποδόσεις δικτύων, πρώτα θα εξετάσουμε με βάση το μέσο τετραγωνικό σφάλμα και έπειτα με βάση την επιτυχία.



α/α	Αριθμός κρυφών επιπέδων	Νευρώνες στο κρυφό/ά επίπεδο	Εποχές εκπαίδευσης	Ρυθμός μάθησης	Επιτυχία (%)	Μέσο τετραγωνικό σφάλμα
1	1	25	1000	0,15	95,6	0,0384
2	1	25	2000	0,15	96,6	0,0316
3	1	50	2000	0,15	98,37	0,0177
4	2	50/20	2000	0,15	97,6	0,0203
5	1	60	2000	0,15	98,12	0,0172
6	1	70	2000	0,15	98	0,0182
7	1	99	2000	0,15	97,71	0,0198
8	1	35	2000	0,15	97,69	0,0194

Πίνακας 3.3 :Διάφορες αρχιτεκτονικές και αποτελέσματα ελέγχου

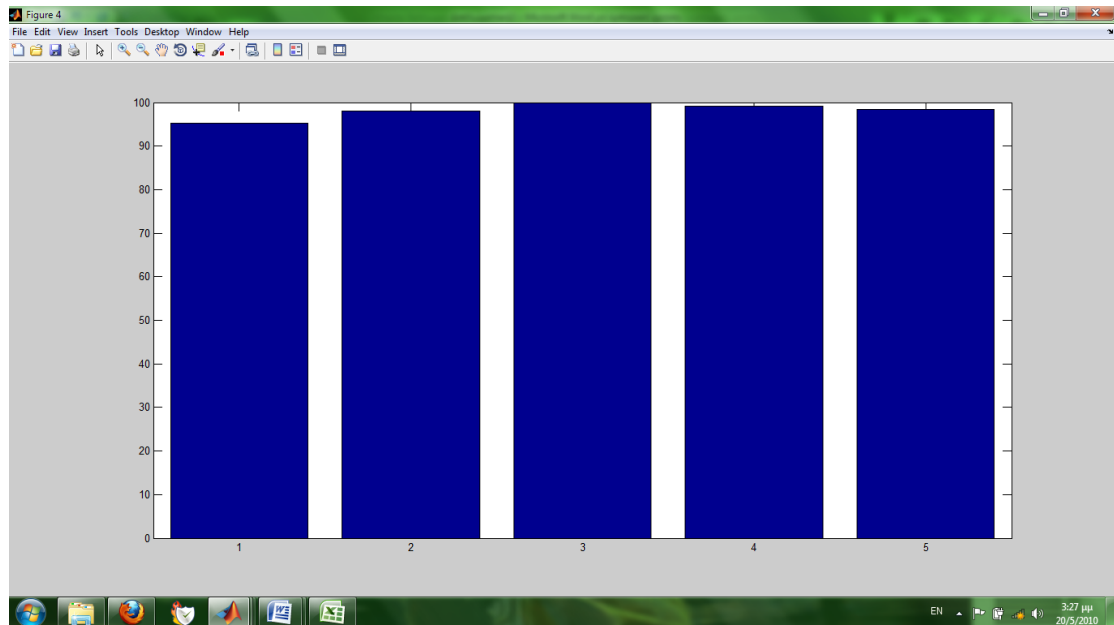


Σχήμα 3.4 : mse mlp με 1 hidden layer, 50 neurons, 2000 epochs, 0,15 learning rate, 0,9 momentum

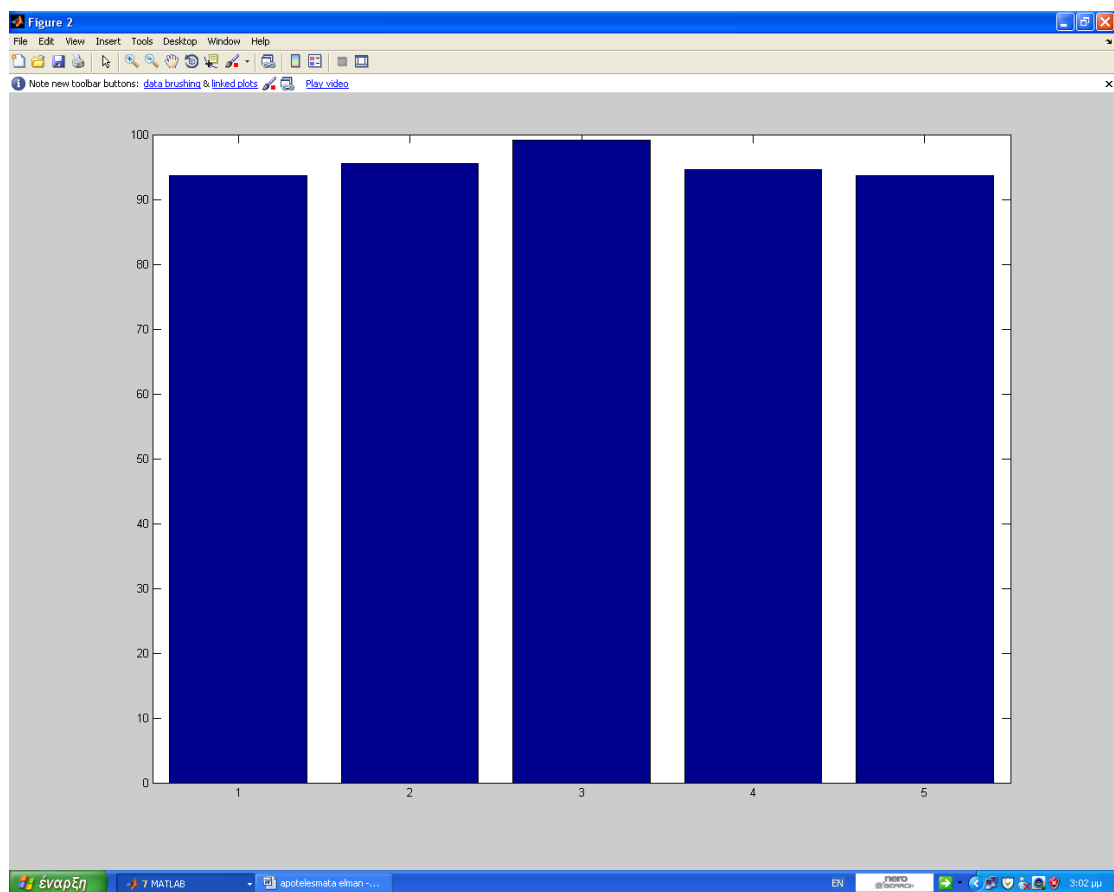
### 3.3.4 Υλοποίηση με είσοδο μόνο των φωνητικών χαρακτηριστικών

Με βάση την παραπάνω αρχιτεκτονική θα ελέγξουμε την απόδοση του νευρωνικού δικτύου με είσοδο 32 μεταβλητών, μια για κάθε φωνητικό χαρακτηριστικό. Μετά την εκπαίδευσή του δικτύου πετύχαμε 94,55% ποσοστό επιτυχίας και 0,0232 μέσο τετραγωνικό σφάλμα. Παρατηρούμε ότι χωρίς τα fars χάνουμε περίπου 4% σε επιτυχία, ποσοστό αρκετά σημαντικό για το πρόβλημα μας αλλά ταυτόχρονα τάξης μεγέθους τέτοιας ώστε να μας δημιουργεί το εξής εύλογο ερώτημα. Αν δεν μπορούμε να παραλείψουμε όλα τα fars, μήπως μπορούμε να παραλείψουμε κάποια, κι αν ναι, τότε ποια και πόσα; Θα προσπαθήσουμε να απαντήσουμε στα

επόμενα κεφάλαια έτσι ώστε να μειωθεί η πολυπλοκότητα και ο χρόνος απόκρισης του προβλήματος μας.



Σχήμα 3.5 : precision mlp με 1 hidden layer, 50 neurons, 2000 epochs, 0,15 learning rate, 0,9 momentum



Σχήμα 3.5 : precision mlp εκπαιδευμένο μόνο με τα φωνητικά χαρακτηριστικά

## Κεφάλαιο 4: Στατιστική ανάλυση των fars

Όπως είδαμε στο προηγούμενο κεφάλαιο η επιτυχία αναγνώρισης του νευρωνικού μας δικτύου είναι τόσο καλή που μας επιτρέπει να προσπαθήσουμε να μειώσουμε την πολυπλοκότητά του. Η ιδέα είναι να μειωθούν οι εισοδοί του δικτύου και οι νευρώνες που χρησιμοποιούνται στο κρυφό επίπεδο, ώστε να έχουμε λιγότερα συναισθηματικά βάρη και επομένως λιγότερους υπολογισμούς έως ότου το δίκτυο να αποκριθεί στην εμφάνιση ενός προτύπου στη είσοδό του.

### 4.1 Εισαγωγή

Στο κεφάλαιο 2 είδαμε τον τρόπο με τον οποίο υπολογίζονται τα fars. Λόγω της φύσης τους λοιπόν γίνεται εύκολα η υπόθεση ότι σε μια φυσική ομιλία ενός ανθρώπου, όπως αυτή που μαγνητοσκοπήσαμε, η μετατόπιση του κάθε σημείου FP, μιας και μιλάμε για σημεία στο πρόσωπο, ίσως σχετίζεται με την μετατόπιση κάποιου άλλου FP. Για παράδειγμα, όταν ο ομιλητής νιώσει έκπληξη και σηκώσει τα φρύδια του, δε θα μετακινηθούν με παρόμοιο τρόπο τα μεσαία σημεία και του αριστερού και του δεξιού φρυδιού; Αν λοιπόν, στα δεδομένα εισόδου του δικτύου, δύο ή παραπάνω μεταβλητές ακολουθούν την ίδια αυξομείωση στο πέρασμα του χρόνου, γιατί απαιτούνται και οι δύο στην εκπαίδευση του δικτύου και κατά συνέπεια στην λειτουργία του; Στο κεφάλαιο αυτό θα εξεταστεί η στατιστική συσχέτιση των fars και στη συνέχεια θα εκπαιδεύσουμε το δίκτυο με εισόδους τις ασυσχέτιστες μεταβλητές. Κατόπιν θα ελέγχουμε την απόδοσή του.

### 4.2 Συσχέτιση δύο μεταβλητών

Στην παράγραφο αυτή θα μελετήσουμε από θεωρητική σκοπιά την στατιστική συσχέτιση δύο μεταβλητών ώστε να καταλάβουμε την έννοια της.

Θεωρούμε δύο τυχαίες μεταβλητές  $X$ ,  $Y$  και  $n$  ζεύγη παρατηρήσεων  $(x_1, y_1)$ ,  $(x_2, y_2)$ , ...,  $(x_n, y_n)$ , από τυχαίο δείγμα μεγέθους  $n$ .

Αναφερόμαστε, δηλαδή, σε μη πειραματικά δεδομένα (ο ερευνητής δεν προκαθορίζει-ελέγχει τις τιμές καμιάς από τις δύο μεταβλητές) όπως,

- $X$  το ύψος των φοιτητών ενός πανεπιστημιακού τμήματος και  $Y$  το βάρος τους
- $X$  οι ώρες μελέτης των φοιτητών ενός πανεπιστημιακού τμήματος και  $Y$  η απόδοση τους σε ένα τεστ
- Έτσι όπως στην περίπτωση μας, για παράδειγμα  $X$  η μετατόπιση της αριστερής γωνίας των χειλιών ενός ανθρώπου που μιλάει και  $Y$  η μετατόπιση της δεξιάς γωνίας.

Δεν αναφερόμαστε όμως σε περιπτώσεις όπως,

- $X$  ο αριθμός των ανοιχτών ταμείων ενός υποκαταστήματος τραπεζής (που καθορίζει ο διευθυντής) και  $Y$  ο χρόνος αναμονής των πελατών
- $X$  η ποσότητα λιπάσματος (που καθορίζει ο ερευνητής) και  $Y$  η απόδοση του αγρού

- $X$  το ύψος της διαφημιστικής δαπάνης ενός προϊόντος (που καθορίζει μια επιχείρηση) και  $Y$  το ύψος των πωλήσεων του προϊόντος.

Στις περιπτώσεις όπου από τον πληθυσμό επιλέγουμε ένα τυχαίο δείγμα και σε κάθε μονάδα του δείγματος μελετάμε δύο ή περισσότερα χαρακτηριστικά αναζητούμε μέτρα τα οποία να μπορούν να εκφράσουν και να ποσοτικοποιήσουν την πιθανή συµµεταβολή - συσχέτιση των χαρακτηριστικών. Ποσό ισχυρή είναι η συµµεταβολή των μεταβλητών;

Ένας απλός τρόπος για να αποκτήσουμε μια πρώτη ιδέα για το αν και πώς δυο μεταβλητές συµµεταβάλλονται-συσχετίζονται, είναι να κατασκευάσουμε το διάγραμμα διασποράς (Scatter Diagram). Να αναπαραστήσουμε δηλαδή τα ζεύγη των παρατηρήσεων σε ένα διάγραμμα. Ας δούμε ένα παράδειγμα: Στον πίνακα που ακολουθεί φαίνονται οι παρατηρήσεις για το ύψος και το βάρος 16 εργατών μιας βιομηχανίας.

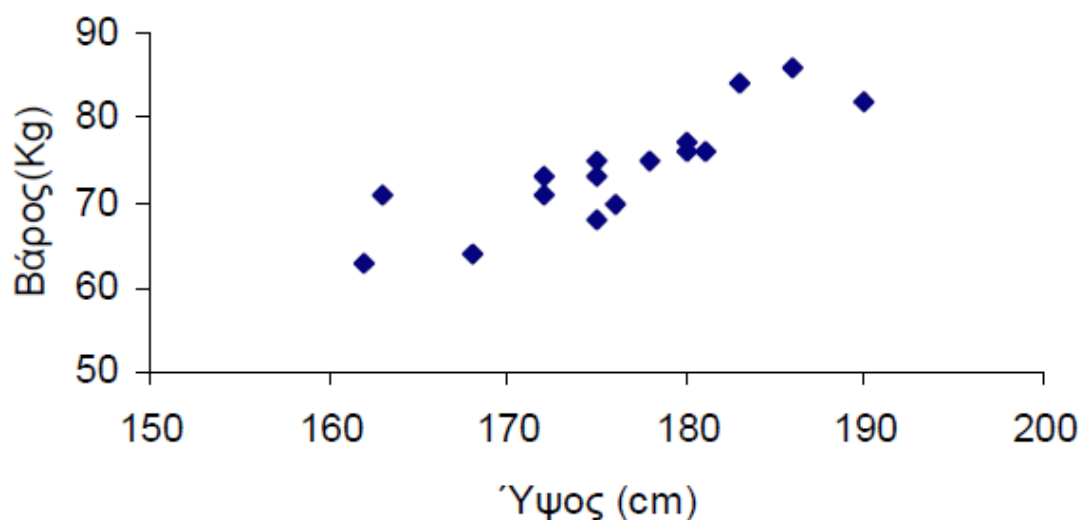
	1	2	3	4	5	6	7	8
Ύψος (cm)	183	162	172	181	180	168	176	180
Βάρος (Kg)	84	63	71	76	77	64	70	76

	9	10	11	12	13	14	15	16
Ύψος (cm)	190	175	178	175	186	172	175	163
Βάρος (Kg)	82	68	75	73	86	73	75	71

Πίνακας 4.1 : παρατηρήσεις για το ύψος και το βάρος 16 εργατών μιας βιομηχανίας.

Από το *διάγραμμα διασποράς* φαίνεται ότι οι εργάτες στο δείγμα που έχουν μεγαλύτερο ύψος έχουν και μεγαλύτερο βάρος. Φαίνεται, δηλαδή, να υπάρχει μια ανάλογη σχέση μεταξύ του ύψους και του βάρους των εργατών.

Πόσο ισχυρή είναι όμως αυτή η συσχέτιση; Πώς μπορεί, δηλαδή, να μετρηθεί; Στο πλαίσιο της εργασίας, θα ασχοληθούμε με ένα μέτρο συσχέτισης: με το *συντελεστή γραμμικής συσχέτισης του Pearson*.



Σχήμα 4.1 : διάγραμμα διασποράς του πίνακα 4.1

Ο δειγματικός συντελεστής γραμμικής συσχέτισης του Pearson συμβολίζεται με  $r$  και ορίζεται από τον τύπο:

$$r = \frac{s_{xy}}{s_x \cdot s_y}$$

όπου,

$$s_{xy} = \text{Cov}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n-1} = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{n-1}$$

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \text{ και } s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$$

Επομένως

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2} \cdot \sqrt{\sum_{i=1}^n y_i^2 - n \cdot \bar{y}^2}}$$

Στους παραπάνω τύπους με  $\bar{x}$ ,  $\bar{y}$  συμβολίζονται οι μέσες τιμές των μεταβλητών  $X$ ,  $Y$ . Δηλαδή  $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$  και αντίστοιχα για  $\bar{y}$ .

Ας δούμε τώρα την ερμηνεία και τις ιδιότητες του συντελεστή γραμμικής συσχέτισης  $r$ :

- Ο συντελεστής γραμμικής συσχέτισης  $r$  δίνει ένα μέτρο του μεγέθους της γραμμικής συσχέτισης μεταξύ δύο μεταβλητών.
- Παίρνει τιμές στο κλειστό διάστημα  $[-1, 1]$
- Αν  $r = \pm 1$  υπάρχει **τέλεια γραμμική** συσχέτιση.  
Αν  $-0,3 \leq r < 0,3$  **δεν υπάρχει γραμμική** συσχέτιση. Αυτό, όμως, δεν σημαίνει ότι δεν υπάρχει άλλου είδους συσχέτιση μεταξύ των δύο μεταβλητών.  
Αν  $-0,5 < r \leq -0,3$  ή  $0,3 \leq r < 0,5$  υπάρχει **ασθενής γραμμική** συσχέτιση.  
Αν  $-0,7 < r \leq -0,5$  ή  $0,5 \leq r < 0,7$  υπάρχει **μέση γραμμική** συσχέτιση.  
Αν  $-0,8 < r \leq -0,7$  ή  $0,7 \leq r < 0,8$  υπάρχει **ισχυρή γραμμική** συσχέτιση.  
Αν  $-1 < r \leq -0,8$  ή  $0,8 \leq r < 1$  υπάρχει **πολύ ισχυρή γραμμική** συσχέτιση.
- Θετικές τιμές του  $r$  δεν υποδηλώνουν, κατ' ανάγκην μεγαλύτερο βαθμό γραμμικής συσχέτισης από το βαθμό γραμμικής συσχέτισης που υποδηλώνουν αρνητικές τιμές του  $r$ . Ο βαθμός γραμμικής συσχέτισης καθορίζεται από την απόλυτη τιμή του  $r$  και

όχι από το πρόσημο του  $r$ . Το πρόσημο του  $r$  καθορίζει το είδος, μόνο, της συσχέτισης (θετική ή αρνητική). Μας πληροφορεί δηλαδή για το αν αύξηση της μιας μεταβλητής αντιστοιχεί σε αύξηση ή σε μείωση της άλλης [8].

### 4.3 Συσχέτιση των faps

Σύμφωνα με τους παραπάνω τύπους υπολογίζουμε τον συντελεστή συσχέτισης  $r$  μεταξύ των 17 faps ανά δύο. Οι τιμές που προκύπτουν παρουσιάζονται στους δύο παρακάτω πίνακες στους οποίους η πρώτη στήλη και γραμμή αναφέρονται στον αύξοντα αριθμό του κάθε far, έτσι όπως αυτά παρουσιάστηκαν στον πίνακα 3.2.

	1	2	3	4	5	6	7	8
2	0,095							
3	<b>0,723</b>	<b>0,751</b>						
4	0,221	0,187	0,267					
5	0,252	0,381	0,422	0,313				
6	0,263	0,430	0,470	0,342	<b>0,568</b>			
7	0,057	0,282	0,230	0,057	0,077	-0,008		
8	0,105	0,199	0,208	0,099	0,048	0,065	<b>0,605</b>	
9	0,063	0,287	0,236	0,077	0,064	-0,019	<b>0,931</b>	<b>0,625</b>
10	0,122	0,185	0,209	0,127	0,034	0,066	<b>0,612</b>	<b>0,926</b>
11	0,035	0,231	0,180	0,028	0,030	-0,060	<b>0,914</b>	<b>0,549</b>
12	0,096	0,192	0,197	0,095	0,016	0,018	<b>0,613</b>	<b>0,951</b>
13	0,077	0,099	0,124	0,171	0,195	0,139	0,037	-0,095
14	0,147	0,201	0,240	0,242	0,213	0,302	-0,161	-0,161
15	0,174	0,308	0,325	0,238	0,274	0,259	0,297	0,260
16	0,345	<b>0,625</b>	<b>0,656</b>	0,386	<b>0,577</b>	<b>0,654</b>	0,014	0,113
17	0,374	<b>0,642</b>	<b>0,686</b>	0,402	<b>0,595</b>	<b>0,634</b>	0,095	0,084

Πίνακας 4.2 Συντελεστής συσχέτισης 1<sup>ου</sup> έως 8<sup>ου</sup> far με τα υπόλοιπα

	9	10	11	12	13	14	15	16
10	<b>0,652</b>							
11	<b>0,912</b>	<b>0,558</b>						
12	<b>0,636</b>	<b>0,912</b>	0,570					
13	0,073	-0,064	-0,037	-0,114				
14	-0,158	-0,129	-0,190	-0,195	0,381			
15	0,286	0,270	0,327	0,301	-0,376	-0,325		
16	-0,014	0,112	-0,042	0,087	-0,019	0,259	0,457	
17	0,075	0,081	0,033	0,062	0,079	0,233	0,431	<b>0,869</b>

*Πίνακας 4.3. Συντελεστής συσχέτισης 9<sup>ου</sup> έως 16<sup>ου</sup> fap με τα υπόλοιπα*

Είναι σαφές , και από τους τύπους που παρουσιάσαμε στην προηγούμενη παράγραφο , ότι οι τιμές του συντελεστή συσχέτισης είναι ανά δύο ίδιες και γι' αυτό το λόγο δεν συμπεριλαμβάνουμε στον πίνακα τις δυαδικές τιμές. Με έντονη γραφή φαίνονται οι τιμές εκείνες για τις οποίες έχουμε μέση γραμμική συσχέτιση ή ισχυρότερη. (  $|r| > 0,5$  ).

#### 4.4 Ομαδοποίηση των faps με βάση τη συσχέτισή τους

Όπως προκύπτει από τους πίνακες 4.2 και 4.3 μπορούμε να διαχωρίσουμε τα faps σε τρεις κατηγορίες με βάση το συντελεστή συσχέτισης τους σε σχέση με τα υπόλοιπα.

- **Ασυσχέτιστα (  $|r| < 0,5$  )**

Σε αυτήν την κατηγορία ανήκουν τα faps εκείνα που δεν έχουν γραμμική συσχέτιση ή έχουν ασθενή γραμμική συσχέτιση με τα υπόλοιπα και είναι τα παρακάτω:

α/α	Όνομα
4	Fx_widening_mouth
13	F37_squeeze_left_eyebrow
14	F38_squeeze_right_eyebrow
15	Fx_wrinkles_between_eyebrows inv

- **Πολύ ισχυρά συσχετισμένα  $r > 0,8$**

Σε αυτήν την κατηγορία ανήκουν τα faps εκείνα που έχουν πολύ ισχυρή γραμμική συσχέτιση με κάποια από τα υπόλοιπα. Εάν θεωρήσουμε ότι όλα τα χαρακτηριστικά που συσχετίζονται μεταξύ τους με  $|r| > 0,8$  αποτελούν μια ομάδα, τότε προκύπτουν οι παρακάτω ομάδες:

A) 7 με 9      7 με 11      9 με 11,  $r > 0,9$

α/α	Όνομα
7	F31_raise_left_inner_eyebrow
9	F33_raise_left_medium_eyebrow
11	F35_raise_left_outer_eyebrow

B) 8 με 10      8 με 12      10 με 12,  $r > 0,9$

α/α	Όνομα
8	F31_raise_right_inner_eyebrow
10	F33_raise_right_medium_eyebrow
12	F35_raise_right_outer_eyebrow

Γ) 16 με 17

α/α	Όνομα
16	F59_raise_left_outer_cornerlip inv
17	F60_raise_right_outer_cornerlip inv

- **Συσχετισμένα ( $0,5 < |r| < 0,8$ )**

Σε αυτήν την κατηγορία ανήκουν τα faps εκείνα που έχουν μέση γραμμική ή ισχυρή γραμμική συσχέτιση με κάποια από τα υπόλοιπα. Εάν θεωρήσουμε ότι όλα τα χαρακτηριστικά που συσχετίζονται μεταξύ τους με  $0,5 < |r| < 0,8$  αποτελούν μια ομάδα, τότε προκύπτουν οι παρακάτω ομάδες :

A)

α/α	Όνομα
1	F3_open_jaw
3	F5_raise_bottom_midlip inv



B)

α/α	Όνομα
2	F4_lower_top_midlip
3	F5_raise_bottom_midlip inv
16	F59_raise_left_outer_cornerlip inv
17	F60_raise_right_outer_cornerlip inv

Γ)

α/α	Όνομα
5	F19_21_close_left_eye inv
6	F20_22_close_right_eye inv
16	F59_raise_left_outer_cornerlip inv
17	F60_raise_right_outer_cornerlip inv

Δ)

α/α	Όνομα
7	F31_raise_left_inner_eyebrow
8	F32_raise_right_inner_eyebrow
9	F33_raise_left_medium_eyebrow
10	F34_raise_right_medium_eyebrow
11	F35_raise_left_outer_eyebrow
12	F36_raise_right_outer_eyebrow

Παρατηρούμε ότι οι ομάδες που προκύπτουν , τόσο για τις ισχυρά συσχετισμένες, όσο και για τις υπόλοιπες, αποτελούνται από “ συγγενικά ” faps, αποτέλεσμα το οποίο και περιμέναμε.

#### **4.5 Εκπαίδευση και έλεγχος απόδοσης mlp με είσοδο που έχει μεταβλητές επιλεγμένες κάθε φορά, με βάση την συσχέτιση τους.**

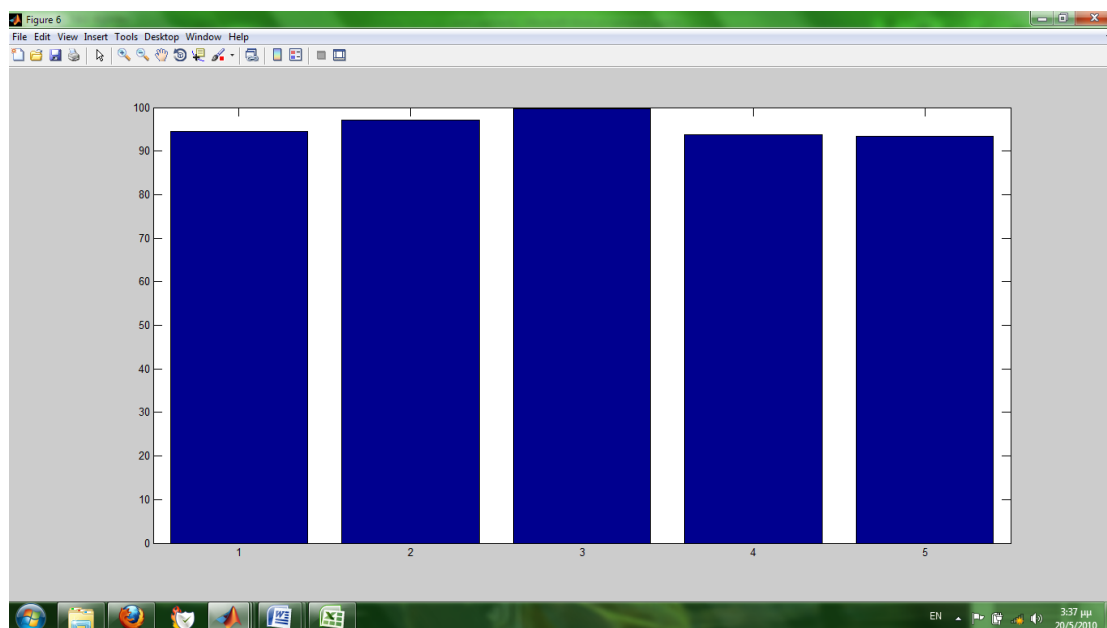
Όπως είπαμε στην εισαγωγή του κεφαλαίου, αλλά και κρίνοντας από τα αποτελέσματα της προηγούμενης παραγράφου, καταλήγουμε να εκπαιδεύσουμε το δίκτυο μας αφαιρώντας συσχετισμένες μεταβλητές, ώστε να μειώσουμε την πολυπλοκότητα. Είναι σαφές ότι οι ασυσχέτιστες μεταβλητές θα συμμετέχουν πάντα στην εκπαίδευση. Αρχικά κάνουμε δοκιμές σε δίκτυο το οποίο έχει ίδια αρχιτεκτονική με αυτό που έχει επιλεγεί στο Κεφάλαιο 3 ως βέλτιστο και με τους ίδιους κανόνες εκπαίδευσης. Αυτό γίνεται για να μπορούμε να συγκρίνουμε “δίκαια” το δίκτυο με τις μειωμένες εισόδους έναντι του βέλτιστου. Στην συνέχεια ,

εάν τα αποτελέσματα είναι ικανοποιητικά , μπορούμε να “ κόψουμε ” νευρώνες, επομένως συναπτικά βάρη, επομένως υπολογισμούς, επομένως πολυπλοκότητα και χρόνο απόκρισης.

Παρακάτω θα παρουσιάσουμε μια σειρά από πειράματα που κάναμε, τροφοδοτώντας το δίκτυο με διαφορετικές μεταβλητές εισόδου. Να υπενθυμίσουμε ότι η αρχιτεκτονική του δικτύου έχει πενήντα νευρώνες στο κρυφό επίπεδο και η εκπαίδευση γίνεται για 2000 εποχές, με 0,15 ρυθμό μάθησης.

#### 4.5.1 Πείραμα 1

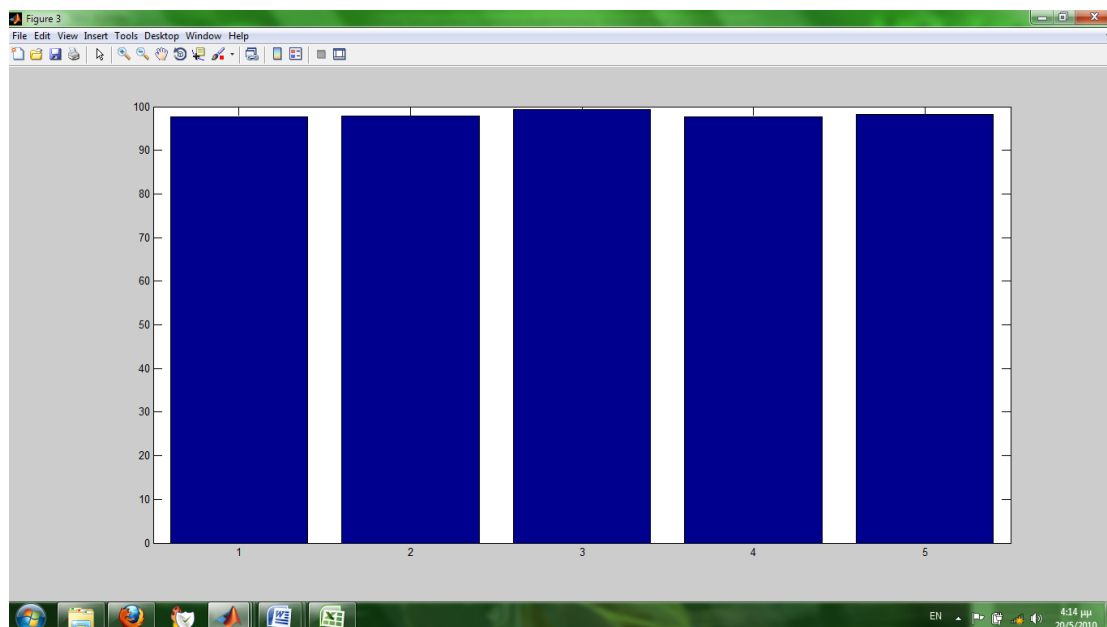
Αρχικά συμπεριλάβαμε όλες τις μεταβλητές εισόδου, κρατώντας όμως μόνο μια μεταβλητή από κάθε ομάδα των πολύ ισχυρά συσχετισμένων μεταβλητών. Συγκεκριμένα συμπεριλάβαμε από την ομάδα (Α) το χαρακτηριστικό 9 (F33\_raise\_left\_medium\_eyebrow) και αφαιρέσαμε τα χαρακτηριστικά 7 και 11 (F31\_raise\_left\_inner\_eyebrow και F35\_raise\_left\_outer\_eyebrow). Από την ομάδα (Β) συμπεριλάβαμε το χαρακτηριστικό 10 (F33\_raise\_right\_medium\_eyebrow) και αφαιρέσαμε τα χαρακτηριστικά 8 και 12 (F31\_raise\_right\_inner\_eyebrow και F35\_raise\_right\_outer\_eyebrow). Τέλος, από την ομάδα (Γ) συμπεριλάβαμε το χαρακτηριστικό 17 (F60\_raise\_right\_outer\_cornerlip inv) και αφαιρέσαμε το 16 (F59\_raise\_left\_outer\_cornerlip inv). Αφαιρώντας αυτές τις 5 μεταβλητές, εκπαιδεύουμε το δίκτυο και μετά τον έλεγχο είχαμε επιτυχία 95.56% και μέσο τετραγωνικό σφάλμα 0.0345. Επίσης στο Σχήμα 4.2 φαίνεται και η ακρίβεια του δικτύου.



Σχήμα 4.2 : Ακρίβεια πειράματος 1

#### 4.5.2 Πείραμα 2

Στην συνέχεια , μιας και η επιτυχία του προηγούμενου πειράματος ήταν καλή, θα αφαιρέσουμε περισσότερες μεταβλητές εισόδου. Συγκεκριμένα συμπεριλαμβάνουμε τις ασυσχέτιστες μεταβλητές και επίσης μια μεταβλητή από κάθε ομάδα των συσχετισμένων μεταβλητών. Αυτές είναι , από την ομάδα (Α) το far 1 (F3\_open\_jaw) , από την (Β) το 3 (F5\_raise\_bottom\_midlip inv) , από την (Γ) το 6 (F20\_22\_close\_right\_eye inv ) και από την (Δ) το 10 (F34\_raise\_right\_medium\_eyebrow). Με αυτές τις μεταβλητές εισόδου είχαμε επιτυχία 98.18% και μέσο τετραγωνικό σφάλμα 0.017. Η ακρίβεια φαίνεται στο Σχήμα 4.3. Τα παραπάνω αποτελέσματα είναι ισοδύναμα του αρχικού μας δικτύου με τις 49 μεταβλητές εισόδου. Επομένως μπορούμε να αφαιρέσουμε κι άλλες μεταβλητές.

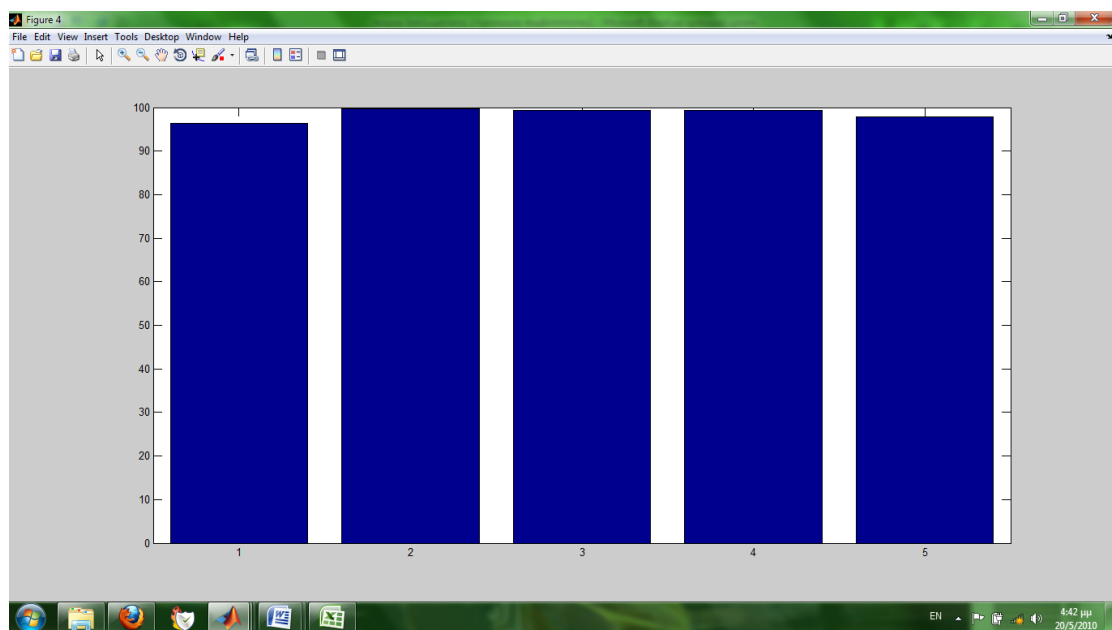


Σχήμα 4.3 : Ακρίβεια πειράματος 2

#### 4.5.3 Πείραμα 3

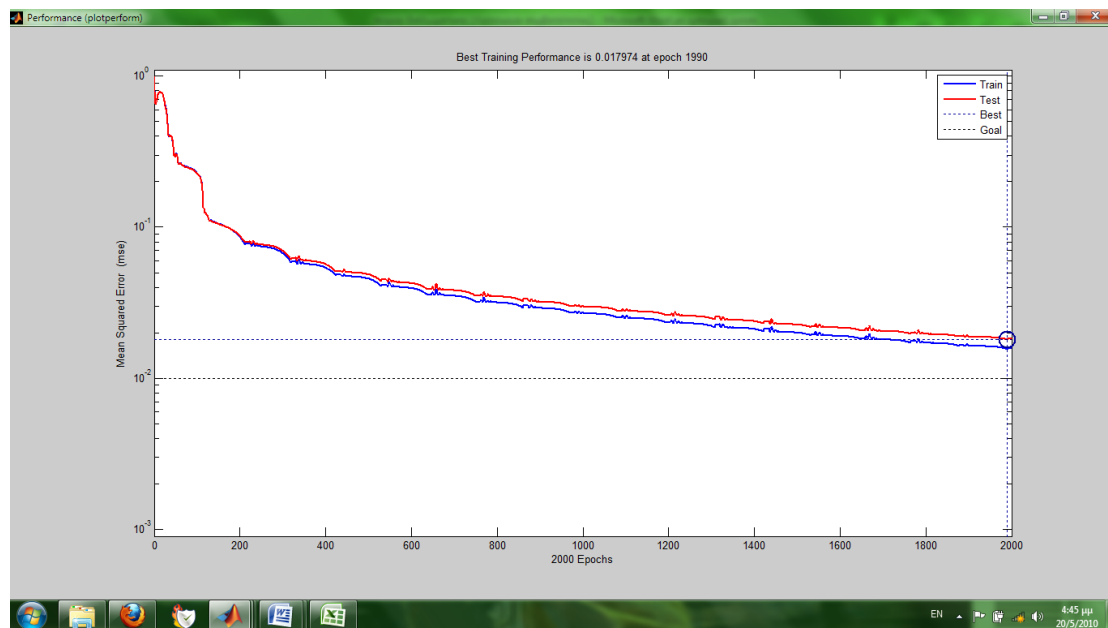
Στο τελευταίο πείραμα θα συμπεριλάβουμε μόνο τις ασυσχέτιστες μεταβλητές. Μόνο με τα χαρακτηριστικά της φωνής και τα τέσσερα ασυσχέτιστα fars παίρνουμε 98.51% επιτυχία και μέσο τετραγωνικό σφάλμα ίσο με 0.0156. Στα διαγράμματα που ακολουθούν βλέπουμε την ακρίβεια του δικτύου αυτού (Σχήμα 4.4) και την καμπύλη του μέσου τετραγωνικού σφάλματος σε συνάρτηση με τις εποχές εκπαίδευσης (Σχήμα 4.5). Η επίδοση που έχουμε πετύχει είναι εντυπωσιακή και καλύτερη της αρχικής μας επίδοσης. Μιας και κρατήσαμε την

αρχιτεκτονική (50 νευρώνες στο κρυφό επίπεδο) του αρχικού βέλτιστου δικτύου προκειμένου να συγκρίνουμε τις δυο επιδόσεις είναι λογικό να περιμένουμε ένα καλό αποτέλεσμα. Αυτό γιατί , ναι μεν έχουμε λιγότερα χαρακτηριστικά εισόδου ,αλλά από την άλλη έχουμε ένα πολύπλοκο δίκτυο για να υπολογίσει την έξοδο μας, το οποίο έχει μεγαλύτερη υπολογιστική ικανότητα. Εφόσον ο στόχος μας είναι να δούμε πόσο μπορούμε να μειώσουμε την πολυπλοκότητα , χωρίς να χάσουμε ιδιαίτερα σε επιτυχία, στην επόμενη παράγραφο θα πειραματιστούμε με διαφορετικές αρχιτεκτονικές δικτύων με είσοδο 36 μεταβλητών (των 32 χαρακτηριστικών φωνής και των τεσσάρων ασυσχέτιστων προσώπου).



Σχήμα 4.4 : Ακρίβεια πειράματος 3

Τέλος, από τα διαγράμματα ακρίβειας των παραπάνω πειραμάτων (Σχήματα 4.1 – 4.2) είναι εμφανές ότι σε αντίθεση με τα δίκτυα των πειραμάτων 1 και 2 , τα οποία αναγνωρίζουν πολύ καλά την κλάση 3 αλλά όχι τόσο καλά τις υπόλοιπες, το δίκτυο του πειράματος 3 είναι ακριβέστατο στις κλάσεις 2,3,4 , προβλέπει πολύ καλά την κλάση 5, ενώ υστερεί μόνο στην κλάση 1. Την ίδια ακριβώς συμπεριφορά είχε παρουσιάσει και το αρχικό μας δίκτυο .



Σχήμα 4.5 : Μέσο τετραγωνικό σφάλμα σε κάθε εποχή του πειράματος 3

#### 4.6 Αναζήτηση βέλτιστης αρχιτεκτονικής με είσοδο τα χαρακτηριστικά φωνής και τα ασυσχέτιστα fars.

Συγκρίνουμε διάφορες αρχιτεκτονικές με τα 32 χαρακτηριστικά φωνής και τα 4 ασυσχέτιστα fars. Παρατίθεται πίνακας(Πίνακας 4.4) με τις παραμέτρους του κάθε δικτύου που δοκιμάσαμε.

α/α	Νευρώνες στο κρυφό επίπεδο	Εποχές εκπαίδευσης	Ρυθμός μάθησης	Επιτυχία (%)	Μέσο τετραγωνικό σφάλμα
1	25	2000	0,15	95,45	0,0296
2	25	3000	0,15	96,29	0,0238
3	30	2000	0,15	96,73	0,0249
4	35	2000	0,15	97,8	0,0220
5	35	2000	0,20	97,41	0,0202
6	35	2600	0,15	97,78	0,202

Πίνακας 4.4: Διάφορες αρχιτεκτονικές και αποτελέσματα ελέγχου νευρωνικού με είσοδο 36 μεταβλητών

Με βάση τα παραπάνω αποτελέσματα βλέπουμε ότι έχουμε μέσο τετραγωνικό σφάλμα 0,0202 με τις παραμέτρους του δικτύου με τον αύξοντα αριθμό 5, το οποίο έχει επιτυχία 97,41%. Κόβοντας , δηλαδή , 15 νευρώνες από το κρυφό επίπεδο πετυχαίνουμε να κάνουμε κατηγοριοποίηση των δεδομένων με λιγότερο από 0.03 στο μέσο τετραγωνικό σφάλμα και 1% σφάλμα στην επιτυχία. Το αποτέλεσμα μας είναι αρκετά καλό και αποδεικνύει ότι η ανάλυση συσχέτισης που κάναμε , μας οδήγησε στο να πλησιάσουμε τον αρχικό μας στόχο, να μειώσουμε την πολυπλοκότητα του προβλήματος μας χωρίς σημαντικές απώλειες.

## Κεφάλαιο 5: Ανάλυση ευαισθησίας σε νευρωνικό δίκτυο εκπαιδευμένο για αναγνώριση συναισθήματος

### 5.1 Εισαγωγή

Η εφαρμογή αλγορίθμων εξαγωγής κανόνων στα φυσικά προβλήματα είναι η βασική δυσκολία στην προσπάθεια σύνθεσης αντιπροσωπευτικού συνόλου δεδομένων. Οι ερευνητές συχνά δυσκολεύονται να εντοπίσουν εκείνες τις μεταβλητές που επηρεάζουν πραγματικά την έξοδο του προβλήματος. Τα νευρωνικά δίκτυα μαθαίνουν από τα δεδομένα εισόδου (σε συνδυασμό με τα δεδομένα εξόδου). Όταν, όμως, ο όγκος δεδομένων είναι μεγάλος, είναι δύσκολο να δει κανείς τις σχέσεις και αλληλεπιδράσεις μέσα σε αυτόν. Συνεπώς, μπορεί έτσι να συμπεριληφθούν περιττά δεδομένα εισόδου ή να παραλειφθούν μία ή περισσότερες σημαντικές παράμετροι. Η ανάλυση ευαισθησίας εφαρμοσμένη σε ένα εκπαιδευμένο νευρωνικό δίκτυο είναι ένα αποδοτικό εργαλείο για τον εντοπισμό των μεταβλητών εισόδων που επηρεάζουν σημαντικά οποιαδήποτε από τις δυνατές εξόδους καθώς και αυτών με την μικρότερη επιρροή στο τελικό αποτέλεσμα. Οι εισοδοί που αποδεικνύεται ότι δεν επηρεάζουν την έξοδο μπορούν να αφαιρεθούν ώστε να μειωθεί η διάσταση του προβλήματος με αποτέλεσμα να μειωθεί η πολυπλοκότητα του προβλήματος μας.

### 5.2 Ανάλυση ευαισθησίας (Θεωρητικό υπόβαθρο)

Παρακάτω αναλύουμε την τεχνική που αναπτύχθηκε για νευρωνικό δίκτυο προσοτροφodότησης .

Για το νευρωνικό δίκτυο τριών επιπέδων έχουμε:

- $\vec{z} = (z_1, \dots, z_i, \dots, z_I)$
- $\vec{y} = (y_1, \dots, y_j, \dots, y_J)$
- $\vec{o} = (o_1, \dots, o_k, \dots, o_K)$

που εκπροσωπούν το επίπεδο εισόδου, την έξοδο του κρυμμένου επιπέδου και το επίπεδο εξόδου αντίστοιχα. Θεωρούμε το ζεύγος εκπαίδευσης  $p = (\vec{z}^{(p)}, t^{(p)})$  όπου  $t = (t_1, \dots, t_k, \dots, t_K)$  οι επιθυμητές εξοδοί. Για δεδομένο ζεύγος  $p$  ορίζεται η ευαισθησία  $s_{ki}^{(p)}$  μιας εκπαιδευμένης εξόδου ως προς μία είσοδο  $z_i$  ως εξής:

$$s_{ki}^{(p)} = \frac{\partial o_k}{\partial z_i} = o'_k \sum_{j=1}^J w_{kj} \frac{\partial y_j}{\partial z_i} = o'_k \sum_{j=1}^J w_{kj} y'_j v_{ji} \quad (1)$$

όπου:  $y_j$  η έξοδος του  $j$ -κρυφού νευρώνα,  $o'_k$  η παράγωγος της συνάρτησης ενεργοποίησης του επιπέδου εξόδου:

$$o_k = f\left(\sum_{j=1}^J w_{kj} y_j + b_k\right) \quad (2)$$

και  $y_j$  η παράγωγος της συνάρτησης ενεργοποίησης του κρυφού επιπέδου:

$$y_j = f\left(\sum_{i=1}^J v_{ji} z_i + b_j\right) \quad (3)$$

Επίσης  $w_{kj}$  είναι το βάρος ανάμεσα στον κρυφό νευρώνα  $y_j$  και την έξοδο  $o_k$ ,  $b_k$  το bias της k-εξόδου,  $v_{ji}$  το βάρος ανάμεσα στην είσοδο  $z_i$  και τον κρυφό νευρώνα  $y_j$  και  $b_j$  το bias του κρυφού νευρώνα  $j$ . Η εξίσωση (1) παριστάνει την ευαισθησία της εξόδου  $o_k$  σε σχέση με την είσοδο  $z_i$  για το ζεύγος  $p$ . Για το ζεύγος αυτό, ο πίνακας ευαισθησίας  $s^{(p)}$  που αποτελείται από στοιχεία  $s_{ki}^{(p)}$  ορίζεται ως εξής:

$$S^{(p)} = O' W Y' V \quad (4)$$

όπου  $W(K \times J)$  και  $V(J \times I)$  είναι οι πίνακες βαρών επιπέδου εξόδου και κρυφού επιπέδου αντίστοιχα και  $O'(K \times K)$  και  $Y'(J \times J)$  ορίζονται ως:

$$\begin{aligned} O' &= \text{diag}(o'_1, \dots, o'_K) \\ Y' &= \text{diag}(y'_1, \dots, y'_J) \end{aligned} \quad (5)$$

Η εξίσωση (4) δίνει τον πίνακα ευαισθησίας για ένα συγκεκριμένο ζεύγος εκπαίδευσης  $p$ . Όμως, κάθε ζεύγος  $p$  παράγει έναν πίνακα ευαισθησίας  $s^{(p)}$ . Για να εφαρμοστεί η ανάλυση ευαισθησίας ο πίνακας ευαισθησίας  $s^{(p)}$  πρέπει να υπολογιστεί για ολόκληρο το σύνολο δεδομένων εισόδου.

Ο μέσος τετραγωνικός πίνακας ευαισθησίας  $S_{avg}$  ορίζεται ως εξής:

$$S_{ki,avg} = \sqrt{\frac{\sum_1^p [S_{ki}^p]^2}{p}} \quad (6)$$

Για να γίνει σωστή σύγκριση ανάμεσα στις εισόδους, είναι απαραίτητο να κανονικοποιηθούν εισοδοι και έξοδοι στην ίδια κλίμακα. Αυτό γίνεται με τον τύπο:

$$S_{ki,avg} = S_{ki,avg} \frac{(\max_{p=1,...,P} \{z_i^{(p)}\} - \min_{p=1,...,P} \{z_i^{(p)}\})}{(\max_{p=1,...,P} \{o_k^{(p)}\} - \min_{p=1,...,P} \{o_k^{(p)}\})} \quad (7)$$

Χρησιμοποιώντας την σχέση (6) η διάσταση του διανύσματος εισόδου  $\vec{z}$  μπορεί να μειωθεί αφαιρώντας εκείνες τις εισόδους  $\vec{z}_i$  που δεν επηρεάζουν πολύ τις εξόδους. Ορίζεται η σπουδαιότητα  $\Phi_i$  της εισόδου  $\vec{z}_i$  για όλες τις εξόδους ως:

$$\Phi_i = \max_{k=1,...,K} \{S_{ki,avg}\} \quad (8)$$

Έπειτα, τα στοιχεία του διανύσματος  $\vec{\Phi}$  κατανέμονται σε αύξουσα σειρά έτσι ώστε τα πρώτα στοιχεία του να αντιστοιχούν στις εισόδους με την μικρότερη επίδραση στις εξόδους.

### 5.3 Αλγόριθμος μείωσης δεδομένων εισόδου

Παρακάτω παρουσιάζονται δύο αλγόριθμοι που στοχεύουν στη μείωση χαρακτηριστικών εισόδων οι οποίοι έχουν κοινά τα τέσσερα πρώτα βήματα και διαφοροποιούνται στην συνέχεια. (Ο αλγόριθμος που θα χρησιμοποιηθεί στην δική μας πειραματική μελέτη είναι ο δεύτερος).

#### 5.3.1 Αλγόριθμος 1

*Βήμα 1:* Εκπαίδευση νευρωνικού δικτύου με τα αρχικά δεδομένα εισόδου.

*Βήμα 2:* Υπολογισμός όλων των ευαισθησιών μεταξύ εισόδων και εξόδων για κάθε ζεύγος εκπαίδευσης χρησιμοποιώντας την (4).

*Βήμα 3:* Υπολογισμός του πίνακα ευαισθησίας για όλες τις παραμέτρους εκπαίδευσης με χρήση της (6).

*Βήμα 4:* Υπολογισμός της σπουδαιότητας της κάθε εισόδου για όλες τις εξόδους μέσω της (8).

*Βήμα 5:* Αφαίρεση των παραμέτρων εισόδου με την μικρότερη σπουδαιότητα.

*Βήμα 6:* Επανεκπαίδευση του νευρωνικού με τις εναπομένουσες εισόδους.

*Βήμα 7:* Επανάληψη των βημάτων 1 έως 6 έως ότου να μην υπάρχουν άλλες εισοδοί με μικρή επιρροή στην έξοδο

#### 5.3.2 Αλγόριθμος 2

*Βήμα 1:* Εκπαίδευση νευρωνικού δικτύου με τα αρχικά δεδομένα εισόδου.

*Βήμα 2:* Υπολογισμός όλων των ευαισθησιών μεταξύ εισόδων και εξόδων για κάθε ζεύγος εκπαίδευσης χρησιμοποιώντας την (4).

*Βήμα 3:* Υπολογισμός του πίνακα ευαισθησίας για όλες τις παραμέτρους εκπαίδευσης με χρήση της (6).

*Βήμα 4:* Υπολογισμός της σπουδαιότητας της κάθε εισόδου για όλες τις εξόδους μέσω της (8).

*Βήμα 5:* Έλεγχος απόδοσης του νευρωνικού χρησιμοποιώντας μόνο τα σπουδαιότερα χαρακτηριστικά [5].



## 5.4 Υλοποίηση αλγορίθμου μείωσης δεδομένων εισόδου

Θα χρησιμοποιήσουμε τον δεύτερο αλγόριθμο που παρουσιάστηκε στην προηγούμενη παράγραφο.

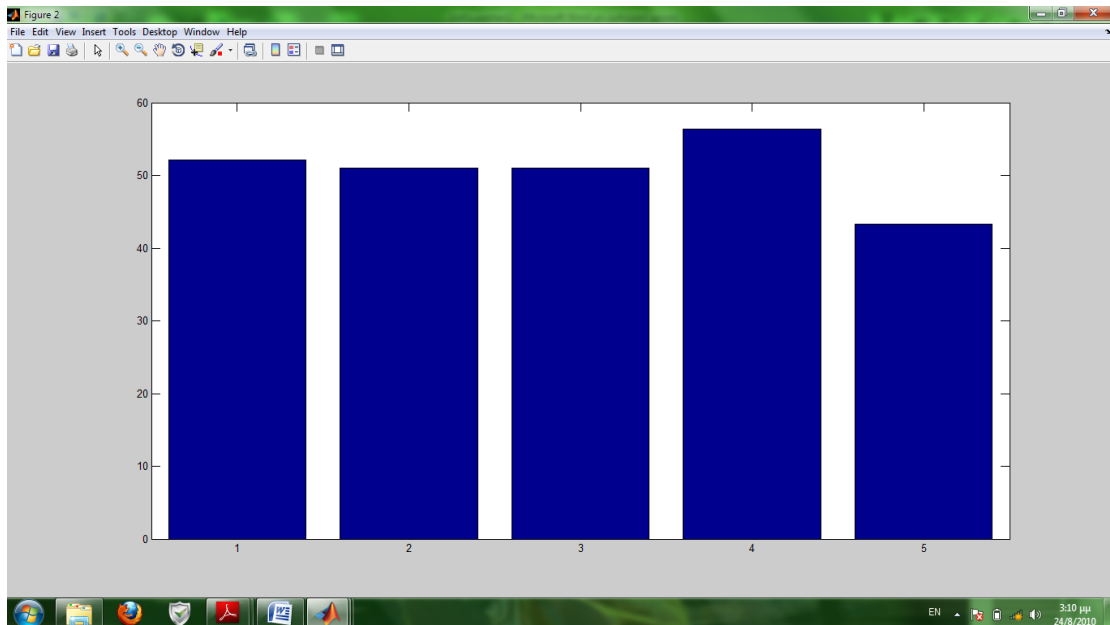
- *Βήμα 1:* Εκπαίδευση νευρωνικού δικτύου με τα αρχικά δεδομένα εισόδου. Εκπαιδεύουμε ένα feed- forward νευρωνικό δίκτυο με δεδομένα εισόδου μόνο τα 17 χαρακτηριστικά προσώπου, μιας και αυτών των χαρακτηριστικών τη σπουδαιότητα στην έξοδο θέλουμε να υπολογίσουμε. Επειδή οι παράμετροι (πίνακες βαρών) του εκπαιδευμένου δικτύου θα χρησιμοποιηθούν στη συνέχεια για την ανάλυση ευαισθησίας , όσο πιο αποδοτικό είναι το δίκτυο τόσο πιο αξιόπιστη και η ανάλυση. Έτσι εκπαιδεύουμε κάποια δίκτυα με διαφορετικές παραμέτρους εκπαίδευσης και αφού τα αξιολογήσουμε, επιλέγουμε το αποδοτικότερο για να συνεχίσουμε την ανάλυσή μας. Παρακάτω τα αποτελέσματα των πειραμάτων (πίνακας 5.1).

α/α	Νευρώνες στο κρυφό επίπεδο	Εποχές εκπαίδευσης	Ρυθμός μάθησης	Επιτυχία (%)	Μέσο τετραγωνικό σφάλμα
1	20	1000	0,15	50,74	0,132
2	20	2000	0,15	51,93	0,126
3	30	2000	0,15	51,90	0,124
4	20	2000	0,20	49,73	0,128
5	15	2000	0,15	50,00	0,125
6	25	2000	0,15	53,41	0,125

*Πίνακας 5.1 : Διάφορες αρχιτεκτονικές και αποτελέσματα ελέγχου για νευρωνικό δίκτυο με είσοδο μόνο τα χαρακτηριστικά προσώπου.*

Όπως το περιμέναμε ένα νευρωνικό δίκτυο που προσπαθεί να προβλέψει την συναισθηματική κατάσταση του ομιλούντος με είσοδο μόνο τα χαρακτηριστικά προσώπου , “ μάλλον ” δεν τα καταφέρνει και τόσο καλά. Όλα τα παραπάνω δίκτυα έχουν επιτυχία γύρω στο 50% , ποσοστό πολύ κακό. Επίσης δεν κάναμε πειράματα με μεγαλύτερο αριθμό εποχών εκπαίδευσης διότι μετά τις περίπου 1500 εποχές, ούτε το μέσο τετραγωνικό σφάλμα μειώνεται. Παρ’ όλα αυτά, ο αρχικός μας σκοπός δεν ήταν να φτιάξουμε ένα δίκτυο που να λύνει το αρχικό μας πρόβλημα, αλλά ένα δίκτυο, του οποίου τους πίνακες βαρών θα χρησιμοποιήσουμε για την ανάλυση ευαισθησίας. Άρα συνεχίζουμε στο δεύτερο βήμα του αλγορίθμου, έχοντας όμως υπ’ όψιν μας ότι οι τιμές που θα λάβουμε θα έχουν ίσως μεγάλο σφάλμα. Επειδή μας ενδιαφέρει η ταξινόμηση των τιμών της σπουδαιότητας της κάθε εισόδου και όχι η ακριβής αποτίμηση τους , σκοπεύουμε να εκμεταλλευτούμε τις μεγάλες διαφορές που τυχόν προκύψουν ώστε το σφάλμα να μην υπεισέρχεται. Στα παρακάτω βήματα θα χρησιμοποιηθούν για τον υπολογισμό των αντίστοιχων μεγεθών οι

πίνακες βαρών και κατωφλίων του εκπαιδευμένου δικτύου με  $\alpha/\alpha$  6 , το οποίο είναι και το αποδοτικότερο.



Σχήμα 5.1: Ακρίβεια του νευρωνικού δικτύου με  $\alpha/\alpha$  6 του πίνακα 5.1.

- Βήμα 2:** Υπολογισμός όλων των ευαισθησιών μεταξύ εισόδων και εξόδων για κάθε ζεύγος εκπαίδευσης.

Χρησιμοποιώντας την εξίσωση (4) της προηγούμενης παραγράφου υπολογίζουμε τις ευαισθησίες για κάθε ζεύγος εκπαίδευσης και προκύπτει ο πίνακας  $s^{(p)}$ . Λόγω του μεγέθους του (21875x17) αυτός ο πίνακας παραλείπεται. Εξάλλου δεν παρέχει καμία ευανάγνωστη πληροφορία που πρέπει να αποτυπωθεί.
- Βήμα 3:** Υπολογισμός του πίνακα ευαισθησίας για όλες τις παραμέτρους εκπαίδευσης .

Χρησιμοποιώντας τις εξισώσεις (6) και (7) προκύπτει ο παρακάτω πίνακας  $S_{ik,avg}$  ( πίνακας 5.2) όπου από 1 έως 17 είναι τα fars σε αντιστοιχία με τον πίνακα 3.2 και από 1 έως 5 οι συναισθηματικές καταστάσεις, όπως παρουσιάστηκαν στην παράγραφο 3.2. ( Για διευκόλυνση ανάγνωσης και αποτύπωσης ο πίνακας 5.2 ( $S_{ik,avg}$ ) είναι ο ανεστραμμένος που υπολογίζεται από την εξίσωση (7) δηλ.  $S_{ki,avg}$ .)

Από τις τιμές αυτές γίνεται άμεσα αντιληπτό ότι κάποια χαρακτηριστικά επηρεάζουν πολύ κάποιες εξόδους , κάποια επηρεάζουν αρκετά όλες τις εξόδους και κάποια έχουν μικρή επιρροή στην απόκριση του δικτύου. Ακόμα και μόνο μια έξοδος να επηρεάζεται πολύ από κάποιο far, το far αυτό θεωρείται σημαντικό στην ανάλυσή μας. Αυτό γίνεται αντιληπτό στο

επόμενο βήμα όπου ταξινομούμε την σπουδαιότητα του κάθε far, δηλαδή την μέγιστη τιμή της ευαισθησίας που παρουσιάζεται σε κάθε far, και κρίνουμε ποια χαρακτηριστικά θα συμπεριλάβουμε στην είσοδο με βάση αυτήν την ταξινόμηση.

i\k	1	2	3	4	5
1	0,484	0,410	0,277	0,304	0,336
2	0,483	0,391	0,497	0,358	0,365
3	0,664	0,617	0,592	0,409	0,473
4	0,262	0,540	0,225	0,328	0,316
5	0,485	0,340	0,298	0,456	0,253
6	0,339	0,512	0,379	0,293	0,251
7	0,459	0,691	0,458	0,511	0,507
8	0,529	0,598	1,649	1,383	0,828
9	0,743	0,327	0,451	0,326	0,541
10	0,380	0,730	1,356	1,727	0,875
11	0,790	0,567	0,640	0,607	0,666
12	0,521	0,294	0,393	0,798	0,303
13	0,602	0,678	0,529	0,760	0,389
14	0,950	0,553	0,403	0,482	0,412
15	1,406	1,021	1,214	1,058	0,649
16	0,671	0,833	0,862	1,058	0,542
17	0,707	0,553	0,496	0,669	0,490

Πίνακας 5.2: Πίνακας ευαισθησίας για όλα τα fars για κάθε έξοδο ( $S_{ik,avg}$ )

- **Βήμα 4:** Υπολογισμός της σπουδαιότητας της κάθε εισόδου για όλες τις εξόδους.

Αφού υπολογίσουμε τη σπουδαιότητα του κάθε far για όλες τις εξόδους μέσω της (8) προκύπτουν οι τιμές του πίνακα 5.3. Στον πίνακα συμπεριλαμβάνουμε και τα ονόματα του κάθε far στην δεξιά στήλη ώστε να έχουμε και λογική εποπτεία των αποτελεσμάτων μας. Όπως έχουμε εξηγήσει και νωρίτερα, ουσιαστικά μας ενδιαφέρει η ποιοτική κατηγοριοποίηση των χαρακτηριστικών σε σημαντικά και όχι σημαντικά. Με βάση αυτήν την παραδοχή μπορούμε να διακρίνουμε τρεις ομάδες σπουδαιότητας. Στην πρώτη ομάδα ανήκουν τα χαρακτηριστικά με σπουδαιότητα περίπου 0,5, στην δεύτερη με σπουδαιότητα περίπου 0,7 και στην τρίτη και τελευταία, όσα έχουν σπουδαιότητα περίπου 1 και πάνω. Η τρίτη ομάδα, στην οποία ανήκουν τα σημαντικά fars και είναι αυτή που θα μας απασχολήσει στη συνέχεια, αποτελείται από τα παρακάτω fars:

α/α	Όνομα
14	F38_squeeze_right_eyebrow
16	F59_raise_left_outer_cornerlip inv
15	Fx_wrinkles_between_eyebrows inv
8	F32_raise_right_inner_eyebrow
10	F34_raise_right_medium_eyebrow

Παρατηρούμε ότι σημαντικό ρόλο παίζουν τελικά, η συμπεριφορά του δεξιού φρυδιού , οι κίνηση των μυών ανάμεσα στα φρύδια και το σήκωμα του άκρου μέρους του αριστερού χειλιού. Αυτό το αποτέλεσμα δημιουργεί το εύλογο ερώτημα για την καθολικότητα της σπουδαιότητας των faps. Μήπως αυτά είναι τα σημαντικά χαρακτηριστικά για τον συγκεκριμένο ομιλητή; Μήπως είναι σημαντικά για έναν , για παράδειγμα, αριστερόχειρα; Είναι τελικά αυτά τα χαρακτηριστικά , σημαντικά για όλους; Το παραπάνω ερώτημα πρέπει να μελετηθεί μελλοντικά , διότι μπορεί το δικό μας σύνολο δεδομένων να αφορά μια ομιλία, αλλά το σύστημα που θέλουμε να κατασκευάσουμε θα πρέπει να είναι ικανό να προβλέπει με επιτυχία τις συναισθηματικές καταστάσεις ενός τυχαίου ομιλητή.

Εφόσον έχουμε , πλέον, ολοκληρώσει την ταξινόμηση, με βάση τη σπουδαιότητα, των χαρακτηριστικών του προσώπου, μπορούμε να προχωρήσουμε στο τελευταίο βήμα του αλγορίθμου μας, για να δούμε αν μπορούμε να μειώσουμε την πολυπλοκότητα του αρχικού μας δικτύου.

i	$\Phi_i$	Όνομα fap
1	0,484	F3_open_jaw
5	0,484	F19_21_close_left_eye inv
2	0,497	F4_lower_top_midlip
6	0,511	F20_22_close_right_eye inv
4	0,540	Fx_widening_mouth
3	0,663	F5_raise_bottom_midlip inv
7	0,690	F31_raise_left_inner_eyebrow
17	0,706	F60_raise_right_outer_cornerlip
9	0,742	F33_raise_left_medium_eyebrow
13	0,760	F37_squeeze_left_eyebrow
11	0,789	F35_raise_left_outer_eyebrow
12	0,797	F36_raise_right_outer_eyebrow
14	0,949	F38_squeeze_right_eyebrow
16	1,057	F59_raise_left_outer_cornerlip inv
15	1,405	Fx_wrinkles_between_eyebrows inv
8	1,648	F32_raise_right_inner_eyebrow
10	1,726	F34_raise_right_medium_eyebrow

*Πίνακας 5.3: Πίνακας σπουδαιότητας του κάθε fap*

- **Βήμα 5:** Έλεγχος απόδοσης του νευρωνικού χρησιμοποιώντας μόνο τα σπουδαιότερα χαρακτηριστικά.

Σε αυτό το βήμα θα κινηθούμε όπως και στο προηγούμενο κεφάλαιο, όταν αρχίσαμε να “κόβουμε” συσχετισμένα fars. Θα συμπεριλάβουμε στις εισόδους των δικτύων που θα ελέγξουμε, τους συνδυασμούς των σημαντικών fars που μας ενδιαφέρουν καθώς και όλα τα χαρακτηριστικά της φωνής. Τα δίκτυα θα έχουν την ίδια αρχιτεκτονική με το αποδοτικότερο του κεφαλαίου 3 (1 κρυφό επίπεδο 50 νευρώνων και 2000 εποχές εκπαίδευσης), έτσι ώστε να μπορεί να γίνει άμεση σύγκριση απόδοσης. Στην συνέχεια, για τον συνδυασμό εισόδων με καλύτερη απόδοση, θα εξετάσουμε την “ελαφρύτερη” αρχιτεκτονική δικτύου που μπορούμε να πετύχουμε, με αυτόν τον συνδυασμό εισόδων χωρίς μεγάλες απώλειες.

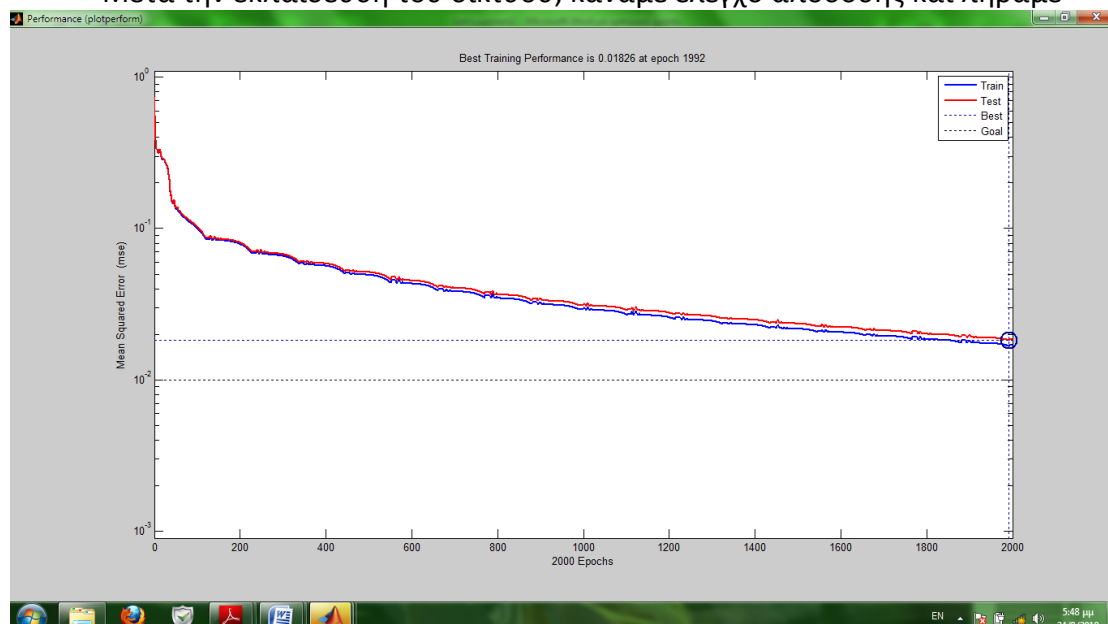
Τώρα, θα πρέπει να καταλήξουμε στους συνδυασμούς των σημαντικών fars, που θα χρησιμοποιήσουμε. Παρατηρώντας τα αποτελέσματα που πήραμε στην σπουδαιότητα του κάθε far στον πίνακα 5.3 και αναλογιζόμενοι ότι θα πρέπει να δώσουμε ιδιαίτερη βάση στις διαφορές των τιμών, όπως είπαμε, καταλήγουμε στους εξής 2 συνδυασμούς.

1. Όλα τα χαρακτηριστικά φωνής και τα 5 σημαντικότερα fars ( $\Phi > 0,9$ ).
2. Όλα τα χαρακτηριστικά φωνής και τα 3 σημαντικότερα fars ( $\Phi > 1,4$ ).

Το να χρησιμοποιήσουμε σε κάποιο συνδυασμό περισσότερα fars δεν έχει κάποια αξία για την ανάλυση μας, διότι όλα τα υπόλοιπα έχουν περίπου την ίδια σπουδαιότητα και επομένως είναι ασήμαντα σε σύγκριση με τα 5 πιο σημαντικά.

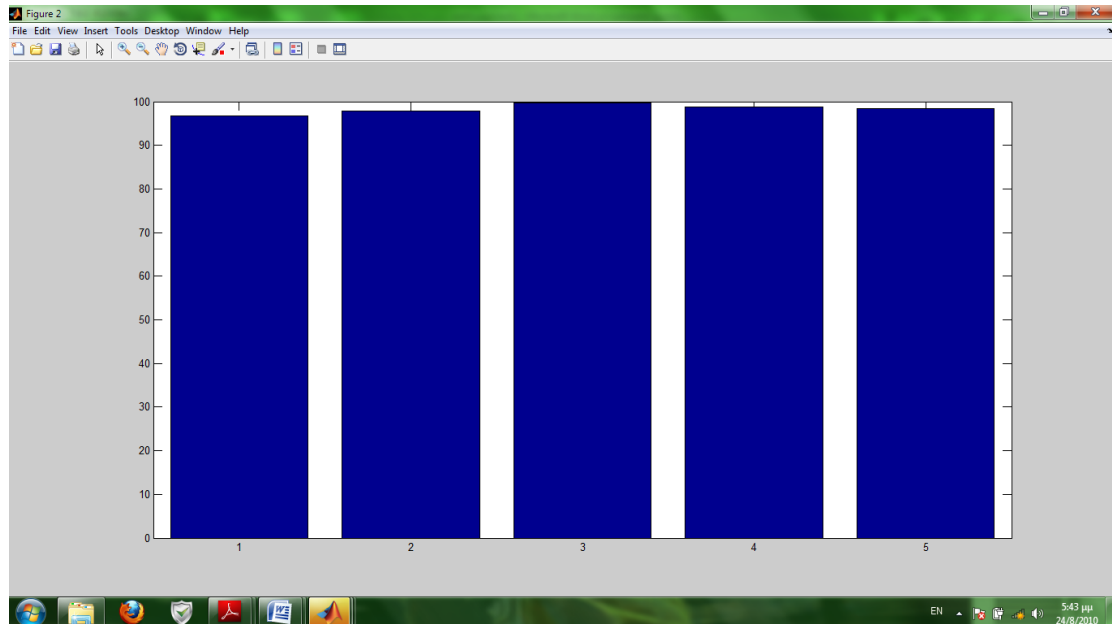
1. Εκπαίδευση και έλεγχος απόδοσης feed- forward νευρωνικού δικτύου με είσοδο όλα τα χαρακτηριστικά φωνής και τα 5 σημαντικότερα fars.

Μετά την εκπαίδευση του δικτύου, κάναμε έλεγχο απόδοσης και πήραμε



*Σχήμα 5.2: Μέσο τετραγωνικό σφάλμα σε συνάρτηση με τις εποχές εκπαίδευσης του νευρωνικού δικτύου με είσοδο τα χαρακτηριστικά φωνής και τα 5 πιο σημαντικά fars*

98,3% επιτυχία και το μέσο τετραγωνικό σφάλμα μετά από 2000 εποχές κατέληξε στην τιμή 0,0168. Καταφέραμε δηλαδή , όπως και με το δίκτυο που περιείχε τα χαρακτηριστικά φωνής και τα τέσσερα ασυσχέτιστα faps, να πετύχουμε μία απόδοση ισάξια του αρχικού βέλτιστου δικτύου που είχε είσοδο με όλα τα χαρακτηριστικά.



*Σχήμα 5.3: Ακρίβεια του νευρωνικού δικτύου με είσοδο τα χαρακτηριστικά φωνής και τα 5 πιο σημαντικά faps*

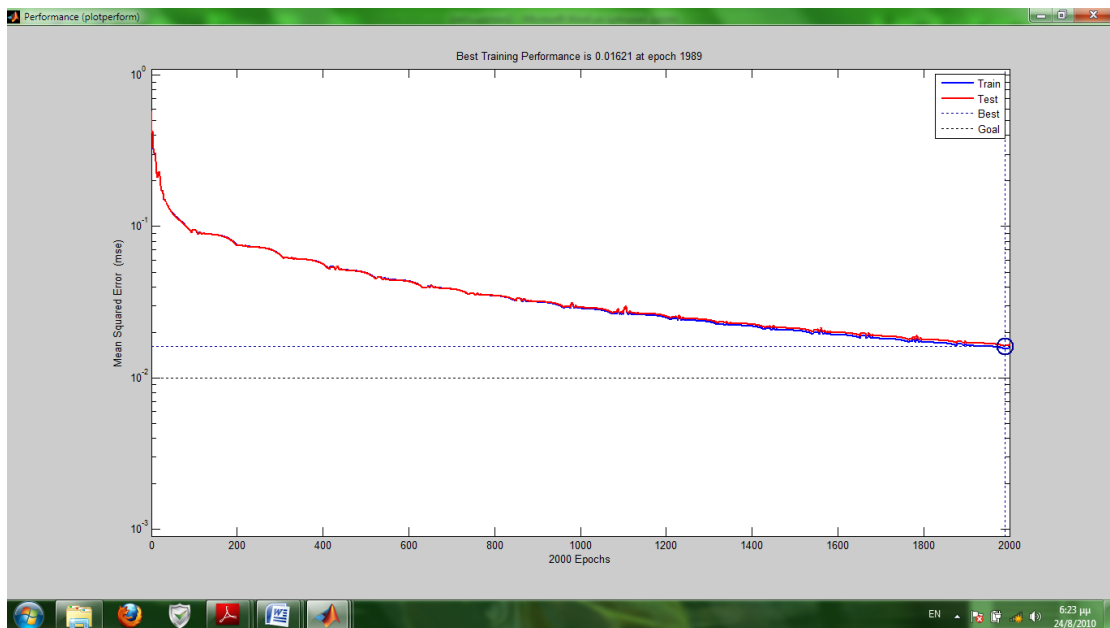
Όσον αφορά την ακρίβεια , παρατηρούμε ότι το δίκτυο αυτό έχει την ίδια συμπεριφορά με το αρχικό βέλτιστο δίκτυο. Αναγνωρίζει , δηλαδή, καλά όλες τις κλάσεις με μόνη “παραφωνία ” την κλάση 1.

Τέλος, αξίζει να αναφερθεί ότι, προς επαλήθευση της ανάλυσης ευαισθησίας που κάναμε , εκπαιδεύσαμε και ελέγξαμε ένα ίδιο δίκτυο με το παραπάνω, μόνο που αντί να χρησιμοποιήσουμε στην είσοδο τα πέντε σημαντικότερα faps , χρησιμοποιήσαμε τα πέντε πιο ασήμαντα. Αυτό το δίκτυο είχε τελικά 97,4% επιτυχία και 0,018 μέσο τετραγωνικό σφάλμα. Μια διαφορά δηλαδή περίπου 1% σε επιτυχία σε σχέση με τον “σοβαρό” αντίπαλό του. Η διαφορά αυτή δεν είναι πολύ μεγάλη αλλά μας επιβεβαιώνει, ότι η ανάλυση μας είναι σωστή.

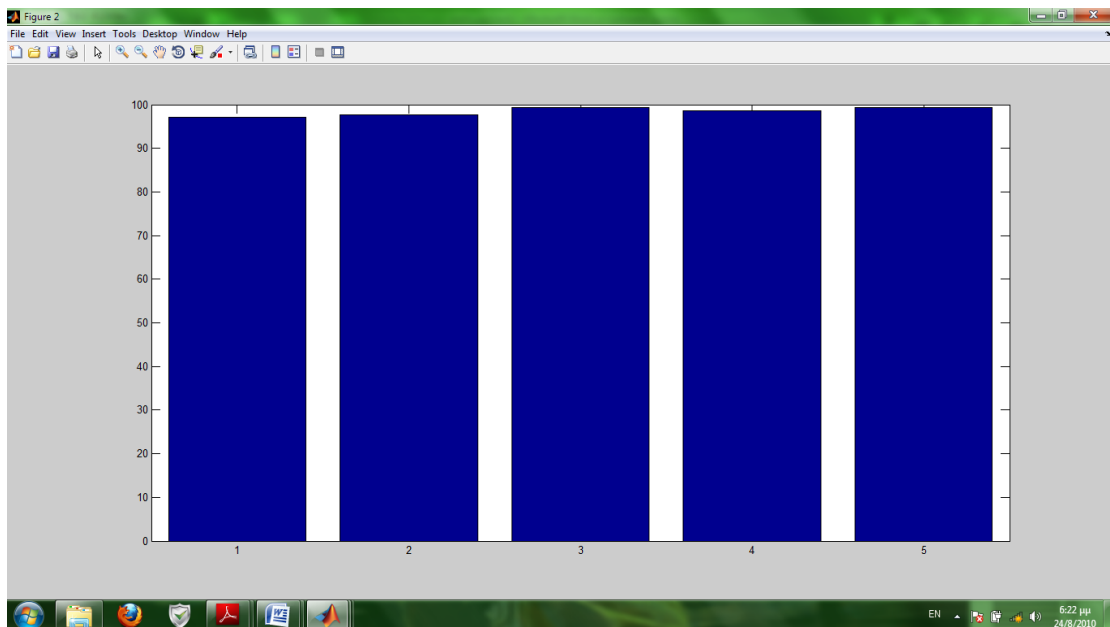
2. Εκπαίδευση και έλεγχος απόδοσης feed- forward νευρωνικού δικτύου με είσοδο όλα τα χαρακτηριστικά φωνής και τα 3 σημαντικότερα faps.

Μετά την εκπαίδευση και τον έλεγχο σε αυτό το δίκτυο πετύχαμε 98,4% επιτυχία και το μέσο τετραγωνικό σφάλμα έφτασε στο 0,0155. Και αυτό το αποτέλεσμα είναι εντυπωσιακό και εξηγείται , όπως αναφέραμε και στο προηγούμενο κεφάλαιο, από την μεγάλη υπολογιστική ικανότητα του δικτύου με τους 50 νευρώνες στο κρυφό επίπεδο, για πρόβλεψη προβλήματος με διάνυσμα εισόδου 36 μεταβλητών. Έχοντας στα χέρια μας

αυτά τα αποτελέσματα θα επιχειρήσουμε στην επόμενη παράγραφο να μειώσουμε την πολυπλοκότητα του δικτύου και θα αναζητήσουμε ποια είναι εκείνη η αρχιτεκτονική που μας δίνει την μεγαλύτερη επιτυχία.



Σχήμα 5.4: Μέσο τετραγωνικό σφάλμα σε συνάρτηση με τις εποχές εκπαίδευσης του νευρωνικού δικτύου με είσοδο τα χαρακτηριστικά φωνής και τα 3 πιο σημαντικά *faps*



Σχήμα 5.5: Ακρίβεια του νευρωνικού δικτύου με είσοδο τα χαρακτηριστικά φωνής και τα 3 πιο σημαντικά *faps*

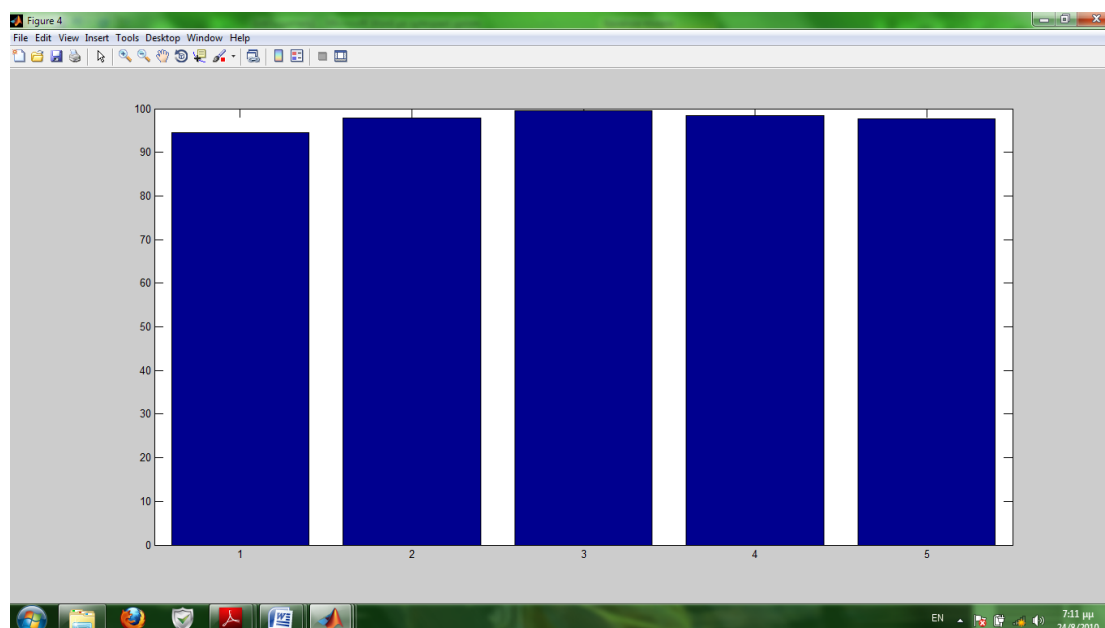
## 5.5 Αναζήτηση βέλτιστης αρχιτεκτονικής

Συγκρίνουμε διάφορες αρχιτεκτονικές με τα 32 χαρακτηριστικά φωνής και τα 3 πιο σημαντικά fars. Παρατίθεται πίνακας(Πίνακας 5.4) με τις παραμέτρους του κάθε δικτύου που δοκιμάσαμε.

α/α	Νευρώνες στο κρυφό επίπεδο	Εποχές εκπαίδευσης	Ρυθμός μάθησης	Επιτυχία (%)	Μέσο τετραγωνικό σφάλμα
1	25	2000	0,15	96,2	0,0268
2	25	3000	0,15	96,68	0,0261
3	30	2000	0,15	96,18	0,0242
4	35	2000	0,15	96,54	0,0197
5	35	2000	0,20	97,07	0,0192
6	35	2600	0,20	97,55	0,0178

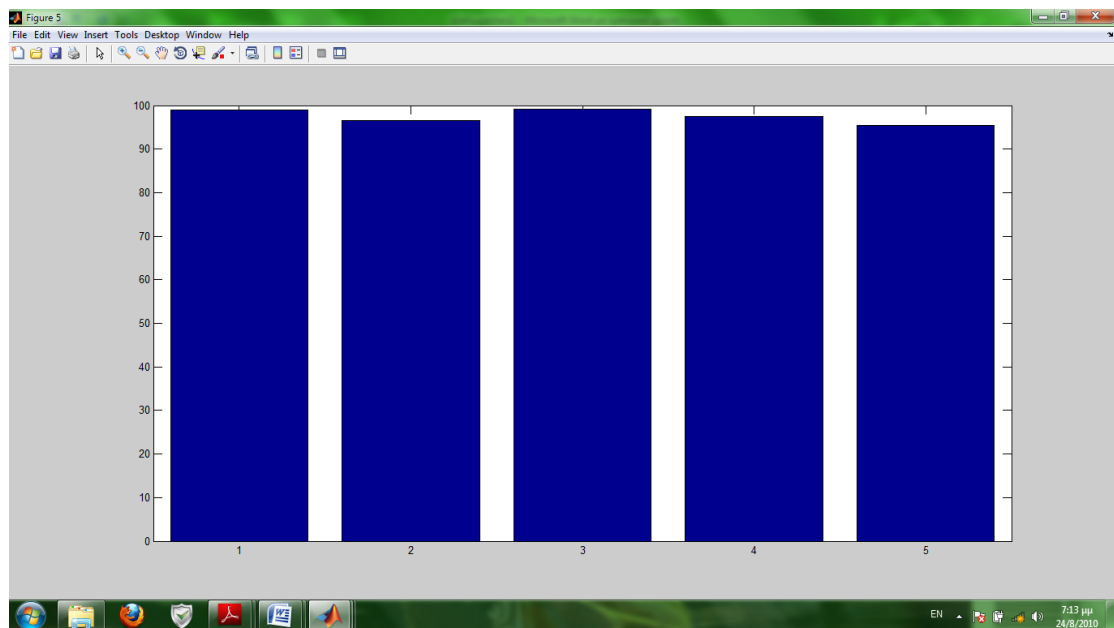
Πίνακας 5.4

Με βάση τα παραπάνω αποτελέσματα βλέπουμε ότι πετυχαίνουμε επιτυχία 97,55% με τις παραμέτρους του δικτύου με τον αύξοντα αριθμό 6. Κόβοντας , δηλαδή , 15 νευρώνες από το κρυφό επίπεδο πετυχαίνουμε κατηγοριοποίηση των δεδομένων με λιγότερο από 1% σφάλμα στην επιτυχία. Βέβαια σε σχέση με το αρχικό δίκτυο σε αυτήν την περίπτωση αυξήσαμε τον αριθμό των εποχών. Το αποτέλεσμα μας είναι αρκετά καλό και αποδεικνύει ότι η ανάλυση ευαισθησίας που κάναμε , μας οδήγησε στο να πλησιάσουμε τον αρχικό μας στόχο, να μειώσουμε την πολυπλοκότητα του προβλήματος μας χωρίς σημαντικές απώλειες.



Σχήμα 5.6: Ακρίβεια του νευρωνικού δικτύου με α/α 6 του πίνακα 5.4





Σχήμα 5.7: Ανάκληση του νευρωνικού δικτύου με α/α 6 του πίνακα 5.4.

### 5.6 Εκπαίδευση και έλεγχος νευρωνικού δικτύου με εισόδους επιλεγμένες με βάση την ανάλυση ευαισθησίας σε συνδυασμό με την στατιστική συσχέτιση των faps.

Με βάση τα αποτελέσματα του προηγούμενου κεφαλαίου, που αφορούσε την συσχέτιση των faps, και την ανάλυση ευαισθησίας που είδαμε σε αυτό το κεφάλαιο, θα προσπαθήσουμε σε αυτήν την παράγραφο να κάνουμε συνδυασμούς εισόδων σε διαφορετικές αρχιτεκτονικές δικτύων προκειμένου να πετύχουμε υψηλή απόδοση.

Όσον αφορά τους τελικούς συνδυασμούς εισόδων που δοκιμάσαμε και πως καταλήξαμε σε αυτούς, τους παρουσιάζουμε παρακάτω. Να σημειωθεί ότι τα δίκτυα για τους παρακάτω συνδυασμούς εισόδων αφορούν αρχιτεκτονική 50 νευρώνων στο κρυφό επίπεδο και 2000 εποχές εκπαίδευσης, όπως το αρχικό βέλτιστο δίκτυο, ώστε να γίνει άμεση σύγκριση απόδοσης. Επίσης όλα τα δίκτυα προφανώς περιέχουν τα χαρακτηριστικά φωνής.

Θυμίζουμε ότι τα 4 ασυσχέτιστα faps είναι τα

α/α	Όνομα
4	Fx_widening_mouth
13	F37_squeeze_left_eyebrow
14	F38_squeeze_right_eyebrow
15	Fx_wrinkles_between_eyebrows inv

Και τα 5 σημαντικότερα τα

14	F38_squeeze_right_eyebrow
16	F59_raise_left_outer_cornerlip inv
15	Fx_wrinkles_between_eyebrows inv
8	F32_raise_right_inner_eyebrow
10	F34_raise_right_medium_eyebrow

Παρατηρούμε ότι τα χαρακτηριστικά Fx\_wrinkles\_between\_eyebrows inv και F38\_squeeze\_right\_eyebrow εμφανίζονται και στις 2 κατηγορίες, επομένως καταλήγουμε ότι πρέπει να κάνουμε συνδυασμούς με τα χαρακτηριστικά 4, 13, 14, 15, 16, 8 και 10 δηλ. :

α/α	Όνομα
4	Fx_widening_mouth
13	F37_squeeze_left_eyebrow
14	F38_squeeze_right_eyebrow
15	Fx_wrinkles_between_eyebrows inv
16	F59_raise_left_outer_cornerlip inv
8	F32_raise_right_inner_eyebrow
10	F34_raise_right_medium_eyebrow

1. 1<sup>ος</sup> συνδυασμός:

Παρατηρούμε ότι στο προηγούμενο κεφάλαιο είχε προκύψει, το χαρακτηριστικό 8 (F32\_raise\_right\_inner\_eyebrow) να είναι πολύ ισχυρά συσχετισμένο με το χαρακτηριστικό 10 (F34\_raise\_right\_medium\_eyebrow). Επομένως μπορούμε να παραλείψουμε ένα από τα δύο και επιλέγουμε να παραλείψουμε το χαρακτηριστικό 8, ως το λιγότερο σημαντικό. Έτσι με είσοδο τα χαρακτηριστικά 4, 13, 14, 15, 16 και 10 και αφού εκπαιδεύσουμε το δίκτυο των 50 νευρώνων στο κρυφό επίπεδο για 2000 εποχές παίρνουμε μέσο τετραγωνικό σφάλμα 0,0206 και επιτυχία 97,76%.

2. 2<sup>ος</sup> συνδυασμός:

Μειώνοντας τις μεταβλητές εισόδου μας, σε αυτόν τον συνδυασμό παραλείπουμε ακόμα ένα far και επιλέγουμε το χαρακτηριστικό 16 ως το πιο ασήμαντο ,μη συσχετισμένο. Έτσι με είσοδο τα χαρακτηριστικά 4, 13, 14, 15 και 10 παίρνουμε μέσο τετραγωνικό σφάλμα 0,0175 και επιτυχία 98,26%.

3. 3<sup>ος</sup> συνδυασμός

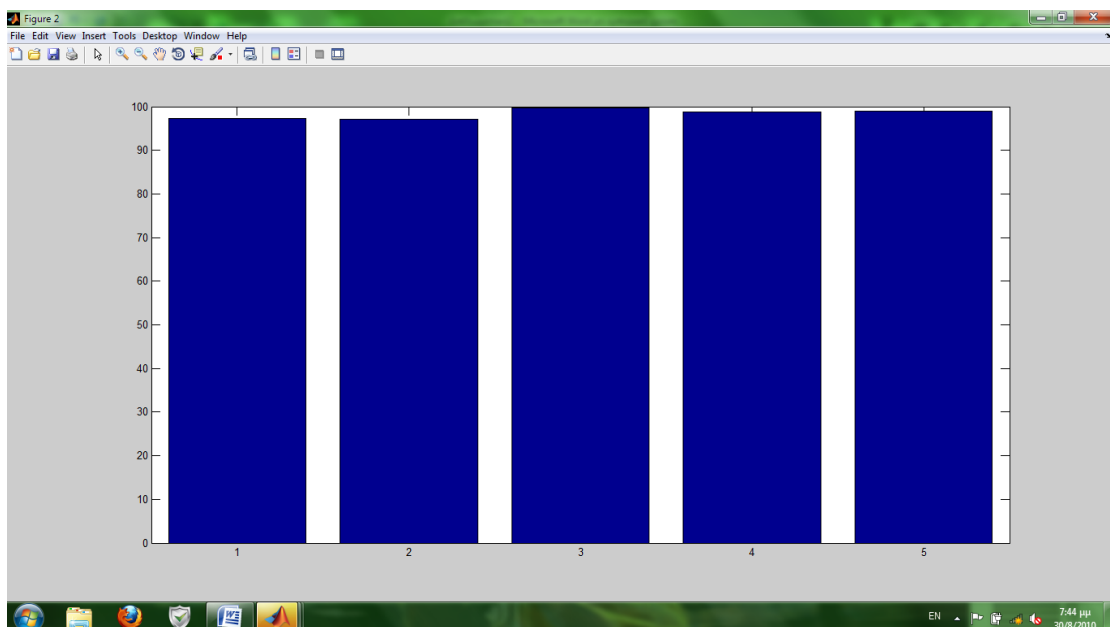
Σε αυτόν τον συνδυασμό παραλείπουμε το χαρακτηριστικό 4 ως το πιο ασήμαντο από τα ασυσχέτιστα. Με είσοδο, λοιπόν, τα χαρακτηριστικά 13,

14, 15, 16 και 10 παίρνουμε μέσο τετραγωνικό σφάλμα 0,0158 και επιτυχία 98,51%.

#### 4. 4<sup>ος</sup> συνδυασμός

Σε αυτόν τον συνδυασμό παραλείπουμε και το χαρακτηριστικό 4 ως το πιο ασήμαντο από τα ασυσχέτιστα, και το χαρακτηριστικό 16. Με είσοδο, λοιπόν, τα χαρακτηριστικά 13, 14, 15 και 10 παίρνουμε μέσο τετραγωνικό σφάλμα 0,0190 και επιτυχία 97,66%.

Παρατηρούμε ότι για τον τρίτο συνδυασμό έχουμε πολύ καλά αποτελέσματα και μάλιστα ισάξια με αυτά που είχαμε πετύχει με αντίστοιχη αρχιτεκτονική με εισόδους μόνο τα 4 ασυσχέτιστα faps (mse = 0,0156 και επιτυχία = 98,51%) και μόνο τα 3 σημαντικότερα faps (mse = 0,0155 και επιτυχία = 98,4%).



Σχήμα 5.8: Ακρίβεια του νευρωνικού δικτύου με εισόδους τον συνδυασμό 3

Ο συνδυασμός 3 περιέχει 5 faps, ενώ οι συνδυασμοί εισόδων μόνο με τα 4 ασυσχέτιστα και μόνο με τα 3 σημαντικότερα περιέχουν, όπως είναι προφανές, λιγότερα. Αυτομάτως, παρότι η απόδοση των δικτύων με τους παραπάνω συνδυασμούς εισόδων είναι ισάξια, τα δύο τελευταία καθίστανται καλύτερα ως λιγότερο πολύπλοκα, μιας και λιγότεροι εισοδοί σημαίνει πολύ λιγότεροι υπολογισμοί για την εξαγωγή του κάθε fap, όπως είδαμε στο κεφάλαιο 2, αλλά και λιγότεροι υπολογισμοί στην εκπαίδευση και έπειτα στην απόκριση του δικτύου αφού θα έχει λιγότερα συνναπτικά βάρη. Παρ' όλα αυτά θα επιχειρήσουμε να "ψαλιδίσουμε" νευρώνες σε δίκτυο με είσοδο τον συνδυασμό 3, μιας και σε περίπτωση που επιτύχουμε καλύτερη απόδοση με λιγότερους νευρώνες απ' ό,τι στα υπόλοιπα, όπως είναι φυσικό θα προτιμήσουμε παραπάνω εισόδους με λιγότερους νευρώνες.

### 5.6.1 Εκπαίδευση και έλεγχος νευρωνικού δικτύου με μεταβλητές εισόδου όλα τα χαρακτηριστικά φωνής, τα τρία σημαντικότερα ασυσχέτιστα χαρακτηριστικά προσώπου και τα δύο, μη συσχετισμένα μεταξύ τους, σημαντικότερα χαρακτηριστικά προσώπου.

Η καλύτερη απόδοση που έχουμε πετύχει ως τώρα, όσον αφορά "πριονισμένα" δίκτυα είναι αυτή του δικτύου με 35 νευρώνες στο κρυφό επίπεδο με είσοδο τα χαρακτηριστικά φωνής και τα 3 σημαντικότερα faps. ( $mse = 0,0178$  και επιτυχία 97,55). Το δίκτυο αυτό είχε εκπαιδευτεί για 2600 εποχές με προσαρμοσμένο ρυθμό μάθησης 0,2 και ορμή 0,9. Παρακάτω θα επιχειρήσουμε να ξεπεράσουμε αυτήν την επίδοση με λιγότερους νευρώνες, αλλιώς όπως είπαμε και παραπάνω δεν θα έχουμε όφελος. Στον πίνακα 5.5 παρουσιάζονται τα δίκτυα που εκπαιδεύτηκαν καθώς και τα αποτελέσματα που πήραμε από τον έλεγχο τους.

α/α	Νευρώνες στο κρυφό επίπεδο	Εποχές εκπαίδευσης	Ρυθμός μάθησης	Επιτυχία (%)	Μέσο τετραγωνικό σφάλμα
1	35	2000	0,15	96,66	0,0229
2	35	3000	0,15	97,89	0,1600
3	35	2000	0,20	96,54	0,0222
4	30	2000	0,15	97,76	0,0236
5	30	3000	0,15	96,5	0,0201
6	25	2000	0,15	96,82	0,2500
7	30	3000	0,20	97,39	0,0208

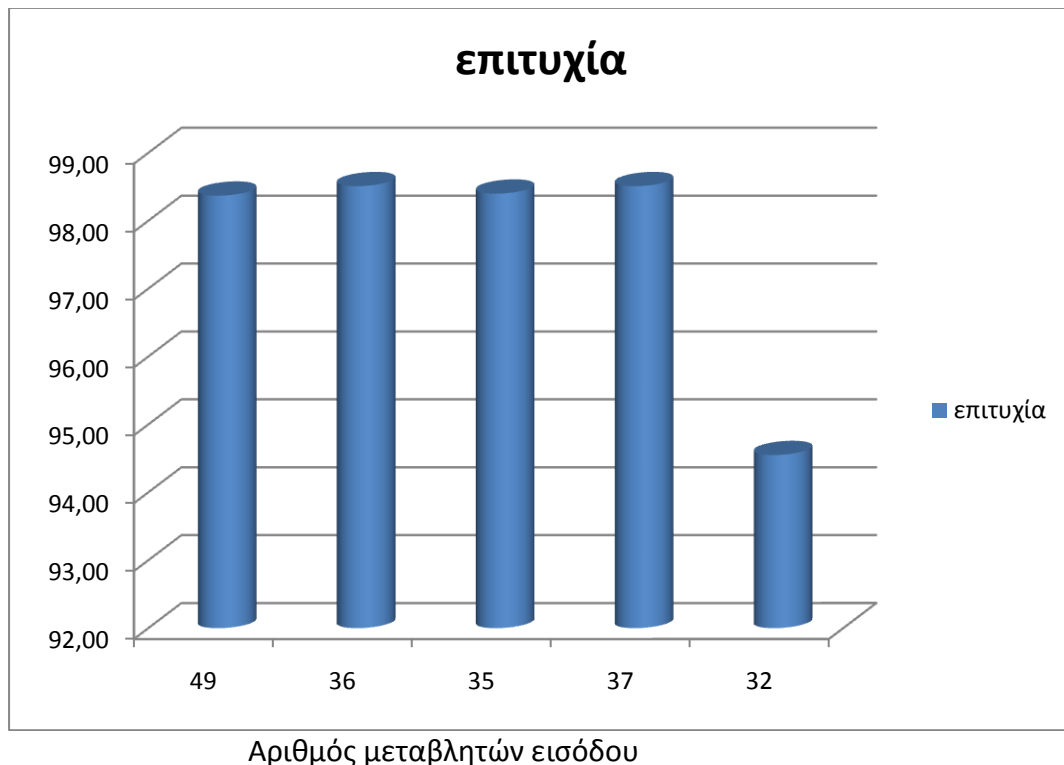
Πίνακας 5.5

Για το δίκτυο με αύξοντα αριθμό 2, σημειώνουμε την καλύτερη απόδοση ( $mse = 0,016$  και επιτυχία = 97,89% ), που είναι ελαφρώς καλύτερη από την απόδοση του δικτύου που αναφέραμε παραπάνω. Παρόλα αυτά, αφενός η απόδοση αυτή επιτεύχθηκε με παραπάνω εποχές εκπαίδευσης και αφετέρου δεν καταφέραμε να ξεπεράσουμε το φράγμα των 35 νευρώνων. Έτσι, το δίκτυο με εισόδους τα χαρακτηριστικά φωνής και τα 3 σημαντικότερα faps κρίνεται αποδοτικότερο, ως αυτό με τις λιγότερες εισόδους, τα λιγότερα συναπτικά βάρη και επομένως το λιγότερο πολύπλοκο.

## Κεφάλαιο 6: Επίλογος

### 6.1 Συμπεράσματα

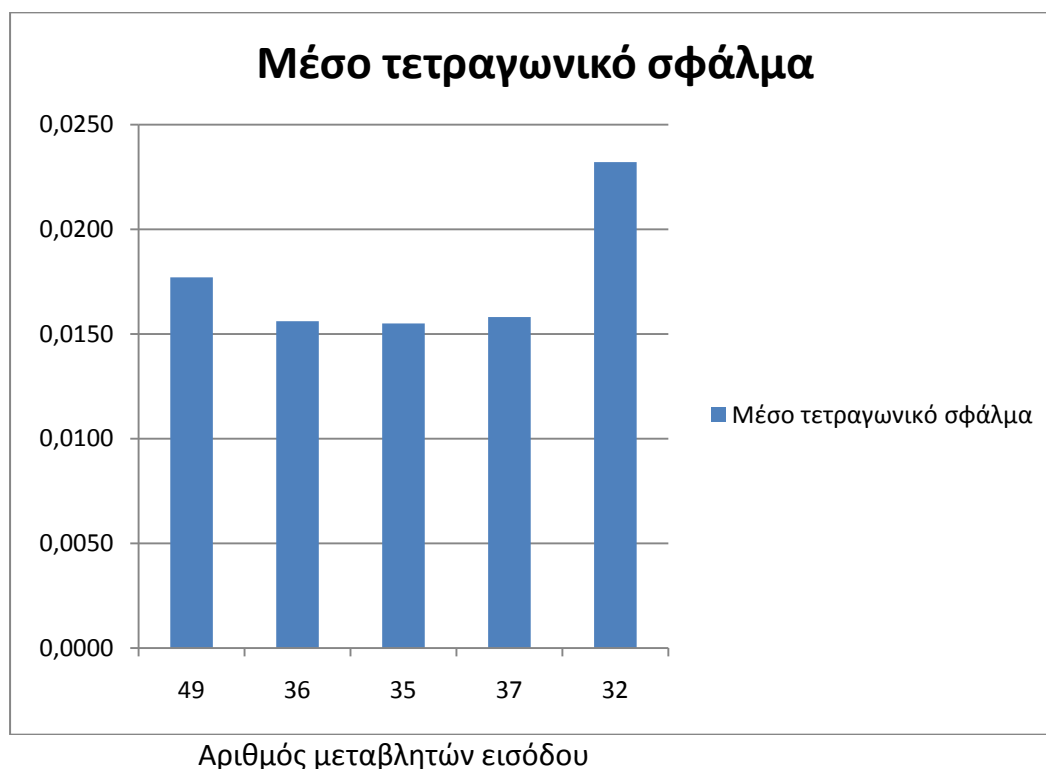
Συμπερασματικά , παραθέτουμε συγκριτικά διαγράμματα των αποδόσεων που πετύχαμε με κάθε βασική υλοποίηση. Συγκεκριμένα στα σχήματα 6.1 και 6.2 απεικονίζεται η επιτυχία και το μέσο τετραγωνικό σφάλμα , αντίστοιχα, των πολυστρωματικών νευρωνικών δικτύων με 50 νευρώνες στο κρυφό επίπεδο , που έχουν εκπαιδευτεί για 2000 εποχές.



Σχήμα 6.1: Επιτυχία των δικτύων με 50 νευρώνες στο κρυφό επίπεδο

Κάθε δίκτυο έχει διαφορετικό αριθμό μεταβλητών εισόδου, όπως προέκυψε από τις διαδικασίες μείωσης δεδομένων. Θυμίζουμε ότι τα χαρακτηριστικά της φωνής είναι σε όλα τα δίκτυα 32, ενώ αυτό που διαφοροποιεί τον αριθμό μεταβλητών είναι τα χαρακτηριστικά του προσώπου. Η πρώτη στήλη των διαγραμμάτων αντιπροσωπεύει το αρχικό βέλτιστο δίκτυο που περιλαμβάνει όλα (17) τα χαρακτηριστικά προσώπου (σύνολο 49 μεταβλητές). Η δεύτερη στήλη αντιπροσωπεύει το δίκτυο που υλοποιήθηκε μετά τον υπολογισμό της συσχέτισης των χαρακτηριστικών προσώπου και η είσοδός του συμπεριλαμβάνει μόνο τα ασυσχέτιστα (4) χαρακτηριστικά προσώπου (σύνολο 36 μεταβλητές). Η τρίτη στήλη αντιπροσωπεύει το δίκτυο που υλοποιήθηκε μετά την ανάλυση ευαισθησίας και συμπεριλαμβάνει τα 3 σημαντικότερα χαρακτηριστικά προσώπου (σύνολο 35 μεταβλητές). Η τέταρτη στήλη αντιπροσωπεύει το δίκτυο που υλοποιήθηκε μετά από τον συνδυασμό των παραπάνω αναλύσεων, όπως παρουσιάστηκε στην παράγραφο 5.6.1, και περιέχει 5 χαρακτηριστικά προσώπου (σύνολο 37 μεταβλητές). Το δίκτυο που αναπαριστά η

τελευταία στήλη δεν περιέχει στην είσοδο του κανένα χαρακτηριστικό προσώπου, παρά μόνο τα χαρακτηριστικά φωνής και παρατίθεται καθαρά για συγκριτικούς λόγους.



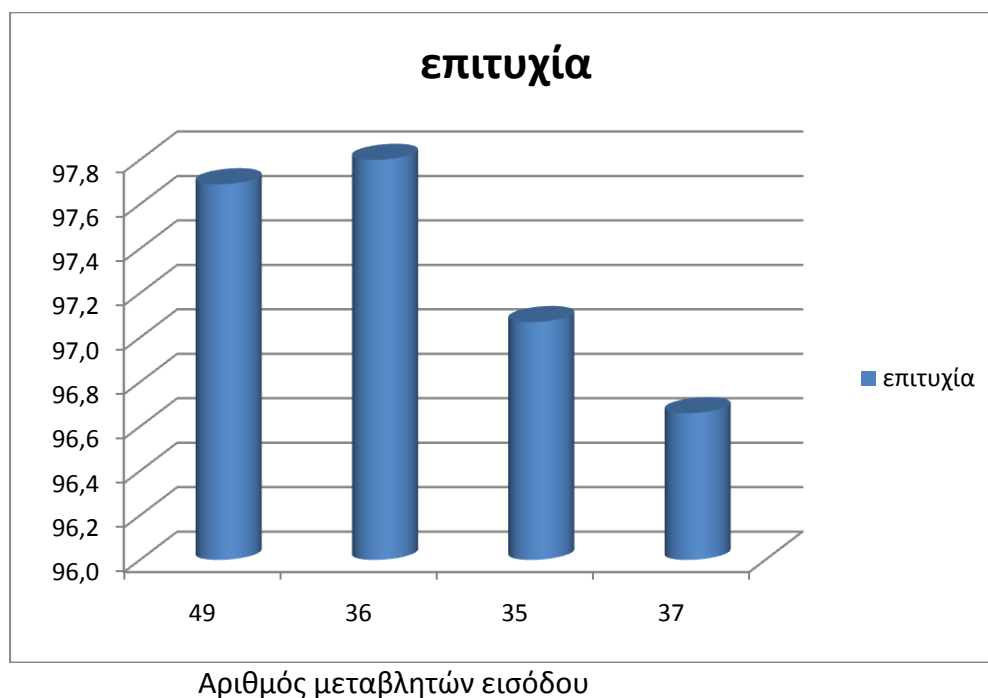
*Σχήμα 6.2: Μέσο τετραγωνικό σφάλμα των δικτύων με 50 νευρώνες στο κρυφό επίπεδο*

Από τα παραπάνω , και όπως έχει επισημανθεί και στις εκάστοτε ενότητες της εργασίας ξεχωριστά, καταλήγουμε στο γεγονός ότι οι διαδικασίες που ακολουθήθηκαν προκειμένου να μειωθούν οι εισόδοι κρίνονται απολύτως επιτυχείς αφού οι αποδόσεις των δικτύων με μειωμένες εισόδους είναι ακόμα καλύτερες και από την απόδοση του αρχικού δικτύου μιας και έχουν περίπου την ίδια επιτυχία , αλλά μικρότερο σφάλμα. Τώρα, όσον αφορά τη διάκριση μεταξύ των δικτύων με μειωμένες εισόδους, από τη στιγμή που οι αποδόσεις τους δεν διαφέρουν, “νικητής” αναδεικνύεται το δίκτυο με είσοδο τα 32 χαρακτηριστικά φωνής και τα 3 σημαντικότερα χαρακτηριστικά προσώπου, επειδή έχει τις λιγότερες μεταβλητές εισόδου και επομένως είναι λιγότερο πολύπλοκο.

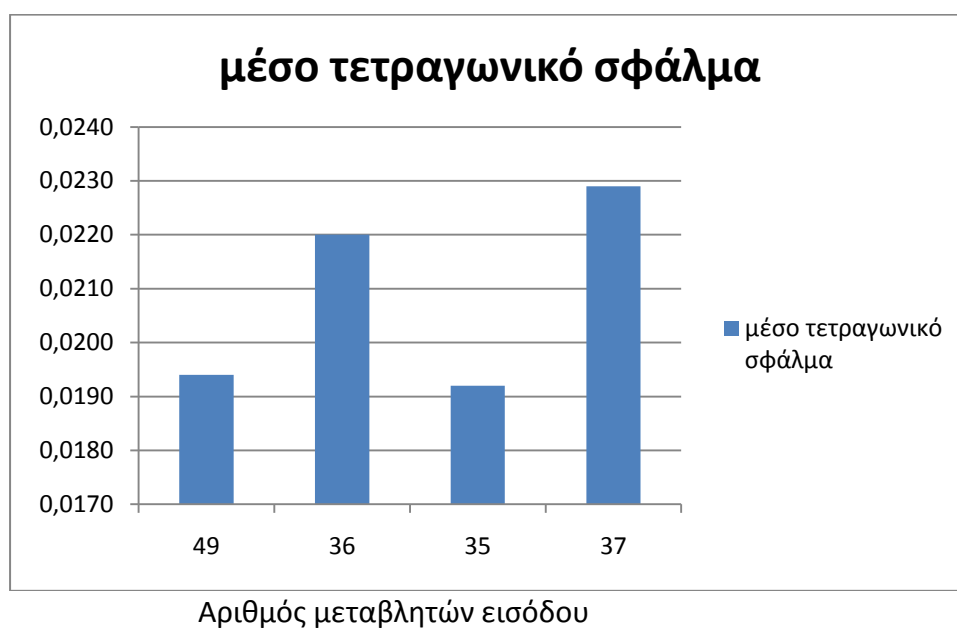
Μετά το συμπέρασμα της επιτυχούς ανάλυσης που πετύχαμε για την μείωση των εισόδων, παρατίθενται διαγράμματα που αφορούν την εκ νέου προσπάθεια για την ελαχιστοποίηση της πολυπλοκότητας του δικτύου μας, η οποία επιχειρήθηκε με αφαίρεση νευρώνων από το κρυφό επίπεδο. Στα σχήματα 6.3 και 6.4 οι στήλες αντιπροσωπεύουν μεγέθη απόδοσης δικτύων (επιτυχία, μέσο τετραγωνικό σφάλμα) με 35 νευρώνες στο κρυφό επίπεδο, τα οποία εκπαιδεύτηκαν για 2000 εποχές και έχουν αριθμό μεταβλητών εισόδου κατ’ αντιστοιχία με τα σχήματα 6.1 και 6.2 με το δίκτυο με είσοδο μόνο των χαρακτηριστικών φωνής να παραλείπεται.

Από το διάγραμμα επιτυχίας παρατηρούμε ότι το δίκτυο με τις ασυσχέτιστες μεταβλητές σημειώνει την ίδια περίπου επιτυχία με τα δίκτυα των 50 νευρώνων στο

κρυφό επίπεδο , αλλά έχει μεγαλύτερο σφάλμα. Το δίκτυο με τις 37 μεταβλητές εισόδου υστερεί και στα δύο μεγέθη.



Σχήμα 6.3: Επιτυχία των δικτύων με 35 νευρώνες στο κρυφό επίπεδο



Σχήμα 6.4: Μέσο τετραγωνικό σφάλμα των δικτύων με 35 νευρώνες στο κρυφό επίπεδο

Τέλος, το δίκτυο με τα 3 σημαντικότερα χαρακτηριστικά προσώπου υστερεί σε επιτυχία περίπου 1% αλλά έχει σαφώς πολύ μικρότερο σφάλμα από τα άλλα δύο δίκτυα το οποίο, αν και χειρότερο, είναι συγκρίσιμο με το σφάλμα των δικτύων με τους 50 νευρώνες στο κρυφό επίπεδο. Έχοντας ήδη εξηγήσει ότι το μέσο τετραγωνικό σφάλμα είναι το πρωταρχικό μας μέτρο σύγκρισης όταν η ακρίβεια και

η επιτυχία έχουν μικρές διαφορές, συμπεραίνουμε ότι το δίκτυο με τους 35 νευρώνες στο κρυφό επίπεδο με είσοδο 35 μεταβλητών είναι άκρως αποδοτικό. Κάνοντας, δηλαδή, την σύμβαση να χάσουμε κάπως σε απόδοση, καταφέρνουμε να μειώσουμε δραματικά την πολυπλοκότητα του προβλήματός μας, κερδίζοντας σε υπολογιστική μνήμη και ταχύτητα επεξεργασίας.

Συνοψίζοντας τα παραπάνω, καταφέραμε από το αρχικό πρόβλημα της αναγνώρισης συναισθήματος που χρειαζόταν 49 χαρακτηριστικά στη είσοδο του, να έχουμε την ίδια απόδοση με 35 χαρακτηριστικά. Με αυτά τα δεδομένα γίνεται σαφές ότι, αφενός η επεξεργασία εικόνας γίνεται απλούστερη και πολύ πιο γρήγορα, αφού δεν απαιτείται πλέον να υπολογίζονται 14 χαρακτηριστικά και αφετέρου το σύστημα που αναγνωρίζει το συναίσθημα εκπαιδεύεται και αποκρίνεται ταχύτερα. Επιπλέον εάν θελήσουμε να 'θυσιάσουμε' ένα μικρό μέρος επιτυχίας του συστήματός μας, που για τους περισσότερους σκοπούς που θα μπορούσε να χρησιμοποιηθεί το σύστημα, ευχαρίστως θα το θυσιάζαμε, είδαμε πειραματικά ότι η πολυπλοκότητα του συστήματος μπορεί να μειωθεί κι άλλο ('ψαλιδίζοντας' νευρώνες) με αποτέλεσμα την ακόμα πιο ταχεία απόκριση.

## 6.2 Μελλοντικές επεκτάσεις

Στο κεφάλαιο 4 μελετήσαμε την συσχέτιση των faps με αφορμή την ιδιότητα τους, όντας μετατοπίσεις σημείων του προσώπου, κάποια από αυτά να μεταβάλλονται στο χρόνο με παρόμοιο τρόπο. Μια άλλη ιδιότητα τους, που αποτελεί τροφή για μελλοντική επανεξέταση του προβλήματος, είναι ότι, ως φυσικά μεγέθη, είναι συνεχή στο χρόνο. Η ιδιότητα της συνέχειας μας δίνει την ιδέα ότι, ίσως να ήταν χρήσιμη η εισαγωγή μνήμης. Δε θα διευκόλυνε η συμβολή της προηγούμενης κατάστασης στην αναγνώριση της επόμενης, εφόσον έχουν ντετερμινιστική σχέση; Στην εργασία μας χρησιμοποιήσαμε δίκτυα προσοτροφodότησης, επομένως μια μελλοντική επέκταση θα μπορούσε να είναι η προσέγγιση του προβλήματος με ένα αναδρομικό νευρωνικό δίκτυο, που θα μπορεί να συνδυάζει επιβλεπόμενη μάθηση και μνήμη. Ένα τέτοιο δίκτυο είναι το δίκτυο Elman.

Η εκπαίδευση του δικτύου με δεδομένα εισόδου, που θα έχουν προκύψει από πολλές μαγνητοσκοπημένες ομιλίες διαφορετικών ανθρώπων, θα παρείχε μια πιο αντικειμενική αξιολόγηση της απόδοσης του δικτύου. Επιπλέον, μια ανάλυση ευαισθησίας σε ένα τέτοιο δίκτυο θα μας απαντούσε στο ερώτημα για την καθολικότητα των σημαντικών faps. Είναι τα ίδια faps τα σημαντικότερα για κάθε ομιλητή; Ή ποία faps είναι σημαντικά για κάποια ομάδα ομιλητών και ποιο το γνώρισμα που τους κατατάσσει σε αυτήν την ομάδα; Με διαφορετικούς ομιλητές οι απαντήσεις σε αυτά τα ερωτήματα θα ήταν εφικτές και θα μας παρείχαν πληροφoρία, που θα μας έδινε την δυνατότητα να φτιάξουμε ένα πιο 'έξυπνο' σύστημα.

Όταν βλέπουμε κάποιον χαρούμενο ή λυπημένο, αναγνωρίζουμε το συναίσθημα του επειδή κάποιος μας το έχει μάθει ή έχουμε μάθει να το αναγνωρίζουμε μόνοι μας. Ακολουθούν τα χαρακτηριστικά που χρησιμοποιούμε για να αναγνωρίσουμε το συναίσθημα κάποια χωρική συμπεριφορά ή είναι χαοτικά; Η εκπαίδευση και ο έλεγχος συστημάτων που υλοποιούνται με αλγορίθμους που χρησιμοποιούν μη επιβλεπόμενη μάθηση θα ήταν μια πολύ ενδιαφέρουσα επέκταση της εργασίας



μας. Θα μπορούσε να ελεγχθεί το πρόβλημα με κλασσικούς clusterers και επεκτάσεις αυτών για πιο ειδικά αποτελέσματα.

## Βιβλιογραφία

- [1] Κωνσταντίνος Διαμαντάρας : “Τεχνητά Νευρωνικά δίκτυα”, εκδόσεις Κλειδάριθμος, Αθήνα 2007
- [2] Σπύρος Τζαφέστας : “Υπολογιστική Νοημοσύνη”, εκδόσεις ΕΜΠ, Αθήνα 2002
- [3] Βλαχάβας Ιωάννης, Κεφαλάς Πέτρος, Βασιλειάδης Νικόλαος, Κόκκορας Φώτης, Σακελλαρίου Ηλίας : “Τεχνητή Νοημοσύνη”, Εκδόσεις Β. Γκιούρδας, 2006
- [4] Κωνσταντίνος Γεωργατζής : “Νευροασαφές σύστημα αναλογικού συμπερασμού”, Διπλωματική εργασία ΕΜΠ, Αθήνα 2009
- [5] Σαρλίν Χονδρού, Μηνάς Περτσελάκης, Γιώργος Καρυδάκης, Κώστας Καρπούζης, Στέφανος Κόλλιας : “Ανάλυση ευαισθησίας σε αναδρομικό νευρωνικό δίκτυο εκπαιδευμένο για αναγνώριση συναισθήματος”, 3ο Πανελλήνιο Συνέδριο Φοιτητών Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
- [6] Κωνσταντίνος Καρπούζης : “ Προσαρμοζόμενα παραμετρικά μοντέλα ανθρώπινων προσώπων και σωμάτων και εφαρμογές στην επικοινωνία ανθρώπου μηχανής”, Διδακτορική διατριβή ΕΜΠ, Αθήνα 2001
- [7] Αμαρυλλίς – Όλγα Θ. Ραουζαίου : “Αναπαράσταση Συναισθήματος και Σύνθεση Εκφράσεων Προσώπου και Χειρονομιών σε Εικονικούς Χαρακτήρες”, Διδακτορική διατριβή ΕΜΠ, Αθήνα 2006
- [8] Γ. Παπαδόπουλος : “Συσχέτιση δύο μεταβλητών”, [www.aua.gr/gpapadopoulos](http://www.aua.gr/gpapadopoulos)
- [9] A. Raouzaïou, K. Karpouzis and S. Kollias, "Online Gaming and Emotion Representation", International Workshop Very Low Bitrate Video Coding 2003 (VLBV03), 18-19 September, Madrid, Spain
- [10] K. Karpouzis, A. Raouzaïou, P. Tzouveli and S. Kollias, "MPEG-4: One multimedia standard to unite all", IEEE International Conference on Multimedia & Expo (ICME), July 6-9, Baltimore, MD, USA