



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
Τμήμα Μηχανολόγων Μηχανικών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δέσποινα Καραμυχαλίδου

*«Εκπαίδευση νευρωνικών δικτύων
τύπου RBF με μεθόδους
βελτιστοποίησης»*

Εξεταστική επιτροπή
Αλέξανδρος Αλεξανδρίδης, Καθηγητής ΠΔΑ (Επιβλέπων)
Αλέξης Κερμανίδης, Επ. Καθηγητής ΠΘ
Γιώργος Σαχαρίδης, Επ. Καθηγητής ΠΘ

Ημερομηνία Υποβολής: 15 - 06 - 2018
Ακαδημαϊκό έτος: 2017 - 2018
Βόλος

© 2018 Δέσποινα Καραμιχαλίδου

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανολόγων Μηχανικών της Πολυτεχνικής Σχολής του Πανεπιστημίου Θεσσαλίας δεν υποδηλώνει αποδοχή των απόψεων του συγγραφέα (Ν. 5343/32 αρ. 202 παρ. 2)

Ευχαριστίες

Με την ολοκλήρωση της διπλωματικής μου εργασίας, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Αλέξανδρο Αλεξανδρίδη για την ευκαιρία που μου έδωσε αναθέτοντας μου το συγκεκριμένο θέμα, καθώς, επίσης, και για την εμπιστοσύνη που μου έδειξε. Η πολύτιμη βοήθειά του και οι γνώσεις που απέκτησα χάρη σε αυτόν, καθ' όλη τη διάρκεια μιας άψογης συνεργασίας, σε αμέτρητες ώρες διδακτικών συζητήσεων, είναι πολύτιμες. Η επιστημονική και ηθική ενθάρρυνση, αλλά και η συνεισφορά του στην εκπόνηση της παρούσας μελέτης είναι ανεκτίμητη.

Είμαι ευγνώμων στον καθηγητή κ. Γρηγόρη Χαϊδεμενόπουλο για τη σύντομη αλλά κατά τα άλλα σπουδαία συνεργασία με τον ίδιο και την ομάδα του. Η καθοδήγησή του σε μια δύσκολη απόφαση στάθηκε αφορμή για την αφετηρία ενός πολλά υποσχόμενου ταξιδιού.

Ευχαριστώ τους καθηγητές κ. Αλέξη Κερμανίδη και κ. Γιώργο Σαχαρίδη για το χρόνο και την υπομονή που διέθεσαν για τη μελέτη και την αξιολόγηση της εργασίας.

Ακόμα, θα ήθελα να ευχαριστήσω όλους τους καθηγητές και συμφοιτητές ανεξαιρέτως για όλα αυτά που μου πρόσφεραν, ο καθένας με το δικό του τρόπο, στα χρόνια φοίτησής μου.

Θα ήθελα να ευχαριστήσω την οικογένεια μου για την κατανόηση και την ενθάρρυνση που έδειξαν όλο αυτό το διάστημα, αλλά και την οικονομική υποστήριξη που μου παρείχαν όλα αυτά τα χρόνια.

Φεύγοντας, λοιπόν, όλο και περισσότερο συνειδητοποιώ πως δεν παίρνω μαζί μου μόνο γνώσεις ή αναμνήσεις, αλλά κάτι σπουδαιότερο, πολυτιμότερο που έχει χαραχτεί βαθειά μέσα στην καρδιά μου, ενώ συνεχώς ένα πράγμα μου έρχεται στο μυαλό... *«Εὔρες φίλον καλόν, εὔρες θησαυρόν...»*

Καραμιχαηλίδου Δέσποινα
Βόλος, Ιούνιος 2018

«...πάντα ἐν σοφίᾳ ἐποίησας...»
(Ψλμ ΡΔ)

Περίληψη

Η παρούσα εργασία ασχολείται με την ανάπτυξη μεθόδων για την εκπαίδευση νευρωνικών δικτύων τύπου RBF (Radial Basis Function), με σκοπό την κατασκευή υπολογιστικών μοντέλων υψηλής ακρίβειας σε μικρό υπολογιστικό χρόνο. Η προτεινόμενη μέθοδος εκπαίδευσης βασίζεται στον αλγόριθμο βελτιστοποίησης Levenberg-Marquardt, ο οποίος χρησιμοποιείται για τον προσδιορισμό των παραμέτρων του δικτύου, δηλαδή των συντεταγμένων των κέντρων RBF, των συναπτικών βαρών και των ευρών των συναρτήσεων βάσης. Αναπτύχθηκαν δύο παραλλαγές της μεθόδου, στην πρώτη εκ των οποίων πραγματοποιείται διαχωρισμός των συνόλων των παραμέτρων του συνολικού προβλήματος της εκπαίδευσης. Η απαλοιφή των γραμμικών παραμέτρων, δηλαδή των βαρών του δικτύου, και η ανάπτυξη της προβολής μεταβλητής (Variable Projection) συναρτήσεων των μη γραμμικών παραμέτρων, δηλαδή των κέντρων των μονάδων RBF, οδηγεί σε δύο μικρότερα υποπροβλήματα, ένα για το κάθε σύνολο παραμέτρων. Τα κέντρα ενημερώνονται σε κάθε επανάληψη του αλγορίθμου Levenberg-Marquardt, ενώ τα βάρη προέρχονται από τη λύση ενός γραμμικού προβλήματος ελαχίστων τετραγώνων. Στη δεύτερη παραλλαγή, χωρίς να προηγείται διαχωρισμός των συνόλων των παραμέτρων, το πρόβλημα προς επίλυση αφορά το πλήρες συναρτησιακό (Full Functional), με τα κέντρα και τα βάρη να ενημερώνονται σε κάθε επανάληψη του αλγορίθμου Levenberg-Marquardt. Οι θέσεις των αρχικών κέντρων προέρχονται από τον αλγόριθμο των ασαφών μέσων, ενώ οι αρχικές τιμές των ευρών των συναρτήσεων βάσης από τη μέθοδο των p -κοντινότερων γειτόνων. Η προτεινόμενη μεθοδολογία εκπαίδευσης δοκιμάστηκε σε σύνολα δεδομένων της βιβλιογραφίας, πραγματικά και προσομοιωμένα, ενώ έγινε σύγκριση με κλασικές μεθόδους εκπαίδευσης δικτύων RBF, αλλά και με διαφορετικές μεθόδους μηχανικής μάθησης. Τα αποτελέσματα δείχνουν την υπεροχή των προτεινόμενων μεθόδων έναντι των υπολοίπων τεχνικών, λαμβάνοντας υπόψη την ακρίβεια των μοντέλων σε συνδυασμό με τους υπολογιστικούς χρόνους.

Λέξεις-κλειδιά: αλγόριθμος ασαφών μέσων, δίκτυα ακτινικής συνάρτησης βάσης, αλγόριθμος Levenberg-Marquardt, μέθοδος προβολής μεταβλητής, τεχνητά νευρωνικά δίκτυα, υπολογιστική νοημοσύνη

Abstract

The objective in this study is the development of methods for training radial basis function (RBF) networks, in order to produce models with increased accuracy in short computational times. The proposed methodology is based on the Levenberg-Marquardt optimization algorithm, which is used to determine the network parameters, such as the coordinates of RBF centers, the synaptic weights and the widths of RBFs. Two variants of the method were developed, the first of which deals with the separation of the sets of variables of the overall training problem. The elimination of linear parameters, i.e. network weights, and the development of Variable Projection as a function of nonlinear parameters, i.e. the RBF centers, leads to two smaller subproblems, one for each set of parameters. The centers are updated in each iteration of the Levenberg-Marquardt algorithm, while the synaptic weights are calculated using linear regression. In the second variant, where the sets of parameters are not separated, the problem to be solved is called Full Functional; in this case, both the centers and the synaptic weights are updated in each iteration of the Levenberg-Marquardt algorithm. The initial locations of the RBF centers are derived from the fuzzy means algorithm, while the initial values of the widths of RBFs are derived from the p -nearest neighbor heuristic. The proposed training methodology was tested in real-world and synthetic benchmark datasets and is also compared with other neural network training techniques. The results show the superiority of the proposed methods over the other techniques, taking into account the accuracy of the models in combination with the computational times.

Keywords: artificial neural networks, computational Intelligence, fuzzy means algorithm, Levenberg-Marquardt algorithm, radial basis function networks, Variable Projection

Περιεχόμενα

Περίληψη	i
Περιεχόμενα	iii
Κατάλογος Σχημάτων	v
Κατάλογος Πινάκων	vii
Κεφάλαιο 1 Εισαγωγή	1
1.1 Κίνητρο της εργασίας	1
1.2 Συμβολή της εργασίας	2
1.3 Δομή εργασίας	3
Κεφάλαιο 2 Υπολογιστική Νοημοσύνη και Νευρωνικά Δίκτυα	4
2.1 Εισαγωγή στην Τεχνητή και Υπολογιστική Νοημοσύνη	4
2.2 Βασικές Μέθοδοι της Υπολογιστικής Επιστήμης	8
2.2.1 Εξελικτικός Υπολογισμός	9
2.2.2 Νοημοσύνη Σμήνους	10
2.2.2.1 Βελτιστοποίηση Σμήνους Σωματιδίων	11
2.2.2.2 Βελτιστοποίηση με Αποκίες Μυρμηγκιών	12
2.2.3 Ασαφή Συστήματα	13
2.3 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα	16
2.3.1 Βιολογικοί Νευρώνες και Βιολογικά Νευρωνικά Δίκτυα	16
2.3.2 Τεχνητός Νευρώνας	17
2.3.2.1 Αναλογίες με τους Βιολογικούς Νευρώνες	18
2.3.3 Μοντέλο Τεχνητού Νευρώνα	18
2.3.3.1 Συναρτήσεις ενεργοποίησης	19
2.3.3.2 Γεωμετρική ερμηνεία ενός Τεχνητού Νευρώνα	22
2.3.4 Δομή ενός Τεχνητού Νευρωνικού Δικτύου	23
2.3.5 Αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων	25
2.3.6 Ιδιότητες Τεχνητών Νευρωνικών Δικτύων	26
2.3.7 Είδη Μηχανικής Μάθησης	28
2.3.7.1 Επποπευόμενη μάθηση	31
2.3.7.2 Μη επποπευόμενη μάθηση	31
Κεφάλαιο 3 Νευρωνικά Δίκτυα Ακτινικής Βάσης (Radial Basis Function Networks)	32
3.1 Δομή ενός δικτύου RBF	32
3.2 Συναρτήσεις βάσης ακτινικής συμμετρίας	34
3.3 Πλεονεκτήματα δικτύων RBF	36
3.4 Εκπαίδευση ενός δικτύου RBF	36
3.4.1 Μορφοποίηση του προβλήματος της εκπαίδευσης ενός δικτύου RBF	36
3.4.2 Υπολογισμός των συναπτικών βαρών ενός δικτύου RBF	38

3.4.3	Υπολογισμός των κέντρων ενός δικτύου RBF	39
3.4.3.1	Αλγόριθμος k-means	39
3.4.3.2	Αλγόριθμος fuzzy means	44
3.4.4	Υπολογισμός των ευρών των συναρτήσεων βάσης ενός δικτύου RBF	49
Κεφάλαιο 4	Αλγόριθμοι Βελτιστοποίησης	50
4.1	Εισαγωγικές έννοιες	50
4.2	Μορφοποίηση προβλήματος βελτιστοποίησης	50
4.3	Μέθοδοι κατάβασης	53
4.3.1	Υπολογισμός διεύθυνσης κατάβασης	54
4.3.1.1	Μέθοδος απότομης κατάβασης	54
4.3.1.2	Μέθοδος Newton	55
4.3.2	Υπολογισμός μεγέθους βήματος	56
4.4	Μέθοδοι περιοχής εμπιστοσύνης	57
4.5	Μέθοδοι για την επίλυση μη γραμμικών προβλημάτων ελαχίστων τετραγώνων	59
4.5.1	Μέθοδος Gauss - Newton	59
4.5.2	Μέθοδος Levenberg-Marquardt	60
4.6	Κριτήρια τερματισμού επαναληπτικών μεθόδων	61
4.7	Ειδική κατηγορία προβλημάτων μη γραμμικών ελαχίστων τετραγώνων	62
4.7.1	Μέθοδος Variable Projection	62
4.7.2	Μέθοδοι για την επίλυση της ελαχιστοποίησης του μειωμένου συναρτησιακού Variable Projection	66
Κεφάλαιο 5	Προτεινόμενη μέθοδος εκπαίδευσης δικτύων RBF	67
5.1	Μορφοποίηση του προβλήματος της εκπαίδευσης ενός δικτύου RBF	67
5.2	Εφαρμογή της μεθόδου Variable Projection	69
5.2.1	Υπολογισμός της κλίσης των συναρτήσεων βάσης	69
5.2.2	Υπολογισμός της Ιακωβιανής του πίνακα $\Phi(c)$	69
5.2.3	Υπολογισμός των ευρών των συναρτήσεων βάσης	70
5.2.4	Υπολογισμός των συναπτικών βαρών	71
5.2.5	Υπολογισμός του κόστους	72
5.2.6	Υπολογισμός των κέντρων των συναρτήσεων βάσης	72
5.2.7	Εφαρμογή του αλγορίθμου Levenberg-Marquardt	73
5.2.8	Περιγραφή του αλγορίθμου Variable Projection - Levenberg-Marquardt	75
5.3	Εκπαίδευση ενός δικτύου RBF βάσει των μεθόδων Variable-Projection και Levenberg-Marquardt	76
5.4	Περίπτωση πλήρους συναρτησιακού	77
5.5	Περίπτωση μεταβλητών ευρών για τις συναρτήσεις βάσης	78
Κεφάλαιο 6	Πειραματική διαδικασία	79
6.1	Περιγραφή των benchmark datasets	79
6.2	Παρουσίαση και συζήτηση αποτελεσμάτων	80
Κεφάλαιο 7	Συμπεράσματα	93
Βιβλιογραφία	95

Κατάλογος Σχημάτων

2.1	Τεχνητή Νοημοσύνη και άλλες επιστήμες.	5
2.2	Συγγραφείς μυθιστορημάτων ως πηγή έμπνευσης για την Τεχνητή Νοημοσύνη.	5
2.3	Βασικές μέθοδοι της Υπολογιστικής Νοημοσύνης.	8
2.4	Χρωμοσώματα και γονίδια.	9
2.5	Διαδικασία εξέλιξης σε έναν γενετικό αλγόριθμο.	10
2.6	Αλλαγή θέσης ενός σωματιδίου του πληθυσμού.	11
2.7	Αναζήτηση της βέλτιστης λύσης από τα σωματίδια του πληθυσμού.	12
2.8	Αναζήτηση του συντομότερου μονοπατιού από τα μυρμήγκια.	13
2.9	Διαφορά κλασικού συνόλου από ένα ασαφές σύνολο.	14
2.10	Συνάρτηση συγγένειας για ένα ασαφές σύνολο.	15
2.11	Δομή και σύνδεση βιολογικών νευρώνων.	16
2.12	Σκαρίφημα βιολογικού νευρώνα.	17
2.13	Τεχνητός νευρώνας.	17
2.14	Αναλογίες μεταξύ τεχνητού και βιολογικού νευρώνα.	18
2.15	Μοντέλου τεχνητού νευρώνα.	18
2.16	Συναρτήσεις ενεργοποίησης.	19
2.17	Συνάρτηση ράμπας.	20
2.18	Σιγμοειδής συνάρτηση.	20
2.19	Σιγμοειδής συνάρτηση και παράγωγος.	21
2.20	Συνάρτηση Gaussian.	21
2.21	Γεωμετρική ερμηνεία διανύσματος βαρών.	22
2.22	Ρόλος της πόλωσης.	22
2.23	Τυπική δομή ενός νευρωνικού δικτύου.	23
2.24	Τεχνητό νευρωνικό δίκτυο δόμης 3-4-1.	23
2.25	Τεχνητά νευρωνικά δίκτυα πλήρους και μερικής διασύνδεσης.	25
2.26	Τεχνητά νευρωνικά δίκτυα πρόσθιας τροφοδότησης και ανατροφοδότησης.	25
2.27	Βαθμός μάθησης.	26
2.28	Ροή σφάλματος στα σύνολα επικύρωσης και εκπαίδευσης.	27
2.29	Μορφές συνόλου δεδομένων.	28
2.30	Είδη μηχανικής μάθησης.	28
2.31	Κατηγοριοποίηση σε γνωστές ομάδες - πρόβλημα εποπτευόμενης μάθησης.	29
2.32	Κατηγοριοποίηση σε άγνωστες ομάδες - πρόβλημα μη εποπτευόμενης μάθησης.	30
2.33	Εκπαιδευόμενος στην ενισχυτική μάθηση.	30
2.34	Προβλήματα εποπτευόμενης μάθησης.	31
2.35	Πρόβλημα μη εποπτευόμενης μάθησης.	31
3.1	Τυπική αρχιτεκτονική ενός δικτύου RBF.	32
3.2	Συνολική επιφάνεια ορισμένη από τις συναρτήσεις βάσης.	33
3.3	Συναρτήσεις βάσης τοπικού χαρακτήρα.	34
3.4	Συναρτήσεις βάσης μη τοπικού χαρακτήρα.	35
3.5	Εφαρμογή του αλγορίθμου k-means σε ένα τυχαίο σύνολο δεδομένων.	41
3.6	Επαναλήψεις του αλγορίθμου στο τυχαίο σύνολο δεδομένων.	41

3.7	Επαναλήψεις του αλγορίθμου στο τυχαίο σύνολο δεδομένων.	42
3.8	Επαναλήψεις του αλγορίθμου στο τυχαίο σύνολο δεδομένων.	42
3.9	Συνολική βελτίωση των θέσεων των κέντρων στο τυχαίο σύνολο δεδομένων με τον αλγόριθμο k-means.	43
3.10	Ασαφές σύνολο.	44
3.11	Ασαφής διαχωρισμός του χώρου μιας μεταβλητής.	45
3.12	Διαχωρισμός ενός διδιάστατου χώρου εισόδου με ασαφείς υποχώρους.	46
3.13	Υπολογισμός των βαθμών αλήθειας μιας μεταβλήτης με τα ασαφή σύνολα.	47
3.14	Επιλογή ευρών συναρτήσεων βάσης.	49
4.1	Σχηματιζόμενη γωνία μιας διεύθυνσης με τη κλίση σε συγκεκριμένο σημείο.	54
4.2	Ορισμός διεύθυνσης κατάβασης.	55
4.3	Τυπική μορφή του μοντέλου συναρτήσεων του βήματος.	57
4.4	Προβολή στο χώρο στηλών ενός πίνακα 3 επί 2.	64
4.5	Χώρος στηλών ενός πίνακα 3 επί 2 με το ορθογώνιο συμπλήρωμα του.	65
5.1	Τυπική δομή ενός δικτύου RBF.	68
6.1	Μεταβολή του κόστους στο σύνολο επικύρωσης για τους διάφορους αλγορίθμους με εκκίνηση από τον αλγόριθμο k-means	88
6.2	Μεταβολή του κόστους στο σύνολο επικύρωσης για τις διάφορες μεθοδολογίες με εκκίνηση κέντρων από τον αλγόριθμο fuzzy means	89

Κατάλογος Πινάκων

6.1	Benchmark σύνολα δεδομένων.	79
6.2	Τιμές παραμέτρων για τους διάφορους αλγορίθμους.	80
6.3	Αποτελέσματα των αλγορίθμων k-means και fuzzy means ως προς το κόστος στα σύνολα εκπαίδευσης, επικύρωσης και δοκιμής.	81
6.4	Αποτελέσματα των αλγορίθμων k-means και fuzzy means ως προς το υπολογιστικό κόστος για την εύρεση του βέλτιστου αριθμού κέντρων N ή του βέλτιστου διαχωρισμού s	81
6.5	Αποτελέσματα των αλγορίθμων k-means και fuzzy means ως προς το υπολογιστικό κόστος της εκπαίδευσης.	81
6.6	Αποτελέσματα του αλγορίθμου Variable Projection - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means και για διάφορες τιμές της παραμέτρου τ , με σταθερές τιμές για τα εύρη σ	82
6.7	Αποτελέσματα του αλγορίθμου Variable Projection - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means και για διάφορες τιμές της παραμέτρου τ , με σταθερές τιμές για τα εύρη σ	82
6.8	Αποτελέσματα του αλγορίθμου Variable Projection - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means και για διάφορες τιμές της παραμέτρου τ , με μεταβλητές τιμές για τα εύρη σ	83
6.9	Αποτελέσματα του αλγορίθμου Variable Projection - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means και για διάφορες τιμές της παραμέτρου τ , με μεταβλητές τιμές για τα εύρη σ	83
6.10	Αποτελέσματα του αλγορίθμου Full functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means και για διάφορες τιμές της παραμέτρου τ , με σταθερές τιμές για τα εύρη σ	84
6.11	Αποτελέσματα του αλγορίθμου Full functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means και για διάφορες τιμές της παραμέτρου τ , με σταθερές τιμές για τα εύρη σ	84
6.12	Αποτελέσματα του αλγορίθμου Full functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means και για διάφορες τιμές της παραμέτρου τ , με μεταβλητές τιμές για τα εύρη σ	85
6.13	Αποτελέσματα του αλγορίθμου Full functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means και για διάφορες τιμές της παραμέτρου τ , με μεταβλητές τιμές για τα εύρη σ	85
6.14	Αποτελέσματα των αλγορίθμων Variable Projection, Full Functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means ως προς το βέλτιστο μέσο σφάλμα στο σύνολο δοκιμής για μεταβλητά και σταθερά εύρη σ	86
6.15	Αποτελέσματα των αλγορίθμων Variable Projection, Full Functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means ως προς το βέλτιστο σφάλμα στο σύνολο δοκιμής για μεταβλητά και σταθερά εύρη σ	87
6.16	Αποτελέσματα των προτεινόμενων μεθοδολογιών ως προς το βέλτιστο σφάλμα στο σύνολο δοκιμής συγκρινόμενα με άλλες μεθοδολογίες ως προς τα κόστη στα σύνολα εκπαίδευσης, επικύρωσης και εφαρμογής.	91

6.17 Αποτελέσματα των προτεινόμενων μεθοδολογιών ως προς το βέλτιστο σφάλμα στο σύνολο δοκιμής συγκρινόμενα με άλλες μεθοδολογίες ως προς τον υπολογιστικό χρόνο.	91
---	----

Κεφάλαιο 1

Εισαγωγή

Την Υπολογιστική Νοημοσύνη συγκροτούν εργαλεία τα οποία προέρχονται από τη μελέτη των φυσικών συστημάτων και, συγκεκριμένα, των βιολογικών οργανισμών. Η μελέτη της βιολογικής εξέλιξης, της νοημοσύνης που επιδεικνύουν τα σμήνη οργανισμών, της ανθρώπινης σκέψης και των βιολογικών νευρωνικών δικτύων οδηγεί, αντίστοιχα, στην ανάπτυξη υπολογιστικών μοντέλων του Εξελικτικού Υπολογισμού, της Νοημοσύνης Σμήνους, της Ασαφούς Λογικής και των Τεχνητών Νευρωνικών Δικτύων, τα οποία μπορούν να χρησιμοποιηθούν για την επίλυση σύνθετων προβλημάτων [Engelbrecht07].

Τα Τεχνητά Νευρωνικά Δίκτυα αποτελούν ένα σημαντικό εργαλείο της Υπολογιστικής Νοημοσύνης. Οι ικανότητες του εγκεφάλου για μάθηση, απομνημόνευση και γενίκευση, σε συνδυασμό με τα χαρακτηριστικά ενός σύνθετου, μη γραμμικού, παράλληλου και κατανεμημένου συστήματος, αποτελούν σπουδαία κίνητρα για τη μοντελοποίηση του ανθρώπινου εγκεφάλου. Ωστόσο, η οποιαδήποτε προσπάθεια για την πλήρη αντιγραφή της δομής και της λειτουργίας του εγκεφάλου είναι πολύ δύσκολο έργο, με τα Τεχνητά Νευρωνικά Δίκτυα να είναι περιορισμένα σε μέγεθος και λειτουργικότητα, ενώ υπάρχουν οι βασικές αναλογίες με τα βιολογικά νευρωνικά δίκτυα [Engelbrecht07], [Βλαχάβας06].

Τα Τεχνητά Νευρωνικά Δίκτυα αποτελούνται από ένα σύνολο διατεταγμένων υπολογιστικών μονάδων επεξεργασίας σε συγκεκριμένα πρότυπα, ενώ χρησιμοποιούνται για την επίλυση προβλημάτων κατηγοριοποίησης, παλινδρόμησης ή πρόβλεψης χρονοσειρών, αλλά και σε άλλες πρακτικές εφαρμογές [Βλαχάβας06]. Οι τιμές των παραμέτρων που απαρτίζουν το νευρωνικό δίκτυο προέρχονται μετά από μια διαδικασία που είναι γνωστή ως εκπαίδευση.

Έχουν αναπτυχθεί διάφορα είδη και αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων. Μια σημαντική παραλλαγή αφορά τα νευρωνικά δίκτυα τα οποία διαχειρίζονται μια συνάρτηση βάσης ακτινικής συμμετρίας και είναι αυτά με τα οποία ασχολείται η συγκεκριμένη εργασία. Τα δίκτυα συνάρτησης βάσης ακτινικής συμμετρίας (radial basis function networks, RBFNs) χαρακτηρίζονται από την απλότητα της δομής τους και από την ταχύτητα των αλγορίθμων εκπαίδευσης, ενώ έχουν χρησιμοποιηθεί εκτενώς στη μοντελοποίηση συστημάτων και αναγνώριση προτύπων [Alexandridis03], [Alexandridis13]. Οι παράμετροι που χαρακτηρίζουν αυτά τα δίκτυα είναι τα κέντρα, ή και τα εύρη, των μονάδων RBF καθώς και τα συναπτικά βάρη σύνδεσης κρυφού επιπέδου - επιπέδου εξόδου. Για τη σωστή λειτουργία ενός δικτύου RBF είναι απαραίτητο το στάδιο της εκπαίδευσης, μέσω του οποίου γίνονται γνωστές οι τιμές των παραμέτρων.

1.1 Κίνητρο της εργασίας

Η εκπαίδευση συνίσταται στη μάθηση των παραμέτρων του δικτύου. Έχουν προταθεί διάφορες μέθοδοι για την εκπαίδευση ενός δικτύου RBF, με την κύρια διαφορά τους να έγκειται στον αριθμό των παραμέτρων που αυτές διαχειρίζονται. Ωστόσο, η εκπαίδευση αυτών των δικτύων είναι συνήθως γρηγορότερη από ό,τι είναι στις αρχιτεκτονικές άλλων δικτύων, καθώς αυτή μπορεί να πραγματοποιηθεί σε δύο στάδια [Alexandridis03], [Engelbrecht07]. Στο πρώτο στάδιο μαθαίνονται οι παράμετροι που σχετίζονται με τις κρυφές μονάδες μέσω μιας τεχνικής μη εποπτευόμενης μάθησης, ενώ στο δεύτερο στάδιο μαθαίνονται τα συναπτικά βάρη του δικτύου μέσω μιας τεχνικής εποπτευόμενης μάθησης, χωρίς

να αποκλείονται άλλες μέθοδοι εκπαίδευσης.

Η επιλογή των θέσεων των κέντρων και ο αριθμός των μονάδων RBF θεωρείται σημαντική για την απόδοση του δικτύου [Chen91]. Στις πρώτες μελέτες, [Michelli86], [Poggio90], σε κάθε δεδομένο του συνόλου εκπαίδευσης αντιστοιχούσε και ένα κέντρο, ωστόσο με τον τρόπο αυτό ο αριθμός των παραμέτρων ήταν υψηλός, με τα παραγόμενα δίκτυα να είναι υψηλής πολυπλοκότητας, κυρίως σε μεγάλα σύνολα δεδομένων [Alexandridis13]. Αν ληφθεί υπόψη ότι τα κέντρα των μονάδων πρέπει να καλύπτουν επαρκώς τα δεδομένα εκπαίδευσης, τότε, μάλλον, θα πρέπει να χρησιμοποιηθούν πιο αποτελεσματικοί τρόποι επιλογής όπως είναι, για παράδειγμα, η ομαδοποίηση του χώρου εισόδου μέσω του αλγορίθμου *k*-means [Moody89], [Darken90]. Ωστόσο, λόγω ορισμένων μειονεκτημάτων αυτού του αλγορίθμου, μεταξύ των οποίων είναι η επαναληπτική του φύση, αλλά και η ανικανότητα του για τον αυτόματο προσδιορισμό του αριθμού των κέντρων, έχουν εφαρμοστεί και άλλες μέθοδοι επιλογής όπως είναι ο αλγόριθμος *fuzzy means*, ο οποίος υπολογίζει τον αριθμό των κέντρων και τις θέσεις βασιζόμενος στον ασαφή διαχωρισμό του χώρου εισόδου [Sarimveis02], [Alexandridis03], [Alexandridis13]. Ο αλγόριθμος αυτός έχει βρει πολλές επιτυχημένες εφαρμογές [Alexandridis13(2)], [Alexandridis13(3)], [Alexandridis14], [Alexandridis14(2)], [Alexandridis16], [Alexandridis17], [Stogiannos18].

Επίσης, μπορούν να χρησιμοποιηθούν μέθοδοι στις οποίες το κρυφό επίπεδο κατασκευάζεται με τρόπο επαυξητικό, όπως στο [Chen91], όπου μέσω ορθογώνιων ελαχίστων τετραγώνων επιλέγονται τα κέντρα των μονάδων μέχρι να κατασκευαστεί ένα επαρκές δίκτυο.

Ακόμα, δεν αποκλείονται οι μέθοδοι εκπαίδευσης ενός βήματος στις οποίες όλες οι παράμετροι μαθαίνονται σε ένα στάδιο. Στο [Moody89(2)], [Poggio89] γίνεται χρήση της μεθόδου απότομης κατάβασης για την προσαρμογή των παραμέτρων του δικτύου, ωστόσο με αργούς χρόνους σύγκλισης [Wettschereck92].

Για την εκπαίδευση ενός δικτύου RBF μπορούν να χρησιμοποιηθούν και συνδυασμοί των μεθόδων της Υπολογιστικής Νοημοσύνης. Συγκεκριμένα στο [Alexandridis13] γίνεται εκπαίδευση ενός δικτύου RBF με χρήση της μεθόδου βελτιστοποίησης με σμήνη σωματιδίων για τη βελτιστοποίηση του ασαφούς διαχωρισμού του χώρου εισόδου, ενώ στο [Sarimveis04] γίνεται χρήση γενετικών αλγορίθμων για τη βέλτιστη επιλογή του αριθμού και των κέντρων των μονάδων RBF.

Τα εύρη των συναρτήσεων βάσης μπορούν να υπολογιστούν μέσω μιας ευρετικής μεθόδου που αφορά τους *p*-κοντινότερους γείτονες, καθώς οι τιμές τους πρέπει να επιλεγθούν έτσι, ώστε να υπάρχει αλληλοκάλυψη με τις *p*-κοντινότερες μονάδες RBF [Leonard91].

Λόγω της γραμμικής σύνδεσης του κρυφού επιπέδου και του επιπέδου εξόδου σε ένα δίκτυο RBF, οι περισσότερες μέθοδοι περιλαμβάνουν τον υπολογισμό των συναπτικών βαρών μέσω της επίλυσης ενός απλού γραμμικού προβλήματος ελαχίστων τετραγώνων [Alexandridis13].

Περισσότερες πληροφορίες για τις μεθόδους εκπαίδευσης ενός δικτύου RBF μπορούν να βρεθούν στα [Alexandridis03], [Alexandridis13], [Engelbrecht07].

1.2 Συμβολή της εργασίας

Η προτεινόμενη μέθοδος εκπαίδευσης για ένα δίκτυο RBF βασίζεται στην απλή παρατήρηση της μορφοποίησης του προβλήματος της εκπαίδευσης, το οποίο ορίζεται ως ένα άθροισμα τετραγώνων μη γραμμικών συναρτήσεων, ενώ οι γραμμικές (βάρη δικτύου) και οι μη γραμμικές παράμετροι (κέντρα δικτύου) είναι διαχωρίσιμες.

Ο διαχωρισμός αυτός οδηγεί στην ιδέα της ελαχιστοποίησης ενός μειωμένου συναρτησιακού, του λεγόμενου *Variable Projection*, το οποίο είναι συνάρτηση μόνο των μη γραμμικών παραμέτρων. Ο αριθμός και οι αρχικές θέσεις των μονάδων RBF προέρχονται από τους αλγορίθμους *k*-means και *fuzzy means*, ενώ μέσω της ελαχιστοποίησης του μειωμένου συναρτησιακού ενημερώνονται οι θέσεις των κέντρων μέσω μιας μεθόδου βελτιστοποίησης. Η μέθοδος βελτιστοποίησης που χρησιμοποιείται αφορά μια τροποποιημένη εκδοχή της μεθόδου Gauss-Newton, τη Levenberg-Marquardt. Σε κάθε ενημέρωση των κέντρων, τα βάρη του δικτύου υπολογίζονται μέσω ενός απλού γραμμικού προβλήματος ελαχίστων τετραγώνων. Έτσι, σε κάθε επανάληψη του αλγορίθμου υπάρχει εγγύηση για ολικό βέλτιστο στα βάρη για τα συγκεκριμένα κέντρα που έχουν επιλεγεί, με το τελικό αποτέλεσμα να εξαρτάται από τη θέση των

κέντρων και, τελικά, από το τοπικό ελάχιστο στο οποίο αυτά καταλήγουν. Οι αρχικές τιμές των ευρών των συναρτήσεων βάσης προέρχονται από τη μέθοδο των p -κοντινότερων γειτόνων.

Στην περίπτωση που δεν πραγματοποιείται κάποια απαλοιφή των γραμμικών παραμέτρων, τότε το πρόβλημα προς επίλυση αφορά την ελαχιστοποίηση του πλήρους συναρτησιακού στο οποίο τα κέντρα, τα βάρη ή, ακόμα, και τα εύρη του δικτύου, μεταβάλλονται καταλλήλως μέσω του αλγορίθμου βελτιστοποίησης Levenberg-Marquardt. Με τη μέθοδο αυτή, δεν υπάρχει εγγύηση για ολικό βέλτιστο στα βάρη, οι τιμές τους μεταβάλλονται ταυτοχρόνως με τα κέντρα, ενώ όλα μαζί κινούνται προς κάποιο τοπικό ελάχιστο. Στην περίπτωση αυτή αναμένεται να χρειάζονται περισσότερες επαναλήψεις για σύγκλιση.

1.3 Δομή εργασίας

Η εργασία αναπτύσσεται σε 7 κύρια κεφάλαια:

Στο επόμενο κεφάλαιο γίνεται μια εισαγωγή στις βασικές μεθόδους της Υπολογιστικής Νοημοσύνης όπως είναι ο Εξελικτικός Υπολογισμός, η Νοημοσύνη Σμήνους και τα Τεχνητά Νευρωνικά, με όλο το ενδιαφέρον να επικεντρώνεται στα τελευταία. Παρουσιάζονται οι αναλογίες με τους βιολογικούς νευρώνες και ορίζεται το μοντέλο του τεχνητού, ενώ γίνεται μια αναφορά στις αρχιτεκτονικές των δικτύων καθώς επίσης και στα είδη των μεθόδων μηχανικής μάθησης.

Στο κεφάλαιο 3 ακολουθεί μια περιγραφή της δομής των δικτύων τύπου RBF, των ειδών των συναρτήσεων βάσης καθώς και των πλεονεκτημάτων αυτών των δικτύων. Στο ίδιο κεφάλαιο μορφοποιείται το πρόβλημα της εκπαίδευσης ενός δικτύου RBF ενώ γίνεται και μια περιγραφή των μεθόδων εκπαίδευσης που χρησιμοποιούνται για την εκμάθηση των παραμέτρων. Στο ίδιο κεφάλαιο αναπτύσσονται οι αλγόριθμοι k -means και fuzzy means, οι οποίοι αποτελούν τη βάση για την προτεινόμενη μέθοδο εκπαίδευσης.

Στο κεφάλαιο 4 γίνεται μια σύντομη αναφορά στους διάφορους αλγόριθμους βελτιστοποίησης που μπορούν να χρησιμοποιηθούν για τη βελτίωση των τιμών των παραμέτρων του δικτύου ενώ αναπτύσσεται η προτεινόμενη μέθοδος εκπαίδευσης, η οποία αποτελεί ένα συνδυασμό των μεθόδων Levenberg-Marquardt και Variable Projection.

Στο κεφάλαιο 5 εφαρμόζεται η προτεινόμενη μέθοδος εκπαίδευσης καθώς και παραλλαγές αυτής, οι οποίες αφορούν την περίπτωση σταθερών ευρών, ή μεταβλητών βαρών, βάση των οποίων αποτελεί ο αλγόριθμος Levenberg-Marquardt.

Στο κεφάλαιο 6 παρουσιάζονται τα αποτελέσματα των μεθόδων σε μια σειρά συνόλων δεδομένων της βιβλιογραφίας, ενώ ακολουθεί και μια σύγκριση με άλλες μεθοδολογίες της βιβλιογραφίας που παρουσιάζονται εκεί.

Στο τέλος, παρουσιάζονται τα συμπεράσματα για τις εν λόγω μεθόδους εκπαίδευσης ενώ δίνεται το έναυσμα για περαιτέρω μελλοντική εργασία.

Κεφάλαιο 2

Υπολογιστική Νοημοσύνη και Νευρωνικά Δίκτυα

2.1 Εισαγωγή στην Τεχνητή και Υπολογιστική Νοημοσύνη

Οποιαδήποτε τεχνική ωθεί τον υπολογιστή να μιμηθεί την ανθρώπινη σκέψη ή συμπεριφορά καλείται Τεχνητή Νοημοσύνη (Artificial Intelligence). Είναι μέρος της επιστήμης των υπολογιστών, μιας και αφορά φαινόμενα που σχετίζονται με αυτούς. Ακόμα, είναι μέρος της ψυχολογίας και της γνωσιολογικής επιστήμης, αφού οι εργασίες που εκτελούνται στους υπολογιστές απαιτούν κάποιο είδος ευφυΐας ή σκέψης, όπως θα συνέβαινε αν τις εκτελούσαν οι άνθρωποι [Simon95].

Ένας από τους πρώτους ορισμούς για την Τεχνητή Νοημοσύνη, διατυπωμένος από τους Barr και Feigenbaum είναι ο εξής [Barr81]:

«Η Τεχνητή Νοημοσύνη είναι ο τομέας της επιστήμης των υπολογιστών που ασχολείται με το σχεδιασμό ευφυών υπολογιστικών συστημάτων, δηλαδή συστημάτων που παρουσιάζουν χαρακτηριστικά που σχετίζονται με την ανθρώπινη συμπεριφορά»

Η κατανόηση της ανθρώπινης σκέψης και συμπεριφοράς αποτελεί προσπάθεια πολλών χρόνων [Russell03]. Το δύσκολο αυτό ερώτημα του τι είναι η νοημοσύνη προσπάθησαν να το εξηγήσουν οι φιλόσοφοι της Αρχαίας Ελλάδας, όπως ο Αριστοτέλης ή ο Ηράκλειτος, περιγράφοντας διάφορες λειτουργίες που σχετίζονται με αυτή: μηχανισμός της μάθησης, του συλλογισμού αλλά και άλλες [Βλαχάβας06]. Σύμφωνα με το λεξικό του Κέιμπριτζ¹ η νοημοσύνη ορίζεται ως:

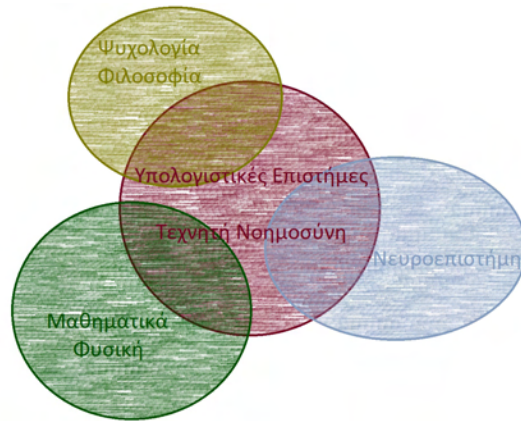
«η ικανότητα για μάθηση, κατανόηση και κρίση ή αιτιολογημένη έκφραση γνώμης»

δηλαδή περιλαμβάνει ικανότητες οι οποίες συνδέονται με την επίλυση προβλημάτων, τη μάθηση από εμπειρία και άλλα [Βλαχάβας06].

Η Τεχνητή Νοημοσύνη ξεκίνησε από τα μέσα του περασμένου αιώνα, με την ονομασία της να οφείλεται στον J. McCarthy το 1956, ενώ η έρευνα στη νέα αυτή περιοχή φαίνεται να έχει ξεκινήσει νωρίτερα [Russell03], [Βλαχάβας06]. Έχοντας ως πειραματικές διατάξεις «νοήμονες» υπολογιστές, χρησιμοποιήθηκε και ως μέσο για την έρευνα της ανθρώπινης νοημοσύνης και συμπεριφοράς [Buchanan05].

Τα αρχικά βήματα της Τεχνητής Νοημοσύνης ξεκινούν από σημαντικά ερωτήματα τα οποία προέρχονται από τα πεδία διάφορων επιστημών όπως της φιλοσοφίας, των μαθηματικών, της ψυχολογίας, της μηχανικής των υπολογιστών και της νευροεπιστήμης [Russell03]. Στο σχήμα 2.1 ορίζεται το πεδίο της Τεχνητής Νοημοσύνης ως μέρος αυτών των επιστημών.

¹ <https://dictionary.cambridge.org/dictionary/english/intelligence>



Σχήμα 2.1: Τεχνητή Νοημοσύνη και άλλες επιστήμες.

Μερικά από αυτά τα ερωτήματα είναι τα εξής [Russell03]:

Φιλοσοφία :

«Από πού έρχεται η γνώση και πώς η γνώση οδηγεί στην πράξη;»

Μαθηματικά :

«Τι μπορεί να υπολογιστεί αληθιά και πώς ορθολογικοποιούμε με αβέβαιες πληροφορίες;»

Ψυχολογία :

«Πώς οι άνθρωποι σκέφτονται και πράττουν;»

Τεχνολογία Υπολογιστών :

«Πώς μπορεί να κατασκευαστεί ένας αποτελεσματικός υπολογιστής;»

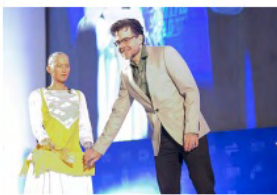
Νευροεπιστήμες :

«Πώς ο εγκέφαλος επεξεργάζεται τις πληροφορίες;»

Συμβολές στην εξέλιξη της Τεχνητής Νοημοσύνης φαίνεται να υπάρχουν και από τη φαντασία και τα μυθιστορήματα, με τις πρώτες αναφορές να προέρχονται από την ελληνική μυθολογία [Βλαχάβας06], [Russell03], [Buchanan05]. Οι συγγραφείς μυθιστορημάτων αποτέλεσαν πηγή έμπνευσης για πολλούς ερευνητές της Τεχνητής Νοημοσύνης, όπως ο Ιούλιος Βέρν (1828–1905), στον οποίο οφείλεται η ονομα-



(α) Ιούλιος Βέρν και το ρομπότ ονομασμένο προς τιμήν του.



(β) Φρανκ Μπάουμ με το μάγο του Οζ.

Σχήμα 2.2: Συγγραφείς μυθιστορημάτων ως πηγή έμπνευσης για την Τεχνητή Νοημοσύνη.

σία του ρομπότι του ερευνητή Ντέιβιντ Χάνσον (σχήμα 2.2α), το οποίο μιλάει και συμπεριφέρεται όπως ένας άνθρωπος, βάσει των ερεθισμάτων που λαμβάνει αλλά και άλλοι μεταξύ των οποίων και ο Φρανκ Μπάουμ (1856-1919) με το βιβλίο του «Ο μάγος του Οζ» (σχήμα 2.2β), όπου το 1907 περιέγραψε τον άνθρωπο μηχανή ως

« Έντονης απόκρισης, δημιουργός σκέψης, ομιλία τέλειου μηχανικού ανθρώπου... Σκέφτεται, μιλάει, ενεργεί, κάνει τα πάντα εκτός από το να ζει»

Μέσα από την πιθανότητα ύπαρξης τέτοιων «έξυπνων» μηχανών σίγουρα κορυφώθηκε η φαντασία αλλά προετοιμάστηκε και η έναρξη της προσπάθειας για την κατανόηση της ανθρώπινης νοημοσύνης [Buchanan05].

Πριν, τα μόνα συστήματα που παρουσίαζαν κάποια νοημοσύνη ήταν οι έμβιοι οργανισμοί και κυρίως οι άνθρωποι. Τώρα, αυξήθηκε η οικογένεια αυτή με την εισαγωγή και «έξυπνων» μηχανών [Simon95].

Οι ορισμοί για την Τεχνητή Νοημοσύνη ποικίλουν. Σύμφωνα με τους Russell και Norvig [Russell03] μπορούν να διακριθούν σε δύο κατηγορίες αναλόγως της προσέγγισης επί του συγκεκριμένου θέματος: η μια κατηγορία ασχολείται με τη σκέψη και η άλλη με τη συμπεριφορά. Επιπλέον διαχωρισμός αυτών γίνεται σε ακόμα δύο κατηγορίες: σε αυτή που μέτρο είναι η ανθρώπινη απόδοση, όπου χρησιμοποιούνται εμπειρικές παρατηρήσεις και υποθέσεις της ανθρώπινης συμπεριφοράς και σε αυτή που μέτρο αποτελεί κάποια ιδανική απόδοση, όπου χρησιμοποιούνται τα μαθηματικά και η τεχνολογία.

Αναλόγως, λοιπόν, της προσέγγισης του πεδίου της Τεχνητής Νοημοσύνης κάποιοι ορισμοί μπορεί να είναι οι εξής:

Συστήματα που σκέφτονται όπως οι άνθρωποι:

«Η αυτοματοποίηση δραστηριοτήτων που σχετίζονται με την ανθρώπινη σκέψη και δραστηριότητες όπως είναι η λήψη αποφάσεων, η επίλυση προβλημάτων, η μάθηση...»
(Bellman, 1978)

Συστήματα που σκέφτονται λογικά:

«Η μελέτη των υπολογισμών που καθιστούν εφικτή την αντίληψη, τη λογική σκέψη και την αντίδραση»
(Winston, 1992)

Συστήματα που συμπεριφέρονται όπως οι άνθρωποι:

«Η μελέτη του πώς να κάνουμε τους υπολογιστές να κάνουν πράγματα στα οποία, αυτήν τη στιγμή, οι άνθρωποι είναι καλύτεροι»
(Rich and Knight, 1991)

Συστήματα που δρουν λογικά:

«Η τεχνητή νοημοσύνη... αφορά ευφυή συμπεριφορά τεχνουργημάτων»
(Nilsson, 1998)

Η διαφορά μεταξύ της ανθρώπινης και λογικής συμπεριφοράς βασίζεται στο γεγονός ότι οι άνθρωποι κάνουν λάθη, δεν είναι τέλειοι [Russell03].

Η Τεχνητή Νοημοσύνη μπορεί να προσεγγιστεί, ακόμα, με δύο τρόπους: την Κλασική ή Συμβολική Τεχνητή Νοημοσύνη και την Υπολογιστική Νοημοσύνη ή Μη Συμβολική Τεχνητή Νοημοσύνη. Η πρώτη εστιάζει στην κατανόηση και την εξομίωση της ανθρώπινης νοημοσύνης την οποία και μοντελοποιεί με αλγορίθμους με χρήση συμβόλων, εν αντιθέσει με τη δεύτερη, η οποία, με αριθμούς πια και όχι με σύμβολα, προσπαθεί να μιμηθεί βιολογικά συστήματα [Βλαχάβας06]. Η παρούσα εργασία ασχολείται με τη δεύτερη προσέγγιση, αυτή της Υπολογιστικής Νοημοσύνης.

Υπάρχει μια ασάφεια σχετικά με τη σχέση που υπάρχει ανάμεσα στους όρους Τεχνητή Νοημοσύνη και Υπολογιστική Νοημοσύνη. Σε κάποια εγχειρίδια, όπως επισημαίνει ο Duch [Duch03], δίνεται η εντύπωση ότι αυτοί οι δύο όροι ταυτίζονται ως έννοιες. Ο όρος Υπολογιστική Νοημοσύνη επινοήθηκε από τους Cercone και MacCalla [Cercone194] το 1985 και ορίστηκε έτσι, ώστε να δωθεί έμφαση στα διαφοροποιημένα υπολογιστικά μοντέλα που αυτή χρησιμοποιεί [Back97].

Ο Duch συγκέντρωσε κάποιους ορισμούς για την Υπολογιστική Νοημοσύνη από διάφορα βιβλία :

Computational Intelligence: A Logical Approach

«Η Υπολογιστική Νοημοσύνη είναι η μελέτη του σχεδιασμού έξυπνων μηχανών [...] Ο κεντρικός επιστημονικός στόχος της Υπολογιστικής Νοημοσύνης είναι η κατανόηση των αρχών, οι οποίες καθιστούν μια νοητική συμπεριφορά δυνατή, σε φυσικά ή τεχνητά συστήματα»

Computational Intelligence for Decision Support

«Η Υπολογιστική Νοημοσύνη είναι το πεδίο μελέτης κατασκευής έξυπνων πρακτόρων»

Computational Intelligence: An Introduction

«Η μελέτη υπολογιστικών συστημάτων που είναι εξοπλισμένα με προσαρμοστικούς μηχανισμούς (ικανότητα μάθησης - μάθηση μηχανών) με νοητική συμπεριφορά (ανακάλυψη, γενίκευση) σε σύνθετα και μη στάσιμα περιβάλλοντα (αντιμετώπιση νέων καταστάσεων - εύρωστα συστήματα). Η Υπολογιστική Νοημοσύνη συνδυάζει τεχνητά νευρωνικά δίκτυα, εξελικτικό υπολογισμό, νοημοσύνη σμήνους και ασαφή συστήματα»

Παρόλο που στόχος και των δύο αφορά τη νοημοσύνη σε μηχανές, ώστε τελικά αυτές να χρησιμοποιούνται για την επίλυση προβλημάτων που λύνουν και οι άνθρωποι, υπάρχει διαφορά στα χρησιμοποιούμενα από αυτές εργαλεία και συγκεκριμένα στη αριθμητική αναπαράσταση της γνώσης, η οποία διαφοροποιείται στην Κλασική Συμβολική Τεχνητή Νοημοσύνη [Bezdek94]. Ο Bezdek χαρακτηρίζει την Υπολογιστική Νοημοσύνη ως υποσύνολο της Τεχνητής Νοημοσύνης στην οποία ικανοποιούνται ορισμένα χαρακτηριστικά [Back97].

Αν τα προβλήματα μπορούν να χωριστούν σε αυτά που απαιτούν χαμηλού επιπέδου γνώσεις, όπως είναι η αναγνώριση αντικειμένων, η εύρεση δομών σε σύνολα δεδομένων, και σε αυτά που απαιτούν υψηλότερου επιπέδου γνώσεις, όπως είναι η συστηματική σκέψη και ο ορθός λογισμός, τότε τα πρώτα θα αφορούσαν την Υπολογιστική Νοημοσύνη, η οποία χρησιμοποιεί προσεγγίσεις βιολογικών συστημάτων αλλά και στατιστικά μοντέλα για τη λύση τέτοιων προβλημάτων [Duch03].

Η Υπολογιστική Νοημοσύνη, ως κλάδος της επιστήμης των υπολογιστών, ασχολείται με τη μελέτη προβλημάτων για τα οποία δεν υπάρχει κάποιος αποτελεσματικός αλγόριθμος, ή, ακόμα, όταν είναι αδύνατη η μορφοποίησή του. Εστιάζει σε μεθόδους εμπνευσμένες από τη φύση, αφού οι βιολογικοί οργανισμοί έρχονται σε επαφή με τέτοια προβλήματα καθημερινά: προσαρμογή του εγκεφάλου σε νέες καταστάσεις, αναγνώριση αντικειμένων και άλλα [Duch03].

Μακροπρόθεσμος στόχος της Υπολογιστικής Επιστήμης είναι η δημιουργία συστημάτων, τα οποία θα μπορούν να ανταγωνιστούν τους ανθρώπους σε πιο σύνθετα προβλήματα, πέρα από τα προαναφερθέντα. Αν και προσπάθειά της είναι η προσομοίωση φυσικών συστημάτων, ακόμα είναι μακριά από αυτά [Duch03].

Ο άνθρωπος έχει μάθει πολλά παρατηρώντας τη φύση και για τους ειδικούς της Υπολογιστικής Νοημοσύνης οι βιολογικές επιρροές είναι πολύ σημαντικές. Εμπνευσμένος από τις γνώσεις αυτές προσπαθεί να μοντελοποιήσει τα διάφορα φυσικά και βιολογικά συστήματα, τα οποία αποτελούν τη βάση για την αρχή της δημιουργίας «έξυπνων» μηχανών, όπου, τελικά, με τη χρήση των νέων αυτών «έξυπνων» αλγοριθμικών μοντέλων θα είναι σε θέση να λύνει σύνθετα προβλήματα [Duch03].

2.2 Βασικές Μέθοδοι της Υπολογιστικής Επιστήμης

Αυτοί οι «έξυπνοι» αλγόριθμοι περιλαμβάνουν (σχήμα 2.3) [Engelbrecht07]:

1. Εξελικτικό Υπολογισμό (Evolutionary Computation)
2. Νοημοσύνη Σμήνους (Swarm Intelligence)
3. Ασαφή Συστήματα (Fuzzy Systems)
4. Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)



Σχήμα 2.3: Βασικές μέθοδοι της Υπολογιστικής Νοημοσύνης.

Κάθε μια από τις παραπάνω μεθόδους, όπως άλλωστε δηλώνει και η ονομασία τους, είναι εμπνευσμένη από βιολογικά συστήματα.

Συγκεκριμένα:

- τα Τεχνητά Νευρωνικά Δίκτυα αφορούν τη μοντελοποίηση βιολογικών νευρωνικών συστημάτων
- ο Εξελικτικός Υπολογισμός ασχολείται με τη μοντελοποίηση της φυσικής εξέλιξης
- η Νοημοσύνη Σμήνους μοντελοποιεί την κοινωνική συμπεριφορά των οργανισμών που ζουν σε σμήνη και, τέλος,
- τα Ασαφή Συστήματα ξεκίνησαν από τη μελέτη του τρόπου της αλληλεπίδρασης των οργανισμών με το περιβάλλον τους

Στις επόμενες ενότητες ακολουθεί μια σύντομη περιγραφή των παραπάνω μεθόδων, με την περισσότερη έμφαση να δίνεται στα Τεχνητά Νευρωνικά Δίκτυα.

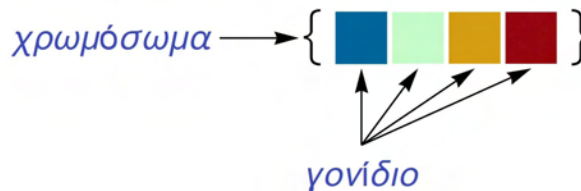
2.2.1 Εξελικτικός Υπολογισμός

Μια από τις μεθόδους της Υπολογιστικής Νοημοσύνης είναι αυτή του εξελικτικού υπολογισμού. Αλγόριθμοι εμπνευσμένοι από τη φυσική εξέλιξη των ειδών χρησιμοποιούνται για τη λύση διάφορων προβλημάτων όπως, για παράδειγμα, σύνθετα προβλήματα βελτιστοποίησης, υποσχόμενοι μια όσο το δυνατόν καλύτερη λύση σε λογικό χρονικό διάστημα [Back97], [Βλαχάβας06].

Ο εξελικτικός υπολογισμός φαίνεται να ξεκίνησε από τα τέλη της δεκαετίας του 50', όμως, παρέμεινε άγνωστο το πεδίο αυτό στην ευρύτερη επιστημονική κοινότητα για αρκετό καιρό, με την ουσιαστική πρόοδο να έχει ξεκινήσει μερικές δεκαετίες αργότερα [Back97].

Οι γενετικοί αλγόριθμοι είναι μια από τις κατηγορίες του εξελικτικού υπολογισμού. Αυτοί οι αλγόριθμοι μοντελοποιούν τη γενετική εξέλιξη, οπότε θα υπάρχουν και οι αντίστοιχες αναλογίες. Στη φύση ισχύει ο κανόνας της φυσικής επιλογής: ο ισχυρότερος επιβιώνει, ο αδύναμος πεθαίνει. Ο προηγούμενος κανόνας αποτελεί τον κυριότερο μηχανισμό της εξέλιξης. Αυτός, λοιπόν, που επιβιώνει πολλαπλασιάζεται μέσω της αναπαραγωγής δίνοντας απογόνους [Βλαχάβας06]. Τα χαρακτηριστικά του απογόνου προκύπτουν από τους προγόνους του και, επομένως, αυτός ο απόγονος που κληρονομεί καλύτερα χαρακτηριστικά, δηλαδή ο ισχυρότερος απόγονος, έχει τις μεγαλύτερες πιθανότητες για επιβίωση και για αναπαραγωγή [Engelbrecht07]. Οι επιρροές που δέχεται κάθε άτομο ενός πληθυσμού προέρχονται τόσο από άλλα άτομα του ίδιου ή διαφορετικού πληθυσμού αλλά και από το ίδιο του το περιβάλλον και μεταβολές αυτού, με κάποιες από αυτές να αφορούν τον ανταγωνισμό για το φαγητό, τους διάφορους εχθρούς, το κλίμα, τη δυσκολία εύρεσης τροφής και άλλα [Back97].

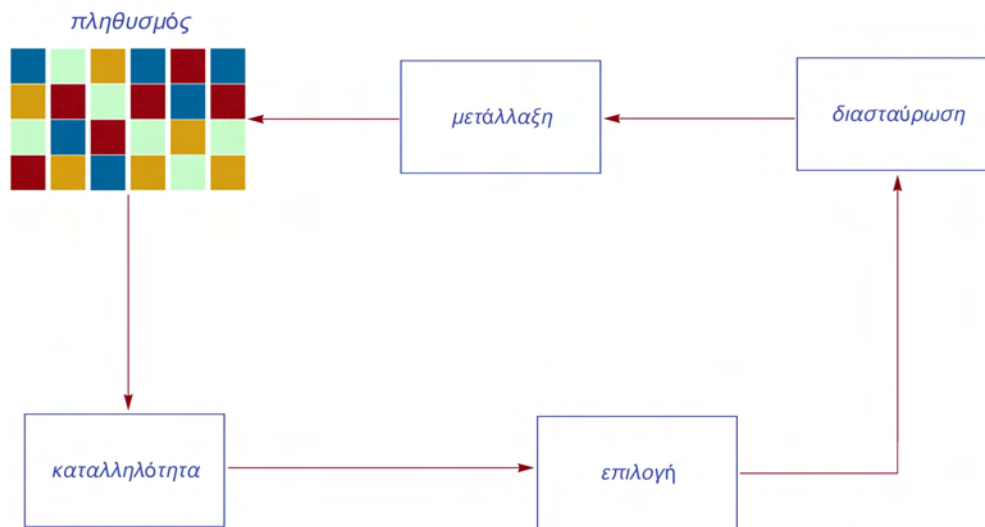
Στον γενετικό αλγόριθμο χρησιμοποιείται ένας πληθυσμός ατόμων, με το κάθε άτομο στο πληθυσμό αυτό να χαρακτηρίζεται ως χρωμόσωμα (σχήμα 2.4). Ο πληθυσμός αυτός αποτελείται από τις υποψήφια λύσεις του προβλήματος. Κάθε χαρακτηριστικό των χρωμοσωμάτων αναφέρεται ως γονίδιο, με όλες τις πιθανές καταστάσεις του γονιδίου να αποκαλούνται αλληλόμορφα [Engelbrecht07].



Σχήμα 2.4: Χρωμοσώματα και γονίδια.

Τα άτομα του πληθυσμού ανταγωνίζονται και ζευγαρώνουν για την αναπαραγωγή των απογόνων, με τα τελευταία να έχουν στοιχεία των προγόνων τους. Τα άτομα, δηλαδή τα χρωμοσώματα, που θα ζευγαρώσουν θα είναι τα πλέον καταλληλότερα. Ο νέος πληθυσμός που θα προκύψει αποτελείται από άτομα με καλύτερα, συνήθως, χαρακτηριστικά. Πέρα από τη διαδικασία της αναπαραγωγής που οδηγεί σε απογόνους, μια ακόμα λειτουργία που λαμβάνει χώρα σε ένα γενετικό αλγόριθμο είναι η μετάλλαξη. Αυτή προκαλεί βελτιώσεις και λάθη που αποτελούν σημαντικό παράγοντα για την εξέλιξη της ζωής. Μέσα από τη μετάλλαξη αλλάζει η δομή των χρωμοσωμάτων, αλλάζουν δηλαδή τα χαρακτηριστικά των ατόμων, οδηγώντας σε νέες λύσεις [Engelbrecht07], [Καβαλλιεράτου], [Βλαχάβας06].

Η εξέλιξη που λαμβάνει χώρα σε ένα γενετικό αλγόριθμο παρουσιάζεται επιγραμματικά στο σχήμα 2.5:



Σχήμα 2.5: Διαδικασία εξέλιξης σε έναν γενετικό αλγόριθμο.

Οι λύσεις του πληθυσμού βαθμολογούνται και ο αριθμός αυτός δείχνει το βαθμό καταλληλότητας, ώστε μια λύση να χαρακτηρίζεται ως αποδεκτή. Για το σκοπό αυτό χρησιμοποιείται μια συνάρτηση, η συνάρτηση καταλληλότητας, η οποία αξιολογεί τις λύσεις αυτές. Όσο μεγαλύτερη είναι η τιμή που αντιστοιχεί σε κάποια λύση, τόσο πιο κατάλληλη είναι, ενώ στόχος είναι η μεγιστοποίηση της συνάρτησης καταλληλότητας. Στη συνέχεια, επιλέγονται τα καταλληλότερα χρωμοσώματα τα οποία θα υπόκεινται σε μια ακολουθία διασταυρώσεων και μεταλλάξεων για τη δημιουργία του νέου πληθυσμού. Επιπλέον, αξίζει να σημειωθεί ότι η αναζήτηση στο χώρο των υποψηφίων λύσεων είναι παράλληλη, με την έννοια ότι κάθε άτομο, κάθε πιθανή λύση, αναζητεί ξεχωριστά από τις άλλες [Βλαχάβας06], [Fonseca95].

Μια άλλη κατηγορία αυτής της μεθόδου, ειδική περίπτωση των γενετικών αλγορίθμων, είναι ο γενετικός προγραμματισμός, με τα άτομα εδώ να είναι ολόκληρα προγράμματα. Ο Friedber το 1958 επιχείρησε την αυτόματη εξαγωγή σύνθετων προγραμμάτων στη Fortran με το συνδυασμό μικρότερων, όμως, τις περισσότερες φορές ήταν μη εκτελέσιμα [Βλαχάβας06].

Οι εξελικτικοί αλγόριθμοι αποτελούν σημαντικό εργαλείο για την επίλυση προβλημάτων που αφορούν την ταξινόμηση, την ομαδοποίηση και την εξόρυξη δεδομένων [Engelbrecht07], [Βλαχάβας06], ενώ σημαντικός είναι ο ρόλος τους στη διαχείριση μιας πολυαντικειμενικής συνάρτησης [Back97], [Fonseca95]. Χαρακτηρίζονται ως εύρωστα συστήματα με ευελιξία και προσαρμοστικότητα, όμως, παρουσιάζουν κάποια μειονεκτήματα όπως είναι ο υψηλός υπολογιστικός χρόνος και η ρύθμιση πολλών παραμέτρων [Back97], [Kotti11], [Engelbrecht07].

2.2.2 Νοημοσύνη Σμήνους

Μια άλλη βασική μέθοδος της Υπολογιστικής Νοημοσύνης αφορά τη νοημοσύνη σμήνους. Πρόκειται για μια μετεωρετική μέθοδο, η οποία βασίζεται στη νοημοσύνη σμήνων, ομάδων δηλαδή που παρουσιάζουν μια «έξυπνη» κοινωνική συμπεριφορά [Engelbrecht07]. Παρουσιάζει κάποια πλεονεκτήματα έναντι της προηγούμενης μεθόδου, λόγω της ευελιξίας του χειρισμού από πρόβλημα σε πρόβλημα, μεταξύ των οποίων είναι η ρύθμιση λιγότερων παραμέτρων, οι υπολογιστικές απαιτήσεις και η γρηγορότερη σύγκλιση [Kotti11], [Abraham06], [You08].

Όπως υποδηλώνει η ονομασία της μεθόδου, η οποία οφείλεται στους G. Beni, J.Wang το 1989 [Beni89] σε εφαρμογές στα πλαίσια ρομποτικών συστημάτων, είναι εμπνευσμένη από τη μελέτη των οργανισμών που ζουν σε ομάδες ή αποικίες. Οι οργανισμοί αυτοί συνθέτουν έναν πληθυσμό και η κοινωνική συμπεριφορά των μεμονωμένων αυτών ατόμων επιτελεί την εκπλήρωση ενός στόχου προς το

κοινωνικό συμφέρον του πληθυσμού. Παραδείγματα αποτελούν η αναζήτηση του βέλτιστου μονοπατιού των μυρμηγκιών από τη φωλιά μέχρι κάποια πηγή φαγητού, ή η αποδημία πουλιών σε ένα μακρινό μέρος, αλλά και η δημιουργία φωλιών από τις σφήκες. Σε όλα τα παραπάνω παραδείγματα, χωρίς κανένα από τα μεμονωμένα άτομα να έχουν κάποιο ηγετικό ρόλο, επιτυγχάνεται ένας κοινός στόχος [Garg09], [Blum08].

Οι ερευνητές, εμπνευσμένοι από τέτοιες «έξυπνες» συμπεριφορές σμήνων, έδωσαν το έναυσμα για το σχεδιασμό αλγορίθμων που να τις μοντελοποιούν. Οι κυριότεροι εξ αυτών αφορούν τη Βελτιστοποίηση Σμήνους Σωματιδίων (Particle Swarm Optimization) και τη Βελτιστοποίηση με Αποικίες Μυρμηγκιών (Ant Colony Optimization) [Engelbrecht07].

2.2.2.1 Βελτιστοποίηση Σμήνους Σωματιδίων

Η βελτιστοποίηση σμήνους σωματιδίων είναι μια στοχαστική μέθοδος βελτιστοποίησης βασισμένη στην κοινωνική συμπεριφορά και στη νοημοσύνη των οργανισμών για την αναζήτηση λύσεων για κάποιο κοινό σκοπό [Engelbrecht07], [Garg09], [Blum08]. Η μέθοδος αυτή ορίστηκε από τους Kennedy και Eberhart το 1995 [Kennedy95]. Από την έναρξή της απέκτησε δημοσιότητα ως μια εύρωστη και αποτελεσματική μέθοδος για την επίλυση προβλημάτων βελτιστοποίησης [Blum08].

Αυτή η μέθοδος παρουσιάζει αρκετές ομοιότητες με τον εξελικτικό υπολογισμό, παραδείγματος χάριν με τους γενετικούς αλγόριθμους [Kott11]. Και εδώ, το σύστημα συνίσταται από ένα πληθυσμό του οποίου τα άτομα θα αναφέρονται ως σωματίδια. Τα σωματίδια αυτά αναπαριστούν τις πιθανές λύσεις του προβλήματος και κινούνται στο χώρο αναζήτησης ψάχνουν για κάποια καλή λύση, ακόμα και το βέλτιστο, αν αυτό υπάρχει. Δεν υπάρχουν οι λειτουργίες της μετάλλαξης και της διασταύρωσης για την εισαγωγή νέων λύσεων, αντιθέτως το κάθε σωματίδιο θα κινείται στον εν λόγω χώρο σύμφωνα με τις πληροφορίες που έχει εκείνη τη στιγμή [Kott11].

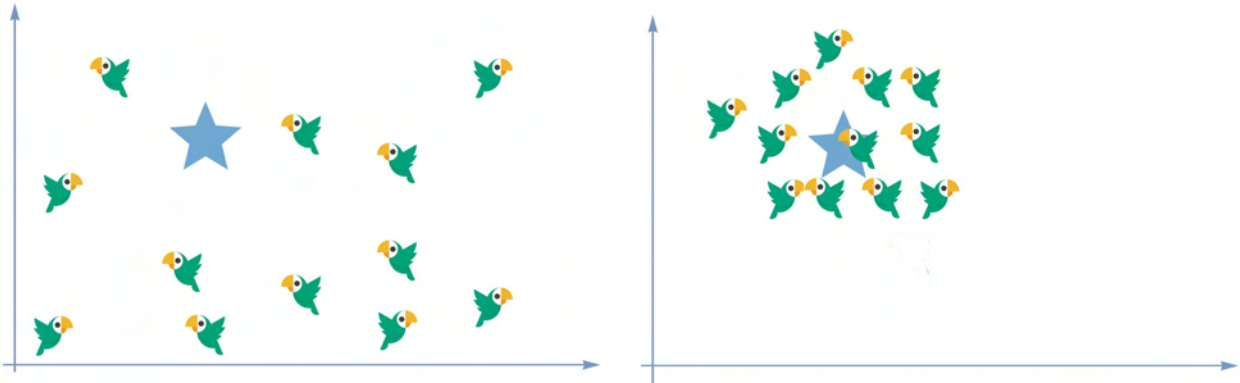
Καταρχάς, κάθε σωματίδιο χαρακτηρίζεται από μια ταχύτητα και από μια θέση. Η επόμενη θέση του σωματιδίου θα προσαρμόζεται σύμφωνα με την τρέχουσα θέση και την υπολογισμένη, βάσει των πληροφοριών, ταχύτητα. Οι πληροφορίες αυτές αφορούν την καλύτερη θέση που έχει βρει το ίδιο μέχρι τώρα, αλλά και την καλύτερη θέση των γειτόνων του.



Σχήμα 2.6: Αλλαγή θέσης ενός σωματιδίου του πληθυσμού.

Η ταχύτητα, τελικά, θα ενημερώνεται σύμφωνα με την τρέχουσα ταχύτητα του σωματιδίου και από δύο ακόμα συνιστώσες: τη διαφορά της τρέχουσας θέσης τόσο από τη δική του βέλτιστη όσο και από τη βέλτιστη γειτονική θέση, όπως φαίνεται στο σχήμα 2.6.

Έχοντας γνώση τόσο των δικών τους βέλτιστων θέσεων όσο και των γειτονικών, το σμήνος θα καταλήγει όλο και πιο πολύ στην περιοχή του χώρου που περιέχει καλές πλέον λύσεις (σχήμα 2.7) [Garg09], [Blum08].



Σχήμα 2.7: Αναζήτηση της βέλτιστης λύσης από τα σωματίδια του πληθυσμού.

Μέσα από αυτή τη στοχαστική διαδικασία τα άτομα του πληθυσμού αναζητούν μια καλή λύση κάνοντας χρήση της γνώσης των καλύτερων προσωπικών θέσεων και των καλύτερων γειτονικών. Η απόδοση κάθε σωματιδίου (πόσο κοντά είναι σε μια καλή θέση) μετράται από μια συνάρτηση καταλληλότητας που σχετίζεται με το προς λύση πρόβλημα [Engelbrecht07].

Τέτοιων αλγορίθμοι βρίσκουν εφαρμογή σε προβλήματα βελτιστοποίησης, στο σχεδιασμό δικύων μεταφοράς [Babazadeh11], στην προσέγγιση συνάρτησης, στην ομαδοποίηση, στη λύση εξισώσεων και σε άλλα [Engelbrecht07]. Περισσότερες πληροφορίες μπορούν να βρεθούν στο [Saggu13].

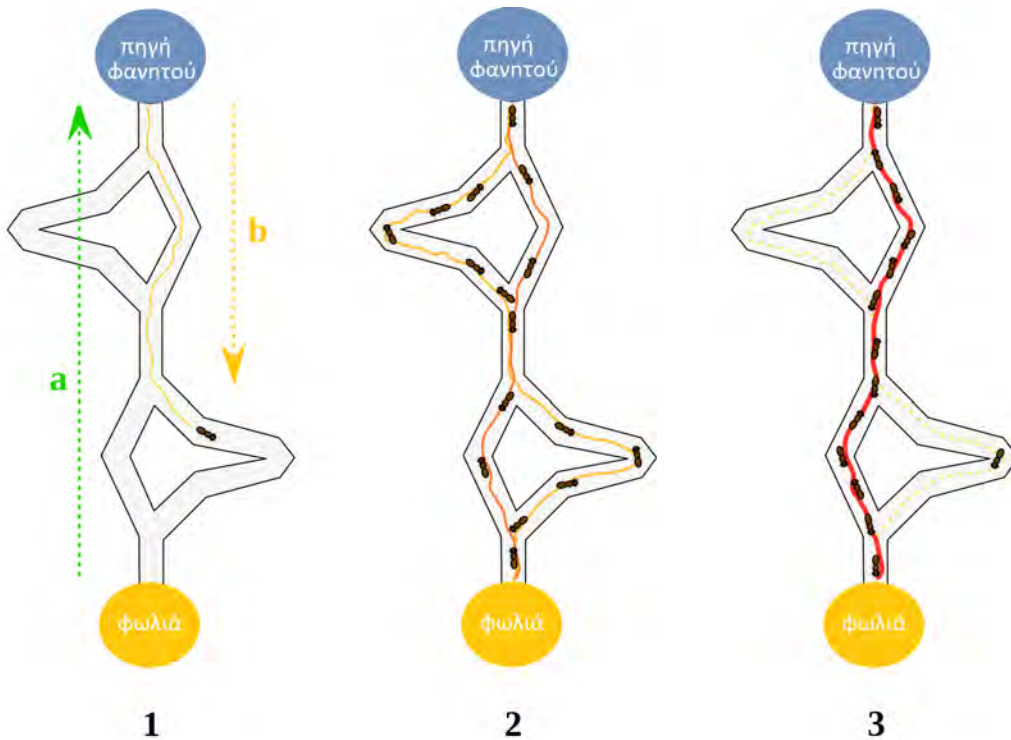
2.2.2.2 Βελτιστοποίηση με Αποικίες Μυρμηγκιών

Η μέθοδος βελτιστοποίησης με αποικίες μυρμηγκιών αποτελεί μια άλλη υποβέλτιστη μειωτική μέθοδο βελτιστοποίησης [Garg09], [Fox07]. Πρωτάθηκε για πρώτη φορά μέσα από τη διδακτορική διατριβή του Dorigo [Dorigo92] το 1992 [Fox07]. Εμπνευσμένος από την αναζήτηση του συντομότερου μονοπατιού των μυρμηγκιών από τη φωλιά έως την πηγή φαγητού, ο πρώτος αλγόριθμος αφορούσε την αναζήτηση βέλτιστου μονοπατιού σε ένα γράφημα. Βασίζεται σε πιθανοτικές τεχνικές για την επίλυση προβλημάτων που αφορούν την εύρεση καλών, σύντομων δηλαδή, διαδρομών [Garg09]. Ενώ στην αρχή τέτοιοι αλγόριθμοι χρησιμοποιήθηκαν για την επίλυση διακριτών προβλημάτων, η προσαρμογή τους για την επίλυση συνεχών προβλημάτων βελτιστοποίησης λαμβάνει διαρκώς αυξημένη προσοχή [Blum08].

Η αρχική ιδέα βασίζεται στη μοντελοποίηση της αναζήτησης σύντομου μονοπατιού από μια αποικία μυρμηγκιών προς μια πηγή φαγητού καθώς μέσα από εργαστηριακές μελέτες έχει βρεθεί πως τα μυρμηγκία τείνουν να διαλέξουν τη συντομότερη [Garg09]. Τα μυρμηγκία επικοινωνούν μέσω ενός κοινού φερομόνης και μέσα από τα ίχνη αυτά οδηγούνται το ένα πίσω από το άλλο στο ίδιο μονοπάτι [Fox07], [Garg09].

Μέσα από την αρχικά τυχαία επιλογή μονοπατιού από τη φωλιά έως την πηγή φαγητού, το μυρμηγκί επιστρέφει πίσω στην αποικία αφήνοντας πίσω ένα ίχνος φερομόνης. Κάποια από τα μυρμηγκία τείνουν να βαδίζουν προς αυτό. Κάποια θα ακολουθήσουν άλλο δρόμο. Όσο πιο μεγάλο είναι ένα μονοπάτι τόσο πιο γρήγορα θα χάνεται το ίχνος της φερομόνης, έτσι ώστε να μην ακολουθείται πλέον από τα μυρμηγκία. Το ίχνος της φερομόνης ενισχύεται στα σύντομα μονοπάτια ενώ, στο τέλος, υπερσχύει το συντομότερο εξ αυτών, όπως φαίνεται στο σχήμα 2.8². Έτσι αποφεύγεται η σύγκλιση σε τοπική βέλτιστη λύση στον αλγόριθμο [Garg09].

²https://upload.wikimedia.org/wikipedia/commons/thumb/a/af/Aco_branches.svg/2000px-Aco_branches.svg.png



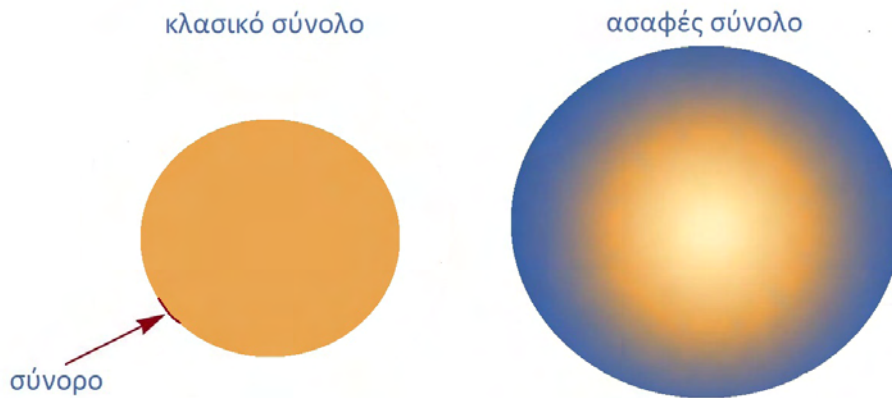
Σχήμα 2.8: Αναζήτηση του συντομότερου μονοπατιού από τα μυρμηγκία.

Ο αλγόριθμος βασίζεται σε αυτή τη μορφή επικοινωνίας των μυρμηγκιών, ενώ μιμείται αυτή τη διαδικασία με τα λεγόμενα προσομοιωμένα μυρμηγκία, τα οποία κινούμενα στο χώρο αναζήτησης αναπαριστούν τις πιθανές λύσεις [Garg09]. Μετράται η ποιότητα των θέσεων τους και, συνεπώς, των λύσεων, έτσι ώστε στις τελευταίες επαναλήψεις όλο και περισσότερα να δίνουν καλύτερες λύσεις [Kassabalidis01]. Κάποιες από τις εφαρμογές αυτού του είδους αλγορίθμων αναφέρονται στο [Fox07], [Saggu13].

2.2.3 Ασαφή Συστήματα

Στην κλασική θεωρία συνόλων τα αντικείμενα χαρακτηρίζονται ως μέρος ενός συνόλου ή όχι, με τις τιμές που μπορεί να πάρει ένα στοιχείο να είναι αντίστοιχα ή «1» ή «0» [Engelbrecht07]. Για παράδειγμα, το γραφείο θα ανήκει στο σύνολο «γραφεία». Ενδέχεται, όμως, να υπάρχουν και άλλα αντικείμενα που μπορούν να χρησιμοποιηθούν ως γραφεία όπως είναι, για παράδειγμα, το τραπέζι, ή ο πάγκος της κουζίνας. Επομένως, το νέο σύνολο που προκύπτει θα είναι διαφορετικό από το προηγούμενο [Βλαχάβας06].

Μερικές καταστάσεις, δηλαδή, δεν επιδέχονται ακριβή απάντηση καθώς τα κριτήρια που χαρακτηρίζουν ένα αντικείμενο ως μέρος ενός συνόλου δεν είναι αυστηρά [Βλαχάβας06]. Ένα τέτοιο σύνολο ονομάζεται ασαφές σύνολο και εκεί ακριβώς έγκειται η διαφορά με ένα κλασικό σύνολο: στο τελευταίο τα σύνορα του συνόλου θα είναι ευδιάκριτα, ενώ στο πρώτο θα ορίζονται ασαφώς, όπως φαίνεται στο σχήμα 2.9.



Σχήμα 2.9: Διαφορά κλασικού συνόλου από ένα ασαφές σύνολο.

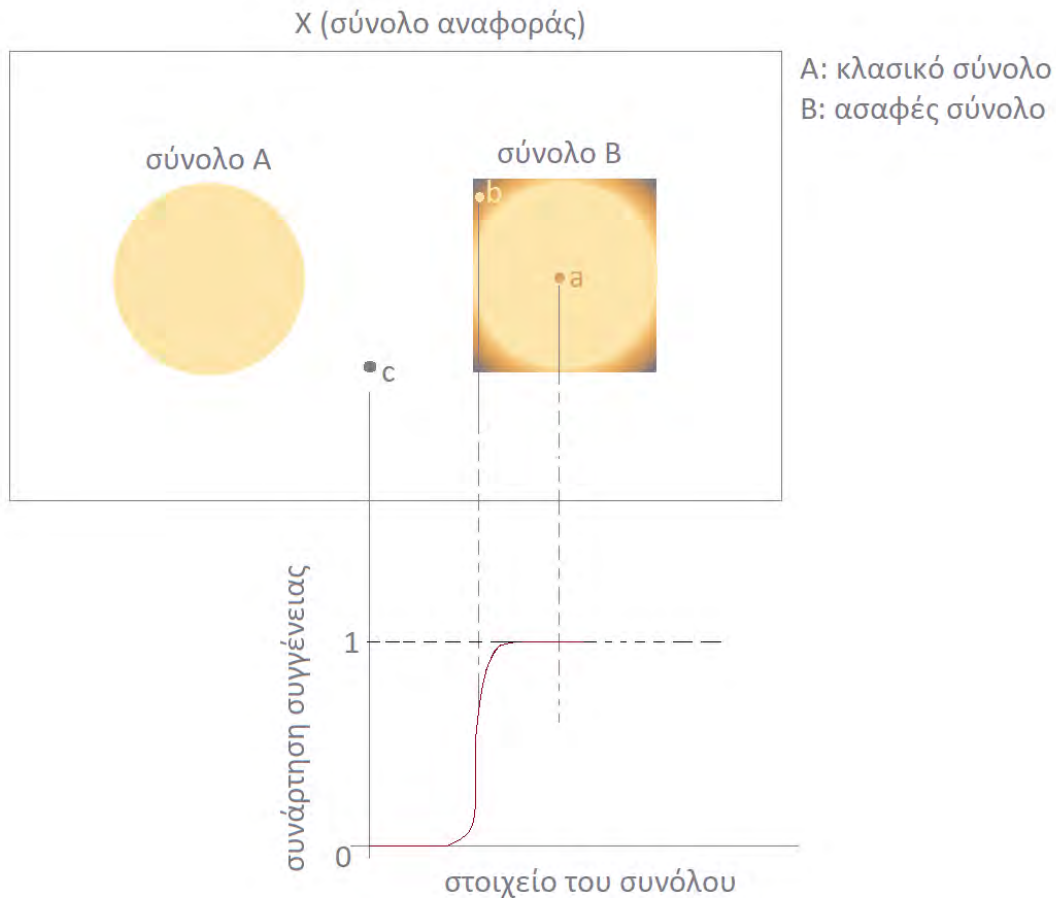
Δηλαδή σε ένα ασαφές σύνολο ένα αντικείμενο θα ανήκει στο σύνολο αυτό με ένα κάποιο βαθμό βεβαιότητας [Engelbrecht07].

Για παράδειγμα, στην πρόταση «έχει ζέστη έξω», κανείς θα μπορούσε να βγάλει συμπεράσματα για το σημερινό καιρό, όμως, τα συμπεράσματα αυτά θα συνοδεύονται από κάποια ασάφεια καθώς ο καθένας έχει διαφορετική αντίληψη. Πόση ζέστη έχει; Τι ορίζει την πολλή ή τη λίγη ζέστη; Φαίνεται ότι η ασάφεια είναι αποτέλεσμα της προσπάθειας ποσοτικοποίησης ποιοτικών πληροφοριών, προερχόμενη από μη ακριβή δεδομένα. Ωστόσο, αν και ο περιορισμός της ασάφειας επιτυγχάνεται με κάποια απόδοση τιμών, μπορεί ακόμα να εξαχθούν λάθος συμπεράσματα. Αν κανείς θεωρεί μια ζεστή μέρα με θερμοκρασία 35 βαθμούς κελσίου, τότε δεν θα ήταν σωστό να βγει το συμπέρασμα ότι μια μέρα με 34 βαθμούς κελσίου δεν είναι μια ζεστή μέρα [Βλαχάβας06].

Δεν θα πρέπει η αβεβαιότητα στα συστήματα αυτά να συγχέεται με τη στατιστική αβεβαιότητα [Engelbrecht07]. Από τη ρίψη ενός αμερόληπτου ζαριού εκ των προτέρων θα είναι γνωστή η πιθανότητα έκβασης για την κάθε πλευρά: 16.67%. Όμως η μη-στατιστική αβεβαιότητα, η αβεβαιότητα δηλαδή που χαρακτηρίζει τα ασαφή συστήματα, είναι μια εγγενής ιδιότητα και δεν μπορεί να μεταβληθεί ή να επιλυθεί μέσω των παρατηρήσεων, εν αντιθέσει με τη στατιστική αβεβαιότητα.

Τα ασαφή σύνολα και η ασαφής λογική, η οποία βασίζεται στα προηγούμενα, πρωτοοριζόμενη από τον L.Zadeh το 1965 [Zadeh65], επιτρέπουν τον προσεγγιστικό συλλογισμό, μοντελοποιούν δηλαδή την κοινή λογική: αβέβαια γεγονότα επεξεργάζονται με τη λογική αυτή, ώστε να βγουν συμπεράσματα με κάποιο βαθμό βεβαιότητας συνδεδεμένο με το κάθε γεγονός [Engelbrecht07]. Η ασαφής λογική, ως υπεрсύνολο της κλασικής λογικής, επιτρέπει το χειρισμό καταστάσεων, οι οποίες επιδέχονται απαντήσεις ανάμεσα στα απόλυτα 0 ή 1 [Βλαχάβας06].

Ένα αντικείμενο θα ανήκει σε ένα ασαφές σύνολο με μια βεβαιότητα, η οποία μετράται με μια τιμή. Η τιμή αυτή ονομάζεται βαθμός αληθείας (degree of truth) και παίρνει τιμές στο διάστημα $[0, 1]$. Ο βαθμός αληθείας χαρακτηρίζει το βαθμό της συγγένειας του στοιχείου με το εν λόγω σύνολο και προκύπτει από τη συνάρτηση συγγένειας (membership function). Το ευρύτερο σύνολο που περιλαμβάνει όλα τα στοιχεία στα οποία μπορεί να γίνει αναφορά ονομάζεται σύνολο αναφοράς (universe of discourse) [Βλαχάβας06]. Στο σχήμα 2.10 παρουσιάζεται μια τυπική μορφή της συνάρτησης συγγένειας για ένα ασαφές σύνολο.



Σχήμα 2.10: Συνάρτηση συγγένειας για ένα ασαφές σύνολο.

Αν και τα ασαφή σύνολα εμφανίστηκαν το 1968, χρειάστηκε κάποιος χρόνος για να περάσουν από τη θεωρία στη πράξη, όπου πλέον από το 1980 και μετά, στην Ιαπωνία και, έπειτα, στην Ευρώπη στις αρχές της δεκαετίας του 90', χρησιμοποιήθηκαν στις αυτοματοποιήσεις αλλά και στο βιομηχανικό έλεγχο [Jakel04]. Τα ασαφή συστήματα έχουν χρησιμοποιηθεί με επιτυχία στον έλεγχο συστημάτων, στο σύστημα φρένων στα αυτοκίνητα και στον έλεγχο φαναριών [Engelbrecht07]. Ακόμα, στην αυτόματη εστίαση φωτογραφικών μηχανών ή στα πλυντήρια ρούχων με προσαρμοσμένο πρόγραμμα πλύσης, αναλόγως την ποσότητα των ρούχων ή την κατάσταση στην οποία βρίσκονται [Βλαχάβας06]. Επίσης, στην ταξινόμηση, στην ομαδοποίηση ή στην προσέγγιση συνάρτησης. Για παράδειγμα στην ταξινόμηση, στην οποία πρέπει να δωθεί σε κάθε αντικείμενο μια ετικέτα, οι ταξινομητές την ορίζουν με το βαθμό συγγένειας του αντικειμένου που έχει με την κάθε κλάση. Περισσότερες πληροφορίες μπορούν να βρεθούν στο [Jakel04].

2.3 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα

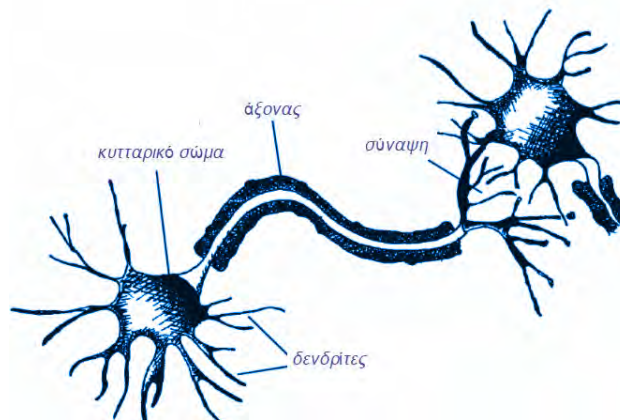
2.3.1 Βιολογικοί Νευρώνες και Βιολογικά Νευρωνικά Δίκτυα

Οι άνθρωποι χαρακτηρίζονται με κάποιο βαθμό αντίληψης και ως εκ τούτου είναι ικανοί για διάφορες εργασίες όπως είναι η διαδικασία λήψης αποφάσεων, η αναγνώριση προσώπων ή αντικειμένων και άλλες, για τις οποίες ακόμα και ο πιο σύγχρονος υπολογιστής μάλλον υστερεί αυτών των δυνατοτήτων. Ο άνθρωπος καθημερινά λαμβάνει πληροφορίες και μέσω των αισθήσεων του, χωρίς να απαιτείται κάποια σπουδαία προγενέστερη προσπάθεια, αναγνωρίζει την πηγή των δεδομένων [Brian96]. Η πληροφορία που εισέρχεται στον εγκέφαλο διανέμεται σε όλο τον όγκο του, όπου και επεξεργάζεται παράλληλα [Βλαχάβας06]. Ο ανθρώπινος εγκέφαλος χαρακτηρίζεται από τις δυνατότες για μάθηση, τη μετέπειτα απομνημόνευση, ακόμα και τη γενίκευση των γνώσεων του. Τα προηγούμενα καλά χαρακτηριστικά του εγκεφάλου μαζί με την γρήγορη επεξεργασία των πληροφοριών στο παράλληλο και καταμεμημένο υπολογιστικό σύστημα, αποτέλεσαν πηγή έμπνευσης για τη μοντελοποίηση των βιολογικών νευρωνικών συστημάτων και την ανάπτυξη των Τεχνητών Νευρωνικών Δικτύων, ενώ μια πρώτη προσέγγιση αυτών προτάθηκε από τον Rosenblatt το 1958, ορίζοντας το λεγόμενο perceptron [Engelbrecht07], [Βλαχάβας06].

Ο ανθρώπινος εγκέφαλος, εκτός των άλλων, αποτελείται από νευρώνες, οι οποίοι αποτελούν τη δομική μονάδα του. Πρόκειται για ειδικού τύπου κύτταρα, τα οποία είναι υπεύθυνα για τη λήψη, την επεξεργασία και τη μετάδοση πληροφοριών μέσω ηλεκτρικών και χημικών σημάτων [Nguyen], ενώ αποτελούν το λειτουργικό πυρήνα του εγκεφάλου [Brotherson18]. Από τη στιγμή που γεννιέται ένας άνθρωπος αρχίζει να μορφοποιείται ο εγκέφαλός του βάσει εμπειριών. Τυπικά ο εγκέφαλος ενός νεογέννητου έχει την ικανότητα μάθησης, όμως, προηγουμένως απαιτείται η σύζευξη των νευρώνων, κάτι το οποίο επιτυγχάνεται μέσω των πρώιμων εμπειριών του [Brotherson18].

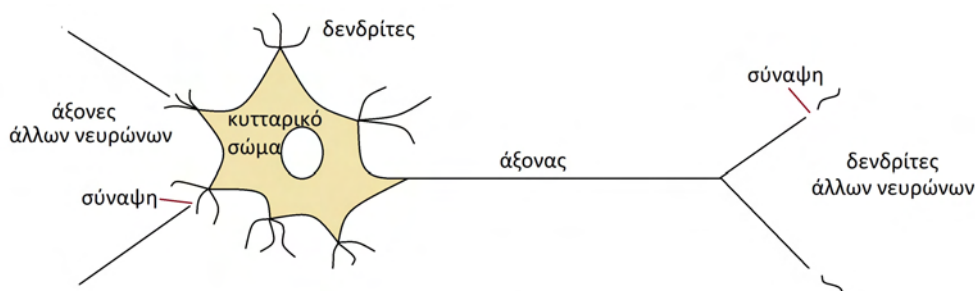
Ένας άνθρωπος γεννιέται με περισσότερους από 100 δις. νευρώνες, χωρίς να δημιουργούνται νέοι στη διάρκεια της ζωής του [Brotherson18]. Ο αριθμός αυτός υπολογίζεται να είναι περίπου 10 φορές μεγαλύτερος του συνολικού αριθμού των αστεριών στο γαλαξία [Brotherson18]. Κάθε νευρώνας είναι συνδεδεμένος με άλλους, δημιουργώντας έτσι 150 τρις. διασυνδέσεις μεταξύ τους [Drach2005]. Όλοι αυτοί οι νευρώνες και οι διασυνδέσεις δημιουργούν χιλιάδες νευρωνικά δίκτυα στον εγκέφαλο [Engelbrecht07]. Για το λόγο αυτό είναι αδύνατο να επιτευχθεί η πραγματική αντιγραφή του ανθρώπινου εγκεφάλου με τα Τεχνητά Νευρωνικά Δίκτυα να αποτελούνται από κάποιες χιλιάδες νευρώνες με το πολύ 1 εκ. συνάψεις με περιορισμένη λειτουργικότητα [Βλαχάβας06]. Χρησιμοποιούνται μικρά Τεχνητά Νευρωνικά Δίκτυα που επιτελούν την εκπλήρωση απλών εφαρμογών όπως, για παράδειγμα, σε προβλήματα κατηγοριοποίησης, αναγνώρισης ή πρόβλεψης τα οποία υπάρχουν σε πολλές ανθρώπινες δραστηριότητες [Βλαχάβας06].

Η δομή και η σύνδεση των βιολογικών νευρώνων απεικονίζονται στο σχήμα 2.11 [Brotherson18].



Σχήμα 2.11: Δομή και σύνδεση βιολογικών νευρώνων.

Ένας βιολογικός νευρώνας (σχήμα 2.12) αποτελείται από το κυτταρικό σώμα, τους δενδρίτες και τον άξονα. Το κυτταρικό σώμα αποτελεί τον πυρήνα του νευρώνα, οι δενδρίτες, τα κλαδιά που αναδύονται από το σώμα, λαμβάνουν σήματα από τους γειτονικούς νευρώνες, ενώ ο άξονας αποτελεί την έξοδο του νευρώνα μέσω του οποίου αποστέλλονται σήματα σε όλους τους συνδεδεμένους σε αυτόν νευρώνες. Η διασύνδεση ενός νευρώνα με έναν άλλον πραγματοποιείται μέσω του άξονα του ενός και τους δενδρίτες του άλλου και συγκεκριμένα μέσω των συνάψεων, το απειροελάχιστο κενό ανάμεσα στους δενδρίτες [Brotherson18], [Engelbrecht07], [Βλαχάβας06].

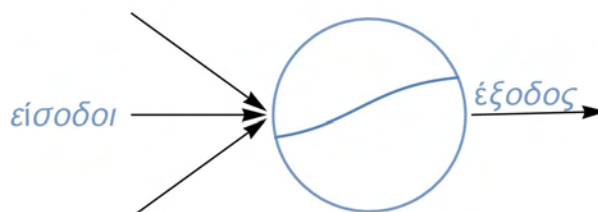


Σχήμα 2.12: Σκαρίφημα βιολογικού νευρώνα.

Τα σήματα που εισέρχονται από τους δενδρίτες στο κυτταρικό σώμα του νευρώνα συνδυάζονται και το επεργασμένο πια σήμα αν ξεπεράσει κάποια συγκεκριμένη τιμή, την τιμή κατωφλίου, τότε μεταδίδεται από τον άξονα σε όλους τους συνδεδεμένους σε αυτόν δενδρίτες, διαμέσου των συνάψεων. Στις συνάψεις και στη μεταβολή της αγωγιμότητας τους οφείλεται η ικανότητα μάθησης και μνήμης καθώς μέσω αυτών και των χημικών διαδικασιών που πραγματοποιούνται, επιταχύνεται ή επιβραδύνεται η ροή των ηλεκτρικών φορτιών από τους άξονες στους δενδρίτες και μετέπειτα στο σώμα του νευρώνα. Επομένως, η επεξεργασία στο σώμα του νευρώνα μπορεί να διεγείρει ή να αναστείλλει το εισερχόμενο συνδυασμένο σήμα. Η μεταφορά στον άξονα, και κατ' επέκταση στους υπόλοιπους νευρώνες, γίνεται όταν ο νευρώνας «πυροδοτεί», δηλαδή όταν ξεπεραστεί η τιμή κατωφλίου [Βλαχάβας06].

2.3.2 Τεχνητός Νευρώνας

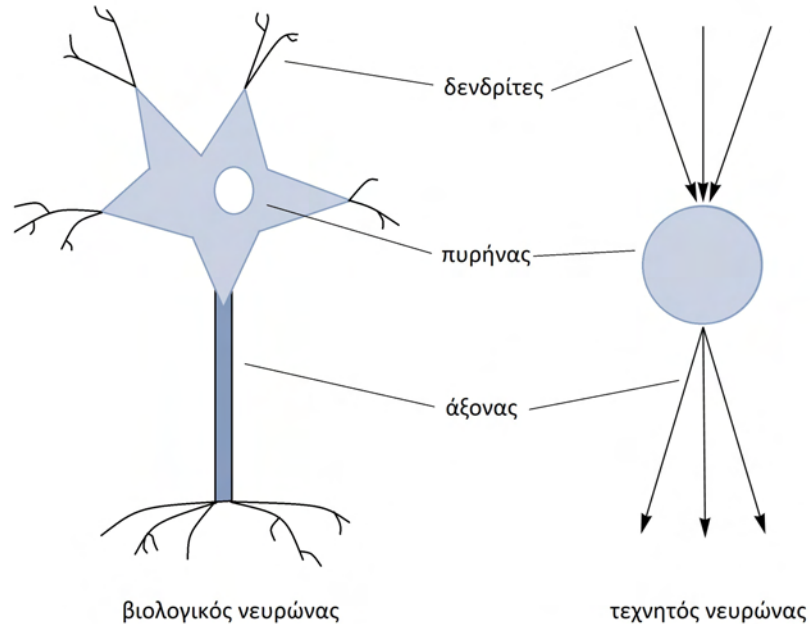
Σε αυτήν ακριβώς τη λειτουργία των βιολογικών νευρωνικών δικτύων βασίζεται η μοντελοποίηση των τεχνητών (σχήμα 2.13). Κι εδώ, ο τεχνητός, πλέον, νευρώνας θα δέχεται πληροφορίες (σήματα), είτε από το εξωτερικό περιβάλλον του, είτε από το εσωτερικό, από άλλους δηλαδή νευρώνες, θα επεξεργάζεται τις πληροφορίες αυτές και όταν «πυροδοτεί», θα στέλνει σε όλους τους συνδεδεμένους σε αυτόν νευρώνες σήμα [Engelbrecht07].



Σχήμα 2.13: Τεχνητός νευρώνας.

2.3.2.1 Αναλογίες με τους Βιολογικούς Νευρώνες

Ο τεχνητός νευρώνας αποτελεί ένα υπολογιστικό μοντέλο βασισμένο στους βιολογικούς, οπότε θα υπάρχουν και οι αντίστοιχες αναλογίες, όπως φαίνεται στο σχήμα 2.14.

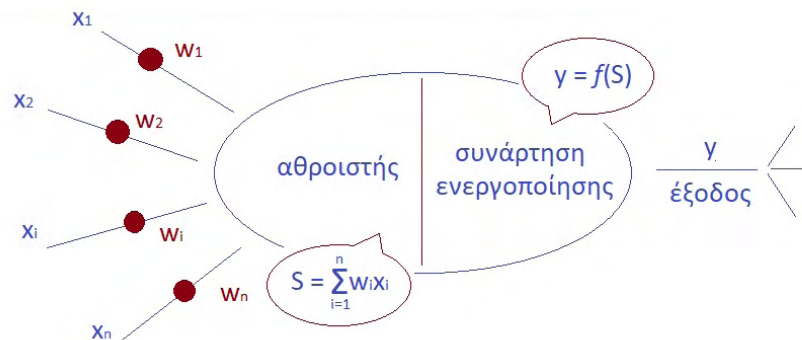


Σχήμα 2.14: Αναλογίες μεταξύ τεχνητού και βιολογικού νευρώνα.

Κάθε μέρος του τεχνητού νευρώνα σχετίζεται άμεσα με αυτά του βιολογικού, ασχέτως της περίπλοκης μορφής του τελευταίου [Debes95]. Οι δενδρίτες πλέον θα αποτελούν την είσοδο, το κυτταρικό σώμα τη μονάδα επεξεργασίας των εισόδων και, τέλος, ο άξονας την έξοδο του νευρώνα. Η σύναψη θα ονομάζεται βάρος και η τιμή μαζί με το πρόσημο αυτού θα μεταβάλλουν την είσοδο καθώς, επίσης, θα δηλώνουν την αναστολή ή τη διέγερση των σημάτων.

2.3.3 Μοντέλο Τεχνητού Νευρώνα

Συγκεκριμένα, ένας τεχνητός νευρώνας θα έχει την παρακάτω μορφή [Βλαχάβας06]:



Σχήμα 2.15: Μοντέλου τεχνητού νευρώνα.

Ένας τεχνητός νευρώνας δέχεται κάποιες εισόδους $x_1, x_2 \dots x_n$, οι οποίες θα μεταβάλλονται σύμφωνα με

την τιμή του αντίστοιχου βάρους, $w_1, w_2 \dots w_n$. Όλα τα σήματα που καταφθάνουν τελικά στο νευρώνα θα αθροίζονται μέσω ενός σταθμισμένου αθροίσματος:

$$S = \sum_{i=1}^n w_i x_i = w_1 x_1 + w_2 x_2 \dots w_n x_n \quad (2.1)$$

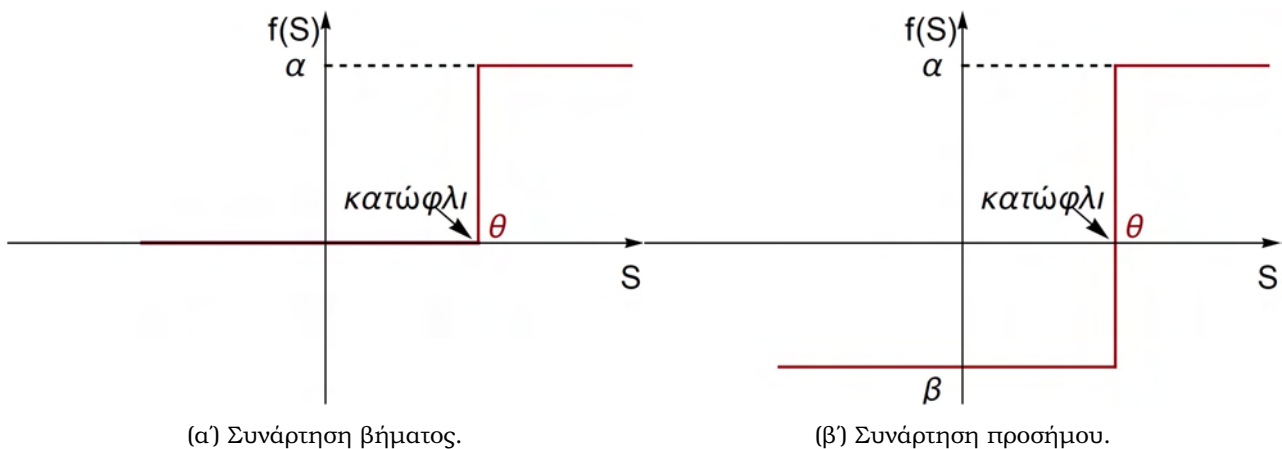
Αυτό πραγματοποιείται στο πρώτο μέρος, τον αθροιστή (sum). Στη συνέχεια το σταθμισμένο πλέον άθροισμα θα αποτελεί την είσοδο στο δεύτερο μέρος του τεχνητού νευρώνα, τη συνάρτηση ενεργοποίησης (activation function). Η επεξεργασία του αθροίσματος των σημάτων γίνεται μέσω αυτής της συνάρτησης, η οποία δίνει την έξοδο του νευρώνα. Αν και μπορεί σε κάθε νευρώνα να είναι συνδεδεμένοι πολλοί άλλοι, η τιμή της εξόδου θα είναι ίδια για όλους τους συνδεδεμένους σε αυτόν νευρώνες [Βλαχάδας06].

2.3.3.1 Συναρτήσεις ενεργοποίησης

Η συνάρτηση ενεργοποίησης θα καθορίζει την ενεργοποίηση ή κατάσταση (τιμή) της εξόδου του νευρώνα, αφού συνδυάσει την κάθε είσοδο με το αντίστοιχο βάρος [Debes95]. Υπάρχουν πολλών ειδών συναρτήσεις ενεργοποίησης που μπορούν να χρησιμοποιηθούν, με τις περισσότερες εξ αυτών να είναι γνησίως αύξουσες [Debes95], [Engelbrecht07]. Όπως συμβαίνει και με τους βιολογικούς νευρώνες, με τα ενισχυμένα σήματα εισόδου να επιδρούν αναλόγως, έτσι, κι εδώ, ενισχύεται η δραστηριότητα ενός τεχνητού νευρώνα με τις μεγάλες εισόδους [Debes95]. Ωστόσο, οι βιολογικοί νευρώνες έχουν ένα άνω όριο στη συχνότητα πυροδότησης και, επομένως, θα υπάρχει, συνήθως, ένα άνω όριο στις εξόδους των συναρτήσεων ενεργοποίησης [Debes95], [Sarpeshkar92].

Μερικές συναρτήσεις ενεργοποίησης είναι οι εξής [Βλαχάδας06], [Engelbrecht07]:

- η συνάρτηση βήματος (step function) ή η συνάρτηση προσήμου (sign function),



Σχήμα 2.16: Συναρτήσεις ενεργοποίησης.

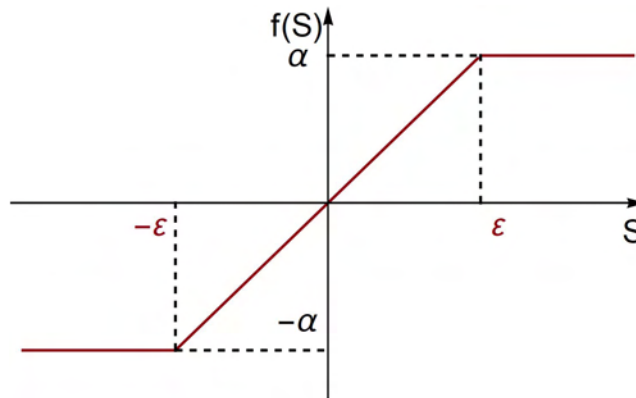
Στην πρώτη περίπτωση, η συνάρτηση δίνει έξοδο μόνο αν το αποτέλεσμα του αθροιστή είναι μεγαλύτερο από κάποια τιμή κατωφλίου θ (σχήμα 2.16α)

$$f(S) = \begin{cases} \alpha, & S \geq \theta \\ 0, & S < \theta \end{cases} \quad (2.2)$$

ενώ στη δεύτερη, η έξοδος μπορεί να είναι είτε θετική, είτε αρνητική, αν, αντίστοιχα, η τιμή του αθροιστή είναι μεγαλύτερη ή μικρότερη από την τιμή κατωφλίου (σχήμα 2.16β).

$$f(S) = \begin{cases} \alpha, & S \geq \theta \\ \beta, & S < \theta \end{cases} \quad (2.3)$$

• η **συνάρτηση ράμπας (ramp function)**, η οποία αποτελεί συνδυασμό γραμμικών και βηματικών συναρτήσεων (σχήμα 2.17).



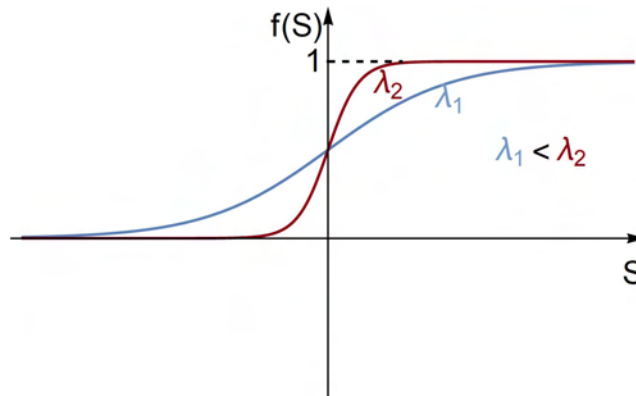
Σχήμα 2.17: Συνάρτηση ράμπας.

όπου αναλόγως με την τιμή που εισάγεται στον αθροιστή αυτός μπορεί να δώσει

$$f(S) = \begin{cases} \alpha, & S \geq +\epsilon \\ S, & -\epsilon < S < \epsilon \\ -\alpha, & S \leq -\epsilon \end{cases} \quad (2.4)$$

• **σιγμοειδής συνάρτηση (sigmoid function)**

Υπάρχουν και συναρτήσεις ενεργοποίησης οι οποίες έχουν τις επιθυμητές ιδιότητες της συνέχειας και της παραγωγίσιμης σε όλο το πεδίο τιμών εισόδου. Τέτοιες είναι, για παράδειγμα, οι σιγμοειδείς συναρτήσεις (σχήμα 2.18).



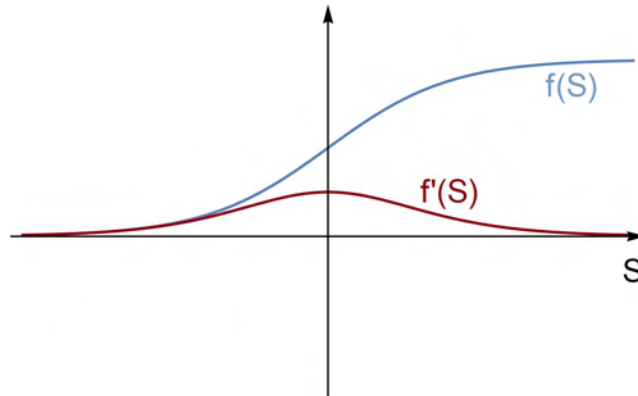
Σχήμα 2.18: Σιγμοειδής συνάρτηση.

Η σιγμοειδής συνάρτηση ορίζεται ως

$$f(S) = \frac{1}{1 + e^{-\lambda S}} \quad (2.5)$$

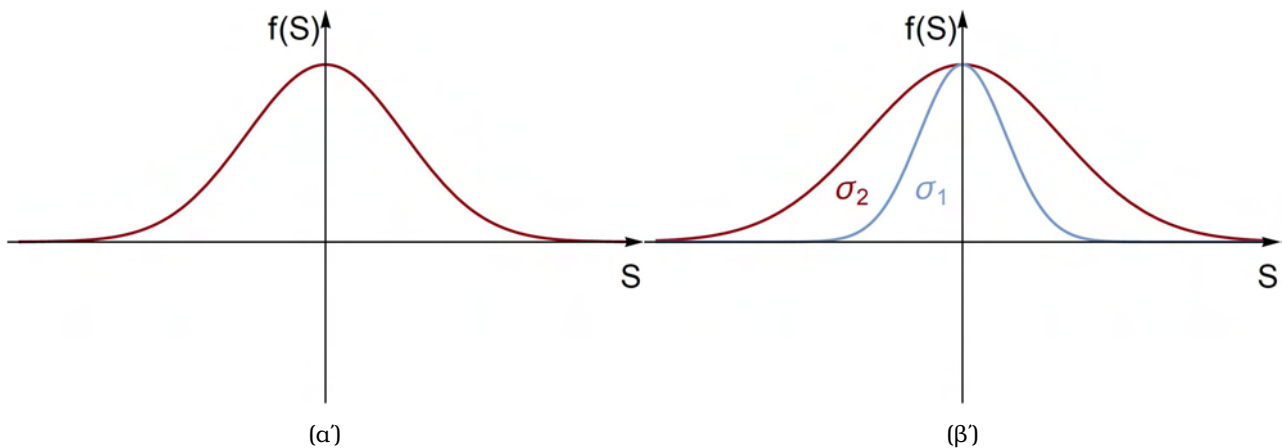
και αποκαλείται έτσι, λόγω της μορφής της η οποία μοιάζει με «S». Η παράμετρος λ ($\lambda > 0$) καθορίζει το ρυθμό μετάβασης από τις χαμηλές στις υψηλές τιμές. Όσο μεγαλύτερη είναι η τιμή του, τόσο πιο απότομα γίνεται αυτή η μετάβαση, ενώ για πολύ μεγάλες τιμές η συνάρτηση αυτή προσεγγίζει τη βηματική.

Η σιγμοειδής συνάρτηση περιορίζει τις τιμές εξόδου ενώ ενδιαφέρουσα είναι και η μορφή της παραγώγου, η οποία μοιάζει με μορφή «καμπάνας» (σχήμα 2.19).



Σχήμα 2.19: Σιγμοειδής συνάρτηση και παράγωγος.

Αυτές οι μορφές των συναρτήσεων είναι σημαντικές καθώς υπάρχει μια ενδιάμεση περιοχή τιμών στην οποία η έξοδος είναι αξιολογήσιμη, όπως συμβαίνει και με την περίπτωση της συνάρτησης Gaussian (σχήμα 2.20α')



Σχήμα 2.20: Συνάρτηση Gaussian.

όπου

$$f(S) = e^{-\frac{S^2}{\sigma^2}} \quad (2.6)$$

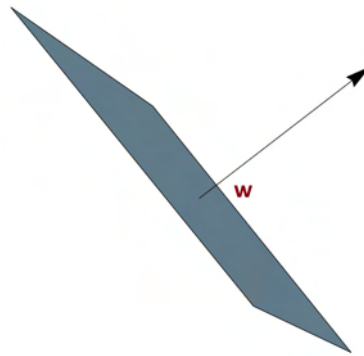
Το σ είναι η τυπική απόκλιση ενώ το S το μέσο. Το μέγιστο δίνεται για $S = 0$ ενώ όσο αυξάνεται ή μειώνεται το S , τόσο θα μειώνεται η τιμή της συνάρτησης. Αναλόγως με τις τιμές του εύρους, η μορφή της συνάρτησης θα αλλάζει καθώς μικρές τιμές καθιστούν την καμπύλη πιο ευρεία, ώστε να υπάρχει ένα μεγαλύτερο μέρος των περιοχών που να είναι αξιολογήσιμες, όπως φαίνεται στο σχήμα 2.20β' στο οποίο $\sigma_2 > \sigma_1$.

2.3.3.2 Γεωμετρική ερμηνεία ενός Τεχνητού Νευρώνα

Ένας νευρώνας ορίζει ένα υπερεπίπεδο στο χώρο με τόσες διαστάσεις όσος είναι και ο αριθμός των εισόδων ενώ χωρίζει το χώρο αυτό σε δύο περιοχές [Βλαχάβας06]. Όλα τα σημεία που ανήκουν στο υπερεπίπεδο αυτό ικανοποιούν την παρακάτω σχέση [Chrysala10]:

$$\sum_{i=1}^n w_i x_i = 0 \quad (2.7)$$

Γεωμετρικά το διάνυσμα των βαρών είναι κάθετο σε αυτό το υπερεπίπεδο, όπως φαίνεται στο σχήμα 2.21.



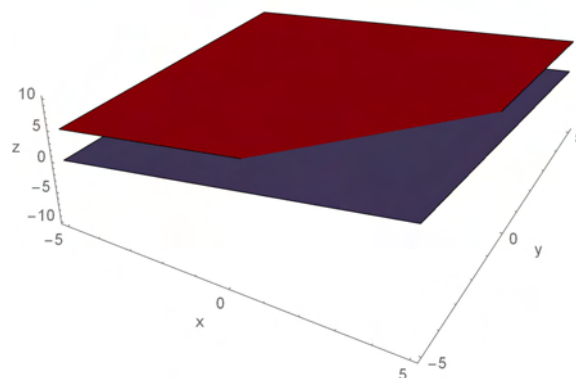
Σχήμα 2.21: Γεωμετρική ερμηνεία διανύσματος βαρών.

Οι τιμές των βαρών καθορίζουν την κλίση του υπερεπιπέδου, το οποίο σύμφωνα με την παραπάνω εξίσωση θα είναι υποχρεωμένο να διέρχεται πάντα από την αρχή των αξόνων. Για να μη συμβαίνει αυτό, τις περισσότερες φορές προστίθεται και μια εξωτερική είσοδος, η οποία πάντα θα έχει τιμή εισόδου 1 με το αντίστοιχο βάρος να ορίζει τη θέση του υπερεπιπέδου στο χώρο, ώστε να μην περιορίζεται η θέση της συνάρτησης ενεργοποίησης [Βλαχάβας06].

Συνολικά, ο αθροιστής θα είναι:

$$S = \sum_{i=1}^n w_i x_i + w_0 x_0 \quad (2.8)$$

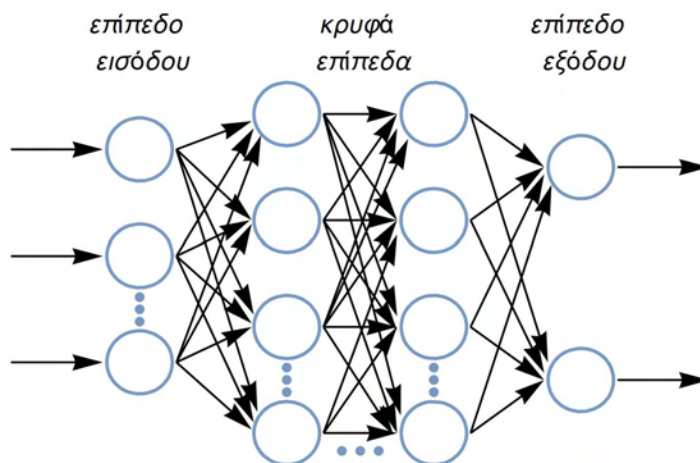
Το βάρος w_0 ονομάζεται πόλωση (bias) και ορίζει τη θέση του υπερεπιπέδου στο χώρο (σχήμα 2.22).



Σχήμα 2.22: Ρόλος της πόλωσης.

2.3.4 Δομή ενός Τεχνητού Νευρωνικού Δικτύου

Ένα Τεχνητό Νευρωνικό Δίκτυο θα αποτελείται από πολλούς νευρώνες διατεταγμένους σε δομές παρόμοιες με αυτές των βιολογικών νευρωνικών δικτύων [Βλαχάδας06], [Kang05]. Η επεξεργασία των δεδομένων θα λαμβάνει χώρα σε αυτό το πολυστρωματικό σύστημα νευρώνων, με μια τυπική του μορφή να απεικονίζεται στο σχήμα 2.23 [Βλαχάδας06]:

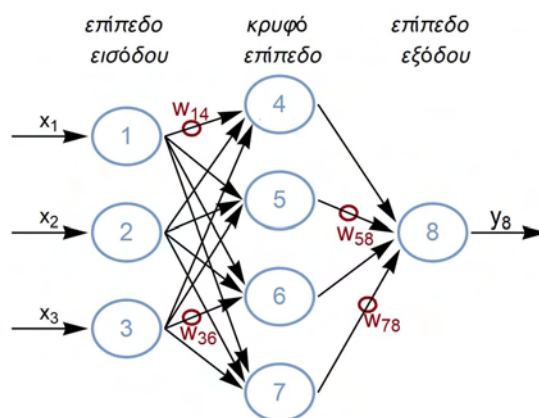


Σχήμα 2.23: Τυπική δομή ενός νευρωνικού δικτύου.

Η δομή ενός Τεχνητού Νευρωνικού Δικτύου αποτελείται από 3 επίπεδα. Το πρώτο επίπεδο ονομάζεται επίπεδο εισόδου και είναι αυτό το οποίο παραλαμβάνει τα σήματα από το περιβάλλον, χωρίς, ωστόσο, να γίνεται καμία επεξεργασία αυτών. Η μόνη εργασία που επιτελεί το αδρανές αυτό επίπεδο είναι η μεταφορά των δεδομένων εισόδου στους νευρώνες του επόμενου επιπέδου και, επομένως, δεν πρόκειται για νευρώνες, αλλά για κάποιες απλές μονάδες, οι οποίες στέλνουν τα σήματα στις κύριες μονάδες επεξεργασίας. Γι' αυτό, πολλές φορές το επίπεδο εισόδου δεν αναφέρεται ως επίπεδο, αφήνοντας πλέον 2 κύρια επίπεδα.

Τα επόμενα επίπεδα είναι ενεργά με την έννοια ότι γίνεται επεξεργασία των δεδομένων. Τα ενδιάμεσα επίπεδα ονομάζονται κρυφά επίπεδα και μπορεί να είναι 1 ή και περισσότερα. Ονομάζονται έτσι επειδή η έξοδος από τους νευρώνες που ανήκουν σε αυτά είναι κρυφή, δεν είναι φανερή ως εξωτερικό αποτέλεσμα, αλλά ως εσωτερικό, το οποίο στέλνεται στους επακόλουθους νευρώνες. Το τελευταίο επίπεδο ονομάζεται επίπεδο εξόδου και δίνει την έξοδο του δικτύου.

Για παράδειγμα, στο δίκτυο του σχήματος 2.24



Σχήμα 2.24: Τεχνητό νευρωνικό δίκτυο δόμης 3-4-1.

υπάρχουν συνολικά 8 νευρώνες. Το επίπεδο εισόδου αποτελείται από 3, το κρυφό επίπεδο από 4 και το επίπεδο εξόδου από 1 νευρώνα. Πρόκειται δηλαδή για ένα ΤΝΔ (3, 4, 1). Η έξοδος αυτού του δικτύου προκύπτει με την εξής διαδικασία: οι εισοδοί x_1, x_2, x_3 (διάνυσμα εισόδου) περνούν στο επίπεδο εισόδου, χωρίς κάποια τροποποίηση, και στη συνέχεια διαμερίζονται σε όλους τους νευρώνες του κρυφού επιπέδου με το αντίστοιχο βάρος. Σε κάθε έναν από αυτούς τους νευρώνες ο αθροιστής ορίζει την είσοδο στη συνάρτηση ενεργοποίησης, η οποία και αφήνει ένα αποτέλεσμα. Το αποτέλεσμα αυτό προωθείται στο νευρώνα του επόμενου επιπέδου, τροποποιείται από τα αντίστοιχα βάρη και ορίζει τη συνολική είσοδο στον νευρώνα αυτό, όπου εκεί, μετά από κάποια επεξεργασία, δίνεται η τελική έξοδος y_8 (διάνυσμα εξόδου) [Βλαχάβας06].

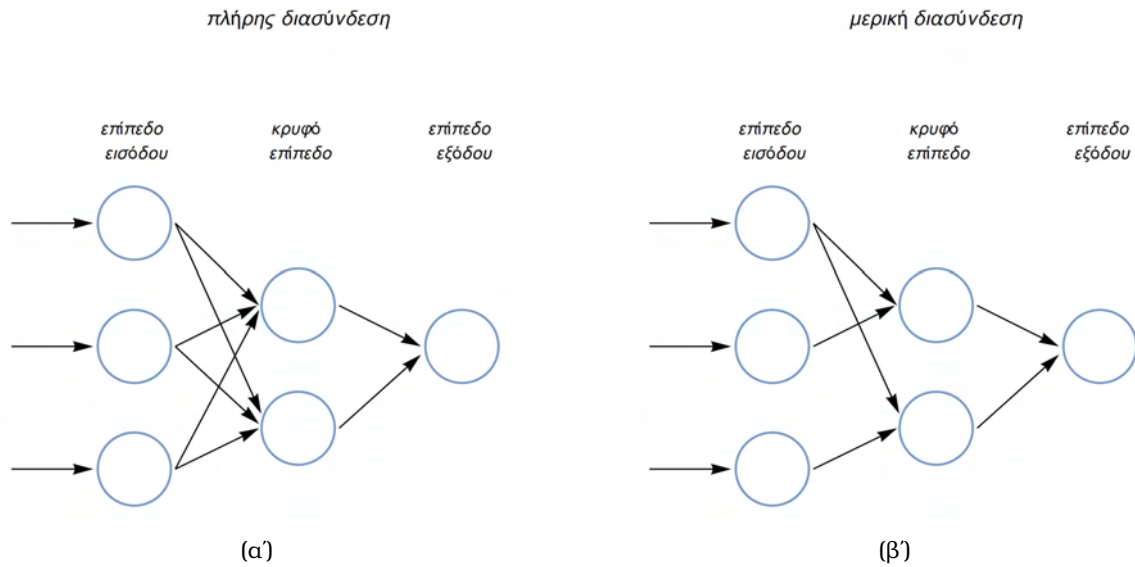
Η παραπάνω διαδικασία είναι γνωστή ως ανάκληση (recall). Αφού γίνουν γνωστές οι παράμετροι του δικτύου, τότε είναι δυνατόν να υπολογιστεί ένα διάνυσμα εξόδου για κάποιο διάνυσμα εισόδου. Σημαντικό, όμως, είναι προηγουμένως να περάσει το δίκτυο από τη διαδικασία της εκπαίδευσης (training), μέσω της οποίας καθίστανται γνωστές οι τιμές των παραμέτρων. Δηλαδή μέσα από τη μάθηση (learning) από ένα σύνολο δεδομένων εκπαίδευσης, τροποποιούνται οι τιμές των παραμέτρων του δικτύου, έτσι ώστε για κάθε ένα από τα παραδείγματα εκπαίδευσης, η είσοδος να δίνει συγκεκριμένη έξοδο. Τα κριτήρια τερματισμού για έναν αλγόριθμο μάθησης μπορεί να είναι οι μέγιστες επαναλήψεις εκπαίδευσης ή μείωση του σφάλματος σε επιθυμητά επίπεδα [Engelbrecht07].

Η μάθηση από τα παραδείγματα πραγματοποιείται σε μια σειρά κύκλων εκπαίδευσης, οι οποίοι ονομάζονται και εποχές (epochs). Στην αρχή του κάθε κύκλου εισάγονται τα παραδείγματα εκπαίδευσης και αναλόγως της μεθόδου τροποποίησης των βαρών διακρίνονται δύο τεχνικές μάθησης: η μάθηση δέσμης (batch learning) και η επαυξητική μάθηση (incremental learning). Στην πρώτη περίπτωση, τα βάρη διαμορφώνονται στο τέλος του κύκλου σύμφωνα με τη συσσωρευμένη μεταβολή των βαρών που προκάλεσε το κάθε σφάλμα σε κάθε ένα από τα παραδείγματα, αφότου έχουν περάσει όλα τα παραδείγματα. Στη δεύτερη περίπτωση, τα βάρη μεταβάλλονται μετά το πέρασμα του κάθε παραδείγματος. Η πρώτη μέθοδος είναι γρηγορότερη, ωστόσο απαιτεί μεγαλύτερες απαιτήσεις σε μνήμη, ενώ μπορεί να χρησιμοποιηθεί και συνδυασμός αυτών [Βλαχάβας06], [Engelbrecht07].

Ακόμα, σημαντική θεωρείται η προεπεξεργασία των δεδομένων πριν ξεκινήσει η εκπαίδευση του δικτύου. Σε μερικές πρακτικές εφαρμογές η προεπεξεργασία των δεδομένων αποτελεί το σημαντικότερο παράγοντα που καθορίζει την απόδοση του τελικού συστήματος [Bishop95]. Η επεξεργασία αυτή περιλαμβάνει, εκτός των άλλων, το διαχωρισμό του συνόλου δεδομένων σε τρία ανεξάρτητα σύνολα αλλά και την κανονικοποίηση των δεδομένων, καθώς μπορεί οι τιμές στο ίδιο σύνολο να διαφέρουν αρκετές τάξεις μεγέθους. Μέσω, λοιπόν, ενός γραμμικού μετασχηματισμού και το διαχωρισμό του συνόλου δεδομένων μπορεί να εξασφαλιστεί μια σωστή εκπαίδευση.

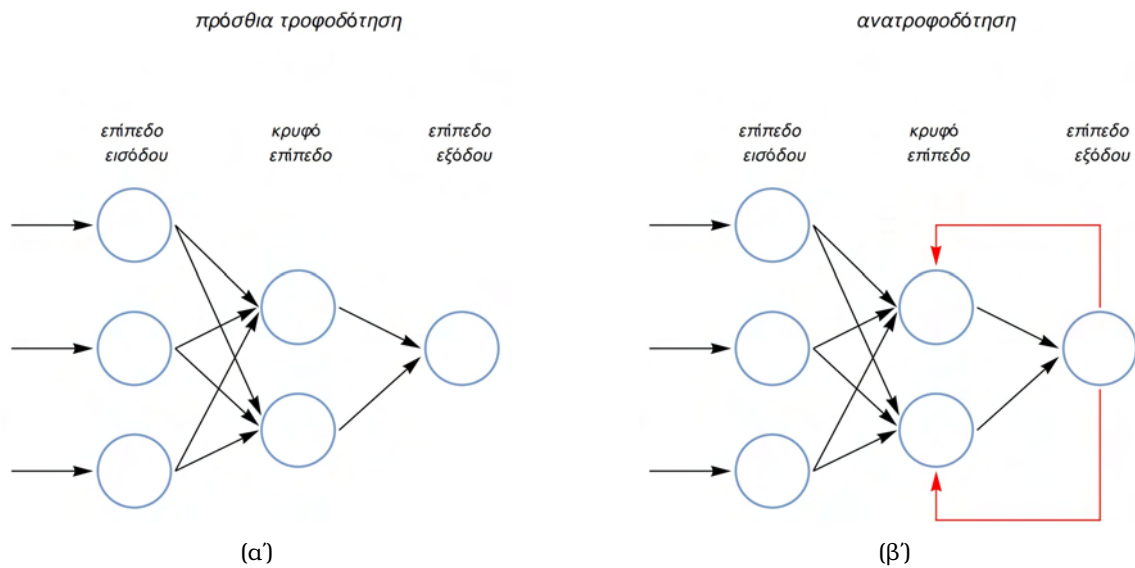
2.3.5 Αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων

Έχουν αναπτυχθεί διάφορες αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων αναλόγως της σύνδεσης που λαμβάνει χώρα και της τοπολογίας του δικτύου [Βλαχάβας06], [Engelbrecht07]. Στο δίκτυο μπορεί να υπάρχει πλήρης διασύνδεση (full connection) (σχήμα 2.25α), με την έννοια ότι κάθε νευρώνας ενός στρώματος θα είναι συνδεδεμένος με όλους τους νευρώνες του επόμενου στρώματος. Διαφορετικά θα υπάρχει μερική σύνδεση (partial connection) (σχήμα 2.25β).



Σχήμα 2.25: Τεχνητά νευρωνικά δίκτυα πλήρους και μερικής διασύνδεσης.

Ακόμα, όταν η ροή πληροφορίας είναι προς τη μία κατεύθυνση, χωρίς να υπάρχουν συνδέσεις νευρώνων ενός στρώματος με νευρώνες του προηγούμενου, ή και συνδέσεις νευρώνων που ανήκουν στο ίδιο στρώμα, τότε τα δίκτυα αυτά ονομάζονται δίκτυα πρόσθιας τροφοδότησης (feedforward) (σχήμα 2.26α). Διαφορετικά ονομάζονται δίκτυα ανατροφοδότησης (feedback/recurrent) (σχήμα 2.26β), με τους υπολογισμούς να χωρίζονται σε δύο στάδια: υπολογισμοί που αφορούν τα μεγέθη πρόσθιας τροφοδότησης και υπολογισμοί που αφορούν τα μεγέθη ανατροφοδότησης.



Σχήμα 2.26: Τεχνητά νευρωνικά δίκτυα πρόσθιας τροφοδότησης και ανατροφοδότησης.

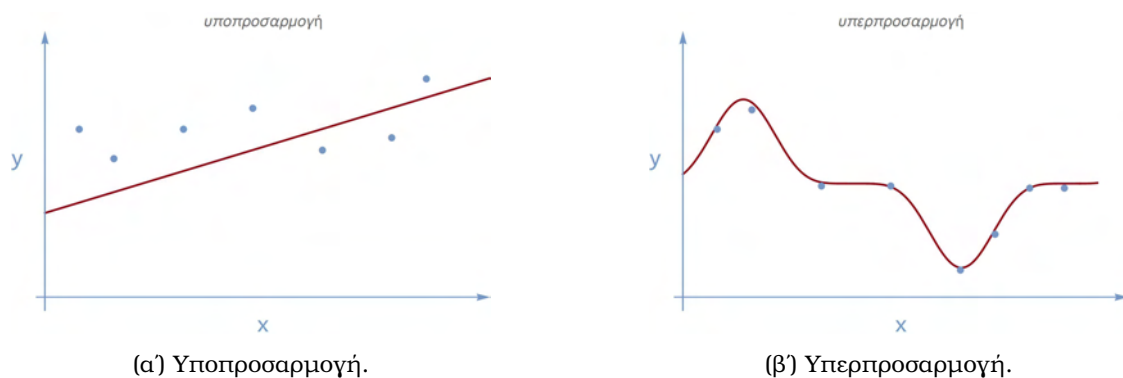
2.3.6 Ιδιότητες Τεχνητών Νευρωνικών Δικτύων

Οι βασικές ιδιότητες των Τεχνητών Νευρωνικών Δικτύων, οι οποίες τα καθιστούν ελκυστικά υπολογιστικά μοντέλα είναι οι εξής [Βλαχάβας06]:

Ικανότητα μάθησης μέσα από παραδείγματα :

Χωρίς να είναι τα μόνα με αυτή την ιδιότητα, τα Τεχνητά Νευρωνικά Δίκτυα μπορούν μέσα από την επεξεργασία των δεδομένων να αποκτήσουν γνώση, η οποία θα περιγράψει με το μοντέλο τις συσχετίσεις μεταξύ εισόδου και εξόδου. Η μάθηση γενικά μπορεί να φτάσει σε δύο ακραίες περιπτώσεις: στη μια να μην έχει γίνει σωστή εκπαίδευση (ατελής μάθηση, *underfitting*) και στην άλλη να έχει γίνει υπερβολική εκπαίδευση (υπερπροσαρμογή, *overfitting*), όπως φαίνεται στα ακόλουθα σχήματα.

Στην πρώτη περίπτωση το δίκτυο φαίνεται να αποτυγχάνει στη μάθηση των συσχετίσεων μεταξύ εισόδου και εξόδου στα δεδομένα εκπαίδευσης (σχήμα 2.27α)



Σχήμα 2.27: Βαθμός μάθησης.

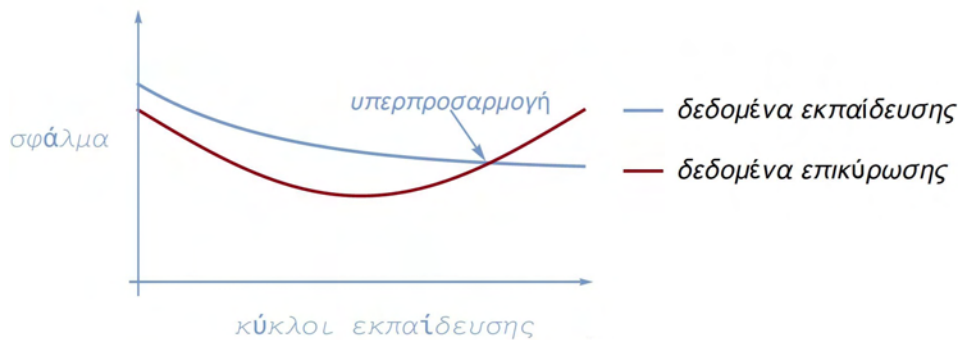
ενώ στη δεύτερη περίπτωση (σχήμα 2.27β) το δίκτυο έχει μοντελοποιήσει υπερβολικά τα δεδομένα εκπαίδευσης μαζί με το τυχόν θόρυβο που αυτά περιέχουν. Αν και δίνει σωστά αποτελέσματα σε αυτά, αδυνατεί να προβλέψει αποτελέσματα για ένα νέο σύνολο δεδομένων. Τα περίπλοκα δίκτυα οδηγούν σε αυτή την κατάσταση.

Σημαντικός θεωρείται ο αριθμός των δεδομένων με τον οποίο το δίκτυο θα εκπαιδευτεί αλλά και ο αριθμός των νευρώνων στα επίπεδα επεξεργασίας, ο οποίος προσαρμόζεται με τον αριθμό δεδομένων και την κατάσταση στην οποία βρίσκονται (π.χ θόρυβος), ώστε η εκπαίδευση να γίνεται με σωστό τρόπο.

Γενίκευση γνώσης:

Μια ακόμα χρήσιμη ιδιότητα που προκύπτει από τα παραπάνω είναι η ικανότητα της γενίκευσης της γνώσης. Πρόκειται για την ικανότητα του δικτύου να μπορεί να βγάλει σωστά συμπεράσματα σε δεδομένα που δεν έχει δει ξανά και ως εκ τούτου μια υπερπροσαρμοσμένη εκπαίδευση καθιστά δύσκολη τη γενίκευση [Agatonovic00]. Για να μπορεί να γίνει σωστά μια εκπαίδευση είναι απαραίτητος ο διαχωρισμός των δεδομένων σε 3 μέρη: στα δεδομένα εκπαίδευσης (*training dataset*), με τα οποία γίνεται η εκπαίδευση, δηλαδή η τροποποίηση των παραμέτρων, ώστε να ελαχιστοποιηθεί το σφάλμα σε αυτά, στο σύνολο επικύρωσης (*validation dataset*), ώστε να καθοριστεί η απόδοση του δικτύου σε παραδείγματα που δεν έχει δει και, τέλος, στο σύνολο εφαρμογής, ή δοκιμής, (*testing dataset*) για τον τελικό έλεγχο της απόδοσης του δικτύου.

Για παράδειγμα, η εξέλιξη του σφάλματος στο σύνολο εκπαίδευσης και επικύρωσης μπορεί να είναι αυτή που φαίνεται στο σχήμα 2.28.



Σχήμα 2.28: Ροή σφάλματος στα σύνολα επικύρωσης και εκπαίδευσης.

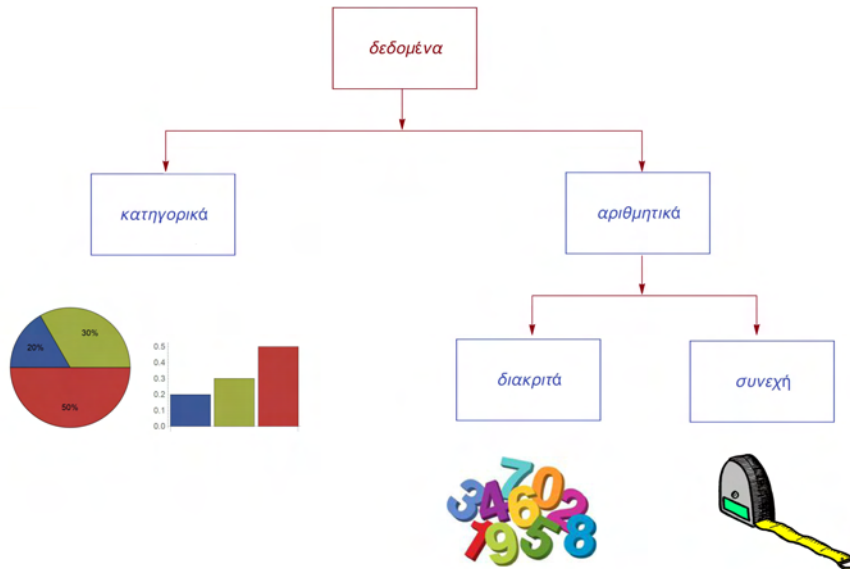
Όσο αυξάνεται ο αριθμός των κύκλων εκπαίδευσης, θα μειώνεται το σφάλμα στα δεδομένα εκπαίδευσης, αλλά υπάρχει περίπτωση να γίνει υπερπροσαρμογή, ώστε να παύει να υπάρχει καλή γενίκευση με το σφάλμα στα δεδομένα επικύρωσης να αυξάνεται, μειώνοντας έτσι την απόδοση του δικτύου. Για το λόγο αυτό, ως κριτήριο τερματισμού αφήνεται το ελάχιστο σφάλμα στα δεδομένα επικύρωσης. Ωστόσο, μπορεί πάλι να υπάρχει ο κίνδυνος για υπερπροσαρμογή, αυτή τη φορά στα δεδομένα επικύρωσης, οπότε αφότου ολοκληρωθεί η εκπαίδευση και επιλεγεί το μοντέλο, η τελική απόδοση του θα εκτιμηθεί από το σύνολο εφαρμογής.

Ανοχή σε σφάλματα :

Αφού τα Τεχνητά Νευρωνικά Δίκτυα βασίζονται στα βιολογικά νευρωνικά δίκτυα και στη λειτουργία του ανθρώπινου εγκεφάλου, ο οποίος θεωρείται ως κατανεμημένο και υπολογιστικό σύστημα, τότε η πληροφορία θα να είναι καταμερισμένη σε όλο το δίκτυο και θα αποθηκεύεται μέσω των συσχετίσεων που δημιουργούνται από τα δεδομένα εκπαίδευσης. Ένα «κακό» σημείο του δικτύου (είτε νευρώνας, είτε συνδέσεις) δεν είναι ικανό να το ορίσει ως ακατάλληλο. Έτσι, τα Τεχνητά Νευρωνικά Δίκτυα αποκαλούνται και ως κατανεμημένη μνήμη και ως μνήμη συσχέτισης με ανοχές σε δομικά σφάλματα.

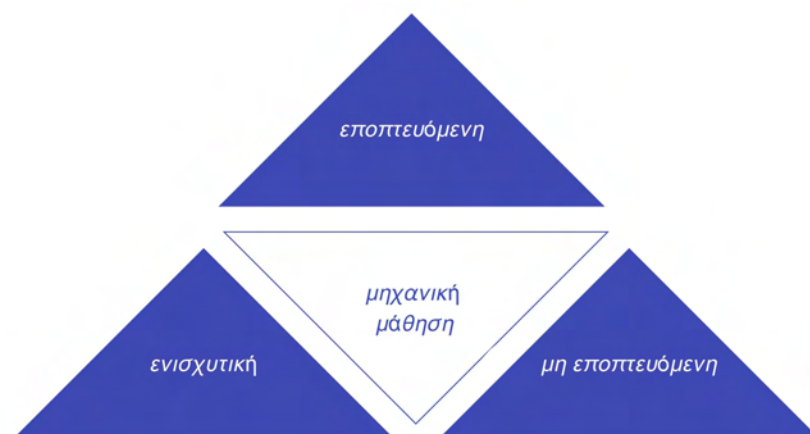
2.3.7 Είδη Μηχανικής Μάθησης

Το είδος της μάθησης εξαρτάται από τα διαθέσιμα δεδομένα εκπαίδευσης, δηλαδή από τη φύση του προβλήματος από την οποία προέρχονται [Βλαχάδας06]. Κάθε παράδειγμα σε αυτό το σύνολο δεδομένων αναπαρίσταται από το ίδιο σύνολο χαρακτηριστικών, οι οποίες μπορεί να είναι συνεχείς τιμές, δυαδικές τιμές ή κατηγορικές μεταβλητές, όπως φαίνεται στο σχήμα 2.29 [Kotsiantis04].



Σχήμα 2.29: Μορφές συνόλου δεδομένων.

Υπάρχουν διάφορες μορφές συνόλου δεδομένων [Lison12]. Μπορεί σε αυτό το σύνολο για κάθε είσοδο να υπάρχει και η αντίστοιχη γνωστή έξοδος, δηλαδή μια έξοδος με γνωστή ετικέτα (label). Από την άλλη, μπορεί και να μην υπάρχει γνωστή έξοδος. Το πρώτο είδος μάθησης καλείται εποπτευόμενη μάθηση (supervised learning), ενώ το δεύτερο είδος μη εποπτευόμενη μάθηση (unsupervised learning). Υπάρχει και μια ακόμα τεχνική μάθησης για την οποία θα γίνει μόνο αναφορά. Πρόκειται για την ενισχυτική μάθηση (reinforcement learning) στην οποία δεν είναι προσβάσιμη κάποια «σωστή» έξοδος, αλλά είναι μετρήσιμη η ποιότητα της εξόδου για την αντίστοιχη είσοδο. Στο σχήμα 2.30 απεικονίζονται τα είδη των μεθόδων της μηχανικής μάθησης.



Σχήμα 2.30: Είδη μηχανικής μάθησης.

Χαρακτηριστικά των μεθόδων μάθησης:

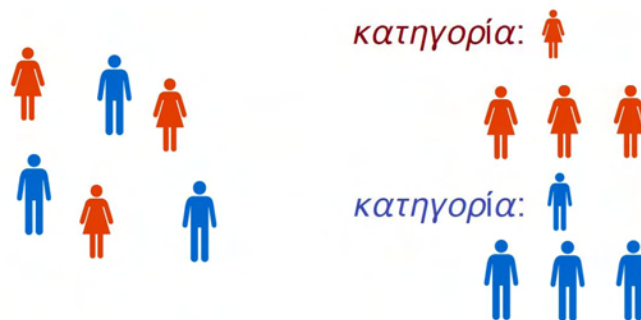
Μάθηση	Δεδομένα	Στόχος
εποπτευόμενη	με ετικέτα	πρόβλεψη αποτελέσματος
μη εποπτευόμενη	χωρίς ετικέτα	εύρεση «κρυφών» δομών

Στην εποπτευόμενη μάθηση το σύνολο των δεδομένων εκπαίδευσης είναι στη μορφή ζευγαριών (είσοδος, έξοδος). Εδώ στόχος είναι το σύστημα να μάθει μια συνάρτηση, η οποία για μια δεδομένη είσοδο να προβλέπει την τιμή της εξόδου. Έξοδος διαφορετική από την αναμενόμενη αφήνει ένα σφάλμα (error), το οποίο μαζί με έναν αλγόριθμο εκπαίδευσης επιτρέπει την κατάλληλη τροποποίηση των βαρών. Η ονομασία αυτής της μεθόδου μάθησης οφείλεται στην «παρουσία» κάποιου επιβλέποντος, αφού είναι γνωστή η τιμή που πρέπει να δώσει το σύστημα για μια δεδομένη είσοδο [Βλαχάδας06].

Σε μερικές περιπτώσεις η «σωστή» έξοδος μπορεί να δηλώνεται εκ των προτέρων όπως, παραδείγματος χάριν, στο φίλτρο ηλεκτρονικών μηνυμάτων με τον τονισμό των «επικίνδυνων» φράσεων-κλειδιά. Στο εν λόγω φίλτρο, μπορεί να υπάρχουν πολλές παράμετροι στη ζητούμενη συνάρτηση, οι οποίες θα πρέπει να ρυθμιστούν μέσω της εκπαίδευσης, ώστε τα αποτελέσματα που προκύπτουν να ταιριάζουν στα δεδομένα. Μπορεί με κάθε «επικίνδυνη» λέξη να συνδέεται ένας συντελεστής, κάποιο βάρος: όσο μεγαλύτερο είναι το βάρος των αντίστοιχων φράσεων, τόσο μεγαλύτερη θα είναι πιθανότητα το συγκεκριμένο μήνυμα να χαρακτηριστεί ως σπαμ [Lison12].

$$\text{πιθανότητα}_{\text{σπαμ}} = \text{βάρος}_1 \cdot \text{φράση}_1 + \text{βάρος}_2 \cdot \text{φράση}_2 + \dots$$

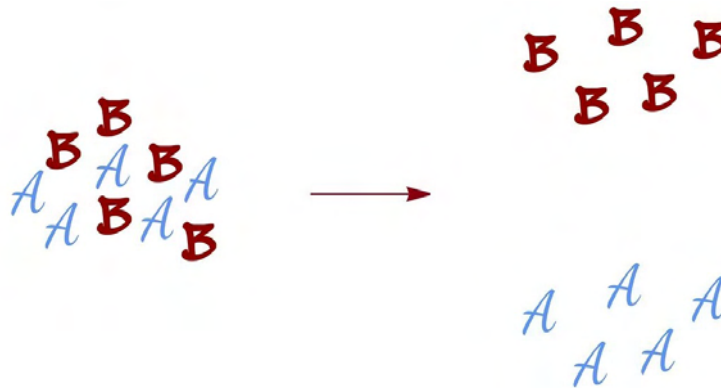
Ακόμα, το πρόβλημα κατηγοριοποίησης αντικειμένων σε γνωστές ομάδες ανήκει στην εποπτευόμενη μάθηση (σχήμα 2.31), αφού η μάθηση γίνεται από παραδείγματα για τα οποία είναι γνωστές οι κατηγορίες [Sebastiani02].



Σχήμα 2.31: Κατηγοριοποίηση σε γνωστές ομάδες - πρόβλημα εποπτευόμενης μάθησης.

Στη μη εποπτευόμενη μάθηση υπάρχει ένα σύνολο δεδομένων για το οποίο η είσοδος δεν συνοδεύεται από κάποια γνωστή έξοδο. Στόχος είναι η εύρεση τυχόν μοτίβων, συσχετίσεων που εμφανίζονται στα δεδομένα για τη δημιουργία προτύπων [Βλαχάδας06]. Υπάρχει κάποια σύνδεση μεταξύ των δεδομένων; Αν ναι, τότε, μέσα από τη μάθηση και την αυτοοργάνωση του δικτύου από τα δεδομένα γίνεται κάποια ομαδοποίηση σε άγνωστες αλλά χρήσιμες ομάδες ή κατηγορίες, με ορισμένα κοινά χαρακτηριστικά [Kotsiantis04].

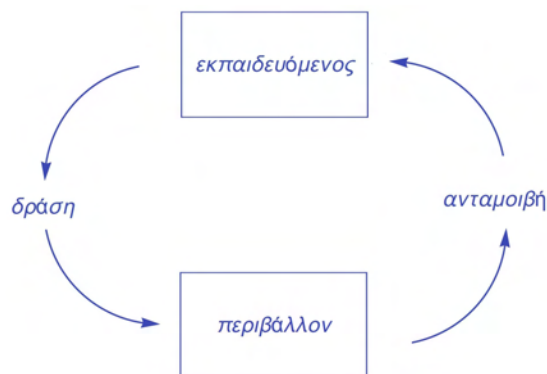
Για παράδειγμα, ο διαχωρισμός των γραμμάτων σε ομάδες βάσει ορισμένων κοινών χαρακτηριστικών (σχήμα 2.32) [Dougherty13]:



Σχήμα 2.32: Κατηγοριοποίηση σε άγνωστες ομάδες - πρόβλημα μη εποπτευόμενης μάθησης.

Σε κάποιες εφαρμογές όπως στις εικονικές αναπαραστάσεις, όπου τα βίντεο είναι πολύ περισσότερων διαστάσεων από ό,τι είναι οι απλές εικόνες, θα χρειάζονται πολλά δεδομένα των οποίων να είναι γνωστή η ετικέτα που, ίσως, και να αυξάνει το κόστος της μάθησης. Με τη μη εποπτευόμενη μάθηση μαθαίνει η μηχανή να βρίσκει χρήσιμες δομές, χωρικές και χρονικές, στα παραδείγματα [Srivastava15].

Στην ενισχυτική μάθηση δεν υπάρχει κάποια σωστή απάντηση για κάποια είσοδο. Ο εκπαιδευόμενος, συνδεδεμένος στο περιβάλλον του, μαθαίνει από τις εμπειρίες του και μέσω αυτών παίρνει αποφάσεις. Σε κάθε αλληλεπίδραση με το περιβάλλον δέχεται μια είσοδο και διαλέγει μια δράση για να δώσει μια έξοδο [Kaelbling96]. Η δράση του αυτή θα έχει συνέπειες: θα του επιστρέφεται κάποιο είδος ανταμοιβής (σχήμα 2.33).



Σχήμα 2.33: Εκπαιδευόμενος στην ενισχυτική μάθηση.

Δηλαδή μέσα από τις δράσεις που θα πάρει ο εκπαιδευόμενος θα αλλάξει η κατάσταση του περιβάλλοντος και η αλλαγή αυτή θα δίνεται στον εκπαιδευόμενο ως μια τιμή, η οποία καλείται σήμα ενίσχυσης. Οι αποφάσεις του εκπαιδευόμενου στοχεύουν, μέσα από διαδικασίες δοκιμής και σφάλματος στο περιβάλλον του, σε μια μακροπρόθεσμη αύξηση του αθροίσματος αυτών των σημάτων ενίσχυσης.

Αναλόγως με τη φύση του προβλήματος εφαρμόζεται και η κατάλληλη τεχνική μάθησης. Στην παρούσα εργασία γίνεται χρήση των δύο πρώτων μεθόδων μάθησης: της εποπτευόμενης και της μη εποπτευόμενης μάθησης.

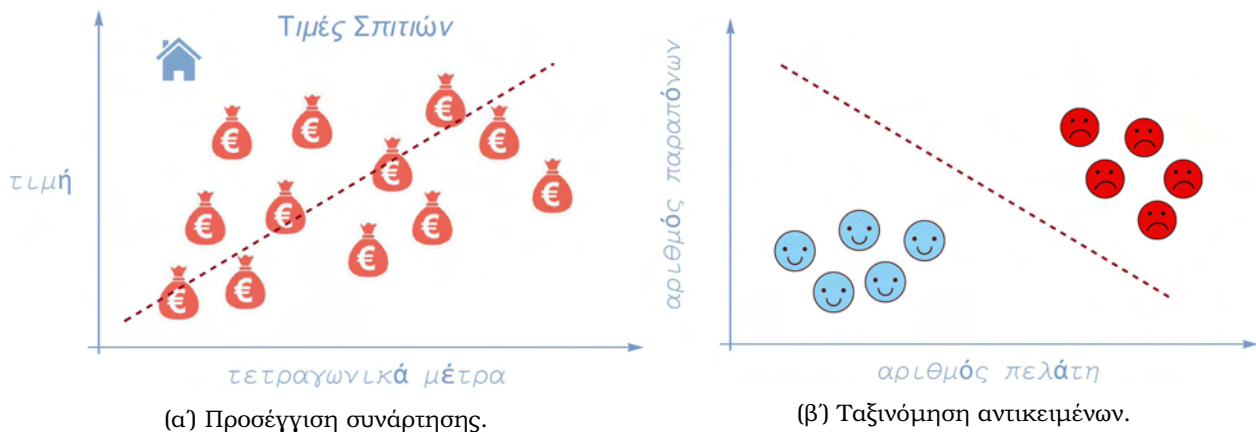
Ένα σημαντικό στάδιο που αφορά τις αρχικές διαδικασίες που πρέπει να ακολουθηθούν για την εφαρμογή αλγορίθμων μηχανικής μάθησης είναι αυτό της αναγνώρισης του προβλήματος [Dufourq17]. Πέρα από την επιλογή μεταξύ των ειδών μάθησης, υπάρχει και περαιτέρω διάκριση όπως, για παράδειγμα, στην περίπτωση της εποπτευόμενης μάθησης.

2.3.7.1 Εποπτευόμενη μάθηση

Ο εν λόγω επιλεγμένος αλγόριθμος εποπτευόμενης μάθησης θα αφορά παλινδρόμηση ή κατηγοριοποίηση; Και τα δύο αυτά προβλήματα σχετίζονται με τη δημιουργία ενός μοντέλου πρόβλεψης με τη διαφορά να εντοπίζεται στο σκοπό του μοντέλου αυτού. Στο πρώτο πρόβλημα, αυτό της παλινδρόμησης, η προβλεπόμενη από το μοντέλο τιμή είναι συνεχής, εν αντιθέσει με το δεύτερο πρόβλημα, όπου το μοντέλο κατηγοριοποίησης στοχεύει στο διαχωρισμό αντικειμένων σε γνωστές ομάδες.

Η πρόβλεψη της τιμής ενός ακινήτου μπορεί να θεωρηθεί ως ένα πρόβλημα παλινδρόμησης. Δοθέντος ενός συνόλου παραδειγμάτων που αφορούν κάποια χαρακτηριστικά όπως τα τετραγωνικά μέτρα και την τιμή πώλησης για κάθε παράδειγμα, μπορεί να βρεθεί μια συνάρτηση, η οποία να προβλέπει τις τιμές των ακινήτων βάσει των χαρακτηριστικών τους (σχήμα 2.34α).

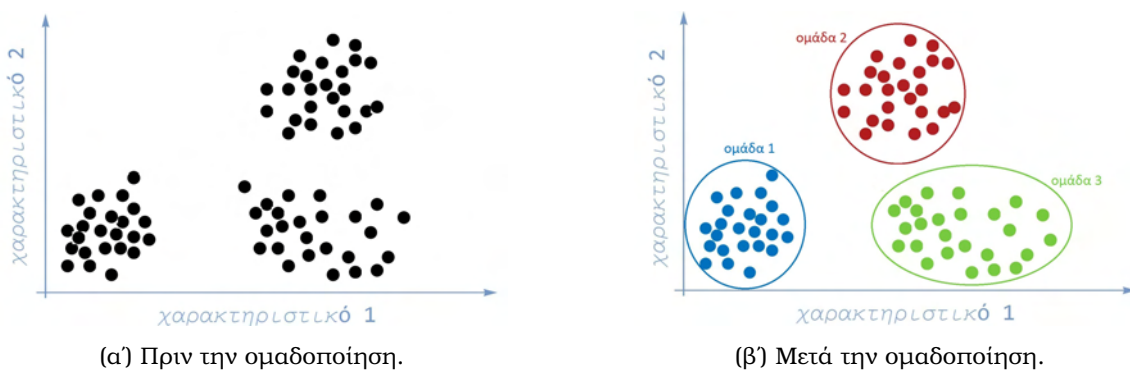
Από την άλλη, η ταξινόμηση πελατών σε ευχαριστημένους και μη, στην ουρά αναμονής κάποιας υπηρεσίας, αποτελεί ένα παράδειγμα προβλήματος κατηγοριοποίησης (σχήμα 2.34β).



Σχήμα 2.34: Προβλήματα εποπτευόμενης μάθησης.

2.3.7.2 Μη εποπτευόμενη μάθηση

Η μη εποπτευόμενη μάθηση μπορεί να χρησιμοποιηθεί στην εύρεση ομάδων, οι οποίες είναι άγνωστες εκ των προτέρων και χαρακτηρίζονται από κοινά στοιχεία, όπως φαίνεται στο σχήμα 2.35.



Σχήμα 2.35: Πρόβλημα μη εποπτευόμενης μάθησης.

Εδώ ο διαχωρισμός των αντικειμένων έγινε σε ομάδες, οι οποίες χαρακτηρίζονται από κάποια κοινά στοιχεία. Η πρώτη ομάδα αποτελείται από αντικείμενα τα οποία έχουν χαμηλές τιμές των χαρακτηριστικών 1 και 2 σε αντίθεση με την τρίτη ομάδα, η οποία χαρακτηρίζεται από μεγαλύτερες τιμών του 1ου χαρακτηριστικού. Τέλος, η δεύτερη ομάδα έχει αντικείμενα με ενδιάμεσες τιμές του 1ου χαρακτηριστικού και μεγαλύτερες τιμές του 2ου χαρακτηριστικού από τις άλλες δύο ομάδες.

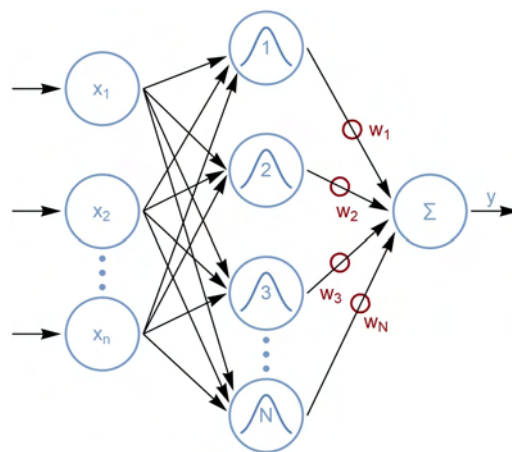
Κεφάλαιο 3

Νευρωνικά Δίκτυα Ακτινικής Βάσης (Radial Basis Function Networks)

Μια σημαντική παραλλαγή των Τεχνητών Νευρωνικών Δικτύων που προτάθηκε από τους Broomhead και Lowe [Broomhead88], αλλά και ανεξάρτητα από άλλους ερευνητές στο τέλος της δεκαετίας του 80', είναι αυτή των Νευρωνικών Δικτύων Ακτινικής Βάσης (Radial Basis Function Networks, RBFNs). Πρόκειται για ένα δίκτυο πρόσθιας τροφοδότησης με τους νευρώνες του κρυφού επιπέδου να διαχειρίζονται μια ακτινική, μη γραμμική συνάρτηση βάσης (RBF) στην οποία και οφείλεται η προέλευση της ονομασίας αυτών των δικτύων [Engelbrecht07], [Bors], [Tao93].

3.1 Δομή ενός δικτύου RBF

Η αρχιτεκτονική ενός δικτύου RBF είναι η αυτή που απεικονίζεται στο σχήμα 3.1 [Alexandridis13].



Σχήμα 3.1: Τυπική αρχιτεκτονική ενός δικτύου RBF.

Παρόμοια με αυτή των απλών δικτύων πρόσθιας τροφοδότησης, αποτελείται από ένα επίπεδο εισόδου, από ένα κρυφό επίπεδο και από ένα επίπεδο εξόδου. Και εδώ το πρώτο επίπεδο διαδίδει απλώς τα σήματα εισόδου στο δίκτυο, δηλαδή τις n κανονικοποιημένες μεταβλητές εισόδου στους N κόμβους του κρυφού επιπέδου. Στο τελευταίο γίνεται η επεξεργασία των σημάτων, με τους κρυφούς νευρώνες να διαχειρίζονται ακτινικές συναρτήσεις βάσης ή συναρτήσεις πυρήνα, ϕ , όπως αυτές αποκαλούνται. Κάθε κρυφός νευρώνας περιγράφεται από ένα κέντρο, το οποίο έχει τόσες διαστάσεις όσες είναι και ο αριθμός των μεταβλητών εισόδου, καθιστώντας έτσι τη σύνδεση μεταξύ εισόδου και κρυφού επιπέδου μη γραμμική. Ωστόσο, η σύνδεση μεταξύ του κρυφού επιπέδου και του επιπέδου εξόδου είναι γραμμική.

με την έξοδο του δικτύου να είναι απλώς ο γραμμικός συνδυασμός των εξόδων των κρυφών επιπέδων στον οποίο τα συναπτικά βάρη αποτελούν τους γραμμικούς συντελεστές.

Τα δίκτυα RBF χαρακτηρίζονται ως δίκτυα τοπικής προσέγγισης καθώς η έξοδος του δικτύου προσδιορίζεται από συγκεκριμένους κρυφούς νευρώνες, οι οποίοι δρουν τοπικά, σε συγκεκριμένα πεδία (receptive fields), ενώ τα δίκτυα, για παράδειγμα, πρόσθιας τροφοδότησης πολλαπλών επιπέδων χαρακτηρίζονται από «ολική» (global) δράση καθώς η έξοδος του δικτύου καθορίζεται από όλους τους νευρώνες σε κάποιο βαθμό [Xie11].

Ο τοπικός χαρακτήρας της ισχύος των συναρτήσεων βάσης οδηγεί στην ανάγκη υπολογισμού της απόστασης μιας μονάδας RBF με ένα παράδειγμα εισόδου. Ο υπολογισμός αφορά την Ευκλείδεια απόσταση ενός κέντρου $\mathbf{c} \in \mathbb{R}^n$ από κάποιο παράδειγμα εκπαίδευσης $\mathbf{x} \in \mathbb{R}^n$ [Alexandridis13]:

$$r = \|\mathbf{x} - \mathbf{c}\|_2 \quad (3.1)$$

Τα κυριότερα σημεία ενός δικτύου RBF είναι τα εξής:

- οι νευρώνες του κρυφού επιπέδου διαχειρίζονται μια συνάρτηση ϕ ακτινικής συμμετρίας. Η έξοδος ενός κρυφού νευρώνα j υπολογίζεται ως εξής:

$$h_{j,i}(\mathbf{x}_i) = \phi_j(\|\mathbf{x}_i - \mathbf{c}_j\|_2) \quad (3.2)$$

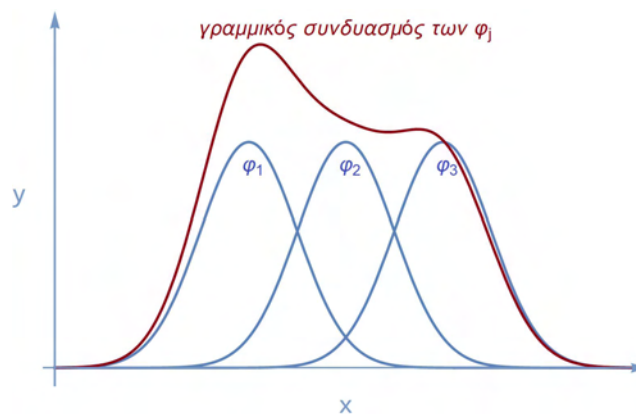
για κάθε ένα από τα παραδείγματα εισόδου $i = 1, 2, \dots, m$

- κάθε νευρώνας j του κρυφού επιπέδου χαρακτηρίζεται από ένα κέντρο, $\mathbf{c}_j \in \mathbb{R}^n$, το οποίο αποτελεί το κέντρο της συνάρτησης βάσης ενώ μερικές συναρτήσεις χαρακτηρίζονται και από μια τιμή εύρους, $\sigma_j \in \mathbb{R}$

- η έξοδος ενός δικτύου RBF για ένα συγκεκριμένο παράδειγμα εισόδου, i , υπολογίζεται ως εξής:

$$y_i(\mathbf{x}_i) = \sum_{j=1}^N w_j \phi_j(\mathbf{x}_i, \mathbf{c}_j) \quad (3.3)$$

Η απλή αρχιτεκτονική των δικτύων RBF τα καθιστά ελκυστικά στο σχεδιασμό και τη μοντελοποίηση μη γραμμικών συστημάτων ενώ η εκπαίδευση τους συνίσταται στην εύρεση μιας πολυδιάστατης επιφάνειας, προερχόμενη από το σταθμισμένο άθροισμα απλούστερων επιφανειών που ορίζονται από τις συναρτήσεις βάσης των κρυφών νευρώνων (σχήμα 3.2), η οποία θα προσεγγίζει όσο το δυνατόν καλύτερα τα δεδομένα εκπαίδευσης [Alexandridis13].



Σχήμα 3.2: Συνολική επιφάνεια ορισμένη από τις συναρτήσεις βάσης.

Αυτή η δομή των δικτύων RBF χωρίζει το σχεδιασμό τους σε 3 στάδια: στην εύρεση του μεγέθους του δικτύου, στην ανάθεση αρχικών τιμών στις παραμέτρους και, τέλος, στην εκπαίδευση του δικτύου [Xie11].

3.2 Συναρτήσεις βάσης ακτινικής συμμετρίας

Πρόκειται για αυστηρώς θετικές με ακτινική συμμετρία συναρτήσεις, οι οποίες χαρακτηρίζονται από ένα διάνυσμα κέντρου, $\mathbf{c}_j \in \mathbb{R}^n$, και κάποιες από αυτές από μια τιμή εύρους, $\sigma_j \in \mathbb{R}$. Συνήθως, οι συναρτήσεις αυτές έχουν τοπικό χαρακτήρα, δηλαδή παρουσιάζουν ένα μοναδικό μέγιστο στο κέντρο με την τιμή να μειώνεται ραγδαία από αυτό. Αυτό που υποδηλώνει η συνάρτηση βάσης είναι η εγγύτητα που παρουσιάζει ένα ορισμένο διάνυσμα εισόδου με το κέντρο της και, επομένως, τη δραστηριότητα του αντίστοιχου κόμβου. Η τιμή του εύρους της συναρτήσεως του νευρώνα j , σ_j , ορίζει το εύρος του δραστικού πεδίου της συνάρτησεως στο χώρο εισόδου για αυτόν το νευρώνα [Engelbrecht07].

Προηγουμένως ορίστηκε μέσω της Ευκλείδειας απόστασης η εγγύτητα ενός παραδείγματος εκπαίδευσης με κάποιο κέντρο ενός κρυφού νευρώνα:

$$\|\mathbf{x}(k) - \mathbf{c}(l)\|_2 = \sqrt{\sum_{i=1}^n (x_i(k) - c_i(l))^2}, \quad k = 1, 2, \dots, m \quad (3.4)$$

όπου m ο συνολικός αριθμός δεδομένων και n ο αριθμός των μεταβλητών. Το διάνυσμα εισόδου $k = 1, 2, \dots, m$ είναι:

$$\mathbf{x}^T(k) = [x_1(k), x_2(k), \dots, x_n(k)] \quad (3.5)$$

ενώ το διάνυσμα του κέντρου $l = 1, 2, \dots, N$:

$$\mathbf{c}^T(l) = [c_1(l), c_2(l), \dots, c_n(l)] \quad (3.6)$$

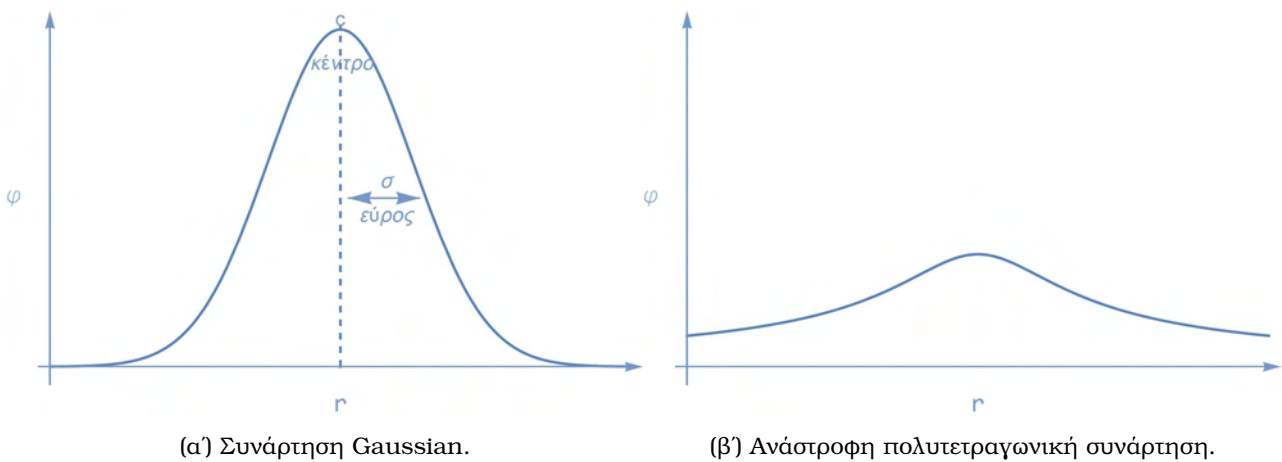
Η παραπάνω απόσταση ονομάζεται δραστηριότητα του κόμβου l [Alexandridis13] και θα συμβολίζεται ως r_l .

Τύποι συναρτήσεων βάσης

Έχουν προταθεί διάφορες συναρτήσεις βάσης [Engelbrecht07], [Bishop95] με την πιο κοινή να είναι η συνάρτηση Gaussian (σχήμα 3.3α):

$$\phi(r) = e^{-\frac{r^2}{2\sigma^2}} \quad (3.7)$$

για την οποία ισχύει ότι $\phi \rightarrow 0$ καθώς $|r| \rightarrow \infty$



Σχήμα 3.3: Συναρτήσεις βάσης τοπικού χαρακτήρα.

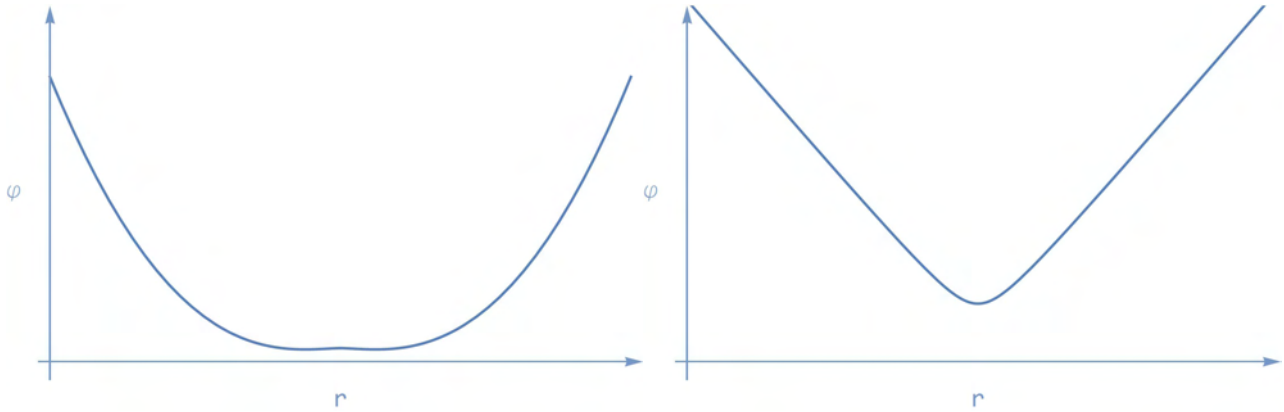
ενώ και η ανάστροφη πολυτετραγωνική (inverse multiquadric) συνάρτηση (σχήμα 3.3β) διατηρεί την ίδια ιδιότητα:

$$\phi(r) = \frac{1}{\sqrt{r^2 + \sigma^2}} \quad (3.8)$$

Πέρα από τις συναρτήσεις με τοπικό χαρακτήρα χρησιμοποιούνται και αυτές για τις οποίες ισχύει ότι $\phi \rightarrow \infty$ καθώς $|r| \rightarrow \infty$.

Τέτοιες, για παράδειγμα, είναι η συνάρτηση thin-plate-spline (σχήμα 3.4α):

$$\phi(r) = r^2 \log r \quad (3.9)$$



(α) Συνάρτηση thin-plate-spline.

(β) Πολυτετραγωνική συνάρτηση.

Σχήμα 3.4: Συναρτήσεις βάσης μη τοπικού χαρακτήρα.

και η πολυτετραγωνική (multiquadratic) συνάρτηση (σχήμα 3.4β):

$$\phi(r) = \sqrt{r^2 + \sigma^2} \quad (3.10)$$

Επομένως, οι σημαντικές παράμετροι που επηρεάζουν την απόδοση ενός δικτύου RBF είναι οι εξής:

- ο αριθμός των συναρτήσεων βάσης, δηλαδή ο αριθμός των κρυφών νευρώνων που θα χρησιμοποιηθούν
- η τοποθεσία των RBF στο χώρο εισόδου, και
- τα εύρη για κάποιες συναρτήσεις

Όσες περισσότερες μονάδες RBF υπάρχουν, τόσο πιο καλή θα είναι η προσέγγιση της συνάρτησης. Ωστόσο, παραπάνω μονάδες που δε χρειάζονται αυξάνουν τις υπολογιστικές απαιτήσεις. Η τοποθεσία των συναρτήσεων βάσης στο χώρο εισόδου καθορίζεται από τις θέσεις των κέντρων τους. Οι συναρτήσεις βάσης πρέπει να κατανεμηθούν ομοιόμορφα στο χώρο εισόδου, ώστε να καλύπτουν επαρκώς τα δεδομένα. Το εύρος ορισμένων συναρτήσεων βάσης ορίζει το πεδίο ισχύος αυτής και, επομένως, όσο μεγαλύτερο είναι, τόσο μεγαλύτερο θα είναι το αντίστοιχο πεδίο που αυτή θα καλύπτει [Engelbrecht07].

Η εκπαίδευση ενός δικτύου RBF, τελικά, συνίσταται στον υπολογισμό των εξής παραμέτρων [Alexandridis13]:

- αριθμό μονάδων RBF
- συντεταγμένες κέντρων (ή και εύρη)
- βάρη σύνδεσης μεταξύ κρυφού επιπέδου και επιπέδου εξόδου

Όλες αυτές οι παράμετροι είναι σημαντικές για ένα δίκτυο RBF και ο τρόπος καθορισμού τους θα οδηγήσει σε διαφορετικά δίκτυα. Αναλόγως, λοιπόν, με την εκπαίδευση, την αρχιτεκτονική και τον τύπο των συναρτήσεων βάσης που χρησιμοποιούνται σε ένα δίκτυο RBF, θα διαφέρει και η αντίστοιχη απόδοση και χαρακτηριστικά του [Tao93].

3.3 Πλεονεκτήματα δικτύων RBF

Ορισμένα γενικά πλεονεκτήματα των δικτύων RBF: η καλή γενίκευση, καθώς ένα δίκτυο RBF αποδίδει καλά για δεδομένα που δεν έχει δει, η μεγάλη ανοχή σε θόρυβο που περιέχεται στα δεδομένα εισόδου, η οποία ενισχύει τη σταθερότητα αυτών των συστημάτων και η διατήρηση της ιδιότητας της ολικής προσέγγισης όπως τα άλλα δίκτυα πρόσθιας τροφοδότησης [Yu11]. Επιπλέον, η απλή και σταθερή αρχιτεκτονική τους αποτελούμενη από 3 επίπεδα καθιστά εύκολο το σχεδιασμό και την εκπαίδευση με τους αντίστοιχους αλγορίθμους εκπαίδευσης να συνοδεύονται από γρηγορότερους χρόνους και καλύτερη συνολική απόδοση εν συγκρίσει με τα δίκτυα πρόσθιας τροφοδότησης πολλαπλών επιπέδων [Alexandridis13].

3.4 Εκπαίδευση ενός δικτύου RBF

Οι παράμετροι ενός δικτύου RBF καθορίζουν άμεσα την ακρίβεια και την απόδοση του και, επομένως, κρίνεται απαραίτητη η σωστή επιλογή αυτών. Η επιλογή των κατάλληλων τιμών των παραμέτρων γίνεται μέσω της εκπαίδευσης του δικτύου. Η εκπαίδευση συνίσταται στις μεθόδους που χρησιμοποιούνται για την εύρεση των βέλτιστων τιμών των παραμέτρων. Έχουν προταθεί διάφορες μέθοδοι εκπαίδευσης, με τη διαφορά τους να έγκειται κυρίως στον αριθμό των παραμέτρων που αυτές διαχειρίζονται. Η μια κατηγορία αφορά τη μάθηση μόνο των βαρών υποθέτοντας σταθερές τις παραμέτρους των συναρτήσεων βάσης (κέντρα ή και εύρη). Η άλλη αφήνει ελεύθερες όλες τις παραμέτρους να μεταβληθούν και να επιλεγθούν μέσω των αντίστοιχων κατάλληλων αλγορίθμων. Υπάρχει και μια επιπλέον κατηγορία, η οποία μπορεί να θεωρηθεί συνδυασμός των δύο προηγούμενων. Σε αυτήν η εκπαίδευση πραγματοποιείται σε δύο στάδια: το πρώτο στάδιο αφορά την επιλογή των κέντρων με μια μέθοδο μη εποπτευόμενης μάθησης και το δεύτερο την επιλογή των βαρών μεταξύ κρυφού επιπέδου και επιπέδου εξόδου με μια μέθοδο εποπτευόμενης μάθησης, όπως είναι, για παράδειγμα, μια μέθοδος βελτιστοποίησης [Engelbrecht07].

3.4.1 Μορφοποίηση του προβλήματος της εκπαίδευσης ενός δικτύου RBF

Σκοπός ενός δικτύου RBF είναι η προσέγγιση μιας συνάρτησης μέσω του συνδυασμού μεμονωμένων συναρτήσεων, ώστε για ένα διάνυσμα εισόδου να προκύπτει η κατάλληλη έξοδος [Alexandridis13].

Συγκεκριμένα, το πρόβλημα είναι το εξής:

Δοθέντος ενός συνόλου δεδομένων εκπαίδευσης με m παραδείγματα, όπου το κάθε παράδειγμα αποτελείται από n κανονικοποιημένες μεταβλητές και το κρυφό στρώμα από N κρυφούς νευρώνες, ζητείται μια συνάρτηση $f : \mathbb{R}^n \rightarrow \mathbb{R}$, η οποία για συγκεκριμένη είσοδο να δίνει συγκεκριμένη έξοδο. Το σύνολο δεδομένων εκπαίδευσης είναι:

$$(\mathbf{x}_i, y_i), \quad i = 1, \dots, m \quad (3.11)$$

όπου $\mathbf{x}_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$

Κανονικοποίηση δεδομένων

Η κανονικοποίηση των μεταβλητών εισόδου και εξόδου είναι απαραίτητη πριν ξεκινήσει το στάδιο της εκπαίδευσης καθώς οι τιμές τους μπορεί να διαφέρουν σημαντικά, επιφέροντας και κάποια επακόλουθα προβλήματα στις διάφορες αριθμητικές μεθόδους που χρησιμοποιούνται. Η κανονικοποίηση αφορά τη μετατροπή των τιμών όλων των μεταβλητών των δεδομένων εκπαίδευσης σε ένα συγκεκριμένο εύρος. Στη συγκεκριμένη εργασία επιλέγεται το $[a, b] = [-1, 1]$. Οποιοδήποτε δεδομένο σε μια μεταβλητή, είτε εισόδου είτε εξόδου, θα πρέπει να τοποθετηθεί σε τέτοια τιμή, ώστε να ισχύει:

$$\frac{x_{-1,1} - a}{b - a} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.12)$$

ή

$$x_{-1,1} = a + (b - a) \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.13)$$

όπου $x_{-1,1}$ η κανονικοποιημένη τιμή της μεταβλητής x ενώ x_{max} και x_{min} είναι η μέγιστη και η ελάχιστη τιμή της μεταβλητής, αντίστοιχα.

Η κανονικοποίηση γίνεται με τα δεδομένα εκπαίδευσης, με τις μεταβλητές στα δεδομένα επικύρωσης και εφαρμογής να υπόκεινται στον ίδιο μετασχηματισμό. Δηλαδή τα όρια των μεταβλητών x_{max}, x_{min} θα είναι αυτά που θα προκύψουν από την κανονικοποίηση των μεταβλητών των δεδομένων εκπαίδευσης.

Ωστόσο, το δίκτυο θα εκπαιδευτεί βάσει των κανονικοποιημένων μεταβλητών, δηλαδή οι συσχετίσεις που θα μάθει θα αφορούν τα κανονικοποιημένα δεδομένα x, y . Άρα, αντίστοιχα, θα πρέπει να γίνει ο αντίστροφος μετασχηματισμός, η λεγόμενη αποκανονικοποίηση των αποτελεσμάτων, με τον εξής τρόπο:

$$y = y_{min} + (y_{-1,1} - a) \frac{x_{max} - x_{min}}{b - a} \quad (3.14)$$

Η έξοδος του δικτύου για το κάθε παράδειγμα εκπαίδευσης θα δίνεται από

$$f(\mathbf{x}_i, \mathbf{c}, \mathbf{w}) = \sum_{j=1}^N w_j \phi(\mathbf{x}_i, \mathbf{c}_j), \quad i = 1, 2, \dots, m \quad (3.15)$$

όπου $w_j \in \mathbb{R}$, $\mathbf{c}_j \in \mathbb{R}^n$, τα βάρη σύνδεσης κρυφού επιπέδου με το επίπεδο εξόδου και τα κέντρα των μονάδων RBF, αντίστοιχα.

Χρήσιμη είναι η αναπαράσταση των παραπάνω σε μητρική μορφή, η οποία ακολουθείται στη συνέχεια.

Τα διανύσματα εισόδου και τα κέντρα του δικτύου δίνονται παρακάτω:

$$x_{i,j} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{bmatrix}, \quad c_{i,j} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N1} & c_{N2} & c_{N3} & \cdots & c_{Nn} \end{bmatrix} \quad (3.16)$$

Για το κάθε διάνυσμα εισόδου υπάρχει και μια έξοδος:

$$[\mathbf{y}]^T = [y_1 \quad y_2 \quad \cdots \quad y_m] \quad (3.17)$$

Τα βάρη του δικτύου:

$$[\mathbf{w}]^T = [w_1 \quad w_2 \quad \cdots \quad w_N] \quad (3.18)$$

Για κάθε διάνυσμα εισόδου θα υπολογίζεται ο βαθμός εγγύτητας με το κέντρο του κάθε νευρώνα στο κρυφό επίπεδο, δίνοντας ένα μητρώο $\Phi_{m \times N}$:

$$\Phi = \begin{bmatrix} \phi(\mathbf{x}_1, \mathbf{c}_1) & \phi(\mathbf{x}_1, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_1, \mathbf{c}_N) \\ \phi(\mathbf{x}_2, \mathbf{c}_1) & \phi(\mathbf{x}_2, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_2, \mathbf{c}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\mathbf{x}_m, \mathbf{c}_1) & \phi(\mathbf{x}_m, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_m, \mathbf{c}_N) \end{bmatrix} \quad (3.19)$$

Για ένα συγκεκριμένο παράδειγμα εκπαίδευσης i , το υπόλοιπο (residual) ορίζεται ως:

$$r_i(\mathbf{c}, \mathbf{w}) = y_i - f(\mathbf{x}_i, \mathbf{c}, \mathbf{w}) = y_i - \sum_{j=1}^N w_j \phi(\mathbf{x}_i, \mathbf{c}_j), \quad i = 1, 2, \dots, m \quad (3.20)$$

Στόχος είναι η ελαχιστοποίηση των υπολοίπων, οπότε το παραπάνω πρόβλημα μπορεί να μορφοποιηθεί ως ένα πρόβλημα ελαχίστων τετραγώνων

$$\min_{\mathbf{c}, \mathbf{w}} e(\mathbf{c}, \mathbf{w}) = \min_{\mathbf{c}, \mathbf{w}} \frac{1}{2} \sum_{i=1}^m r_i(\mathbf{c}, \mathbf{w})^2 \quad (3.21)$$

ή σε μητρική μορφή:

$$\min_{\mathbf{c}, \mathbf{w}} e(\mathbf{c}, \mathbf{w}) = \min_{\mathbf{c}, \mathbf{w}} \frac{1}{2} \|\mathbf{y} - \Phi(\mathbf{c}, \mathbf{w})\mathbf{w}\|_2^2 \quad (3.22)$$

3.4.2 Υπολογισμός των συναπτικών βαρών ενός δικτύου RBF

Αν με κάποιο τρόπο έχουν οριστεί τα κέντρα των συναρτήσεων βάσης (ή και τα εύρη), τότε η εκπαίδευση του δικτύου συνίσταται στην εύρεση του διανύσματος των βαρών, \mathbf{w} , το οποίο προκύπτει με την ελαχιστοποίηση των υπολοίπων [Alexandridis13]:

$$\min_{\mathbf{w}} e(\mathbf{w}) = \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{y} - \Phi(\mathbf{c}, \mathbf{w})\mathbf{w}\|_2^2 \quad (3.23)$$

Πρόκειται για ένα απλό γραμμικό πρόβλημα ελαχίστων τετραγώνων με τη λύση του να προκύπτει ως εξής [Bindel12]:

Αν

$$\|\mathbf{r}\|_2^2 = \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 = (\mathbf{b} - \mathbf{A}\mathbf{x})(\mathbf{b} - \mathbf{A}\mathbf{x})^T \quad (3.24)$$

με $\mathbf{A}_{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$, τότε σε ένα συγκεκριμένο \mathbf{x} , η κατευθυντική παράγωγος (directional derivative) σε μια διεύθυνση $\delta\mathbf{x} \in \mathbb{R}^n$ θα είναι:

$$\nabla_{\mathbf{x}} \|\mathbf{r}\|_2^2 \cdot \delta\mathbf{x} = -2\delta\mathbf{x}^T \mathbf{A}^T (\mathbf{b} - \mathbf{A}\mathbf{x}) = -2\delta\mathbf{x}^T (\mathbf{A}^T \mathbf{b} - \mathbf{A}^T \mathbf{A}\mathbf{x}) \quad (3.25)$$

Το ελάχιστο συμβαίνει όταν η παραπάνω ποσότητα ισούται με το μηδέν για οποιοδήποτε $\delta\mathbf{x}$.

Δηλαδή,

$$2\delta\mathbf{x}^T (\mathbf{A}^T \mathbf{b} - \mathbf{A}^T \mathbf{A}\mathbf{x}) = 0 \quad (3.26)$$

ή αλλιώς

$$\mathbf{A}^T \mathbf{A}\mathbf{x} = \mathbf{A}^T \mathbf{b} \quad (3.27)$$

Οι παραπάνω εξισώσεις ονομάζονται «κανονικές» εξισώσεις (normal equations).

Επομένως, το διάνυσμα των βαρών θα προέρχεται από τη λύση των γραμμικών εξισώσεων:

$$\Phi^T \Phi \mathbf{w} = \Phi^T \mathbf{y} \quad (3.28)$$

ή για αντιστρέψιμο $\Phi^T \Phi$

$$\mathbf{w} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{y} \quad (3.29)$$

όπου το διάνυσμα \mathbf{y} είναι το διάνυσμα με τις τιμές-στόχο, με το μητρώο Φ να περιέχει τις εξόδους των κρυφών νευρώνων για όλα τα παραδείγματα εκπαίδευσης.

Ο πίνακας

$$\Phi^+ = (\Phi^T \Phi)^{-1} \Phi^T \quad (3.30)$$

ονομάζεται Moore-Penrose ψευδοαντίστροφος του Φ .

Συνεπώς, έχοντας τα κέντρα, ή σε μερικές συναρτήσεις βάσης και τα εύρη, τότε η εκπαίδευση του δικτύου συνίσταται απλώς στην επίλυση αυτών των γραμμικών εξισώσεων (3.1). Τα κέντρα ενός δικτύου RBF μπορούν να προκύψουν με διάφορους τρόπους, κάποιους από τους οποίους παρουσιάζονται στη συνέχεια [Alexandridis13].

3.4.3 Υπολογισμός των κέντρων ενός δικτύου RBF

Ο αρχικός τρόπος επιλογής των κέντρων αφορούσε την ανάθεση σε κάθε ένα από τα παραδείγματα του συνόλου εκπαίδευσης και ενός κέντρου RBF. Όμως, με τον τρόπο αυτό κανείς οδηγείται σε τεράστια δίκτυα με ανάλογο αριθμό των παραμέτρων που πρέπει να ρυθμιστούν, αυξάνοντας έτσι τις απαιτήσεις σε μνήμη, κυρίως σε μεγάλα σύνολα δεδομένων [Alexandridis13]. Πέρα από τις υψηλές απαιτήσεις μνήμης, η ανάθεση κέντρων σε κάθε ένα από τα παραδείγματα εκπαίδευσης μπορεί να οδηγήσει σε υπερπροσαρμογή και, επομένως, σε όχι καλή γενίκευση [Bors], [Tao93]. Για το λόγο αυτό, η επιλογή του αριθμού των κέντρων αφορά έναν αριθμό μικρότερο των παραδειγμάτων και η επιλογή των θέσεων τους πρέπει να γίνεται έτσι, ώστε όλα τα δεδομένα στο χώρο εισόδου να καλύπτονται επαρκώς από αυτά. Πέρα από την τυχαία επιλογή των θέσεων των κέντρων από τα διαθέσιμα δεδομένα, χρησιμοποιείται και ένας αυτοοργανωμένος τρόπος ομαδοποίησης ο λεγόμενος k -μέσοι ή k -means, με τα κέντρα αυτών των ομάδων να αποτελούν τα κέντρα των συναρτήσεων βάσης του δικτύου [Alexandridis13].

3.4.3.1 Αλγόριθμος k -means

Ο όρος αυτός ορίστηκε από τον Macqueen [Macqueen1967], ενώ φαίνεται να είχε χρησιμοποιηθεί νωρίτερα από τον Lloyd το 1957, με την εργασία του να δημοσιεύεται το 1982 [Lloyd82]. Η ομαδοποίηση με k -means χρησιμοποιείται για να βρεθεί ένα σύνολο κέντρων, k , για το οποίο ελαχιστοποιείται το άθροισμα των αποστάσεων όλων των δεδομένων εκπαίδευσης από τα εγγύτερα σε αυτά κέντρα [Moody89]. Δηλαδή απαιτείται η ελαχιστοποίηση του:

$$E = \sum_j \sum_i \|x_i - x_j\|_2^2 \quad (3.31)$$

όπου $j = 1, \dots, N$, τα εγγύτερα κέντρα στα δεδομένα εισόδου $i = 1, \dots, m$.

Η ομαδοποίηση αφορά το διαμερισμό ενός συνόλου δεδομένων σε μη αλληλοκαπτόμενες ομάδες, τα λεγόμενα clusters. Οι ομάδες αυτές θα περιλαμβάνουν στοιχεία τα οποία θα είναι πιο «όμοια» μεταξύ τους από ό,τι με στοιχεία άλλης ομάδας. Πιο «όμοια» στοιχεία σημαίνει ότι τα στοιχεία μιας ομάδας είναι κοντά πάντα με κάποιο μέτρο εγγύτητας [Faber94].

Το πρόβλημα διαχωρισμού k -means στην ουσία αφορά το εξής πρόβλημα:

Δοθέντος m δεδομένων στον n -διάστατο χώρο \mathbb{R}^n και έναν ακέραιο αριθμό k , ζητείται το σύνολο των κέντρων k στο \mathbb{R}^n , ώστε να ελαχιστοποιηθεί η μέση τετραγωνική απόσταση από το κάθε δεδομένο στο εγγύτερο σε αυτό κέντρο [Kanungo02].

Ο αλγόριθμος του Lloyd είναι δημοφιλής καθότι βασίζεται στην απλή παρατήρηση ότι η βέλτιστη τοποθέτηση των κέντρων είναι στο σημείο βάρους της κάθε ομάδας. Πρόκειται για έναν επαναληπτικό αλγόριθμο, ο οποίος αναθέτει τα δεδομένα σε ομάδες, με την κάθε ομάδα να χαρακτηρίζεται από ένα κέντρο. Σε κάθε δεδομένο αντιστοιχεί και ένα κέντρο, αυτό που είναι στην κοντινότερη απόσταση, ενώ μέσω των επαναλήψεων στόχος είναι να ελαχιστοποιηθεί το σφάλμα E , με τα κέντρα να μη μετακινούνται άλλο [Faber94]. Αποτελεί ένα μη εποπτευόμενο τρόπο μάθησης, αφού δεν υπάρχουν εκ των προτέρων οι ομάδες αυτές. Για να είναι σωστός ο διαχωρισμός πρέπει ένα κέντρο μιας συγκεκριμένης ομάδας να αναπαριστά όσο το δυνατόν καλύτερα τα δεδομένα που ανήκουν στην ομάδα αυτή [Faber94].

Το πρώτο βήμα του αλγορίθμου είναι η επιλογή του αριθμού των κέντρων k καθώς και των αρχικών θέσεων τους, βασίζομενη είτε σε εμπειρική γνώση ή παρατηρήσεις, είτε απλά δίνοντας τυχαίες τιμές σε αυτά [Morissette13]. Στη συνέχεια, κάθε ένα από τα δεδομένα ανατίθεται σε μια ομάδα, την κοντινότερη σε αυτό, βάσει ενός μετρικού απόστασης. Εδώ χρησιμοποιείται η τετραγωνισμένη Ευκλείδεια απόσταση. Κάθε ομάδα θα αποτελείται από στοιχεία και όλα αυτά θα είναι μέρος του υποχώρου που σχηματίζεται από το κέντρο της ομάδας. Επόμενο βήμα είναι η ενημέρωση των κέντρων χρησιμοποιώντας το μέσο όλων των στοιχείων της κάθε ομάδας. Οι επαναλήψεις συνεχίζονται μέχρις ότου να μην αλλάζουν πλέον οι θέσεις των κέντρων [Faber94].

Ο ψευδοκώδικας που περιγράφει τον εν λόγω αλγόριθμο είναι ο εξής:

Αλγόριθμος k-means

Είσοδος: χώρος εισόδου $\mathbf{x}_i \in \mathcal{R}^n, i = 1, 2, \dots, m$

ακέραιος αριθμός K , αρχικές θέσεις κέντρων $\mathbf{c}_j^0 \in \mathcal{R}^n, j = 1, 2, \dots, K$

Έξοδος: τελικές θέσεις κέντρων $\mathbf{c}_j \in \mathcal{R}^n, j = 1, 2, \dots, K$

1: **επανάλαβε**

2: **για** $i = 1, 2, \dots, m$

3: υπολογισμός του εγγύτερου κέντρου μ^j

$$\mu^j = \arg \min_{j=1,2,\dots,K} [\|\mathbf{x}_i - \mathbf{c}_j\|_2^2]$$

4: ανάθεση δεδομένου \mathbf{x}_i στο σύνολο του κέντρου μ^j

$$\begin{aligned} \mu^j &\leftarrow \mathbf{x}_i \\ l^j &\leftarrow l^j + 1 \end{aligned}$$

5: **τέλος**

6: **για** $j = 1, 2, \dots, K$

7: ανανέωση των κέντρων \mathbf{c}_j

με τα στοιχεία \mathbf{x}_i που ανήκουν στην ομάδα j

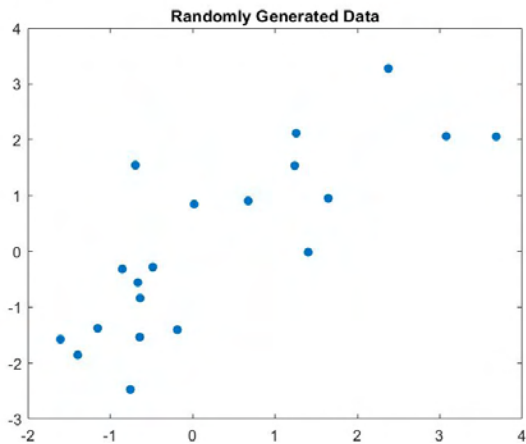
$$\mathbf{c}_j = \frac{1}{l^j} \sum_{i=1}^{l^j} \mathbf{x}_i$$

8: **τέλος**

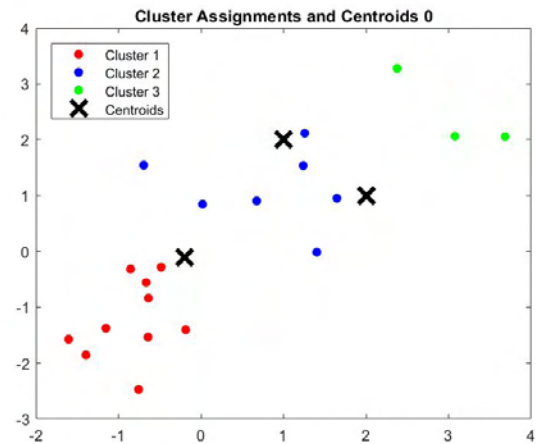
9: **μέχρι** να μη γίνονται αλλαγές στα κέντρα

Παρακάτω γίνεται μια εφαρμογή του αλγορίθμου για την καλύτερη κατανόηση του σε ένα τυχαίο σύνολο δεδομένων, το οποίο αποτελείται από 20 παραδείγματα εκπαίδευσης στο διδιάστατο χώρο για ευκολότερη αναπαράσταση. Επιλέγονται 3 ομάδες ($k = 3$).

Από την τυχαία αναπαραγωγή των δεδομένων και με την αρχικά τυχαία τοποθέτηση των κέντρων, κανείς λαμβάνει τα εξής γραφήματα:



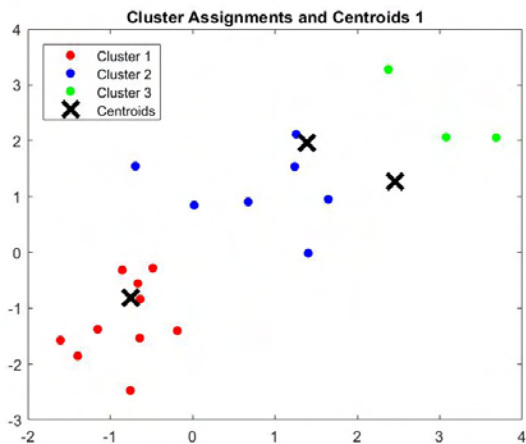
(α) Τυχαίο σύνολο δεδομένων.



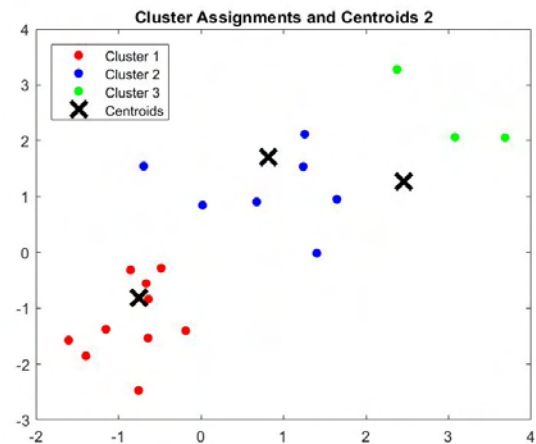
(β) Ανάθεση δεδομένων στις ομάδες.

Σχήμα 3.5: Εφαρμογή του αλγορίθμου k-means σε ένα τυχαίο σύνολο δεδομένων.

Με τις επαναλήψεις να τρέχουν τροποποιούνται οι θέσεις των κέντρων έτσι, ώστε τα σημεία των ομάδων να αντιπροσωπεύονται όσο το δυνατόν καλύτερα από αυτά, όπως φαίνεται στα ακόλουθα γραφήματα.



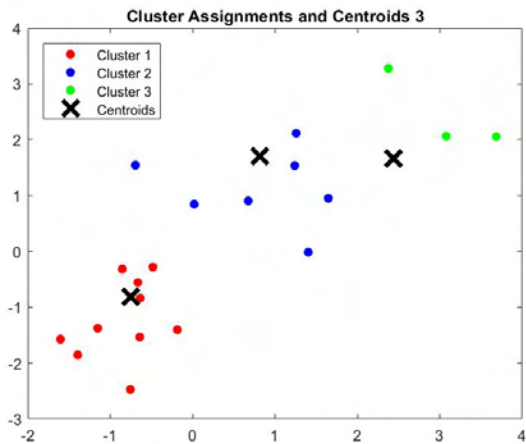
(α) Πρώτη επανάληψη.



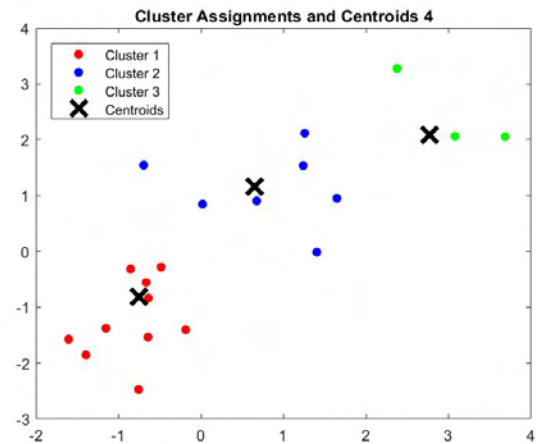
(β) Δεύτερη επανάληψη.

Σχήμα 3.6: Επαναλήψεις του αλγορίθμου στο τυχαίο σύνολο δεδομένων.

Το κέντρο της ομάδας 1 φαίνεται να μην αλλάζει πλέον θέση. Ωστόσο, οι υπόλοιπες ομάδες χρειάζονται ακόμα βελτίωση, η οποία και επέρχεται στις επόμενες επαναλήψεις του αλγορίθμου.

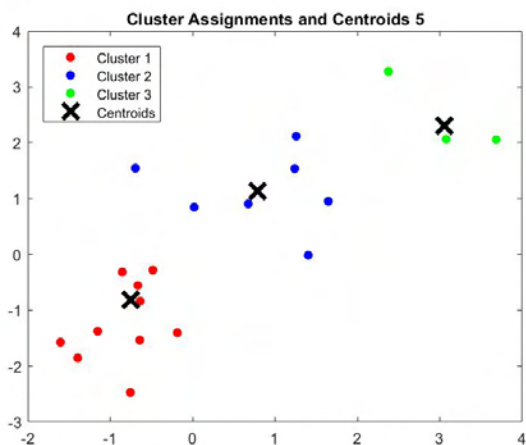


(α) Τρίτη επανάληψη.

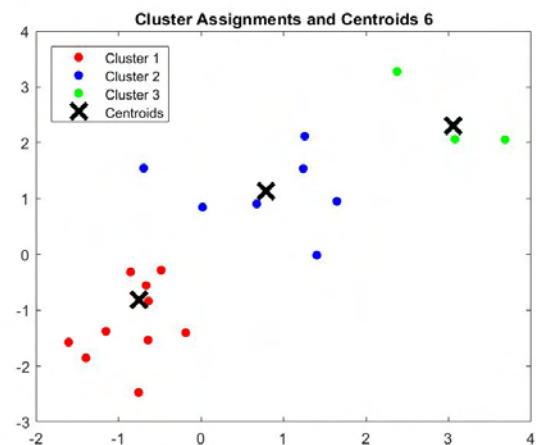


(β) Τέταρτη επανάληψη.

Σχήμα 3.7: Επαναλήψεις του αλγορίθμου στο τυχαίο σύνολο δεδομένων.



(α') Πέμπτη επανάληψη.

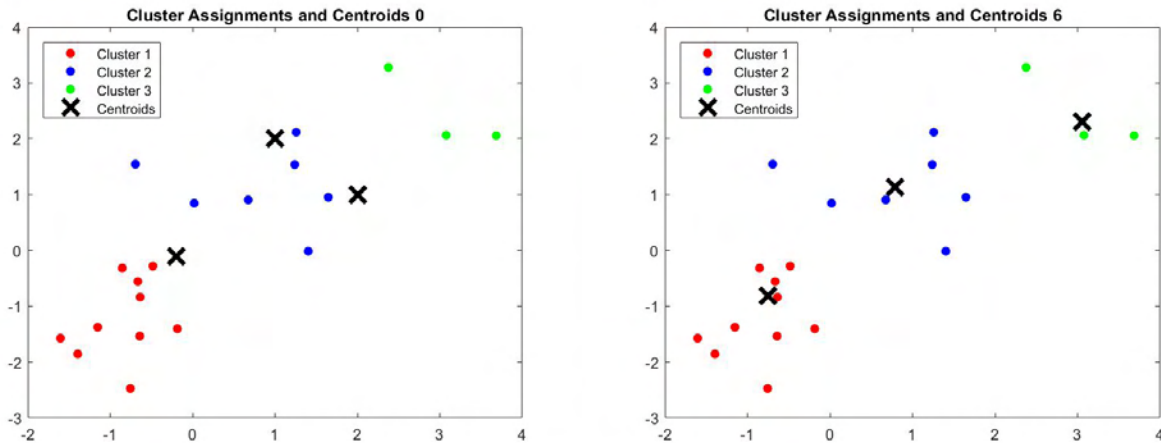


(β') Έκτη και τελευταία επανάληψη.

Σχήμα 3.8: Επαναλήψεις του αλγορίθμου στο τυχαίο σύνολο δεδομένων.

Ο αλγόριθμος φαίνεται να σταματάει με τις παραπάνω θέσεις των κέντρων, όπου σε σχέση με την αρχική τυχαία τοποθέτηση των κέντρων παρατηρείται σημαντική βελτίωση.

Συνολικά, από την αρχική τυχαία επιλογή (σχήμα 3.9α) μέχρι την επανάληψη στις θέσεις που τερματίζει ο αλγόριθμος (σχήμα 3.9β):



(α) Τοποθέτηση κέντρων στην αρχή του αλγορίθμου. (β) Τοποθέτηση κέντρων με το πέρας του αλγορίθμου.

Σχήμα 3.9: Συνολική βελτίωση των θέσεων των κέντρων στο τυχαίο σύνολο δεδομένων με τον αλγόριθμο k -means.

Από το παραπάνω παράδειγμα μπορούν να βγουν σημαντικά συμπεράσματα για τον αλγόριθμο. Η τελική θέση των κέντρων εξαρτάται από την αρχική επιλογή, με την τυχειότητα αυτή να μπορεί ίσως να περιοριστεί με έναν αριθμό εφαρμογών, όπου κάθε φορά τα κέντρα θα ξεκινούν από διαφορετικές θέσεις. Ακόμα, ο σωστός αριθμός των ομάδων και, επομένως, των κέντρων για το δίκτυο RBF, είναι άγνωστος εξ αρχής, οπότε θα χρειαστεί μια σειρά δοκιμών για να βρεθεί ο κατάλληλος αριθμός k μέσω δοκιμών και σφαλμάτων, ενώ λόγω της επαναληπτικής φύσης του αλγορίθμου μπορεί να οδηγήσει σε μεγάλους χρόνους σύγκλισης [Alexandridis13]. Ο αλγόριθμος αυτός συνήθως καταλήγει σε τοπικό ελάχιστο, δηλαδή σε έναν υποβέλτιστο τρόπο διαχωρισμού του χώρου εισόδου [Morissette13].

Για το λόγο αυτό, για να ξεπεραστούν τα παραπάνω προβλήματα, για την εύρεση των κέντρων ενός δικτύου RBF χρησιμοποιείται και μια άλλη μέθοδος, η οποία κάνει χρήση της ασαφούς λογικής και των ασαφών συνόλων. Αυτός ο τρόπος διαχωρισμού ονομάζεται αλγόριθμος fuzzy means [Sarimveis02], [Alexandridis03], [Alexandridis13].

3.4.3.2 Αλγόριθμος fuzzy means

Ο διαχωρισμός του χώρου εισόδου με τον αλγόριθμο k-means γίνεται σε διάκριτες ομάδες με το κάθε στοιχείο μιας ομάδας να ανήκει αποκλειστικά σε αυτήν και σε καμία άλλη. Με βάση την ασαφή λογική, όπου ένα στοιχείο ανήκει σε ένα ασαφές σύνολο με κάποιο βαθμό αλήθειας, τα στοιχεία μπορούν να ανήκουν σε παραπάνω της μιας ομάδας ως προς κάποιο βαθμό [Bora14]. Η λογική αυτή πρωτοορίστηκε από τον Bezdek το 1981 [Bezdek81] ως βελτίωση νωρίτερων προσπαθειών του Dunn το 1973 [Dunn73]. Αυτός ο τρόπος διαχωρισμού συνοδεύεται από κάποια πλεονεκτήματα εις βάρος του προηγούμενου όσον αφορά τον υπολογιστικό χρόνο, αφού εδώ δεν υπάρχει τυχαιότητα, όσες φορές και να «τρέξει» ο αλγόριθμος θα καταλήξει στα ίδια κέντρα και, επιπλέον, δεν χρειάζονται δοκιμές για την εύρεση του βέλτιστου αριθμού κέντρων μιας και γίνεται αυτόματα ο προσδιορισμός του [Sarimveis02], [Alexandridis03], [Alexandridis13].

Ό,τι ακολουθεί βασίζεται κατά κόρον στα άρθρα των [Sarimveis02], [Alexandridis03], [Alexandridis13]:

Ο αλγόριθμος συνίσταται στο διαχωρισμό του χώρου της κάθε μεταβλητής σε ασαφή σύνολα (fuzzy sets). Έστω ότι υπάρχουν N κανονικοποιημένες μεταβλητές για το χώρο εισόδου με συνολικά K παραδείγματα. Ο χώρος της κάθε μεταβλητής x_i ($i = 1, 2, \dots, N$) χωρίζεται σε c_i τριγωνικά σύνολα. Ο αριθμός των συνόλων σε κάθε διάσταση μπορεί να είναι ίδιος, οπότε ο διαχωρισμός αυτός αναφέρεται ως συμμετρικός, ή μπορεί να είναι διαφορετικός, οπότε θα αναφέρεται ως μη συμμετρικός. Στη συγκεκριμένη εργασία, γίνεται χρήση του συμμετρικού διαχωρισμού του χώρου εισόδου, στον οποίο κάθε μεταβλητή χωρίζεται στον ίδιο αριθμό ασαφών συνόλων.

Σε κάθε διάσταση, λοιπόν, θα υπάρχουν c ασαφή σύνολα:

$$\begin{aligned} i = 1 & \quad A_1^1, A_1^2, \dots, A_1^c \\ i = 2 & \quad A_2^1, A_2^2, \dots, A_2^c \\ & \quad \vdots \\ i = N & \quad A_N^1, A_N^2, \dots, A_N^c \end{aligned} \quad (3.32)$$

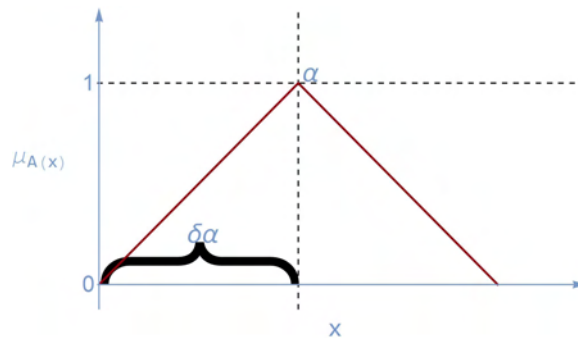
Κάθε σύνολο χαρακτηρίζεται με μια συνάρτηση συγγένειας της οποίας η τιμή (βαθμός αλήθειας) θα συμβολίζει το βαθμό συγγένειας ενός σημείου με το εν λόγω σύνολο, κατά πόσο δηλαδή μπορεί να θεωρηθεί στοιχείο του συνόλου [Βλαχάβας06].

Η συνάρτηση συγγένειας έχει την εξής μορφή:

$$\mu_A(x) = \begin{cases} 1 - \frac{|x - a|}{\delta a}, & x \in [a - \delta a, a + \delta a] \\ 0, & \text{αλλιώς} \end{cases} \quad (3.33)$$

Κάθε ασαφές σύνολο θα το περιγράψουν 2 ποσότητες: το κέντρο του συνόλου a και η ποσότητα δa η οποία είναι το μισό του αντίστοιχου εύρους.

Για παράδειγμα, ένα ασαφές σύνολο απεικονίζεται στο σχήμα 3.10.

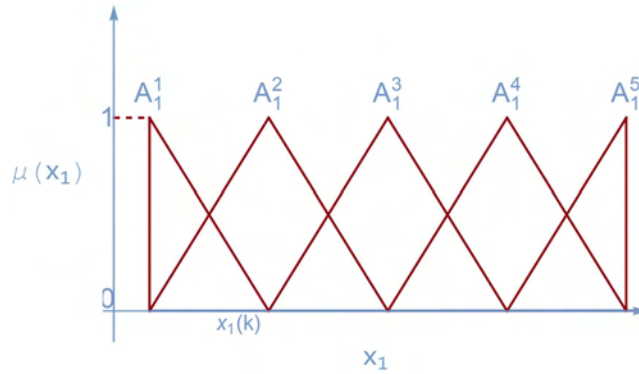


Σχήμα 3.10: Ασαφές σύνολο.

Ο διαχωρισμός του χώρου μιας μεταβλητής ορίζει τον λεγόμενο ασαφή διαχωρισμό (fuzzy partitioning), ο οποίος περιγράφεται ως εξής:

$$T_i = \{A_i^1, A_i^2, \dots, A_i^c\}, \quad 1 \leq i \leq N \quad (3.34)$$

Για παράδειγμα, για τη μεταβλητή 1 και για $c = 5$



Σχήμα 3.11: Ασαφής διαχωρισμός του χώρου μιας μεταβλητής.

Με

$$T_1 = \{A_1^1, A_1^2, \dots, A_1^5\} \quad (3.35)$$

Ο συνδυασμός των ασαφών συνόλων κάθε μεταβλητής ορίζει τους ασαφείς υποχώρους (fuzzy subspaces). Επειδή κάθε μεταβλητή έχει c ασαφή σύνολα, ο συνολικός αριθμός των ασαφών υποχώρων θα ορίζεται ως:

$$C = \prod_{i=1}^N c_i \quad (3.36)$$

Αυτός ο αριθμός υποδηλώνει το διαχωρισμό του συνολικού χώρου εισόδου. Δηλαδή ο χώρος εισόδου θα είναι χωρισμένος σε C ασαφείς υποχώρους.

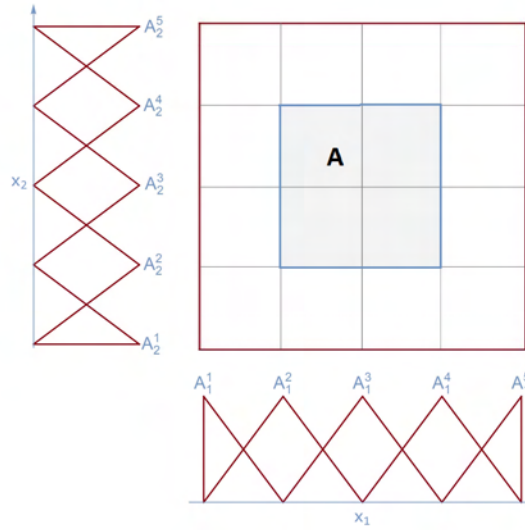
Για παράδειγμα, ο l -οστός ασαφής υποχώρος A^l ($1 \leq l \leq C$) είναι συνδυασμός N ασαφών συνόλων σε κάθε διάσταση και συμβολίζεται ως:

$$A_l = [A_1^l, A_2^l, \dots, A_N^l] \quad (3.37)$$

όπου

$$\begin{aligned} A_1^l &\in T_1 \\ A_2^l &\in T_2 \\ &\vdots \\ A_N^l &\in T_N \end{aligned} \quad (3.38)$$

Ο διαχωρισμός ενός διδιάστατου χώρου με ασαφείς υποχώρους φαίνεται στο σχήμα 3.12, με το χώρο της κάθε μεταβλητής εισόδου να είναι χωρισμένος σε 5 ασαφή σύνολα, δίνοντας έτσι συνολικά $C = 25$ ασαφείς υποχώρους.



Σχήμα 3.12: Διαχωρισμός ενός διδιάστατου χώρου εισόδου με ασαφείς υποχώρους.

Συγκεκριμένα, ο υποχώρος A που ορίζεται από τα ασαφή σύνολα A_1^3 και A_2^3 είναι ο εξής:

$$A = [A_1^3, A_2^3] \quad (3.39)$$

και ορίζει ένα ορθογώνιο το οποίο θα περιγράφεται από ένα κέντρο και από το αντίστοιχο εύρος. Δηλαδή,

$$\mathbf{a} = [a_1^3, a_2^3], \quad \delta \mathbf{a} = [\delta a_1, \delta a_2], \quad \text{ή} \quad \delta \mathbf{a} = [\delta a] \quad (3.40)$$

Πιο γενικά, ένας υποχώρος A^l θα ορίζεται από:

$$\mathbf{a} = [a_1^l, a_2^l, \dots, a_N^l], \quad \delta \mathbf{a} = [\delta a_1, \delta a_2, \dots, \delta a_N], \quad \text{ή} \quad \delta \mathbf{a} = [\delta a] \quad (3.41)$$

ή

$$A^l = \{\mathbf{a}^l, \delta \mathbf{a}\} \quad (3.42)$$

Στόχος, λοιπόν, είναι να βρεθούν αυτοί οι ασαφείς υποχώροι των οποίων το \mathbf{a}^l θα αποτελεί τα κέντρα των μονάδων RBF και θα είναι τέτοιои, ώστε να καλύπτουν ομοιόμορφα την κατανομή των δεδομένων εισόδου. Για το σκοπό αυτό, ορίζεται η έννοια της πολυδιάστατης συνάρτησης συμμετοχής ενός διανύσματος εισόδου $\mathbf{x}(k)$ με έναν ασαφή υποχώρο A^l ως:

$$\mu_{A^l}(\mathbf{x}(k)) = \begin{cases} 1 - rd^l(\mathbf{x}(k)), & rd^l(\mathbf{x}(k)) \leq 1 \\ 0, & \text{αλλιώς} \end{cases} \quad (3.43)$$

όπου $rd^l(\mathbf{x}(k))$ είναι η σχετική Ευκλείδεια απόσταση μεταξύ του A^l και του $\mathbf{x}(k)$:

$$rd^l(\mathbf{x}(k)) = \frac{[\sum_{i=1}^N (a_i^l - x_i(k))^2]^{1/2}}{[\sum_{i=1}^N (\delta a_i)^2]^{1/2}} \quad (3.44)$$

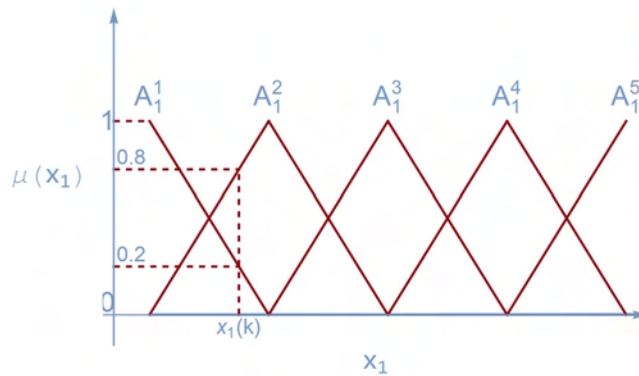
Στο συμμετρικό διαχωρισμό ($\delta a_1 = \delta a_2 = \dots = \delta a_N = \delta a$) το προηγούμενο γίνεται :

$$rd^l(\mathbf{x}(k)) = \frac{[\sum_{i=1}^N (a_i^l - x_i(k))^2]^{1/2}}{\sqrt{N}\delta a} \quad (3.45)$$

Σύμφωνα με τον ορισμό της πολυδιάστατης συνάρτησης συμμετοχής, εκείνος ο ασαφής υποχώρος με τη χαμηλότερη Ευκλείδεια σχετική απόσταση για το συγκεκριμένο διάνυσμα εισόδου θα δίνει τη μεγαλύτερη τιμή σε αυτή.

Η επιλογή των ασαφών συνόλων σε κάθε διάσταση προκύπτει από την τιμή της συνάρτησης συμμετοχής στην εν λόγω διάσταση, ενώ αυτή θα καθορίζει, τελικά, την επιλογή των ασαφών υποχώρων. Σε κάθε διάσταση υπολογίζεται ο βαθμός συμμετοχής ενός παραδείγματος εισόδου με το κάθε ασαφές σύνολο αυτής της διάστασης. Όσο πιο κοντά είναι το εν λόγω παράδειγμα σε ένα ασαφές σύνολο, τόσο πιο μεγάλος θα είναι ο βαθμός συμμετοχής με τη μέγιστη τιμή που μπορεί να πάρει αυτός να είναι η 1. Όσο πιο μακριά θα είναι το σημείο, τόσο πιο πολύ θα μειώνεται ο βαθμός συμμετοχής μέχρι την οριακή τιμή 0. Τελικά, θα επιλέγεται αυτό το ασαφές σύνολο το οποίο δίνει τη μεγαλύτερη τιμή της συνάρτησης συμμετοχής, όπως φαίνεται στο ακόλουθο παράδειγμα.

Ένα συγκεκριμένο παράδειγμα εκπαίδευσης k στη διεύθυνση $i = 1$ θα δίνει τη μεγαλύτερη τιμή του



Σχήμα 3.13: Υπολογισμός των βαθμών αλήθειας μιας μεταβλήτης με τα ασαφή σύνολα.

βαθμού αλήθειας σε εκείνο το ασαφές σύνολο που είναι εγγύτερα σε αυτό, ενώ στα πλέον μακρινά θα δίνει την τιμή 0. Τελικά, εδώ επιλέγεται το ασαφές σύνολο A_1^2 , μιας και δίνει τη μεγαλύτερη τιμή στη συνάρτηση συμμετοχής, όπως φαίνεται στο σχήμα 3.13.

Αν οι παραπάνω υπολογισμοί επαναληφθούν για όλες τις διαστάσεις, τότε θα προκύψει ο ασαφής υποχώρος που θα περιγράφει καλύτερα το συγκεκριμένο διάνυσμα εισόδου.

Οι επιλεγμένοι ασαφείς υποχώροι πρέπει να καλύπτουν επαρκώς τα διανύσματα εισόδου. Αν οι υπάρχοντες ασαφείς υποχώροι δεν τα καλύπτουν, τότε εισάγεται νέος. Για τον έλεγχο κάλυψης ενός διανύσματος εισόδου από κάποιον ασαφή υποχώρο χρησιμοποιείται η πολυδιάστατη συνάρτηση συγγένειας με το εν λόγω ασαφή υποχώρο: αν η σχετική Ευκλείδεια απόσταση $rd^l(\mathbf{x}(k))$ είναι μεγαλύτερη της μονάδας, τότε το συγκεκριμένο διάνυσμα εισόδου δεν καλύπτεται από τον l ασαφή υποχώρο.

Η διαδικασία έχει ως εξής:

Για κάθε μια από τις κανονικοποιημένες μεταβλητές εισόδου N γίνεται διαχωρισμός των πεδίων τιμών σε s ασαφή σύνολα στα οποία ανατίθενται οι αντίστοιχες τιμές κέντρου και εύρους. Οι υπολογισμοί ξεκινούν με το πρώτο παράδειγμα εκπαίδευσης $k = 1$ ενώ αφορούν το σχηματισμό του πρώτου ασαφούς υποχώρου, δηλαδή του πρώτου κέντρου του δικτύου, $L = 1$. Για κάθε μεταβλητή i εισόδου ($i = 1, 2, \dots, N$) υπολογίζεται ένας βαθμός αλήθειας από τη συνάρτηση συμμετοχής με το κάθε ασαφές σύνολο στη διάσταση i . Αυτό το j ασαφές σύνολο που δίνει το μεγαλύτερο βαθμό συμμετοχής στη διεύθυνση i θα είναι μέρος του ασαφούς υποχώρου που θα σχηματιστεί, το οποίο στη διεύθυνση i θα έχει ως συντεταγμένη το κέντρο του συγκεκριμένου ασαφούς συνόλου. Οι υπολογισμοί αυτοί συνεχίζονται για

όλες τις διευθύνσεις από $1, \dots, N$, όπου διαμορφώνεται ο πρώτος ασαφής υποχώρος και, τελικά, το πρώτο κέντρο του δικτύου.

Στη συνέχεια, εισάγονται τα επόμενα παραδείγματα εκπαίδευσης, $k = 2, \dots, K$. Για το επόμενο παράδειγμα γίνεται ένας έλεγχος ο οποίος αφορά την κάλυψη του από τους υπάρχοντες ασαφείς υποχώρους (κέντρα του δικτύου). Αν αυτό το παράδειγμα καλύπτεται από κάποιο κέντρο, κάτι το οποίο δηλώνεται με τη σχετική Ευκλείδεια απόσταση και, τελικά, με την πολυδιάστατη συνάρτηση συμμετοχής, τότε δεν προστίθεται νέο κέντρο L . Αν δεν υπάρχει σχετική Ευκλείδεια απόσταση μικρότερη ή ίση της μονάδος, τότε προστίθεται νέο κέντρο και ξεκινούν οι υπολογισμοί για να βρεθούν οι συντεταγμένες του, όπως περιγράφηκε προηγουμένως. Έπειτα, σειρά έχει το επόμενο παράδειγμα εκπαίδευσης με τη διαδικασία να συνεχίζεται μέχρι να περάσουν όλα τα δεδομένα εκπαίδευσης από μια φορά το καθένα.

Η μη επαναληπτική φύση του αλγορίθμου καθιστά τον υπολογισμό των κέντρων του δικτύου μια αρκετά γρήγορη διαδικασία, ακόμα και στην περίπτωση μεγάλων συνόλων δεδομένων, ενώ αλγόριθμος που εκτελεί την επιλογή των ασαφών υποχώρων και, τελικά, των κέντρων του δικτύου είναι ο εξής [Sarimveis02], [Alexandridis03], [Alexandridis13]:

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

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**

3.4.4 Υπολογισμός των ευρών των συναρτήσεων βάσης ενός δικτύου RBF

Αφού έχουν καθοριστεί τα κέντρα των μονάδων RBF, σειρά έχουν τα εύρη. Τα εύρη μιας ακτινικής συμμετρικής συνάρτησης βάσης υποδηλώνουν την περιοχή της ισχύος της και επιλέγονται έτσι, ώστε οι μονάδες RBF να καλύπτουν όσο το δυνατόν καλύτερα το χώρο εισόδου, δίνοντας μια ομαλή συνάρτηση προσέγγισης [Leonard91]. Αυτό σημαίνει ότι κάθε μονάδα RBF θα πρέπει να ενεργοποιεί και τουλάχιστον ακόμα μια σε κάποιο βαθμό. Ωστόσο, η τιμή του εύρους πρέπει να κρατιέται χαμηλή για να διατηρηθεί ο τοπικός χαρακτήρας των συναρτήσεων βάσης, αφού μια συνάρτηση πρέπει να καλύπτει μια συγκεκριμένη περιοχή: αυτή την οποία αναπαριστά [Leonard91].

Μια επιλογή των ευρών μπορεί να γίνει βάσει της μέγιστης απόστασης των κέντρων, με την κάθε συνάρτηση βάσης ($k = 1, 2, \dots, K$) να έχει την ίδια τιμή εύρους, η οποία υπολογίζεται ως εξής:

$$\sigma = \frac{d_{max}}{\sqrt{N}} \quad (3.46)$$

όπου N ο αριθμός των μονάδων RBF και d_{max} η μέγιστη απόσταση μεταξύ των κέντρων [Engelbrecht07].

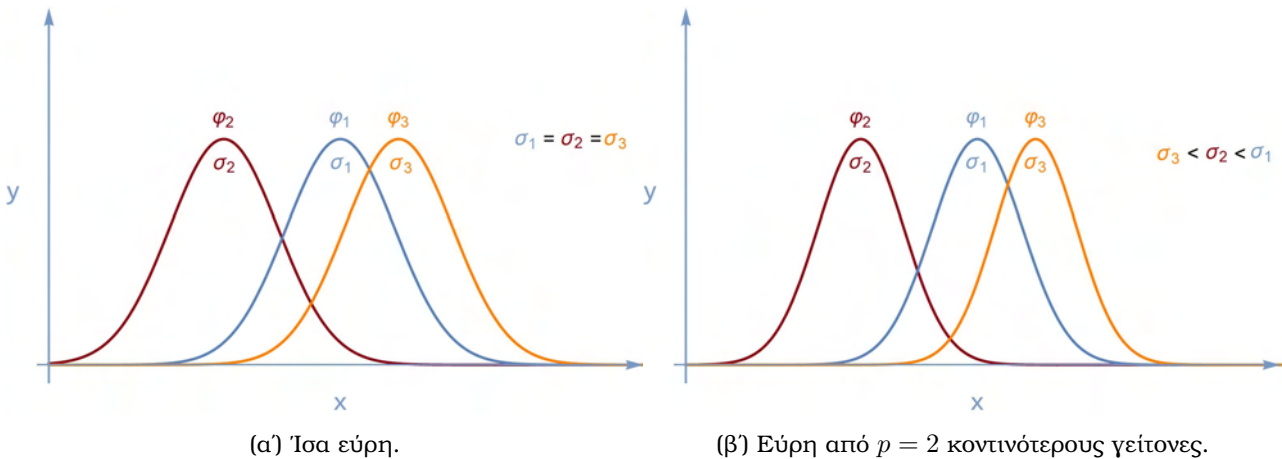
Ωστόσο, μια άλλη επιλογή των ευρών, αυτή η οποία χρησιμοποιείται στην παρούσα εργασία, βασίζεται σε έναν ευρετικό τρόπο υπολογισμού των p -κοντινότερων γειτόνων (p -nearest neighbors), με το εύρος για κάθε συνάρτηση βάσης να επιλέγεται έτσι, ώστε το πεδίο της συγκεκριμένης μονάδας RBF να αλληλοκαλύπτεται με τις p εγγύτερες RBF σε αυτήν [Alexandridis13], [Moody89].

Το εύρος σ_j για τη j μονάδα RBF υπολογίζεται ως εξής:

$$\sigma_j = \sqrt{\frac{1}{p} \sum_{i=1}^p \|c_j - c_i\|_2^2} \quad (3.47)$$

όπου τα c_i είναι οι κοντινότεροι γείτονες του c_j .

Παρακάτω φαίνονται σχηματικά οι δύο περιπτώσεις επιλογής των ευρών στην περίπτωση της συνάρτησης Gaussian, με ίσα εύρη (σχήμα 3.14α) και με εύρη από τη μέθοδο των p -κοντινότερων γειτόνων (σχήμα 3.14β), για $p = 2$.



Σχήμα 3.14: Επιλογή ευρών συναρτήσεων βάσης.

Κεφάλαιο 4

Αλγόριθμοι Βελτιστοποίησης

4.1 Εισαγωγικές έννοιες

Σε όλα τα προβλήματα βελτιστοποίησης στόχος είναι να βρεθεί εκείνο το σύνολο παραμέτρων το οποίο δίνει τη βέλτιστη τιμή μιας συνάρτησης. Η συνάρτηση αυτή αποκαλείται αντικειμενική συνάρτηση (objective function) και ορίζει με ένα ποσοτικό μέγεθος την απόδοση του συστήματος που μελετάται. Η βελτιστοποίηση μπορεί να αφορά προβλήματα ελαχιστοποίησης ή μεγιστοποίησης, με τα τελευταία να μπορούν να οριστούν ως τα πρώτα, απαιτώντας απλώς την ελαχιστοποίηση της αντίθετης αντικειμενικής. Οι παράμετροι, ή αλλιώς μεταβλητές, του προβλήματος μπορεί να συνοδεύονται από κάποιους περιορισμούς, διακρίνοντας έτσι τα προβλήματα βελτιστοποίησης σε αυτά που έχουν περιορισμούς και σε αυτά που δεν έχουν. Ωστόσο, σε μερικές περιπτώσεις είναι δυνατή η μετατροπή ενός προβλήματος με περιορισμούς σε ένα πρόβλημα χωρίς περιορισμούς με δύο τρόπους: ο πρώτος τρόπος απαιτεί την επαναμορφοποίηση του προβλήματος εισάγοντας τους περιορισμούς στην αντικειμενική συνάρτηση με κάποιο είδος «ποινής», ώστε να αποθαρρύνονται οι παραβιάσεις των περιορισμών, ενώ ο δεύτερος τρόπος υποθέτει την ασφαλή παράβλεψή τους όταν, παραδείγματος χάριν, δεν επηρεάζουν τη λύση του προβλήματος. Επιπλέον διαχωρισμός γίνεται σε γραμμικά και μη γραμμικά προβλήματα με τα τελευταία να περιέχουν τουλάχιστον μια μη γραμμικότητα, είτε στην αντικειμενική συνάρτηση, είτε σε κάποιον από τους περιορισμούς [Nocedal99].

Η αντικειμενική συνάρτηση μαζί με τις μεταβλητές και τους τυχόν περιορισμούς που τις συνοδεύουν συνιστούν το μοντέλο το οποίο περιγράφει το πρόβλημα βελτιστοποίησης. Εφόσον έχει οριστεί το μοντέλο, επόμενο βήμα είναι η χρήση ενός αλγορίθμου βελτιστοποίησης για να βρεθεί μια λύση του προβλήματος. Μιας και δεν υπάρχει ένας συγκεκριμένος αλγόριθμος που να χρησιμοποιείται σε όλα τα προβλήματα βελτιστοποίησης, πρέπει μέσα από μια συλλογή αλγορίθμων να επιλεγθεί ο κατάλληλος. Η επιλογή θεωρείται κρίσιμη καθώς αυτή θα καθορίσει τελικά το χρόνο επίλυσης ή και, ακόμα, την εύρεση λύσης [Nocedal99].

Μια λύση θα θεωρείται πραγματική λύση του προβλήματος εφόσον ικανοποιεί ορισμένες μαθηματικές εκφράσεις. Αυτές οι εκφράσεις συνιστούν τις συνθήκες βελτιστοποίησης. Αν η επιλεγμένη λύση δεν τις ικανοποιεί τότε μπορεί να προκύψουν χρήσιμες πληροφορίες για το πώς αυτή μπορεί να βελτιωθεί, για παράδειγμα μέσω κάποιας ανάλυσης ευαισθησίας [Nocedal99].

4.2 Μορφοποίηση προβλήματος βελτιστοποίησης

Από μαθηματική άποψη, ένα πρόβλημα βελτιστοποίησης ορίζεται ως:

$$\min_x f(\mathbf{x}) \quad \text{s.t. } c_i \in C \quad (4.1)$$

όπου \mathbf{x} είναι το διάνυσμα των παραμέτρων του προβλήματος, f είναι η αντικειμενική συνάρτηση και c_i είναι οι συναρτήσεις που περιγράφουν τους περιορισμούς, τους οποίους πρέπει να ικανοποιούν οι παράμετροι. Για ένα πρόβλημα βελτιστοποίησης χωρίς περιορισμούς το σύνολο C αποτελεί το κένο

σύνολο $C = \emptyset$, οπότε το πρόβλημα μορφοποιείται ως εξής:

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad (4.2)$$

Στην παρούσα εργασία το πρόβλημα εκπαίδευσης ενός νευρωνικού δικτύου RBF, το οποίο περιγράφηκε στο προηγούμενο κεφάλαιο, μορφοποιείται ως ένα πρόβλημα ελαχίστων τετραγώνων, οπότε ότι ακολουθεί θα βασίζεται στον εξής συμβολισμό, προσαρμοσμένα από το βιβλίο του Hansen κ.ά. [Hansen12]:

$$\min_{\mathbf{x}} f(\mathbf{x}) \equiv \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|_2^2 = \min_{\mathbf{x}} \frac{1}{2} \sum_{i=1}^m r_i(\mathbf{x})^2 \quad (4.3)$$

όπου $\mathbf{x} \in \mathcal{R}^n$ και $\mathbf{r}(\mathbf{x}) = (r_1(\mathbf{x}), \dots, r_m(\mathbf{x}))^T$, οι παράμετροι και τα υπόλοιπα του προβλήματος αντίστοιχα, με

$$r_i(\mathbf{x}) = y_i - M(\mathbf{x}, t_i) \quad i = 1, \dots, m \quad (4.4)$$

Η συνάρτηση M είναι το μοντέλο που χρησιμοποιείται, έτσι ώστε για δεδομένες τιμές t_i να δίνονται ως αποτέλεσμα τα αντίστοιχα y_i . Σε αναλογία με τις συναρτήσεις βάσης ενός δικτύου RBF, τα ζευγάρια (t_i, y_i) , $i = 1, \dots, m$, θα αποτελούν το σύνολο δεδομένων εκπαίδευσης ενώ οι παράμετροι του προβλήματος μπορεί να είναι τα κέντρα ή και τα βάρη του δικτύου.

Εφόσον δεν είναι γνωστή εκ των προτέρων η μορφή της αντικειμενικής συνάρτησης, αλλά είναι γνωστή η τιμή ή και η παράγωγος σε έναν συγκεκριμένο αριθμό σημείων (δεδομένων), η εύρεση ενός ολικού ελάχιστου είναι πιθανόν να είναι υπολογιστικά ακριβή ή ακόμα και αδύνατη, οπότε οι αλγόριθμοι που χρησιμοποιούνται επιδιώκουν την εύρεση μιας τοπικής λύσης.

Οι συνθήκες βελτιστοποίησης χρησιμοποιούνται για να χαρακτηρίσουν ένα σημείο \mathbf{x}^* ως τοπικό ελάχιστο και βασίζονται στις πρώτες και δεύτερες παραγώγους της $f(\mathbf{x})$, οι οποίες ορίζονται στη συνέχεια. Η κλίση (gradient) μιας βαθμωτής συνάρτησης $f(\mathbf{x})$, $\nabla f(\mathbf{x})$, όπου $\mathbf{x} \in \mathcal{R}^n$, ορίζεται ως:

$$\nabla f(\mathbf{x})^T = \left[\frac{\partial f(\mathbf{x})}{\partial x_1} \quad \frac{\partial f(\mathbf{x})}{\partial x_2} \quad \dots \quad \frac{\partial f(\mathbf{x})}{\partial x_n} \right] \quad (4.5)$$

ενώ ο Εσσιανός (Hessian) της $f(\mathbf{x})$, $\nabla^2 f(\mathbf{x})$, ορίζεται ως:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix} \quad (4.6)$$

ή

$$[\nabla^2 f(\mathbf{x})]_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}, \quad i, j = 1, 2, \dots, n \quad (4.7)$$

Συνθήκες βελτιστοποίησης

Αναγκαία συνθήκη πρώτης τάξης:

το \mathbf{x}^* είναι ένα στάσιμο σημείο, δηλαδή η κλίση της f στο \mathbf{x}^* είναι 0, $\nabla f(\mathbf{x}^*) = 0$.

Ικανοποιητική συνθήκη δεύτερης τάξης:

ο Εσσιανός H , $\nabla^2 f(\mathbf{x}^*)$, είναι θετικά ορισμένος, δηλαδή είναι συμμετρικός με θετικές ιδιοτιμές ενώ ισχύει για κάθε διάνυσμα $\mathbf{z} \in \mathcal{R}^n$ ($\mathbf{z} \neq \mathbf{0}$)

$$\mathbf{z}^T H \mathbf{z} > 0 \quad (4.8)$$

Στην προκειμένη περίπτωση, η συνάρτηση f ορίζεται ως άθροισμα τετραγώνων μη γραμμικών συναρτήσεων

$$f(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^m r_i(\mathbf{x})^2 \quad (4.9)$$

με

$$r_i(\mathbf{x}) = y_i - M(\mathbf{x}, t_i) \quad i = 1, \dots, m \quad (4.10)$$

τότε για συγκεκριμένα i, j κανείς λαμβάνει:

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = \sum_{i=1}^m r_i(\mathbf{x}) \frac{\partial r_i(\mathbf{x})}{\partial x_j} \quad (4.11)$$

Οπότε ο πίνακας της Ιακωβιανής του διανύσματος των υπολοίπων $\mathbf{r}(\mathbf{x})$ θα είναι ο πίνακας με στοιχεία:

$$[J(\mathbf{x})]_{ij} = \frac{\partial r_i(\mathbf{x})}{\partial x_j}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n \quad (4.12)$$

δηλαδή,

$$J = \begin{bmatrix} \frac{\partial r_1(\mathbf{x})}{\partial x_1} & \frac{\partial r_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial r_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial r_2(\mathbf{x})}{\partial x_1} & \frac{\partial r_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial r_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_m(\mathbf{x})}{\partial x_1} & \frac{\partial r_m(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial r_m(\mathbf{x})}{\partial x_n} \end{bmatrix} \quad (4.13)$$

Για την εξαγωγή του Εσσιανού κανείς χρειάζεται τη δεύτερη παράγωγο της f .

Από την

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = \sum_{i=1}^m r_i(\mathbf{x}) \frac{\partial r_i(\mathbf{x})}{\partial x_j} \quad (4.14)$$

προκύπτει

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_k \partial x_l} = \sum_{i=1}^m \frac{\partial r_i(\mathbf{x})}{\partial x_k} \frac{\partial r_i(\mathbf{x})}{\partial x_l} + \sum_{i=1}^m r_i(\mathbf{x}) \frac{\partial^2 r_i(\mathbf{x})}{\partial x_k \partial x_l} \quad (4.15)$$

Σε μητρική μορφή τα παραπάνω γράφονται ως:

$$\nabla f(\mathbf{x}) = J(\mathbf{x})^T \mathbf{r}(\mathbf{x}) \quad (4.16)$$

$$\nabla^2 f(\mathbf{x}) = J(\mathbf{x})^T J(\mathbf{x}) + \sum_{i=1}^m r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x}) \quad (4.17)$$

Όπου οι συνθήκες βελτιστοποίησης θα απαιτούν

$$\nabla f(\mathbf{x}) = \mathbf{0} \quad (4.18)$$

$$\nabla^2 f(\mathbf{x}) \text{ θετ.ορισμένος} \quad (4.19)$$

Στην περίπτωση της μη γραμμικής βελτιστοποίησης χρησιμοποιούνται επαναληπτικές μέθοδοι οι οποίες ξεκινούν από μια αρχική λύση \mathbf{x}_0 και αναπαράγουν μια ακολουθία λύσεων \mathbf{x}_k , η οποία πρέπει να καταλήξει σε κάποιο τοπικό ελάχιστο. Αυτές οι μέθοδοι αναπτύσσονται στη συνέχεια [Madsen04].

4.3 Μέθοδοι κατάβασης

Οι περισσότερες μέθοδοι για την εύρεση μιας τοπικής λύσης απαιτούν μια φθίνουσα συνθήκη (descending condition) στην οποία πρέπει να ισχύει:

$$f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k) \quad (4.20)$$

Με την απαίτηση της ικανοποίησης της παραπάνω συνθήκης αποτρέπεται η κατάληξη σε τοπικό μέγιστο αλλά και είναι λιγότερο πιθανό να συγκλίνει ο αλγόριθμος σε κάποιο σαγματικό σημείο.

Η μετακίνηση σε μια νέα θέση \mathbf{x}_{k+1} απαιτεί:

- μια διεύθυνση κατάβασης (descent direction), η διεύθυνση δηλαδή στην οποία μειώνεται η τιμή της συνάρτησης

- ένα μέγεθος βήματος, ώστε η τιμή της συνάρτησης να μειώνεται σε κάποιο ικανοποιητικό βαθμό

Δηλαδή,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + a\mathbf{h}_d \quad (4.21)$$

όπου a είναι το εν λόγω μέγεθος βήματος και \mathbf{h}_d η διεύθυνση κατάβασης.

Από το θεώρημα Taylor

$$f(\mathbf{x}_k + a\mathbf{h}) = f(\mathbf{x}_k) + a\mathbf{h}^T \nabla f(\mathbf{x}) + \mathcal{O}(a^2) \quad (4.22)$$

ή για ένα a ικανοποιητικά μικρό με $a > 0$

$$f(\mathbf{x}_{k+1}) \approx f(\mathbf{x}_k) + a\mathbf{h}^T \nabla f(\mathbf{x}) \quad (4.23)$$

Από το παραπάνω εύκολα συμπεραίνεται ότι μια διεύθυνση \mathbf{h} θα είναι μια διεύθυνση κατάβασης της f στο \mathbf{x} αν ισχύει:

$$\mathbf{h}^T \nabla f(\mathbf{x}) < 0 \quad (4.24)$$

Έχοντας μια διεύθυνση κατάβασης, η επιλογή του βήματος, δηλαδή η απόσταση που θα διανυθεί σε αυτή τη διεύθυνση \mathbf{h} ξεκινώντας από την τρέχουσα θέση \mathbf{x} , ώστε να προκύπτει επαρκής μείωση στην τιμή της f , μπορεί να βρεθεί λύνοντας το εξής μονοδιάστατο πρόβλημα:

$$a = \arg \min_{a>0} f(\mathbf{x} + a\mathbf{h}) \quad (4.25)$$

Λύνοντας το παραπάνω επακριβώς κατακτάται όλο το κέρδος που προτείνει η διεύθυνση κατάβασης \mathbf{h} , ωστόσο με κάποιο επακόλουθο υπολογιστικό κόστος. Από την άλλη, μπορεί και να μη χρειάζεται η εύρεση του πραγματικού ελαχίστου και για το λόγο αυτό, συνήθως, επαρκεί μια προσεγγιστική λύση την οποία και επιδιώκουν οι λεγόμενοι αλγόριθμοι *line search*, όπου μέσα από κάποιες δοκιμές για το μέγεθος του βήματος επιλέγεται τελικά αυτό που δίνει τη μεγαλύτερη δυνατή μείωση στη τιμή της συνάρτησης στη νέα θέση [Nocedal99]. Έπειτα, στη νέα θέση υπολογίζεται μια νέα διεύθυνση \mathbf{h} και ένα μέγεθος βήματος a με τη διαδικασία να συνεχίζεται μέχρι να ικανοποιηθεί κάποιο κριτήριο τερματισμού.

Ένας υποτυπώδης αλγόριθμος για την εν λόγω διαδικασία μπορεί να είναι ο εξής [Madsen04]:

Αλγόριθμος μεθόδου κατάβασης

1. $k = 0, \mathbf{x}_i = \mathbf{x}_0$
 2. **για** $k = 0, 1, 2, \dots$
 3. εύρεση διεύθυνσης κατάβασης \mathbf{h}_d
 4. εύρεση μεγέθους βήματος a
 5. ενημέρωση νέας θέσης
 $\mathbf{x}_{k+1} = \mathbf{x}_k + a\mathbf{h}_d$
 6. **τέλος**
-

4.3.1 Υπολογισμός διεύθυνσης κατάβασης

Για να βρεθεί η νέα θέση, λοιπόν, απαιτείται, αρχικά, μια διεύθυνση κατάβασης και, έπειτα, ένα μέγεθος βήματος. Οι διευθύνσεις \mathbf{h}_d μπορούν να υπολογιστούν με τις εξής μεθόδους: τη μέθοδο απότομης κατάβασης (steepest descent method) και τη μέθοδο Newton [Nocedal99], [Hansen12], [Madsen04].

4.3.1.1 Μέθοδος απότομης κατάβασης

Προηγούμενως, μέσω του θεωρήματος Taylor για μικρά a , προέκυψε ότι μια διεύθυνση \mathbf{h} είναι διεύθυνση κατάβασης \mathbf{h}_d αν ισχύει το εξής:

$$\mathbf{h}^T \nabla f(\mathbf{x}) < 0 \quad (4.26)$$

Αυτός είναι και ο ρυθμός της αλλαγής της f στη διεύθυνση \mathbf{h} στο σημείο \mathbf{x} . Επομένως, ζητείται εκείνη η μοναδιαία διεύθυνση που να δίνει το μεγαλύτερο ρυθμό μείωσης. Αυτή προκύπτει λύνοντας το εξής πρόβλημα ελαχιστοποίησης [Nocedal99]:

$$\min_{\|\mathbf{h}\|=1} \mathbf{h}^T \nabla f(\mathbf{x}) \quad (4.27)$$

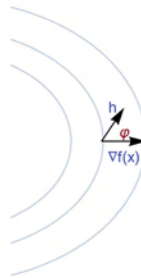
Από το εσωτερικό γινόμενο των διανυσμάτων κανείς λαμβάνει:

$$\mathbf{h}^T \nabla f(\mathbf{x}) = \cos(\phi) \|\nabla f(\mathbf{x})\| \|\mathbf{h}\| \quad (4.28)$$

ή

$$\mathbf{h}^T \nabla f(\mathbf{x}) = \cos(\phi) \|\nabla f(\mathbf{x})\| \quad (4.29)$$

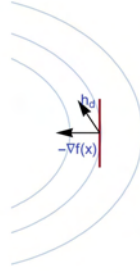
όπου ϕ είναι η γωνία που σχηματίζεται μεταξύ της διεύθυνσης \mathbf{h} και της κλίσης (gradient) $\nabla f(\mathbf{x})$, όπως φαίνεται στο σχήμα 4.1.



Σχήμα 4.1: Σχηματιζόμενη γωνία μιας διεύθυνσης με τη κλίση σε συγκεκριμένο σημείο.

Από τα παραπάνω κανείς παρατηρεί ότι για να ισχύει η συνθήκη κατάβασης πρέπει $\cos(\phi) < 0$.

Δηλαδή η διεύθυνση \mathbf{h} πρέπει να βρίσκεται στο χώρο που ορίζεται από το επίπεδο της εφαπτομένης και περιέχει την αρνητική κλίση $-\nabla f(\mathbf{x})$ για να είναι μια διεύθυνση κατάβασης \mathbf{h}_d (σχήμα 4.2).



Σχήμα 4.2: Ορισμός διεύθυνσης κατάβασης.

Τελικά, από τα παραπάνω κανείς καταλήγει ότι η διεύθυνση που δίνει τη μέγιστη κάθοδο (steepest descent) είναι για $\phi = \pi$, ώστε η νέα θέση να κινείται προς

$$\mathbf{h}_{sd} = -\nabla f(\mathbf{x}) \quad (4.30)$$

ή αλλιώς

$$\mathbf{x}_{k+1} = \mathbf{x}_k - a\nabla f(\mathbf{x}) \quad (4.31)$$

Η παραπάνω επιλογή της διεύθυνσης είναι η πιο προφανής επιλογή καθώς από όλες τις διευθύνσεις ζητείται αυτή που μειώνει την τιμή της συνάρτησης με το γρηγορότερο ρυθμό [Nocedal99]. Λέγεται ότι οφείλεται στον Cauchy το 1847 [Lemarechal12]. Θεωρείται η καλύτερη επιλογή τοπικά, ενώ μπορεί να συνδυαστεί με διάφορους τρόπους επιλογής του βήματος a . Η μέθοδος συγκλίνει αλλά αργά. Για πολλά προβλήματα χρησιμοποιείται στα αρχικά στάδια της επαναληπτικής διαδικασίας, έχοντας αρκετά καλή απόδοση [Madsen04]. Ένα πλεονέκτημα αυτής της μεθόδου είναι ότι απαιτούνται μόνο υπολογισμοί της κλίσης ∇f και όχι δευτέρων παραγώγων, ωστόσο μπορεί να είναι επικίνδυνα αργή [Nocedal99].

4.3.1.2 Μέθοδος Newton

Μια άλλη μέθοδος για την επιλογή της διεύθυνσης είναι η κλασική μέθοδος Newton και θεωρείται από τις σημαντικότερες [Nocedal99]. Η μέθοδος αυτή βασίζεται στη προσέγγιση της $f(\mathbf{x})$ με σειρά Taylor 2ης τάξης.

Από το ανάπτυγμα Taylor, κρατώντας όρους μέχρι και δεύτερης τάξης και υποθέτοντας πάντα ότι η υπό μελέτη συνάρτηση είναι δύο φορές παραγωγίσιμη, προκύπτει:

$$f(\mathbf{x} + \mathbf{h}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})^T \mathbf{h} + \frac{1}{2} \mathbf{h}^T H \mathbf{h} \quad (4.32)$$

Οι πρώτης τάξης συνθήκες απαιτούν:

$$\nabla f(\mathbf{x} + \mathbf{h}) = \mathbf{0} \quad (4.33)$$

από όπου προκύπτει ότι

$$\nabla f(\mathbf{x}) + H \mathbf{h} = \mathbf{0} \quad (4.34)$$

Η συνθήκη δεύτερης τάξης απαιτεί ο H να είναι θετικά ορισμένος, οπότε η διεύθυνση Newton είναι:

$$\mathbf{h}_N = -H^{-1} \nabla f(\mathbf{x}) \quad (4.35)$$

όπου $H = \nabla^2 f(\mathbf{x}_k)$.

Οπότε η νέα θέση θα προκύπτει από

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{h} \quad (4.36)$$

δηλαδή,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\nabla^2 f(\mathbf{x}_k))^{-1} \nabla f(\mathbf{x}) \quad (4.37)$$

Η διεύθυνση Newton είναι αξιόπιστη μόνο σε μικρά $\|\mathbf{h}\|$, ώστε το τετραγωνικό μοντέλο να είναι αντιπροσωπευτικό της f γύρω από το σημείο \mathbf{x} [Nocedal99].

Το $\nabla^2 f(\mathbf{x}_k)$ δίνεται από

$$\nabla^2 f(\mathbf{x}) = \mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) + \sum_{i=1}^m r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x}) \quad (4.38)$$

ενώ για εύκολο συμβολισμό η δεύτερη ποσότητα στο δεξί μέλος θα αναφέρεται ως :

$$\mathbf{S}(\mathbf{x}_k) = \sum_{i=1}^m r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x}) \quad (4.39)$$

Αντί να υπολογιστεί ο αντίστροφος μπορεί απλώς να λυθεί το εξής :

$$(\mathbf{J}(\mathbf{x}_k))^T \mathbf{J}(\mathbf{x}_k) + \mathbf{S}(\mathbf{x}_k) \mathbf{h} = -\mathbf{J}(\mathbf{x}_k)^T \mathbf{r}(\mathbf{x}_k) \quad (4.40)$$

Η διεύθυνση \mathbf{h}_N θα είναι μια διεύθυνση κατάβασης όσο το $\nabla^2 f(\mathbf{x}_k)$ είναι θετικά ορισμένο.

Εδώ το μέγεθος βήματος αφήνεται μονάδα σε κάθε επανάληψη, ωστόσο η μέθοδος Newton μπορεί να συνδυαστεί με μεθόδους line-search όταν δε δίνεται επαρκής μείωση στην f , ή και, ακόμα, όταν δεν ικανοποιείται η συνθήκη κατάβασης [Nocedal99].

Η μέθοδος Newton έχει μεγάλο ρυθμό σύγκλισης ενώ είναι γρήγορη στα τελευταία στάδια καθώς προσεγγίζεται η περιοχή που περιέχει τη λύση \mathbf{x}^* [Madsen04], [Nocedal99]. Υπάρχει καλή σύγκλιση, όμως, αυτή συνοδεύεται με ένα υπολογιστικό κόστος, αφού πρέπει να υπολογιστούν mn^2 παράγωγοι στον όρο S και για το λόγο αυτό δεν προτιμάται στα μη γραμμικά ελάχιστα τετράγωνα [Hansen12]. Ακόμα, μπορεί να μην υπάρχουν καν οι δεύτερες παράγωγοι, οπότε χρησιμοποιούνται προσεγγίσεις αυτών, δίνοντας τις μεθόδους quasi-Newton [Hansen12].

4.3.2 Υπολογισμός μεγέθους βήματος

Εφόσον εξαχθεί μια διεύθυνση κατάβασης με μια από τις παραπάνω μεθόδους, το επόμενο βήμα αυτών των αλγορίθμων είναι ο υπολογισμός ενός κατάλληλου βήματος, a , το οποίο προκύπτει λύνοντας το εξής πρόβλημα ελαχιστοποίησης [Madsen04]:

$$\min_a f(\mathbf{x} + a\mathbf{h}) \quad (4.41)$$

Χρησιμοποιώντας εκ νέου το θεώρημα Taylor η συνάρτηση $f(\mathbf{x} + a\mathbf{h})$ μπορεί να προσεγγιστεί ως

$$f(\mathbf{x} + a\mathbf{h}) \approx \phi(a) \equiv f(\mathbf{x}) + h^T \nabla f(\mathbf{x}) \quad (4.42)$$

όπου το $\phi(a)$ είναι μόνο συνάρτηση του a :

$$\phi(a) = f(\mathbf{x}) + a\mathbf{h}^T \nabla f(\mathbf{x}) \quad (4.43)$$

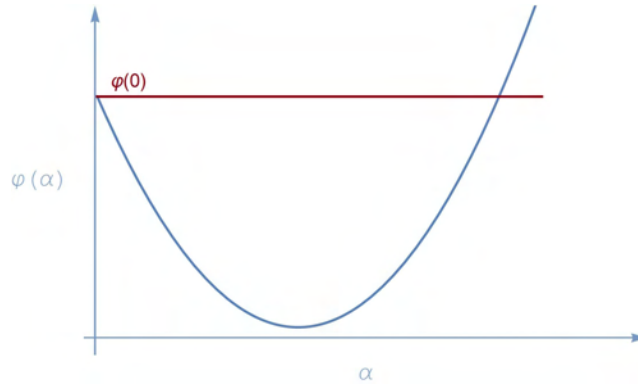
Αφού η \mathbf{h} είναι διεύθυνση κατάβασης, τότε :

$$\left. \frac{d\phi(a)}{da} \right|_{a=0} = \mathbf{h}^T \nabla f(\mathbf{x}) < 0 \quad (4.44)$$

ενώ για ικανοποιητικά μικρά a , θα ισχύει ότι:

$$\phi(a) < \phi(0) \quad (4.45)$$

Όπως φαίνεται στο σχήμα 4.3, μια τυπική μορφή αυτής της συνάρτησης μπορεί να είναι η εξής:



Σχήμα 4.3: Τυπική μορφή του μοντέλου συναρτήσεως του βήματος.

Φαίνεται ότι η επιλογή του κατάλληλου βήματος a περιορίζεται από:

- μικρές τιμές του a : καθώς μπορεί να μην επιφέρουν ικανοποιητική μείωση στην τιμή της συνάρτησης
- μεγάλες τιμές του a : καθώς μπορεί να μην ικανοποιείται πλέον η συνθήκη κατάβασης

Συνεπώς, πρέπει να επιλέγεται μια σωστή τιμή για το μέγεθος του βήματος, ώστε σε κάθε βήμα να ικανοποιείται η συνθήκη κατάβασης και να προκύπτει επαρκής μείωση στην f . Περισσότερες πληροφορίες σχετικά με τους τρόπους επιλογής βήματος μπορούν να βρεθούν στο [Nocedal99], [Madsen04].

Μέχρι τώρα, στις παραπάνω μεθόδους, επιλέγεται μια διεύθυνση \mathbf{h} στην οποία θα κινηθεί από τη τρέχουσα θέση \mathbf{x}_k στην επόμενη \mathbf{x}_{k+1} , διανύοντας μια συγκεκριμένη απόσταση η οποία προκύπτει από το μέγεθος βήματος a , ώστε στη νέα θέση να μειώνεται επαρκώς η τιμή της συνάρτησης. Ωστόσο, υπάρχει και μια άλλη στρατηγική για την εύρεση της νέας θέσης, η μέθοδος στην οποία υπάρχει μια περιοχή εμπιστοσύνης (trust-region) [Nocedal99].

4.4 Μέθοδοι περιοχής εμπιστοσύνης

Στη μέθοδο αυτή κατασκευάζεται ένα μοντέλο της συνάρτησης, η συμπεριφορά του οποίου γύρω από το τρέχον σημείο θα είναι παρόμοια με αυτή, ενώ υπάρχει ένα όριο μέσα στο οποίο θεωρείται αντιπροσωπευτικό. Δηλαδή υπάρχει μια περιοχή εμπιστοσύνης στην οποία το μοντέλο μπορεί να θεωρηθεί καλός προσεγγιστής της συνάρτησης. Η μέθοδος αυτή ονομάζεται έτσι για ακριβώς αυτόν το λόγο [Nocedal99], [Madsen04]. Οι αρχικές μελέτες για αυτές τις μεθόδους ξεκίνησαν από τον Powell το 1970 [Powell70] ενώ από τις επόμενες δεκαετίες λαμβάνει διαρκούς προσοχής από πολλούς ερευνητές [Yuan].

Το μοντέλο που προσεγγίζει τη συνάρτηση θα συμβολίζεται με L και συνήθως είναι τετραγωνικό. Προκύπτει από το ανάπτυγμα Taylor, όπου για μικρά $\|\mathbf{h}\|$ κανείς λαμβάνει:

$$f(\mathbf{x} + \mathbf{h}) \approx L(\mathbf{h}) \equiv f(\mathbf{x}) + \mathbf{h}^T \nabla f(\mathbf{x}) + \frac{1}{2} \mathbf{h}^T \mathbf{B} \mathbf{h} \quad (4.46)$$

ο πίνακας \mathbf{B} είναι συμμετρικός και περιέχει τις δεύτερες παραγώγους της f .

Στόχος, λοιπόν, είναι να βρεθεί η διεύθυνση \mathbf{h} , ώστε στη νέα θέση

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{h} \quad (4.47)$$

να μειώνεται η τιμή της συνάρτησης.

Το μοντέλο θα είναι καλό, δηλαδή θα είναι καλός προσεγγιστής της συνάρτησης, σε μια συγκεκριμένη περιοχή, η οποία περιγράφεται με ένα μικρό αριθμό Δ , όπου $\Delta > 0$. Γεωμετρικά, στις δύο

διαστάσεις, η περιοχή εμπιστοσύνης μπορεί να θεωρηθεί ως ένα κύκλος με κέντρο το \mathbf{x} και ακτίνα το Δ . Επομένως, ζητείται εκείνο το \mathbf{h} που να ελαχιστοποιεί το $L(\mathbf{h})$, λαμβάνοντας υπόψη τον προηγούμενο περιορισμό.

Το πρόβλημα μπορεί να μορφοποιηθεί ως:

$$\min_{\|\mathbf{h}\| < \Delta} L(\mathbf{h}) = \min_{\|\mathbf{h}\| < \Delta} f(\mathbf{x}) + \mathbf{h}^T \nabla f(\mathbf{x}) + \frac{1}{2} \mathbf{h}^T B \mathbf{h} \quad (4.48)$$

Αν το \mathbf{h} που προκύπτει οδηγεί σε μείωση της τιμής της συνάρτησης, τότε γίνεται αποδεκτό και ίσως είναι δυνατή η αύξηση του μεγέθους της περιοχής εμπιστοσύνης, δηλαδή αύξηση του Δ , ώστε τα μεγαλύτερα βήματα να οδηγήσουν με λιγότερες επαναλήψεις σε κάποια λύση. Ωστόσο, η αντίθετη περίπτωση, η οποία επέρχεται μετά από ένα μεγάλο \mathbf{h} , καθιστά το μοντέλο σε εκείνο το σημείο όχι καλό προσεγγιστή της συνάρτησης, οπότε πρέπει να συρρικνωθεί η περιοχή εμπιστοσύνης, δηλαδή να μειωθεί το Δ το οποίο θα οδηγήσει σε μικρότερα \mathbf{h} .

Αυτό ακριβώς επιτυγχάνεται με τον αλγόριθμο, όπου μετά την κάθε ενημέρωση της τρέχουσας θέσης μετράται η ποιότητα του μοντέλου, ώστε να επιλεγεί η κατάλληλη τιμή για το μέγεθος της περιοχής εμπιστοσύνης.

Ο εν λόγω αλγόριθμος περιέχει τα εξής βήματα:

Αλγόριθμος περιοχής εμπιστοσύνης

1. $k = 0, \mathbf{x}_0$
 2. **για** $k = 0, 1, 2, \dots$
 3. εύρεση διεύθυνσης \mathbf{h}

$$\min_{\|\mathbf{h}\| < \Delta} f(\mathbf{x}) + \mathbf{h}^T \nabla f(\mathbf{x}) + \frac{1}{2} \mathbf{h}^T B \mathbf{h}$$
 4. **αν** $f(\mathbf{x} + \mathbf{h}) < f(\mathbf{x})$
 5. ενημέρωση νέας θέσης

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{h}$$
 5. **τέλος**
 6. ενημέρωση Δ
 7. **τέλος**
-

Η ποιότητα του μοντέλου μετράται με έναν λόγο [Madsen04], το λόγο της πραγματικής μείωσης στην τιμή της συνάρτησης προς τη μείωση που προβλέπει το μοντέλο, όπου αναλόγως της τιμής που λαμβάνει, γίνεται και η κατάλληλη τροποποίηση του μεγέθους της περιοχής εμπιστοσύνης. Δηλαδή,

$$\rho_k = \frac{\text{πραγματική μείωση}}{\text{προβλεπόμενη μείωση}} \quad (4.49)$$

ή

$$\rho_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})}{L(\mathbf{0}) - L(\mathbf{h})} \quad (4.50)$$

όπου $L(\mathbf{0}) = f(\mathbf{x}_k)$. Ο παρονομαστής είναι εξορισμού θετικός, οπότε το πρόσημο του αριθμητή θα ορίζει το πρόσημο του λόγου ρ . Αρνητικός λόγος επέρχεται μετά από μεγάλα βήματα τα οποία οδηγούν σε

$$f(\mathbf{x}_k) < f(\mathbf{x}_{k+1}) \quad (4.51)$$

Το βήμα γίνεται αποδεκτό μόνο με θετικές τιμές του λόγου. Ωστόσο, οι μεταβολές του Δ επέρχονται σε ορισμένα πεδία τιμών. Μεγάλες τιμές του λόγου ρ υποδεικνύουν δυνατή την αύξηση του μεγέθους της περιοχής εμπιστοσύνης, ενώ μικρές τιμές, μείωση του μεγέθους και ανάπτυξη μικρότερων βημάτων.

ενημέρωση του Δ
αν $\rho < a$
$\Delta = \Delta/c_1$
αλλιώς αν $\rho > b$
$\Delta = c_2\Delta$
τέλος

Ακόμα, ο αλγόριθμος δεν είναι ευαίσθητος στις επιλογές των τιμών πάνω ή κάτω από τις οποίες μεταβάλλεται η τιμή του Δ , a και b αντίστοιχα, αν και πρέπει να επιλέγονται έτσι τα c_1 και c_2 , ώστε να μην υπόκεινται σε διακυμάνσεις οι τιμές του Δ [Madsen04]. Περισσότερες πληροφορίες σχετικά με τις μεθόδους trust-region μπορούν να βρεθούν στο [Yuan].

4.5 Μέθοδοι για την επίλυση μη γραμμικών προβλημάτων ελαχίστων τετραγώνων

Οι μέθοδοι που παρουσιάστηκαν έως τώρα μπορούν να χρησιμοποιηθούν για την επίλυση αυτού του είδους προβλημάτων, ωστόσο, όπως θα φανεί αργότερα, εφαρμόζονται αυτές που εκμεταλλεύονται τα ιδιαίτερα χαρακτηριστικά αυτών. Μια μέθοδος είναι αυτή που βασίζεται στην κλασική μέθοδο Newton και αποτελεί τη βάση για τις αποτελεσματικές μεθόδους που χρησιμοποιούνται σε μη γραμμικά προβλήματα ελαχίστων τετραγώνων. Η μέθοδος αυτή ονομάζεται μέθοδος Gauss-Newton [Madsen04].

4.5.1 Μέθοδος Gauss - Newton

Η μέθοδος Gauss - Newton συνιστάται για συναρτήσεις οι οποίες είναι άθροισμα τετραγώνων συναρτήσεων. Προτείνει ένα γραμμικό μοντέλο για την f , το οποίο προκύπτει από το ανάπτυγμα Taylor, όπου για μικρά $\|\mathbf{h}\|$, κανείς λαμβάνει

$$f(\mathbf{x} + \mathbf{h}) \approx l(\mathbf{h}) \equiv f(\mathbf{x}) + \mathbf{J}(\mathbf{x})\mathbf{h} \quad (4.52)$$

Πρόκειται για μία μέθοδο εναλλακτική της μεθόδου Newton, αφού δεν λαμβάνονται πλέον υπόψη οι δεύτερες παράγωγοι ενώ ο Εσσιανός του μοντέλου μπορεί να προσεγγιστεί ως:

$$\mathbf{H} \approx (\mathbf{J}(\mathbf{x}_k))^T \mathbf{J}(\mathbf{x}_k) \quad (4.53)$$

Με αντίστοιχο τρόπο κανείς καταλήγει στο εξής πρόβλημα

$$(\mathbf{J}(\mathbf{x}_k)^T \mathbf{J}(\mathbf{x}_k))\mathbf{h} = -\mathbf{J}(\mathbf{x}_k)^T \mathbf{r}(\mathbf{x}_k) \quad (4.54)$$

όπου πάλι

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{h} \quad (4.55)$$

Τελικά, κανείς πρέπει να λύσει το ακόλουθο γραμμικό πρόβλημα ελαχίστων τετραγώνων:

$$\min_{\mathbf{h}} \|\mathbf{A}\mathbf{h} - \mathbf{b}\| \quad (4.56)$$

όπου

$$\mathbf{A} = \mathbf{J}(\mathbf{x}_k), \quad \mathbf{b} = -(\mathbf{r}(\mathbf{x}_k)) \quad (4.57)$$

οπότε

$$\mathbf{h} = (\mathbf{J}(\mathbf{x}_k)^T \mathbf{J}(\mathbf{x}_k))^{-1} \mathbf{J}(\mathbf{x}_k) \mathbf{r}(\mathbf{x}_k) \quad (4.58)$$

Η διεύθυνση \mathbf{h} αποτελεί μια διεύθυνση κατάβασης για θετικά ορισμένο Εσσιανό ενώ πάλι, σε αντιστοιχία με τη μέθοδο Newton, θα είναι αξιόπιστη μόνο σε κοντινές αποστάσεις, λόγω του τοπικού χαρακτήρα ισχύος του μοντέλου. Το μέγεθος βήματος μπορεί να είναι τέτοιο, ώστε πλέον να μην αντιστοιχεί σε σημείο στο οποίο να ισχύει η συνθήκη κατάβασης. Η μέθοδος αυτή μπορεί να συνδυαστεί με μεθόδους line-search, ώστε να εγγυάται σύγκλιση, εφόσον ικανοποιούνται ορισμένες συνθήκες. Περισσότερες πληροφορίες μπορούν να βρεθούν στο [Madsen04].

Σε αυτή την περίπτωση, η νέα θέση θα δίνεται από:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + a\mathbf{h} \quad (4.59)$$

όπου το a υπολογίζεται σύμφωνα με τις απαιτήσεις προηγούμενης ενότητας.

Ο αλγόριθμος μπορεί να βελτιωθεί με τη χρήση ενός όρου απόσβεσης, ώστε να ελέγχεται το βήμα για να εξασφαλιστεί επαρκής μείωση στη συνάρτηση και να ισχύει η συνθήκη κατάβασης, οδηγώντας στη λεγόμενη μέθοδο Levenberg-Marquardt [Madsen04].

4.5.2 Μέθοδος Levenberg-Marquardt

Στη μέθοδο Gauss-Newton μπορεί το μέγεθος βήματος \mathbf{h} να γίνει πολύ μεγάλο, ώστε να οδηγηθεί ο αλγόριθμος σε «κακές» διευθύνσεις και να μην ισχύει το μοντέλο προσέγγισης της f . Για το λόγο αυτό, εισάγεται στο γραμμικό πρόβλημα ελαχιστοποίησης ένας περιορισμός για το μέγεθος του βήματος, δηλαδή θα υπάρχει πλέον ένα άνω όριο στο μέγεθος του βήματος. Με την τεχνική αυτή βελτιώνεται η ποιότητα του βήματος. Πρόκειται ουσιαστικά για μια μέθοδο trust region και συγκεκριμένα για μια εκδοχή της μεθόδου Gauss-Newton [Hansen12] [Madsen04], η οποία ορίστηκε αρχικά από τον Levenberg το 1944 [Levenberg44] και αργότερα από τον Marquardt το 1963 [Marquardt63].

Σε αναλογία με πριν, το πρόβλημα μορφοποιείται ως εξής:

$$\min_{\|\mathbf{h}\| < c} \|\mathbf{r}(\mathbf{x}_k) + \mathbf{J}(\mathbf{x}_k)\mathbf{h}\|^2 \quad (4.60)$$

Ο περιορισμός αυτός μπορεί να εισαχθεί με τη χρήση πολλαπλασιαστή Langrange, έστω λ , στην αντικειμενική συνάρτηση, ώστε το πρόβλημα ελαχιστοποίησης πλέον να είναι [Hansen12]:

$$\min_{\mathbf{h}} \|\mathbf{J}(\mathbf{x}_k)\mathbf{h} + \mathbf{r}(\mathbf{x}_k)\|^2 + \lambda \|\mathbf{h}\|^2 \quad (4.61)$$

Το παραπάνω έχει δύο ισοδύναμες μορφές:

$$(\mathbf{J}(\mathbf{x}_k)^T \mathbf{J}(\mathbf{x}_k) + \lambda_k I)\mathbf{h} = -\mathbf{J}(\mathbf{x}_k)^T \mathbf{r}(\mathbf{x}_k) \quad (4.62)$$

ή

$$\min_{\mathbf{h}} \left\| \begin{pmatrix} \mathbf{J}(\mathbf{x}_k) \\ \lambda_k I \end{pmatrix} \mathbf{h} - \begin{pmatrix} -\mathbf{r}(\mathbf{x}_k) \\ \mathbf{0} \end{pmatrix} \right\|_2^2 \quad (4.63)$$

Με τη μέθοδο αυτή κανείς διαχειρίζεται και περιπτώσεις στις οποίες η Ιακωβιανή J δεν είναι πλήρους τάξης (rank-deficient) ή ιδιόμορφα προβλήματα (ill-conditioned), αφού η ποσότητα στο αριστερό μέλος που πολλαπλασιάζεται με το \mathbf{h} θα είναι πάντα θετικά ορισμένη για $\lambda > 0$. Η διεύθυνση \mathbf{h} θα είναι διεύθυνση κατάβασης και η μέθοδος θα είναι καλώς ορισμένη [Hansen12].

Η μέθοδος Levenberg-Marquardt μπορεί να θεωρηθεί ως συνδυασμός της μεθόδου Newton και της μεθόδου απότομης κατάβασης που παρουσιάστηκαν νωρίτερα. Η παράμετρος λ καθορίζει τόσο τη διεύθυνση όσο και το μέγεθος του βήματος, ώστε για μικρές τιμές της παραμέτρου να αποκτάται το βήμα Newton, αφού η διεύθυνση θα είναι ανάλογη του $\sim \frac{1}{\lambda}$, ενώ για μεγάλες τιμές της παραμέτρου η ενημέρωση της θέσης να πραγματοποιείται με βήμα απότομης κατάβασης [Madsen04]. Η εναλλαγή

αυτή είναι σημαντική καθώς είναι βολικό να χρησιμοποιείται η μέθοδος απότομης κατάβασης με αργά και σταθερά βήματα στις αρχικές επαναλήψεις του αλγορίθμου, ενώ στα τελευταία στάδια, όταν η λύση θα βρίσκεται στη γειτονιά ενός τοπικού βέλτιστου, να αλλάζει σε Newton για να επιτυγχάνονται μεγάλα βήματα και, επομένως, γρήγορη σύγκλιση.

Ο Marquardt όρισε για πρώτη φορά μια μεθοδολογία για την επιλογή της παραμέτρου λ_k η οποία βασίζεται σε ένα λόγο, το λόγο ενίσχυσης ρ . Αφορά το λόγο μεταξύ της πραγματικής μείωσης προς τη προβλεπόμενη μείωση [Hansen12].

Δηλαδή,

$$\rho_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})}{f(\mathbf{x}_k) - \frac{1}{2} \|\mathbf{J}(\mathbf{x}_k)\mathbf{h} + \mathbf{r}(\mathbf{x}_k)\|^2} \quad (4.64)$$

Ο παρονομαστής είναι εξορισμού θετικός και υποδηλώνει τη μείωση στην f από το τοπικό γραμμικό μοντέλο.

Η λογική για την επιλογή του λ είναι η εξής: αν το μοντέλο δεν είναι καλός προσεγγιστής της f , δηλαδή αν ο λόγος ρ είναι μικρός, ή, ακόμα, και αρνητικός, τότε πρέπει να μειωθεί το εύρος εμπιστοσύνης, δηλαδή να χρησιμοποιηθεί ένα μικρότερο βήμα, ή, διαφορετικά, να αυξηθεί το λ . Σε μεγάλες τιμές του λόγου η προσέγγιση θεωρείται καλή, οπότε το βήμα μπορεί να αυξηθεί, ή, διαφορετικά, να μειωθεί το λ .

Η ενημέρωση του λ μαζί με τις τυπικές τιμές των σταθερών γίνεται ως εξής:

ενημέρωση του λ
αν $\rho < 0.25$ $\lambda = 2\lambda$
αλλιώς αν $\rho > 0.75$ $\lambda = \lambda/3$
τέλος αν $\rho > 0$ ενημέρωσε το \mathbf{x}
τέλος

4.6 Κριτήρια τερματισμού επαναληπτικών μεθόδων

Τα κριτήρια τερματισμού για τις επαναληπτικές μεθόδους μπορεί να είναι τα εξής [Hansen12]:

- σύγκλιση ακολουθίας: $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \leq \varepsilon_1$
- έλεγχος τιμής συνάρτησης: $\|f(\mathbf{x}_{k+1})\| \leq \varepsilon_2$
- μέτρο κλίσης συνάρτησης: $\|\nabla f(\mathbf{x}_{k+1})\| \leq \varepsilon_3$
- μέγιστες επαναλήψεις: $k > k_{max}$

όπου ε_i τα αντίστοιχα όρια ανοχής (μικροί, θετικοί αριθμοί) και οι μέγιστες επαναλήψεις k_{max} επιλέγονται από το χρήστη.

4.7 Ειδική κατηγορία προβλημάτων μη γραμμικών ελαχίστων τετραγώνων

Μια ειδική κατηγορία προβλημάτων μη γραμμικών ελαχίστων τετραγώνων (non linear least squares, NLLS) είναι αυτή στην οποία η συνάρτηση f που μελετάται είναι γραμμικός συνδυασμός μη γραμμικών συναρτήσεων [Hansen12], με το μοντέλο M να είναι της μορφής:

$$M(\boldsymbol{\omega}, \boldsymbol{\alpha}, t) = \sum_{j=1}^n \omega_j \phi_j(\boldsymbol{\alpha}, t) \quad (4.65)$$

όπου $\boldsymbol{\omega} \in \mathbb{R}^n$ και $\boldsymbol{\alpha} \in \mathbb{R}^k$ είναι οι παράμετροι του προβλήματος, με τις συναρτήσεις ϕ να είναι μη γραμμικές.

Η κατηγορία αυτή συμπίπτει με την εκπαίδευση νευρωνικών δικτύου RBF που παρουσιάστηκε στο προηγούμενο κεφάλαιο, αφού για ένα συγκεκριμένο παράδειγμα εκπαίδευσης i το υπόλοιπο ορίζεται ως:

$$r_i(\mathbf{c}, \boldsymbol{\omega}) = y_i - f(\mathbf{x}_i, \mathbf{c}, \boldsymbol{\omega}) = y_i - \sum_{j=1}^N w_j \phi(\mathbf{c}_j, \mathbf{x}_i), \quad i = 1, \dots, m \quad (4.66)$$

όπου $w_j \in \mathbb{R}$, $\mathbf{c}_j \in \mathbb{R}^n$, $\mathbf{x}_i \in \mathbb{R}^n$, τα βάρη, τα κέντρα του δικτύου και τα δεδομένα εισόδου αντίστοιχα, ωστόσο θα ακολουθηθεί ο γενικός συμβολισμός με τις άγνωστες παραμέτρους να συμβολίζονται ως $\boldsymbol{\omega} \in \mathbb{R}^n$ οι γραμμικές και $\boldsymbol{\alpha} \in \mathbb{R}^k$ οι μη γραμμικές, με $t_i \in \mathbb{R}$, αφού ακολουθείται παρόμοια διαδικασία.

Οπότε,

$$r_i(\boldsymbol{\omega}, \boldsymbol{\alpha}) = y_i - M(\boldsymbol{\omega}, \boldsymbol{\alpha}, t_i), \quad i = 1, \dots, m \quad (4.67)$$

ή

$$r_i(\boldsymbol{\omega}, \boldsymbol{\alpha}) = y_i - \sum_{j=1}^n \omega_j \phi_j(\boldsymbol{\alpha}, t), \quad i = 1, \dots, m \quad (4.68)$$

και σε μητρική μορφή

$$\mathbf{r}(\boldsymbol{\omega}, \boldsymbol{\alpha}) = \mathbf{y} - \Phi(\boldsymbol{\alpha})\boldsymbol{\omega} \quad (4.69)$$

με τον πίνακα Φ να είναι ένας πίνακας με διαστάσεις $m \times n$.

Κανείς καταλήγει στο εξής πρόβλημα μη γραμμικών ελαχίστων τετραγώνων:

$$\min_{\boldsymbol{\omega}, \boldsymbol{\alpha}} \frac{1}{2} \|\mathbf{y} - \Phi(\boldsymbol{\alpha})\boldsymbol{\omega}\|^2 \quad (4.70)$$

4.7.1 Μέθοδος Variable Projection

Ένας αλγόριθμος που να διαχειρίζεται τέτοιου είδους προβλήματα ορίστηκε από τους Golub και Pereyra το 1973, κινούμενοι στα πλαίσια της δουλειάς των Guttman και Scolnik [Guttman73], [Golub73], [Golub03]. Για την επίλυση του μη γραμμικού προβλήματος ελαχίστων τετραγώνων, όπου το μοντέλο είναι ένας γραμμικός συνδυασμός μη γραμμικών συναρτήσεων, αξιοποιείται η ρητή σύζευξη μεταξύ των γραμμικών $\boldsymbol{\omega}$ και μη γραμμικών παραμέτρων $\boldsymbol{\alpha}$, ώστε το αρχικό πρόβλημα να αναχθεί σε δύο υποπροβλήματα, τα οποία λύνονται διαδοχικά [Hansen12], [Kaufman75].

Το ένα υποπρόβλημα θα είναι ένα μη γραμμικό πρόβλημα ελαχίστων τετραγώνων για την εύρεση των μη γραμμικών παραμέτρων. Συνήθως είναι πιο περίπλοκο, ωστόσο περιλαμβάνει μόνο τις μη γραμμικές παραμέτρους, οπότε ο χώρος των αγνώστων παραμέτρων θα είναι μικρότερος από τον αρχικό. Το δεύτερο υποπρόβλημα θα είναι ένα απλό, γραμμικό πρόβλημα ελαχίστων τετραγώνων για την εύρεση των γραμμικών παραμέτρων. Δηλαδή η ύπαρξη δύο συνόλων παραμέτρων, ενός γραμμικού και ενός μη γραμμικού, τα οποία εξαρτώνται το ένα από το άλλο, οδηγεί στη ρητή εξάλειψη του ενός και κατόπιν στην ελαχιστοποίηση του εναπομείνοντος συναρτησιακού, το οποίο εξαρτάται μόνο από τις εναπομείνουσες

μεταβλητές ενώ οδηγεί σε ένα καλύτερο ορισμένο πρόβλημα (better-conditioned) στο οποίο μπορεί να υπάρχουν καλύτεροι χρόνοι σύγκλισης [Golub03].

Η ιδέα είναι η εξής [Golub73]:

Στόχος είναι η εύρεση των παραμέτρων ω, α που να ελαχιστοποιούν το συναρτησιακό:

$$r(\omega, \alpha) = \mathbf{y} - \Phi(\alpha)\omega \quad (4.71)$$

Δηλαδή το πρόβλημα μορφοποιείται ως:

$$\min_{\omega, \alpha} \frac{1}{2} \|\mathbf{y} - \Phi(\alpha)\omega\|^2 \quad (4.72)$$

Αν με κάποιο τρόπο είναι γνωστές οι μη γραμμικές παράμετροι του προβλήματος α , τότε

$$\min_{\omega} \frac{1}{2} \|\mathbf{y} - \Phi(\alpha)\omega\|^2 \quad (4.73)$$

με τη λύση του παραπάνω προβλήματος να προκύπτει από τις κανονικές εξισώσεις:

$$\omega = \Phi(\alpha)^+ \mathbf{y} \quad (4.74)$$

όπου ο πίνακας $\Phi(\alpha)^+$ είναι ο Moore-Penrose γενικευμένος ψευδοαντίστροφος του $\Phi(\alpha)$

$$\Phi(\alpha)^+ = (\Phi(\alpha)^T \Phi(\alpha))^{-1} \Phi(\alpha)^T \quad (4.75)$$

Άρα, με αντικατάσταση στο αρχικό πρόβλημα

$$\min_{\alpha} \frac{1}{2} \|\mathbf{y} - \Phi(\alpha)\Phi(\alpha)^+ \mathbf{y}\|^2 \quad (4.76)$$

και με ανακατάταξη των όρων προκύπτει το εξής μη γραμμικό πρόβλημα ελαχίστων τετραγώνων:

$$\min_{\alpha} \frac{1}{2} \|(I - \Phi(\alpha)\Phi(\alpha)^+) \mathbf{y}\|^2 \quad (4.77)$$

Δηλαδή μετά την αντικατάσταση των γραμμικών παραμέτρων ω , πλέον το συναρτησιακό θα είναι μόνο συνάρτηση των μη γραμμικών παραμέτρων α .

Ο όρος μέσα στην παρένθεση συμβολίζεται με $r_2(\alpha)$,

$$r_2(\alpha) = (I - \Phi(\alpha)\Phi(\alpha)^+) \mathbf{y} \quad (4.78)$$

και καλείται προβολή μεταβλητής - variable projection (VP) του \mathbf{y} .

Στο σημείο αυτό θα χρειαστούν ορισμένα στοιχεία από τη θεωρία της γραμμικής άλγεβρας προσαρμοσμένα από το βιβλίο του [Gilbert08] για την καλύτερη κατανόηση κάποιων σημείων της μεθόδου.

Για την επίλυση ενός μη συμβιβαστού συστήματος $A\mathbf{x} = \mathbf{b}$, όπου ο πίνακας A είναι $m \times n$, δεν υπάρχει ένα $\mathbf{x} \in \mathbb{R}^n$ που να ταιριάζει ακριβώς με τα δεδομένα $\mathbf{b} \in \mathbb{R}^m$. Διαφορετικά, το διάνυσμα \mathbf{b} δεν θα ανήκει στο χώρο στηλών του A , το οποίο συμβολίζεται με $R(A)$. Υπενθυμίζεται ότι ο χώρος στηλών ενός πίνακα A $m \times n$ αποτελείται από όλους τους συνδυασμούς στηλών του A , είναι υποχώρος του \mathbb{R}^m και θεωρείται ένας από τους τέσσερις θεμελιώδεις χώρους.

Για το λόγο αυτό, πρέπει να υπολογιστεί ένα διάνυσμα $\bar{\mathbf{x}}$ για το οποίο θα ελαχιστοποιείται το σφάλμα $E = \|A\bar{\mathbf{x}} - \mathbf{b}\|$, ή, διαφορετικά, η απόσταση του \mathbf{b} από το χώρο στηλών του A . Το πλησιέστερο σημείο στο \mathbf{b} , το οποίο θα ανήκει στο χώρο στηλών του A , είναι η προβολή του \mathbf{b} , η οποία θα συμβολίζεται με \mathbf{p} , πάνω στο χώρο στηλών του A .

Η γραφική απεικόνιση είναι πάντα χρήσιμη, οπότε θα χρησιμοποιηθεί στη συνέχεια με ένα παράδειγμα στις τρεις διαστάσεις του χώρου.

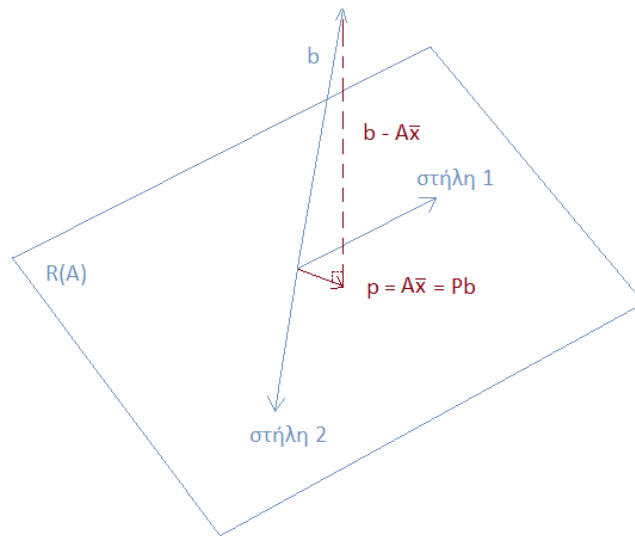
Έστω ότι ο A είναι ένας πίνακας 3×2

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \quad (4.79)$$

με τα δεδομένα b

$$b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad (4.80)$$

Ο χώρος στηλών του A και η προβολή του b πάνω στο χώρο αυτόν απεικονίζονται στο σχήμα 4.4.:



Σχήμα 4.4: Προβολή στο χώρο στηλών ενός πίνακα 3 επί 2.

Ένας άλλος θεμελιώδης χώρος είναι ο αριστερός μηδενοχώρος του A . Ο μηδενοχώρος του A περιέχει όλες τις λύσεις του $Ax = 0$, συμβολίζεται με $N(A)$ και είναι υποχώρος του \mathbb{R}^n . Επομένως, ο αριστερός μηδενοχώρος του A θα είναι ο μηδενοχώρος του A^T , συμβολίζεται με $N(A^T)$ και θα περιέχει όλες τις λύσεις του $A^T y = 0$ ή $y^T A = 0$ ($y \in \mathbb{R}^m$) και είναι υποχώρος του \mathbb{R}^m .

Ένα σημαντικό στοιχείο που προκύπτει από τα παραπάνω είναι ότι ο αριστερός μηδενοχώρος και ο χώρος στηλών του A είναι ορθογώνιοι υποχώροι του \mathbb{R}^m .

Προκύπτει από τον ορισμό του αριστερού μηδενοχώρου:

$$N(A^T) = \{y^T A = 0, y \in \mathbb{R}^m\} \quad (4.81)$$

Σε μητρική μορφή:

$$y^T \begin{bmatrix} \sigma & \sigma & \cdots & \sigma \\ \tau & \tau & \cdots & \tau \\ \eta & \eta & \cdots & \eta \\ \lambda & \lambda & \cdots & \lambda \\ \eta & \eta & \cdots & \eta \end{bmatrix} = 0 \quad (4.82)$$

ή

$$\begin{aligned} \mathbf{y}^T \cdot (\text{στήλη } 1) &= 0 \\ \mathbf{y}^T \cdot (\text{στήλη } 2) &= 0 \\ &\vdots \\ \mathbf{y}^T \cdot (\text{στήλη } n) &= 0 \end{aligned} \tag{4.83}$$

οπότε πράγματι ο αριστερός μηδενοχώρος είναι ορθογώνιος στο χώρο στηλών του A .

Πίσω στο σχήμα, λοιπόν, από όπου προκύπτει το εξής: το διάνυσμα $\mathbf{b} - A\bar{\mathbf{x}}$ είναι κάθετο στο χώρο στηλών του A και επειδή το ορθογώνιο συμπλήρωμα του χώρου στηλών του A είναι ο αριστερός μηδενοχώρος του A , τότε το διάνυσμα $\mathbf{b} - A\bar{\mathbf{x}}$ περιέχεται στο χώρο αυτό, δηλαδή

$$A^T(\mathbf{b} - A\bar{\mathbf{x}}) = 0 \tag{4.84}$$

ή

$$A^T A\bar{\mathbf{x}} = A^T \mathbf{b} \tag{4.85}$$

είναι οι «κανονικές» εξισώσεις που επιλύουν το $\|\mathbf{b} - A\mathbf{x}\|$. Προέρχεται από το αγγλικό "normal equations" για ακριβώς αυτό το λόγο, καθώς το διάνυσμα του σφάλματος πρέπει να είναι κάθετο στο χώρο στηλών του A .

Τελικά, προκύπτει το ζητούμενο $\bar{\mathbf{x}}$

$$\bar{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} \tag{4.86}$$

ενώ το πλησιέστερο σημείο στο \mathbf{b} το οποίο ανήκει στο χώρο στηλών του A είναι το

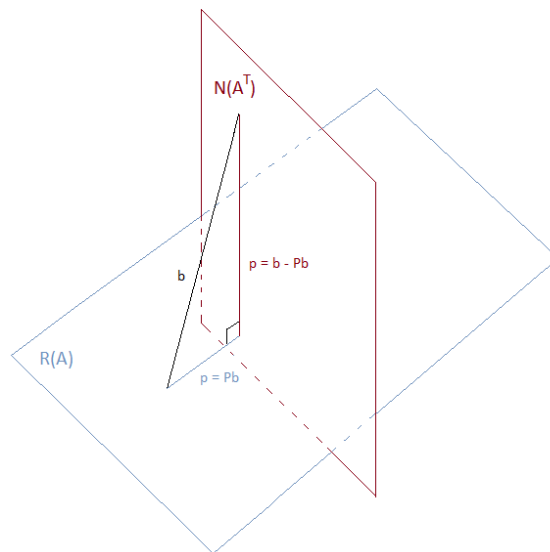
$$\mathbf{p} = A\bar{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{b} \tag{4.87}$$

Ο πίνακας

$$P = A(A^T A)^{-1} A^T = AA^+ \tag{4.88}$$

ονομάζεται πίνακας προβολής και προβάλλει ένα διάνυσμα \mathbf{b} στο χώρο στηλών του A ενώ ο A^+ είναι ο Moore-Penrose ψευδοαντίστροφος του A .

Δηλαδή το \mathbf{p} θα είναι συνιστώσα του \mathbf{b} στο χώρο στηλών του A , με το $\mathbf{b} - P\mathbf{b}$ να είναι η συνιστώσα του \mathbf{b} στον αριστερό μηδενοχώρο του A που είναι το ορθογώνιο στο χώρο στηλών του A , όπως φαίνεται στο σχήμα 4.5.



Σχήμα 4.5: Χώρος στηλών ενός πίνακα 3 επί 2 με το ορθογώνιο συμπλήρωμα του.

Επομένως, και ο πίνακας $I - P$ είναι ένας πίνακας προβολής που προβάλλει το \mathbf{b} στο ορθογώνιο συμπλήρωμα του χώρου στηλών του A , ο οποίος συμβολίζεται με P^\perp και για τον οποίο ισχύει ότι:

$$P^\perp = I - P \quad (4.89)$$

Άρα, στην προκειμένη περίπτωση, όπου το συναρτησιακό προς ελαχιστοποίηση είναι της μορφής

$$\mathbf{r}_2(\boldsymbol{\alpha}) = (I - \Phi(\boldsymbol{\alpha})\Phi(\boldsymbol{\alpha})^+) \mathbf{y} \quad (4.90)$$

ο όρος μέσα στην παρένθεση

$$I - \Phi(\boldsymbol{\alpha})\Phi(\boldsymbol{\alpha})^+ \quad (4.91)$$

θα συμβολίζεται με $P_{\Phi(\boldsymbol{\alpha})}^\perp$ και θα είναι ο προβολέας στο ορθογώνιο συμπλήρωμα του χώρου στηλών του $\Phi(\boldsymbol{\alpha})$, δηλαδή στον αριστερό μηδενοχώρο του $\Phi(\boldsymbol{\alpha})$. Σε αυτό, λοιπόν, οφείλεται η ονομασία της εν λόγω μεθόδου: προβολή μεταβλητής - variable projection. Το $\mathbf{r}_2(\boldsymbol{\alpha})$ από εδώ και πέρα θα καλείται συναρτησιακό VP.

Επομένως, με τη μέθοδο variable projection η επίλυση του αρχικού συναρτησιακού

$$\min_{\boldsymbol{\omega}, \boldsymbol{\alpha}} \frac{1}{2} \|\mathbf{r}_2(\boldsymbol{\alpha})\|^2 = \min_{\boldsymbol{\omega}, \boldsymbol{\alpha}} \frac{1}{2} \|\mathbf{y} - \Phi(\boldsymbol{\alpha})\boldsymbol{\omega}\|^2 \quad (4.92)$$

συνίσταται στην επίλυση του μειωμένου συναρτησιακού VP

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \|(I - \Phi(\boldsymbol{\alpha})\Phi(\boldsymbol{\alpha})^+) \mathbf{y}\|^2 \quad (4.93)$$

και στην επίλυση του

$$\boldsymbol{\omega} = \Phi(\boldsymbol{\alpha})^+ \mathbf{y} \quad (4.94)$$

4.7.2 Μέθοδοι για την επίλυση της ελαχιστοποίησης του μειωμένου συναρτησιακού Variable Projection

Το μη γραμμικό πρόβλημα ελαχίστων τετραγώνων (4.93) μπορεί να λυθεί με τις γενικές μεθόδους αριθμητικής βελτιστοποίησης [Golub03]. Κάποιες από τις μεθόδους που μπορούν να χρησιμοποιηθούν είναι η κλασική μέθοδος Newton, ή και οι εκδοχές τις, καθώς και η μέθοδος Gauss-Newton, ή η μέθοδος Levenberg-Marquardt, παραλλαγή της προηγούμενης [Hansen12]. Οι μέθοδοι διακρίνονται σε αυτές που δεν χρησιμοποιούν παραγώγους και σε αυτές που χρησιμοποιούν. Περισσότερες πληροφορίες μπορούν να βρεθούν στο [Golub73], [Golub03].

Στη συγκεκριμένη εργασία θα χρησιμοποιηθούν οι μέθοδοι που απαιτούν υπολογισμό της παραγώγου και που παρουσιάστηκαν προηγουμένως για το εξής πρόβλημα ελαχιστοποίησης:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \|\mathbf{r}_2(\boldsymbol{\alpha})\|^2 = \min_{\boldsymbol{\alpha}} \frac{1}{2} \|(I - \Phi(\boldsymbol{\alpha})\Phi(\boldsymbol{\alpha})^+) \mathbf{y}\|^2 \quad (4.95)$$

όπου πλέον το συναρτησιακό VP θα συμβολίζεται ως:

$$\mathbf{R}(\boldsymbol{\alpha}) = P_{\Phi(\boldsymbol{\alpha})}^\perp \mathbf{y} \quad (4.96)$$

Σύμφωνα με τους Golub και Pereyra [Golub73] η Ιακωβιανή του $\mathbf{R}(\boldsymbol{\alpha})$ δίνεται από

$$DP_{\Phi(\boldsymbol{\alpha})}^\perp \mathbf{y} = -DP_{\Phi(\boldsymbol{\alpha})} \mathbf{y} \quad (4.97)$$

όπου

$$DP_{\Phi(\boldsymbol{\alpha})} \mathbf{y} = \left(P_{\Phi(\boldsymbol{\alpha})}^\perp D(\Phi(\boldsymbol{\alpha}))\Phi(\boldsymbol{\alpha})^+ + (P_{\Phi(\boldsymbol{\alpha})}^\perp D(\Phi(\boldsymbol{\alpha}))\Phi(\boldsymbol{\alpha})^+)^T \right) \mathbf{y} \quad (4.98)$$

Η Kaufman [Kaufman75] αγνοεί το δεύτερο όρο προσεγγίζοντας την ιακωβιανή ως:

$$D\mathbf{R}(\boldsymbol{\alpha}) \approx - \left(P_{\Phi(\boldsymbol{\alpha})}^\perp D(\Phi(\boldsymbol{\alpha}))\Phi(\boldsymbol{\alpha})^+ \right) \mathbf{y} \quad (4.99)$$

εξασφαλίζοντας συντομότερους υπολογιστικούς χρόνους, χωρίς όμως την ιδιαίτερη αύξηση στον αριθμό των επαναλήψεων [Golub03].

Κεφάλαιο 5

Προτεινόμενη μέθοδος εκπαίδευσης δικτύων RBF

Στο κεφάλαιο 3 έγινε γνωστό ότι η εκπαίδευση ενός νευρωνικού δικτύου RBF συνίσταται στον υπολογισμό των εξής παραμέτρων:

- αριθμό νευρώνων του κρυφού επιπέδου, δηλαδή αριθμό των μονάδων RBF
- θέσεις κέντρων των μονάδων RBF (σε μερικές συναρτήσεις βάσης και τα εύρη)
- τα βάρη σύνδεσης του κρυφού επιπέδου με το επίπεδο εξόδου

Τα κέντρα προέρχονται από τις τεχνικές *k-means* και *fuzzy means* που παρουσιάστηκαν στο εν λόγω κεφάλαιο, οι οποίες αποτελούν μεθόδους μη εποπτευόμενης μάθησης καθώς δε λαμβάνεται υπόψη η έξοδος. Στις συναρτήσεις RBF που χαρακτηρίζονται και από μια τιμή εύρους χρησιμοποιείται η ευρετική μέθοδος με τους *p*-κοντινότερους γείτονες.

Ένας τρόπος για να ληφθεί υπόψη η έξοδος και να είναι εγγυημένο το ολικό βέλτιστο για τα βάρη είναι να σκεφτεί κανείς τη διαδικασία της εκπαίδευσης ως ένα πρόβλημα βελτιστοποίησης και να το μορφοποιήσει με ένα κλασικό αλγόριθμο βελτιστοποίησης, παράμετροι του οποίου θα είναι τα κέντρα του δικτύου. Δηλαδή σε κάθε επανάληψη του αλγορίθμου θα ανανεώνονται τα κέντρα βάσει ενός κόστους και στη συνέχεια τα βάρη μέσω ελαχίστων τετραγώνων για τους λόγους που αναφέρθηκαν.

Η εν λόγω μέθοδος θα περιλαμβάνει χονδρικά τα παρακάτω: θα ξεκινάει από μια αρχική τιμή για τα κέντρα, θα υπολογίζει τα βάρη με ελάχιστα τετράγωνα ενώ σε κάθε επανάληψη θα ανανεώνει τα κέντρα βάσει μιας τεχνικής βελτιστοποίησης και θα υπολογίζει εκ νέου τα βάρη με ελάχιστα τετράγωνα.

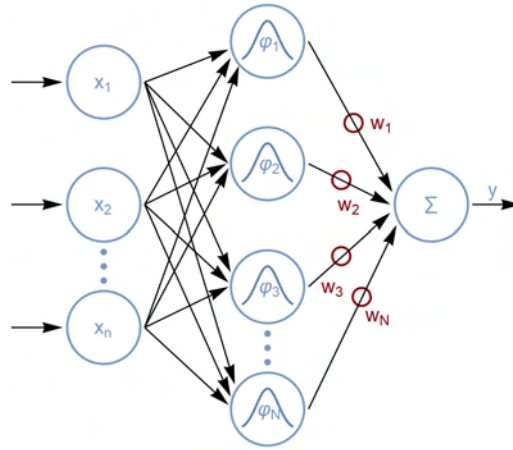
Η μέθοδος που χρησιμοποιείται στην παρούσα εργασία θα περιγραφεί στη συνέχεια, ενώ πρώτα αναγκαία θεωρείται η σύνδεση της με τις βασικές έννοιες των νευρωνικών δικτύων RBF που παρουσιάστηκαν νωρίτερα, ώστε να υπάρχει μια συνέχεια για τον αναγνώστη για την καλύτερη κατανόηση αυτής.

5.1 Μορφοποίηση του προβλήματος της εκπαίδευσης ενός δικτύου RBF

Σε ένα δίκτυο RBF (σχήμα 5.1) στόχος της εκπαίδευσης είναι να βρεθούν τα κέντρα, τα εύρη των συναρτήσεων βάσης και τα βάρη του δικτύου, ώστε η συνάρτηση f να ταιριάζει με τα δεδομένα εξόδου \mathbf{y} για τα συγκεκριμένα δεδομένα εισόδου \mathbf{x} .

Για το λόγο αυτό, απαιτείται η ελαχιστοποίηση του

$$\min_{\mathbf{c}, \mathbf{w}} e(\mathbf{c}, \mathbf{w}) = \min_{\mathbf{c}, \mathbf{w}} \frac{1}{2} \sum_{i=1}^m r_i(\mathbf{c}, \mathbf{w})^2 \quad (5.1)$$



Σχήμα 5.1: Τυπική δομή ενός δικτύου RBF.

όπου το i -οστό υπόλοιπο για το i -οστό παράδειγμα εκπαίδευσης είναι

$$r_i(\mathbf{c}, \mathbf{w}) = y_i - f(\mathbf{x}_i, \mathbf{c}, \mathbf{w}) = y_i - \sum_{j=1}^N w_j \phi(\mathbf{x}_i, \mathbf{c}_j), \quad i = 1, \dots, m \quad (5.2)$$

ή σε μητρική μορφή

$$\min_{\mathbf{c}, \mathbf{w}} e(\mathbf{c}, \mathbf{w}) = \min_{\mathbf{c}, \mathbf{w}} \frac{1}{2} \|\mathbf{y} - \Phi(\mathbf{c})\mathbf{w}\|_2^2 \quad (5.3)$$

Από τις συναρτήσεις ϕ που παρουσιάστηκαν νωρίτερα, επιλέγεται η συνάρτηση Gaussian, μιας και είναι συνεχώς παραγωγίσιμη, της μορφής:

$$\phi(\mathbf{x}, \mathbf{c}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|_2^2}{2\sigma^2}\right) \quad (5.4)$$

στην οποία η ποσότητα στον αριθμητή του εκθετικού είναι η τετραγωνισμένη Ευκλείδεια απόσταση, η οποία για το νευρώνα j και για ένα συγκεκριμένο παράδειγμα εκπαίδευσης k υπολογίζεται ως:

$$\|\mathbf{x}_k - \mathbf{c}_j\|_2 = \sqrt{\sum_{i=1}^n (x_{k,i} - c_{j,i})^2}, \quad k = 1, 2, \dots, m \quad (5.5)$$

ενώ η ποσότητα αυτή ορίζει τη δραστηριότητα του νευρώνα j .

Η απόδοση του δικτύου μετράται σύμφωνα με την τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Square Error, RMSE) το οποίο ορίζεται ως:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \bar{y}_i)^2} \quad (5.6)$$

όπου y_i είναι οι πραγματικές τιμές και \bar{y}_i οι προβλεπόμενες από το μοντέλο.

Ακόμα, για του λόγους που αναφέρθηκαν στο κεφάλαιο 2, χρησιμοποιείται και ένας σταθερός όρος (bias term). Πρόκειται στην ουσία για έναν επιπλέον νευρώνα στο κρυφό επίπεδο, ο οποίος θα δίνει σταθερή έξοδο ίση με 1. Αυτό δηλώνεται απλά με προσθήκη μιας μοναδιαίας στήλης στον πίνακα Φ , δηλαδή:

$$\Phi = \begin{bmatrix} \phi(\mathbf{x}_1, \mathbf{c}_1) & \phi(\mathbf{x}_1, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_1, \mathbf{c}_N) & 1 \\ \phi(\mathbf{x}_2, \mathbf{c}_1) & \phi(\mathbf{x}_2, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_2, \mathbf{c}_N) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi(\mathbf{x}_m, \mathbf{c}_1) & \phi(\mathbf{x}_m, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_m, \mathbf{c}_N) & 1 \end{bmatrix} \quad (5.7)$$

5.2 Εφαρμογή της μεθόδου Variable Projection

Επειδή το μοντέλο είναι ένας γραμμικός συνδυασμός μη γραμμικών συναρτήσεων, το πρόβλημα εκπαίδευσης ενός δικτύου RBF (εξ. 5.3) θα συνιστά ένα μη γραμμικό πρόβλημα ελαχίστων τετραγώνων για το οποίο μπορεί να εφαρμοστεί η μέθοδος Variable Projection που περιγράφηκε στο προηγούμενο κεφάλαιο.

5.2.1 Υπολογισμός της κλίσης των συναρτήσεων βάσης

Για την προτεινόμενη μέθοδο εκπαίδευσης αναγκαία είναι η κλίση της συναρτήσεως βάσης ως προς τα κέντρα και για το λόγο αυτό ακολουθείται η αντίστοιχη προεργασία για την εύκολη εξαγωγή αυτής. Η δραστηριότητα θέτεται ίση με

$$r = \|\mathbf{x} - \mathbf{c}\|_2 \quad (5.8)$$

ενώ

$$u = r^2 \quad (5.9)$$

Οπότε

$$l(u) = \exp\left(-\frac{u}{2\sigma^2}\right) \quad (5.10)$$

Η $l(u)$ είναι παραγωγίσιμη ως προς u και, επομένως, κανείς λαμβάνει το εξής:

$$\frac{dl}{du} = -\frac{1}{2\sigma^2} \exp\left(-\frac{u}{2\sigma^2}\right) \quad (5.11)$$

Η συνάρτηση

$$\phi(\mathbf{x}, \mathbf{c}) = l(\|\mathbf{x} - \mathbf{c}\|_2^2) \quad (5.12)$$

είναι παραγωγίσιμη ως συνδυασμός δύο διαφορίσιμων συναρτήσεων.

Επομένως, με αλυσιδωτή παραγωγή κανείς λαμβάνει:

$$\frac{\partial \phi}{\partial \mathbf{c}} = \frac{\partial l}{\partial u} \frac{\partial u}{\partial \mathbf{c}} \quad (5.13)$$

όπου

$$\frac{\partial u}{\partial \mathbf{c}} = \frac{\partial \|\mathbf{x} - \mathbf{c}\|_2^2}{\partial \mathbf{c}} = \frac{\partial [(\mathbf{x} - \mathbf{c})(\mathbf{x} - \mathbf{c})^T]}{\partial \mathbf{c}} = -(\mathbf{x} - \mathbf{c}) - (\mathbf{x} - \mathbf{c}) = 2(\mathbf{c} - \mathbf{x}) \quad (5.14)$$

Άρα,

$$\frac{\partial \phi}{\partial \mathbf{c}} = \frac{\partial l}{\partial u} \frac{\partial u}{\partial \mathbf{c}} = 2 \frac{dl}{du} (\mathbf{c} - \mathbf{x}) \quad (5.15)$$

ή

$$\nabla_{\mathbf{c}} \phi(\mathbf{x}, \mathbf{c}) = 2 \left[-\frac{1}{2\sigma^2} \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|_2^2}{2\sigma^2}\right) \right] (\mathbf{c} - \mathbf{x}) \quad (5.16)$$

Το παραπάνω είναι χρήσιμο αποτέλεσμα καθώς αποτελεί τη βάση για το σχηματισμό της Ιακωβιανής του πίνακα Φ .

5.2.2 Υπολογισμός της Ιακωβιανής του πίνακα $\Phi(\mathbf{c})$

Στη μέθοδο αυτή, απαραίτητος είναι ο υπολογισμός της Ιακωβιανής του πίνακα Φ . Πρώτα, όμως, πρέπει να οριστεί η Ιακωβιανή μιας στήλης j , έστω $\Phi_j(\mathbf{c}_j)$.

Ο πίνακας Φ

$$\Phi = \begin{bmatrix} \phi(\mathbf{x}_1, \mathbf{c}_1) & \phi(\mathbf{x}_1, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_1, \mathbf{c}_j) & \cdots & \phi(\mathbf{x}_1, \mathbf{c}_N) \\ \phi(\mathbf{x}_2, \mathbf{c}_1) & \phi(\mathbf{x}_2, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_2, \mathbf{c}_j) & \cdots & \phi(\mathbf{x}_2, \mathbf{c}_N) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi(\mathbf{x}_m, \mathbf{c}_1) & \phi(\mathbf{x}_m, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_m, \mathbf{c}_j) & \cdots & \phi(\mathbf{x}_m, \mathbf{c}_N) \end{bmatrix} \quad (5.17)$$

με την εν λόγω στήλη j

$$\Phi_j(\mathbf{c}_j) = \begin{bmatrix} \phi(\mathbf{x}_1, \mathbf{c}_j) \\ \phi(\mathbf{x}_2, \mathbf{c}_j) \\ \vdots \\ \phi(\mathbf{x}_m, \mathbf{c}_j) \end{bmatrix} \quad (5.18)$$

Η Ιακωβιανή επομένως του $\Phi_j(\mathbf{c}_j)$, $D\Phi_j(\mathbf{c}_j)$, θα είναι ένας $m \times n$ πίνακας. Δηλαδή,

$$D\Phi_j(\mathbf{c}_j) = \begin{bmatrix} \nabla_{\mathbf{c}} \phi(\mathbf{x}_1, \mathbf{c}_j)^T \\ \nabla_{\mathbf{c}} \phi(\mathbf{x}_2, \mathbf{c}_j)^T \\ \vdots \\ \nabla_{\mathbf{c}} \phi(\mathbf{x}_m, \mathbf{c}_j)^T \end{bmatrix} = \begin{bmatrix} \frac{\partial \phi(x_{11}, c_{j1})}{\partial c_{j1}} & \frac{\partial \phi(x_{12}, c_{j2})}{\partial c_{j2}} & \frac{\partial \phi(x_{13}, c_{j3})}{\partial c_{j3}} & \cdots & \frac{\partial \phi(x_{1n}, c_{jn})}{\partial c_{jn}} \\ \frac{\partial \phi(x_{21}, c_{j1})}{\partial c_{j1}} & \frac{\partial \phi(x_{22}, c_{j2})}{\partial c_{j2}} & \frac{\partial \phi(x_{23}, c_{j3})}{\partial c_{j3}} & \cdots & \frac{\partial \phi(x_{2n}, c_{jn})}{\partial c_{jn}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi(x_{m1}, c_{j1})}{\partial c_{j1}} & \frac{\partial \phi(x_{m2}, c_{j2})}{\partial c_{j2}} & \frac{\partial \phi(x_{m3}, c_{j3})}{\partial c_{j3}} & \cdots & \frac{\partial \phi(x_{mn}, c_{jn})}{\partial c_{jn}} \end{bmatrix} \quad (5.19)$$

για όλα τα $j = 1, 2, \dots, N$.

Για παράδειγμα, έστω ότι στο δίκτυο υπάρχουν $N = 20$ κέντρα με τα δεδομένα εκπαίδευσης να είναι $m = 100$ με $n = 4$ μεταβλητές το καθένα.

Για το κέντρο $j = 3$ η Ιακωβιανή, $D\Phi_3(\mathbf{c}_3)$, θα είναι η εξής:

$$D\Phi_3(\mathbf{c}_3) = \begin{bmatrix} \nabla_{\mathbf{c}} \phi(\mathbf{x}_1, \mathbf{c}_3)^T \\ \nabla_{\mathbf{c}} \phi(\mathbf{x}_2, \mathbf{c}_3)^T \\ \vdots \\ \nabla_{\mathbf{c}} \phi(\mathbf{x}_{100}, \mathbf{c}_3)^T \end{bmatrix} = \begin{bmatrix} \frac{\partial \phi(x_{11}, c_{31})}{\partial c_{31}} & \frac{\partial \phi(x_{12}, c_{32})}{\partial c_{32}} & \frac{\partial \phi(x_{13}, c_{33})}{\partial c_{33}} & \frac{\partial \phi(x_{14}, c_{34})}{\partial c_{34}} \\ \frac{\partial \phi(x_{21}, c_{31})}{\partial c_{31}} & \frac{\partial \phi(x_{22}, c_{32})}{\partial c_{32}} & \frac{\partial \phi(x_{23}, c_{33})}{\partial c_{33}} & \frac{\partial \phi(x_{24}, c_{34})}{\partial c_{34}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \phi(x_{m1}, c_{31})}{\partial c_{31}} & \frac{\partial \phi(x_{m2}, c_{32})}{\partial c_{32}} & \frac{\partial \phi(x_{m3}, c_{33})}{\partial c_{33}} & \frac{\partial \phi(x_{m4}, c_{34})}{\partial c_{34}} \end{bmatrix} \quad (5.20)$$

Όπου το στοιχείο 2, 3 σύμφωνα με την (5.16) θα υπολογίζεται ως:

$$\frac{\partial \phi(x_{23}, c_{33})}{\partial c_{33}} = \left[-\frac{1}{\sigma_3^2} \exp\left(-\frac{\|x_{23} - c_{33}\|_2^2}{2\sigma_3^2}\right) \right] (c_{33} - x_{23}) \quad (5.21)$$

Τελικά, η Ιακωβιανή του $\Phi(\mathbf{c})$, $D\Phi(\mathbf{c})$, θα περιλαμβάνει N τέτοιους πίνακες, οπότε ο

$$D\Phi(\mathbf{c}) = [D\Phi_1(\mathbf{c}_1) \quad D\Phi_2(\mathbf{c}_2) \quad \dots \quad D\Phi_N(\mathbf{c}_N)] \quad (5.22)$$

θα είναι ένας πίνακας $m \times Nn$.

Στο προηγούμενο παράδειγμα αυτός ο πίνακας θα ήταν ο εξής:

$$D\Phi(\mathbf{c}) = [D\Phi_1(\mathbf{c}_1) \quad D\Phi_2(\mathbf{c}_2) \quad \dots \quad D\Phi_{20}(\mathbf{c}_{20})] \quad (5.23)$$

Δηλαδή θα αποτελείται από 100×80 στοιχεία.

5.2.3 Υπολογισμός των ευρών των συναρτήσεων βάσης

Τα εύρη του δικτύου θεωρούνται σταθερά και υπολογίζονται με τη μέθοδο των p -κοντινότερων γειτόνων, όπως παρουσιάστηκε στο κεφάλαιο 3.

Δηλαδή το εύρος σ_j για τη j μονάδα RBF υπολογίζεται από την (3.47) ως εξής:

$$\sigma_j = \sqrt{\frac{1}{p} \sum_{i=1}^p \| \mathbf{c}_j - \mathbf{c}_i \|_2^2} \quad (5.24)$$

με τα \mathbf{c}_i να είναι οι κοντινότεροι γείτονες του \mathbf{c}_j . Στην συγκεκριμένη εργασία επιλέγεται $p = 2$.

5.2.4 Υπολογισμός των συναπτικών βαρών

Για κάποια κέντρα \mathbf{c} μπορούν να υπολογιστούν τα βάρη \mathbf{w} του δικτύου και, επομένως, η ελαχιστοποίηση του $e(\mathbf{c}, \mathbf{x})$ ανάγεται στο εξής πρόβλημα:

$$\min_{\mathbf{c}} h(\mathbf{c}) \quad (5.25)$$

όπου

$$h(\mathbf{c}) = \min_{\mathbf{w}} e(\mathbf{c}, \mathbf{w}) = \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{y} - \Phi(\mathbf{c})\mathbf{w}\|_2^2 \quad (5.26)$$

Σε αυτή τη μέθοδο, οι γραμμικές παράμετροι του προβλήματος, τα βάρη του δικτύου δηλαδή, μπορούν να υπολογιστούν από τη λύση ενός γραμμικού προβλήματος ελαχίστων τετραγώνων.

Δηλαδή,

$$\mathbf{w}(\mathbf{c}) = \arg \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{y} - \Phi(\mathbf{c})\mathbf{w}\|_2^2 \quad (5.27)$$

το οποίο δίνει τις «κανονικές» συνθήκες:

$$\Phi(\mathbf{c})^T \Phi(\mathbf{c})\mathbf{w}(\mathbf{c}) = \Phi(\mathbf{c})^T \mathbf{y} \quad (5.28)$$

Για αποτελεσματικούς υπολογισμούς συνιστάται η χρήση της "thin" παραγοντοποίησης QR του $\Phi(\mathbf{c})$, όπως αυτή αναφέρεται στους Golub και Van Loan [Golub96] καθώς αν ο πίνακας $\Phi(\mathbf{c})$ είναι πλήρους τάξης τότε

$$\Phi(\mathbf{c}) = QR = [Q_1 \quad Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1 \quad (5.29)$$

όπου $Q_1 \in \mathbb{R}^{m \times N}$, $Q_2 \in \mathbb{R}^{m \times (m-N)}$ είναι ορθογώνιοι ενώ ο $R_1 \in \mathbb{R}^{N \times N}$ είναι άνω τριγωνικός με θετικά στοιχεία στη διαγώνιο ενώ ισχύει ότι

$$Q^T Q = Q Q^T = I \quad (5.30)$$

από όπου μπορεί να αναχθεί μια χρήσιμη σχέση

$$[Q_1 \quad Q_2] \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} = Q_1 Q_1^T + Q_2 Q_2^T \quad (5.31)$$

Τελικά,

$$Q_1 Q_1^T + Q_2 Q_2^T = I \quad (5.32)$$

Άρα,

$$\Phi(\mathbf{c})^T \Phi(\mathbf{c}) = (Q_1 R_1)^T Q_1 R_1 = R_1^T Q_1^T Q_1 R_1 \quad (5.33)$$

και επειδή ο Q_1 είναι ορθογώνιος τότε

$$Q_1^T Q_1 = I \quad (5.34)$$

Επομένως,

$$\Phi(\mathbf{c})^T \Phi(\mathbf{c}) = (Q_1 R_1)^T Q_1 R_1 = R_1^T R_1 \quad (5.35)$$

Αντικατάσταση στις «κανονικές» συνθήκες:

$$R_1^T R_1 \mathbf{w}(\mathbf{c}) = R_1^T Q_1^T \mathbf{y} \quad (5.36)$$

ο R_1 είναι τετραγωνικός και αντιστρέψιμος, οπότε:

$$\mathbf{w}(\mathbf{c}) = R_1^{-1} Q_1^T \mathbf{y} \quad (5.37)$$

και επειδή ο ψευδοαντίστροφος του Φ ορίζεται ως:

$$\Phi^+ = (\Phi(\mathbf{c})^T \Phi(\mathbf{c}))^{-1} \Phi(\mathbf{c})^T \quad (5.38)$$

τότε με τις κατάλληλες πράξεις προκύπτει για τον ψευδοαντίστροφο ότι

$$\Phi^+ = R_1^{-1} Q_1^T \quad (5.39)$$

Με τη χρήση της εν λόγω παραγοντοποίησης απλοποιούνται οι σχέσεις για το ψευδοαντίστροφο του $\Phi(\mathbf{c})$ και, επομένως, για τη επίλυση των «κανονικών» συνθηκών, αλλά ακόμα και για τον υπολογισμό του κόστους και, όπως θα φανεί στη συνέχεια, για τους πίνακες προβολής του $\Phi(\mathbf{c})$.

5.2.5 Υπολογισμός του κόστους

Το κόστος που θα δίνουν οι παράμετροι του δικτύου $h(\mathbf{c})$ θα υπολογίζεται ως εξής:

$$h(\mathbf{c}) = \frac{1}{2} \|\Phi(\mathbf{c})\mathbf{w} - \mathbf{y}\|_2^2 \quad (5.40)$$

ενώ με την παραγοντοποίηση QR αυτό γίνεται:

$$h(\mathbf{c}) = \frac{1}{2} \|Q_1 R_1 \mathbf{w} - \mathbf{y}\|_2^2 \quad (5.41)$$

όμως από την (5.37)

$$\mathbf{w}(\mathbf{c}) = R_1^{-1} Q_1^T \mathbf{y} \quad (5.42)$$

Αρα,

$$h(\mathbf{c}) = \frac{1}{2} \|Q_1 R_1 R_1^{-1} Q_1^T \mathbf{y} - \mathbf{y}\|_2^2 \quad (5.43)$$

ή

$$h(\mathbf{c}) = \frac{1}{2} \|Q_1 Q_1^T \mathbf{y} - \mathbf{y}\|_2^2 = \frac{1}{2} \|(I - Q_1 Q_1^T) \mathbf{y}\|_2^2 \quad (5.44)$$

Λόγω της (5.32) ισχύει ότι:

$$I = Q_1 Q_1^T + Q_2 Q_2^T \quad (5.45)$$

Επομένως,

$$h(\mathbf{c}) = \frac{1}{2} \|(I - Q_1 Q_1^T) \mathbf{y}\|_2^2 = \frac{1}{2} \|(Q_2 Q_2^T) \mathbf{y}\|_2^2 \quad (5.46)$$

και επειδή για έναν ορθογώνιο πίνακα ισχύει ότι $\|Q\|_2 = 1$, τότε:

$$h(\mathbf{c}) = \frac{1}{2} \|Q_2^T \mathbf{y}\|_2^2 \quad (5.47)$$

5.2.6 Υπολογισμός των κέντρων των συναρτήσεων βάσης

Στο κεφάλαιο που προηγήθηκε έγινε περιγραφή της μεθόδου Variable-Projection στην οποία το αρχικό πρόβλημα της εύρεσης των δύο συνόλων παραμέτρων (εξ. 4.92) ανάγεται σε δύο υποπροβλήματα. Το ένα αφορά τον υπολογισμό των γραμμικών παραμέτρων, δηλαδή των βαρών, μέσω ενός απλού γραμμικού προβλήματος (εξ. 4.94), ενώ το άλλο υποπρόβλημα αφορά τον υπολογισμό των μη γραμμικών παραμέτρων, δηλαδή των κέντρων, μέσω ενός μη γραμμικού προβλήματος (εξ. 4.93).

Από το

$$h(\mathbf{c}) = \frac{1}{2} \|\Phi(\mathbf{c})\mathbf{w} - \mathbf{y}\|_2^2 \quad (5.48)$$

και από τον πίνακα προβολής στο χώρο στηλών του $\Phi(\mathbf{c})$

$$P_{\Phi(\mathbf{c})} = \Phi(\Phi^T\Phi)^{-1}\Phi^T = \Phi\Phi^+ \quad (5.49)$$

Με την παραγοντοποίηση QR ο πίνακας προβολής παίρνει την εξής μορφή:

$$P_{\Phi(\mathbf{c})} = Q_1R_1R_1^{-1}Q_1^T = Q_1Q_1^T \quad (5.50)$$

Τότε το κόστος μπορεί να οριστεί ως:

$$h(\mathbf{c}) = \frac{1}{2}\|(I - Q_1Q_1^T)\mathbf{y}\|_2^2 = \frac{1}{2}\|(I - P_{\Phi})\mathbf{y}\|_2^2 \quad (5.51)$$

Η ποσότητα $I - P_{\Phi}$ είναι η προβολή στο ορθογώνιο συμπλήρωμα του χώρου στηλών του $\Phi(\mathbf{c})$ (δηλαδή στον αριστερό μηδενοχώρο του $\Phi(\mathbf{c})$), όπου από την (4.89):

$$P_{\Phi(\mathbf{c})}^{\perp} = I - P_{\Phi(\mathbf{c})} \quad (5.52)$$

Άρα,

$$h(\mathbf{c}) = \frac{1}{2}\|P_{\Phi(\mathbf{c})}^{\perp}\mathbf{y}\|_2^2 \quad (5.53)$$

ή

$$\min_{\mathbf{c}} \frac{1}{2}\|P_{\Phi(\mathbf{c})}^{\perp}\mathbf{y}\|_2^2 \quad (5.54)$$

Η ποσότητα μέσα στη νόρμα είναι το προβαλλόμενο υπόλοιπο και συμβολίζεται με

$$\mathbf{R}(\mathbf{c}) = P_{\Phi(\mathbf{c})}^{\perp}\mathbf{y} \quad (5.55)$$

5.2.7 Εφαρμογή του αλγορίθμου Levenberg-Marquardt

Για την επίλυση της (5.54) χρησιμοποιείται η μέθοδος Levenberg-Marquardt, η οποία αποτελεί παραλλαγή της μεθόδου Gauss-Newton που παρουσιάστηκε προηγουμένως. Όπως περιγράφηκε νωρίτερα, τελικά, κανείς πρέπει να λύσει ένα γραμμικό πρόβλημα το οποίο προκύπτει για μικρά $\|\mathbf{d}\|$ και από την ανάπτυξη του (5.55):

$$\min_{\mathbf{d}} \frac{1}{2}\|\mathbf{R}(\mathbf{c}) + D\mathbf{R}(\mathbf{c})\mathbf{d}\|_2^2 \quad (5.56)$$

Με την Ιακωβιανή να προσεγγίζεται από την (4.99) ως :

$$D\mathbf{R}(\mathbf{c}) \approx -\left(P_{\Phi(\mathbf{c})}^{\perp}D(\Phi(\mathbf{c}))\Phi(\mathbf{c})^+\right)\mathbf{y} \quad (5.57)$$

Αν ληφθεί υπόψη ότι

$$\mathbf{w}(\mathbf{c}) = R_1^{-1}Q_1^T\mathbf{y} = \Phi^+\mathbf{y} \quad (5.58)$$

τότε

$$D\mathbf{R}(\mathbf{c}) = -P_{\Phi(\mathbf{c})}^{\perp}D(\Phi(\mathbf{c}))\mathbf{w}(\mathbf{c}) \quad (5.59)$$

Επομένως, με αντικατάσταση των παραπάνω στην (5.56)

$$\min_{\mathbf{d}} \frac{1}{2}\|\mathbf{R}(\mathbf{c}) + D\mathbf{R}(\mathbf{c})\mathbf{d}\|_2^2 = \min_{\mathbf{d}} \frac{1}{2}\|P_{\Phi(\mathbf{c})}^{\perp}\mathbf{y} - P_{\Phi(\mathbf{c})}^{\perp}D(\Phi(\mathbf{c}))\mathbf{w}(\mathbf{c})\mathbf{d}\|_2^2 = \min_{\mathbf{d}} \frac{1}{2}\|P_{\Phi(\mathbf{c})}^{\perp}(\mathbf{y} - D(\Phi(\mathbf{c}))\mathbf{w}(\mathbf{c})\mathbf{d})\|_2^2 \quad (5.60)$$

Από τις 4.89, 5.32, 5.50

$$P_{\Phi(\mathbf{c})}^{\perp} = Q_2Q_2^T \quad (5.61)$$

Άρα,

$$\min_{\mathbf{d}} \frac{1}{2} \left\| Q_2 Q_2^T (\mathbf{y} - D(\Phi(\mathbf{c})) \mathbf{w}(\mathbf{c}) \mathbf{d}) \right\|_2^2 = \min_{\mathbf{d}} \frac{1}{2} \left\| Q_2^T (\mathbf{y} - D(\Phi(\mathbf{c})) \mathbf{w}(\mathbf{c}) \mathbf{d}) \right\|_2^2 \quad (5.62)$$

Με το τελευταίο αποτέλεσμα να προκύπτει λόγω της ιδιότητας του ορθογώνιου πίνακα.

Συνεπώς, για να βρεθεί η διεύθυνση \mathbf{d} πρέπει να λυθεί το ακόλουθο γραμμικό πρόβλημα ελαχίστων τετραγώνων

$$\min_{\mathbf{d}} \frac{1}{2} \left\| A(\mathbf{c}) \mathbf{d} - \mathbf{b} \right\|_2^2 \quad (5.63)$$

όπου

$$A(\mathbf{c}) = -Q_2^T D(\Phi(\mathbf{c})) \mathbf{w}(\mathbf{c}) \mathbf{d} \in \mathfrak{R}^{(m-N) \times Nn}$$

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

Με τη μέθοδο Levenberg-Marquardt το παραπάνω γίνεται:

$$\min_{\mathbf{d}} \frac{1}{2} \left\| A(\mathbf{c}) \mathbf{d} - \mathbf{b} \right\|_2^2 + \frac{\lambda}{2} \left\| \mathbf{d} \right\|_2^2 \quad (5.64)$$

Επίλυση για τη διεύθυνση \mathbf{d}

Η 5.64 μπορεί να λυθεί από το

$$(A(\mathbf{c})^T A(\mathbf{c}) + \lambda I) \mathbf{d} = A(\mathbf{c})^T \mathbf{b} \quad (5.65)$$

χρησιμοποιώντας παραγοντοποίηση Cholesky και την εκτέλεση εμπρόσθιων και οπίσθιων υπολογισμών για την εύρεση του \mathbf{d} . Για $A(\mathbf{c})^T A(\mathbf{c}) + \lambda I$ συμμετρικό και θετικά ορισμένο πίνακα, η επίλυση της παραπάνω συνίσταται στη λύση του $L \mathbf{y} = A(\mathbf{c})^T \mathbf{b}$ για το \mathbf{y} με εμπρόσθια αντικατάσταση και στη λύση του $L^T \mathbf{d} = \mathbf{y}$ για το \mathbf{d} με οπίσθια αντικατάσταση.

Δηλαδή,

$$\mathbf{d} = (L^T)^{-1} L^{-1} A(\mathbf{c})^T \mathbf{b} \quad (5.66)$$

Διαφορετικά η 5.64 μπορεί να μορφοποιηθεί ως ένα γραμμικό πρόβλημα ελαχίστων τετραγώνων:

$$\min_{\mathbf{d}} \frac{1}{2} \left\| \begin{pmatrix} A(\mathbf{c}) \\ \sqrt{\lambda} I \end{pmatrix} \mathbf{d} - \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \right\|_2^2 \quad (5.67)$$

5.2.8 Περιγραφή του αλγορίθμου Variable Projection - Levenberg-Marquardt

Ο αλγόριθμος ξεκινάει με μια αρχική επιλογή των κέντρων και με μια αρχική τιμή για την παράμετρο $\lambda > 0$. Αφού έχει σχηματιστεί ο πίνακας Φ , ακολουθεί η παραγοντοποίηση QR και ο υπολογισμός των βαρών για τα επιλεγμένα κέντρα από τη λύση των κανονικών εξισώσεων. Έπειτα, υπολογίζεται το κόστος που δίνουν αυτές οι παράμετροι. Σε κάθε επανάληψη του αλγορίθμου, για τα τρέχοντα κέντρα και βάρη, υπολογίζεται μια διεύθυνση \mathbf{d} όπως περιγράφηκε νωρίτερα, και, έπειτα, ελέγχεται αν στη νέα θέση των κέντρων που προτείνει αυτή η διεύθυνση υπάρχει μείωση στην τιμή του κόστους και κατά πόσο θεωρείται καλή η μείωση αυτή, έτσι ώστε να μεταβληθεί καταλλήλως η τιμή του λ . Μεγάλη τιμή του r_k υποδηλώνει ότι το γραμμικό μοντέλο είναι καλός προσεγγιστής της συνάρτησης κόστους, οπότε η επόμενη επανάληψη θα ξεκινάει με μικρότερο λ , έτσι ώστε το βήμα να είναι πιο κοντά σε βήμα Gauss-Newton. Μικρές τιμές του r_k , ή ακόμα και αρνητικές, δείχνουν ότι το γραμμικό μοντέλο δεν είναι καλός προσεγγιστής, οπότε πρέπει να αυξηθεί η τιμή του λ , ώστε να περιοριστεί το μέγεθος του βήματος και να πλησιάζει αυτό της απότομης κατάβασης. Η ενημέρωση των κέντρων γίνεται μόνο στην περίπτωση όπου $r_k > 0$, δηλαδή, μόνο εάν στη νέα θέση $\mathbf{c}_k + \mathbf{d}_k$ το κόστος είναι μικρότερο. Στην περίπτωση αυτή, λοιπόν, υπολογίζονται εκ νέου τα βάρη από τη λύση των κανονικών εξισώσεων. Ο αλγόριθμος συνεχίζεται μέχρι να ικανοποιηθεί κάποιο κριτήριο τερματισμού.

Ο ψευδοκώδικας που περιγράφει τον αλγόριθμο Variable Projection με Levenberg-Marquardt είναι ο παρακάτω:

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 $\Phi(\mathbf{c}^0)$

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

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 $\Phi(\mathbf{c}^k + \mathbf{d}^k)$

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

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

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

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

5.3 Εκπαίδευση ενός δικτύου RBF βάσει των μεθόδων Variable-Projection και Levenberg-Marquardt

Για τη σωστή εκπαίδευση του δικτύου εισάγονται ως δεδομένα δύο διαφορετικά σύνολα: το σύνολο εκπαίδευσης και το σύνολο επικύρωσης. Αν και η εκπαίδευση πραγματοποιείται με αυτά τα δύο, η ύπαρξη ενός τρίτου ανεξάρτητου συνόλου, του συνόλου δοκιμής, είναι κρίσιμη και είναι αυτό το οποίο, τελικά, καθορίζει την απόδοση του δικτύου. Η επιλογή των παραμέτρων του δικτύου γίνεται βάσει του συνόλου επικύρωσης, ωστόσο η διαδικασία βελτιστοποίησης που ακολουθείται μπορεί να οδηγήσει σε υπερπροσαρμογή του μοντέλου σε αυτό το σύνολο καθιστώντας έτσι απαραίτητη την ύπαρξη του συνόλου εφαρμογής [Alexandridis13].

Οι παράμετροι που πρέπει να μάθει το δίκτυο είναι ο αριθμός των μονάδων RBF, τα κέντρα και τα εύρη των συναρτήσεων βάσης καθώς και τα συναπτικά βάρη σύνδεσης κρυφού επιπέδου - επιπέδου εξόδου.

Η επιλογή των ευρών των συναρτήσεων βάσης πραγματοποιείται με τη μέθοδο των p -κοντινότερων γειτόνων, όπως αυτή περιγράφηκε στο κεφάλαιο 3, ενώ τα βάρη θα προέρχονται από τη λύση των κανονικών εξισώσεων του γραμμικού προβλήματος ελαχίστων τετραγώνων.

Για την αρχική επιλογή των κέντρων χρησιμοποιήθηκαν δύο επιλογές: η πρώτη αφορά την εφαρμογή της μεθόδου k -means και η δεύτερη την εφαρμογή της μεθόδου fuzzy means. Στην περίπτωση της μεθόδου k -means πρέπει πρώτα να βρεθεί ο αριθμός και οι θέσεις των κέντρων μέσω μιας επαναληπτικής διαδικασίας. Η διαδικασία περιλαμβάνει τα στάδια δοκιμής και σφάλματος μέσα σε ένα εύρος διαφορετικών αριθμών κέντρων $[N_{\alpha\rho\chi\iota\kappa\acute{o}}, N_{\tau\epsilon\lambda\iota\kappa\acute{o}}]$. Ακόμα, λόγω της τυχαιότητας που χαρακτηρίζει τη μέθοδο αυτή, πρέπει σε κάθε έναν αριθμό κέντρων N μέσα σε αυτό το εύρος να γίνει μια πληθώρα δοκιμών. Στη συγκεκριμένη εργασία επιλέγεται για κάθε αριθμό κέντρων να γίνουν 20 δοκιμές. Αφότου ολοκληρωθούν οι δοκιμές αυτές σε κάθε αριθμό κέντρων, τελικά επιλέγεται το μοντέλο εκείνο που δίνει τη χαμηλότερη μέση τιμή RMSE στο σύνολο επικύρωσης. Έχοντας πλέον τον αριθμό των μονάδων RBF και τις θέσεις των κέντρων μπορεί να εφαρμοστεί ο αλγόριθμος Variable-Projection-Levenberg-Marquardt. Με τις επαναλήψεις του αλγορίθμου ενημερώνονται οι τιμές των κέντρων βάσει του κόστους στο σύνολο εκπαίδευσης και των βαρών από τη λύση των κανονικών εξισώσεων. Αν και η ενημέρωση των κέντρων γίνεται βάσει του κόστους στο σύνολο εκπαίδευσης, η επιλογή του μοντέλου γίνεται βάσει του κόστους στο σύνολο επικύρωσης.

Ως αντικειμενική συνάρτηση ορίζεται η ποσότητα

$$h = \frac{1}{2} \|\mathbf{y} - \Phi\mathbf{w}\|_2^2 \quad (5.68)$$

όπως ορίστηκε προηγουμένως.

Οι επαναλήψεις ολοκληρώνονται όταν ικανοποιηθούν τα επιλεγμένα κριτήρια τερματισμού.

Αυτά τα κριτήρια αφορούν:

- μέγιστο αριθμό αποτυχημένων επαναλήψεων του αλγορίθμου, δηλαδή διαδοχικές επαναλήψεις στις οποίες δεν έγινε κάποια ενημέρωση στα κέντρα
- μέγιστο αριθμό αποτυχημένων επαναλήψεων του αλγορίθμου για βελτίωση του κόστους στο σύνολο επικύρωσης, δηλαδή διαδοχικές επαναλήψεις στις οποίες δεν έγινε κάποια βελτίωση του κόστους στο σύνολο αυτό

Το πρώτο κριτήριο αποτρέπει τις άσκοπες επαναλήψεις ενώ το δεύτερο την υπερπροσαρμογή του μοντέλου στο σύνολο εκπαίδευσης.

Στην περίπτωση της μεθόδου fuzzy means δεν υπάρχει αυτή η τυχαιότητα που χαρακτηρίζει τη μέθοδο k -means. Ωστόσο, πρέπει να υπολογιστεί ο κατάλληλος τρόπος διαχωρισμός του χώρου εισόδου και να επιλεγθεί ο κατάλληλος μέσα από μια διαδικασία επιλογής βάσει του ελαχίστου κόστους στο σύνολο επικύρωσης σε ένα συγκεκριμένο εύρος $[s_{\alpha\rho\chi\iota\kappa\acute{o}}, s_{\tau\epsilon\lambda\iota\kappa\acute{o}}]$ για τον ασαφή διαχωρισμό, με την υπόλοιπη διαδικασία να μη διαφοροποιείται.

5.4 Περίπτωση πλήρους συναρτησιακού

Στην περίπτωση του πλήρους συναρτησιακού (full functional, FF), δηλαδή στην περίπτωση που δεν πραγματοποιείται απαλοιφή των βαρών, τότε το πρόβλημα προς επίλυση είναι το εξής:

$$\min_{\mathbf{c}, \mathbf{w}} e(\mathbf{c}, \mathbf{w}) = \min_{\mathbf{c}, \mathbf{w}} \frac{1}{2} \sum_{i=1}^m r_i(\mathbf{c}, \mathbf{w})^2 = \frac{1}{2} \|\mathbf{r}(\mathbf{c}, \mathbf{w})\|^2 \quad (5.69)$$

με

$$\mathbf{r}(\mathbf{c}, \mathbf{w}) = \mathbf{y} - \Phi(\mathbf{c})\mathbf{w} \quad (5.70)$$

Η διαδικασία είναι παρόμοια με αυτή που προηγήθηκε καθώς κι εδώ η γραμμικοποίηση γύρω από τα τρέχοντα \mathbf{c}, \mathbf{w} οδηγεί στο:

$$\min_{\mathbf{c}, \mathbf{w}} \frac{1}{2} \|\mathbf{r}(\mathbf{c}, \mathbf{w}) + D\mathbf{r}(\mathbf{c}, \mathbf{w})\mathbf{d}\|^2 \quad (5.71)$$

με την Ιακωβιανή $D\mathbf{r}(\mathbf{c}, \mathbf{w})$ να προκύπτει από [Golub73]:

$$D\mathbf{r}(\mathbf{c}, \mathbf{w}) = - [D\Phi(\mathbf{c})\mathbf{w} \quad \Phi(\mathbf{c})] \in \mathfrak{R}^{m \times (Nn+N)} \quad (5.72)$$

με το $D\Phi(\mathbf{c})\mathbf{w} \in \mathfrak{R}^{m \times Nn}$ να είναι η Ιακωβιανή ως προς τα κέντρα \mathbf{c} και με το $\Phi(\mathbf{c}) \in \mathfrak{R}^{m \times N}$ να είναι η Ιακωβιανή ως προς τα βάρη \mathbf{w} .

Με τη μέθοδο Levenberg-Marquardt (5.64) κανείς πρέπει να λύσει το εξής:

$$\min_{\mathbf{d}} e(\mathbf{c}, \mathbf{w}) = \min_{\mathbf{d}} \frac{1}{2} \|\mathbf{A}\mathbf{d} - \mathbf{b}\|^2 + \frac{\lambda}{2} \|\mathbf{d}\|^2 \quad (5.73)$$

με αντίστοιχο τρόπο όπως και στο variable projection για να βρεθεί η διεύθυνση $\mathbf{d} \in \mathfrak{R}^{(Nn+N) \times 1}$

$$(\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})\mathbf{d} = \mathbf{A}^T \mathbf{b} \quad (5.74)$$

με τις μεθόδους που αναφέρθηκαν νωρίτερα, όπου πλέον εδώ:

$$\begin{aligned} \mathbf{A}(\mathbf{c}) &= D\mathbf{r}(\mathbf{c}, \mathbf{w}) \in \mathfrak{R}^{m \times (Nn+N)} \\ \mathbf{b} &= -\mathbf{r}(\mathbf{c}, \mathbf{w}) \in \mathfrak{R}^{m \times 1} \end{aligned} \quad (5.75)$$

Ο Golub [Golub73] προτείνει οι αρχικές γραμμικές παραμέτρους, δηλαδή τα αρχικά βάρη, να προέρχονται πάλι από τη λύση ελαχίστων τετραγώνων

$$\mathbf{w}^0 = \Phi(\mathbf{c}^0)^+ \mathbf{y} \quad (5.76)$$

Ο αλγόριθμος που εκτελεί το πλήρες συναρτησιακό είναι παρόμοιος με αυτό του μειωμένου, με τη διαφορά να έγκειται στο γεγονός ότι πλέον και τα βάρη του δικτύου θα υπολογίζονται βάσει της μεθόδου Levenberg-Marquardt.

5.5 Περίπτωση μεταβλητών ευρών για τις συναρτήσεις βάσης

Μέχρι τώρα έγινε η υπόθεση ότι οι τιμές των ευρών που προκύπτουν από τους p -κοντινότερους γείτονες παραμένουν σταθερές και δε μεταβάλλονται κατά τη διάρκεια του αλγορίθμου. Ωστόσο, καθώς αλλάζουν οι θέσεις των κέντρων των συναρτήσεων βάσης, κανείς θα περίμενε να μεταβάλλονται και τα αντίστοιχα εύρη, μιας και πάντα πρέπει οι μονάδες RBF να είναι έτσι τοποθετημένες, ώστε να αντιπροσωπεύεται ο χώρος εισόδου όσο το δυνατόν καλύτερα από αυτές.

Κανείς θα μπορούσε, λοιπόν, στην ίδια διαδικασία να διαχειρίζεται και τα εύρη σ_j ως παραμέτρους του προβλήματος

$$\min_{\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma}} e(\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma}) = \min_{\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma}} \frac{1}{2} \sum_{i=1}^m r_i(\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma})^2 = \min_{\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma}} \frac{1}{2} \|\mathbf{r}(\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma})\|^2 \quad (5.77)$$

με

$$\mathbf{r}(\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma}) = \mathbf{y} - \Phi(\mathbf{c}, \boldsymbol{\sigma})\mathbf{w} \quad (5.78)$$

Για την επίλυση του

$$\min_{\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma}} \frac{1}{2} \|\mathbf{r}(\mathbf{c}, \mathbf{w}) + D\mathbf{r}(\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma})\mathbf{d}\|^2 \quad (5.79)$$

κανείς χρειάζεται την Ιακωβιανή ως προς τα εύρη.

Από

$$\Phi = \begin{bmatrix} \phi(\mathbf{x}_1, \mathbf{c}_1) & \phi(\mathbf{x}_1, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_1, \mathbf{c}_j) & \cdots & \phi(\mathbf{x}_1, \mathbf{c}_N) \\ \phi(\mathbf{x}_2, \mathbf{c}_1) & \phi(\mathbf{x}_2, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_2, \mathbf{c}_j) & \cdots & \phi(\mathbf{x}_2, \mathbf{c}_N) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi(\mathbf{x}_m, \mathbf{c}_1) & \phi(\mathbf{x}_m, \mathbf{c}_2) & \cdots & \phi(\mathbf{x}_m, \mathbf{c}_j) & \cdots & \phi(\mathbf{x}_m, \mathbf{c}_N) \end{bmatrix} \quad (5.80)$$

προκύπτει ότι η ζητούμενη Ιακωβιανή ως προς τα εύρη θα είναι ένα πίνακας $m \times N$ με στοιχεία:

$$\left[D\Phi\boldsymbol{\sigma} \right]_{i,j} = \frac{\partial \phi(\mathbf{x}_i, \mathbf{c}_j, \sigma_j)}{\sigma_j}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, N \quad (5.81)$$

Το στοιχείο 2, 3 του προηγούμενου παραδείγματος υπολογίζεται ως:

$$\frac{\partial \phi(\mathbf{x}_2, \mathbf{c}_3, \sigma_3)}{\partial \sigma_3} = \frac{\|\mathbf{x}_2 - \mathbf{c}_3\|_2^2}{\sigma_3^3} \exp\left(-\frac{\|\mathbf{x}_2 - \mathbf{c}_3\|_2^2}{2\sigma_3^2}\right) \quad (5.82)$$

Η διαδικασία που ακολουθείται είναι παρόμοια καθώς κανείς πρέπει να αλλάξει μόνο τα A και b , ώστε να τα προσαρμόσει στο εκάστοτε πρόβλημα.

Δηλαδή στην περίπτωση του VP για να βρεθεί η διεύθυνση $\mathbf{d} \in \mathbb{R}^{(Nn+N) \times 1}$ πρέπει να λυθεί το παρακάτω

$$(A^T A + \lambda I)\mathbf{d} = A^T \mathbf{b} \quad (5.83)$$

με τις μεθόδους που αναφέρθηκαν νωρίτερα, όπου πλέον εδώ:

$$\begin{aligned} A &= D\mathbf{r}(\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma}) \in \mathbb{R}^{m \times (Nn+N)} \\ \mathbf{b} &= -\mathbf{r}(\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma}) \in \mathbb{R}^{m \times 1} \end{aligned} \quad (5.84)$$

ενώ η Ιακωβιανή $D\mathbf{r}(\mathbf{c}, \mathbf{w})$ προκύπτει από:

$$D\mathbf{r}(\mathbf{c}, \mathbf{w}, \boldsymbol{\sigma}) = - \left[D\Phi(\mathbf{c})\mathbf{w} \quad D\Phi(\boldsymbol{\sigma})\mathbf{w} \right] \in \mathbb{R}^{m \times (Nn+N)} \quad (5.85)$$

με το $D\Phi(\mathbf{c})\mathbf{w} \in \mathbb{R}^{m \times Nn}$ να είναι η Ιακωβιανή ως προς τα κέντρα, \mathbf{c} , και με το $D\Phi(\boldsymbol{\sigma}) \in \mathbb{R}^{m \times N}$ να είναι η Ιακωβιανή ως προς τα εύρη, $\boldsymbol{\sigma}$.

Ο αλγόριθμος πλέον θα μεταβάλλει και τις τιμές των ευρών των συναρτήσεων βάσης σύμφωνα με τη μέθοδο Levenberg-Marquardt, ενώ μπορεί να συνδυαστεί τόσο με τη μέθοδο του μειωμένου, ή του πλήρους συναρτησιακού.

Κεφάλαιο 6

Πειραματική διαδικασία

6.1 Περιγραφή των benchmark datasets

Η προτεινόμενη μέθοδος εκπαίδευσης δικτύων RBF εφαρμόστηκε σε μια πληθώρα διαφόρων συνόλων δεδομένων, συνθετικά και πραγματικά, όπως αυτά αναφέρονται στην εργασία των [Alexandridis13].

Στη συνέχεια, ακολουθεί μια σύντομη περιγραφή των συνόλων:

- Auto MPG: στόχος είναι η πρόβλεψη της κατανάλωσης καυσίμου ενός αυτοκινήτου χρησιμοποιώντας ως είσοδο κάποια από τα χαρακτηριστικά του.
- Boston Housing: στόχος είναι η πρόβλεψη των τιμών για τα σπίτια στη Βοστώνη, βάσει των χαρακτηριστικών τους. Τα δεδομένα προήλθαν από την απογραφή του 1970.
- CPU Small: στόχος είναι η πρόβλεψη του μέρους του χρόνου που τρέχουν τα CPUs σε κατάσταση χρήστη, λαμβάνοντας υπόψη έναν περιορισμένο αριθμό χαρακτηριστικών.
- Friedman: το σύνολο αυτό, αντιθέτως με τα προηγούμενα, είναι συνθετικό και προέρχεται από τη συνάρτηση

$$y = 5[2\sin(\pi x_1 x_2) + 4(x_3 - 0.5)^2 + 2x_4 + x_5] + \epsilon \quad (6.1)$$

όπου ϵ ο θόρυβος Gaussian $\sim N(0, 0.8)$. Οι τιμές για τις μεταβλητές εισόδου προέρχονται από μια ομοιόμορφη κατανομή στο $[0, 1]$.

Στον πίνακα 6.1 αναφέρονται επιγραμματικά τα σύνολα δεδομένων στα οποία έγιναν οι δοκιμές μαζί με τον αριθμό των μεταβλητών εισόδου και τον αριθμό των παραδειγμάτων που τα συνοδεύουν.

Πίνακας 6.1: Benchmark σύνολα δεδομένων.

Σύνολο δεδομένων	Αριθμός μεταβλητών εισόδου	Αριθμός παραδειγμάτων
Πραγματικά Σύνολα Δεδομένων		
Auto MPG	6	392
Boston Housing	13	506
CPU Small	12	8192
Συνθετικά Σύνολα Δεδομένων		
Friedman	5	1000

6.2 Παρουσίαση και συζήτηση αποτελεσμάτων

Ο τρόπος διαχωρισμού των συνόλων δεδομένων είναι αυτός που αναφέρεται στο [Alexandridis13] καθώς ένα σύνολο χωρίζεται σε ένα σύνολο εκπαίδευσης (training dataset), ένα σύνολο επικύρωσης (validation dataset) και σε ένα σύνολο δοκιμής (testing dataset) σε συγκεκριμένη αναλογία: 50%, 25%, 25%, αντίστοιχα.

Οι παράμετροι για τους αλγόριθμους που χρησιμοποιούνται στις μεθόδους εκπαίδευσης αναγράφονται στον πίνακα 6.2.

Πίνακας 6.2: Τιμές παραμέτρων για τους διάφορους αλγόριθμους.

Παράμετρος	Σύμβολο	Τιμή
ελάχιστος αριθμός ασαφών συνόλων	s_{min}	5
μέγιστος αριθμός ασαφών συνόλων	s_{max}	25
ελάχιστος αριθμός κέντρων	N_{min}	N_{min}
μέγιστος αριθμός κέντρων	N_{max}	N_{max}
μέθοδος Levenberg-Marquardt	λ	$\tau \cdot \max(A^T A)$

Οι αλγόριθμοι που χρησιμοποιούνται για την εύρεση του βέλτιστου αριθμού και των θέσεων των κέντρων και, τελικά, για την επιλογή του μοντέλου είναι ο αλγόριθμος k-means και ο αλγόριθμος fuzzy means. Το μοντέλο επιλέγεται βάσει του ελάχιστου κόστους στο σύνολο επικύρωσης. Στον αλγόριθμο k-means η αναζήτηση επικεντρώνεται σε ένα συγκεκριμένο εύρος αριθμού κέντρων για κάθε σύνολο δεδομένων της βιβλιογραφίας $[N_{min}, N_{max}]$. Ωστόσο, λόγω της στοχαστικής φύσης του αλγορίθμου, απαιτείται μια σειρά δοκιμών, οπότε η επιλογή του μοντέλου γίνεται βάσει του ελάχιστου μέσου κόστους στο σύνολο επικύρωσης. Στον αλγόριθμο fuzzy means η αναζήτηση επικεντρώνεται σε ένα συγκεκριμένο εύρος αριθμού διαχωρισμού του χώρου εισόδου $[s_{min}, s_{max}]$.

Εφόσον έχει βρεθεί το μοντέλο με έναν από τους παραπάνω τρόπους, τότε είναι δυνατόν να εφαρμοστεί είτε η μέθοδος Variable Projection είτε η μέθοδος Full Functional, συνδυασμένες με τον αλγόριθμο Levenberg-Marquardt, ενώ χρησιμοποιούνται διάφορες τιμές της παραμέτρου λ , όπως προτείνεται στο [Madsen04].

Για λόγους σύγκρισης, παρατίθενται τα αποτελέσματα της εργασίας των [Alexandridis13], ονόματι PSO-NSFM, αλλά και αποτελέσματα διαφόρων μεθόδων μηχανικής μάθησης και συγκεκριμένα των μεθόδων SFM και MLP. Στη μέθοδο PSO-NSFM η εκπαίδευση του δικτύου γίνεται βάσει ενός μη συμμετρικού διαχωρισμού του χώρου εισόδου, βελτιωμένος με τη χρήση βελτιστοποίησης σμήνων σωματιδίων. Στην περίπτωση του SFM η επιλογή του μοντέλου γίνεται βάσει της παραμέτρου s για το συμμετρικό διαχωρισμό των χωρών των μεταβλητών εισόδου. Ως συνάρτηση βάσης χρησιμοποιείται η thin-plate-spline. Όσον αφορά τη μέθοδο MLP, πρόκειται για πρόσθιας τροφοδότησης δίκτυα δύο επιπέδων εκπαιδευμένα βάσει του αλγορίθμου Levenberg-Marquardt, ενώ οι δοκιμές για την επιλογή του μοντέλου έγιναν για κάθε δυνατό συνδυασμό αριθμών κέντρων σε ένα συγκεκριμένο εύρος. Σε όλες τις μεθόδους η εκπαίδευση γίνεται βάσει του συνόλου εκπαίδευσης και η επιλογή του δικτύου βάσει του συνόλου επικύρωσης.

Αρχικά, στους παρακάτω πίνακες, παρουσιάζονται τα αποτελέσματα για απλή εφαρμογή των αλγορίθμων k-means και fuzzy means, δηλαδή χωρίς τις επαναλήψεις του αλγορίθμου Levenberg-Marquardt, ως προς την απόδοση (πίνακας 6.3) και τον υπολογιστικό χρόνο (πίνακες 6.4-6.5). Αυτοί οι πίνακες αφορούν τη διαδικασία επιλογής του μοντέλου, όπως αυτή περιγράφηκε προηγουμένως.

Πίνακας 6.3: Αποτελέσματα των αλγορίθμων k-means και fuzzy means ως προς το κόστος στα σύνολα εκπαίδευσης, επικύρωσης και δοκιμής.

Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Ασαφής διαχωρισμός	Αριθμός μονάδων
Auto MPG	k-means	2.6152 ($2.6167 \pm 5.25 \times 10^{-2}$)	2.3743 ($2.5490 \pm 7.78 \times 10^{-2}$)	2.0782 ($2.1681 \pm 9.76 \times 10^{-2}$)	-	49
	fuzzy means	2.6174	2.3417	2.0659	10	43
Boston Housing	k-means	1.9253 ($2.0911 \pm 35.32 \times 10^{-2}$)	2.8498 ($3.0900 \pm 13.09 \times 10^{-2}$)	2.6298 ($2.9436 \pm 19.37 \times 10^{-2}$)	-	128
	fuzzy means	3.2695	2.9723	3.0851	11	93
CPU Small	k-means	2.7046 ($2.6759 \pm 2.78 \times 10^{-2}$)	3.3665 ($3.4548 \pm 5.94 \times 10^{-2}$)	3.2512 ($3.2383 \pm 3.20 \times 10^{-2}$)	-	487
	fuzzy means	2.8591	3.3315	3.1668	17	239
Friedman	k-means	0.8639 ($0.8284 \pm 2.87 \times 10^{-2}$)	1.1623 ($1.2187 \pm 3.05 \times 10^{-2}$)	1.2388 ($1.1856 \pm 4.25 \times 10^{-2}$)	-	145
	fuzzy means	0.8278	1.0894	1.1039	7	119

Στην περίπτωση του k-means, δίνεται το βέλτιστο αποτέλεσμα μαζί με τη μέση τιμή και την τυπική απόκλιση των 20 τρεξιμάτων. Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής, λαμβάνοντας υπόψη το μέσο όρο των 20 τρεξιμάτων για τον k-means.

Πίνακας 6.4: Αποτελέσματα των αλγορίθμων k-means και fuzzy means ως προς το υπολογιστικό κόστος για την εύρεση του βέλτιστου αριθμού κέντρων N ή του βέλτιστου διαχωρισμού s .

Συν. δεδομένων	Υπολογιστικός χρόνος αναζήτησης βέλτιστου N ή s (s)	
	Αλγόριθμος	
	k-means	fuzzy means
Auto MPG	25	0.6
Boston Housing	90	1.3
CPU small	988	48
Friedman	100	6

Πίνακας 6.5: Αποτελέσματα των αλγορίθμων k-means και fuzzy means ως προς το υπολογιστικό κόστος της εκπαίδευσης.

Συν. δεδομένων	Υπολογιστικός χρόνος Εκπαίδευσης (s)	
	Αλγόριθμος	
	k-means	fuzzy means
Auto MPG	0.003 ($0.003 \pm 2.7 \times 10^{-4}$)	0.002
Boston Housing	0.01 ($0.01 \pm 0.3 \times 10^{-2}$)	0.01
CPU small	1.3 ($1.3 \pm 1.3 \times 10^{-2}$)	0.7
Friedman	0.02 ($0.02 \pm 9.95 \times 10^{-4}$)	0.02

Στην περίπτωση του k-means, δίνεται το βέλτιστο αποτέλεσμα μαζί με τη μέση τιμή και την τυπική απόκλιση των 20 τρεξιμάτων.

Στους επόμενους πίνακες (6.6 - 6.13) παρουσιάζονται τα αποτελέσματα των μεθόδων Variable Projection και Full Functional, συνδυασμένες με τον αλγόριθμο Levenberg-Marquardt, για εκκίνηση κέντρων από τους αλγορίθμους k-means και fuzzy means, τόσο για μεταβλητά όσο και για σταθερά εύρη των συναρτήσεων βάσης.

Πίνακας 6.6: Αποτελέσματα του αλγορίθμου Variable Projection - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means και για διάφορες τιμές της παραμέτρου τ , με σταθερές τιμές για τα εύρη σ .

Variable Projection, k-means, σταθερό σ , $\lambda = \tau \cdot \max(A^T A)$						
Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Επαναλήψεις	Χρόνος (s)
Auto MPG, (49)	$\tau = 10^{-1}$	2.6152 (2.5803±13.05×10 ⁻²)	2.3743 (2.5458±7.29×10 ⁻²)	2.0782 (2.1770±9.34×10 ⁻²)	0 (0±0)	0.03 (0.04±0.72×10 ⁻²)
	$\tau = 10^0$	2.6152 (2.5646±9.67×10 ⁻²)	2.3743 (2.5373±6.85×10 ⁻²)	2.0782 (2.1690±9.87×10 ⁻²)	0 (0±0)	0.04 (0.05±0.79×10 ⁻²)
	$\tau = 10^3$	2.5965 (2.5480±7.37×10 ⁻²)	2.3734 (2.5338±6.72×10 ⁻²)	2.0795 (2.1665 ±9.50×10 ⁻²)	6 (5±4)	0.08 (0.08±3.66×10 ⁻²)
	k-means	2.6152 (2.6167±5.25×10 ⁻²)	2.3743 (2.5490±7.78×10 ⁻²)	2.0782 (2.1681±9.76×10 ⁻²)	-	0.003 (0.003±2.7×10 ⁻⁴)
Friedman, (145)	$\tau = 10^{-3}$	0.7113 (0.8102±5.16×10 ⁻²)	1.0910 (1.2070±4.25×10 ⁻²)	1.0991 (1.1947±6.73×10 ⁻²)	1 (0±0)	0.5 (0.5±5.91×10 ⁻²)
	$\tau = 10^0$	0.7233 (0.6908±4.35×10 ⁻²)	1.1311 (1.1741±3.00×10 ⁻²)	1.1650 (1.1410 ±2.80×10 ⁻²)	1 (1±1)	0.4 (0.5±4.93×10 ⁻²)
	$\tau = 10^3$	0.6974 (0.7042±4.32×10 ⁻²)	1.1339 (1.1779±2.93×10 ⁻²)	1.1543 (1.1449±3.10×10 ⁻²)	10 (9±2)	1 (0.9±15.3×10 ⁻²)
	k-means	0.8639 (0.8284±2.87×10 ⁻²)	1.1623 (1.2187±3.05×10 ⁻²)	1.2388 (1.1856±4.25×10 ⁻²)	-	0.02 (0.02±9.95×10 ⁻⁴)
Boston Housing, (128)	$\tau = 10^{-3}$	1.9253 (2.0911±35.32×10 ⁻²)	2.8498 (3.0900±13.09×10 ⁻²)	2.6298 (2.9436±19.37×10 ⁻²)	0 (0±0)	0.7 (0.7±5.2×10 ⁻²)
	$\tau = 10^0$	0.9493 (1.4591±45.99×10 ⁻²)	2.8001 (2.9703±16.15×10 ⁻²)	3.1358 (3.0037±14.82×10 ⁻²)	2 (1±1)	0.9 (0.8±12.6×10 ⁻²)
	$\tau = 10^3$	0.9685 (1.6090±64.74×10 ⁻²)	2.7523 (2.9871±16.41×10 ⁻²)	3.1487 (3.0361±18.52×10 ⁻²)	11 (7±5)	1.9 (1.4±56.49×10 ⁻²)
	k-means	1.9253 (2.0911±35.32×10 ⁻²)	2.8498 (3.0900±13.09×10 ⁻²)	2.6298 (2.9436 ±19.37×10 ⁻²)	-	0.01 (0.01±0.30×10 ⁻²)
CPU Small, (487)	$\tau = 10^{-1}$	2.3433 (2.2799±5.45×10 ⁻²)	3.1631 (3.2844±6.20×10 ⁻²)	3.0825 (3.0851 ±3.58×10 ⁻²)	1 (1±0)	57 (59±3)
	$\tau = 10^0$	2.3092 (2.3151±7.65×10 ⁻²)	3.1609 (3.2989±6.29×10 ⁻²)	3.0773 (3.1007±3.44×10 ⁻²)	3 (2±1)	73 (69±5)
	$\tau = 10^3$	2.2558 (2.3066±4.31×10 ⁻²)	3.1666 (3.3029±6.42×10 ⁻²)	3.0794 (3.1011±3.67×10 ⁻²)	13 (12±1)	152 (143±6)
	k-means	2.7046 (2.6759±2.78×10 ⁻²)	3.3665 (3.4548±5.94×10 ⁻²)	3.2512 (3.2383±3.20×10 ⁻²)	-	1.3 (1.3±1.30×10 ⁻²)

Από τα 20 τρεξίματα δίνονται τα αποτελέσματα του βέλτιστου μαζί με τη μέση τιμή και την τυπική απόκλιση. Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής, λαμβάνοντας υπόψη το μέσο όρο των 20 τρεξιμάτων.

Πίνακας 6.7: Αποτελέσματα του αλγορίθμου Variable Projection - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means και για διάφορες τιμές της παραμέτρου τ , με σταθερές τιμές για τα εύρη σ .

Variable Projection, fuzzy means, σταθερό σ , $\lambda = \tau \cdot \max(A^T A)$						
Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Επαναλήψεις	Χρόνος (s)
Auto MPG, (10,43)	$\tau = 10^{-1}$	2.6174	2.3417	2.0659	0	0.1
	$\tau = 10^0$	2.6174	2.3417	2.0659	0	0.1
	$\tau = 10^3$	2.6174	2.3417	2.0659	0	0.1
	fuzzy means	2.6174	2.3417	2.0659	-	0.002
Friedman, (7,119)	$\tau = 10^{-3}$	0.8278	1.0894	1.1039	0	0.4
	$\tau = 10^0$	0.7110	1.0421	1.0782	1	0.4
	$\tau = 10^3$	0.6905	1.0399	1.0885	10	0.8
	fuzzy means	0.8278	1.0894	1.1039	-	0.02
Boston Housing, (11,93)	$\tau = 10^{-3}$	3.2695	2.9723	3.0851	0	0.5
	$\tau = 10^0$	2.3566	2.7838	2.8368	1	0.5
	$\tau = 10^3$	2.8778	2.8316	2.9130	8	1
	fuzzy means	3.2695	2.9723	3.0851	-	0.01
CPU Small, (17,239)	$\tau = 10^{-1}$	2.5744	3.2297	3.0449	1	30
	$\tau = 10^0$	2.6098	3.2049	3.0200	2	30
	$\tau = 10^3$	2.5623	3.2015	3.0096	12	67
	fuzzy means	2.8591	3.3315	3.1668	-	0.7

Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής.

Πίνακας 6.8: Αποτελέσματα του αλγορίθμου Variable Projection - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means και για διάφορες τιμές της παραμέτρου τ , με μεταβλητές τιμές για τα εύρη σ .

Variable Projection, k-means, μεταβλητό σ , $\lambda = \tau \cdot \max(A^T A)$						
Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Επαναλήψεις	Χρόνος (s)
Auto MPG, (49)	$\tau = 10^{-1}$	2.6152 (2.5524±15.92×10 ⁻²)	2.3743 (2.5350±6.90×10 ⁻²)	2.0782 (2.1935±10.70×10 ⁻²)	0 (0±1)	0.03 (0.03±1×10 ⁻²)*
	$\tau = 10^0$	2.3497 (2.4853±15.32×10 ⁻²)	2.3732 (2.5267±6.80×10 ⁻²)	2.0276 (2.1890±9.40×10 ⁻²)	1 (0±1)	0.06 (0.06±1.03×10 ⁻²)
	$\tau = 10^3$	2.5432 (2.5648±9.01×10 ⁻²)	2.3618 (2.5325±7.10×10 ⁻²)	2.0949 (2.1654 ±8.89×10 ⁻²)	8 (4±4)	0.1 (0.08±3.62×10 ⁻²)
	k-means	2.6152 (2.6167±5.25×10 ⁻²)	2.3743 (2.5490±7.78×10 ⁻²)	2.0782 (2.1681±9.76×10 ⁻²)	-	0.003 (0.003±2.7×10 ⁻⁴)
Friedman, (145)	$\tau = 10^{-3}$	0.6184 (0.6256±1.72×10 ⁻²)	1.0783 (1.1542±2.89×10 ⁻²)	1.1449 (1.1545±3.58×10 ⁻²)	1 (1±0)	0.6 (0.6±4.46×10 ⁻²)
	$\tau = 10^0$	0.6416 (0.6713±4.01×10 ⁻²)	1.0949 (1.1396±2.88×10 ⁻²)	1.1305 (1.1341 ±3.76×10 ⁻²)	2 (2±1)	0.5 (0.5±5.64×10 ⁻²)
	$\tau = 10^3$	0.6358 (0.6686±3.53×10 ⁻²)	1.0986 (1.1421±3.10×10 ⁻²)	1.1618 (1.1381±3.74×10 ⁻²)	11 (11±1)	1.2 (1.2±16.48×10 ⁻²)
	k-means	0.8639 (0.8284±2.87×10 ⁻²)	1.1623 (1.2187±3.05×10 ⁻²)	1.2388 (1.1856±4.25×10 ⁻²)	-	0.02 (0.02±9.95×10 ⁻⁴)
Boston Housing, (128)	$\tau = 10^{-3}$	1.2977 (1.6071±62.18×10 ⁻²)	2.7638 (3.0329±16.02×10 ⁻²)	2.7495 (3.0402±21.96×10 ⁻²)	1 (1±1)	0.9 (0.9±11.49×10 ⁻²)
	$\tau = 10^0$	1.0645 (1.6274±59.53×10 ⁻²)	2.7736 (3.0089±14.82×10 ⁻²)	2.7700 (2.9666±17.59×10 ⁻²)	4 (1±1)	1.2 (1±15.34×10 ⁻²)
	$\tau = 10^3$	1.0205 (1.8136±53.19×10 ⁻²)	2.7817 (3.0322±14.61×10 ⁻²)	2.9182 (2.9623±18.97×10 ⁻²)	14 (5±5)	2.8 (1.6±72.25×10 ⁻²)
	k-means	1.9253 (2.0911±35.32×10 ⁻²)	2.8498 (3.0900±13.09×10 ⁻²)	2.6298 (2.9436 ±19.37×10 ⁻²)	-	0.01 (0.01±0.30×10 ⁻²)
CPU Small, (487)	$\tau = 10^{-1}$	2.3032 (2.2800±3.36×10 ⁻²)	3.1249 (3.2339±5.84×10 ⁻²)	3.0573 (3.0512 ±4.00×10 ⁻²)	1 (1±0)	66 (68±2)
	$\tau = 10^0$	2.2287 (2.2786±6.55×10 ⁻²)	3.1288 (3.2484±6.27×10 ⁻²)	3.0544 (3.0651±3.57×10 ⁻²)	3 (3±1)	83 (82±6)
	$\tau = 10^3$	2.3033 (2.2895±5.34×10 ⁻²)	3.1352 (3.2563±6.46×10 ⁻²)	3.0670 (3.0758±3.18×10 ⁻²)	12 (12±1)	164 (165±6)
	k-means	2.7046 (2.6759±2.78×10 ⁻²)	3.3665 (3.4548±5.94×10 ⁻²)	3.2512 (3.2383±3.20×10 ⁻²)	-	1.3 (1.3±1.30×10 ⁻²)

Από τα 20 τρεξίματα δίνονται τα αποτελέσματα του βέλτιστου μαζί με τη μέση τιμή και την τυπική απόκλιση. Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής, λαμβάνοντας υπόψη το μέσο όρο των 20 τρεξιμάτων.

Πίνακας 6.9: Αποτελέσματα του αλγορίθμου Variable Projection - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means και για διάφορες τιμές της παραμέτρου τ , με μεταβλητές τιμές για τα εύρη σ .

Variable Projection, fuzzy means, μεταβλητό σ , $\lambda = \tau \cdot \max(A^T A)$						
Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Επαναλήψεις	Χρόνος (s)
Auto MPG, (10,43)	$\tau = 10^{-1}$	2.6174	2.3417	2.0659	0	0.3
	$\tau = 10^0$	2.6174	2.3417	2.0659	0	0.2
	$\tau = 10^3$	2.6174	2.3417	2.0659	0	0.1
	fuzzy means	2.6174	2.3417	2.0659	-	0.002
Friedman, (7,119)	$\tau = 10^{-3}$	0.8278	1.0894	1.1039	0	0.4
	$\tau = 10^0$	0.6972	1.0117	1.0605	2	1.1
	$\tau = 10^3$	0.7065	1.0154	1.0600	11	1.6
	fuzzy means	0.8278	1.0894	1.1039	-	0.02
Boston Housing, (11,93)	$\tau = 10^{-3}$	1.9240	2.8401	3.1102	1	1
	$\tau = 10^0$	2.5399	2.8539	3.0231	1	0.6
	$\tau = 10^3$	2.9822	2.8526	2.9975	8	1.1
	fuzzy means	3.2695	2.9723	3.0851	-	0.01
CPU Small, (17,239)	$\tau = 10^{-1}$	2.3597	3.1257	2.9826	3	35
	$\tau = 10^0$	2.5120	3.1362	2.9638	4	43
	$\tau = 10^3$	2.4891	3.1391	2.9586	14	98
	fuzzy means	2.8591	3.3315	3.1668	-	0.7

Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής.

Πίνακας 6.10: Αποτελέσματα του αλγορίθμου Full functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means και για διάφορες τιμές της παραμέτρου τ , με σταθερές τιμές για τα εύρη σ .

Full Functional, k-means, σταθερό σ , $\lambda = \tau \cdot \max(A^T A)$						
Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Επαναλήψεις	Χρόνος (s)
Auto MPG, (49)	$\tau = 10^{-1}$	2.5776 (2.5032±10.53×10 ⁻²)	2.3720 (2.5143±5.97×10 ⁻²)	2.0487 (2.1112±7.86×10 ⁻²)	2 (2±2)	0.05 (0.06±1.45×10 ⁻²)
	$\tau = 10^0$	2.5760 (2.5008±10.26×10 ⁻²)	2.3727 (2.5136±5.92×10 ⁻²)	2.0478 (2.1098 ±7.97×10 ⁻²)	5 (5±3)	0.07 (0.08±2.40×10 ⁻²)
	$\tau = 10^3$	2.5740 (2.5408±11.64×10 ⁻²)	2.3727 (2.5225±6.75×10 ⁻²)	2.0467 (2.1281±9.35×10 ⁻²)	15 (9±8)	0.1 (0.1±5.60×10 ⁻²)
	k-means	2.6152 (2.6167±5.25×10 ⁻²)	2.3743 (2.5490±7.78×10 ⁻²)	2.0782 (2.1681±9.76×10 ⁻²)	-	0.003 (0.003±2.7×10 ⁻³)
Friedman, (145)	$\tau = 10^{-3}$	0.6413 (0.6767±3.76×10 ⁻²)	1.1363 (1.1660±2.01×10 ⁻²)	1.1113 (1.1301 ±2.98×10 ⁻²)	3 (2±1)	0.5 (0.5±5.14×10 ⁻²)
	$\tau = 10^0$	0.6682 (0.6822±3.41×10 ⁻²)	1.1339 (1.1650±2.00×10 ⁻²)	1.1073 (1.1317±2.87×10 ⁻²)	12 (12±1)	1 (1±6.87×10 ⁻²)
	$\tau = 10^3$	0.8639 (0.8284±2.87×10 ⁻²)	1.1623 (1.2187±3.05×10 ⁻²)	1.2388 (1.1856±4.25×10 ⁻²)	0 (0±0)	0.4 (0.4±3.25×10 ⁻²)
	k-means	0.8639 (0.8284±2.87×10 ⁻²)	1.1623 (1.2187±3.05×10 ⁻²)	1.2388 (1.1856±4.25×10 ⁻²)	-	0.02 (0.02±9.95×10 ⁻⁴)
Boston Housing, (128)	$\tau = 10^{-3}$	0.9885 (1.2909±39.68×10 ⁻²)	2.6650 (2.8699±11.54×10 ⁻²)	2.8534 (2.9285±14.06×10 ⁻²)	5 (2±1)	1.5 (1.1±20.65×10 ⁻²)
	$\tau = 10^0$	0.9786 (1.2587±31.63×10 ⁻²)	2.6637 (2.8749±12.04×10 ⁻²)	2.8547 (2.9224±12.78×10 ⁻²)	12 (10±2)	2 (2±27.1×10 ⁻²)
	$\tau = 10^3$	0.9725 (1.6940±65.90×10 ⁻²)	2.6629 (2.9623±15.44×10 ⁻²)	2.8592 (2.9218 ±17.28×10 ⁻²)	22 (10±11)	3.7 (1.9±1.3)
	k-means	1.9253 (2.0911±35.32×10 ⁻²)	2.8498 (3.0900±13.09×10 ⁻²)	2.6298 (2.9436±19.37×10 ⁻²)	-	0.01 (0.01±0.30×10 ⁻²)
CPU Small, (487)	$\tau = 10^{-1}$	2.1813 (2.1876±7.32×10 ⁻²)	3.1555 (3.2505±5.31×10 ⁻²)	3.0255 (3.0305±3.15×10 ⁻²)	10 (10±1)	90 (92±5)
	$\tau = 10^0$	2.2163 (2.1993±7.68×10 ⁻²)	3.1539 (3.2496±5.26×10 ⁻²)	3.0266 (3.0304 ±3.07×10 ⁻²)	13 (13±1)	107 (110±6)
	$\tau = 10^3$	2.2430 (2.6542±10.07×10 ⁻²)	3.2342 (3.4467±7.64×10 ⁻²)	3.0037 (3.2284±6.12×10 ⁻²)	23 (1±5)	160 (43±28)
	k-means	2.7046 (2.6759±2.78×10 ⁻²)	3.3665 (3.4548±5.94×10 ⁻²)	3.2512 (3.2383±3.20×10 ⁻²)	-	1.3 (1.3±1.30×10 ⁻²)

Από τα 20 τρεξίματα δίνονται τα αποτελέσματα του βέλτιστου μαζί με τη μέση τιμή και την τυπική απόκλιση. Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής, λαμβάνοντας υπόψη το μέσο όρο των 20 τρεξιμάτων.

Πίνακας 6.11: Αποτελέσματα του αλγορίθμου Full functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means και για διάφορες τιμές της παραμέτρου τ , με σταθερές τιμές για τα εύρη σ .

Full Functional, fuzzy means, σταθερό σ , $\lambda = \tau \cdot \max(A^T A)$						
Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Επαναλήψεις	Χρόνος (s)
Auto MPG, (10,43)	$\tau = 10^{-1}$	2.5245	2.3400	2.0223	4	0.1
	$\tau = 10^0$	2.6174	2.3417	2.0659	0	0.1
	$\tau = 10^3$	2.6174	2.3417	2.0659	0	0.1
	fuzzy means	2.6174	2.3417	2.0659	-	0.002
Friedman, (7,119)	$\tau = 10^{-3}$	0.7253	1.0312	1.0691	2	0.4
	$\tau = 10^0$	0.7377	1.0287	1.0683	11	0.3
	$\tau = 10^3$	0.8278	1.0894	1.1039	0	0.8
	fuzzy means	0.8278	1.0894	1.1039	-	0.02
Boston Housing, (11,93)	$\tau = 10^{-3}$	2.2339	2.7178	2.7621	2	0.6
	$\tau = 10^0$	2.1965	2.7270	2.7893	10	1.2
	$\tau = 10^3$	2.5325	2.7273	2.7936	19	1.8
	fuzzy means	3.2695	2.9723	3.0851	-	0.01
CPU Small, (17,239)	$\tau = 10^{-1}$	2.4375	3.1025	2.9145	15	38
	$\tau = 10^0$	2.4972	3.1063	2.9249	16	37
	$\tau = 10^3$	2.8591	3.3315	3.1668	0	12
	fuzzy means	2.8591	3.3315	3.1668	-	0.7

Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής.

Πίνακας 6.12: Αποτελέσματα του αλγορίθμου Full functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means και για διάφορες τιμές της παραμέτρου τ , με μεταβλητές τιμές για τα εύρη σ .

Full Functional, k-means, μεταβλητό σ , $\lambda = \tau \cdot \max(A^T A)$						
Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Επαναλήψεις	Χρόνος (s)
Auto MPG, (49)	$\tau = 10^{-1}$	2.5236 (2.4914±9.64×10 ⁻²)	2.3572 (2.5049±5.74×10 ⁻²)	2.0039 (2.1001 ±8.64×10 ⁻²)	3 (2±1)	0.08 (0.08±1.64×10 ⁻²)
	$\tau = 10^0$	2.5303 (2.4987±9.87×10 ⁻²)	2.3587 (2.5047±5.68×10 ⁻²)	2.0075 (2.1047±8.14×10 ⁻²)	6 (5±2)	0.1 (0.1±2.41×10 ⁻²)
	$\tau = 10^3$	2.5276 (2.5178±7.21×10 ⁻²)	2.3586 (2.5067±6.31×10 ⁻²)	2.0062 (2.1031±8.48×10 ⁻²)	16 (13±6)	0.3 (0.2±6.25×10 ⁻²)
	k-means	2.6152 (2.6167±5.25×10 ⁻²)	2.3743 (2.5490±7.78×10 ⁻²)	2.0782 (2.1681±9.76×10 ⁻²)	-	0.003 (0.003±2.7×10 ⁻³)
Friedman, (145)	$\tau = 10^{-3}$	0.6530 (0.6720±2.73×10 ⁻²)	1.0900 (1.1297±2.24×10 ⁻²)	1.1137 (1.1217±2.74×10 ⁻²)	2 (2±1)	0.6 (0.6±6.46×10 ⁻²)
	$\tau = 10^0$	0.6674 (0.6746±1.61×10 ⁻²)	1.0928 (1.1288±2.08×10 ⁻²)	1.1153 (1.1197 ±2.63×10 ⁻²)	11 (11±1)	1.2 (1.2±12.73×10 ⁻²)
	$\tau = 10^3$	0.6851 (0.8138±5.77×10 ⁻²)	1.1319 (1.2114±3.62×10 ⁻²)	1.0919 (1.1833±4.56×10 ⁻²)	20 (2±6)	1.8 (0.6±47.04×10 ⁻²)
	k-means	0.8639 (0.8284±2.87×10 ⁻²)	1.1623 (1.2187±3.05×10 ⁻²)	1.2388 (1.1856±4.25×10 ⁻²)	-	0.02 (0.02±9.95×10 ⁻⁴)
Boston Housing, (128)	$\tau = 10^{-3}$	0.9336 (1.5351±60.62×10 ⁻²)	2.6756 (2.9100±12.88×10 ⁻²)	2.8482 (2.8786±13.72×10 ⁻²)	5 (2±2)	1.6 (1.1±24.86×10 ⁻²)
	$\tau = 10^0$	0.9162 (1.4658±65.72×10 ⁻²)	2.6714 (2.9198±14.28×10 ⁻²)	2.8319 (2.8527 ±19.22×10 ⁻²)	12 (8±5)	2.7 (2±71.85×10 ⁻²)
	$\tau = 10^3$	0.9098 (1.8849±57.42×10 ⁻²)	2.6708 (3.0164±17.35×10 ⁻²)	2.8337 (2.9005±21.18×10 ⁻²)	22 (5±9)	4 (1.6±1.3)
	k-means	1.9253 (2.0911±35.32×10 ⁻²)	2.8498 (3.0900±13.09×10 ⁻²)	2.6298 (2.9436±19.37×10 ⁻²)	-	0.01 (0.01±0.30×10 ⁻²)
CPU Small, (487)	$\tau = 10^{-1}$	2.2833 (2.1572±10.43×10 ⁻²)	3.1182 (3.2009±4.97×10 ⁻²)	2.9972 (2.9936 ±2.58×10 ⁻²)	9 (10±2)	105 (115±13)
	$\tau = 10^0$	2.2082 (2.1675±9.27×10 ⁻²)	3.1162 (3.2007±4.96×10 ⁻²)	2.9840 (2.9947±2.57×10 ⁻²)	13 (14±1)	136 (136±3)
	$\tau = 10^3$	1.9673 (2.6404±16.08×10 ⁻²)	3.1683 (3.4434±8.67×10 ⁻²)	2.9491 (3.2256±7.20×10 ⁻²)	26 (1±6)	218 (51±40)
	k-means	2.7046 (2.6759±2.78×10 ⁻²)	3.3665 (3.4548±5.94×10 ⁻²)	3.2512 (3.2383±3.20×10 ⁻²)	-	1.3 (1.3±1.30×10 ⁻²)

Από τα 20 τρεξίματα δίνονται τα αποτελέσματα του βέλτιστου μαζί με τη μέση τιμή και την τυπική απόκλιση. Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής, λαμβάνοντας υπόψη το μέσο όρο των 20 τρεξιμάτων.

Πίνακας 6.13: Αποτελέσματα του αλγορίθμου Full functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means και για διάφορες τιμές της παραμέτρου τ , με μεταβλητές τιμές για τα εύρη σ .

Full Functional, fuzzy means, μεταβλητό σ , $\lambda = \tau \cdot \max(A^T A)$						
Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Επαναλήψεις	Χρόνος (s)
Auto MPG, (10,43)	$\tau = 10^{-1}$	2.6174	2.3417	2.0659	0	0.1
	$\tau = 10^0$	2.6174	2.3417	2.0659	0	0.1
	$\tau = 10^3$	2.6174	2.3417	2.0659	0	0.1
	fuzzy means	2.6174	2.3417	2.0659	-	0.002
Friedman, (7,119)	$\tau = 10^{-3}$	0.6924	1.0080	1.0546	4	0.6
	$\tau = 10^0$	0.7155	1.0077	1.0512	13	1.3
	$\tau = 10^3$	0.8278	1.0894	1.1039	0	0.4
	fuzzy means	0.8278	1.0894	1.1039	-	0.02
Boston Housing, (11,93)	$\tau = 10^{-3}$	1.4068	2.7393	3.3513	5	1
	$\tau = 10^0$	2.3564	2.7209	2.8753	9	1.2
	$\tau = 10^3$	2.3406	2.7206	2.8754	19	2
	fuzzy means	3.2695	2.9723	3.0851	-	0.01
CPU Small, (17,239)	$\tau = 10^{-1}$	2.5159	3.0827	2.9227	15	42
	$\tau = 10^0$	2.4836	3.0798	2.9207	20	49
	$\tau = 10^3$	2.8591	3.3315	3.1668	0	14
	fuzzy means	2.8591	3.3315	3.1668	-	0.7

Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής.

Στους επόμενους πίνακες συνοψίζονται τα βέλτιστα αποτελέσματα των μεθόδων Variable Projection και Full Functional, συνδυασμένες με τον αλγόριθμο Levenberg-Marquardt, για εκκίνηση κέντρων από τους αλγορίθμους k-means (πίνακας 6.14) και fuzzy means (πίνακας 6.15), για εκείνες τις τιμές του τ , οι οποίες δίνουν το ελάχιστο σφάλμα στο σύνολο δοκιμής, τόσο για σταθερά όσο και για μεταβλητά εύρη, σ . Στην περίπτωση του k-means επιλέγονται εκείνα με το ελάχιστο μέσο κόστος στο σύνολο δοκιμής.

Πίνακας 6.14: Αποτελέσματα των αλγορίθμων Variable Projection, Full Functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο k-means ως προς το βέλτιστο μέσο σφάλμα στο σύνολο δοκιμής για μεταβλητά και σταθερά εύρη σ .

Συν. δεδομένων	Αλγόριθμος	Variable Projection, Full Functional, k-means, $\lambda = \tau \cdot \max(A^T A)$			Επαναλήψεις	Χρόνος (s)
		RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής		
Auto MPG, (49)	$\tau = 10^3$, VP, σταθ. σ	2.5965 (2.5480±7.37×10 ⁻²)	2.3734 (2.5338±6.72×10 ⁻²)	2.0795 (2.1665±9.50×10 ⁻²)	6 (5±4)	0.08 (0.08±3.66×10 ⁻²)
	$\tau = 10^3$, VP, μετ. σ	2.5432 (2.5648±9.01×10 ⁻²)	2.3618 (2.5325±7.10×10 ⁻²)	2.0949 (2.1654±8.89×10 ⁻²)	8 (4±4)	0.1 (0.08±3.62×10 ⁻²)
	$\tau = 10^4$, FF, σταθ. σ	2.5760 (2.5008±10.26×10 ⁻²)	2.3727 (2.5136±5.92×10 ⁻²)	2.0478 (2.1098±7.97×10 ⁻²)	5 (5±3)	0.07 (0.08±2.40×10 ⁻²)
	$\tau = 10^4$, FF, μετ. σ	2.5276 (2.5178±7.21×10 ⁻²)	2.3586 (2.5067±6.31×10 ⁻²)	2.0062 (2.1031±8.48×10 ⁻²)	16 (13±6)	0.3 (0.2±6.25×10 ⁻²)
	k-means	2.6152 (2.6167±5.25×10 ⁻²)	2.3743 (2.5490±7.78×10 ⁻²)	2.0782 (2.1681±9.76×10 ⁻²)	-	0.003 (0.003±2.7×10 ⁻⁴)
Friedman, (145)	$\tau = 10^0$, VP, σταθ. σ	0.7233 (0.6908±4.35×10 ⁻²)	1.1311 (1.1741±3.00×10 ⁻²)	1.1650 (1.1410±2.80×10 ⁻²)	1 (1±1)	0.4 (0.5±4.93×10 ⁻²)
	$\tau = 10^0$, VP, μετ. σ	0.6416 (0.6713±4.01×10 ⁻²)	1.0949 (1.1396±2.88×10 ⁻²)	1.1305 (1.1341±3.76×10 ⁻²)	2 (2±1)	0.5 (0.5±5.64×10 ⁻²)
	$\tau = 10^{-3}$, FF, σταθ. σ	0.6413 (0.6767±3.76×10 ⁻²)	1.1363 (1.1660±2.01×10 ⁻²)	1.1113 (1.1301±2.98×10 ⁻²)	3 (2±1)	0.5 (0.5±5.14×10 ⁻²)
	$\tau = 10^0$, FF, μετ. σ	0.6674 (0.6746±1.61×10 ⁻²)	1.0928 (1.1288±2.08×10 ⁻²)	1.1153 (1.1197±2.63×10 ⁻²)	11 (11±1)	1.2 (1.2±12.73×10 ⁻²)
	k-means	0.8639 (0.8284±2.87×10 ⁻²)	1.1623 (1.2187±3.05×10 ⁻²)	1.2388 (1.1856±4.25×10 ⁻²)	-	0.02 (0.02±9.95×10 ⁻⁴)
Boston Housing, (128)	$\tau = 10^{-3}$, VP, σταθ. σ	1.9253 (2.0911±35.32×10 ⁻²)	2.8498 (3.0900±13.09×10 ⁻²)	2.6298 (2.9436±19.37×10 ⁻²)	0 (0±0)	0.7 (0.7±5.2×10 ⁻²)
	$\tau = 10^3$, VP, μετ. σ	1.0205 (1.8136±53.19×10 ⁻²)	2.7817 (3.0322±14.61×10 ⁻²)	2.9182 (2.9623±18.97×10 ⁻²)	14 (5±5)	2.8 (1.6±72.25×10 ⁻²)
	$\tau = 10^3$, FF, σταθ. σ	0.9725 (1.6940±65.90×10 ⁻²)	2.6629 (2.9623±15.44×10 ⁻²)	2.8592 (2.9218±17.28×10 ⁻²)	22 (10±11)	3.7 (1.9±1.3)
	$\tau = 10^0$, FF, μετ. σ	0.9162 (1.4658±65.72×10 ⁻²)	2.6714 (2.9198±14.28×10 ⁻²)	2.8319 (2.8527±19.22×10 ⁻²)	12 (8±5)	2.7 (2±71.85×10 ⁻²)
	k-means	1.9253 (2.0911±35.32×10 ⁻²)	2.8498 (3.0900±13.09×10 ⁻²)	2.6298 (2.9436±19.37×10 ⁻²)	-	0.01 (0.01±0.30×10 ⁻²)
CPU Small, (487)	$\tau = 10^{-1}$, VP, σταθ. σ	2.3433 (2.2799±5.45×10 ⁻²)	3.1631 (3.2844±6.20×10 ⁻²)	3.0825 (3.0851±3.58×10 ⁻²)	1 (1±0)	57 (59±3)
	$\tau = 10^{-1}$, VP, μετ. σ	2.3032 (2.2800±3.36×10 ⁻²)	3.1249 (3.2339±5.84×10 ⁻²)	3.0573 (3.0512±4.00×10 ⁻²)	1 (1±0)	66 (68±2)
	$\tau = 10^0$, FF, σταθ. σ	2.2163 (2.1993±7.68×10 ⁻²)	3.1539 (3.2496±5.26×10 ⁻²)	3.0266 (3.0304±3.07×10 ⁻²)	13 (13±1)	107 (110±6)
	$\tau = 10^0$, FF, μετ. σ	2.2833 (2.1572±10.43×10 ⁻²)	3.1182 (3.2009±4.97×10 ⁻²)	2.9972 (2.9936±2.58×10 ⁻²)	9 (10±2)	105 (115±13)
	k-means	2.7046 (2.6759±2.78×10 ⁻²)	3.3665 (3.4548±5.94×10 ⁻²)	3.2512 (3.2383±3.20×10 ⁻²)	-	1.3 (1.3±1.30×10 ⁻²)

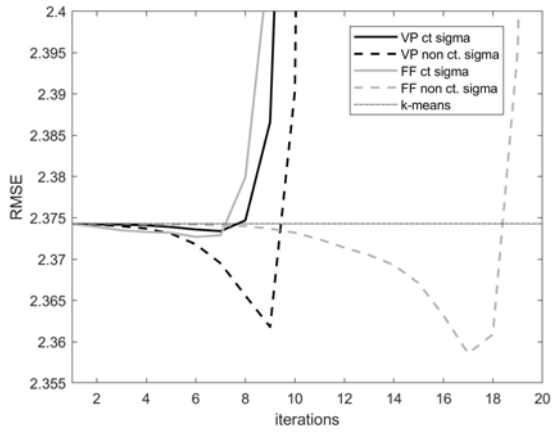
Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής, λαμβάνοντας υπόψη το μέσο όρο των 20 τρεξιμάτων.

Πίνακας 6.15: Αποτελέσματα των αλγορίθμων Variable Projection, Full Functional - Levenberg Marquardt για εκκίνηση κέντρων από τον αλγόριθμο fuzzy means ως προς το βέλτιστο σφάλμα στο σύνολο δοκιμής για μεταβλητά και σταθερά εύρη σ .

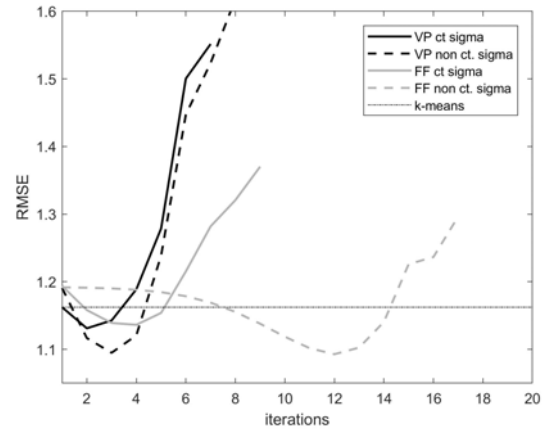
Variable Projection, Full Functional, fuzzy means, $\lambda = \tau \cdot \max(A^T A)$						
Συν. δεδομένων	Αλγόριθμος	RMSE Εκπαίδευσης	RMSE Επικύρωσης	RMSE Δοκιμής	Επαναλήψεις	Χρόνος (s)
Auto MPG, (10,43)	τ , VP, σταθ. σ	2.6174	2.3417	2.0659	0	0.1
	τ , VP, μετ. σ	2.6174	2.3417	2.0659	0	0.1
	$\tau = 10^{-1}$, FF, σταθ. σ	2.5245	2.3400	2.0223	4	0.1
	τ , FF, μετ. σ	2.6174	2.3417	2.0659	0	0.1
	fuzzy means	2.6174	2.3417	2.0659	-	0.002
Friedman, (7,119)	$\tau = 10^0$, VP, σταθ. σ	0.7110	1.0421	1.0782	1	0.4
	$\tau = 10^3$, VP, μετ. σ	0.7065	1.0154	1.0600	11	1.6
	$\tau = 10^0$, FF, σταθ. σ	0.7377	1.0287	1.0683	11	0.3
	$\tau = 10^0$, FF, μετ. σ	0.7155	1.0077	1.0512	13	1.3
	fuzzy means	0.8278	1.0894	1.1039	-	0.02
Boston Housing, (11,93)	$\tau = 10^0$, VP, σταθ. σ	2.3566	2.7838	2.8368	1	0.5
	$\tau = 10^3$, VP, μετ. σ	2.9822	2.8526	2.9975	8	1.1
	$\tau = 10^{-3}$, FF, σταθ. σ	2.2339	2.7178	2.7621	2	0.6
	$\tau = 10^0$, FF, μετ. σ	2.3564	2.7209	2.8753	9	1.2
	fuzzy means	3.2695	2.9723	3.0851	-	0.01
CPU Small, (17,239)	$\tau = 10^3$, VP, σταθ. σ	2.5623	3.2015	3.0096	12	67
	$\tau = 10^3$, VP, μετ. σ	2.4891	3.1391	2.9586	14	98
	$\tau = 10^{-1}$, FF, σταθ. σ	2.4375	3.1025	2.9145	15	38
	$\tau = 10^{-1}$, FF, μετ. σ	2.5159	3.0827	2.9227	15	42
	fuzzy means	2.8591	3.3315	3.1668	-	0.7

Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής.

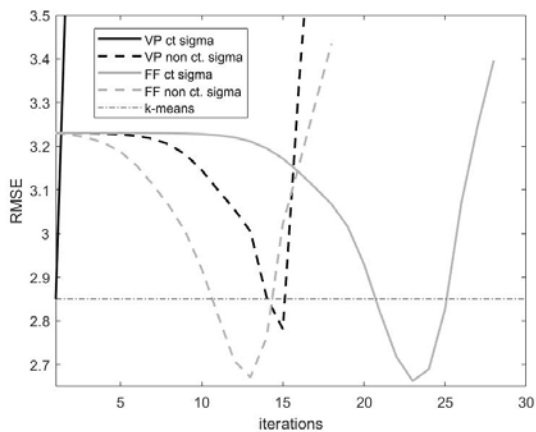
Στα επόμενα διαγράμματα (σχήμα 6.1 - 6.2) παρουσιάζεται η πορεία που ακολουθεί το σφάλμα στο σύνολο επικύρωσης για κάθε μια από τις μεθοδολογίες. Στην περίπτωση της εκκίνησης με κέντρα από τον αλγόριθμο k-means, αναπαριστάται εκείνη η πορεία που αντιστοιχεί στη βέλτιστη τιμή από τα 20 τρεξίματα.



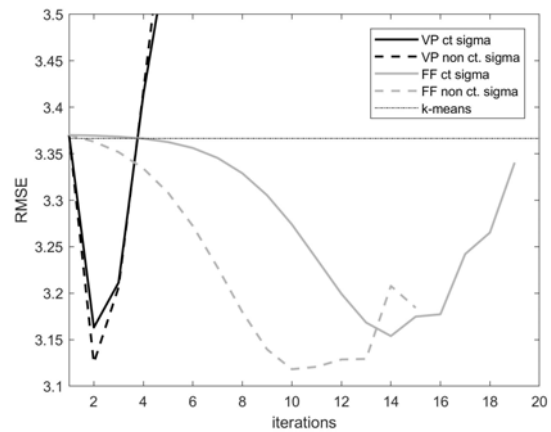
(α) AutoMPG



(β) Friedman

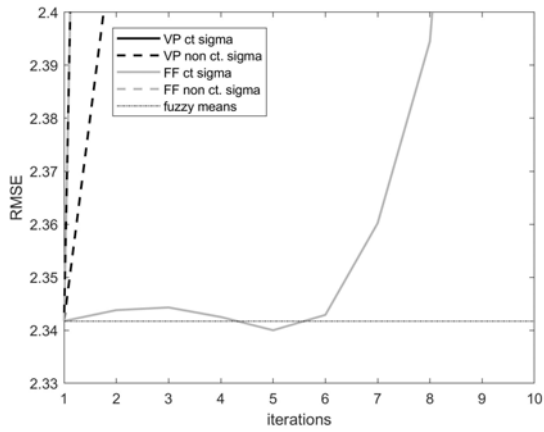


(γ) Boston Housing

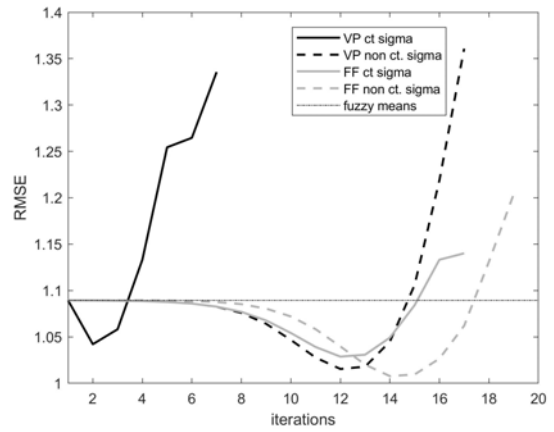


(δ) CPUSmall

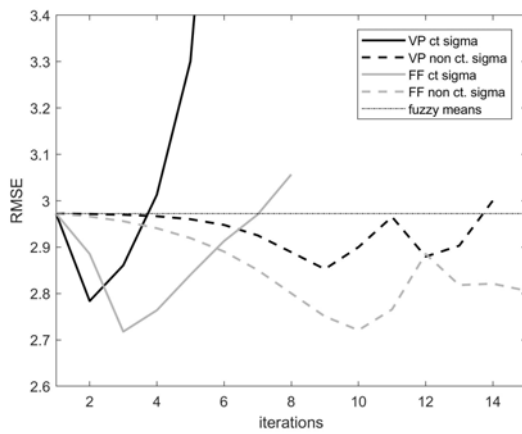
Σχήμα 6.1: Η μεταβολή στη βέλτιστη λύση του κόστους στο σύνολο επικύρωσης για τις διάφορες μεθοδολογίες με εκκίνηση κέντρων από τον αλγόριθμο k-means.



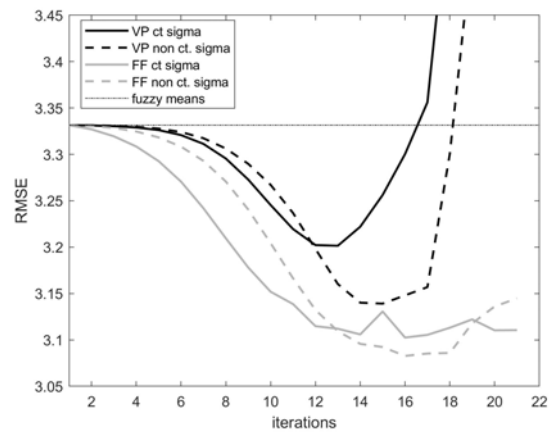
(α) AutoMPG



(β) Friedman



(γ) Boston Housing



(δ) CPU Small

Σχήμα 6.2: Η μεταβολή στη βέλτιστη λύση του κόστους στο σύνολο επικύρωσης για τους διάφορους αλγόριθμους με εκκίνηση από τον αλγόριθμο fuzzy means.

Στην περίπτωση της εκκίνησης των κέντρων με τον αλγόριθμο k-means, λόγω της στοχαστικής του φύσης, σε κάθε εφαρμογή του αλγορίθμου επιστρέφονται διαφορετικά αποτελέσματα. Για το λόγο αυτό και για την απόκτηση συνεπών αποτελεσμάτων, πρέπει να γίνει μια πληθώρα εφαρμογών των οποίων τα αποτελέσματα αναγράφονται με τη μέγιστη τιμή και τη μέση τιμή μαζί με την τυπική απόκλιση σε παρένθεση. Σε όλες τις περιπτώσεις φαίνεται η τιμή της τυπικής απόκλισης να είναι σχετικά μικρή, διασφαλίζοντας έτσι την εγγυρότητα των αποτελεσμάτων.

Για να εξεταστεί η τυχόν επίδραση της παραμέτρου λ στον αλγόριθμο Levenberg-Marquardt, δοκιμάστηκαν διάφορες τιμές σε κάθε μεθοδολογία και στα 4 σύνολα δεδομένων της βιβλιογραφίας. Αν σκεφτεί κανείς τον αλγόριθμο Levenberg-Marquardt ως μια μέθοδο περιοχής εμπιστοσύνης, η οποία εναλλάσσεται από τη μέθοδο απότομης κατάβασης (μεγάλες τιμές λ) στη μέθοδο Newton (μικρές τιμές λ), τότε δικαιολογείται το γεγονός ότι εκκίνηση με μικρές τιμές λ έχει ως αποτέλεσμα λιγότερες επαναλήψεις με μεγάλα βήματα, ενώ μεγάλες τιμές λ περισσότερες επαναλήψεις με μικρά βήματα. Από τα σύνολα δεδομένων της βιβλιογραφίας φαίνεται να μην υπάρχει μια συγκεκριμένη τιμή λ στην οποία όλες οι μεθοδολογίες να ανταποκρίνονται καλά όσον αφορά τα σφάλματα στο σύνολο επικύρωσης και δοκιμής ή, και ακόμα, να εξελίσσεται το σφάλμα με τον ίδιο τρόπο αυξάνοντας ή μειώνοντας την τιμή του λ . Ωστόσο, για τις διάφορες τιμές της παραμέτρου λ φαίνεται στις περισσότερες περιπτώσεις να ξεπερνιέται το μέσο ή το βέλτιστο σφάλμα του συνόλου δοκιμής για εκκίνηση κέντρων από k-means, ή το σφάλμα για εκκίνηση κέντρων από fuzzy means, κάτι το οποίο επιβεβαιώνεται από τους πίνακες 6.14 και 6.15.

Από τα αποτελέσματα των προηγούμενων σελίδων κανείς μπορεί να δει ότι, γενικά, η επιλογή της εκκίνησης των κέντρων με τον αλγόριθμο fuzzy means οδηγεί σε καλύτερα αποτελέσματα όσον αφορά τις βέλτιστες τιμές του RMSE στο σύνολο επικύρωσης και δοκιμής στις περισσότερες των περιπτώσεων. Αν ληφθεί υπόψη η μέση τιμή του RMSE στις μεθοδολογίες με εκκίνηση κέντρων από τον αλγόριθμο k-means, τότε κανείς παρατηρεί την υπεροχή του fuzzy means έναντι του προηγούμενου. Ακόμα, με αυτή την επιλογή περιορίζεται το μέγεθος των δικτύων σε όλα τα σύνολα δεδομένων της βιβλιογραφίας, κάτι το οποίο αντανάκλα στους απαιτούμενους υπολογιστικούς χρόνους της εκπαίδευσης, κυρίως στα μεγάλα σύνολα δεδομένων.

Η εισαγωγή μεταβλητών ευρών για τις συναρτήσεις βάσης οδηγεί, γενικά, σε βελτιωμένα αποτελέσματα, χωρίς να λείπουν βεβαίως κι εδώ οι ανάλογες εξαιρέσεις σε συγκεκριμένες περιπτώσεις.

Όσον αφορά τις δύο κύριες μεθοδολογίες, του μειωμένου συναρτησιακού (VP) και του πλήρους συναρτησιακού (FF), υπάρχουν ορισμένες σημαντικές παρατηρήσεις. Το πρώτο που παρατηρεί κανείς είναι οι αυξημένες επαναλήψεις στη δεύτερη περίπτωση, ενώ όσον αφορά τα αποτελέσματα του RMSE στα σύνολα επικύρωσης και δοκιμής παρουσιάζονται βελτιωμένα σε σχέση με την πρώτη περίπτωση στην πλειονότητα των δοκιμών.

Θεωρητικά, η μέθοδος VP σε κάθε επανάληψη του αλγορίθμου Levenberg-Marquardt έχει εξασφαλισμένες τις βέλτιστες τιμές για τα βάρη, μιας και αυτά προέρχονται από τη λύση του γραμμικού προβλήματος ελαχίστων τετραγώνων για τις δεδομένες θέσεις των κέντρων. Η ποιότητα του τοπικού ελαχίστου για τα κέντρα είναι αυτή που θα καθορίσει το κατά πόσο θα αξιοποιηθεί το πλεονέκτημα των βέλτιστων τιμών των βαρών και, τελικά, την απόδοση του δικτύου. Αν οι θέσεις των κέντρων, στην οποία θα καταλήξουν οι επαναλήψεις του αλγορίθμου Levenberg-Marquardt, δεν είναι καλές, τότε θα υποβαθμίζεται, κατά κάποιο τρόπο, το συνολικό αποτέλεσμα.

Από την άλλη, στην περίπτωση της μεθόδου FF, δεν υπάρχει εγγυημένο βέλτιστο για τα βάρη, καθώς σε κάθε επανάληψη του αλγορίθμου Levenberg-Marquardt μεταβάλλονται ταυτοχρόνως οι θέσεις των κέντρων και οι τιμές των βαρών. Στην περίπτωση αυτή, είναι πιθανόν να βρεθεί ένα συνολικά καλύτερο τοπικό ελάχιστο για τα κέντρα και τα βάρη, ώστε το τελικό αποτέλεσμα να ξεπεράσει αυτό της μεθόδου VP ως προς την απόδοση του δικτύου.

Γενικά κανείς θα περίμενε στις περισσότερες των περιπτώσεων η μέθοδος VP να παράγει καλύτερο αποτέλεσμα σε σχέση με την FF, δεδομένου ότι σε κάθε επανάληψη είναι εξασφαλισμένη η εύρεση του ολικού ελαχίστου ως προς τα βάρη, κάτι που πάντως δεν επαληθεύεται στα συγκεκριμένα 4 σύνολα δεδομένων που δοκιμάστηκαν. Για την εξαγωγή πιο ασφαλών συμπερασμάτων απαιτούνται δοκιμές σε περισσότερα σύνολα δεδομένων, ενώ θα πρέπει επίσης να διερευνηθούν πιο εντατικά σημαντικές παράμετροι που επηρεάζουν τα αποτελέσματα, όπως η ρύθμιση των λειτουργικών παραμέτρων των αλγορίθμων (π.χ του λ), καθώς και η επιλογή των κριτηρίου τερματισμού.

Όσον αφορά τους υπολογιστικούς χρόνους, φαίνεται να σχετίζονται ίσως και με την επιλογή της τιμής της παραμέτρου λ , καθώς αυτή καθορίζει το μέγεθος των βημάτων και, τελικά, τον αριθμό των επαναλήψεων του αλγορίθμου Levenberg-Marquardt.

Σύγκριση με άλλες μεθοδολογίες εκπαίδευσης της βιβλιογραφίας

Στους παρακάτω πίνακες γίνεται σύγκριση των μεθοδολογιών της παρούσας εργασίας με τις μεθοδολογίες που περιγράφονται στο [Alexandridis13], τόσο για την απόδοση του δικτύου (πίνακας 6.16) όσο και για τον υπολογιστικό χρόνο (πίνακας 6.17). Επιλέγονται τα αποτελέσματα εκείνα τα οποία δίνουν τις βέλτιστες τιμές στο σύνολο δοκιμής για τις δύο κύριες μεθοδολογίες, Variable Projection και Full Functional. Ωστόσο, στις μεθοδολογίες της παρούσας εργασίας ο υπολογιστικός χρόνος αφορά μόνο την εκπαίδευση του δικτύου, των υπολογισμό δηλαδή των παραμέτρων, χωρίς να περιλαμβάνονται οι χρόνοι για τη διευρέυνση του βέλτιστου αριθμού κέντρων, οπότε αναμένονται και οι σχετικές διαφορές. Στο σημείο αυτό, πρέπει να τονιστεί ότι οι μετρήσεις που αφορούν τους υπολογιστικούς χρόνους έχουν πραγματοποιηθεί σε διαφορετικούς υπολογιστές.

Πίνακας 6.16: Αποτελέσματα των προτεινόμενων μεθοδολογιών ως προς το βέλτιστο σφάλμα στο σύνολο δοκιμής συγκρινόμενα με άλλες μεθοδολογίες ως προς τα κόστη στα σύνολα εκπαίδευσης, επικύρωσης και εφαρμογής.

Συν. δεδομένων	Αλγόριθμος	RMSE Επικύρωσης	RMSE Δοκιμής	Ασαφής διαχωρισμός	Αριθμός μονάδων
Auto MPG	PSO-NSFM	2.11 ($2.14 \pm 3.19 \times 10^{-2}$)	1.95 ($2.09 \pm 8.96 \times 10^{-2}$)	[27 4 4 32 4 8]	36 (39±24)
	SFM	2.22	2.12	10	41
	MLP	2.51	2.34	-	[23 19]
	$\tau = 10^{-1}$, FF fuzzy means, σταθ. σ	2.3400	2.0223	10	43
	τ , VP fuzzy means	2.3417	2.0659	10	43
Boston Housing	PSO-NSFM	2.35 ($2.38 \pm 3.13 \times 10^{-2}$)	2.72 ($3.12 \pm 2.26 \times 10^{-1}$)	[48 50 39 4 50 24 4 6 28 45 18 5 11]	138 (141±17)
	SFM	2.77	2.93	16	146
	MLP	2.47	3.38	-	[18 19]
	$\tau = 10^{-3}$, FF fuzzy means, σταθ. σ	2.7178	2.7621	11	93
	$\tau = 10^0$, VP fuzzy means, σταθ. σ	2.7838	2.8368	11	93
CPU Small	PSO-NSFM	2.80 ($2.83 \pm 1.34 \times 10^{-2}$)	2.81 ($2.83 \pm 2.39 \times 10^{-2}$)	[4 7 18 15 16 32 24 29 26 16 35 20]	467 (463±261)
	SFM	2.87	2.84	23	480
	MLP	3.01	3.02	-	[24 30]
	$\tau = 10^{-1}$, FF fuzzy means, σταθ. σ	3.1025	2.9145	17	239
	$\tau = 10^0$, VP fuzzy means, μετ. σ	3.1391	2.9586	17	239
Friedman	PSO-NSFM	9.08×10^{-1} ($9.33 \times 10^{-1} \pm 1.61 \times 10^{-2}$)	1.06 ($1.07 \pm 3.03 \times 10^{-2}$)	[8 4 4 32 4]	114 (116±66)
	SFM	9.64×10^{-1}	1.09	8	179
	MLP	1.04	1.08	-	[15 15]
	$\tau = 10^0$, FF fuzzy means, μετ. σ	1.0077	1.0512	7	119
	$\tau = 10^0$, VP fuzzy means, μετ. σ	1.0154	1.0600	7	119

Οι αριθμοί σε έντονη γραμματοσειρά υποδηλώνουν το βέλτιστο αποτέλεσμα ανά σύνολο στο σύνολο δοκιμής.

Πίνακας 6.17: Αποτελέσματα των προτεινόμενων μεθοδολογιών ως προς το βέλτιστο σφάλμα στο σύνολο δοκιμής συγκρινόμενα με άλλες μεθοδολογίες ως προς τον υπολογιστικό χρόνο.

Συν. δεδομένων / Αλγόριθμος	Υπολογιστικός χρόνος (s)				
	PSO-NSFM (complete)	SFM	MLP	FF fuzzy means	VP fuzzy means
Auto MPG	92 (388±113)	8	1174	0.1	0.1
Boston Housing	2588 (2824±877)	13	2402	0.6	0.5
CPU small	4505 (2377±1053)	790	2814	38	98
Friedman	133 (436±123)	47	1300	1.3	1.6

Οι μεθοδολογίες της παρούσας εργασίας φαίνεται να ξεπερνούν τα δίκτυα MLP όσον αφορά τις τιμές κόστους στο σύνολο δοκιμής σε όλες τις περιπτώσεις. Επίσης, σε 3 από τα 4 σύνολα οι προτεινόμενες μέθοδοι ξεπερνούν τον αλγόριθμο SFM. Σχεδόν σε όλες τις περιπτώσεις φαίνεται να ξεπερνιέται η μέση τιμή του PSO-NSFM στο σύνολο δοκιμής, ενώ σε ορισμένες περιπτώσεις και η βέλτιστη τιμή αυτού του συνόλου. Θα πρέπει να τονισθεί ότι σε όλες τις μεθόδους που έχουν χρησιμοποιηθεί για σύγκριση, έχει

εφαρμοστεί ξεχωριστή διαδικασία βελτιστοποίησης του μεγέθους του δικτύου, κάτι που δεν έχει γίνει για τις προτεινόμενες μεθοδολογίες και θα μπορούσε να βελτιώσει σημαντικά τα αποτελέσματα. Όσον αφορά τον υπολογιστικό χρόνο, δε μπορεί να γίνει ακριβής σύγκριση, λόγω του διαφορετικού τρόπου υπολογισμού στις παραπάνω μεθοδολογίες, ωστόσο σημειώνεται ότι οι προτεινόμενες μέθοδοι υπερέχουν κατά πολλές τάξεις μεγέθους, ειδικά σε σχέση με τον αλγόριθμο PSO-NSFM και τα δίκτυα MLP.

Κεφάλαιο 7

Συμπεράσματα

Η εκπαίδευση ενός νευρωνικού δικτύου τύπου RBF είναι μια πολύπλοκη διαδικασία, καθώς για τη σωστή λειτουργία του πρέπει να υπολογιστούν οι κατάλληλες τιμές των παραμέτρων, ώστε να αποκτήσει η επιθυμητή απόδοση του δικτύου σε λογικό χρονικό διάστημα.

Η εκπαίδευση ενός δικτύου RBF μπορεί να θεωρηθεί ως ένα πολυδιάστατο πρόβλημα βελτιστοποίησης με ορισμένες ιδιαιτερότητες [Alexandridis13]. Ο αριθμός των μονάδων RBF, οι θέσεις των κέντρων, τα εύρη αλλά και τα συναπτικά βάρη επηρεάζουν τόσο την απόδοση όσο και τα χαρακτηριστικά του δικτύου. Σημαντική κρίνεται η αρχική επιλογή των κέντρων, ο αριθμός τους και οι θέσεις, [Chen91], οπότε οι αλγόριθμοι εκπαίδευσης πρέπει να εστιάζουν στην κατάλληλη επιλογή τους ή και τη μετέπειτα τροποποίηση των θέσεων προκειμένου να κατασκευαστεί ένα επαρκές μοντέλο.

Αντικείμενο της συγκεκριμένης εργασίας αποτελεί η ανάπτυξη μεθόδων για την εκπαίδευση ενός δικτύου RBF, βάση των οποίων αποτελεί ο αλγόριθμος βελτιστοποίησης Levenberg-Marquardt. Οι θέσεις των κέντρων των μονάδων RBF, τα συναπτικά βάρη και τα εύρη των συναρτήσεων βάσης υπολογίζονται μέσω αυτού του αλγορίθμου. Οι μέθοδοι διαφέρουν στον τρόπο υπολογισμού των παραμέτρων. Στη μια περίπτωση, πραγματοποιείται απαλοιφή των βαρών και κατόπιν ελαχιστοποίηση του μειωμένου συναρτησιακού προβολής μεταβλητής (Variable Projection, VP), το οποίο περιλαμβάνει μόνο τα κέντρα. Σε κάθε επανάληψη του αλγορίθμου, ενημερώνονται οι θέσεις των κέντρων ενώ τα βάρη του δικτύου υπολογίζονται από τη λύση ενός γραμμικού προβλήματος ελαχίστων τετραγώνων για τις τρέχουσες θέσεις των κέντρων. Στην άλλη περίπτωση, δεν πραγματοποιείται διαχωρισμός των συνόλων των παραμέτρων, οπότε το πρόβλημα προς ελαχιστοποίηση αφορά το πλήρες συναρτησιακό (Full Functional, FF). Οι θέσεις των κέντρων και οι τιμές των βαρών ενημερώνονται σε κάθε επανάληψη του αλγορίθμου. Οι αρχικές θέσεις των κέντρων προέρχονται από τους αλγόριθμους k-means και fuzzy means, ενώ οι τιμές των ευρών των συναρτήσεων βάσης από τη μέθοδο των p -κοντινότερων γειτόνων.

Στη μέθοδο VP, κάθε επανάληψη του αλγορίθμου Levenberg-Marquardt εξασφαλίζει τις βέλτιστες τιμές για τα βάρη για τις συγκεκριμένες θέσεις των κέντρων. Η ποιότητα του τοπικού ελαχίστου των θέσεων των κέντρων είναι αυτή που, τελικά, καθορίζει τη συνολική απόδοση του δικτύου. Αντιθέτως, στη μέθοδο FF, αν και δεν υπάρχει εγγύηση για βέλτιστες τιμές βαρών για τις τοποθεσίες των κέντρων που προκύπτουν σε κάθε επανάληψη, είναι πιθανόν το συνολικό ελάχιστο που πετυχαίνουν τα βάρη και τα κέντρα να δίνει καλύτερα αποτελέσματα, ενώ παρατηρείται μια αντίστοιχη αύξηση στον αριθμό των επαναλήψεων που απαιτούνται για σύγκλιση.

Οι δοκιμές των παραπάνω μεθοδολογιών πραγματοποιήθηκαν σε 4 σύνολα δεδομένων της βιβλιογραφίας, πραγματικών και συνθετικών, και για διάφορες τιμές της παραμέτρου λ του αλγορίθμου Levenberg-Marquardt. Αναλόγως της μεθοδολογίας και της τιμής του λ , τα αποτελέσματα θα διαφέρουν, ενώ η σύγκρισή τους με άλλες μεθοδολογίες της βιβλιογραφίας τις καθιστά ιδιαίτερα ανταγωνιστικές, όσον αφορά τις τιμές RMSE στα σύνολα επικύρωσης και δοκιμής, σε συνδυασμό με τους υπολογιστικούς χρόνους.

Αν και τα πρώτα αποτελέσματα που παρουσιάζονται σε αυτή τη διπλωματική εργασία κρίνονται ικανοποιητικά, υπάρχει περιθώριο για βελτίωσή τους. Η μελλοντική έρευνα που θα γίνει προς αυτή την κατεύθυνση περιλαμβάνει:

-
- διερεύνηση της επιλογής των λειτουργικών παραμέτρων του αλγορίθμου (π.χ του λ) καθώς και των κριτηρίων τερματισμού
 - τροποποίηση των μεθόδων για την αυτόματη επιλογή του αριθμού των κέντρων, π.χ με παλινδρόμηση LASSO αντί για απλής γραμμικής παλινδρόμησης
 - εισαγωγή περιορισμών για τις θέσεις των κέντρων
 - δοκιμές σε μεγαλύτερο αριθμό συνόλων δεδομένων
 - επιλογή διαφορετικών συναρτήσεων βάσης

Βιβλιογραφία

- [Abraham06] Abraham, A., Guo, H., Liu, H. (2006). Swarm Intelligence: Foundations, Perspectives and Applications. In Nedhah, N., Mourelle, L. (Eds.) *Swarm Intelligent Systems*, (3-25). Berlin: Springer Heidelberg.
- [Agatonovic00] Agatonovic-Kustrin, S., Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, 22(5):717-727.
- [Alexandridis03] Alexandridis, A., Sarimveis, H., Bafas, G. (2003). A new algorithm for on-line structure and parameter adaptation of RBF networks. *Neural Networks*, 16(7):1003-1017.
- [Alexandridis13] Alexandridis, A., Chondrodima E., Sarimveis, H. (2013). 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*, 24(2):219-230.
- [Alexandridis13(2)] Alexandridis, A. (2013). Evolving RBF neural networks for adaptive soft-sensor design. *International Journal of Neural Systems*, 23(6):1350029.
- [Alexandridis13(3)] Alexandridis, A., Stogiannos M., Kyriou A., Sarimveis H. (2013). An offset-free neural controller based on approximating the inverse process dynamics. *Journal of Process Control*, 23(7):968{979.
- [Alexandridis14] Alexandridis A., Chondrodima, E. (2014). A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing. *Journal of Biomedical Informatics*, 49: 61-72.
- [Alexandridis14(2)] Alexandridis A, Chondrodima, E., Efthimiou, E., Papadakis G., Vallianatos F., Triantis D. (2014). Large earthquake occurrence estimation based on radial basis function neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5443-5453.
- [Alexandridis16] Alexandridis, A., Chondrodima E., Sarimveis H. (2016). Cooperative learning for radial basis function networks using particle swarm optimization. *Applied Soft Computing*, 49:485-497.
- [Alexandridis17] Alexandridis, A., Chondrodima E., Giannopoulos N., Sarimveis H. (2017). A Fast and Efficient Method for Training Categorical Radial Basis Function Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 28(11):2831 - 2836.
- [Babazadeh11] Babazadeh A., Poorzahedy H., Nikoosokhan S. (2011). Application of particle swarm optimization to transportation network design problem. *Journal of King Saud University - Science*, 23(3):293-300.

-
- [Back97] Back, T., Hammel U., Schwefel H.P. (1997). Evolutionary computation: Comments on the history and current state. *IEEE Trans. Evolutionary Computation*, 1(1):3-17.
- [Barr81] Barr, A., & Feigenbaum, E. (1981). *The Handbook of Artificial Intelligence*. Kaufmann, Inc.
- [Beni89] Beni, G., Wang, J. (1989). Swarm Intelligence in Cellular Robotic Systems. *Proceedings NATO Advanced Workshop on Robots and Biological Systems*. Tuscany, Italy, 1989.
- [Bezdek81] Bezdek, J.C . (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. *Plenum Press, New York*.
- [Bezdek94] Bezdek, J. C. (1994). What is computational intelligence? In Zurada, J. M. , Marks II R. J., & Robinson Ch. J. (Eds.) *Computational Intelligence: Imitating Life*, (1-12), New York: IEEE Press.
- [Bindel12] Bindel, D. (2012). Intro to Scientific Computing. Department of Computer Science, Cornell University. Lecture notes, <http://www.cs.cornell.edu/~bindel/class/cs3220-s12/notes/lec10.pdf>
- [Bishop95] Bishop C.M. (1995). *Neural Networks for Pattern Recognition*. Oxford: The Clarendon Press
- [Βλαχάβας06] Βλαχάβας, Ι., Κεφάλας, Π., Βασιλειάδης, Ν., Κόκκορας Φ, Σακελλαρίου, Η. (2006). Τεχνητή Νοημοσύνη (Γεκδ.). Αθήνα: Γκιούρδας Β.
- [Blum08] Blum, C., Li, X. (2008). Swarm Intelligence in Optimization. In Blum, C., Merkle, D. (Eds.) *Swarm Intelligence: Introduction and Applications.*, (43-85). Berlin: Springer-Verlag Heidelberg.
- [Bora14] Bora, D.A, Gupta, A.K. (2014). A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm. *International Journal of Computer Trends and Technology (IJCTT)*, 10(2):108-113.
- [Bors] Bors, A. G. *Introduction of the Radial Basis Function (RBF) Network*.
- [Brian96] Ripley, B.D. (1996). *Patter Recognition and Neural Networks*. Oxford University Press.
- [Buchanan05] Buchanan, B. G. (2005). A (Very) Brief History of Artificial Intelligence. *AI Magazine*, 26(4):53-60.
- [Broomhead88] Broomhead, D.S, Lowe, D. (1988). Multivariate Functional Interpolation and Adaptive Networks. *Complex Systems*, 2:321-355.
- [Brotherson18] Brotherson, Sean. (2018). Understanding Brain Development in Young Children.
- [Cerconel94] Cercone, N. (1994). Ten years of computational intelligence. *Computational Intelligence*, 10(4):i-vi.
- [Chen91] Chen, S., Cowan, C. F. N., Grant. P. M.(1991). Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks. *IEEE Transactions on Neural Networks*. 15:209{18.
- [Chrupala10] Chrupala, G., Stroppa, N. (2010). Linear models for classification. Saarland University. META Workshop. <http://grzegorz.chrupala.me/papers/ml4nlp/linear-classifiers.pdf>
-

- [Darken90] C. Darken and J. Moody. (1990). Fast adaptive K-means clustering: Some empirical results. *in Proc. IEEE Int. Joint Conf. Neural Netw. San Diego, CA* pp. 233{238.
- [Debes95] Debes, K., Koenig A., Gross, H-M. (1995). Transfer Functions in Artificial Neural Networks. Department of Neuroinformatics and Cognitive Robotics, Technical University Ilmenau.
- [Dorigo92] Dorigo, M. (1992). *Optimization, learning and natural algorithms*. PhD thesis, Dip. Elettronica e Informazione, Politecnico di Milano, Italy.
- [Dougherty13] Dougherty, G. (2013). Pattern Recognition and Classification: An Introduction. In Nedhah, N., Mourelle, L. (Eds.) *Swarm Intelligent Systems*, (3-25). Springer-Verlag New York.
- [Drach2005] Garg, D.A. (2005). Do we have brain to spare? *Neurology, 2004-2005*, 64
- [Duch03] Duch, W. (2003). What is Computational Intelligence and what could it become?. *Computational Intelligence Methods and Applications*, Nanyang Technol. Univ.
- [Dunn73] Dunn, J.C . (1973). Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*. 3(2):32-57.
- [Dufourq17] Dufourq, E., Bassett B.A. (2017). Automated Problem Identification: Regression vs Classification via Evolutionary Deep Networks. *CoRR*.
- [Engelbrecht07] Engelbrecht, A. P. (2007). *Computational Intelligence: An introduction (2 ed.)*. John Wiley & Sons, Ltd.
- [Faber94] Faber, V. (1994). Clustering and the Continuous k-means Algorithm. *Los Alamos Science*, 22:138-144.
- [Fonseca95] Fonseca C. M., Fleming P. J. (1994). An Overview of Evolutionary Algorithms in Multiobjective Optimization. *Evolutionary Computation*, 3(1):1-16.
- [Fox07] Fox, B., Xiang, W., Lee, H.P. (2007). Industrial applications of the ant colony optimization. *The International Journal of Advanced Manufacturing Technology*, 31:805-814.
- [Garg09] Garg, A., Gill, P., Rathi, P., Garg, K.K. (2009). An Insight into Swarm Intelligence. *International Journal of Recent Trends in Engineering*, 2(8):42-44.
- [Gilbert08] Gilbert, S. (2008). *Intoduction to Linear Algebra*. Wellesley-Cambridge Press.
- [Golub73] Golub, G.H, Pereyra, V. (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *IAM Journal on Numerical Analysis*, 10(2):413{432.
- [Golub96] Golub, G.H, Van Loan, F. (1996). *Matrix Computations, 3rd Edition*. London: The Johns Hopkins Press Ltd.
- [Golub03] Golub, G.H, Pereyra, V. (2003). Separable nonlinear least squares: the variable projection method and its applications. *Inverse Problems*, 19(2): R1-R26.
- [Guttman73] Guttman I., Pereyra V., Scolnik H.D. (1973). Least squares estimation for a class of nonlinear models. *Technometrics*. 15:209{18.
- [Hansen12] Hansen, C., Pereyra, V., Scherer, G. Least Squares Data Fitting with Applications. *Baltimore: Johns Hopkins University Press*, 2012. Project MUSE.

- [Jakel04] Jakel, J., Bretthauer, G. (2004). Fuzzy system applications. In Unbehauen, H. (Eds.) *Control Systems Robotics and Automation*, (1-24). EOLSS, Oxford.
- [Καβαλλιεράτου] Καβαλλιεράτου Ε., Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων, Πανεπιστήμιο Αιγαίου, Διαλέξεις μαθημάτων http://www.icسد.aegean.gr/lecturers/kavallieratou/NN&EP_files/ci_2.pdf
- [Kaelbing96] Kaelbling L. P., Littman M. L., Moore A.W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4:237-285
- [Kaufman75] Kaufman, L. (1973). A variable projection method for solving separable nonlinear least squares problems. *BIT Numerical Mathematics*, 15(1):49{57.
- [Kang05] Kang, S., Isik, C. (2005). Partially Connected Feedforward Neural Networks Structured by Input Types. *IEEE Transactions on Neural Networks*, 16(1):175-184.
- [Kanungo02] Kanungo, T., Mount D.M., Netanyahu, N.S, Piatko, C.D., Silverman R., Wu, A.Y. (2002). An Efficient k-Means Clustering Algorithm: Analysis and Implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):881-892.
- [Kassabalidis01] Kassabalidis, M.A. El-Sharkawi, R.J. Marks II, P.Arabshahi, A.A. Gray. (2001). Swarm Intelligence for Routing in Communication Networks. *IEEE Global Telecommunications Conference*. 6:3613-3617.
- [Kennedy95] Kennedy, J., Eberhart, R. (1995). Particle swarm optimization. *Proceedings., IEEE International Conference on Neural Networks*, 1942-1948. Perth, WA, vol. 4.
- [Kotsiantis04] Kotsiantis, S. & Pintelas, P. (2004). Recent advances in clustering: A brief survey. *WSEAS Transactions on Information Science and Applications*, 1(1):73-81.
- [Kotti11] Kotti M. et al. (2011). Comparison between PSO and ACO techniques for analog circuit performance optimization. *ICM 2011 Proceeding*, Hammamet, 1-6.
- [Lemarechal12] Lemarechal, C. (2012). Cauchy and the gradient method. *Doc Math Extra Volume*, 251{254.
- [Leonard91] Leonard, J.A, Kramer, M.A. (1991). Radial basis function networks for classifying process faults. *IEEE Control Syst.*, 11(3):31-38.
- [Levenberg44] Levenberg, K. (1944). A Method for the Solution of Certain Problems in Least Squares. *Quart. Appl. Math.*, 2(3):164{168.
- [Lison12] Lison P., An introduction to machine learning. Language Technology Group (LTG), University of Oslo, Department of Informatics. <http://folk.uio.no/plison/pdfs/talks/machinelearning.pdf>
- [Lloyd82] Lloyd, S.P. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information theory*, 28(2):129-137.
- [Madsen04] Madsen, K., Nielsen, H.B, Tingleff O., Methods for Non-Linear Least Squares Problems 2nd end. *Informatics and Mathematical Modelling*, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby.

-
- [Marquardt63] Marquardt, D. (1963). An Algorithm for Least Squares Estimation on Nonlinear Parameters. *Siam J. Appl. Math.*, 11:431{441.
- [Michelli86] Michelli C. A. (1986). Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constr. Approx.*, 2:11{22
- [Moody89] Moody, J, Darken, C.J. (1989). Fast Learning in Networks of Locally-Tuned Processing Units. *Neural Computation*, 1(2):281-294.
- [Moody89(2)] Moody, J., Darken., C. (1989). Learning with Localized Receptive Fields. In D. Touretzky, G. Hinton, and T. Sejnowski. (Eds.), *Proceedings of the Connectionist Models Summer School*pp. 133{143, San Mateo, C.A. Morgan Kaufmann.
- [Morissette13] Morissette, L., Chartier, S. (2013). The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, 9(1):15-24.
- [Nocedal99] Nocedal, J., Wright, S.J. (1999). *Numerical Optimization*, Springer.
- [Nguyen] Nguyen, T., & Norvig, P. (2010). Total Number of Synapses in the Adult Human Neocortex. *Undergraduate Journal of Mathematical Modeling: One + Two*, 3(1), Article 14.
- [Poggio89] Poggio, T., and F. Girosi. (1989). A Theory of Networks for Approximation and Learning. *Technical Report A.I. Memo 1140, MIT, Cambridge, M.A.*
- [Poggio90] Poggio, T., Girosi, F. (1990). Networks for approximation and learning. *Proc. IEEE*, 78(9):1481{1497
- [Powell70] Powell, M.J.D. (1970). A Hybrid Method for Non-Linear Equations. In P. Rabinowitz (eds.) *Numerical Methods for Non-Linear Algebraic Equations*, Gordon and Breach
- [Russell03] Russell, S.J., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach (2 ed.)*. Pearson Education.
- [Sarimveis02] Sarimveis, H., Alexandridis, A., Tsekouras, G., Bafas, G. (2002). 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*, 41(4):751{759.
- [Sarimveis04] Sarimveis H., Alexandridis A., Mazarakis S., and Bafas G., (2004). A new algorithm for developing dynamic radial basis function neural network models based on genetic algorithms. *Comput. Chem. Eng.*, 28:209{217.
- [Saggu13] Saggu, A., Yadav, P., Roopak, M. (2013). Applications of Swarm Intelligence. *International Journal of Computer Science and Mobile Computing.*, 2(5):353-359.
- [Sarpeshkar92] Sarpeshkar, R., Watts, L., Mead, C. (1992). *Refractory neuron circuits Calif. Inst. Technol.*
- [Sebastiani02] Sebastiani, F. (2005). Machine Learning in Automated Text Categorization. *ACM Comput. Surv.*, 34(1):1-47.
- [Simon95] Simon, H. A. (1995). Artificial intelligence: an empirical science. *Artificial Intelligence*, 77(1):95-127.
-

-
- [Srivastava15] Srivastava, N., Mansimov, E., Salakhutdinov, R. (2015). Unsupervised Learning of Video Representations using LSTMs.
- [Stogiannos18] Stogiannos, M., Alexandridis A., Sarimveis H. (2018) Model predictive control for systems with fast dynamics using inverse neural models. *ISA Transactions*, 72:161-177.
- [Tao93] Tao, K. M., A closer look at the radial basis function (RBF) networks " *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, 401-405. Pacific Grove, CA, vol.1.
- [Wettschereck92] D. Wettschereck and T. Dietterich. Improving the Performance of Radial Basis Function Networks by Learning Center Locations. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, (Eds.), *Advances in Neural Information Processing Systems*, volume 4, pages 1133-1140, San Mateo, C.A., 1992. Morgan Kaufmann
- [Xie11] Xie, T., Yu, H., Wilamowski, B. (2011). Comparison between traditional neural networks and radial basis function networks. *IEEE International Symposium on Industrial Electronics*, 1194-1199. Gdansk.
- [Yu11] Yu, H. Xie, T., Paszczyński, S., Wilamowski, B.M. (2011). Advantages of Radial Basis Function Networks for Dynamic System Design. *IEEE International Symposium on Industrial Electronics*, , 58(12):5438-5450.
- [Yuan] Yuan, Y. *A Review of Trust Region Algorithms for Optimization*
- [You08] You, P.S. (2008). An efficient computational approach for railway booking problems. *European Journal of Operational Research*, 185(2):811-824.
- [Zadeh65] Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8(3):338-353.