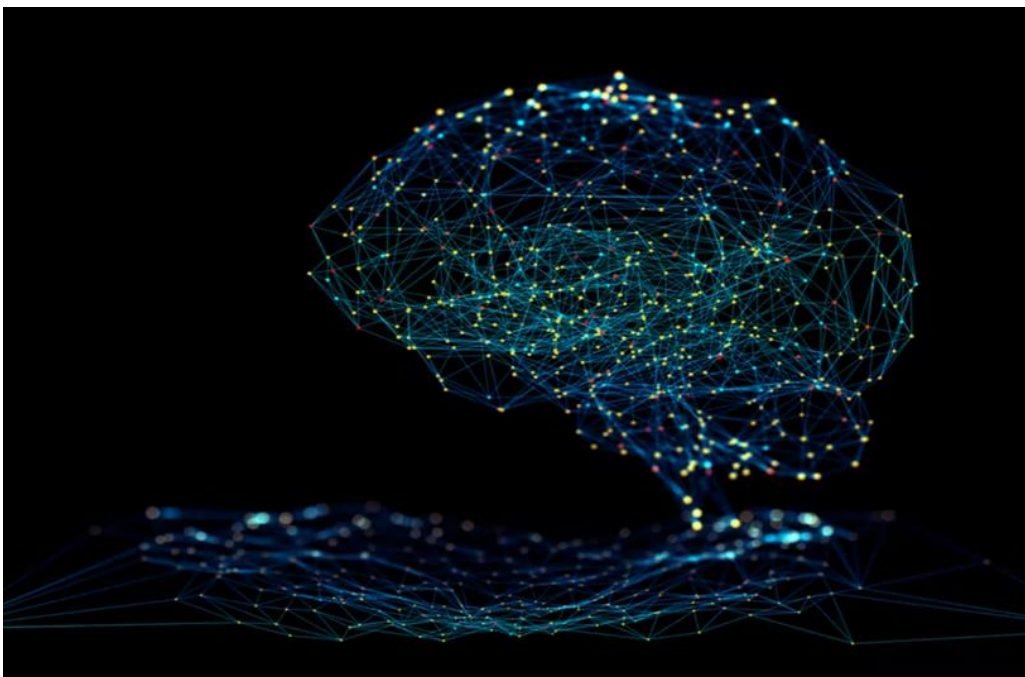




ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ & ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

Διπλωματική Εργασία

**Ανάπτυξη τεχνικών εκπαίδευσης νευρωνικών δικτύων ακτινικής
συνάρτησης βάσης με τον αλγόριθμο Levenberg-Marquardt**



Φοιτητής: Γερολυμάτος Γεώργιος
ΑΜ: 50347899

Επιβλέπων Καθηγητής
Αλεξανδρίδης Αλέξανδρος
Καθηγητής

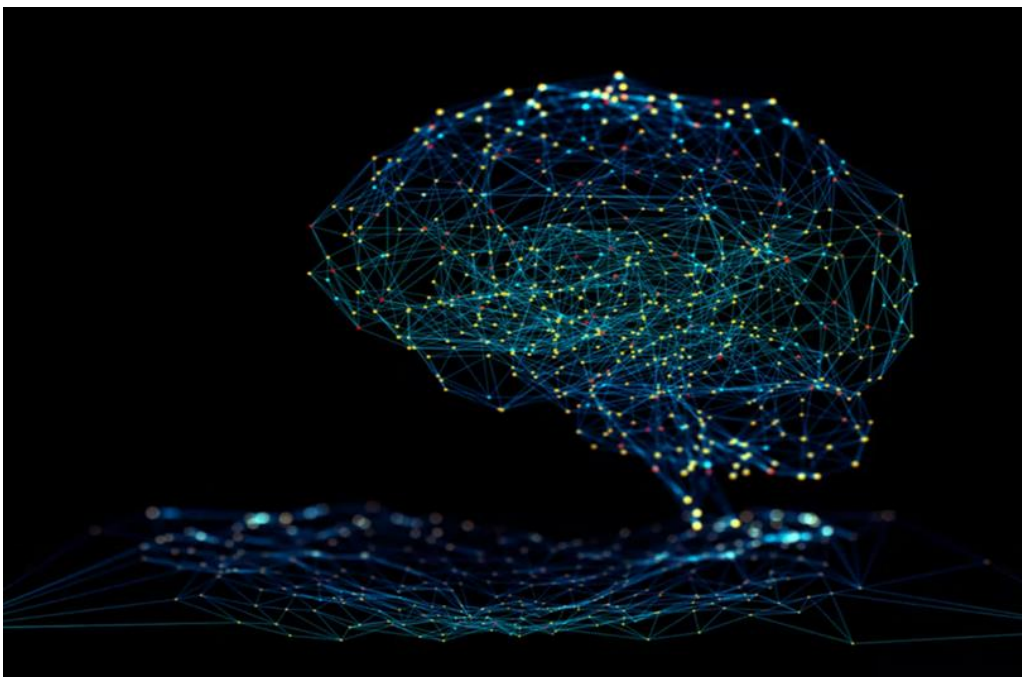
ΑΘΗΝΑ-ΑΙΓΑΛΕΩ, Σεπτέμβριος 2022



UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

Diploma Thesis

**Development of radial basis function neural network training techniques
based on the Levenberg-Marquardt algorithm**



Student: Gerolymatos Georgios
Registration Number: 50347899

Supervisor

Alex Alexandridis
Professor

ATHENS-EGALEO, September 2022

Η Διπλωματική Εργασία έγινε αποδεκτή και βαθμολογήθηκε από την εξής τριμελή
επιτροπή:

Αλεξανδρίδης Αλέξανδρος, Καθηγητής	Παπαδόπουλος Περικλής, Καθηγητής	Φαμέλης Ιωάννης, Καθηγητής
(Υπογραφή)	(Υπογραφή)	(Υπογραφή)

Copyright © Με επιφύλαξη παντός δικαιώματος. All rights reserved.

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και Γερολυμάτος Γεώργιος,
Σεπτέμβριος, 2022**

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 τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον/την συγγραφέα του και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις θέσεις του επιβλέποντος, της επιτροπής εξέτασης ή τις επίσημες θέσεις του Τμήματος και του Ιδρύματος.

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Γερολυμάτος Γεώργιος του Παναγιώτη, με αριθμό μητρώου 50347899 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ,

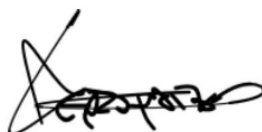
δηλώνω υπεύθυνα ότι:

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

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

Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου, παρά μόνο και έπειτα από αίτησή μου στη Βιβλιοθήκη και έγκριση του επιβλέποντος καθηγητή.»

Ο Δηλών
Γερολυμάτος Γεώργιος



*" We can only see a short
distance ahead, but we can
see plenty there that needs
to be done."
-- Alan Turing --*

Ευχαριστίες

Τα λόγια δεν μπορούν να εκφράσουν την ευγνωμοσύνη μου προς τον επιβλέποντα καθηγητή μου, Δρ. Αλέξανδρο Αλεξανδρίδη, για τη συνεχή καθοδήγησή του και την αμέριστη προσφορά του. Η ανεπιτήδευτη προσέγγισή του στην έρευνα και στην επιστήμη αποτελεί πηγή έμπνευσης και αντικατοπτρίζεται από ένα απλό αλλά σαφή τρόπο διδασκαλίας, το οποίο είναι κάτι που ελπίζω να μεταφέρω σε όλη τη διάρκεια της επιστημονικής μου σταδιοδρομίας.

Ευχαριστώ θερμά για τη συνεχή υποστήριξη καθ' όλης της διάρκειας της εκπόνησης της εργασίας, της υποψήφιας διδάκτορα, Δέσποινας Καραμιχαηλίδου, μέλος του ερευνητικού εργαστηρίου Telsip.

Αισθάνομαι επίσης τυχερός που υπήρξα μέλος της ομάδας iReact-NG του ερευνητικού εργαστηρίου Telsip, μια υπέροχη ομάδα με δημιουργικούς και άριστους επιστήμονες.

Δεν μπορώ να παραλείψω τον άνθρωπο που με στήριξε όλα τα χρόνια των σπουδών μου. Χωρίς αυτόν, όλα θα ήταν διαφορετικά. Ορφέα Κ. , φίλε μου, θα σου είμαι παντοτινά ευγνώμων.

Κλείνοντας, η ολοκλήρωση της διπλωματικής μου εργασίας, σηματοδοτεί την επιτυχή ολοκλήρωση των σπουδών μου και έναρξη ενός νέου κεφαλαίου.

Περίληψη

Η εν λόγω εργασία έχει ως αντικείμενο την ανάπτυξη και αξιολόγηση τεχνικών εκπαίδευσης νευρωνικών δικτύων ακτινικής συνάρτησης βάσης (Radial Basis Function Networks, RBFNs), με γνώμονα την δημιουργία μοντέλων υψηλής απόδοσης και ακρίβειας. Η εκπαίδευση βασίζεται στον αλγόριθμο βελτιστοποίησης Levenberg-Marquardt (LM) με τον οποίον γίνεται ο προσδιορισμός των παραμέτρων του μοντέλου. Συγκεκριμένα, των συναπτικών βαρών, των θέσεων των κέντρων RBF και εάν είναι απαραίτητο, των ευρών των συναρτήσεων βάσης. Δοκιμάστηκαν δύο εκδοχές της εν λόγω μεθόδου για την δυνατότητα σύγκρισης των αποτελεσμάτων. Στην πρώτη εκδοχή, η κατάτμηση των συνόλων των παραμέτρων του προβλήματος εκπαίδευσης επιτυγχάνεται με γνώμονα τη γραμμικότητά τους. Η απαλοιφή των γραμμικών παραμέτρων και η ανάπτυξη της προβολής της μεταβλητής (Variable Projection, VP) προς τις μη γραμμικές παραμέτρων, εξελίσσεται σε δυο επιμέρους προβλήματα: το μη γραμμικό για την ενημέρωση των κέντρων σε κάθε επανάληψη της LM, καθώς και τον υπολογισμό των βαρών από την επίλυση ενός απλού γραμμικού προβλήματος ελαχίστων τετραγώνων. Στην δεύτερη εκδοχή δεν προηγείται διάσπαση των συνόλων των παραμέτρων, το πρόβλημα είναι πλήρες συναρτησιακό, με τα βάρη να ενημερώνονται παράλληλα με τα κέντρα στις επαναλήψεις του αλγορίθμου LM. Οι αρχικές θέσεις των κέντρων προέρχονται από τον αλγόριθμο των ασαφών μέσων (Fuzzy Means, FM), τα αρχικά βάρη καθορίζονται από την μέθοδο των ελαχίστων τετραγώνων και τα εύρη των συναρτήσεων βάσης υπολογίζονται με την μέθοδο των κ-κοντινότερων γειτόνων. Οι δύο εκδοχές δοκιμάστηκαν σε 12 προσομοιωμένα αλλά και πραγματικά σύνολα δεδομένων της βιβλιογραφίας και συγκρίθηκαν με άλλες μεθόδους εκπαίδευσης νευρωνικών δικτύων. Τα αποτελέσματα και των δύο εκδοχών κρίθηκαν ικανοποιητικά και ανταγωνιστικά σε σύγκριση με των άλλων μεθόδων. Ιδιαίτερα με την εκδοχή VP δημιουργήθηκαν δίκτυα με υψηλότερη ταχύτητα εκπαίδευσης και ταυτόχρονα μικρότερα σε μέγεθος.

Λέξεις – κλειδιά

Αλγόριθμος Levenberg-Marquardt, ασαφή σύνολα, δίκτυα ακτινικής συνάρτησης βάσης, μέθοδος προβολής μεταβλητής, τεχνητή νοημοσύνη, νευρωνικό δίκτυο,

Abstract

The purpose of this work is to develop and evaluation techniques for training Radial Basis Function neural networks (RBFNs), focusing on the creation of high performance and accuracy models. Training is performed using the Levenberg-Marquardt (LM) optimization algorithm which is used to define the model parameters. In particular the synaptic weights, the positions of the RBF centers and if it is necessary, the widths of the basis functions. Two versions of such method were tested so that the results can be compared. In the first version, the segmentation of the sets in the training problem parameters is achieved by taking into account whether they are linear. The elimination process for the linear parameters and the development of the Variable Projection (VP) towards the nonlinear parameters evolves into two subproblems. The nonlinear subproblem corresponds to updating the centers at each iteration of LM, while the linear one to calculating the weights by solving a simplified linear least squares' problem. In the second version where no separation of the parameter sets is preceded, the problem becomes Fully Functional (FF), with the weights being updated along with the centers in the iterations of the LM algorithm. The initial positions of the centers are obtained from the Fuzzy Means (FM) algorithm, the initial weights are determined by the least squares' method and the widths of the basis functions are computed by the k-nearest neighbors' algorithm. The two versions were tested on 12 simulated and real-world benchmark datasets in the literature and compared with other neural network training methods. The results of both versions were found to be competitive compared to the other methods. Especially the VP version created networks with higher training speed and smaller size.

Keywords

artificial intelligence, fuzzy sets, Levenberg-Marquardt algorithm, neural network, radial basis function network, variable projection

Περιεχόμενα

Κατάλογος σχημάτων.....	11
Αλφαβητικό Ευρετήριο.....	12
1 ΕΙΣΑΓΩΓΗ.....	13
1.1 Το θέμα της διπλωματικής εργασίας	13
1.1 Σκοπός και επιδιώξεις	14
1.2 Υπάρχουσες μέθοδοι εκπαίδευσης.....	14
1.3 Μεθοδολογία της εργασίας.....	15
1.1 Δομή εργασίας.....	15
2 Τεχνητή Νοημοσύνη.....	17
2.1 Ιστορική αναδρομή	18
2.2 Εφαρμογές	19
2.3 Προσεγγίσεις της Τεχνητής Νοημοσύνης	20
2.4 Μάθηση – Απόφαση - Βελτιστοποίηση.....	20
2.5 Αποθήκευση της γνώσης	21
2.6 Μέθοδος Ελαχίστων Τετραγώνων	23
3 Νευρωνικά δίκτυα.....	26
3.1 Η εξέλιξη της τεχνολογίας	26
3.2 Ο νευρώνας	27
3.2.1 Μοντέλο Νευρώνα McCulloch-Pitts	28
3.2.2 Λειτουργία του μοντέλου	28
3.2.3 Συναρτήσεις ενεργοποίησης	29
3.3 Αρχιτεκτονικές τεχνητών νευρωνικών δικτύων.....	31
3.3.1 Νευρωνικά δίκτυα πρόσθιας τροφοδότησης.....	31
3.3.2 Αναδρομικά νευρωνικά δίκτυα	32
3.3.3 Νευρωνικά δίκτυα ακτινικής συνάρτησης βάσης.....	33
3.3.4 Αυτό-οργανώμενα νευρωνικά δίκτυα.....	33
3.4 Perceptron ενός και περισσότερων στοιβάδων (MLP).....	34
3.5 Μηχανική Μάθηση	34
4 Δίκτυα συνάρτησης ακτινικής βάσης (RBF)	37
4.1 Σχεδιασμός δικτύου RBF	37
4.2 Αλγόριθμος k-means.....	38
4.3 Αλγόριθμος ασαφών μέσων (Fuzzy Means,FM)	38
4.4 Στατιστικοί δείκτες.....	41
5 Μέθοδοι Βελτιστοποίησης.....	43
5.1 Λίγο πιο αναλυτικά.....	43
5.1.1 Γραμμικά προβλήματα	44
5.1.2 Μη γραμμικά προβλήματα.....	44
5.2 Βελτιστοποίηση στο δίκτυο RBF	45
5.3 Μέθοδοι κατάβασης	45
5.3.1 Μέθοδος Gauss-Newton	46
5.3.2 Μέθοδος Levenberg-Marquardt, LM	47
5.4 Μέθοδος Variable Projection.....	48
5.5 Παραγοντοποίηση QR.....	49
6 Πειραματική διαδικασία	51

6.1	Η εκπαίδευση στην πράξη	51
6.1.1	Χωρισμός των δεδομένων εισόδου.....	51
6.1.2	Κανονικοποίηση των δεδομένων	52
6.2	1^η Εφαρμογή Variable Projection	52
6.3	2^η Εφαρμογή Πλήρους Συναρτησιακού	55
7	Ολοκλήρωση πειραματικής διαδικασίας	56
7.1	Ανάλυση των benchmark datasets.....	56
7.2	Αποτελέσματα	57
7.3	Συζήτηση Αποτελεσμάτων.....	59
7.4	Συμπεράσματα	59
8	Βιβλιογραφία – Αναφορές - Διαδικτυακές Πηγές	61
Παράρτημα Α : Αλγόριθμος Fuzzy Means		66
Παράρτημα Β : Αλγόριθμος Variable Projection, Levenberg-Marquardt		67

Κατάλογος σχημάτων

Σχήμα 1 Ευθεία $y=ax$	21
Σχήμα 2 Ευθεία $y=ax+b$	22
Σχήμα 3 Καμπύλη $y=ax^2+bx+c$	22
Σχήμα 4 Ο νευρώνας.....	27
Σχήμα 5 Τεχνητός νευρώνας.....	28
Σχήμα 6 Βηματική συνάρτηση.....	29
Σχήμα 7 Γραμμική συνάρτηση.....	30
Σχήμα 8 Σιγμοειδής συνάρτηση.....	30
Σχήμα 9 Δίκτυο με 1 στοιβάδα.....	31
Σχήμα 10 Δίκτυο πολλών στοιβάδων.....	32
Σχήμα 11 Δίκτυο με ανάδραση	32
Σχήμα 12 Δομή ενός δικτύου RBF	33
Σχήμα 13 Αυτο-οργανώμενα νευρωνικά δίκτυα.....	33
Σχήμα 14 Μάθηση με επίβλεψη	35
Σχήμα 15 Μάθηση χωρίς επίβλεψη	36
Σχήμα 16 Μάθηση με ενίσχυση.....	36
Σχήμα 17 Ομαδοποίηση k-means	38
Σχήμα 18 Ασαφές σύνολο.....	39
Σχήμα 19 Ασαφής διαχωρισμός.....	40
Σχήμα 20 Δισδιάστατος ασαφής διαχωρισμός.....	40

Αλφαβητικό Ευρετήριο

AI: Artificial Intelligence

FM: Fuzzy Means

PSO: Particle Swarm Optimization

RBF: Radial Basis Network

IEEE: Institute of Electrical and Electronic Engineers

LM: Levenberg Marquardt

MLP: Multilayer Perceptron

PID: Proportional Integral Derivative

VP: Variable Projection

1 ΕΙΣΑΓΩΓΗ

Ο άνθρωπος έχει μελετήσει και μιμηθεί την αρμονία των φυσικών συστημάτων που λειτουργούν αυτόνομα μετά από δισεκατομμύρια χρόνια έρευνας και ανάπτυξης. Ο όρος ‘βιομιμητική’ περιγράφει τις λύσεις που έχει υιοθετήσει η φύση μεταφράζοντάς τες στην ανάπτυξη αλγοριθμικών μοντέλων. [1]

Ένας τομέας μίμησης των βιολογικών οργανισμών είναι η Υπολογιστική Νοημοσύνη. Συγκεκριμένα είναι μια ομάδα ανάπτυξης μεθόδων των Τεχνητών Νευρωνικών Δικτύων [2], της Ασαφούς Λογικής [3], της Νοημοσύνης Σμήνους [4], [5], [6], του Εξελικτικού Υπολογισμού [7], και του τεχνητού ανοσοποιητικού συστήματος [8], για την επίλυση πολύπλοκων προβλημάτων [9].

Συγκεκριμένα η μοντελοποίηση του ανθρώπινου εγκεφάλου είναι ένας πολλά υποσχόμενος και ενδιαφέρον τομέας. Τα Τεχνητά Νευρωνικά Δίκτυα, τα οποία είναι από τα σημαντικότερα εργαλεία που αναφέρθηκαν, προσπαθούν να προσομοιώσουν τις ικανότητες του ανθρώπινου εγκεφάλου. Κάτι το οποίο είναι πολύ δύσκολο να πραγματοποιηθεί λόγω των περιορισμών τους σε μέγεθος και λειτουργικότητα. [10]

1.1 Το θέμα της διπλωματικής εργασίας

Υπάρχει ήδη μεγάλη ποικιλία διάφορων τύπων και αρχιτεκτονικών των Τεχνητών Νευρωνικών Δικτύων. Το πεδίο της παρούσας εργασίας είναι τα Νευρωνικά Δίκτυα που διαμορφώνουν μια συνάρτηση βάσης ακτινικής συμμετρίας (Radial Basis Network, RBF). Διακρίνονται σε σύγκριση με πιο πολύπλοκα, όπως τα πολυεπίπεδα δίκτυα Perceptron (Multilayer Perceptron, MLP), για την απλή δομή τους καθώς και για την ταχύτητα εκπαίδευσής τους. Αποτελούνται από ένα μόνο κρυφό επίπεδο, συνδεδεμένο γραμμικά με τη μονάδα εξόδου. Έχουν ενσωματωθεί σε πληθώρα εφαρμογών όπως η αναγνώριση προτύπων και η μοντελοποίηση συστημάτων [11], [12].

Η εκπαίδευση ενός δικτύου RBF ισοδυναμεί με τον υπολογισμό των παραμέτρων που ακολουθούν:

- ο αριθμός των μονάδων RBF
- οι θέσεις των RBF κέντρων
- τα συναπτικά βάρη
- τα εύρη των συναρτήσεων βάσης, εφόσον απαιτούνται

Ο χρόνος εκπαίδευσης είναι συνήθως μικρός και διαφοροποιείται από τον αριθμό παραμέτρων που πρέπει να διαχειριστεί ο αλγόριθμος. Μια από τις πιο γρήγορες τεχνικές είναι ο διαχωρισμός της εκπαίδευσης σε δύο μέρη. Στο πρώτο μέρος υπολογίζονται οι παράμετροι των κρυφών μονάδων με μη εποπτευόμενη μάθηση ενώ στο δεύτερο τα συναπτικά βάρη με εποπτευόμενη μάθηση [9], [12].

Η απόδοση του δικτύου επηρεάζεται άμεσα από το πλήθος των μονάδων RBF αλλά και από την επιλογή των θέσεων των κέντρων [13]. Η αρχική προσέγγιση επιλογής των κέντρων ήταν απλή καθώς το κάθε δεδομένο εκπαίδευσης έπαιρνε από ένα κέντρο. Με αυτόν τον τρόπο όμως μεγάλωνε το μέγεθος του δικτύου αυξάνοντας εκθετικά την απαίτηση της μνήμης [11]. Στην παρούσα εργασία, για την εύρεση των παραμέτρων, χρησιμοποιείται η μέθοδος βελτιστοποίησης Levenberg-Marquardt [14], [15]. Η χρήση της μεθόδου LM γίνεται για την διερεύνηση της αποτελεσματικότητας της πάνω στην ταχύτητα εκπαίδευσης όσο και στην ακρίβεια πρόβλεψης.

1.1 Σκοπός και επιδιώξεις

Ο σκοπός της εργασίας είναι η ανάπτυξη τεχνικών εκπαίδευσης δικτύων RBF και η αξιολόγησή τους στην αντιμετώπιση των προκλήσεων που τα διέπουν. Ένα από τα σημαντικότερα προβλήματα που αντιμετωπίζουν σήμερα οι σχεδιαστές νευρωνικών δικτύων είναι η επιλογή του κατάλληλου μεγέθους δικτύου για μια δεδομένη εφαρμογή. Επομένως, η διαδικασία επιλογής της κατάλληλης αρχιτεκτονικής νευρωνικού δικτύου για ένα δεδομένο πρόβλημα εξακολουθεί να είναι ένα αμφιλεγόμενο ζήτημα.

Επίσης, τα δεδομένα συνήθως περιέχουν κάποιο θόρυβο, ο οποίος οδηγεί σε ένα μικρό σφάλμα εκπαίδευσης. Σε αυτές τις περιπτώσεις αυξάνεται το σφάλμα γενίκευσης καθώς το δίκτυο αποκλίνει από την πραγματική συνάρτηση. Το πρόβλημα αυτό ονομάζεται υπερπροσαρμογή και είναι ένα κρίσιμο ζήτημα.

Στην προσπάθεια να βρεθεί λύση για τα δύο αναφερθέντα προβλήματα, εμφανίζεται ένα τρίτο, η αύξηση της υπολογιστικής ισχύος. Ένα εξίσου σημαντικό πρόβλημα που επηρεάζει την ταχύτητα της εκπαίδευσης.

Σκοπός της εργασίας είναι γίνει μια διερεύνηση της αποτελεσματικότητας νέων μεθόδων εκπαίδευσης δικτύων RBF ως προς τον υπολογισμό των παραμέτρων του νευρωνικού δικτύου με το λιγότερο σφάλμα, στον καλύτερο χρόνο.

1.2 Υπάρχουσες μέθοδοι εκπαίδευσης

Σχετικά με το μέγεθος του δικτύου, έχουν προταθεί αρκετές τεχνικές για την ελαχιστοποίηση του. Ένας εκ των οποίων είναι ο αλγόριθμος k-means, ο οποίος τα καταφέρνει κάνοντας συσταδοποίηση του χώρου εισόδου, αλλά αντιμετωπίζει αρκετά προβλήματα όπως την ανικανότητα να προσδιορίσει τον αριθμό των κέντρων [16]. Για αυτόν τον λόγο εφαρμόζονται άλλες μέθοδοι όπως ο αλγόριθμος των ασαφών μέσων, ο οποίος βασίζεται στον ασαφή διαχωρισμό του χώρου εισόδου [11], [12]. Μερικές επιτυχημένες εφαρμογές του αλγορίθμου των ασαφών μέσων είναι [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29].

Ο συνδυασμός των μεθόδων αποτελεί ένα σημαντικό σημείο έρευνας της Υπολογιστικής Νοημοσύνης. Μεταξύ άλλων στα [11], [30], εφαρμόζεται βελτιστοποίηση με σμήνη

σωματιδίων για τον ασαφή διαμερισμό του χώρου εισόδου και στο [13], χρησιμοποιούνται οι γενετικοί αλγόριθμοι για την βέλτιστη επιλογή των κέντρων των μονάδων RBF.

Περαιτέρω πληροφορίες αναφορικά με τις τεχνικές εκπαίδευσης ενός δικτύου RBF διατίθενται στα [9], [11], [12], [31].

1.3 Μεθοδολογία της εργασίας

Ο βασικός αλγόριθμος εκπαίδευσης των νευρωνικών δικτύων RBF που χρησιμοποιείται στην παρούσα εργασία είναι ο αλγόριθμος βελτιστοποίησης Levenberg-Marquardt (LM), μια τροποποιημένη εκδοχή της μεθόδου Gauss-Newton. Εφαρμόζεται σε δύο παραλλαγές για τον καθορισμό των παραμέτρων του δικτύου [14], [15].

Στην πρώτη μέθοδο πραγματοποιείται η διάσπαση των συνόλων των παραμέτρων του προβλήματος εκπαίδευσης. Το πρόβλημα διασπάται σε δύο επιμέρους προβλήματα με την απαλοιφή των γραμμικών παραμέτρων και την ανάπτυξη της προβολής της μεταβλητής προς τις μη γραμμικές παραμέτρους. Η ενημέρωση των κέντρων γίνεται με την επαναληπτική διαδικασία του LM, καθώς τα βάρη που υπολογίζονται με την μέθοδο των ελαχίστων τετραγώνων. Για την βελτιστοποίηση των υπολογισμών χρησιμοποιείται η παραγοντοποίηση QR.

Στην δεύτερη μέθοδο δεν προηγείται διαχωρισμός στο σύνολο των παραμέτρων έχοντας πλέον ένα πλήρες συναρτησιακό πρόβλημα. Με τον αλγόριθμο LM ενημερώνονται και τα βάρη μαζί με τα κέντρα σε κάθε επανάληψη.

Οι αρχικές θέσεις των κέντρων προέρχονται από τον αλγόριθμο των ασαφών μέσων (Fuzzy Means, FM), τα αρχικά βάρη καθορίζονται από την μέθοδο των ελαχίστων τετραγώνων και τα εύρη των συναρτήσεων βάσης υπολογίζονται με την μέθοδο των κ-κοντινότερων γειτόνων.

1.1 Δομή εργασίας

Η εργασία αναπτύσσεται σε 7 κύρια κεφάλαια. Συνεχίζοντας, στο επόμενο κεφάλαιο γίνεται μια μικρή εισαγωγή στην Τεχνητή Νοημοσύνη με μια ιστορική αναδρομή καθώς και των εφαρμογών της. Αναλύονται οι διαφορετικές προσεγγίσεις που έχουν επικρατήσει, δίνεται έμφαση στους τρόπους της μηχανικής μάθησης και στον τρόπο αποθήκευσης της γνώσης.

Στο 3^ο κεφάλαιο ακολουθεί η ανάλυση του νευρώνα και των νευρωνικών δικτύων. Παρουσιάζονται οι αρχιτεκτονικές των νευρωνικών δικτύων καθώς και πιο αναλυτικά τα είδη μάθησης.

Στο 4^ο κεφάλαιο γίνεται η περιγραφή και ο τρόπος εκπαίδευσης ενός δικτύου RBF. Επιπλέον παρουσιάζεται αλγόριθμος των ασαφών μέσων καθώς και ο k-means. Γίνεται επίσης ανάλυση των στατιστικών δεικτών που χρησιμοποιούνται για την αξιολόγηση των μεθόδων.

Στο 5^ο κεφάλαιο εξετάζονται τα γραμμικά και μη γραμμικά προβλήματα και αναλύεται η έννοια της βελτιστοποίησης στο RBF. Στη συνέχεια εισάγονται μέθοδοι βελτιστοποίησης Gauss-Newton, Levenberg-Marquardt, Variable Projection και η παραγοντοποίηση QR.

Στο 6^ο κεφάλαιο αναλύεται η επεξεργασία που χρειάζονται τα δεδομένα εισόδου και εφαρμόζεται η πειραματική διαδικασία με τις δύο μεθόδους εκπαίδευσης.

Στο 7^ο κεφάλαιο παρουσιάζονται τα αποτελέσματα των μεθόδων με τα σύνολα δεδομένων της βιβλιογραφίας σε σύγκριση με άλλες μεθοδολογίες. Τέλος, διατυπώνονται τα πορίσματα για τις υπό εξέταση τεχνικές εκπαίδευσης.

2 Τεχνητή Νοημοσύνη

Με την πάροδο του χρόνου υπήρξαν διάφορες προσεγγίσεις στον ορισμό της ‘Τεχνητής Νοημοσύνης’. Θεωρητικά ο 1^{ος} ορισμός διατυπώθηκε το 1956:

«Πρόκειται για τον τομέα που ασχολείται με την σχεδίαση ευφύων υπολογιστικών συστημάτων που επιδεικνύουν χαρακτηριστικά τα οποία σχετίζονται με την ανθρώπινη νοημοσύνη.» [32].

Τα αποτελέσματα από την αναζήτηση σε κάποια από τα πιο δημοφιλή λεξικά είναι:

«Είναι της κλάδος της επιστήμης των υπολογιστών που ασχολείται με την προσομοίωση της ευφούς συμπεριφοράς σε υπολογιστές» από το Merriam-Webster [33] και

«Η μελέτη του τρόπου παραγωγής μηχανών που έχουν κάποιες από τις ιδιότητες που έχει το ανθρώπινο μυαλό, της ικανότητας κατανόησης της γλώσσας, αναγνώρισης εικόνων, επίλυσης προβλημάτων και μάθησης» από το Cambridge [34].

Οι σύγχρονοι ορισμοί, εξηγούν με περισσότερη σαφήνεια την έννοια της τεχνητής νοημοσύνης αλλά χρειάζεται ακόμη αρκετή μελέτη για να γίνει απόλυτα κατανοητή.

Δύναται η νοημοσύνη να σχετίζεται άμεσα με την ανθρώπινη αλλά μπορεί να εντοπιστεί και στα ζώα. Ενδεικτικά, μπορεί ένα μυρμήγκι να μην είναι ‘έξυπνο’ αλλά μια αποικία μυρμηγκιών καταφέρνει να βρει την βέλτιστη διαδρομή από την τροφή μέχρι την φωλιά. Συνοψίζοντας, μια οντότητα θεωρείται ‘έξυπνη’ όταν έχει τρία χαρακτηριστικά:

1. Ικανότητα μάθησης με την οποία ανταλλάσσει δεδομένα με το περιβάλλον.
2. Λήψη λογικών αποφάσεων μέσα σε ένα αβέβαιο περιβάλλον.
3. Ικανότητα επίλυσης προβλημάτων βρίσκοντας την βέλτιστη λύση.

Ο Άλαν Τούρινγκ, ήταν ένας πρωτοπόρος μηχανικός ηλεκτρονικών υπολογιστών και θεωρείται ότι έθεσε τις βάσεις της τεχνητής νοημοσύνης. Έγινε γνωστός για το τεστ Τούρινγκ (1950) το οποίο προβληματίσε και ενέπνευσε πολλούς επιστήμονες και μηχανικούς. Έτσι επιδίωξαν τη σχεδίαση συστημάτων που αλληλοεπιδρούσαν με πιο φυσικό και ανθρώπινο τρόπο. Το τεστ βασίζεται σε μια απλή υπόθεση [35]:

Αν ένας άνθρωπος μπορεί να συμμετέχει σε μια συζήτηση για πέντε λεπτά χωρίς να καταλάβει ότι μιλάει σε μια μηχανή, τότε ο υπολογιστής περνάει το τεστ.

Με την πάροδο των ετών τα συστήματα έχουν εξελιχθεί, με αποτέλεσμα να έχει αυξηθεί αυτή η δυσκολία διάκρισης της μηχανής από τον άνθρωπο, καθώς έχουν αναπτυχθεί νέες τεχνικές.

2.1 Ιστορική αναδρομή

Ορισμένες σημαντικές ημερομηνίες στην εξέλιξη της τεχνητής νοημοσύνης [36], [37]:

- ➔ 384-322 π.Χ , Ο Αριστοτέλης στην αρχαία Ελλάδα όπου βρίσκεται στο σύγγραμμα του «Αναλυτικά Πρότερα», η Αριστοτέλεια συλλογιστική. Οι Αριστοτέλαιοι συλλογισμοί αποτελούσαν πρότυπα δομής επιχειρημάτων που κατέληγαν πάντα σε σωστά συμπεράσματα όταν ξεκινούσαν από σωστές υποθέσεις.
- ➔ 1854, Ο Τζορτζ Μπουλ παρουσιάζει την άλγεβρα boole, μέσα από την οποία μπορεί κάποιος να αναπαραστήσει οποιοδήποτε λογικό συλλογισμό με την ακολουθία 0/1.
- ➔ 1943, Οι McCulloch και Pitts δημοσίευσαν την 1^η εργασία , ένα μοντέλο τεχνητών νευρώνων που είχε την ικανότητα να μαθαίνει και να υπολογίζει μαθηματικές συναρτήσεις.
- ➔ 1950, Ο Alan Turing δημιουργεί το Turing τεστ.
- ➔ 1951, Οι Minsky και Edmonds , δύο μεταπτυχιακοί από το Princeton υλοποίησαν το πρώτο νευρωνικό δίκτυο SNARC. Χρησιμοποιήθηκαν 3000 λυχνίες και ένας μηχανισμός από αυτόματο πιλότο ,ενός παλιού βομβαρδιστικού, για να φτιάξουν μόλις 40 νευρώνες.
- ➔ 1956, Δημιουργήθηκε στην Μασαχουσέτη το πρώτο εργαστήριο και εκεί θεωρείται ότι γεννήθηκε το πεδίο της τεχνητής νοημοσύνης.
- ➔ 1958, Ο McCarthy δημιουργεί την γλώσσα προγραμματισμού LISP.
- ➔ 1960 – 1980, Έχουμε την ωρίμανση της τεχνητής νοημοσύνης. Αναπτύχθηκε ο νευρώνας Perceptron , φτιάχτηκαν τα έμπειρα συστήματα, δημιουργήθηκε η γλώσσα PROLOG , τα «Ασαφή Σύνολα» και «Ασαφείς Αλγόριθμοι»
- ➔ 1987, το 1^ο Διεθνές Συνέδριο για τα Νευρωνικά Δίκτυα από το Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (IEEE)
- ➔ 1992, το 1^ο Διεθνές Συνέδριο για τα Ασαφή Σύνολα από το IEEE
- ➔ 1990 - σήμερα, ξεκίνησε η εμπορευματοποίηση με υπολογιστικά συστήματα και μηχανές που προσαρμόζονται στο περιβάλλον τους αλλά και εφαρμογές που μαθαίνουν από την εμπειρία τους.

2.2 Εφαρμογές

Στον 21^ο πλέον αιώνα εντοπίζονται πολλές υλοποιήσεις και χρήσεις της τεχνητής νοημοσύνης από τις οποίες αρκετές όλοι χρησιμοποιούνε καθημερινά χωρίς να το γνωρίζουν. Ακόμη και αν στις επόμενες σελίδες γινόταν λεπτομερής αναφορά μόνο σε αυτές στο τέλος, πολλές από αυτές θα είχαν παραλειφθεί. Μερικά παραδείγματα [38]:

➤ Διαδίκτυο

Προσαρμογή των διαφημίσεων ανάλογα με αυτά που έχει αναζητήσει ο χρήστης ή προσαρμογή των τιμών σε μια ιστοσελίδα με βάση την επισκεψιμότητα του χρήστη.

➤ Οχήματα που παίρνουν αυτόνομα αποφάσεις.

Τεχνολογίες εντοπίζονται σε ένα σύγχρονο αυτοκίνητο, όπως έξυπνο φρενάρισμα και αυτόνομη οδήγηση με κάμερες και αισθητήρες.

➤ Ηλεκτρικές και ηλεκτρονικές συσκευές

Η αναγνώριση ενός ανθρώπου σε μια φωτογραφική μηχανή ή ακόμα και η αυτόματη προσαρμογή του φακού ανάλογα με το φως του περιβάλλοντος.

➤ Βέλτιστη δρομολόγηση αεροπλάνων

Λήψη αποφάσεων προσγείωσης ή απογείωσης αεροπλάνων σε ένα αεροδρόμιο, έξυπνη χρήση καυσίμων και αποφυγή εκατοντάδων ιπτάμενων πτηνών στον εναέριο χώρο.

➤ Συστήματα αναγνώρισης

Η οπτική αναγνώριση χαρακτήρων σε ένα έγγραφο ή σε συστήματα ασφαλείας για αναγνώριση του προσώπου, την ίριδας, των αποτυπωμάτων ή της υπογραφής [39].

➤ Συστήματα προβλέψεων

Σε οικονομικούς δείκτες, σε μετεωρολογικές υπηρεσίες για πρόβλεψη καταστροφών ή ακόμα σε προγράμματα που βοηθάνε επιχειρήσεις να πάρουν κρίσιμες αποφάσεις.

2.3 Προσεγγίσεις της Τεχνητής Νοημοσύνης

Η τεχνητή νοημοσύνη μπορεί να χωριστεί σε δύο προσεγγίσεις [9]:

❖ Κλασσική ή αλλιώς συμβολική τεχνητή νοημοσύνη

Ήταν μια αναπαράσταση της ανθρώπινης νοημοσύνης, χρησιμοποιώντας σύμβολα, με συστήματα και αλγορίθμους που βασίζόμενα στην γνώση. Διαμορφώθηκε χρησιμοποιώντας τις 2 γλώσσες προγραμματισμού, LISP και PROLOG. Στην αρχή ήταν μια πολύ καλή ιδέα, αλλά ήταν δύσκολη στην υλοποίηση καθώς προϋπόθετε γνώσεις προγραμματισμού.

❖ Υπολογιστική ή συνδετική ή μη-συμβολική Τεχνητή νοημοσύνη

Η προσπάθεια του ανθρώπου να μιμηθεί τους μηχανισμούς της φύσης, όπως την λειτουργία του ανθρώπινου εγκεφάλου ή την δημιουργία μιας αποικίας μυρμηγκιών. Ο στόχος είναι η προσομοίωση διαδικασιών με αλγορίθμους που θα μιμούνται την έξυπνη λειτουργία της φύσης. Η χρήση της είναι πλέον πιο διαδεδομένη σε σχέση με την προηγούμενη.

2.4 Μάθηση – Απόφαση - Βελτιστοποίηση

Ακολουθεί μια σύντομη αναφορά των τριών χαρακτηριστικών που πρέπει να έχει μια οντότητα για να είναι 'έξυπνη':

➤ Μάθηση

Μάθηση ονομάζεται η διαδικασία κατά την οποία ένα νευρωνικό δίκτυο τροποποιεί τις ελεύθερες παραμέτρους προκειμένου να βελτιωθεί η συμπεριφορά του και να αλληλεπιδράσει καλύτερα με το περιβάλλον του. Εν συντομία αναφέρονται μερικοί βασικοί τρόποι μάθησης [38]:

- Επιβλεπόμενη μάθηση
- Μάθηση με διόρθωση σφάλματος
- Μη επιβλεπόμενη μάθηση
- Ενισχυτική μάθηση

➤ Λήψη αποφάσεων

Στην λήψη αποφάσεων υπάρχουν οι μέθοδοι της Ασαφούς Λογικής (fuzzy logic) με τους οποίους γίνεται η μετάδοση της πληροφορίας σε μια μηχανή. Δηλαδή, ακολουθούνται λογικοί κανόνες που οδηγούν σε λογικά συμπεράσματα. Μεταξύ πολλών, χρησιμοποιείται στον αυτόματο έλεγχο με την αντικατάσταση των κλασσικών ελεγκτών PID, στις επιχειρηματικές αποφάσεις και στην αναγνώριση-ταξινόμηση προτύπων [40].

➤ Βελτιστοποίηση

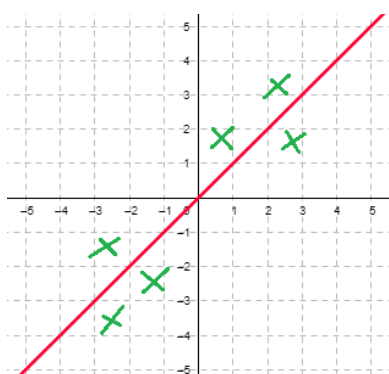
Στην βελτιστοποίηση εντοπίζονται οι εξελικτικοί αλγόριθμοι που μιμούνται τις διαδικασίες της φύσης για να βρουν την βέλτιστη λύση σε ένα πρόβλημα. Ένα βασικό παράδειγμα είναι τα μυρμήγκια στα οποία έχει αποδειχθεί ότι η διαδρομή που σχηματίζουν από την τροφή στην φωλιά τους είναι η βέλτιστη δυνατή. Οι εξελικτικοί αλγόριθμοι προσπαθούν να κωδικοποιήσουν τις πιθανές λύσεις του προβλήματος με έναν τρόπο παρόμοιο με το ανθρώπινο DNA. Έπειτα προσπαθούν να τις εξελίσουν ώστε να ανταποκρίνονται όσο καλύτερα γίνεται στο πρόβλημα προς επίλυση [41].

2.5 Αποθήκευση της γνώσης

Οι ελεύθερες παράμετροι, χρησιμοποιούνται για την αποθήκευση της γνώσης η οποία γίνεται μέσα από την διαδικασία της μάθησης [42].

Αρχικά αναζητείται μια μαθηματική σχέση που να ενώνει την είσοδο με την έξοδο, δηλαδή μια συνάρτηση (Σχήμα 1) σαν την :

$$y = ax \quad (2.1)$$

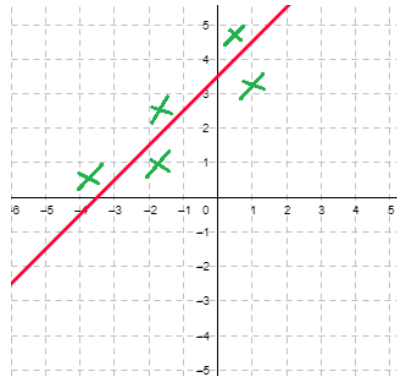


Σχήμα 1 Ευθεία $y=ax$

Η μάθηση της συμπεριφοράς του συστήματος ισοδυναμεί με τον υπολογισμό της ελεύθερης παραμέτρου a η οποία συνεπάγεται με την κλίση της ευθείας

Γίνεται αντιληπτό ότι δεν κατορθώνεται με μία ελεύθερη παράμετρο να περιγράψουν πλήρως τα δεδομένα, οπότε προσθέτετε ακόμα μία (Σχήμα 2):

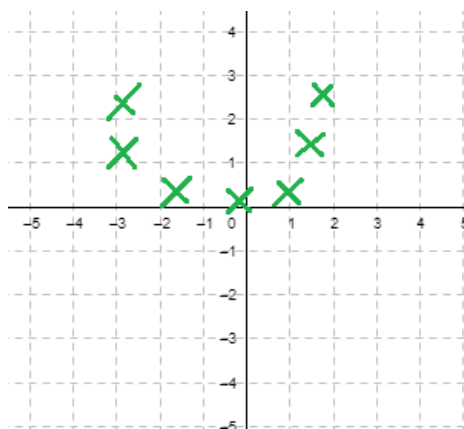
$$y = ax + b \quad (2.2)$$



Σχήμα 2 Ευθεία $y=ax+b$

Όμως για την περιγραφή των δεδομένων του (Σχήμα 3), χρειάζεται πλέον και τρίτη ελεύθερη παράμετρο, έτσι η ευθεία γίνεται καμπύλη με:

$$y = ax^2 + bx + c \quad (2.3)$$



Σχήμα 3 Καμπύλη $y=ax^2+bx+c$

Και στις τρεις προαναφερθείσες περιπτώσεις, πρέπει να βρεθούν οι τιμές των μεταβλητών για να περάσει η ευθεία μέσα από τα δεδομένα. Διακρίνεται ότι όσο περιπλοκότερη είναι η διάταξη των δεδομένων, τόσες περισσότερες παραμέτρους θα χρειαστούν. Παρόλα αυτά, για την εύρεση της καλύτερης ευθείας η οποία θα περνά μέσα από τα δεδομένα, θεωρήθηκε δεδομένη η μορφή της εξίσωσης. Στις περισσότερες περιπτώσεις δεν είναι δυνατόν να προσδιοριστεί η μορφή της εξίσωσης, ένα σημαντικό πρόβλημα καλείται να λύσει η μέθοδος των ελαχίστων τετραγώνων.

2.6 Μέθοδος Ελαχίστων Τετραγώνων

Η μέθοδος ελαχίστων τετραγώνων (Least Mean Square, LMS) αναπτύχθηκε από τους Widrow και Hoff το 1960 και ήταν ο πρώτος γραμμικός αλγόριθμος προσαρμοστικού φιλτραρίσματος για την επίλυση προβλημάτων όπως η πρόβλεψη. Αποτελεί ένα από τα θεμελιώδη εργαλεία που έχει στην διάθεση του ένας μηχανικός και θεωρείται η πιο απλή μέθοδος μηχανικής μάθησης [43].

Αν θεωρηθεί ότι η πρόβλεψη της ευθείας είναι η \bar{y} . Η διαφορά της κανονικής ευθείας από την πρόβλεψη θα είναι το σφάλμα:

$$\varepsilon = |y - \bar{y}| = |y - (ax + b)| \quad (2.4)$$

Γίνεται χρήση της απόλυτης τιμής για την απαλοιφή του προσήμου, δηλαδή αν είναι πάνω ή κάτω από την ευθεία. Με αυτόν τον τρόπο δημιουργείται το πρόβλημα της ασυνέχειας και για αυτόν τον λόγο υψώνεται στο τετράγωνο:

$$\varepsilon^2 = (y - \bar{y})^2 = (y - (ax + b))^2 \quad (2.5)$$

Τελικά το άθροισμα των τετραγώνων των σφαλμάτων θα είναι:

$$SSE = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2 \quad (2.6)$$

οπου το n είναι το πλήθος των δεδομένων.

Το ιδανικό σφάλμα είναι μηδενικό, αυτό θα σήμαινε ότι η ευθεία περνάει ακριβώς μέσα από τα δεδομένα, πράγμα το οποίο συνήθως δεν μπορεί να γίνει και γίνεται συμβιβασμός στον οποίο υπολογίζονται τα a,b όπου ελαχιστοποιούν την ευθεία.

Αρχικά βρίσκονται οι 2 μερικές παραγώγους οι οποίες θα πρέπει να μηδενίζονται για να υπάρξει ακρότατο.

$$\frac{\partial SSE}{\partial a} = 0, \frac{\partial SSE}{\partial b} = 0 \quad (2.7)$$

και έπειτα από πράξεις έχει αποδειχθεί ότι

$$\alpha = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2} \quad (2.8)$$

και

$$b = \bar{y} - \alpha \bar{x} \quad (2.9)$$

Καθώς ανεβαίνει η τάξη του συστήματος, αυξάνεται η πολυπλοκότητα οπότε ενδείκνυται η ομαδοποίηση των δεδομένων με την χρήση πινάκων. Για παράδειγμα, έστω 3 μεταβλητές οι

οποίες θα είναι οι στήλες και έναν μεγάλο αριθμό δεδομένων που θα είναι οι γραμμές. Έτσι η είσοδος θα είναι:

$$x = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} \end{bmatrix} \quad (2.10)$$

Και η έξοδος:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \quad (2.11)$$

με τρεις κλίσεις των ευθειών:

$$\alpha = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad (2.12)$$

Τελικά

$$x \cdot \alpha = \begin{bmatrix} x_{11} \cdot a_1 & x_{12} \cdot a_2 & x_{13} \cdot a_3 \\ x_{21} \cdot a_1 & x_{22} \cdot a_2 & x_{23} \cdot a_3 \\ \vdots & \vdots & \vdots \\ x_{k1} \cdot a_1 & x_{k2} \cdot a_2 & x_{k3} \cdot a_3 \end{bmatrix} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_k \end{bmatrix} = \bar{y} \quad (2.13)$$

Παρατηρείτε ότι η κάθε γραμμή δίνει την πρόβλεψη της ευθείας για την αντίστοιχη μέτρηση.

Όπως αναφέρθηκε και νωρίτερα, πρέπει $Y = \bar{Y}$ η αλλιώς η λύση της εξίσωσης:

$$Y = X \cdot A \quad (2.14)$$

ως προς το A. Εδώ χρειάζονται πράξεις πινάκων και θα γινόταν:

$$X^{-1} \cdot Y = A \quad (2.15)$$

κάτι το οποίο δεν μπορεί να γίνει επειδή ο X δεν είναι τετραγωνικός για να έχει αντίστροφο.

Η λύση είναι ένας έξυπνος τρόπος με τον οποίο ο πίνακας X γίνεται τετραγωνικός.

$$X' \cdot Y = (X' \cdot X) \cdot A \Rightarrow (X' \cdot X)^{-1} \cdot Y = A \quad (2.16)$$

Αλλά για να υπάρχει ο αντίστροφος πίνακας $X' \cdot X$, θα πρέπει η ορίζουσα του να μην είναι μηδέν και να μην υπάρχουν γραμμικά εξαρτημένες γραμμές ή στήλες. Στην περίπτωση που δεν έχει αντίστροφο είναι αδύνατη η χρήση της μεθόδου LMS. Αυτό πρακτικά σημαίνει ότι υπάρχουν άπειρες καλύτερες ευθείες από τις οποίες καμία δεν κάνει.

Επίσης ο πίνακας $X' \cdot X$ ονομάζεται πίνακας συν-διασποράς και περιέχει πληροφορίες για την στατιστική του συστήματος δείχνοντάς πως σχετίζονται μεταξύ τους οι μεταβλητές.

3 Νευρωνικά δίκτυα

Η χρήση των ελαχίστων τετραγώνων για το μετασχηματισμό της πληροφορίας εφαρμόζεται μόνο εάν είναι γνωστός ο τρόπος με τον οποίο σχετίζονται η είσοδος και η έξοδος. Συνήθως, ο τρόπος συσχέτισης είναι άγνωστος και γι' αυτό εφαρμόζονται τα νευρωνικά δίκτυα [44], τα οποία προσπαθούν να μιμηθούν την λειτουργία του ανθρώπινου εγκεφάλου

3.1 Η εξέλιξη της τεχνολογίας

Σύμφωνα με τον γνωστό νόμο του Moore, ο αριθμός των τρανζίστορ που χωράνε σε ένα ολοκληρωμένο κύκλωμα, διπλασιάζεται κάθε 2 χρόνια. Η αύξηση της πυκνότητας των τρανζίστορ αλλά και η ταυτόχρονη παράλληλη σύνδεση πολλών επεξεργαστών έχουν μεγιστοποιήσει την υπολογιστική ισχύ. Η ισχύς αυτή μετριέται με βάση την ταχύτητα που εκτελούνται μαθηματικές πράξεις με μονάδα μέτρησης τα FLOPS (Floating Points Operation Per Second) [45]. Ένας σημερινός υπέρ-υπολογιστής εκτελεί υπολογισμούς γρηγορότερα από τον ανθρώπινο εγκέφαλο ενώ ταυτόχρονα αποθηκεύει περισσότερες πληροφορίες.

Παρά την εξέλιξη της τεχνολογίας, ο ανθρώπινος εγκέφαλος τα καταφέρνει καλύτερα παρά την μικρότερη ταχύτητα επεξεργασίας. Αυτό οφείλεται στο γεγονός ότι ένας υπολογιστής λειτουργεί σειριακά, ενώ ο εγκέφαλος αποτελείται από νευρώνες που συνδέονται παράλληλα, αλλά και μεταξύ τους με υψηλή συνδεσιμότητα, καθιστώντας τον εγκέφαλο να λειτουργεί ως ένα ενιαίο κομμάτι [44].

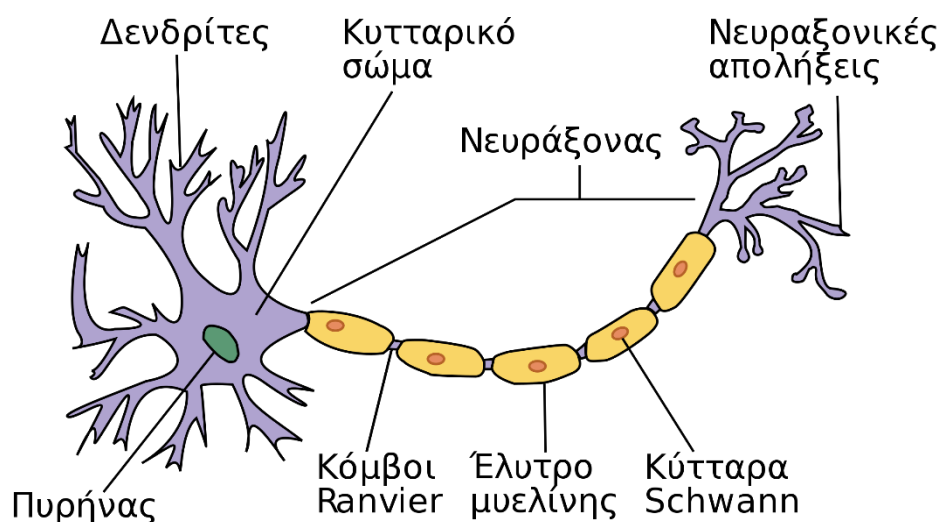
Τα τεχνητά Νευρωνικά δίκτυα είναι μια προσέγγιση της δομής και λειτουργικότητας του εγκεφάλου, για αυτό και το βασικότερο στοιχείο τους είναι οι νευρώνες. Αυτά τα τεχνητά δίκτυα χρησιμοποιούν απλοποιημένα μοντέλα νευρώνων τα οποία έχουν ελάχιστη σχέση με τα βιολογικά. Ωστόσο, ακόμη και αυτά τα απλά μοντέλα είναι σε θέση να σχηματίσουν ενδιαφέροντα δίκτυα εφόσον έχουν δυο ιδιότητες [44].

- ✓ Διαθέτουν μεταβαλλόμενες παραμέτρους ώστε να διευκολύνεται η διαδικασία της μάθησης, η ιδιότητα αυτή καλείται ως πλαστικότητα των νευρώνων.
- ✓ Έχουν έναν μεγάλο αριθμό νευρώνων ώστε να εξασφαλιστεί ο παραλληλισμός της επεξεργασίας και της κατανομής της πληροφορίας.

Το σημαντικότερο πρόβλημα που καλείται να λύσει ο άνθρωπος είναι η εύρεση κατάλληλων αλγόριθμων εκπαίδευσης των δικτύων ώστε να παρομοιάζονται με έξυπνες διαδικασίες. Για να γίνει αυτό, είναι σημαντικό η επιλογή του καταλληλότερου αλγορίθμου εκπαίδευσης. Δηλαδή, αν η εκπαίδευση θα είναι εποπτευόμενη ή αν θα είναι αυτό-οργανωμένη με χρήση κριτηρίων τερματισμού.

3.2 Ο νευρώνας

Ο νευρώνας (Σχήμα 4) είναι ένα από τα βασικότερα στοιχεία στον εγκέφαλο είτε αυτός είναι ενός ανθρώπου ή ενός ζώου. Αποτελείται από τον πυρήνα, τον άξονα, τις συνάψεις, τους δενδρίτες και όλα αυτά μαζί δημιουργούν ένα νευρωνικό δίκτυο [38].



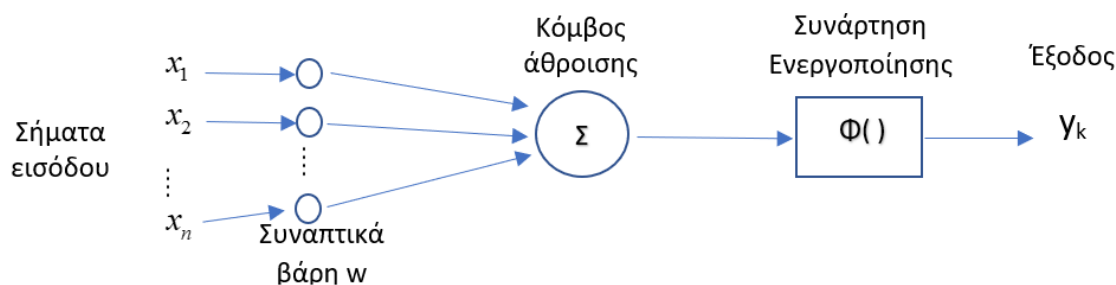
Σχήμα 4 Ο νευρώνας

Ανάμεσα από τις νευροαξονικές απολήξεις ενός νευρώνα και των δενδρίτων του επομένου, υπάρχει μια σύναψη. Ο ηλεκτρικός αυτός παλμός που φτάνει μέσω του νευράξονα, προκαλεί την έκκριση μια χημικής ουσίας που λέγεται διαβιβαστής. Αυτός μεταδίδεται σε υποδοχείς που υπάρχουν στους επόμενους δενδρίτες και έτσι προκαλείται ένα ηλεκτρικό σήμα προς αυτούς. Μετά την συλλογή των σημάτων από τις απολήξεις των προηγούμενων, γίνεται μια ενίσχυση ή υποβάθμιση αναλόγως με την ισχύς της κάθε σύναψης. Με αυτά τα σήματα ο νευρώνας μπορεί να ενεργοποιηθεί ή και όχι. Αν ενεργοποιηθεί τότε το σώμα παρέχει την ενέργεια για την πυροδότηση ενός παλμού, ο οποίος διατρέχει τον άξονα και μεταφέρεται προς τους υπόλοιπους νευρώνες.

Έτσι ο κάθε νευρώνας μέσω των δενδριτών συλλέγει ηλεκτροχημικά σήματα εισόδου. Η πηγή ενέργειας του νευρώνα βρίσκεται στο σώμα και ο άξονας είναι ο δρόμος για την μετάδοση των ηλεκτρικών σημάτων. Η ικανότητα μάθησης προκύπτει σαν αποτέλεσμα εναλλαγής των συνάψεων και η γνώση αποθηκεύεται στην ισχύ των συνάψεων.

3.2.1 Μοντέλο Νευρώνα McCulloch-Pitts

Στα μέσα του 19^{ου} αιώνα οι McCulloch και Pitts όρισαν ένα απλοϊκό μοντέλο για την περιγραφή της νευρωνικής δραστηριότητας. Η κατάσταση του προσδιορίζεται από έναν δυαδικό αριθμό, αν ο νευρώνας είναι ανενεργός έχει έξοδο $y=0$ ενώ αν πυροδοτεί παλμούς έχει έξοδο $y=1$. Συνεχίζοντας, η αντιστοιχισή των συνάψεων γίνεται με τα συνοπτικά βάρη w_i όπου αποτελούνται από θετικούς πραγματικούς αριθμούς όταν είναι για ενίσχυση, ενώ αρνητικούς για απόσβεση [38]:



Σχήμα 5 Τεχνητός νευρώνας

Ο τεχνητός νευρώνας (Σχήμα 5) αποτελείται από [10]:

- Τις συνάψεις από τις οποίες η καθεμία χαρακτηρίζεται από ένα βάρος w . Το κάθε σήμα εισόδου που φτάνει στην σύναψη πολλαπλασιάζεται με το συνταπτικό βάρος και ενισχύεται ή αποσβένεται.
- Έναν αθροιστή ο οποίος προσθέτει τα σταθμισμένα σήματα εισόδου με έναν γραμμικό τρόπο.
- Μία συνάρτηση ενεργοποίησης, η οποία περιορίζει το πλάτος του σήματος εξόδου ανάμεσα σε ένα εύρος τιμών.

3.2.2 Λειτουργία του μοντέλου

Έστω οι είσοδοι του νευρώνα x_1, x_2, \dots, x_n και μετά τα βάρη w_1, w_2, \dots, w_n , η έξοδος στον αθροιστή θα είναι $w_1 \cdot x_1 + \dots + w_n \cdot x_n$. Στη συνέχεια ελέγχεται αν αυτό το άθροισμα υπερβαίνει ένα κατώφλι θ που έχει οριστεί έτσι ώστε να παρθεί η απόφαση αν θα ενεργοποιηθεί ο νευρώνας ή αν θα παραμείνει αδρανής [38], δηλαδή :

$$u = \sum_{i=1}^n w_i x_i - \theta \quad (3.1)$$

και μόνο αν $u > 0$ ο νευρώνας θα έχει έξοδο:

$$y = f(u) = 1 \quad (3.2)$$

Η διέγερση u του νευρώνα προκύπτει από τις εισόδους. Μία από τις πιο γνωστές συναρτήσεις ενεργοποίησης που χρησιμοποιούνται είναι η βηματική 0 ή 1.

Επίσης σε μορφή διανυσμάτων γράφεται ως

$$u = \bar{w}^T \bar{x} - \theta \quad (3.3)$$

όπου τα συνοπτικά βάρη $\bar{w} = [w_1, \dots, w_n]^T$ και το διάνυσμα εισόδου $\bar{x} = [x_1, \dots, x_n]^T$. Όσον αφορά για το κατώφλι θ μπορεί να είναι ένας αρνητικός ή θετικός πραγματικός αριθμός, όπως και τα συνοπτικά βάρη.

Τέλος, γίνεται η πόλωση της συνάρτησης για να έχει μια ελευθερία, έτσι ώστε να μην περνάει από την αρχή των αξόνων προσθέτοντας $x_0 = 1$ και τελικά το επιπλέον συναπτικό βάρος $w_0 = -\theta$. Επομένως η (2.1) γίνεται [\[38\]](#)

$$u = \sum_{i=1}^n w_i x_i + x_0 w_0 = \sum_{i=0}^n w_i x_i \quad (3.4)$$

με

$$w = [w_0, w_1, \dots, w_n]^T \quad (3.5)$$

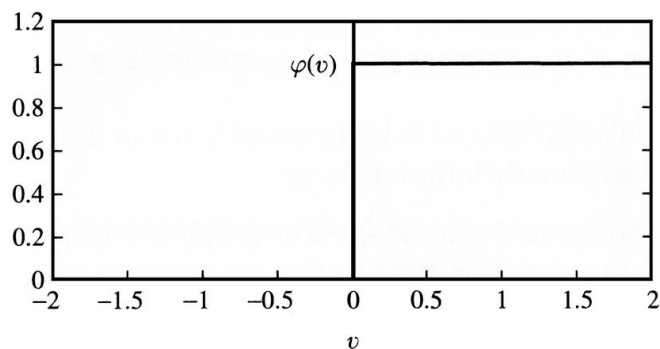
και

$$x = [x_0, x_1, \dots, x_n]^T \quad (3.6)$$

3.2.3 Συναρτήσεις ενεργοποίησης

Σημειώθηκε νωρίτερα ότι η συνάρτηση ενεργοποίησης δέχεται στην είσοδο το άθροισμα των σταθμισμένων σημάτων και το μετασχηματίζει. Υπάρχουν αρκετές συναρτήσεις ενεργοποίησης που μπορούν να χρησιμοποιηθούν στον νευρώνα, με την αναφορά των πιο συνηθισμένων [\[44\]](#) [\[45\]](#):

❖ Βηματική συνάρτηση

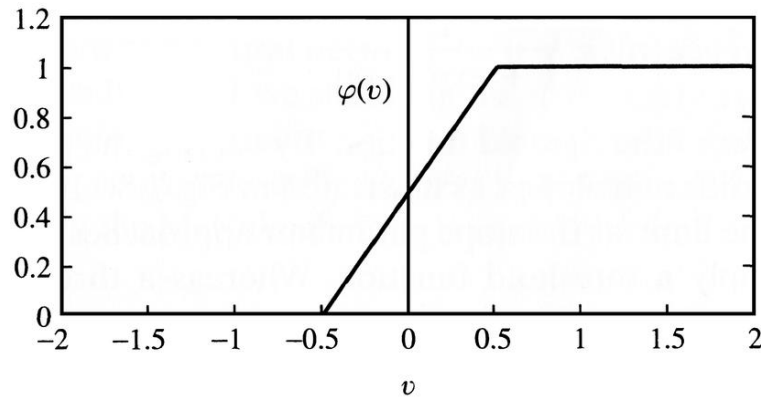


Σχήμα 6 Βηματική συνάρτηση

Αν το δυναμικό ενεργοποίησης είναι θετικό, τότε η έξοδος του νευρώνα γίνεται ίση με 1 ενώ διαφορετικά 0. Ένα μειονέκτημα είναι ότι η παράγωγος της βηματικής είναι μηδέν. (Σχήμα 6)

$$\varphi(u) = \begin{cases} 1 & \text{αν } u \geq 0 \\ 0 & \text{αν } u < 0 \end{cases} \quad (3.7)$$

❖ Γραμμική συνάρτηση

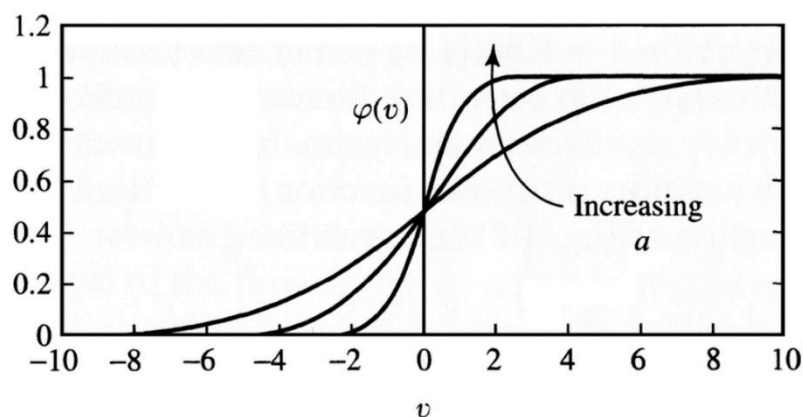


Σχήμα 7 Γραμμική συνάρτηση

Γίνεται πολλαπλασιασμός με μια σταθερά ή αλλιώς έναν γραμμικό ενισχυτή. (Σχήμα 7)

$$\varphi(u) = \begin{cases} 1 & \text{για } u \geq \frac{1}{2} \\ u & \text{για } \frac{1}{2} > u > -\frac{1}{2} \\ 0 & \text{για } u \leq -\frac{1}{2} \end{cases} \quad (3.8)$$

❖ Σιγμοειδής συνάρτηση



Σχήμα 8 Σιγμοειδής συνάρτηση

Ονομάζονται σιγμοειδής επειδή η μορφή τους μοιάζει με ένα τελικό σίγμα. Ένα πλεονέκτημα είναι ότι είναι συνεχής και παραγωγίσιμη. Επίσης για πολύ μεγάλα a γίνεται συνάρτηση κατωφλίου (Σχήμα 8).

$$\varphi(u) = \frac{1}{1 + \exp(-a \cdot u)} \quad (3.9)$$

Στα τεχνητά νευρωνικά δίκτυα συνηθίζεται να χρησιμοποιείται η σιγμοειδής συνάρτηση γιατί δεν έχει πρόβλημα στην παραγωγή της.

3.3 Αρχιτεκτονικές τεχνητών νευρωνικών δικτύων

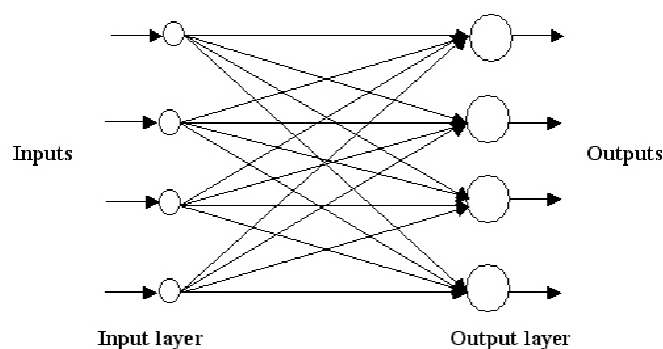
Υπάρχουν διάφορες αρχιτεκτονικές τεχνητών νευρωνικών δικτύων, δηλαδή τρόποι διασύνδεσης των νευρώνων που συνδυάζονται με τις συναρτήσεις ενεργοποίησης. Ανάλογα το πρόβλημα που υπάρχει προς λύση γίνεται και επιλογή της αντίστοιχης αρχιτεκτονικής. Παρακάτω γίνεται αναφορά μερικών αρχιτεκτονικών οι οποίες καθορίζονται από το είδος των νευρώνων και τον τρόπο που συνδέονται μεταξύ τους. Επίσης, σε κάθε περίπτωση διαφοροποιείται και ο τρόπος αποθήκευσης της πληροφορίας [10].

3.3.1 Νευρωνικά δίκτυα πρόσθιας τροφοδότησης

Χωρίζονται σε δύο κατηγορίες, σε δίκτυα μιας στοιβάδας και σε δίκτυα πολλαπλών στοιβάδων [10].

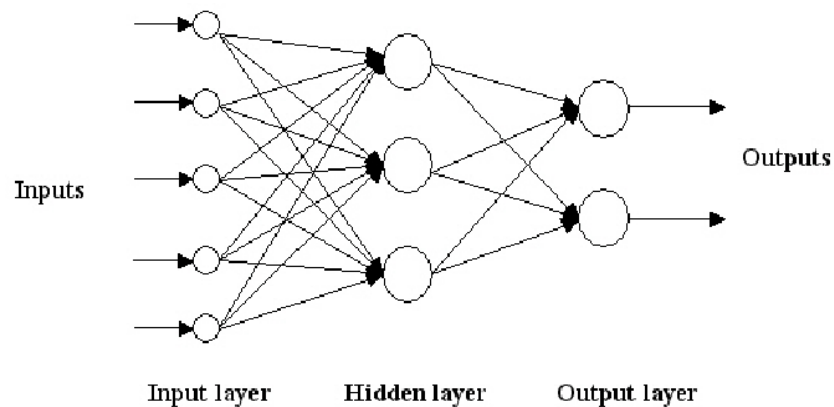
❖ Το απλό δίκτυο μιας στοιβάδας

Τα σήματα μεταδίδονται από την είσοδο στην έξοδο και μόνο προς τη συγκεκριμένη κατεύθυνση. (Σχήμα 9).



Σχήμα 9 Δίκτυο με 1 στοιβάδα

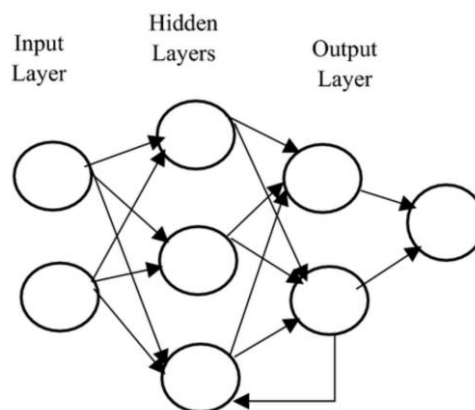
❖ Δίκτυα πολλών στοιβάδων



Σχήμα 10 Δίκτυο πολλών στοιβάδων

Αρχικά έχουν περισσότερες από μία στοιβάδες (Σχήμα 10). Οι στοιβάδες που υπάρχουν ανάμεσα από τις στοιβάδες εισόδου και εξόδου ονομάζονται κρυφές στοιβάδες επειδή δεν είναι άμεσα ορατές από την είσοδο και την έξοδο. Επίσης χαρακτηρίζεται ως μερικώς συνδεδεμένος αν λείπουν κάποιες συνάψεις ή πλήρως συνδεδεμένο αν κάθε κόμβος συνδέεται με όλους τις επομένης. Η ικανότητα εκμάθησης αυξάνεται προσθέτοντας κρυφές στοιβάδες και χρησιμοποιούνται περισσότερο για προβλήματα προσέγγισης συναρτήσεων ή για προβλήματα ταξινόμησης προτύπων [10].

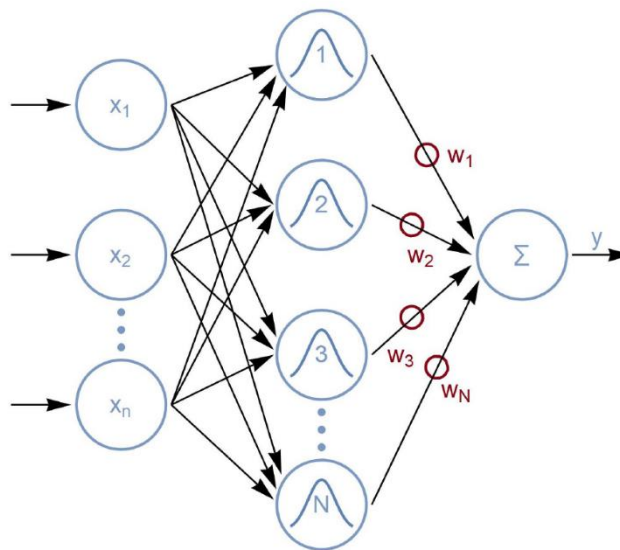
3.3.2 Αναδρομικά νευρωνικά δίκτυα



Σχήμα 11 Δίκτυο με ανάδραση

Στα δίκτυα αυτά (Σχήμα 11) υπάρχει τουλάχιστον ένας βρόγχος ανατροφοδότησης, ο οποίος διαφοροποιεί την διαδικασία εκμάθησης αλλά και της απόδοσης. Λόγω της ανατροφοδότησης παίζει μεγάλο ρόλο η ευστάθεια του συστήματος και πρέπει να σχεδιαστεί με προσοχή. Αυτό το είδος χρησιμοποιείται για μοντελοποίηση δυναμικών συστημάτων [10].

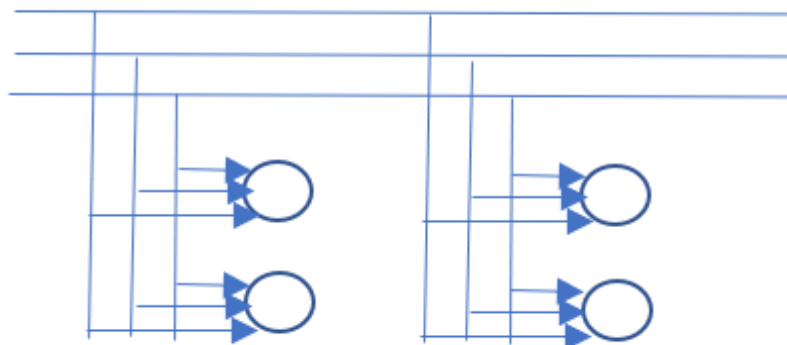
3.3.3 Νευρωνικά δίκτυα ακτινικής συνάρτησης βάσης



Σχήμα 12 Δομή ενός δικτύου RBF

Είναι παρόμοια με τα δίκτυα εμπρόσθιας τροφοδότησης έχοντας μόνο μία κρυφή στοιβάδα και η στοιβάδα εξόδου είναι και κόμβοι άθροισης (Σχήμα 12). Όπως υποδηλώνει το όνομά τους, αξιοποιούν μια συνάρτηση ενεργοποίησης ακτινικής βάσης, η οποία είναι πολύ σημαντική. Ο λόγος είναι επειδή η γνώση αποθηκεύεται και στις παραμέτρους της συνάρτησης εκτός από τις συνάψεις. Χρησιμοποιούνται σε προβλήματα ταξινόμησης προτύπων και ταξινόμησης συναρτήσεων [\[9\]](#), [\[10\]](#), [\[55\]](#).

3.3.4 Αυτό-οργανώμενα νευρωνικά δίκτυα



Σχήμα 13 Αυτό-οργανώμενα νευρωνικά δίκτυα

Διαφορετικός τρόπος προσέγγισης των νευρωνικών δικτύων (Σχήμα 13). Συγκεκριμένα ο κάθε νευρώνας αντιπροσωπεύει μια έξοδο στο δίκτυο. Για αυτόν τον λόγο οι νευρώνες διαμορφώνονται κατά την διάρκεια της εκπαίδευσης. Χρησιμοποιούνται για την επίλυση προβλημάτων αναγνώρισης προτύπων και προβλήματα ομαδοποίησης [\[9\]](#).

3.4 Perceptron ενός και περισσότερων στοιβάδων (MLP)

Το Perceptron είναι ευρέως διαδεδομένο σε όσους ασχολούνται με τα νευρωνικά δίκτυα. Αποτελείται από μόνο έναν νευρώνα και θεωρείται το πιο απλό είδος νευρωνικού δικτύου. Η συνάρτηση ενεργοποίησης είναι η κατωφλίου που σημαίνει ότι δεν παραγωγίζεται λόγω της ασυνέχειας της. Επίσης υπάρχει ένας σταθερός όρος, ο οποίος ονομάζεται πόλωση και εισέρχεται στον κόμβο άθροισης και δίνει μια ελευθερία στην προσεγγιστική συνάρτηση. Μπορεί να ταξινομήσει πρότυπα ανάμεσα σε 2 κλάσεις αρκεί αυτές να είναι γραμμικά διαχωρίσιμες.

Αν οι κλάσεις δεν είναι γραμμικά διαχωρίσιμες τότε η ταξινόμηση γίνεται πιο πολύπλοκη και χρειάζεται περισσότερα δίκτυα Perceptron. Θεωρητικά υπάρχουν περισσότερες μεταβλητές εισόδου και το όριο απόφασης θα πρέπει να σχηματιστεί σε ένα πολυδιάστατο χώρο. Τα δίκτυα με τα περισσότερα Perceptron θα αποτελούνται από πολλές στοιβάδες και χρησιμοποιούνται συνήθως σε δίκτυα με πρόσθια τροφοδότηση.

Με βάση του θεωρήματος Kolmogorov, έχει την δυνατότητα με την κατάλληλη διαδικασία εκπαίδευσης να μάθει οποιαδήποτε συμπεριφορά ενός συστήματος [38], [44].

3.5 Μηχανική Μάθηση

Μηχανική Μάθηση ονομάζεται η διαδικασία κατά την οποία ένα νευρωνικό δίκτυο τροποποιεί τις ελεύθερες παραμέτρους προκειμένου να βελτιωθεί η συμπεριφορά του και να αλληλεπιδράσει καλύτερα με το περιβάλλον του. Σημαντικό εδώ είναι η μελέτη των χρονοσειρών, δεδομένα με χρονικές πληροφορίες, οι οποίες χρησιμοποιούνται για την πρόβλεψη.

Αρχικά, η μηχανή μαθαίνει συλλέγοντας γνώση από τα δεδομένα, ψάχνοντας κάποιο μοτίβο το οποίο στην συνέχεια το χρησιμοποιεί για να προβλέψει κάτι καινούριο σε καινούρια δεδομένα. Τα δεδομένα συλλέγονται από αισθητήρες και η γνώση αποθηκεύεται σε ελεύθερες παραμέτρους.

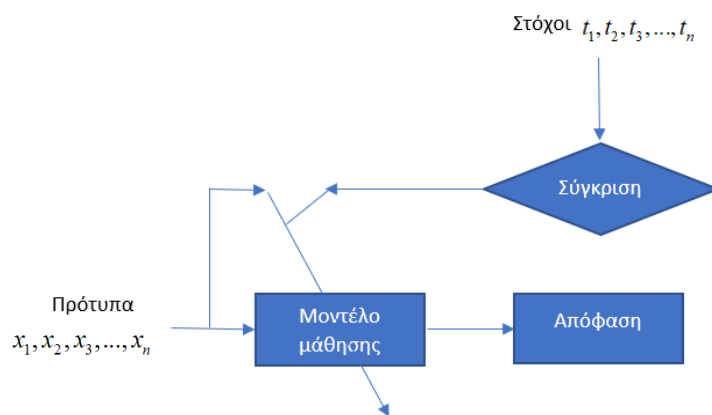
Επίσης η μάθηση μπορεί να είναι σε πραγματικό χρόνο και να γίνεται κάθε φορά που έρχεται ένα νέο δεδομένο ή να γίνει μόνο μια φορά για ένα συγκεκριμένο σύνολο δεδομένων. Όπως θεωρείται και λογικό, η μάθηση σε πραγματικό χρόνο είναι πιο δύσκολη να υλοποιηθεί.

Υπάρχουν διάφοροι αλγόριθμοι μάθησης, μερικοί από αυτούς είναι [38]:

- **Μάθηση με επίβλεψη**

Υπάρχει ένας εκπαιδευτής ο οποίος γνωρίζει για κάθε δεδομένο εισόδου ποιο θα είναι το αντίστοιχο δεδομένο εξόδου. Με μια επαναληπτική διαδικασία, η γνώση του εκπαιδευτή μεταφέρεται στο νευρωνικό δίκτυο, δηλαδή στις ελεύθερες παραμέτρους του. Στο τέλος της εκπαίδευσης αποσυνδέεται ο εκπαιδευτής και πλέον είναι ικανό να κάνει τις προβλέψεις του. Ο λόγος της αναγνώρισης της οφείλεται στην ευκολία που την διέπει γιατί δεν είναι απαραίτητη περεταίρω γνώση του άγνωστου συστήματος.

Το μοντέλο (Σχήμα 14) χρησιμοποιεί μια σειρά από πρότυπα εισόδου $x_1, x_2, x_3, \dots, x_n$ μαζί με τους αντίστοιχους στόχους $t_1, t_2, t_3, \dots, t_n$



Σχήμα 14 Μάθηση με επίβλεψη

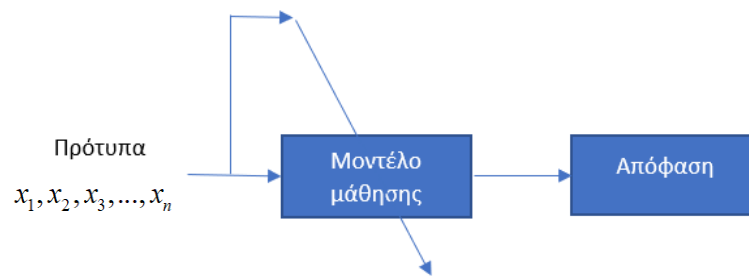
Τα προβλήματα μάθησης με επίβλεψη χωρίζονται σε 2 κατηγορίες :

- ❖ Προβλήματα ταξινόμησης προτύπων στα οποία γίνεται προσπάθεια ταξινόμησης των δεδομένων σε κλάσεις και οι στόχοι είναι σε διακριτές τιμές
- ❖ Προβλήματα παλινδρόμησης τα οποία προσπαθούν να προσεγγίσουν μια συνάρτηση η οποία συνδέει την έξοδο με την είσοδο σε ένα σύστημα, οι στόχοι εδώ είναι συνεχείς τιμές ή απεριόριστο πλήθος από διακριτές τιμές.

- **Μάθηση χωρίς επίβλεψη**

Ο εκπαιδευτής απουσιάζει επειδή δεν υπάρχουν δεδομένα εισόδου. Το νευρωνικό δίκτυο προσπαθεί να ανακαλύψει κάποια στατιστικά των δεδομένων εισόδου. Κάθε νευρώνας ανταγωνίζεται τον άλλον προσπαθώντας να τοποθετηθεί όσο καλύτερα γίνεται μέσα στο σύστημα. Μετά το τέλος της εκπαίδευσης, μόνο ένας νευρώνας είναι ο νικητής, για αυτόν τον λόγο χρησιμοποιείται στα αυτό-οργανωμένα δίκτυα.

Το μοντέλο (Σχήμα 15) κάνει χρήση των προτύπων εισόδου $x_1, x_2, x_3, \dots, x_n$ χωρίς όμως να έχει πληροφορίες σχετικά με τους στόχους.

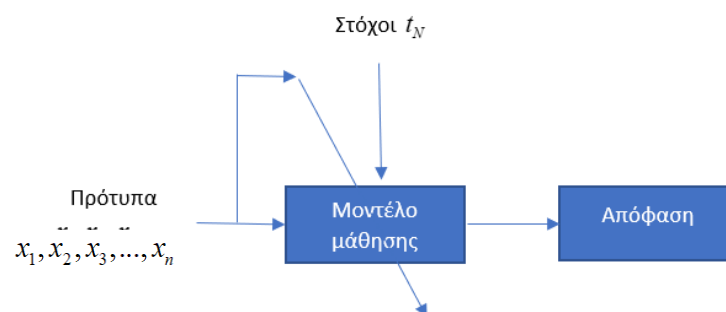


Σχήμα 15 Μάθηση χωρίς επίβλεψη

- **Μάθηση με ενίσχυση**

Η μάθηση γίνεται με έναν κριτή ο οποίος τιμωρεί ή επιβραβεύει την συμπεριφορά του συστήματος. Ο κριτής τροφοδοτείται με δεδομένα εισόδου από το περιβάλλον, επομένως υπάρχει ανατροφοδότηση. Μετά από αρκετές επαναλήψεις το σύστημα σταθεροποιείται και ο κριτής σταματάει να κάνει αλλαγές στις συνάψεις. Η χρήση των δικτύων γίνεται στην εκπαίδευση παλινδρομικών νευρωνικών δικτύων και στην επίλυση προβλημάτων δυναμικού προγραμματισμού.

Σε αυτήν την περίπτωση (Σχήμα 16), θα χρησιμοποιηθεί μια ακολουθία προτύπων εισόδου $x_1, x_2, x_3, \dots, x_n$ και οι στόχοι θα είναι τιμές ανταμοιβής ή τιμωρίας



Σχήμα 16 Μάθηση με ενίσχυση

4 Δίκτυα συνάρτησης ακτινικής βάσης (RBF)

Τα δίκτυα συνάρτησης βάσης χρησιμοποιούνται σε ευρεία κλίμακα και αποτελούν ενδιαφέρον για την ερευνητική κοινότητα. Είναι ένας ειδικός τύπος των δικτύων με εμπρόσθια τροφοδότηση. Ο λόγος που τα καθιστά ευρέως γνωστά είναι η απλή δομή τους, έχουν μόνο ένα κρυφό επίπεδο, με συνέπεια την ταχύτερη εκμάθηση των αλγορίθμων. Έχουν καταφέρει να εξελιχθούν σε ισχυρά υπολογιστικά εργαλεία, προσφέροντας πολύ ανταγωνιστικά αποτελέσματα συγκριτικά με άλλες τεχνικές μηχανικής μάθησης [46].

4.1 Σχεδιασμός δικτύου RBF

Η επιλογή των παραμέτρων του δικτύου RBF είναι το πιο κρίσιμο ζήτημα που πρέπει να αντιμετωπιστεί. Η αρχική επιλογή των κέντρων ήταν απλή, το κάθε δεδομένο εκπαίδευσης αποτελούσε και ένα κέντρο του δικτύου. Τα προβλήματα που δημιουργούνται όμως από αυτή τη τεχνική είναι να υπάρξει κίνδυνος υπερπροσαρμογής (overfitting) ή και προβλήματα μνήμης, κυρίως σε μεγάλα σύνολα δεδομένων [11]. Για να αποφευχθούν τα παραπάνω, θα μπορούσε να χρησιμοποιηθεί μια μέθοδος βελτιστοποίησης, όπως συνήθως γίνεται στα δίκτυα MLP, αλλά εδώ υπάρχει η δυνατότητα της διάσπασης του προβλήματος εκπαίδευσης σε δύο φάσεις [9].

- 1^η φάση: Καθορισμός αριθμού και θέσεων κέντρων

Με την χρήση αλγορίθμων χωρίς επίβλεψη, μέσα από τα δεδομένα εισόδου, καθορίζονται ο αριθμός και οι θέσεις των κέντρων. Μια επιπλέον πληροφορία που μπορεί να χρειαστεί να υπολογισθεί είναι το εύρος των κέντρων, αν χρησιμοποιηθεί Gaussian συνάρτηση ενεργοποίησης. Συνήθως εφαρμόζεται ο αλγόριθμος k-means ή ο αλγόριθμος Fuzzy Means(FM).

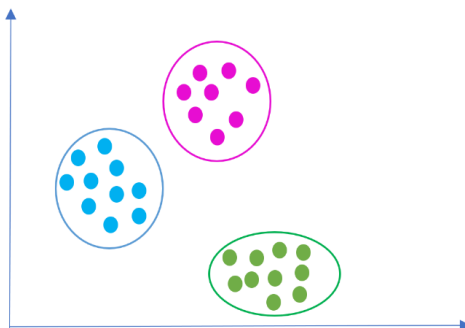
- 2^η φάση: Καθορισμός συναπτικών βαρών

Η σχέση ανάμεσα στην έξοδο του δικτύου και στην έξοδο των νευρώνων της κρυφής στοιβάδας είναι γραμμική. Για αυτόν τον λόγο χρησιμοποιείται η μέθοδος των ελαχίστων τετραγώνων.

Η επιλογή του μεγέθους του δικτύου απαιτεί μια διαδικασία δοκιμής και σφάλματος. Η οποία είναι εφικτή σε δεδομένα που δεν ανανεώνονται, αλλά δυσκολεύει στα συνεχώς μεταβαλλόμενα δεδομένα που υπάρχουν στην online μάθηση.

4.2 Αλγόριθμος k-means

Ο πιο ευρέως διαδεδομένος και απλός αλγόριθμος συσταδοποίησης ή ομαδοποίησης (Σχήμα 17) είναι ο k-μέσων (k-means) . Αναφέρθηκε πρώτη φορά το 1957 από τον Lloyd αλλά δημοσιεύτηκε επίσημα το 1987 [47], [55].



Σχήμα 17 Ομαδοποίηση k-means

Ο λόγος που χρησιμοποιείται συχνά είναι τα δύο σημαντικά χαρακτηριστικά του, είναι πολύ απλός στην εφαρμογή του αλλά ταυτόχρονα αποτελεσματικός. Δεν έχει κάποια συγκεκριμένη συνάρτηση για την κατανομή των δεδομένων. Στην πραγματικότητα είναι μια επαναληπτική διαδικασία που αποτελείται από τα εξής βήματα [16]:

- i. Γίνεται αρχικοποίηση των κέντρων με την επιλογή τυχαίων δεδομένων εισόδου.
- ii. Για κάθε δεδομένο υπολογίζεται το πλησιέστερο σε αυτό κέντρο και τοποθετείται στην ομάδα του συγκεκριμένου κέντρου.
- iii. Υπολογίζεται η μέση τιμή των μεταβλητών του κάθε κέντρου και ενημερώνεται η τιμή του.
- iv. Τα δύο τελευταία βήματα επαναλαμβάνονται έως ότου τα κέντρα συγκλίνουν στις τελικές θέσεις.

Φυσικά όπως σε κάθε μέθοδο υπάρχουν και μειονεκτήματα εκ των οποίων τα βασικότερα προβλήματα είναι η εξάρτηση του από τις αρχικές τυχαίες τιμές και από την επιλογή του αριθμού των κέντρων.

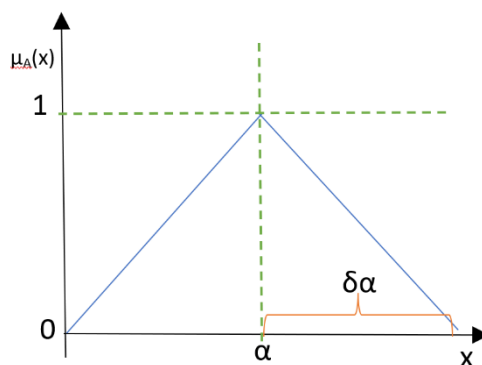
4.3 Αλγόριθμος ασαφών μέσων (Fuzzy Means, FM)

Η θεωρία της ασαφούς λογικής αναπτύχθηκε εξαιτίας της ανάγκης να αποτυπωθούν τα πραγματικά γεγονότα με πιο αντιπροσωπευτικό τρόπο, χωρίς να υπολογίζονται οι διακριτοί περιορισμοί. Η βασική διαφορά είναι οι στο εύρος των τιμών του. Ενώ ήταν 0 ή 1, τώρα

ανήκουν στο κλειστό διάστημα $[0,1]$.Ο αλγόριθμος ασαφών μέσων δημιουργήθηκε για να λύσει ένα από τα κύρια προβλήματα, την επιλογή των αρχικών τιμών των παραμέτρων σχεδιασμού. Αν οι τιμές αυτές δεν επιλεγούν σωστά τότε ο προηγούμενος αλγόριθμος k-means θα οδηγηθεί σε τοπικό ελάχιστο. Εξίσου σημαντικό είναι ότι ο FM δεν έχει το στοιχείο της τυχαιότητας, όσες φορές και να εκτελεστεί θα οδηγηθεί στα ίδια κέντρα.

Αρχικά γίνεται ο διαχωρισμός του χώρου των N μεταβλητών σε ασαφής σύνολα. Έτσι δημιουργούνται c_i σύνολα (τριγωνικά) για κάθε κανονικοποιημένη μεταβλητή x_i , όπου $i=1,2\dots N$. Επίσης, ο διαχωρισμός θα ονομάζεται συμμετρικός αν το πλήθος των συνόλων σε καθεμία διάσταση συμπίπτει, διαφορετικά θα ονομάζεται μη συμμετρικός.

Κάθε σύνολο c (Σχήμα 18) έχει μια συνάρτηση συγγένειας από την οποία προκύπτει ο βαθμός της σχέσης του συνόλου με ένα σημείο και έτσι μια τιμή καθορίζει αν αυτό θα είναι μέρος του συνόλου [55].



Σχήμα 18 Ασαφές σύνολο

Η μορφή αυτής της συνάρτησης θα είναι [10]:

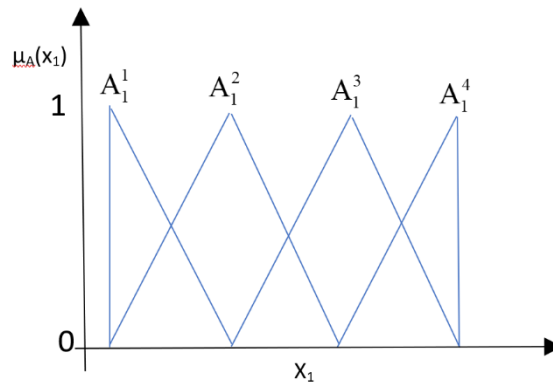
$$\mu_A(x) = \begin{cases} 1 - rd(x) & rd(x) \leq 1 \\ 0 & \text{αλλιώς} \end{cases} \quad (4.1)$$

Όπου η σχετική ευκλείδεια απόσταση

$$rd(x) = \frac{|x-a|}{\delta a} \quad (4.2)$$

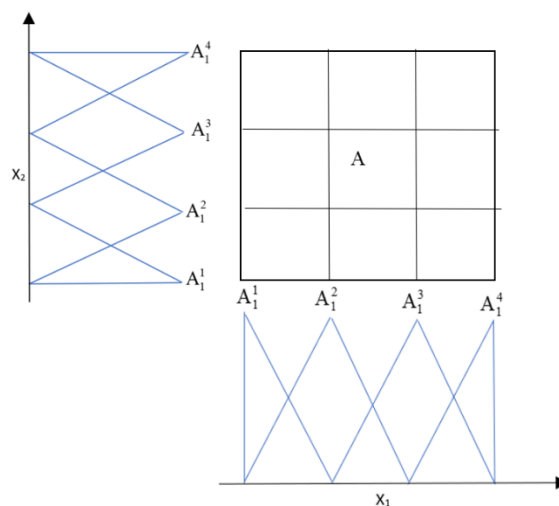
με το a να είναι το κέντρο του συνόλου και δa το ήμισυ του εύρους.

Έτσι για κάθε μία μεταβλητή γίνεται ένας διαχωρισμός σε ασαφή σύνολα (Σχήμα 19). Για παράδειγμα για την μία μεταβλητή αν γίνει χωρισμός σε 4 σύνολα :



Σχήμα 19 Ασαφής διαχωρισμός

Αν οι μεταβλητές είναι δύο, τότε ο χώρος είναι δισδιάστατος (Σχήμα 20) [55]. Για παράδειγμα 2 μεταβλητές με 4 ασαφή σύνολο θα υπάρχουν $C=8$ ασαφή υποχώροι:



Σχήμα 20 Δισδιάστατος ασαφής διαχωρισμός

Με μια επαναληπτική διαδικασία, υπολογίζονται οι βαθμοί συγγένειας των μεταβλητών για όλες τις διαστάσεις και λαμβάνεται ο ασαφής υποχώρος που χαρακτηρίζει καλύτερα τα δεδομένα εισόδου. Με το πέρας της διαδικασίας, οι υποχώροι που έχουν σχηματιστεί θα καλύπτουν όλα τα διανύσματα εισόδου, διότι σε κάθε επανάληψη γίνεται έλεγχος και αν υπάρχουν σύνολα που δεν καλύπτονται, εισάγεται ένας νέος υποχώρος.

Πιο αναλυτικά τα βήματα στην πράξη είναι:

- i. Τα δεδομένα N εισόδου κανονικοποιούνται και βρίσκονται οι τιμές κέντρου a και εύρους da με την χρήση των \max και \min τιμών των συνόλων.
- ii. Αρχικά για το 1^ο κέντρο, υπολογίζεται ο βαθμός συγγένειας για κάθε μεταβλητή εισόδου για κάθε ασαφές σύνολο.

- iii. Βρίσκεται και αποθηκεύεται η διεύθυνση του κέντρου του ασαφούς συνόλου με το μέγιστο βαθμό συγγένειας
- iv. Μόλις ολοκληρωθεί η διαδικασία για όλες τις μεταβλητές εισόδου, δημιουργείται το 1^ο κέντρο RBF.
- v. Συνεχίζεται για το 2^ο έως το K αριθμό δεδομένων, ελέγχοντας για το καθένα αν καλύπτεται από κάποιο κέντρο με χρήση της ευκλείδειας απόστασης. Αν η απόσταση δεν είναι μικρότερη ή ίση του 1, τότε προστίθεται νέο κέντρο και ξαναγίνονται οι υπολογισμοί.
- vi. Όταν τελειώσει ο αριθμός των δεδομένων εισόδου, τότε θα έχουν δημιουργηθεί τα RBF κέντρα, δηλαδή οι συντεταγμένες και ο αριθμός των κέντρων.

Χρησιμοποιείται στο πειραματικό μέρος για την κατασκευή της κρυφής στοιβάδας του δικτύου RBF βρίσκοντας το κατάλληλο μέγεθος του δικτύου καθώς και τις θέσεις των κέντρων. Μετά με την μέθοδο ελαχίστων τετραγώνων θα υπολογιστούν τα γραμμικά βάρη. Περισσότερη βιβλιογραφία για τον FM βρίσκεται στα [\[11\]](#), [\[12\]](#), [\[48\]](#).

4.4 Στατιστικοί δείκτες

Ως κριτήρια αξιολόγησης-επίδοσης των συστημάτων παλινδρόμησης γίνεται χρήση ορισμένων στατιστικών δεικτών. Με αυτόν τον τρόπο ελέγχεται η απόδοση των δικτύων [\[38\]](#).

- ❖ Μέσο τετραγωνικό σφάλμα (Mean Squared Error, MSE)

$$MSE = \frac{\sum_{i=1}^n (y_j - \bar{y}_j)^2}{N} \quad (4.3)$$

Μετράει την μέση ευκλείδεια απόσταση ανάμεσα του διανύσματος εξόδου y_j του μοντέλου και του διανύσματος(στόχου) \bar{y}_j . Το εν λόγω κριτήριο λαμβάνει πάντα μη αρνητικές τιμές και η βέλτιστη τιμή είναι το 0. Λόγω του τετραγώνου είναι ευαίσθητο όταν υπάρχουν μεγάλα σφάλματα αλλά είναι χρήσιμο όταν υπάρχουν σφάλματα μακριά από το μοντέλο. Υπάρχει και το RMSE με την μόνη διαφορά ότι είναι η ρίζα του μέσου τετραγωνικού σφάλματος.

- ❖ Μέσο απόλυτο σφάλμα (Mean Absolute Error, MAE)

$$MAE = \frac{\sum_{i=1}^n |y_j - \bar{y}_j|^2}{N} \quad (4.4)$$

Μετράει την μέση απόλυτη απόσταση μεταξύ των διανυσμάτων y_j και \bar{y}_j . Επίσης, το κριτήριο λαμβάνει πάντα μη αρνητικές τιμές και η βέλτιστη τιμή είναι το 0.

- ❖ Μέσο απόλυτο σχετικό σφάλμα (Mean Absolute Relative Error, MARE)

$$MARE = \frac{\sum_{i=1}^n (y_j - \bar{y}_j)^2}{N y_j} \% \quad (4.5)$$

Είναι ένα σύνθετο κριτήριο το οποίο παρουσιάζει την σχέση σφάλματος με τα δεδομένα, δηλαδή πόσο % λάθος κάνει.

- ❖ Συντελεστής συσχέτισης (Correlation Coefficient, R^2)

$$R^2 = 1 - \frac{SSE}{SST} \quad (4.6)$$

όπου

$$SSE = \sum_{i=1}^n (y_j - \hat{y}_j)^2, SST = \sum_{i=1}^n (y_j - \bar{y}_j)^2 \quad (4.7)$$

Το SSE είναι πόσο η πραγματικότητα από την πρόβλεψη και το SST πόσο απέχουν οι πραγματικές τιμές από τον μέσο όρο. Ο συντελεστής συσχέτισης δείχνει πόσο καλύτερα μπορεί να γίνει η πρόβλεψη του y χρησιμοποιώντας το μοντέλο σε σχέση με την πρόβλεψη του μέσου όρου των μετρήσεων. Το ιδανικό μοντέλο θα είχε την τιμή 1 ενώ το χειρότερο το $-\infty$.

Στην παρούσα εργασία χρησιμοποιήθηκε ο στατιστικός δείκτης RMSE.

5 Μέθοδοι Βελτιστοποίησης

Η βελτιστοποίηση θεωρείται δεδομένη σε όλα τα πεδία της μηχανικής και επιστήμης γενικότερα. Ένας ορισμός της είναι:

«Βελτιστοποίηση ονομάζεται η διαδικασία μέσα από την οποία μπορεί να βρεθεί το καλύτερο αποτέλεσμα για κάποιες δεδομένες συνθήκες.» [49]

Για να συμβεί αυτό, απαιτούνται οι παράμετροι του προβλήματος, οι οποίες μπορεί να πλαισιώνονται από ορισμένους περιορισμούς καθώς και η αντικειμενική συνάρτηση μέσω της οποίας θα επιλυθεί το πρόβλημα ελαχιστοποίησης ή μεγιστοποίησης. Τα προβλήματα μεγιστοποίησης μπορούν να διατυπωθούν ως ελαχιστοποίησης και αν υπάρχουν περιορισμοί γίνεται παραμετροποίηση της αντικειμενικής συνάρτησης προσθέτοντας ένα είδος ποινής. Επίσης διακρίνονται με βάση την γραμμικότητα τους σε γραμμικά ή μη γραμμικά προβλήματα.

5.1 Λίγο πιο αναλυτικά

Ένα πρόβλημα βελτιστοποίησης μαθηματικά διατυπώνεται ως εξής [50]:

$$\min_x f(x) \quad (5.1)$$

Με περιορισμούς :

$$c_i(x) = 0 \quad (5.2)$$

Όπου x ένας πίνακας που περιέχει N μεταβλητές σχεδιασμού του προβλήματος

$$x = [x_1 \quad x_2 \quad \dots \quad x_N] \quad (5.3)$$

Στην πράξη πρέπει να βρεθούν οι τιμές για τις μεταβλητές σχεδιασμού x που δίνουν ελάχιστη τιμή στην αντικειμενική συνάρτηση f .

5.1.1 Γραμμικά προβλήματα

Για την επίλυση τους χρειάζεται ο γραμμικός προγραμματισμός. Ονομάζονται έτσι επειδή η αντικειμενική συνάρτηση και οι περιορισμοί είναι γραμμικές συναρτήσεις των μεταβλητών σχεδιασμού. Το πρόβλημα διατυπώνεται :

$$\min_x f(x) = c_1x_1 + c_2x_2 + \dots c_nx_n \quad (5.4)$$

με τους αντίστοιχους ανισωτικούς ή εξισωτικούς περιορισμούς.

Τα απλά προβλήματα μπορούν να λυθούν με γραφική επίλυση. Η ύπαρξη περισσότερων μεταβλητών σχεδιασμού απαιτεί περισσότερες διαστάσεις καθιστώντας αδύνατη τη γραφική απεικόνιση. Επίσης η ύπαρξη πολλών περιορισμών καθιστά το πρόβλημα πιο δύσκολο. Στο σημείο αυτό εισέρχεται στο πεδίο η υπολογιστική μέθοδος Simplex, η οποία θεωρείται η πιο ευρέως εφαρμοσμένη μέθοδος για την επίλυση γραμμικών προβλημάτων.

Ακόμα αξίζει να αναφερθεί ο ακέραιος και ο μεικτός ακέραιος γραμμικός προγραμματισμός. Στον πρώτο, οι μεταβλητές σχεδιασμού λαμβάνουν μόνο ακέραιες τιμές ενώ στον δεύτερο έχουν γραμμική μορφή. Συμπληρωματικά υπάρχει ο κυρτός και τετραγωνικός προγραμματισμός. Το πλεονέκτημα της κυρτής είναι το γεγονός ότι υπάρχει ένα μοναδικό σημείο που κάνει ελάχιστο ή μέγιστο αντίστοιχα.

5.1.2 Μη γραμμικά προβλήματα

Μέχρι στιγμής αναφέρθηκαν επιγραμματικά τα γραμμικά προβλήματα στα οποία είναι γνωστή η μορφή της συνάρτησης. Στην πράξη όμως δεν συμβαίνει κάτι τέτοιο και ερχόμαστε αντιμέτωποι με προβλήματα μη γραμμικού προγραμματισμού. Είναι πολύ πιο δύσκολα στην επίλυση τους καθώς έχουν πάνω από ένα τοπικό ελάχιστο και δεν υπάρχει κάποια μέθοδος που να εγγυάται την εύρεση ολικού ελαχίστου.

Επιγραμματικά προς το παρόν αναφέρονται μερικές από τις μεθόδους:

- Μέθοδος τάχιστης καθόδου
- Μέθοδος Gauss-Newton
- Μέθοδος Levenberg–Marquardt (LM)
- Μέθοδος των πολλαπλασιαστών Lagrange

5.2 Βελτιστοποίηση στο δίκτυο RBF

Στην εκπαίδευση του δικτύου RBF καλείται κάποιος να λύσει ένα πρόβλημα ελαχίστων τετραγώνων, δηλαδή την ελαχιστοποίηση της αντικειμενικής συνάρτησης [14]:

$$\min_x f(x) = \min_x \frac{1}{2} \sum_{i=1}^m r_i(x)^2 \quad (5.5)$$

Με

$$x \in R^n, \quad r(x) = (r_1(x), \dots, r_m(x))^T \quad (5.6)$$

και

$$r_i(x) = y_i - M(x, t_i), i = 1, \dots, m \quad (5.7)$$

Το μοντέλο που χρησιμοποιείται είναι η συνάρτηση M όπου για δεδομένες τιμές (x, t_i) δίνει αποτέλεσμα τα αντίστοιχα (y_i) . Στην συγκεκριμένη περίπτωση οι δεδομένες τιμές θα είναι τα σύνολα δεδομένων εκπαίδευσης ενώ οι παράμετροι θα είναι τα βάρη ή τα κέντρα. Επειδή δεν είναι γνωστή η μορφή της συνάρτησης f , ο στόχος είναι να βρεθεί ένα τοπικό ελάχιστο ως λύση μιας και το ολικό μπορεί να είναι και αδύνατο να βρεθεί. Για να γίνει αυτό χρησιμοποιούνται οι πρώτες και οι δεύτερες παράγωγοι της $f(x)$ έχοντας το ανάδελτα ή αλλιώς την κλίση της $\nabla f(x)$ και ο πίνακας δευτέρων παραγώντων ή αλλιώς Εσσιανός πίνακας $\nabla^2 f(x)$ όπου για $x \in R^n$:

$$\nabla f(x)^T = \left[\frac{\partial f(x)}{\partial x_1} \quad \frac{\partial f(x)}{\partial x_2} \quad \dots \quad \frac{\partial f(x)}{\partial x_n} \right], \quad (5.8)$$

$$\left[\nabla^2 f(x) \right]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, i, j = 1, 2, \dots, n \quad (5.9)$$

5.3 Μέθοδοι κατάβασης

Οι μέθοδοι καθόδου είναι μια μεγάλη κατηγορία που βασίζεται στην τεχνική κατά την οποία ξεκινούν από μια τιμή του x και τροποποιώντας σταδιακά τις μεταβλητές σχεδιασμού μειώνουν την τιμή της συνάρτησης οπότε:

$$f(x_{k+1}) < f(x_k) \quad (5.10)$$

Η νέα θέση της x_{k+1} χρειάζεται την διεύθυνση κατάβασης και το μέγεθος του βήματος δηλαδή:

$$x_{k+1} = x_k + ah_d \quad (5.11)$$

Με a η τιμή του βήματος και h_d η διεύθυνση κατάβασης η οποία υπολογίζετε με την μέθοδο της απότομης κατάβασης ή με την μέθοδο Newton

Ένας απλός αλγόριθμός κατάβασης σύμφωνα με την βιβλιογραφία [14], [55] είναι:

Αλγόριθμος μεθόδου κατάβασης

1. $k = 0, x_i = x_0$
 2. **για** $k = 0, 1, 2, \dots$
 3. εύρεση διεύθυνσης κατάβασης h_d
 4. εύρεση μεγέθους βήματος a
 5. ενημέρωση νέας θέσης
 $x_{k+1} = x_k + ah_d$
 6. **τέλος**
-

5.3.1 Μέθοδος Gauss-Newton

Η μέθοδος Gauss-Newton χρησιμοποιείται σε συναρτήσεις στις οποίες εντοπίζεται το άθροισμα τετραγώνων των συναρτήσεων. Είναι μια εναλλακτική μέθοδος της Newton και το μοντέλο της προκύπτει από το ανάπτυγμα Taylor [14]:

$$f(x+h) \approx l(h) = f(x) + J(x)h \quad (5.12)$$

Με $J(x)$ να είναι η Ιακωβιανή του διανύσματος και μια διεύθυνση h , όπως αναφέρθηκε και προηγουμένως, ένα βήμα που πρέπει να διανυθεί για την μείωση της τιμής f . Στο τέλος, καλείται προς επίλυση ένα γραμμικό πρόβλημα ελαχίστων τετραγώνων:

$$\min_h \|Ah - b\| \quad (5.13)$$

με

$$A = J(x_k), \quad b = -(r(x_k)) \quad (5.14)$$

Η μέθοδος αυτή προτείνεται για μικρά βήματα h γιατί σε μεγάλα βήματα ο αλγόριθμος μπορεί να οδηγηθεί σε λάθος διευθύνσεις στις οποίες δεν θα ισχύει η προσέγγιση της f . Περισσότερα μπορούν να βρεθούν στην αντίστοιχη βιβλιογραφία [14].

5.3.2 Μέθοδος Levenberg-Marquardt, LM

Η μέθοδος LM λύνει το πρόβλημα της Gauss-Newton με τα μεγάλα βήματα εισάγοντας έναν περιορισμό σχετικά με το μέγεθος του βήματος [15]:

$$\min_{\|h\|<c} \|J(x_k)h + r(x_k)\|^2 \quad (5.15)$$

Προσθέτετε ο περιορισμός του βήματος χρησιμοποιώντας τον πολλαπλασιαστή Lagrange στην αντικειμενική συνάρτηση (5.15) [50]:

$$\min_h \|J(x_k)h + r(x_k)\|^2 + \lambda \|h\|^2 \quad (5.16)$$

Ο λόγος αναφοράς της μεθόδου Gauss-Newton και της απότομης κατάβασης ήταν για να προετοιμαστεί το έδαφος για την εισαγωγή του αλγορίθμου LM, ο οποίος είναι ο συνδυασμός των πλεονεκτημάτων των δύο μεθόδων. Πιο συγκεκριμένα, στην αρχή της επαναληπτικής διαδικασίας για μεγάλα βήματα θα γίνεται απότομη κατάβαση ενώ στην σύγκλιση στο τέλος για μικρά βήματα θα γίνεται Gauss-Newton. [14]

Στην παρούσα εργασία χρησιμοποιήθηκε η μέθοδος LM για την εύρεση των δοκιμαστικών κέντρων ή και βαρών με την αντίστοιχη διεύθυνση οδηγούμενη από το κριτήριο αύξησης ή μείωσης του βήματος.

Το κριτήριο αυξομείωσης του βήματος [50] :

$$\rho_k = \frac{f(x_k) - f(x_{k+1})}{f(x_k) - \frac{1}{2} \|J(x_k)h + r(x_k)\|^2} \quad (5.17)$$

Με τον παρακάτω αλγόριθμο, ο οποίος αυξάνει το βήμα αν το κριτήριο είναι πολύ μικρό και αντίστοιχα το μειώνει αν το κριτήριο είναι μεγάλο. Αν είναι θετικό το κάνει αποδοχή και ενημερώνει τις σταθερές.

ενημέρωση του λ

αν $\rho < 0.25$

$\lambda = 2\lambda$

αλλιώς αν $\rho > 0.75$

$\lambda = \lambda/3$

τέλος

αν $\rho > 0$

ενημέρωσε το x

τέλος

Επίσης ως κριτήριο τερματισμού υπάρχει η σύγκλιση ακολουθίας, ο έλεγχος τιμής της συνάρτησης, το μέτρο κλίσης της και μέγιστες επαναλήψεις. Στην παρούσα εργασία χρησιμοποιήθηκε και η σύγκλιση ακολουθίας για τον τερματισμό των επαναλήψεων της LM μεθόδου ελέγχοντας την ρίζα του μέσου τετραγωνικού σφάλματος RMSE.

5.4 Μέθοδος Variable Projection

Σε ένα μη γραμμικό πρόβλημα ελαχίστων τετραγώνων δηλαδή γραμμικός συνδυασμός μη γραμμικών συναρτήσεων είναι πιο πολύπλοκά τα πράγματα. Μια λύση που έχει προταθεί [50], [51] είναι να σπάσει σε δύο μέρη, αρχικά στο μη γραμμικό πρόβλημα και έπειτα στο γραμμικό. Η επίλυση και των δύο θα είναι με ελάχιστα τετράγωνα απλά το ένα θα έχει μη γραμμικές παραμέτρους ενώ το άλλο γραμμικές. Επίσης με αυτόν τον τρόπο μειώνονται οι χρόνοι σύγκλισης καθώς μειώνεται ο χώρος των άγνωστων παραμέτρων [52].

Τα βήματα είναι τα εξής [53]:

1. Για την ελαχιστοποίηση της συνάρτησης πρέπει να βρεθούν οι παράμετροι c και w

$$r(c, w) = y - \Phi(c)w \quad \text{ή} \quad \min_{c, w} \frac{1}{2} \|y - \Phi(c)w\|^2 \quad (5.18)$$

2. Εφόσον είναι διαθέσιμες οι μη γραμμικές παράμετροι c γίνεται

$$\min_w \frac{1}{2} \|y - \Phi(c)w\|^2 \quad (5.19)$$

3. Μετά την λύση του θα προκύψει

$$w = \Phi(c)^+ y \quad (5.20)$$

με τον

$$\Phi(c)^+ = (\Phi(c)^T \Phi(c))^{-1} \Phi(c)^T \quad (5.21)$$

να είναι ο πίνακας Moore-Penrose, ένας γενικευμένος ψευδο-αντίστροφος.

4. Γίνεται αντικατάσταση της στο αρχικό πρόβλημα και με έναν κοινό παράγοντα το μη γραμμικό πρόβλημα ελαχίστων τετραγώνων γίνεται:

$$\min_c \frac{1}{2} \|(I - \Phi(c)\Phi(c)^+)y\|^2 \quad (5.22)$$

5. Τελικά μετά την επίλυση η προβολή μεταβλητής $r_2(c)$ του y θα είναι:

$$r_2(c) = (I - \Phi(c)\Phi(c)^+)y \quad (5.23)$$

5.5 Παραγοντοποίηση QR

Χρησιμοποιήθηκε η μέθοδος παραγοντοποίησης QR του πίνακα $\Phi(c)$ για να γίνουν οι υπολογισμοί πιο αποδοτικοί [54]:

$$\Phi(c) = QR = [Q_1 \quad Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1 \quad (5.24)$$

με $Q_1 \in R^{m \times N}$, $Q_2 \in R^{m \times (m-N)}$ ορθογώνιοι, ο $R_1 \in R^{N \times N}$ άνω τριγωνικός και ισχύει

$$Q^T Q = Q Q^T = I \quad (5.25)$$

Και

$$Q_1 Q_1^T + Q_2 Q_2^T = I \quad (5.26)$$

Οπότε :

$$\Phi(c)^T \Phi(c) = R_1^T Q_1^T Q_1 R_1 = R_1^T R_1 \quad (5.27)$$

Μεταβαίνοντας στο πρόβλημα της συγκεκριμένης εργασίας, εξηγείται και παρακάτω πιο αναλυτικά, χρειάζεται να υπολογιστούν τα βάρη μέσω της επίλυσης γραμμικού προβλήματος ελαχίστων τετραγώνων για συγκεκριμένα κέντρα.

$$w(c) = \arg \min_w \frac{1}{2} \|y - \Phi(c)w\|_2^2 \Rightarrow \Phi(c)^T \Phi(c)w(c) = \Phi(c)^T y \quad (5.28)$$

Αντικαθιστώντας στην (4.28) την (4.27) έχουμε:

$$R_1^T R_1 w(c) = R_1^T Q_1^T y \Leftrightarrow w(c) = R_1^{-1} Q_1^T y \quad (5.29)$$

Και αντίστοιχα ο ψευδό-αντίστροφος:

$$\Phi^+ = R_1^{-1} Q_1^T \quad (5.30)$$

6 Πειραματική διαδικασία

Στόχος της πειραματικής διαδικασίας ήταν η σύγκριση των δύο μεθόδων που βασίζονται στον αλγόριθμο LM, τόσο μεταξύ τους, αλλά και με άλλες μεθόδους εκπαίδευσης δικτύων RBF. Και οι δύο μέθοδοι προγραμματίστηκαν στο λογισμικό Matlab, όπου για να γίνει η διαδικασία λιγότερο χρονοβόρα όσον αφορά την ανθρώπινη παρέμβαση, ο κώδικας δημιουργήθηκε όσο πιο παραμετρικό γινόταν χρησιμοποιώντας functions. Με αυτόν το τρόπο αυτοματοποιήθηκαν οι αλλαγές στα datasets.

Επίσης οι παράμετροι που χρησιμοποιήθηκαν στις μεθόδους εκπαίδευσης είναι οι εξής:

- Ο αριθμός των ασαφών συνόλων κυμαίνεται από 5 έως και 12 (με βήμα 1)
- Οι τιμές που πήρε η παράμετρος λ της LM είναι $[10^{-6} 10^{-5} 10^{-4} 10^{-3} 10^{-2} 1 10 10^2]$

Συνολικά έγιναν 64 δοκιμές σε κάθε dataset.

6.1 Η εκπαίδευση στην πράξη

Σε αυτό το σημείο παρουσιάζονται ορισμένες λεπτομέρειες που έπρεπε να πραγματοποιηθούν πριν από την έναρξη του πειραματικού μέρους.

6.1.1 Χωρισμός των δεδομένων εισόδου

Συνήθως στα σύνολα δεδομένων υπάρχουν κάποια λανθασμένα δεδομένα τα οποία χαρακτηρίζονται από θόρυβο. Χρειάζεται προσοχή στην εκπαίδευση του δικτύου γιατί μπορεί να επηρεαστεί από αυτά τα δεδομένα. Στην περίπτωση που περνάει από όλα τα δεδομένα έχουμε υπέρ-εκπαίδευση ή αλλιώς υπερ-προσαρμογή. Για να το αποφευχθεί αυτό χωρίζονται με τυχαίο τρόπο τα δεδομένα εισόδου σε 3 υποσύνολα με αναλογία 50%, 25%, 25% [11] :

- Στο υποσύνολο εκπαίδευσης (training subset), το οποίο χρησιμοποιείται αποκλειστικά για την εκπαίδευση του ή αλλιώς τον υπολογισμό των παραμέτρων του δικτύου.
- Στο υποσύνολο αξιολόγησης (validation subset), το οποίο χρησιμοποιείται μετά το πέρας της εκπαίδευσης για την επιλογή του μοντέλου με τα καλύτερα αποτελέσματα, ώστε να αποφευχθεί η υπερπροσαρμογή στα δεδομένα εκπαίδευσης.

- Στο υποσύνολο ελέγχου (testing subset), το οποίο χρησιμοποιείται για την αμερόληπτη αξιολόγηση της απόδοσης του δικτύου. Σημειώνεται ότι το υποσύνολο ελέγχου δεν χρησιμοποιείται καθόλου κατά τη διάρκεια της εκπαίδευσης ή της επιλογής του μοντέλου.

6.1.2 Κανονικοποίηση των δεδομένων

Στα δεδομένα εισόδου υπάρχει συνήθως μια διαφορά στην κλίμακα τους η οποία μπορεί να προκαλέσει προβλήματα στην εκπαίδευση. Για να αποφευχθεί αυτό γίνεται μια κανονικοποίηση των δεδομένων πριν την εκπαίδευση και από-κανονικοποίηση στο τέλος για να έχουν νόημα τα αποτελέσματα. Γίνεται αναφορά δύο μεθόδων:

- Κανονικοποίηση με μηδενικό μέσο και μοναδιαία τυπική απόκλιση

Με αυτήν την τεχνική, οι τιμές κεντράρονται γύρω από τη μέση τιμή με μια μοναδιαία τυπική απόκλιση. Δηλαδή μηδενίζεται ο μέσος όρος και η κατανομή έχει μοναδιαία τυπική απόκλιση. Μετά το τέλος της κανονικοποίησης θα πρέπει να αποθηκεύονται ο μέσος όρος και η τυπική απόκλιση της κάθε μεταβλητής, έτσι ώστε η έξοδος y να μπορεί να από-κανονικοποιηθεί.

- Γραμμικός μετασχηματισμός ανάμεσα σε δύο τιμές

Όλες οι μεταβλητές μετασχηματίζονται ανάμεσα σε ένα συγκεκριμένο εύρος τιμών. Οι παράμετροι κανονικοποίησης αποθηκεύονται για να μπορεί στο τέλος της διαδικασίας να γίνει από-κανονικοποίηση. Το πιο σύνηθες εύρος τιμών είναι $[-1,1]$ το οποίο χρησιμοποιείται στην παρούσα εργασία.

6.2 1^η Εφαρμογή Variable Projection

Όπως έγινε αναφορά νωρίτερα, ο σκοπός είναι η εκπαίδευση ενός δικτύου RBF δηλαδή η εύρεση των κέντρων και των βαρών του δικτύου, το οποίο συνεπάγεται στην ελαχιστοποίηση του :

$$\min_{c,w} e(c,w) = \min_{c,w} \frac{1}{2} \sum_{i=1}^m r_i(c,w)^2 \quad (6.1)$$

Όπου

$$r_i(c, w) = y_i - f(x_i, c, w) = y_i - \sum_{j=1}^N w_j \Phi(x_i, c_j), \quad i = 1, 2, \dots, m \quad (6.2)$$

Και τελικά η αντικειμενική συνάρτηση θα είναι:

$$\min_{c, w} e(c, w) = \min_{c, w} \frac{1}{2} \|y - \Phi(c)w\|_2^2 \quad (6.3)$$

Συνεχίζεται η διαδικασία με τον αλγόριθμο Fuzzy Means που αναλύθηκε, στον οποίο δίνοντας ως είσοδο τα δεδομένα, υπολογίζονται οι θέσεις και ο αριθμός των κέντρων.

Χρησιμοποιήθηκε η πιο δημοφιλής συνάρτηση βάσης, Gaussian, επειδή είναι συνεχώς παραγωγίσιμη. Η Gaussian για την δημιουργία των συναρτήσεων Φ χρειάζεται να είναι γνωστά τα εύρη των κέντρων c .

Για τον υπολογισμό των ευρών έγινε χρήση της μεθόδου των κ-κοντινότερων γειτόνων με το εύρος σ_j για την j μονάδα RBF να υπολογίζεται:

$$\sigma_j = \sqrt{\frac{1}{p} \sum_{i=1}^p \|c_j - c_i\|_2^2} \quad (6.4)$$

Η μορφή των συναρτήσεων Φ θα είναι:

$$\Phi(x_m, c_N) = \exp\left(-\frac{u^2}{2\sigma^2}\right) \quad (6.5)$$

Όπου u είναι η ευκλείδεια απόσταση δεδομένου με κέντρου νευρώνα και υπολογίζεται ως:

$$u = \sqrt{\sum_{i=1}^n (x_{k,i} - c_{j,i})^2}, \quad k = 1, 2, \dots, m \quad (6.6)$$

Μετά τον υπολογισμό των συναρτήσεων Φ , προτείνεται η πρόσθεση μιας στήλης με άσσους. Με αυτόν τον τρόπο γίνεται η 'πόλωση' για τη μετατόπιση της συνάρτησης από την αρχή των αξόνων.

Συνεχίζοντας υλοποιείται η QR παραγοντοποίηση και βάση αυτής υπολογίζονται τα βάρη και οι πρώτες προβλέψεις για τα τρία υποσύνολα. Με βάση αυτές τις προβλέψεις υπολογίζεται και ο δείκτης σφάλματος RMSE, ο οποίος είναι πολύ σημαντικός για την σύγκριση των αποτελεσμάτων αλλά και ως κριτήριο τερματισμού.

Προχωρώντας στο κύριο μέρος της εργασίας, την επαναληπτική διαδικασία Levenberg-Marquardt στην οποία σε κάθε επανάληψη τα κέντρα μετατοπίζονται βάσει μιας διεύθυνσης. Αρχικά υπολογίζεται η Ιακωβιανή, δηλαδή η παραγωγή (ανάδελτα) του πίνακα $\Phi(c)$, η οποία είναι η κλίση των συναρτήσεων βάσης. Σημειώνεται ότι η χρήση της Gaussian ως συνάρτηση ενεργοποίησης γίνεται για τον λόγο ότι είναι παραγωγίσιμη, σε αντίθεση με άλλες συναρτήσεις ενεργοποίησης.

Υπολογίζεται το αρχικό κόστος h , το οποίο θα υπολογιστεί ξανά μετά τις νέες θέσεις των κέντρων και θα αποτελέσει μέρος του κριτηρίου αυξομείωσης του βήματος.:

$$h(c) = \frac{1}{2} \|y - \Phi(c)w\|_2^2 \quad (6.7)$$

Όπου με την παραγοντοποίηση QR και από την (5.29) γίνεται :

$$h(c) = \frac{1}{2} \|Q_1 R_1 R_1^{-1} Q_1^T y - y\|_2^2 = \frac{1}{2} \|I - Q_1 Q_1^T\|_2^2 \quad (6.8)$$

Και σύμφωνα με την (5.26), την (6.7) και επειδή είναι ορθογώνιος ο πίνακας $\|Q\|_2 = 1$

Τότε:

$$h(c) = \frac{1}{2} \|Q_2 Q_2^T y\|_2^2 = \frac{1}{2} \|Q_2^T y\|_2^2 \quad (6.9)$$

Προχωρώντας στην εύρεση της διεύθυνσης d για την μετατόπιση των κέντρων. Με βάση την βιβλιογραφία [51], πρέπει να λυθεί το εξής γραμμικό πρόβλημα ελαχίστων τετραγώνων:

$$\min_d \frac{1}{2} \|A(c)d - b\|_2^2 \quad (6.10)$$

Όπου

$$A(c) = -Q_2^T D(\Phi(c))w(c) \in \mathfrak{R}^{(m-N) \times Nm} \quad (6.11)$$

και

$$b = -Q_2^T y \in \mathfrak{R}^{(m-N) \times 1} \quad (6.12)$$

Και με την μέθοδο LM όπως (5.16) γίνεται:

$$\min_d \frac{1}{2} \|A(c)d - b\|_2^2 + \frac{\lambda}{2} \|d\|_2^2 \quad (6.13)$$

Οι αρχικές τιμές έχουν πλέον ολοκληρωθεί και τώρα υπολογίζονται τα δοκιμαστικά κέντρα, τα οποία θα είναι τα αρχικά με την προσθήκη της διεύθυνσης d . Πρέπει να υπολογιστούν ξανά όλες οι νέες δοκιμαστικές τιμές όπως : το δοκιμαστικό εύρος, ο δοκιμαστικός πίνακας $\Phi(c)$, η παραγοντοποίηση QR καταλήγοντας στο πιο σημαντικό, το δοκιμαστικό κόστος. Με βάση το κόστος υπολογίζεται το κριτήριο (4.17) το οποίο θα καθορίσει την αύξηση ή την μείωση του συντελεστή βήματος. Τα στοιχεία αυξομείωσης και αποδοχής της τιμής έχουν

αναλυθεί. Εφόσον η τιμή γίνει αποδεκτή, ενημερώνονται οι κέντρα και υπολογίζονται ξανά τα βάρη με την μέθοδο των ελάχιστων τετραγώνων καθώς και οι προβλέψεις των τριών συνόλων. Για να ολοκληρωθεί η επαναληπτική διαδικασία, ελέγχεται η σύγκλιση ακολουθίας με το μέσο τετραγωνικό σφάλμα, η οποία διακόπτεται αν για πέντε επαναλήψεις δεν βελτιωθεί.

6.3 2^η Εφαρμογή Πλήρους Συναρτησιακού

Στο Full Functional υπολογίζονται μαζί με τα κέντρα και τα βάρη. Αυτή είναι η βασική διαφορά. Με την προηγούμενη μεθοδολογία υπολογίζονταν τα βάρη με την μέθοδο των ελαχίστων τετραγώνων έχοντας έτσι τα ολικά βέλτιστα αλλά για τα συγκεκριμένα κέντρα. Δεν χρειάζεται επίσης να γίνει διαχωρισμός του προβλήματος με Variable Projection. Το πρόβλημα προς επίλυση θα είναι :

$$\min_{c,w} \frac{1}{2} \|r(c,w) + Dr(c,w)d\|^2 \quad (6.14)$$

Όπου $Dr(c,w)$ η Ιακωβιανή σύμφωνα με [53]:

$$Dr(c,w) = -[D\Phi(c)w \quad \Phi(c)] \in R^{m \times (Nn+N)} \quad (6.15)$$

Με $\Phi(c) \in R^{m \times N}$ η Ιακωβιανή ως προς τα βάρη και $D\Phi(c)w \in R^{m \times Nn}$ η Ιακωβιανή ως προς τα κέντρα. Σημείωση, m είναι το σύνολο των δεδομένων, N το σύνολο των κέντρων και n ο αριθμός των διαστάσεων.

Η διαδικασία αυτή είναι σε γενικές γραμμές παρόμοια με την προηγούμενη αλλά πιο εύκολη καθώς παραμετροποιείται ο κώδικας με την αφαίρεση του VP και προστίεται της Ιακωβιανής για τα βάρη. Η διεύθυνση d πλέον θα δείχνει την κατεύθυνση που θα κινηθεί το διάνυσμα για τα κέντρα αλλά και για τα βάρη.

Όσον αφορά τις αρχικές τιμές των βαρών ο [53] προτείνει να υπολογίζονται από επίλυση ελαχίστων τετραγώνων ενώ οι θέσεις των κέντρων από τον αλγόριθμο Fuzzy Means.

7 Ολοκλήρωση πειραματικής διαδικασίας

7.1 Ανάλυση των benchmark datasets

Αρχικά μια σύντομη περιγραφή (Πίνακας 1) των datasets που χρησιμοποιήθηκαν στην παρούσα εργασία, όπως αυτά αναφέρονται [11]:

- Gas Furnace: μεταβαλλόμενος ρυθμός τροφοδοσίας μεθανίου με σκοπό την πρόβλεψη της συγκέντρωσης διοξειδίου του άνθρακα στα καυσαέρια ενός κλιβάνου αερίου
- Samad: δημιουργήθηκε από την συνάρτηση

$$y = \frac{1}{1 + \exp[-\exp(x_1) + (x_2 - 0.5)^2 + \sin(\pi x_3)]} + \varepsilon$$

με ε ένας Gaussian θόρυβος $\sim N(0, 0.025)$, οι τιμές εισόδου λαμβάνονται από μια ομοιόμορφη κατανομή στο διάστημα $[0, 1]$

- Friedman: πρόβλημα παλινδρόμησης με εισόδους χαρακτηριστικά ομοιόμορφα καταναμημένα στο διάστημα $[0, 1]$ και έξοδο σύμφωνα με τον τύπο $y = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + noise * N(0, 0.8)$.

- Machine CPU: χαρακτηριστικά CPU με στόχο την πρόβλεψη της σχετικής απόδοσης της CPU

- Auto MPG: ορισμένα χαρακτηριστικά αυτοκινήτου με στόχο την πρόβλεψη κατανάλωσης καυσίμου.

- Mackey-Glass: χασοτική χρονοσειρά η οποία έχει χρησιμοποιηθεί εκτενώς για την αξιολόγηση μοντέλων νευρωνικών δικτύων και παράγεται από την διαφορική

$$y(t+1) = (1-b)y(t) + a \frac{y(t-\tau)}{(1+y^{10}(t-\tau))},$$

ο στόχος είναι να γίνει η πρόβλεψη της τρέχουσας τιμής $y(t)$, με βάση τις προηγούμενες έξι τιμές

$$y(t) = \text{RBF}(y(t-1), y(t-2), \dots, y(t-6))$$

με $a=0.2$, $b=0.1$ και $\tau=30$

- Auto Price: χαρακτηριστικά αυτοκινήτου με στόχο να προβλεφθεί η τιμή του
- Boston Housing: χαρακτηριστικά σπιτιών στην Βοστώνη από την απογραφή του 1970 με στόχο την πρόβλεψη της τιμής πώλησης τους.
- CPU Small: περιορισμένος αριθμός χαρακτηριστικών επεξεργαστών CPU με σκοπό την πρόβλεψη του τμήματος χρόνου που τρέχουν σε λειτουργία χρήστη.
- House Price-16H: χαρακτηριστικά με βάση τη δημογραφική δομή και την κατάσταση της αγοράς κατοικιών με στόχο την πρόβλεψη της μέσης τιμής της κατοικίας

- Ailerons-Elevators : χαρακτηριστικά που προέρχονται προέρχονται από την πτήση ενός αεροσκάφους F16, περιγράφουν την κατάσταση του αεροπλάνου, ενώ ο στόχος είναι να προβλεφθεί ο έλεγχος των πηδαλίων και των πτερυγίων του

Πίνακας 1 Τα σύνολα δεδομένων

Πίνακας 1 Σύνολο δεδομένων	Αριθμος μεταβλητών εισόδου	Αριθμός παραδειγμάτων
Gas Furnace	3	150
Samad	3	1500
Friedman	5	1000
Machine CPU	6	209
Auto MPG	6	392
Mackey-Glass	6	1995
Auto Price	15	159
Boston Housing	13	506
CPU Small	12	8192
House Price-16H	16	22784
Elevators	18	16599
Ailerons	40	13750

7.2 Αποτελέσματα

Δημιουργήθηκε ένας συγκεντρωτικός πίνακας αποτελεσμάτων (Πίνακας 2) των μεθόδων της παρούσας εργασίας, VP και FF. Για λόγους σύγκρισης προστέθηκαν τα αποτελέσματα της μεθόδου PSO-NSFM [11] μαζί με τον αλγόριθμο SFM και τα δίκτυα MLP. Περιλαμβάνει επίσης τις σημαντικότερες παραμέτρους όλων των μεθόδων για την λεπτομερέστερη εξέτασή τους. Όσον αφορά την σύγκριση με τις άλλες μεθόδους πρέπει να σημειωθεί ότι οι μετρήσεις των υπολογιστικών χρόνων έχουν πραγματοποιηθεί σε διαφορετικούς υπολογιστές.

Στην μέθοδο PSO-NSFM γίνεται χρήση βελτιστοποίησης σμήνους σωματιδίων για την εκπαίδευση με μη συμμετρικό διαχωρισμό του χώρου εισόδου. Στον αλγόριθμο SFM επιλέγεται το μοντέλο βάση των ασαφών συνόλων με συμμετρικό διαχωρισμό του χώρου εισόδου χρησιμοποιώντας την thin-plate-spline ως συνάρτησης βάσης. Ενώ στα δίκτυα MLP πραγματοποιείται εκπαίδευση βάσει του αλγορίθμου LM σε δίκτυα πρόσθιας τροφοδότησης δύο επιπέδων. Τέλος, η εκπαίδευση σε όλες τις μεθόδους γίνεται βάση του συνόλου δεδομένων εκπαίδευσης και έπειτα η επιλογή γίνεται βάση του συνόλου δεδομένων δοκιμής.

Πίνακας 2 Συγκεντρωτικός πίνακας αποτελεσμάτων

Dataset	Algorithm	RMSE Validation	RMSE Testing	Fuzzy Partition (s)	Nodes	λ	Time (sec)
Gas Furnace	VP	1.64×10^{-1}	3.80×10^{-1}	5	5	10^{-6}	1.6
	FF	1.64×10^{-1}	3.81×10^{-1}	5	5	10^{-6}	2.3
	PSO-NSFM	1.70×10^{-1}	4.27×10^{-1}	[8 5 23]	22	-	77
	SFM	2.19×10^{-1}	4.44×10^{-1}	15	29	-	1.6
	MLP	3.10×10^{-1}	4.60×10^{-1}	-	[8 19]	-	1040
Samad	VP	2.44×10^{-2}	2.71×10^{-2}	9	73	10^{-6}	25.7
	FF	2.47×10^{-2}	2.75×10^{-2}	12	143	10^{-4}	30.3
	PSO-NSFM	2.47×10^{-2}	2.80×10^{-2}	[4 6 38]	89	-	60
	SFM	2.59×10^{-2}	2.83×10^{-2}	11	112	-	34
	MLP	2.49×10^{-2}	2.81×10^{-2}	-	[13 7]	-	1298
Friedman	VP	8.74×10^{-1}	1.06	6	69	10^{-6}	42.9
	FF	8.72×10^{-1}	9.72×10^{-1}	6	69	10^{-5}	75.7
	PSO-NSFM	9.08×10^{-1}	1.06	[8 4 4 32 4]	114	-	133
	SFM	9.64×10^{-1}	1.09	8	179	-	47
	MLP	1.04	1.08	-	[15 15]	-	1300
Machine CPU	VP	2.54×10^1	3.03×10^1	5	11	10^{-4}	1.6
	FF	2.55×10^1	2.78×10^1	6	11	10^{-6}	2.1
	PSO-NSFM	2.28×10^1	2.16×10^1	[12 4 13 6 16 5]	19	-	83
	SFM	2.49×10^1	2.77×10^1	11	22	-	4
	MLP	2.92×10^1	2.99×10^1	-	[21 7]	-	954
Auto MPG	VP	2.28	2.21	6	17	10^{-4}	3.6
	FF	2.17	2.43	10	41	10^{-4}	5.6
	PSO-NSFM	2.11	1.95	[27 4 4 32 4 8]	36	-	92
	SFM	2.22	2.12	10	41	-	8
	MLP	2.51	2.34	-	[23 19]	-	1174
Mackey-Glass	VP	4.81×10^{-2}	5.24×10^{-2}	7	7	10^{-5}	14.5
	FF	4.81×10^{-2}	5.28×10^{-2}	5	5	10^{-5}	10.1
	PSO-NSFM	4.79×10^{-2}	5.30×10^{-2}	[4 4 4 15 5 43]	24	-	539
	SFM	4.92×10^{-2}	5.34×10^{-2}	17	34	-	27
	MLP	4.88×10^{-2}	5.35×10^{-2}	-	[9 13]	-	1140
Auto Price	VP	2.39×10^3	2.60×10^3	10	40	10	3.8
	FF	2.46×10^3	2.51×10^3	9	38	10^{-6}	5.1
	PSO-NSFM	2.17×10^3	2.11×10^3	[4 4 29 50 50 23 7 34 5 10 24 4 35 4 37]	57	-	339
	SFM	2.73×10^3	2.86×10^3	11	44	-	6
	MLP	2.29×10^3	2.38×10^3	-	[17 11]	-	624
Boston Housing	VP	2.69	2.89	5	20	10	13.5
	FF	2.85	2.95	11	93	1	18
	PSO-NSFM	2.35	2.72	[48 50 39 4 50 24 4 6 28 45 18 5 11]	138	-	2588
	SFM	2.77	2.93	16	146	-	13
	MLP	2.47	3.38	-	[18 19]	-	2402
CPU Small	VP	2.90	2.93	10	65	1	1358
	FF	2.97	2.93	12	122	10^{-2}	584
	PSO-NSFM	2.80	2.81	[4 7 18 15 16 32 24 29 26 16 35 20]	476	-	4505
	SFM	2.87	2.84	23	480	-	790
	MLP	3.01	3.02	-	[24 30]	-	2814
House Price-16H	VP	3.60×10^4	3.64×10^4	9	175	10	25119
	FF	3.66×10^4	3.71×10^4	8	124	1	23120
	PSO-NSFM	3.62×10^4	3.61×10^4	[5 9 4 26 39 10 6 4 11 10 24 16 13 12 14 30]	984	-	13201
	SFM	3.64×10^4	3.62×10^4	17	1065	-	3508
	MLP	3.64×10^4	3.63×10^4	-	[37 14]	-	15745
Elevators	VP	2.00×10^{-3}	2.00×10^{-3}	10	165	10^{-2}	10997
	FF	2.00×10^{-3}	2.10×10^{-3}	12	527	10^{-2}	9385
	PSO-NSFM	2.08×10^{-3}	2.11×10^{-3}	[16 17 5 18 49 17 13 7 9 16 15 11 4 11 5 15 16 12]	981	-	7197
	SFM	2.10×10^{-3}	2.12×10^{-3}	18	1776	-	3873
	MLP	2.10×10^{-3}	2.14×10^{-3}	-	[40 24]	-	10715
Ailerons	VP	1.58×10^{-4}	1.56×10^{-4}	6	12	10^{-2}	9040
	FF	1.58×10^{-4}	1.57×10^{-4}	9	69	10^{-5}	10936
	PSO-NSFM	1.57×10^{-4}	1.56×10^{-4}	[4 6 4 4 13 7 5 4 14 11 22 32 5 4 4 10 5 37 12 9 4 50 4 24 50 37 48 44 5 50 4 27 6 50 42 6 4 50 9 9]	237	-	4220
	SFM	1.59×10^{-4}	1.57×10^{-4}	13	311	-	2613
	MLP	1.59×10^{-4}	1.58×10^{-4}	-	[35 25]	-	8141

7.3 Συζήτηση Αποτελεσμάτων

Ξεκινώντας, σαν πρώτη εικόνα από τον δείκτη ελέγχου δοκιμής φαίνεται ο νικητής να είναι η μέθοδος PSO-NSFM σε 6 σύνολα δεδομένων, την δεύτερη θέση παίρνει ο VP σε 4 σύνολα δεδομένων ενώ σε 1 έρχεται ισοπαλία με την PSO-NSFM και τρίτος ο FF σε 1 σύνολο.

Όσον αφορά τον αριθμό των κέντρων παρατηρείται στην μέθοδο PSO-NSFM να είναι πολύ μεγαλύτερος σε σύγκριση με το VP. Ο αριθμός των κέντρων είναι μια πολύ σημαντική παράμετρος καθώς αυξάνει την απαίτηση της μνήμης και του υπολογιστικού χρόνου.

Σχετικά με τους χρόνους, τόσο ο VP όσο και ο FF δείχνουν να τα πηγαίνουν πολύ καλά ανταγωνιζόμενοι ο ένας τον άλλον για τον καλύτερο χρόνο στα περισσότερα σύνολα δεδομένων. Αυτό ενισχύει την προηγούμενη παρατήρηση του μειωμένου αριθμού κέντρων. Παρόλα αυτά και οι δύο μέθοδοι καθυστερούν πολύ όταν ο αριθμός των παραδειγμάτων είναι μεγάλος. Αυτό παρατηρείται στα τρία μεγαλύτερα σύνολα ενώ στα υπόλοιπα είναι πολύ μικρότερος ο χρόνος. Σίγουρα οι υπολογιστές με την πάροδο των χρόνων έχουν εξελιχθεί και δεν μπορεί να γίνει άμεση σύγκριση. Ενδεικτικά όμως στο Boston Housing ο PSO-NSFM χρειάστηκε 2588 δευτερόλεπτα ενώ ο VP μόλις 13, η διαφορά είναι αξιοσημείωτη.

Υπενθυμίζεται ότι η παράμετρος λ είναι σημαντική στην LM καθώς για μεγάλες τιμές οδηγεί στη μέθοδο απότομης κατάβασης ενώ για μικρές στη μέθοδο Newton. Παρότι δεν υπάρχει σύγκριση με των άλλων μεθόδων της βιβλιογραφίας, εξετάστηκαν 8 διαφορετικές τιμές. Τα αποτελέσματα δείχνουν στα περισσότερα σύνολα δεδομένων την τιμή του λ να είναι αρκετά μικρή.

Η μέθοδος VP φαίνεται να υπερισχύει της FF στα περισσότερα σύνολα δεδομένων. Μπορεί στον δείκτη ελέγχου δοκιμής να είναι μικρή η διαφορά τους αλλά η VP έχει στην πλειοψηφία καλύτερους χρόνους και λιγότερα κέντρα. Με βάση την θεωρία είναι λογικό καθώς η VP έχει τα βέλτιστα βάρη σε κάθε επανάληψη της LM γιατί προέρχονται από λύση γραμμικού προβλήματος ελαχίστων τετραγώνων. Αντίθετα, στην FF δεν υπάρχει εγγύηση για το αν τα βάρη είναι βέλτιστα γιατί γίνεται μεταβολή των κέντρων και των βαρών σε κάθε επανάληψη της LM. Παρόλα αυτά σε 3 σύνολα βγήκε καλύτερο το FF και αυτό θα συνέβη διότι βρέθηκε συνολικά καλύτερο τοπικό ελάχιστο ταυτόχρονα για τα βάρη και τα κέντρα.

7.4 Συμπεράσματα

Επαληθεύτηκε το γεγονός ότι η εκπαίδευση ενός νευρωνικού δικτύου ακτινικής βάσης είναι περίπλοκη διαδικασία, δεδομένου ότι πρέπει να υπολογιστούν οι κατάλληλες τιμές των παραμέτρων για την εξασφάλιση της ορθής λειτουργίας του δικτύου σε ένα ικανοποιητικό χρονικό διάστημα.

Σύμφωνα με [11] η εκπαίδευση μπορεί να θεωρηθεί ως ένα πολυδιάστατο πρόβλημα βελτιστοποίησης με ορισμένες ιδιαιτερότητες. Τα συναπτικά βάρη, οι θέσεις των κέντρων, ο αριθμός των μονάδων και τα εύρη επηρεάζουν άμεσα την απόδοση και τα χαρακτηριστικά του. Επίσης είναι σημαντική η επιλογή των αρχικών παραμέτρων του, για αυτόν τον λόγο πρέπει πρώτα να γίνεται η σωστή επιλογή τους και μετά η τροποποίηση τους.

Στην παρούσα εργασία γίνεται ανάπτυξη και αξιολόγηση δύο τεχνικών εκπαίδευσης δικτύων ακτινικής συνάρτησης βάσης με τον αλγόριθμο Levenberg-Marquardt, με τον οποίο υπολογίζονται οι τιμές των παραμέτρων του δικτύου RBF. Στην πρώτη τεχνική, με την μέθοδο Variable Projection, γίνεται διαχωρισμός των γραμμικών και μη γραμμικών παραμέτρων. Αρχικά γίνεται απαλοιφή τα βάρη και έπειτα με την ανάπτυξη της προβολής μεταβλητής προς τις μη γραμμικές παραμέτρους, ελαχιστοποιούνται τα κέντρα. Σε κάθε επανάληψη του αλγορίθμου LM τροποποιούνται τα κέντρα και υπολογίζονται τα βάρη με την επίλυση γραμμικού προβλήματος ελαχίστων τετραγώνων, για τα συγκεκριμένα κέντρα. Στην άλλη τεχνική, το πρόβλημα προς ελαχιστοποίηση είναι πλήρες συναρτησιακό καθώς δεν γίνεται διαχωρισμός των παραμέτρων. Σε κάθε επανάληψη του αλγορίθμου, ενημερώνονται μαζί με τα κέντρα και τα βάρη. Ο αρχικός υπολογισμός των τιμών των κέντρων γίνεται από τον αλγόριθμο fuzzy means και για τις τιμές των ευρών, των συναρτήσεων βάσης, από την μέθοδο p-κοντινότερων γειτόνων.

Με την μέθοδο VP εξασφαλίζεται σε κάθε επανάληψη του αλγορίθμου LM τα ολικά βέλτιστα βάρη για τα συγκεκριμένα κέντρα. Από την άλλη πλευρά η μέθοδος FF δεν εξασφαλίζει ολικά βέλτιστα βάρη επομένως μπορεί να συγκλίνει και σε τοπικά βέλτιστα. Στα αποτελέσματα, όμως, φαίνεται ανταγωνιστικός ως προς την ταχύτητα σύγκλισης αλλά και με μικρή διαφορά στο δείκτη σφάλματος.

Οι δύο τεχνικές δοκιμάστηκαν σε 12 προσομοιωμένα αλλά και πραγματικά σύνολα δεδομένων της βιβλιογραφίας για 8 διαφορετικές τιμές της παραμέτρου λ του αλγορίθμου LM. Σε 8 διαφορετικά σύνολα fuzzy και μετρήθηκαν οι δείκτες σφάλματος των συνόλων καθώς και ο χρόνος σύγκλισης. Η επιλογή της καλύτερης μεθόδου έγινε με βάση με τον δείκτη σφάλματος και κρίθηκαν ιδιαίτερα ανταγωνιστικές σε σύγκριση με άλλες μεθοδολογίες της βιβλιογραφίας.

Τα αποτελέσματα των μεθόδων που χρησιμοποιήθηκαν κρίθηκαν ικανοποιητικά, παρόλα αυτά πάντα θα υπάρχει περιθώριο βελτίωσης. Μελλοντικά θα μπορούσε να :

- χρησιμοποιηθούν διαφορετικές συναρτήσεις βάσης
- προστεθούν περιορισμοί για τις θέσεις των κέντρων
- δοκιμαστούν μεταβαλλόμενα εύρη στις συναρτήσεις
- τροποποιηθούν οι μέθοδοι της αυτόματης επιλογής των κέντρων
- αλλάξουν τα κριτήρια τερματισμού του αλγορίθμου

8 Βιβλιογραφία – Αναφορές - Διαδικτυακές Πηγές

- [1].Harkness Jon., A Lifetime of Connections: Otto Herbert Schmitt, 1913-1998. Physics in Perspective (PIP), (2002), pp. 456–490
- [2].Korovesis N., D. Kandris, G. Koulouras, A. Alexandridis, “Robot Motion Control via an EEG-Based Brain–Computer Interface by Using Neural Networks and Alpha Brainwaves”, **Electronics**, 8(12) (2019), 1387.
- [3].Alexandridis, A.P., C.I. Siettos, H.K. Sarimveis, A.G. Boudouvis, G.V. Bafas, “Modeling of nonlinear process dynamics using Kohonen’s Neural Networks, Fuzzy Systems and Chebyshev Series”, **Computers and Chemical Engineering**, 26 (2002), pp. 479–486.
- [4].Alexandridis, A., E. Paizis, E. Chondrodima, M. Stogiannos, “A particle swarm optimization approach in printed circuit board thermal design”, **Integrated Computer-Aided Engineering**, 24(2) (2017), pp. 143-155.
- [5].Kapnopoulos, A., A. Alexandridis, “A cooperative particle swarm optimization approach for tuning an MPC-based quadrotor trajectory tracking scheme”, **Aerospace Science and Technology**, 127C (2022) 107725.
- [6].Papadimitrakis, M., A. Kapnopoulos, S. Tsavartzidis, A. Alexandridis, “A cooperative PSO algorithm for Volt-VAR optimization in smart distribution grids”, **Electric Power Systems Research**, In Press, 212 (2022), 108618
- [7].Papadimitrakis, N., N. Giamarelos, M. Stogiannos, E.N. Zois, N.A. Livanos, A. Alexandridis, “Metaheuristic search in smart grid: A review with emphasis on planning, scheduling and power flow optimization applications”, **Renewable and Sustainable Energy Reviews**, 145 (2021), 111072.
- [8].Stogiannos, M., A. Alexandridis, H. Sarimveis, “An enhanced decentralized artificial immune-based strategy formulation algorithm for swarms of autonomous vehicles”, **Applied Soft Computing**, 89 (2020), 106135.
- [9].Engelbrecht A. P., Computational Intelligence: An Introduction, Second Edition, (2007).
- [10].Ι. Βλαχάβας, Π. Κεφαλάς, Ν.Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου, Η Τεχνητή Νοημοσύνη, Δ έκδοση, Πανεπιστήμιο Μακεδονίας, (2011).
- [11].Alexandridis A., Chondrodima E., Sarimveis H., Radial Basis Function Network Training Using a Nonsymmetric Partition of the Input Space and Particle Swarm Optimization, **IEEE Transactions on Neural Networks and Learning Systems**, (2013), 24(2):219230.

- [12]. Alexandridis A., Sarimveis, H., Bafas, G., A new algorithm for online structure and parameter adaptation of RBF networks, **Neural Networks**, (2003), 16(7):10031017.
- [13]. Sarimveis H., Alexandridis A., Mazarakis S., and Bafas G., A new algorithm for developing dynamic radial basis function neural network models based on genetic algorithms. **Comput. Chem. Eng.**, (2004), 28:209{217.
- [14]. Madsen K., Nielsen H.B, Tingleff O., Methods for Non-Linear Least Squares Problems 2nd end. Informatics and Mathematical Modelling, **Technical University of Denmark**, (2004), DTU, Richard Petersens Plads, Building 321, DK2800, Kgs. Lyngby.
- [15]. Marquardt D. An Algorithm for Least Squares Estimation on Nonlinear Parameters, **Journal of the Society for Industrial and Applied Mathematics**, (1963), 11:431{441.
- [16]. Lloyd S. P., Least Squares Quantization in PCM, **IEEE Transactions on Information Theory**, (1982), 28(2):129137.
- [17]. Alexandridis A., Evolving RBF neural networks for adaptive soft-sensor design. **International Journal of Neural Systems**, (2013), 23(6):1350029.
- [18]. Alexandridis A., Stogiannos M., Kyriou A., Sarimveis H., An offset-free neural controller based on approximating the inverse process dynamics. **Journal of Process Control**, (2013), 23(7):968{979.
- [19]. Alexandridis A., Chondrodima E., A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing. **Journal of Biomedical Informatics**, (2014), 49:6172.
- [20]. Alexandridis A, Chondrodima E., Efthimiou E., Papadakis G., Vallianatos F., Triantis D., Large earthquake occurrence estimation based on radial basis function neural networks. **IEEE Transactions on Geoscience and Remote Sensing**, (2014), 52(9):54435453.
- [21]. Stogiannos M., Alexandridis A., Sarimveis H., Model predictive control for systems with fast dynamics using inverse neural models. **ISA Transactions**, (2018), 72:161177.
- [22]. Karamichailidou, D., S. Koletsios, A. Alexandridis, “An RBF online learning scheme for non-stationary environments based on fuzzy means and Givens rotations”, **Neurocomputing**, 501C (2022) 370-386.
- [23]. Papadimitrakis, M., A. Alexandridis, “Active vehicle suspension control using road preview model predictive control and radial basis function networks”, **Applied Soft Computing**, 120C (2022), 108646.

- [24]. Karamichailidou, D., A. Alexandridis, G. Anagnostopoulos, G. Syriopoulos, O. Sekkas, “Modeling biogas production from anaerobic wastewater treatment plants using radial basis function networks and differential evolution”, **Computers and Chemical Engineering**, 157C (2022), 107629.
- [25]. Karamichailidou D., V. Kaloutsas, A. Alexandridis, “Wind Turbine Power Curve Modeling Using Radial Basis Function Neural Networks and Tabu Search”, **Renewable Energy**, 163 (2021), 2137-2152.
- [26]. Papadimitrakis, M. M. Stogiannos, H. Sarimveis, A. Alexandridis, “Multi-ship control and collision avoidance using MPC and RBF-based trajectory predictions”, **Sensors**, 21(21) (2021) 6959.
- [27]. Alexandridis, A., M. Stogiannos, N. Papaioannou, E. Zois, H. Sarimveis, “An Inverse Neural Controller Based on the Applicability Domain of RBF Network Models”, **Sensors**, 18(1) (2018), 315.
- [28]. Alexandridis A, E. Chondrodima, “A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing”, **Journal of Biomedical Informatics**, 49 (2014), pp. 61-72.
- [29]. Alexandridis, A., H. Sarimveis, K. Ninos, “RBF network training using a non-symmetric partition of the input space – Application to an MPC configuration”, **Advances in Engineering Software**, 42(10) (2011), pp. 830-837.
- [30]. Alexandridis A., Chondrodima E., Sarimveis H., Cooperative learning for radial basis function networks using particle swarm optimization. **Applied Soft Computing**, (2016), 49:485497.
- [31]. Alexandridis A., Chondrodima E., Giannopoulos N., Sarimveis H., A Fast and Efficient Method for Training Categorical Radial Basis Function Networks. **IEEE Transactions on Neural Networks and Learning Systems**, (2017), 28(11):2831-2836.
- [32]. Barr. A. Feigenbaum, The handbook of Artificial Intelligence, Vols. I, II & III, William Kaufmann, (1981).
- [33]. Artificial intelligence at Merriam Webster Dictionary [Online]. Available: [https://www.merriam-webster.com/dictionary/artificial intelligence](https://www.merriam-webster.com/dictionary/artificial%20intelligence), Accessed 5 April (2022).
- [34]. Artificial intelligence at Cambridge Dictionary [Online]. Available: <https://dictionary.cambridge.org/dictionary/english/artificial-intelligence>, Accessed 5 April (2022).
- [35]. A. M. Turing, Computing Machinery and Intelligence, Mind, New Series, (1950).

- [36]. Bruce MacLennan, History of Artificial Intelligence Before Computers, **University of Tennessee USA**, (2009).
- [37]. Buchanan B. G., A (very) brief history of artificial intelligence, **AI Magazine**, (2005), 26(4):5360.
- [38]. Διαμαντάρας Κωνσταντίνος, Μπότσης Α. Δημήτρης, Μηχανική Μάθηση, Εκδόσεις κλειδάριθμος, 1^η έκδοση, (2019).
- [39]. Zois, E.N., A. Alexandridis, G. Economou, “Writer independent offline signature verification based on asymmetric pixel relations and unrelated training-testing datasets”, **Expert Systems with Applications**, 125 (2019), pp. 14-32
- [40]. M. Hellmann, Fuzzy Logic Introduction, **Jour**, (2001).
- [41]. Dan Simon, Evolutionary Optimization Algorithms, **Wiley**, (2013).
- [42]. Norman Matloff, Statistical Regression and Classification from Linear Models to Machine Learning, **CRC Press**, (2017)
- [43]. Simon Haykin, Bernard Widrow, Least-Mean-Square Adaptive Filters, **Wiley**, (2003).
- [44]. Simon Haykin, Νευρωνικά Δίκτυα και Μηχανική Μάθηση, 3^η έκδοση, Παπασωτηρίου (2010).
- [45]. Arnold Thackray, David C. Brock, Rachel Jones. Technology and Culture Moore’s Law: The Life of Gordon Moore, Silicon Valley’s Quiet Revolutionary, (2015).
- [46]. Bors, A. G. Introduction of the Radial Basis Function (RBF) Network, (2001).
- [47]. Morissete L., Chartierm S., The k-means clustering technique: General considerations and implementation in Mathematica, *Tutorials in Quantitative Methods for Psychology*, (2013), 9(1):1524.
- [48]. Sarimveis H., Alexandridis A., Tsekouras, G., Bafas G., A fast and efficient algorithm for training radial basis function neural networks based on a fuzzy partition of the input space, **Industrial and Engineering Chemistry Research**, (2002), 41(4):751{759.
- [49]. Optimization at Britannica, [Online]. Available: <https://www.britannica.com/science/optimization>, Accessed 20 April 2022.
- [50]. Hansen C., Pereyra V., Scherer G. Least Squares Data Fitting with Applications, **Johns Hopkins University Press**, (2012), Project MUSE.
- [51]. Kaufman L., A variable projection method for solving separable nonlinear least squares problems, **BIT Numerical Mathematics**, (1973), 15(1):49{57.

- [52]. Golub G. H, Pereyra V., Separable nonlinear least squares: the variable projection method and its applications, **Inverse Problems**, (2003), 10(2):413{432.
- [53]. Golub G. H, Pereyra V., The differentiation of pseudoinverses & nonlinear least squares problems whose variables separate, *IAM Journal*, (1973), 15:209{18.
- [54]. Golub G. H, Van Loan F., *Matrix Computations*, 3rd Edition. London: The Johns Hopkins Press Ltd, (1996).
- [55]. Καραμιχαλίδου, Δέσποινα Α., “Εκπαίδευση νευρωνικών δικτύων τύπου RBF με μεθόδους βελτιστοποίησης”, **Πανεπιστήμιο Θεσσαλίας**, (2018).

Παράρτημα Α : Αλγόριθμος Fuzzy Means

Ο αλγόριθμος που πραγματοποιεί την επιλογή των ασαφών υποχώρων και τελικά των κέντρων του μοντέλου έχει ως εξής [11], [12], [48], [55].

Αλγόριθμος fuzzy means

Είσοδος: κανονικοποιημένες μεταβλητές εισόδου N , $i = 1, 2, \dots, N$
των δεδομένων εισόδου K , $k = 1, 2, \dots, K$
ακέραιος αριθμός ασαφών συνόλων s
για το διαχωρισμό της κάθε διεύθυνσης εισόδου

Έξοδος: αριθμός κέντρων δικτύου L

$$\text{θέσεις κέντρων } \hat{U} = [\hat{u}_1, \hat{u}_2, \dots, \hat{u}_L]^T$$

1: διαχωρισμός του χώρου κάθε μεταβλητής εισόδου x_i , $i = 1, 2, \dots, N$
σε s ασαφή σύνολα με κέντρο a και εύρος δa

2: 1^ο δεδομένο εισόδου, $k = 1$

3: υπολογισμοί για το 1^ο κέντρο RBF, $L = 1$

4: **για** $i = 1, 2, \dots, N$

5: υπολογισμός του ασαφούς συνόλου j το οποίο
δίνει τη μεγαλύτερη συγγένεια σε κάθε διεύθυνση i

$$A_i^1 = \{a_i^1, \delta a\} \leftarrow \max_{j=1,2,\dots,s} [\mu_{A_{i,j}}(x_i(1))]$$

6: **τέλος**

7: Πρώτο κέντρο δικτύου, \hat{u}_1

$$\hat{u}_1 = [a_1^1, a_2^1, \dots, a_N^1]$$

8: **για** $k = 1, 2, \dots, K$

9: **αν** το σημείο k δεν καλύπτεται από τα υπάρχοντα κέντρα: $\min_{l=1,2,\dots,L} [r^l(x(k))] > 1$ **τότε**

10: προσθήκη νέου κέντρου: $L \leftarrow L + 1$

11: **για** $i = 1, 2, \dots, N$

12: υπολογισμός του ασαφούς συνόλου j το οποίο
δίνει τη μεγαλύτερη συγγένεια σε κάθε διεύθυνση i

$$A_i^L = \{a_i^L, \delta a\} \leftarrow \max_{j=1,2,\dots,s} [\mu_{A_{i,j}}(x_i(k))]$$

13: **τέλος**

14: Επόμενο κέντρο δικτύου \hat{u}_L

$$\hat{u}_L = [a_1^L, a_2^L, \dots, a_N^L]$$

15: **τέλος**

16: **τέλος**

Παράρτημα Β : Αλγόριθμος Variable Projection, Levenberg-Marquardt

Ο αλγόριθμος με τον οποίον γίνεται το VP και LM είναι ο εξής [55]:

Αλγόριθμος Variable Projection, Levenberg-Marquardt

$$\mathbf{c}_j^0 \in \mathbb{R}^n, j = 1, 2, \dots, N, \lambda_0 > 0$$

1: υπολογισμός της παραγοντοποίησης QR του $\Phi(\mathbf{c}^0)$

$$\Phi(\mathbf{c}^0) = QR = [Q_1 \quad Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1$$

2: υπολογισμός των βαρών $\mathbf{w}^0 = R_1^{-1} Q_1^T \mathbf{y}$

3: υπολογισμός κόστους $h(\mathbf{c}^0) = \frac{1}{2} \|Q_2^T \mathbf{y}\|^2$

4: για $k=0,1,2,\dots$ επανάλαβε

5: σχηματισμός των $A_k = -Q_2^T D(\Phi(\mathbf{c}^k)) \mathbf{w}^k, \mathbf{b}_k = -Q_2^T \mathbf{y}$

6: υπολογισμός της διεύθυνσης $\mathbf{d}^k \in \mathbb{R}^{Nn}$ λύνοντας το εξής κανονικοποιημένο πρόβλημα ελαχίστων τετραγώνων

$$\min_{\mathbf{d}} \frac{1}{2} \|A(\mathbf{c})\mathbf{d} - \mathbf{b}\|_2^2 + \frac{\lambda}{2} \|\mathbf{d}\|_2^2$$

7: υπολογισμός της παραγοντοποίησης QR του $\Phi(\mathbf{c}^k + \mathbf{d}^k)$

$$\Phi(\mathbf{c}^k + \mathbf{d}^k) = \bar{Q}\bar{R} = [\bar{Q}_1 \quad \bar{Q}_2] \begin{bmatrix} \bar{R}_1 \\ 0 \end{bmatrix} = \bar{Q}_1 \bar{R}_1$$

8: υπολογισμός κόστους $h(\mathbf{c}^k + \mathbf{d}^k) = \frac{1}{2} \|\bar{Q}_2^T \mathbf{y}\|^2$

9: $r_k = \frac{h(\mathbf{c}^k) - h(\mathbf{c}^k + \mathbf{d}^k)}{h(\mathbf{c}^k) - \frac{1}{2} \|\mathbf{R}(\mathbf{c}) + D\mathbf{R}(\mathbf{c})\mathbf{d}\|^2}$

10: αν $r_k < \frac{1}{4}$

11: $\lambda_{k+1} = 4\lambda_k$

12: αλλιώς αν $r_k > \frac{3}{4}$

13: $\lambda_{k+1} = \frac{\lambda_k}{2}$

αλλιώς

14: $\lambda_{k+1} = \lambda_k$

15: αν $r_k \leq 0$

16: $\mathbf{c}_{k+1} = \mathbf{c}_k$

αλλιώς

17: $\mathbf{c}_{k+1} = \mathbf{c}_k$

18: $\mathbf{w}_{k+1} = \bar{R}_1^{-1} \bar{Q}_1^T \mathbf{y}$

19: $Q_2 \leftarrow \bar{Q}_2$
