



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών
«Πληροφορική»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Κατηγοριοποίηση πλοίων με την χρήση νευρωνικών δικτύων Ship classification with neural networks
Όνοματεπώνυμο Φοιτητή	Ευθυμίου Ευάγγελος
Πατρώνυμο	Χρυσόστομος
Αριθμός Μητρώου	ΜΠΠΛ/ 14014
Επιβλέπων	Ιωάννης Θεοδωρίδης, Καθηγητής

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

(υπογραφή)

(υπογραφή)

Ιωάννη Θεοδωρίδης
Καθηγητής

Νικόλας Πελέκης
Επίκουρος Καθηγητής

Άγγελος Πικράκης
Επίκουρος Καθηγητής

Ευχαριστίες

Η παρούσα εργασία, αποτελεί διπλωματική εργασία που εκπονήθηκε για τις ανάγκες του μεταπτυχιακού προγράμματος «Μεταπτυχιακό στην Πληροφορική» του τμήματος Πληροφορικής του Πανεπιστημίου Πειραιώς.

Πριν την παρουσίαση του θεωρητικού και του πειραματικού τμήματος της εργασίας, θα ήθελα να ευχαριστήσω ορισμένους ανθρώπους που γνώρισα και που χωρίς την συνεργασία τους, δεν θα ήταν εφικτή η ολοκλήρωση της διπλωματικής εργασίας.

Πρώτα από όλους, θα ήθελα να ευχαριστήσω τον καθηγητή κ. Ιωάννη Θεοδωρίδη για τη βοήθεια και την εμπιστοσύνη που έδειξε στο σχέδιο της διπλωματικής εργασίας και την άμεση βοήθεια που παρείχε κάθε στιγμή. Επίσης, ευχαριστώ για τη βοήθεια τους, το διδάκτορα Χάρη Γεωργίου, που με την καθοριστική συμβολή του βοήθησε στην ολοκλήρωση της εργασίας καθώς και τον υποψήφιο διδάκτορα Ιωάννη Κοντούλη για τη βοήθεια του στην κατανόηση των δεδομένων που χρησιμοποιήθηκαν στο πειραματικό στάδιο.

Τέλος, θα ήθελα να ευχαριστήσω όλους τους καθηγητές του Μεταπτυχιακού προγράμματος, που βοήθησαν στην περαιτέρω ανάπτυξη των γνώσεων μου.

Περίληψη

Σκοπός της εργασίας είναι, η δημιουργία και η χρήση ενός νευρωνικού δικτύου, το οποίο θα χρησιμοποιείται για την κατηγοριοποίηση των πλοίων με βάση τον τρόπο κίνησης τους. Παρακάτω, παρουσιάζονται κάποιες από τις αρχιτεκτονικές των μοντέλων πρόβλεψης, που έχουν επικρατήσει τα τελευταία χρόνια. Επίσης, γίνεται εκτενής αναφορά στον Fuzzy means αλγόριθμο, που χρησιμοποιείται για τη δημιουργία του μοντέλου πρόβλεψης, καθώς επίσης και στον αλγόριθμο one vs all, που θα χρησιμοποιηθεί για την κατηγοριοποίηση των αποτελεσμάτων. Ακόμη, αναφέρεται ο τρόπος με τον οποίο θα δημιουργηθούν τα δεδομένα, όπως επίσης και ο τρόπος επεξεργασίας τους, προκειμένου να εμπεριέχουν όσο το δυνατόν καλύτερη πληροφορία, για να δημιουργηθεί η καλύτερη δυνατή συσχέτιση μεταξύ των δεδομένων εισόδου και εξόδου. Κατά τη διάρκεια της εργασίας, θα ασχοληθούμε με τους συντελεστές αποδοτικότητας του μοντέλου και με τις κινήσεις που θα κάνουμε, προκειμένου μέσα από την βελτίωσή τους, να πετύχουμε τη σωστή λειτουργία του συστήματος. Τέλος, σε περίπτωση που τα αποτελέσματα του συστήματος παρουσιάσουν ικανοποιητική συμπεριφορά, θα προσπαθήσουμε να ελέγξουμε την ευρωστία του, μεταβάλλοντας κάποιες από τις παραμέτρους που έχουν δοθεί, για να προκύψει η βέλτιστη συμπεριφορά. Σε περίπτωση που και μετά την παραπάνω τροποποίηση, το σύστημα αποκρίνεται επιτυχώς, τότε μπορούμε να πούμε πως το δίκτυο έχει μάθει επιτυχώς τη συσχέτιση μεταξύ της εισόδου και της εξόδου.

Abstract

The purpose of this paper is to create and use a neural network, which will be used to classify different types of ships, based on the characterized movement. Nowadays, many architectures of prediction models are used and we could have the opportunity to present them. Fuzzy means, which is a powerful algorithm and based on radial basis action of neural network architecture, will be the core component for the neural network. In order to achieve the correct results, we will use one vs all algorithm, which is used for classification problems, when we have more than two classes as outputs. All data have received from the University of Piraeus antenna, which can encapsulate data around the area of Argosaronikos sea. Firstly, we will train the neural network with all the available datasets, without any process of data. As inputs, we will give the transportation, the speed and the acceleration of every class. After that, we will try to improve the network by processing the input data, in order to get rid of the noise and then we will retrain the network. If the result is successful, we will check the robustness of the network via changing some of the parameters, in order to observe, if the response of the network, will remain in good rate.

Περιεχόμενα

Κεφάλαιο 1	3
Εισαγωγή.....	3
Κεφάλαιο 2	5
Νευρωνικά Δίκτυα.....	5
2.1 Τι είναι τα νευρωνικά δίκτυα	5
2.2 Πλεονεκτήματα των νευρωνικών δικτύων	5
2.3 Δομή βιολογικού νευρώνα.....	6
2.4 Δομή τεχνητού νευρώνα.....	6
2.5 Συναρτήσεις ενεργοποίησης	7
Γραμμική συνάρτηση (Linear function).....	7
Βηματική συνάρτηση (Step function).....	8
Κεκλιμένη συνάρτηση (Ramp function)	8
Σιγμοειδής συνάρτηση (Sigmoid function).....	8
Hyperbolic tangent.....	8
Gaussian function	9
2.6 Διαδικασίες μάθησης	9
Εκπαίδευση με επίβλεψη	10
Εκπαίδευση χωρίς επίβλεψη	10
Ενισχυτική εκπαίδευση	10
2.7 Αρχιτεκτονικές δικτύων.....	11
Δίκτυο πρόσθιας τροφοδότησης (Feedforward neural networks-FFNNs)	11
Δίκτυα πρόσθιας τροφοδότησης πολλών στοιβάδων (MLP)	11
Αναδρομικά δίκτυα	12
Ελάχιστα τετράγωνα.....	13
Συντελεστές αξιολόγησης.....	14
Ιστορική αναδρομή – perceptron του Rosenblatt.....	15
Σύγχρονες εφαρμογές νευρωνικών δικτύων	16
Κεφάλαιο 3	18
Δίκτυα συναρτήσεων ακτινικής βάσης (RBF).....	18
Δίκτυα RBF	18
Αρχιτεκτονικές νευρωνικών δικτύων RBF	18
Συναρτήσεις ενεργοποίησης RBF	19
Γραμμική συνάρτηση.....	19
Κυβική συνάρτηση.....	19
Thin-Plate-Spline	19
Multiquadratic	20
Inverse Multiquadratic	20

Gaussian	20
Logistic	20
Εκπαίδευση νευρωνικών δικτύων RBF	20
Αλγόριθμος διόρθωσης κέντρων	21
Ασαφής λογική (Fuzzy logic)	21
Κεφάλαιο 4	23
Υλοποίηση Μοντέλου	23
Αλγόριθμος των ασαφών μέσων (Fuzzy means algorithm)	23
Αλγόριθμος κατηγοριοποίησης one vs all	25
Αλγόριθμος k-folds cross-validation	26
Επεξεργασία Δεδομένων	27
Κεφάλαιο 5	29
Πειραματική δραστηριότητα.....	29
Δεδομένα χωρίς επεξεργασία.....	29
Επεξεργασμένα δεδομένα	30
Παραμετροποίηση δεδομένων εισόδου	33
Έλεγχος ευρωστίας (Robustness).....	34
Κεφάλαιο 6	38
Συμπεράσματα	38
Βιβλιογραφία.....	39

Κεφάλαιο 1

Εισαγωγή

Το βασικό χαρακτηριστικό της κοινωνίας μας στον 21ο αιώνα είναι η ταχύτητα. Η ταχύτητα με την οποία κινούμαστε και εργαζόμαστε αυξάνεται σχεδόν καθημερινά με θεαματικούς ρυθμούς. Μάλιστα η αύξηση είναι τόσο ραγδαία, που σε πολλές περιπτώσεις θέλουμε να γνωρίζουμε τα αποτελέσματα των ενεργειών μας, πριν ακόμα αυτά συμβούν. Το πιο ενδεικτικό παράδειγμα αφορά το χρηματοοικονομικό κλάδο. Οι χρηματιστές θέλουν να γνωρίζουν από πριν την φορά που θα έχει η μετοχή μιας εταιρείας ή ενός προϊόντος, ούτως ώστε να διαμορφώσουν ανάλογα τις επενδύσεις τους. Με λίγα λόγια, θέλουν να περιορίσουν το ρίσκο που περιέχει η κάθε τους κίνηση. Πέρα από τον οικονομικό τομέα, ένα ακόμα παράδειγμα που συναντάμε στην καθημερινή μας ζωή αφορά την πρόβλεψη του καιρού. Οι αγρότες θέλουν να γνωρίζουν με όσο το δυνατόν μεγαλύτερη ακρίβεια τον καιρό, προκειμένου να προσαρμόσουν τις εργασίες τους όσο καλύτερα γίνεται.

Για την πρόβλεψη τέτοιων καταστάσεων έχει αναπτυχθεί ιδιαίτερα τα τελευταία χρόνια μια κατεύθυνση, η οποία ονομάζεται μηχανική μάθηση, υποκατηγορία της οποίας είναι τα νευρωνικά δίκτυα. Ουσιαστικά, πρόκειται για μαθηματικά μοντέλα τα οποία χρησιμοποιούνται για την πρόβλεψη *μη γραμμικών* γεγονότων. Πολλές αρχιτεκτονικές νευρωνικών δικτύων έχουν δημιουργηθεί και χρησιμοποιούνται ολοένα και περισσότερο με ιδιαίτερα ενθαρρυντικά αποτελέσματα.

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) είναι ισχυρά υπολογιστικά εργαλεία [1], τα οποία προσομοιώνουν τη λειτουργία του ανθρώπινου εγκεφάλου, στο πώς αυτός αλληλοεπιδρά με το περιβάλλον, όπως και τη διαδικασία της εκπαίδευσης του. Τα νευρωνικά δίκτυα έχουν τη δυνατότητα να προσδιορίζουν και να μαθαίνουν περίπλοκες και μη γραμμικές σχέσεις μεταξύ των δεδομένων εισόδου και εξόδου, χωρίς να γνωρίζουν τις σχέσεις που διέπουν το εκάστοτε σύστημα. Λόγω των χαρακτηριστικών αυτών, τα νευρωνικά δίκτυα χρησιμοποιούνται σε πληθώρα γεωλογικών εφαρμογών.

Μια σημαντική εφαρμογή που έχουν χρησιμοποιηθεί νευρωνικά δίκτυα είναι στην πρόβλεψη των βροχοπτώσεων [2]. Επίσης, έχουν χρησιμοποιηθεί για την οπτική προσέγγιση του βάθους της γήινης ατμόσφαιρας, σε συνδυασμό με την πρόβλεψη της αβεβαιότητας [3]. Τρία διαφορετικά μοντέλα νευρωνικών δικτύων έχουν χρησιμοποιηθεί για την πρόβλεψη των ειδών δέντρων που βρίσκονται σε συγκεκριμένα χαρακτηριστικά των δασών, αποδίδοντας ιδιαίτερα υψηλή αξιοπιστία πρόβλεψης [4]. Ένα νευρωνικό δίκτυο με το όνομα Adaptive Neuro-Fuzzy Inference System ή ANFIS, εφαρμόζεται για τη βελτίωση της ακρίβειας των ατμοσφαιρικών χαρακτηριστικών που σχετίζονται με την θερμοκρασία και την υγρασία, που προέρχονται από παρατηρήσεις υπέρυθρων ακτινών στο βυθόμετρο [5]. Μια μέθοδος που αφορά την πρόβλεψη της υγρασίας του εδάφους, υλοποιείται από νευρωνικά δίκτυα, με την βοήθεια φωτογραφιών που λαμβάνονται από έναν γεωστατικό δορυφόρο και παρατηρείται πως σε σύγκριση με άλλου είδους τεχνικές, αυτή των νευρωνικών δικτύων είναι η πλέον κατάλληλη για τη δημιουργία χαρτών υγρασίας [6]. Ακόμη, για την πρόβλεψη της μεταβολής του μεγέθους της θάλασσας με το πέρασμα του χρόνου, έχει δημιουργηθεί ένα μοντέλο που βασίζεται σε ένα συνδυασμό RBF νευρωνικών δικτύων [7]. Επίσης τα νευρωνικά δίκτυα έχουν χρησιμοποιηθεί, για την πρόβλεψη και την εκτίμηση φυσικών καταστροφών, που αφορούν τους κυκλώνες και τα σεισμικά γεγονότα.

Τα νευρωνικά δίκτυα τα τελευταία χρόνια, χρησιμοποιούνται σαν μια διαφορετική προσέγγιση για την πρόβλεψη των σεισμών. Έχουν δημοσιευθεί εργασίες για την πρόβλεψη του μεγέθους σεισμών με χρήση νευρωνικών δικτύων πρόσθιας τροφοδότησης, χρησιμοποιώντας σαν δεδομένα εισόδου τη συγκέντρωση του ραδόνιου, μαζί με την περιοχή που συνέβη το εκάστοτε σεισμικό γεγονός, καθώς και το εστιακό βάθος [9]. Επίσης έχει χρησιμοποιηθεί ένα ακόμα νευρωνικό δίκτυο πρόσθιας τροφοδότησης, το οποίο λαμβάνει σαν εισόδους τα σεισμικά ηλεκτρικά σήματα (Seismic Electric Signals – SES), με σκοπό την πρόβλεψη του μεγέθους της κλίμακας Ρίχτερ σεισμικών γεγονότων [10]. Για την πρόβλεψη του χρονικού διαστήματος που μεσολαβεί μεταξύ δύο σεισμών έχει δημιουργηθεί επίσης ένα νευρωνικό δίκτυο πρόσθιας τροφοδότησης [11]. Δύο πρόσφατες έρευνες που πραγματοποιήθηκαν είχαν ως σκοπό την χρήση νευρωνικών δικτύων για τον υπολογισμό της πιθανότητας εμφάνισης ενός σεισμού μεγάλου μεγέθους με βάση την κλίμακα Ρίχτερ σε ένα προκαθορισμένο χρονικό διάστημα. Η

μέθοδος αυτή αξιολογήθηκε με την πρόβλεψη σεισμών στην Χιλή [12] και την ευρύτερη περιοχή της Ιβηρικής Χερσονήσου [13]. Τέλος έχει παρουσιαστεί μια μελέτη που αφορά την πρόβλεψη του μεγέθους, του μεγαλύτερου σεισμού που θα παρουσιαστεί εντός ενός συγκεκριμένου χρονικού διαστήματος. Σαν είσοδοι στο νευρωνικό στην τελευταία περίπτωση δόθηκαν οχτώ διαφορετικοί δείκτες σεισμικότητας [14].

Σκοπός της μεταπτυχιακής εργασίας είναι η δημιουργία ενός μοντέλου που θα χρησιμοποιείται για την πρόβλεψη του τύπου των σκαφών τα οποία ανιχνεύονται από την κεραία που έχει εγκατασταθεί στις υποδομές που Πανεπιστημίου Πειραιώς. Τα δεδομένα που καλούμαστε να αναλύσουμε, έχουν ληφθεί με τη βοήθεια της κεραίας αυτής και αφορά τα δεδομένα των πλοίων που έχουν καταγραφεί στην ευρύτερη περιοχή του Αργοσαρωνικού.

Τα δεδομένα που λαμβάνει η κεραία και καλούμαστε να επεξεργαστούμε στην παρούσα εργασία είναι η χρονική στιγμή λήψης των δεδομένων, το γεωγραφικό μήκος (latitude) και πλάτος (longitude) που βρίσκεται εκείνη τη στιγμή το πλοίο, καθώς επίσης και την τιμή MMSI που αφορά την ταυτότητα εκπομπής του πλοίου.

Η πρόβλεψη των αποτελεσμάτων θα γίνει με την χρήση του fuzzy means algorithm[30], που υπάγεται στην αρχιτεκτονική RBF (radial basis function – ακτινικών συναρτήσεων βάσης). Υπάγεται στην κατηγορία των δικτύων που χρησιμοποιούν την επιβλεπόμενη μάθηση. Επίσης θα χρησιμοποιηθεί η τεχνική “one vs all” η οποία μας δίνει την δυνατότητα να κατηγοριοποιούμε δεδομένα που περιέχουν περισσότερες από δύο κλάσης. Η αξιολόγηση των αποτελεσμάτων θα γίνει με τον ποσοστιαίο υπολογισμό (επι της εκατό %) της επιτυχούς πρόβλεψης των δεδομένων καθώς επίσης και με τη χρήση του confusion matrix [8] προκειμένου να έχουμε μια καλύτερη εικόνα για την συμπεριφορά των δεδομένων πρόβλεψης.

Σαν αποδεκτή τιμή των δεδομένων, ορίζεται ένα ποσοστό άνω του 85% στα δεδομένα ελέγχου. Θα εξετάσουμε το ποσοστό επιτυχίας των δεδομένων και σε περίπτωση που αυτό καλύπτει το όριο που έχουμε θέσει, θα συνεχίσουμε να εξετάζουμε την αξιοπιστία του συστήματος μεταβάλλοντας τον τρόπο εκπαίδευσης καθώς και το πλήθος των δεδομένων αξιολόγηση, προκειμένου να έχουμε μια καλύτερη εικόνα για την συμπεριφορά του συστήματος.

Η δομή της παρούσας πτυχιακής είναι η ακόλουθη. Αρχικά παρουσιάζεται ένα κεφάλαιο το οποίο είναι εισαγωγικό στην έννοια των νευρωνικών δικτύων, αναλύοντας τον τρόπο λειτουργίας και την αρχιτεκτονική τους. Στο επόμενο κεφάλαιο πραγματοποιείται αναλυτική περιγραφή των RBF νευρωνικών δικτύων, που είναι η αρχιτεκτονική η οποία χρησιμοποιήθηκε για την υλοποίηση της πτυχιακής και θα γίνει αναφορά στους κυριότερους αλγόριθμους εκπαίδευσης που χρησιμοποιούνται. Το μοντέλο που θα χρησιμοποιηθεί για την πρόβλεψη των αποτελεσμάτων και οι αλγόριθμοι που το αποτελούν αναφέρονται στο τέταρτο κεφάλαιο. Το επόμενο κεφάλαιο αναφέρεται στο πειραματικά αποτελέσματα που προέκυψαν και στο τέλος παρουσιάζονται τα συμπεράσματα, οι τρόποι βελτίωσης καθώς επίσης και η βιβλιογραφία πάνω στην οποία βασίστηκε η παρούσα εργασία.

Κεφάλαιο 2

Νευρωνικά Δίκτυα

2.1 Τι είναι τα νευρωνικά δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα – ΤΝΔ (Artificial Neural Networks – ANN), αποτελούν έναν από τους κλάδους της υπολογιστικής νοημοσύνης. Η υπολογιστική νοημοσύνη είναι μια επιστημονική περιοχή που εμφανίστηκε και γνώρισε μεγάλη ανάπτυξη κατά την διάρκεια του 20^{ου} αιώνα και χρησιμοποιείται πλέον σε ένα μεγάλο εύρος εφαρμογών. Τα ΤΝΔ είναι μια σχετικά νέα περιοχή στην επιστήμη και στην μηχανική γνωρίζοντας ραγδαία ανάπτυξη τις τελευταίες δεκαετίες.

Τα ΤΝΔ βασίστηκαν στο γεγονός, ότι ο ανθρώπινος εγκέφαλος εκτελεί τους υπολογισμούς με εντελώς διαφορετικό τρόπο απ' ότι ένας συμβατικός ηλεκτρονικός υπολογιστής. Ουσιαστικά ο εγκέφαλος είναι ένας εξαιρετικά πολύπλοκος, μη γραμμικός, παράλληλος υπολογιστής. Έχει την δυνατότητα να οργανώνει τα δομικά του στοιχεία, γνωστά και ως νευρώνες, με τρόπο ώστε να εκτελούν συγκεκριμένους υπολογισμούς με ταχύτητα πολλαπλάσια από αυτή του γρηγορότερου υπολογιστή.

Ένα ενδιαφέρον παράδειγμα είναι το σύστημα όρασης μιας νυχτερίδας (Sonar). Πρόκειται για ένα ενεργό σύστημα ηχητικού εντοπισμού, παρέχοντας πληροφορίες που αφορούν την απόσταση ενός στόχου, την ταχύτητα, το μέγεθος, καθώς και το αζιμούθιο και το ύψος του από το έδαφος. Όλοι αυτοί οι πολύπλοκοι υπολογισμοί που απαιτούνται για την εξαγωγή αυτών των δεδομένων, λαμβάνουν χώρα στον εγκέφαλο της νυχτερίδας. Πρόκειται για ένα σύστημα εντοπισμού του στόχου που θα ζήλευαν ακόμα και τα πλέον προηγμένα σήμερα ραντάρ [1].

Για ποιο λόγο όμως χρειαζόμαστε ένα τόσο προηγμένο σύστημα όπως αυτό ενός βιολογικού νευρώνα; Ο λόγος που στοχεύουμε στην δημιουργία συστημάτων τα οποία προσομοιώνουν βιολογικά συστήματα είναι η επίλυση σύνθετων προβλημάτων. Πολλές από τις σύγχρονες εφαρμογές χρειάζεται να παίρνουν αποφάσεις προκειμένου να καταλήξουν στην όσο το δυνατόν καλύτερη προσέγγιση της λύσης ενός προβλήματος.

Συνοψίζοντας λοιπόν, ο ορισμός ενός νευρωνικού δικτύου είναι ο εξής *“Ένα νευρωνικό δίκτυο είναι ένας τεράστιος παράλληλος επεξεργαστής με κατανομημένη αρχιτεκτονική, ο οποίος αποτελείται από απλές μονάδες επεξεργασίας και έχει από την φύση του την δυνατότητα να αποθηκεύει εμπειρική γνώση και να την καθιστά διαθέσιμη για χρήση”*. [16]

2.2 Πλεονεκτήματα των νευρωνικών δικτύων

Τα νευρωνικά δίκτυα οφείλουν την υπολογιστική τους ισχύ και την αποτελεσματικότητά τους, αφ' ενός στην παράλληλη κατανομημένη δομή τους και αφ' ετέρου στην ικανότητά τους να μαθαίνουν. Με λίγα λόγια μπορούν να παράγουν λογικές εξόδους, για τιμές εισόδου που δεν έχουν συναντήσει κατά την διάρκεια της εκπαίδευσής τους. Με τον τρόπο αυτό μπορούν να βρίσκουν καλές προσεγγιστικές λύσεις σε πολύπλοκα προβλήματα, τα οποία δεν επιδέχονται κάποιας αναλυτικής λύσης. [17]

Τα νευρωνικά δίκτυα προσφέρουν τις παρακάτω χρήσιμες ιδιότητες και δυνατότητες:

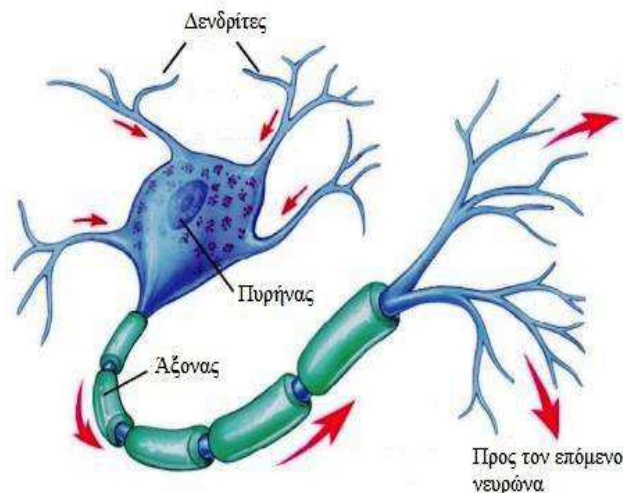
- Μη γραμμικότητα
- Αντιστοίχιση Εισόδου-Εξόδου
- Προσαρμοστικότητα
- Ενδεικτική απόκριση
- Πληροφορία σχετική με το περιεχόμενο
- Ανοχή σε βλάβες
- Δυνατότητα υλοποίησης σε VLSI
- Ομοιομορφία ανάλυσης και σχεδίασης
- Αναλογία με την νευροφυσιολογία του εγκεφάλου

2.3 Δομή βιολογικού νευρώνα

Ο νευρώνας θεωρείται η βασικότερη μονάδα του εγκεφάλου. Ο ανθρώπινος εγκέφαλος αποτελείται από ένα τεράστιο αριθμό νευρώνων. Κάθε νευρώνας συνδέεται με τους υπόλοιπους νευρώνες με συνδέσεις που ονομάζονται συνάψεις. Ο αριθμός των νευρώνων μαζί με τις διασυνδέσεις που έχει, αποτελούν ένα νευρωνικό δίκτυο. Ο εκάστοτε νευρώνας έχει μόνο δύο καταστάσεις, την ενεργό κατάσταση και την κατάσταση ηρεμίας του.

Ένας βιολογικός νευρώνας αποτελείται από τρία βασικά χαρακτηριστικά, όπως παρουσιάζονται και στο Σχ. 2.1. Το κυρίως σώμα, τον άξονα και τους δενδρίτες. Ο κυρίως κορμός αποτελεί το μέρος μέσα στο οποίο βρίσκεται ο πυρήνας του κυττάρου. Στο σημείο αυτό λαμβάνουν χώρα οι χημικές δράσεις που είναι απαραίτητες για την ζωή του κυττάρου. Ο άξονας είναι μια μεγάλη επέκταση, η οποία ξεκινάει από το σώμα και εφάπτεται με άλλους νευρώνες, προκειμένου να μεταφέρει και να λαμβάνει από αυτούς τα εξερχόμενα και εισερχόμενα σήματα αντίστοιχα. Τέλος, οι λεπτές επεκτάσεις που μοιάζουν με διακλαδώσεις δέντρων ονομάζονται δενδρίτες. Οι δενδρίτες όπως και ο άξονας είναι επιφορτισμένοι για την παραλαβή των εισερχόμενων σημάτων που προορίζονται για τον κάθε νευρώνα.

Οι συνδέσεις των νευρώνων, με τους άξονες και τους δενδρίτες, πραγματοποιούνται με επαφές που ονομάζονται συνάψεις. Οι συνάψεις δημιουργούνται στα σημεία που εφάπτονται οι δενδρίτες και είναι υπεύθυνες για την μεταφορά των σημάτων από τον ένα νευρώνα στον άλλο.

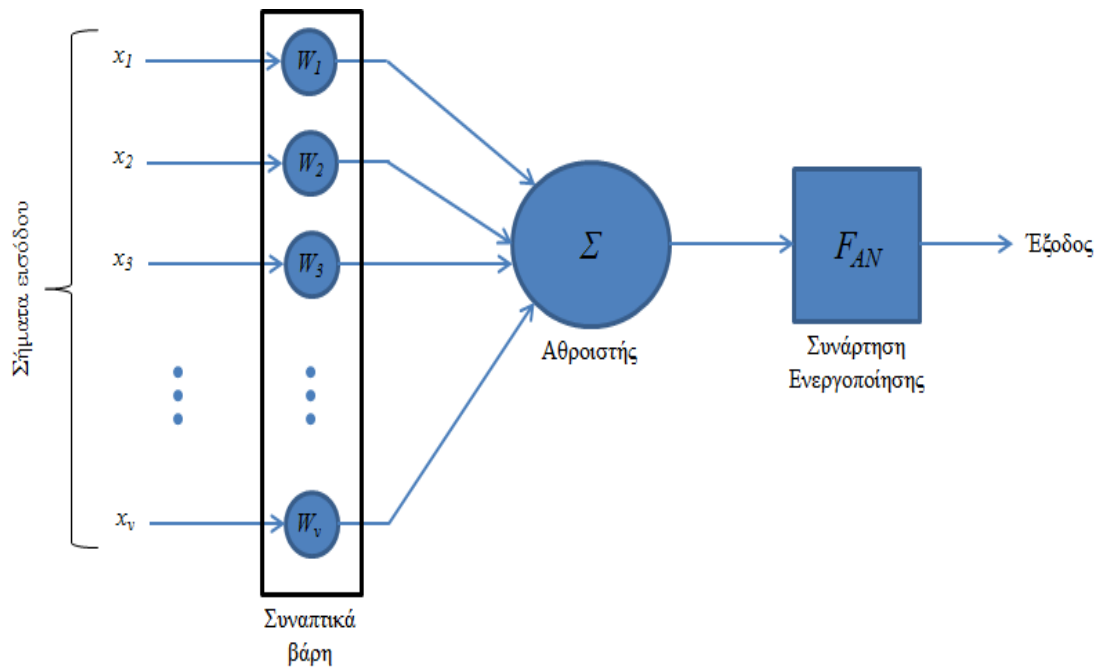


Σχήμα 2.1: Μοντέλο βιολογικού νευρώνα

2.4 Δομή τεχνητού νευρώνα

Ο νευρώνας είναι μια μονάδα επεξεργασίας και είναι θεμελιώδης για την λειτουργία του τεχνητού νευρωνικού δικτύου. Όπως είδαμε για τον βιολογικό νευρώνα, έτσι και ο τεχνητός αποτελείται από τρία βασικά στοιχεία, τις συνάψεις, έναν αθροιστή και μια συνάρτηση ενεργοποίησης.

Οι συνάψεις αποτελούνται από τις τιμές εισόδου που δίνονται σε έναν νευρώνα (x_1, x_2, \dots, x_n) , οι οποίες πολλαπλασιάζονται με μια τιμή w , την οποία ονομάζουμε συναπτικό βάρος. Τα συναπτικά βάρη μας δίνουν μια πληροφορία για το πόσο σημαντική είναι η συνεισφορά του κάθε σήματος στην διαμόρφωση της δομής του δικτύου, για τους δύο νευρώνες που συνδέει. Σε περίπτωση λοιπόν που το w είναι μεγάλο, τότε η συνεισφορά του συγκεκριμένου σήματος είναι μεγάλη και το αντίθετο. Στην συνέχεια, οι τιμές των σημάτων καταλήγουν στον αθροιστή Σ . Το σήμα που προκύπτει περνάει μέσα από μια συνάρτηση ενεργοποίησης η οποία είναι επιφορτισμένη με τον περιορισμό του πλάτους του σήματος εξόδου του κάθε νευρώνα. Οι συναρτήσεις ενεργοποίησης παρουσιάζονται αναλυτικά στην επόμενη παράγραφο. [1]



Σχήμα 2.2: Μοντέλο τεχνητού νευρώνα

2.5 Συναρτήσεις ενεργοποίησης

Σε προηγούμενη ενότητα αναφερόμενοι στον βιολογικό νευρώνα, τονίσαμε ότι υπάρχουν μόνο δύο καταστάσεις στις οποίες μπορεί να βρεθεί, την ενεργή κατάσταση και την μη ενεργή. Προκειμένου να έχουμε το ίδιο αποτέλεσμα και στο τεχνητό νευρωνικό δίκτυο, χρησιμοποιούμε τις *συναρτήσεις ενεργοποίησης*. Πιο συγκεκριμένα με βάση το μοντέλο νευρώνα του Σχ. 2.2, τα σήματα με τα συναπτικά βάρη φτάνουν στο αθροιστή όπου παράγεται ένα σήμα εξόδου. Σε περίπτωση που η τιμή της εξόδου είναι μεγαλύτερη από την τιμή κατωφλίου θ , που έχει οριστεί από την συνάρτηση ενεργοποίησης F_{AN} ο νευρώνας ενεργοποιείται, σε διαφορετική περίπτωση, ο νευρώνας παραμένει ανενεργός.

Υπάρχουν διαφορετικά είδη συναρτήσεων που μπορούμε να χρησιμοποιήσουμε ανάλογα με το ποια ταιριάζει στην εφαρμογή που θέλουμε να υλοποιήσουμε. Γενικά οι συναρτήσεις ενεργοποίησης, εξαιρουμένης της γραμμικής συνάρτησης, είναι μονοτονικά αύξουσες, όπου ισχύει [18]:

$$f_{AN}(-\infty) = 0 \quad \text{ή} \quad f_{AN}(-\infty) = -1 \quad (2.1)$$

και

$$f_{AN}(\infty) = 1 \quad (2.2)$$

Οι συνηθέστερες συναρτήσεις που χρησιμοποιούμε είναι οι εξής:

Γραμμική συνάρτηση (Linear function)

$$f_{AN}(net - \theta) = \lambda(net - \theta) \quad (2.3)$$

Όπου το λ είναι η κλίση της συνάρτησης. Οι γραμμικές συναρτήσεις παράγουν μια γραμμικά διαμορφωμένη έξοδο, όπου το λ είναι σταθερό.

Βηματική συνάρτηση (Step function)

$$f_{AN}(net - \theta) = \begin{cases} \gamma_1 & \text{εάν } net \geq \theta \\ \gamma_2 & \text{εάν } net < \theta \end{cases} \quad (2.4)$$

Η βηματική συνάρτηση παράγει δυο διαφορετικές τιμές εξόδου, ανάλογα με την τιμή του κατωφλίου θ . Συνήθως οι τιμές εξόδου είναι $\gamma_1=1$ και $\gamma_2=0$. Υπάρχει η δυνατότητα να χρησιμοποιήσουμε και διπολική έξοδο, σε αυτή την περίπτωση έχουμε $\gamma_1=1$ και $\gamma_2=-1$.

Κεκλιμένη συνάρτηση (Ramp function)

$$f_{AN}(net - \theta) = \begin{cases} \gamma & \text{εάν } net - \theta \geq \varepsilon \\ net - \theta & \text{εάν } -\varepsilon < net - \theta < \varepsilon \\ -\gamma & \text{εάν } net - \theta \leq -\varepsilon \end{cases} \quad (2.5)$$

Η κεκλιμένη συνάρτηση είναι ένας συνδιασμός της γραμμικής και της βηματικής συνάρτησης.

Σιγμοειδής συνάρτηση (Sigmoid function)

$$f_{AN}(net - \theta) = \frac{1}{1 + e^{-\lambda(net - \theta)}} \quad (2.6)$$

Η σιγμοειδής συνάρτηση αποτελεί μια συνεχή μορφή της κεκλιμένης συνάρτησης, με $f_{AN}(net) \in (0,1)$. Η παράμετρος λ , ελέγχει την κλίση της συνάρτησης. Στις περισσότερες περιπτώσεις έχουμε $\lambda=1$.

Hyperbolic tangent

$$f_{AN}(net - \theta) = \frac{e^{\lambda(net - \theta)} - e^{-\lambda(net - \theta)}}{e^{\lambda(net - \theta)} + e^{-\lambda(net - \theta)}} \quad (2.7)$$

Επίσης προσεγγίζεται ως

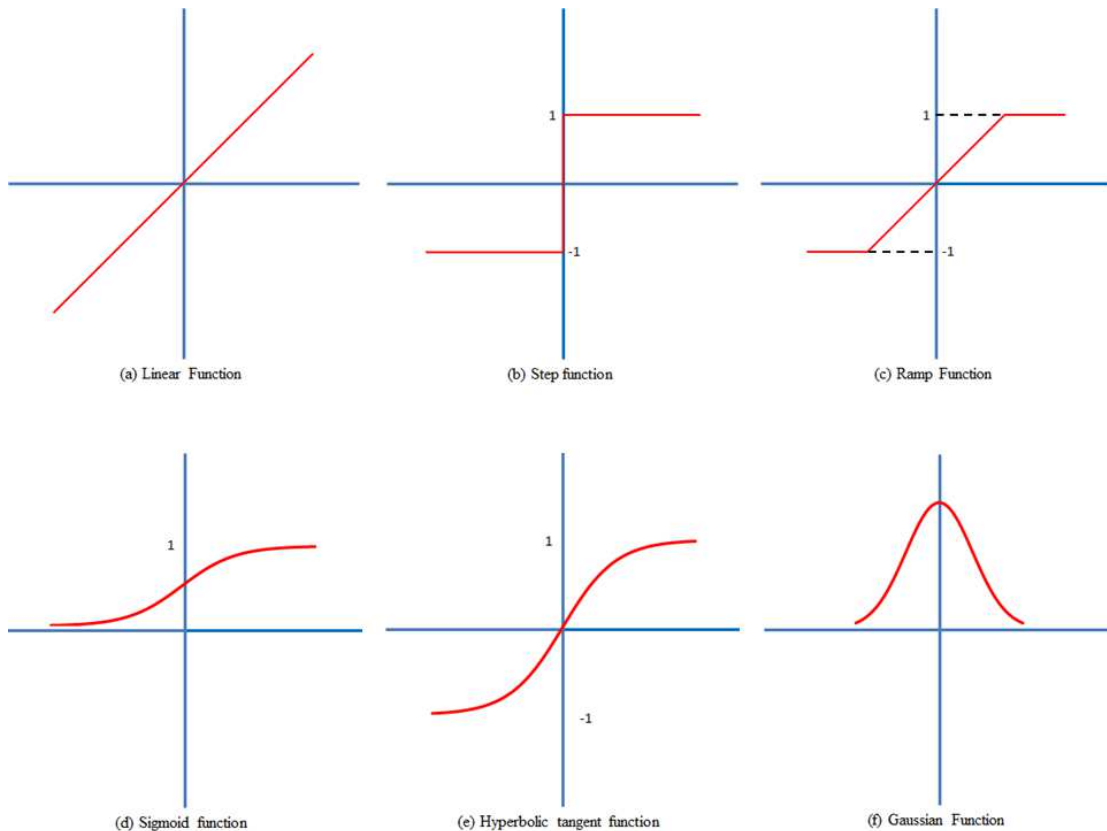
$$f_{AN}(net - \theta) = \frac{2}{1 + e^{-\lambda(net - \theta)}} - 1 \quad (2.8)$$

Η τιμή της εξόδου που προκύπτει βρίσκεται μέσα στο διάστημα $(-1,1)$.

Gaussian function

$$f_{AN}(net - \theta) = e^{-\frac{(net - \theta)^2}{\sigma^2}} \quad (2.9)$$

Όπου το $net - \theta$ είναι ο μέσος και το σ είναι η τυπική απόκλιση της Gaussian κατανομής.



Σχήμα 2.3: Συναρτήσεις ενεργοποίησης

2.6 Διαδικασίες μάθησης

Σκοπός του ΤΝΔ, όπως έχουμε τονίσει, είναι να έχει την δυνατότητα να επιλύει κάποια προβλήματα ή να πραγματοποιεί ορισμένες διεργασίες, όπως για παράδειγμα είναι η αναγνώριση προτύπων. Για να μπορεί το νευρωνικό δίκτυο να φέρει σε πέρας τις παραπάνω λειτουργίες, θα πρέπει να προηγηθεί η διαδικασία της εκπαίδευσης του, που αποτελεί και την σημαντικότερη φάση για την σωστή δημιουργία του.

Ποιος ο λόγος όμως που είναι απαραίτητη η εκπαίδευση του ΤΝΔ; Για να καταλάβουμε το γιατί είναι τόσο χρήσιμη αυτή η διαδικασία θα ανατρέξουμε για μια ακόμα φορά σε ένα παράδειγμα βιολογικού συστήματος. Ο άνθρωπος, λέμε, πως μαθαίνει καθ' όλη την διάρκεια της ζωής. Αυτό σημαίνει πως λαμβάνει ερεθίσματα από το περιβάλλον, τα οποία επεξεργάζεται και καταλήγει σε κάποια συμπεράσματα, π.χ. έχουμε μάθει κατά την διάρκεια της ζωής μας, πως για να διασχίσουμε ένα δρόμο θα πρέπει πρώτα να ελέγξουμε αν περνάνε ή όχι αυτοκίνητα, μηχανές, κτλ. Δηλαδή παίρνουμε δεδομένα από τα αισθητήρια όργανά μας, στην προκειμένη περίπτωση τα μάτια μας, τις πληροφορίες αυτές επεξεργάζεται ο εγκέφαλος μας ο οποίος και έχει εκπαιδευτεί κατάλληλα κατά την διάρκεια της ζωής μας, με σκοπό να πάρει την σωστή απόφαση για το αν θα διασχίσουμε ή όχι τον δρόμο.

Την ίδια διαδικασία ακολουθούμε και για το ΤΝΔ. Ένας από τους πλέον διαδεδομένους τρόπους εκπαίδευσης είναι ο εξής: Δίνουμε στο νευρωνικό δίκτυο ορισμένες αριθμητικές τιμές εισόδου και τις αντίστοιχες τιμές εξόδου. Αυτοί οι αριθμοί εισόδου αποτελούν κάποιο πρότυπο. Σε κάθε πρότυπο εισόδου υπάρχει και μια σωστή απάντηση, η οποία είναι η τιμή που θέλουμε να πάρουμε σαν έξοδο. Κατά την διάρκεια της εκπαίδευσης δίνουμε στο δίκτυο μια ομάδα από τέτοια πρότυπα τα οποία είναι αντιπροσωπευτικά ή παρόμοια με αυτά που θέλουμε να μάθει. Ουσιαστικά είναι σαν να κάνουμε στο δίκτυο κάποια ερώτηση μέσα από τα δεδομένα εισόδου δίνοντας του παράλληλα την σωστή απάντηση με τα δεδομένα εξόδου. Η σημαντική αλλαγή που παρουσιάζει το ΤΝΔ κατά την διάρκεια της εκπαίδευσης του έγκειται στην διαφοροποίηση των τιμών που έχουν τα συναπτικά βάρη, ενώ οι αρχικές τιμές δίνονται τυχαία με την διαδικασία της εκπαίδευσης. Το νευρωνικό δίκτυο διαφοροποιεί τις τιμές αυτές με σκοπό να πετύχει τον καλύτερο συσχετισμό των τιμών εισόδου, με αυτές των τιμών εξόδου. Μόλις ολοκληρωθεί σωστά η διαδικασία της εκπαίδευσης, το δίκτυο θα είναι σε θέση να λύνει ίδιες φύσης προβλήματα, για δεδομένα εισόδου που δεν έχει συναντήσει κατά την διάρκεια της δημιουργίας του.

Όπως μπορούμε λοιπόν να συμπεράνουμε, η διαδικασία της εκπαίδευσης αποτελεί το πλέον κρίσιμο σημείο για την περαιτέρω εύρυθμη λειτουργία και αποτελεσματικότητα του ΤΝΔ. Υπάρχουν τρεις βασικοί τύποι εκπαίδευσης, *η εκπαίδευση με επίβλεψη, η εκπαίδευση χωρίς επίβλεψη και η ενισχυτική μάθηση.*

Εκπαίδευση με επίβλεψη

Αυτός ο τρόπος εκπαίδευσης συχνά συναντάται στην βιβλιογραφία με τον όρο *επιβλεπόμενη μάθηση*. Ο κάθε νευρώνας παρουσιάζει κάποιους περιορισμούς που αφορούν στον τύπο της συνάρτησης που μπορεί να μάθει, δηλαδή μπορεί να χρησιμοποιείται μόνο για τον προσδιορισμό γραμμικών συναρτήσεων. Προκειμένου να πετύχουμε την εκμάθηση μη γραμμικών συναρτήσεων θα πρέπει να χρησιμοποιήσουμε ένα σύνολο νευρώνων το οποίο ονομάζουμε *επίπεδο νευρώνων*. Το να πραγματοποιήσουμε λοιπόν εκπαίδευση σε ένα ολόκληρο επίπεδο νευρώνων, είναι αρκετά πιο σύνθετο από την εκπαίδευση ενός και μόνο νευρώνα.

Στην περίπτωση αυτή δίνουμε στο ΤΝΔ δεδομένα τα οποία αποτελούνται από τις τιμές εισόδου και τις αντίστοιχες τιμές εξόδου. Αυτά τα δεδομένα ονομάζονται *δεδομένα εκπαίδευσης*. Σκοπός της διαδικασίας αυτής είναι η τροποποίηση των συναπτικών βαρών προκειμένου να ελαχιστοποιηθεί η τιμή της εξόδου που δίνει το ΤΝΔ με τις τιμές που εμείς θέλουμε να έχουμε σαν έξοδο (Σχ. 2.4 (α)).

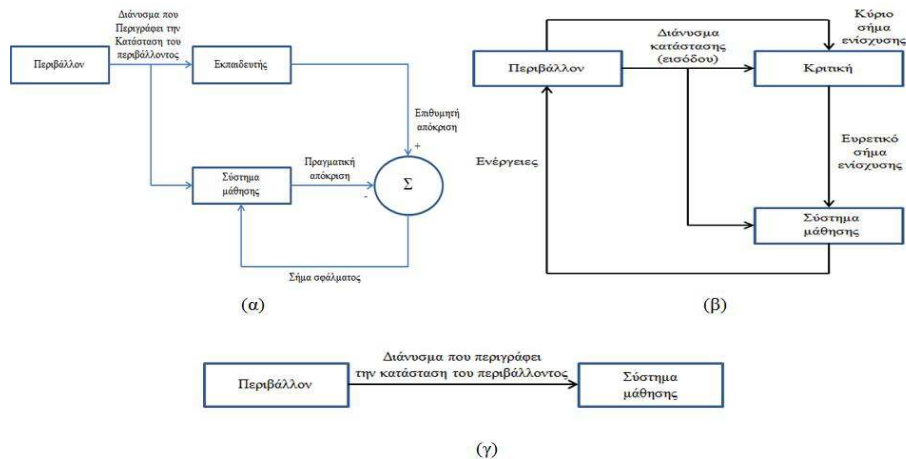
Εκπαίδευση χωρίς επίβλεψη

Οι αλγόριθμοι που χρησιμοποιούνται στην συγκεκριμένη διαδικασία μάθησης αναφέρονται στην βιβλιογραφία και ως αυτό-οργανώμενοι (self-organized). Το όνομα προήλθε από την λογική που ακολουθεί η συγκεκριμένη μέθοδος στην οποία δεν χρειάζεται να είναι παρών ένας εξωτερικός παρατηρητής. Ο στόχος αυτής της διαδικασίας εκπαίδευσης είναι να προκύψει ένα μοτίβο μεταξύ των δεδομένων εισόδου και των δεδομένων εξόδου (Σχ. 2.4 (β)). Πολλοί από τους αλγόριθμους που υπάγονται στην κατηγορία αυτή βασίζονται στην διαδικασία της κατηγοριοποίησης των προτύπων εκπαίδευσης. Οι πλέον διαδεδομένοι αλγόριθμοι στην συγκεκριμένη κατηγορία είναι ο αλγόριθμος Hebbian και ο Min-Max (references).

Ενισχυτική εκπαίδευση

Με τον όρο ενισχυτική μάθηση περιγράφεται μια οικογένεια τεχνικών που εφαρμόζονται με απώτερο σκοπό το σύστημα, στην συγκεκριμένη περίπτωση το νευρωνικό δίκτυο, να μάθει μια συσχέτιση μεταξύ της εισόδου και της εξόδου μέσα από την άμεση αλληλεπίδραση του με το περιβάλλον. Το σύστημα δεν καθοδηγείται από κάποιον εξωτερικό επιβλέποντα για να μάθει σωστά την συσχέτιση, αλλά πρέπει μόνο του να ανακαλύψει τις ενέργειες που θα του αποφέρουν το καλύτερο δυνατό αποτέλεσμα (Σχ. 2.4 (γ)).

Η τεχνική που χρησιμοποιείται για την επίτευξη του στόχου αυτού είναι εμπνευσμένη από την συμπεριφορά της ανταμοιβής και της τιμωρίας που παρατηρείται και στην ανθρώπινη συμπεριφορά.



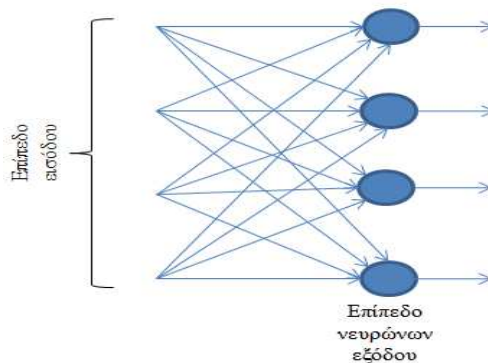
Σχήμα 2.4:(α) Μάθηση με εκπαιδευτή, (β) Μάθηση χωρίς εκπαιδευτή, (γ) Ενισχυτική μάθηση

2.7 Αρχιτεκτονικές δικτύων

Ο τρόπος με τον οποίο είναι δομημένοι οι νευρώνες εντός του νευρωνικού δικτύου είναι στενά συνδεδεμένος με τον τύπο του αλγόριθμου μάθησης που χρησιμοποιείται για την εκπαίδευσή του δικτύου. Οι βασικότερες αρχιτεκτονικές δικτύου είναι οι εξής:

Δίκτυο πρόσθιας τροφοδότησης (Feedforward neural networks-FFNNs)

Η οργάνωση των νευρώνων σε ένα νευρωνικό δίκτυο γίνεται με την μορφή επιπέδων. Στην απλούστερη μορφή ενός τέτοιου δικτύου, συναντάμε μια στοιβάδα εισόδου η οποία συνδέεται απευθείας με την στοιβάδα των νευρώνων εξόδου. Ένα τέτοιο δίκτυο παρουσιάζεται στο Σχ. 2.5.



Σχήμα 2.5: Δίκτυο πρόσθιας τροφοδότησης με ένα επίπεδο.

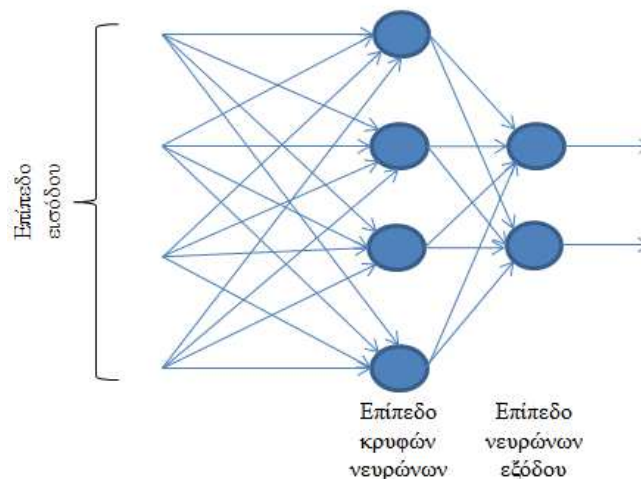
Δίκτυα πρόσθιας τροφοδότησης πολλών στοιβάδων (MLP)

Αποτελεί την δεύτερη αρχιτεκτονική των ΤΝΔ και χαρακτηρίζεται από την παρουσία περισσότερων στοιβάδων μεταξύ του επιπέδου εισόδου και εξόδου τα οποία και ονομάζουμε κρυφές στοιβάδες. Τα ονομάζουμε έτσι γιατί δεν είναι άμεσα ορατά ούτε από την είσοδο, ούτε από την έξοδο του δικτύου (Σχ. 2.6). Η λειτουργία που επιτελούν είναι να παρεμβαίνουν μεταξύ της εισόδου και της εξόδου του δικτύου υλοποιώντας ιδιαίτερα σύνθετους υπολογισμούς. Πως μπορούμε να ξέρουμε όμως πόσα κρυφά επίπεδα πρέπει να χρησιμοποιήσουμε για την κάθε εφαρμογή; Για τα περισσότερα προβλήματα η χρήση ενός κρυφού επιπέδου θεωρείται επαρκής αριθμό. Δύο επίπεδα χρησιμοποιούνται συνήθως για την μοντελοποίηση δεδομένων που

παρουσιάζουν ασυνέχειες. Χρησιμοποιώντας δύο κρυφά επίπεδα επιτυγχάνουμε καλύτερη ευελιξία, όμως πρέπει να είμαστε ιδιαίτερα προσεκτικοί στον σχεδιασμό, καθώς ελλοχεύει ο κίνδυνος να οδηγηθεί το δίκτυο σε κατάρρευση, άρα θα πρέπει να ήμαστε ιδιαίτερα προσεκτικοί κατά την διάρκεια της εκπαίδευσης.

Αυτό που πρέπει τέλος να τονίσουμε για τα πολυεπίπεδα δίκτυα, είναι πως μπορούν να υπάρξουν δύο επιμέρους κατηγορίες σύνδεσης των νευρώνων, τα *πλήρως συνδεδεμένα* όπου ο κάθε ένας κόμβος ενός επιπέδου συνδέεται με όλους τους κόμβους του επόμενου επιπέδου, και τα *μερικώς συνδεδεμένα* στα οποία ενδέχεται να λείπουν μερικές συνδέσεις.

Και για το δίκτυο της πρόσθιας τροφοδότησης όπως και στο πολυεπίπεδο δίκτυο πρόσθιας τροφοδότησης χρησιμοποιούνται οι αλγόριθμοι μάθησης με επίβλεψη, καθώς κατά την σχεδίαση δεν πραγματοποιείται κάποια ανατροφοδότηση.

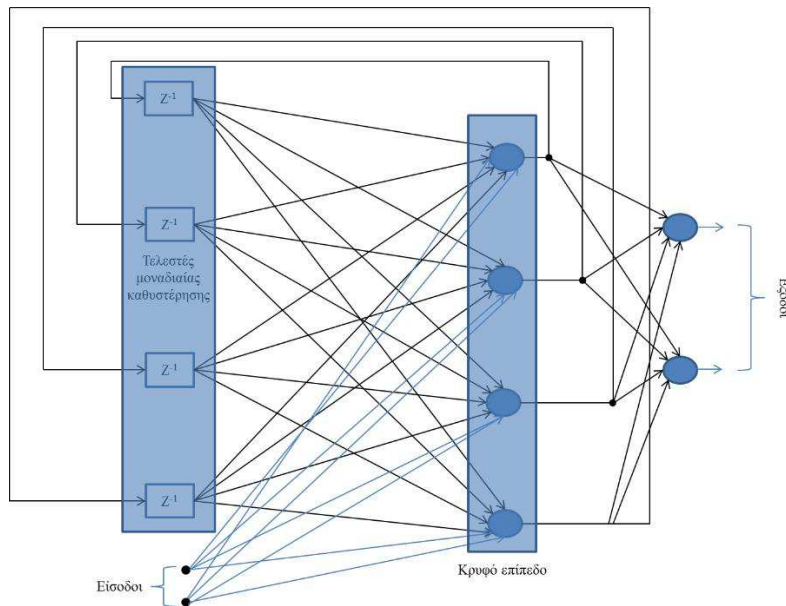


Σχήμα 2.6: Πλήρως συνδεδεμένο δίκτυο πρόσθιας τροφοδότησης με ένα κρυφό επίπεδο.

Αναδρομικά δίκτυα

Η βασική διαφοροποίηση μεταξύ των δικτύων πρόσθιας τροφοδότησης και των αναδρομικών δικτύων είναι πως στα δεύτερα υπάρχει τουλάχιστον ένας κόμβος ανάδρασης. Με τον κόμβο ανάδρασης πετυχαίνουμε την διόρθωση των συναπτικών βαρών των νευρώνων με την μέθοδο της ελαχιστοποίησης του σφάλματος. Απαραίτητη προϋπόθεση για την χρήση της ανάδρασης είναι η ύπαρξη των κλάδων μοναδιαίας καθυστέρησης οι οποίοι έχουν σαν αποτέλεσμα την δημιουργία μη γραμμικής συμπεριφοράς.

Η απλούστερη μορφή ενός αναδρομικού δικτύου είναι το μοντέλο του Elman Σχ. 2.7. Πρόκειται για ένα πολυεπίπεδο δίκτυο το οποίο έχει μόνο μια κρυφή στοιβάδα, όπου στην προκειμένη περίπτωση περιέχει τέσσερις νευρώνες οι οποίοι δέχονται τις τιμές των δύο εισόδων. Οι έξοδοι που παράγουν, ανατροφοδοτούνται την επόμενη χρονική στιγμή (λόγω των τελεστών καθυστέρησης).



Σχήμα 2.7: Αναδρομικό δίκτυο (μοντέλο Elman).

Ελάχιστα τετράγωνα

Τα ελάχιστα τετράγωνα είναι μια μεθοδολογία η οποία χρησιμοποιείται για τον προσδιορισμό των παραμέτρων ενός συστήματος. Ουσιαστικά επιλύεται ένα σύστημα στο οποίο οι εξισώσεις είναι περισσότερες από τους αγνώστους. Με τον όρο ελάχιστα τετράγωνα προσδιορίζουμε πως η συνολική λύση ελαχιστοποιεί το άθροισμα των τετραγώνων των σφαλμάτων κάθε εξίσωσης [19].

Η μεθοδολογία των ελαχίστων τετραγώνων εξετάζει δύο κατηγορίες προβλημάτων, τα γραμμικά και τα μη γραμμικά. Το πρόβλημα των γραμμικών ελαχίστων τετραγώνων εμφανίζεται στον τομέα της στατιστικής ανάλυσης και συνήθως για την επίλυση του προβλήματος αυτού υπάρχει μια σταθερή λύση, δηλαδή υπάρχει μια φόρμουλα η οποία θεωρούμε πως προσομοιώνει σωστά το μοντέλο, ύστερα από έναν πεπερασμένο αριθμό πράξεων.

Η πιο σημαντική εφαρμογή στην οποία χρησιμοποιείται η μέθοδος αυτή είναι η προσαρμογή των δεδομένων [20]. Τα ελάχιστα τετράγωνα υλοποιούνται σε μορφή πινάκων μέσω της ακόλουθης εξίσωσης Εξ. 2.10.

$$A = (X' \cdot X)^{-1} \cdot X' \cdot Y \tag{2.10}$$

όπου X είναι ο προσδιορισμός των τιμών εισόδου μέσα στην οποία συμπεριλαμβάνεται και η στήλη με τους σταθερούς όρους οι οποίοι μας βοηθάει στον υπολογισμό της πόλωσης b, Y είναι η τιμές της εξόδου. Τέλος, ο πίνακας A που προκύπτει περιέχει τις τιμές των συντελεστών καθώς και την τιμή της πόλωσης b.

$$\begin{array}{c}
 \xrightarrow{\text{Αριθμός μεταβλητής}} \\
 X = \begin{bmatrix} x_{11} & \dots & x_{1N} & 1 \\ x_{21} & \dots & x_{2N} & 1 \\ \vdots & \dots & \vdots & 1 \\ x_{P1} & \dots & x_{PN} & 1 \end{bmatrix} \quad \downarrow \text{Αριθμός δεδομένου} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_P \end{bmatrix} \quad \downarrow \text{Αριθμός δεδομένου} \quad A = \begin{bmatrix} a_1 \\ \vdots \\ a_N \\ b \end{bmatrix}
 \end{array} \tag{2.11}$$

Συντελεστές αξιολόγησης

Προκειμένου να αξιολογήσουμε την συμπεριφορά του εκάστοτε μοντέλου υπάρχουν δύο τρόποι. Ο ένας τρόπος είναι με την βοήθεια μια γραφικής παράστασης η οποία θα παρουσιάζει τα πραγματικά αποτελέσματα σε σχέση με τις προβλέψεις, ενώ ο δεύτερος τρόπος είναι η χρήση εξισώσεων που ονομάζουμε συντελεστές αξιολόγησης. Οι συντελεστές αξιολόγησης, βέβαια, είναι αυτοί που έχουν επικρατήσει διότι μας βοηθούν να έχουμε μια αυξημένη ακρίβεια για το πόσο καλό ή όχι είναι ένα νευρωνικό δίκτυο αλλά και πέρα από αυτό, βοηθούν στην πραγματοποίηση σύγκρισης των αποτελεσμάτων δύο ή περισσότερων νευρωνικών δικτύων που δημιουργήθηκαν με διαφορετικό τρόπο, αλλά έχουν τον ίδιο σκοπό [21].

Δύο από τους πιο αξιόπιστους συντελεστές αξιολόγησης είναι το R^2 (R τετράγωνο) και ο συντελεστής $RMSE$ (Root Mean Square Error). Ο συντελεστής R^2 χρησιμοποιείται για να μας δείξει την ευστοχία που έχουν τα αποτελέσματα πρόβλεψης σε σχέση με τα πραγματικά αποτελέσματα. Το εύρος τιμών που μπορεί να λάβει ο συντελεστής R^2 είναι $[-\infty 1]$ όπου η τιμή 1 αντιστοιχεί στο τέλειο μοντέλο. Ο τρόπος υπολογισμού του συντελεστή φαίνεται στην Εξ. 2.12 [22]

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} \quad (2.12)$$

$$SS_{err} = \sum_{i=1}^K (y_i - \hat{y}_i)^2 \quad (2.13)$$

$$SS_{tot} = \sum_{i=1}^K (y_i - \bar{y})^2 \quad (2.14)$$

Στην παραπάνω εξίσωση, το y προσδιορίζει την πραγματική τιμή της εξόδου, το \hat{y} την πρόβλεψη του νευρωνικού δικτύου, και το \bar{y} τον μέσο όρο των πραγματικών τιμών εξόδου. Όσον αφορά την αξιολόγηση των αποτελεσμάτων, για ένα αξιόλογο μοντέλο νευρωνικού δικτύου η τιμή που λαμβάνει ο συντελεστής είναι $R^2 \geq 0,6$ (ανάλογα βέβαια και από το είδος της εφαρμογής που καλούμαστε να υλοποιήσουμε). Η μικρότερη τιμή που ενδέχεται να πάρει είναι ίση με $-\infty$, αλλά και κάποια τιμή κάτω από το 0,5 δεν θεωρείται αξιόλογη.

Ο επόμενος συντελεστής αξιολόγησης που χρησιμοποιήθηκε για τον χαρακτηρισμό του μοντέλου είναι ο $RMSE$ (Root Mean Square Error – Ρίζα Μέσου Τετραγωνικού Σφάλματος). Στην βιβλιογραφία ενδέχεται να το συναντήσουμε και ως $RMSD$ (Root Mean Square Deviation), και αυτός είναι ένας δείκτης που μας δίνει την αξιολόγηση του μοντέλου όσον αφορά την σχέση των πραγματικών τιμών και των προβλέψεων. Είναι μια μέθοδος η οποία μας δίνει ιδιαίτερα σημαντικές πληροφορίες για την ακρίβεια του μοντέλου που κατασκευάσαμε. Ο δείκτης αυτός μπορεί να συγκρίνει τιμές μόνο για μια μεταβλητή εξόδου και όχι για παραπάνω. Αυτός είναι και ένας από τους λόγους που τον χρησιμοποιήσαμε στον χαρακτηρισμό του μοντέλου μας, απ' την στιγμή που στην εφαρμογή μας είχαμε μόνο μια μεταβλητή εξόδου. Η υλοποίηση του παρουσιάζεται στην Εξ. 2.13

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2.15)$$

με y συμβολίζονται οι πραγματικές τιμές εξόδου, με \hat{y} οι τιμές της πρόβλεψης και με n το πλήθος των δεδομένων.

Ο συντελεστής αυτός λαμβάνει τιμές από $[0 +\infty]$ με την τιμή μηδέν να χαρακτηρίζει το τέλειο νευρωνικό δίκτυο.

Ένας τρόπος παρουσίασης των επιδόσεων ανά κλάση ενός κατηγοριοποιητή είναι με τη χρήση του πίνακα σύγχυσης (confusion matrix). Ο Πίνακας Σύγχυσης είναι ένας διδιάστατος πίνακας, όπου οι στήλες αντιστοιχούν στις προβλέψεις και οι γραμμές στις πραγματικές τιμές κλάσης. Στα κελιά του πίνακα αναγράφονται οι αληθινές θετικές, οι αληθινές αρνητικές, οι ψευδείς θετικές και οι ψευδείς αρνητικές προβλέψεις δίνοντας την δυνατότητα του υπολογισμού της ακρίβειας με την οποία το δίκτυο προβλέπει τα δεδομένα .

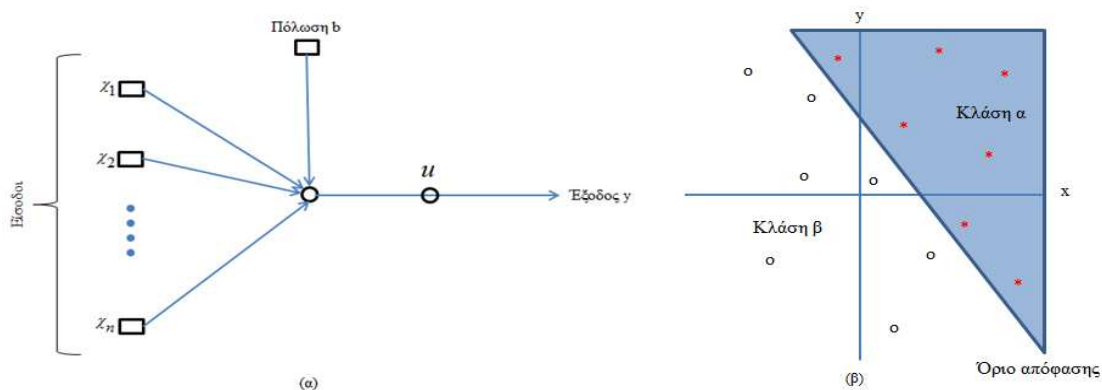
Ιστορική αναδρομή – perceptron του Rosenblatt

Τα ΤΝΔ, όπως αναφέραμε και σε προηγούμενη παράγραφο, ξεκίνησαν να γνωρίζουν μεγάλη άνθιση και να χρησιμοποιούνται σε εφαρμογές από την δεκαετία του 1980. Η πρώτη εμφάνιση τους όμως, έγινε μερικές δεκαετίες πριν, και πιο συγκεκριμένα στις αρχές της δεκαετίας του 40'.

Το 1943 λοιπόν, παρουσιάστηκε το πρώτο μοντέλο ΤΝΔ από τους McCulloch και Pitts. Ο McCulloch ήταν καθηγητής νευροφυσιολογίας και ο Pitts ήταν φοιτητής μαθηματικών. Ήταν οι πρώτοι που προσπάθησαν να προσομοιώσουν με χρήση μαθηματικών μοντέλων τις λειτουργίες των νευρώνων του ανθρώπινου εγκεφάλου.

Την σκυτάλη από τους McCulloch-Pitts παίρνει μερικά χρόνια αργότερα, το 1949, ο Καναδός ψυχολόγος Donald O. Hebb. Ο D. Hebb ασχολήθηκε ιδιαίτερα με τον τομέα της νευροφυσιολογίας και το 1949 εκδίδει το βιβλίο «The Organization of Behavior» [23] στο οποίο παρουσιάζεται ο ισχυρισμός ότι η ανθρώπινη συμπεριφορά είναι απόλυτα συνδεδεμένη με την δομή του ανθρώπινου εγκεφάλου.

Για να γίνει το επόμενο βήμα στον τομέα των ΤΝΔ χρειάστηκε να περάσουν οχτώ χρόνια. Το 1957 ο Αμερικανός ψυχολόγος Frank Rosenblatt [24] παρουσιάζει το μοντέλο του Perceptron, το οποίο τελειοποίησε το 1959, και μέχρι σήμερα καταλαμβάνει μια ειδική θέση στην εξέλιξη των ΤΝΔ. Το perceptron Σχ. 2.8(α) ουσιαστικά αποτελείται από έναν νευρώνα, τις εισόδους με τα συναπτικά βάρη, και μια πόλωση. Με την χρήση του perceptron μπορούμε να πετύχουμε ταξινόμηση δύο κλάσεων όπως φαίνεται στο Σχ. 2.8(β), εφόσον οι κλάσεις είναι γραμμικά διαχωρίσιμες. Επίσης χρησιμοποιώντας περισσότερους νευρώνες και ανάλογες εξόδους μπορούμε να επιτύχουμε ταξινόμηση προτύπων σε περισσότερες από δύο κλάσεις.



Σχήμα 2.8: (α) Μοντέλο perceptron (β) Όριο απόφασης

Την ίδια χρονική περίοδο που τελειοποιήθηκε το perceptron, το 1959 δηλαδή, ο Bernard Widrow καθηγητής ηλεκτρονικός μηχανικός και ο διδακτορικός φοιτητής Marcian Edward "Ted" Hoff, ανέπτυξαν δύο νέα μοντέλα. Το ADALINE (ADaptive LInear NEuron), πρόκειται για ένα νευρωνικό δίκτυο με ένα επίπεδο στο οποίο υπάρχουν αρκετοί νευρώνες. Ο κάθε νευρώνας δέχεται έναν αριθμό εισόδων και παράγει μόνο μια έξοδο. Το επόμενο μοντέλο που ανέπτυξαν ονομάστηκε MADALINE (Multiple ADaptive Linear NEuron), ουσιαστικά είναι ένα δίκτυο βασισμένο στο μοντέλο του ADALINE με την διαφορά ότι εμπεριέχει δύο επίπεδα. Και τα δύο αυτά μοντέλα χρησιμοποιήθηκαν σε πληθώρα πρακτικών προβλημάτων, όπως για να εξαλείψουν την ηχώ σε τηλεφωνικές γραμμές [25].

Σύγχρονες εφαρμογές νευρωνικών δικτύων

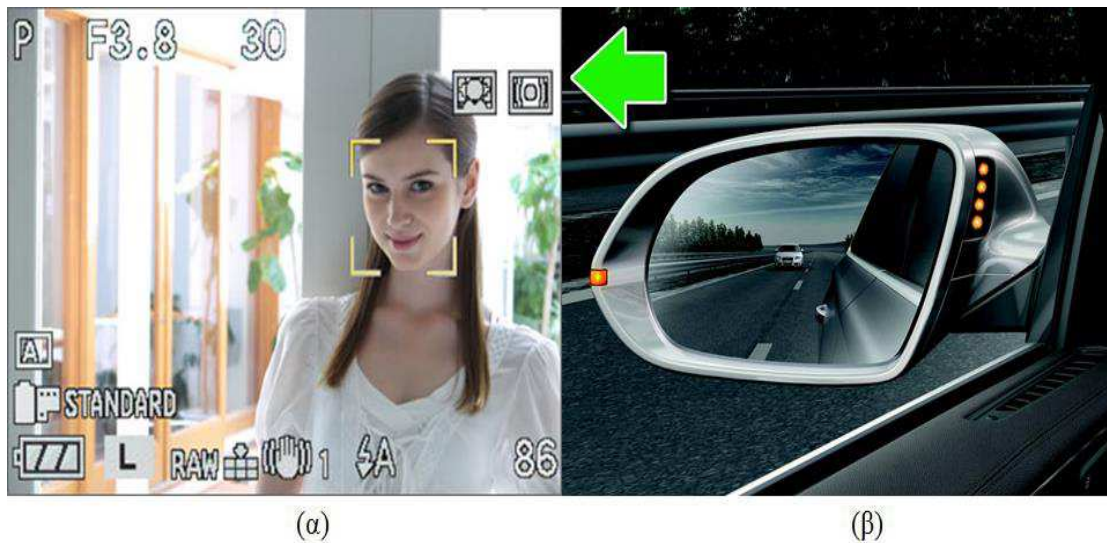
Τα ΤΝΔ τις τελευταίες δύο δεκαετίες έχουν βρει τον δικό τους χώρο στο τομέα της επιστήμης, και της μηχανικής, ενώ παρατηρείται ότι όλο και μεγαλύτερος αριθμός επιστημόνων ασχολούνται με την περαιτέρω ανάπτυξή τους. Χρησιμοποιούνται με μεγάλη επιτυχία σε όλο το φάσμα των επιστημών, όπως στα οικονομικά για την πρόβλεψη του χρηματιστηριακού δείκτη, στην ιατρική για την έγκαιρη διάγνωση ασθενειών, στην ηλεκτρονική για αναγνώριση προτύπων καθώς επίσης έχουν μεγάλη συμβολή στην επικοινωνία ανθρώπου-μηχανής που τα τελευταία χρόνια γνωρίζει ραγδαία εξέλιξη.

Αν και ένα μεγάλο φάσμα της βιομηχανίας χρησιμοποιεί τα ΤΝΔ για την δημιουργία εμπορικών εφαρμογών, παρατηρείται πως υπάρχει μεγάλο περιθώριο για περαιτέρω εξέλιξή τους, κάτι το οποίο μας δίνει την δυνατότητα να αισιοδοξούμε πως μετά από μερικά χρόνια ίσως να ήμασταν σε θέση να τα χρησιμοποιήσουμε για την επίλυση προβλημάτων που ακόμα και σήμερα δυσκολεύουν τους επιστήμονες και μηχανικούς.

Καλό θα ήταν να αναφέρουμε μερικές από τις εφαρμογές που υπάρχουν σήμερα στο εμπόριο ή θα εμφανιστούν σε αυτό το αμέσως επόμενο χρονικό διάστημα. Μια από τις εφαρμογές που συναντάμε στο εμπόριο σήμερα, είναι οι ψηφιακές φωτογραφικές μηχανές με δυνατότητα αναγνώρισης προσώπου (face recognition). Μέσω ενός chip στο εσωτερικό της φωτογραφικής μηχανής, πραγματοποιείται σάρωση της εικόνας, το σημείο δηλαδή που εστιάζει ο φακός. Σε περίπτωση που ανιχνεύσει το σύστημα τα χαρακτηριστικά του ανθρώπινου προσώπου, και μπορεί να ξεχωρίσει το άτομο με βάση τα πρότυπα που έχει μάθει κατά την διάρκεια της εκπαίδευσης. Με τον τρόπο αυτό ο χρήστης έχει την δυνατότητα να ανεβάσει στα μέσα κοινωνικής δικτύωσης φωτογραφίες κοινοποιώντας παράλληλα και τα άτομα τα οποία βρίσκονται μαζί. Σχ. 2.9(α).

Η επόμενη εφαρμογή προέρχεται από τον χώρο της αυτοκινητοβιομηχανίας. Η Audi έχει προβεί τα τελευταία χρόνια στην μελέτη και ανάπτυξη συστήματος «απόφασης αλλαγής λωρίδας». Στους καθρέφτες έχουν τοποθετηθεί αισθητήρες οι οποίοι καταγράφουν τυχόν εμφάνιση οχήματος και ενημερώνουν τον οδηγό προκειμένου αυτός να μην προβεί σε αλλαγή λωρίδας. Και σε αυτή την περίπτωση πρόκειται για ένα νευρωνικό δίκτυο το οποίο πραγματοποιεί αναγνώριση προτύπων και αναγνωρίζει πως αυτό που ανιχνεύεται είναι όχημα και όχι κάποιο σταθερό σημείο όπως για παράδειγμα ένα δέντρο. Επίσης πραγματοποιείται παράλληλα εκτίμηση της ταχύτητας με την οποία κινείται το κάθε όχημα. Σε περίπτωση που ανιχνεύσει το σύστημα ένα όχημα που πλησιάζει με μεγάλη ταχύτητα, ενημερώνεται ο οδηγός με την χρήση LED's που είναι τοποθετημένα στους καθρέφτες Σχ. 2.9(β).

Τέλος, μια εφαρμογή που έχει ξεκινήσει αρκετό καιρό τώρα και αναμένεται να γνωρίσει μεγάλη ανταπόκριση, είναι η μετάφραση των λεγομένων μας σε διαφορετική γλώσσα [26]. Έχει ξεκινήσει ήδη το λανσάρισμα στην αγορά μέσω των κινητών τηλεφώνων. Στην εφαρμογή αυτή χρησιμοποιούνται νευρωνικά δίκτυα τα οποία μαθαίνουν μια συσχέτιση μεταξύ των δύο γλωσσών και μας επιτρέπουν την ακριβή μετάφραση των φράσεων μας στην γλώσσα που έχουμε εμείς επιλέξει (Σχ.2.10.)



Σχήμα 2.9: (α) Ψηφιακή κάμερα με λειτουργία αναγνώρισης προσώπου, (β) Σύστημα απόφασης αλλαγής λωρίδας, της αυτοκινητοβιομηχανίας Audi.



Σχήμα 2.10: Κινητό τηλέφωνο με λειτουργία μετάφρασης μηνυμάτων και ομιλίας.

Μπορούμε να καταλάβουμε λοιπόν, πως με την χρήση των νευρωνικών δικτύων στην σύγχρονη βιομηχανία, μπορούμε να δημιουργήσουμε εφαρμογές χωρίς να χρειαστεί να περιγράψουμε αναλυτικά τις σχέσεις που χαρακτηρίζουν την είσοδο με την έξοδο, πράγμα το οποίο θα ήταν ιδιαίτερα χρονοβόρο αν όχι απίθανο να συμβεί με την χρήση κάποιας άλλης μεθοδολογίας.

Παραπάνω αναφέρθηκαν μόνο μερικές από τις εφαρμογές που υπάρχουν ή βρίσκονται στην διαδικασία της παραγωγής και έχουν σχέση με τα νευρωνικά δίκτυα. Ο κατάλογος είναι πολύ μεγαλύτερος και στο μέλλον αναμένεται να αυξηθεί ακόμα περισσότερο.

Κεφάλαιο 3

Δίκτυα συναρτήσεων ακτινικής βάσης (RBF)

Στο συγκεκριμένο κεφάλαιο, θα παρουσιάσουμε αναλυτικά τα δίκτυα ακτινικής βάσης, τα οποία θα αναφέρονται ως RBF (Radial Basis Function). Άλλωστε αυτός είναι ο τύπος νευρωνικών δικτύων χρησιμοποιήθηκε για την υλοποίηση της εφαρμογής σε αυτή την πτυχιακή. Θα αναλύσουμε την αρχιτεκτονική των δικτύων αυτών, τα πλεονεκτήματα που παρουσιάζουν, καθώς και τους σημαντικότερους αλγόριθμους μάθησης [18].

Δίκτυα RBF

Τα νευρωνικά δίκτυα RBF, τα οποία στην βιβλιογραφία μπορεί να τα συναντήσουμε και ως RBFNN, παρουσιάστηκαν για πρώτη φορά το 1985. Πρόκειται για νευρωνικά δίκτυα πρόσθιας τροφοδότησης που αποτελούνται από τρία επίπεδα, το επίπεδο εισόδου, το κρυφό επίπεδο και το επίπεδο εξόδου. Χρησιμοποιούνται ευρέως στην προσέγγιση συναρτήσεων, στην πρόβλεψη δεδομένων καθώς και στον έλεγχο συστημάτων.

Η πρώτη βασική διαφορά που συναντάμε στα νευρωνικά δίκτυα RBF έγκειται στο γεγονός πως δεν χρησιμοποιούνται συναρτήσεις ενεργοποίησης, όπως έχουμε δει έως τώρα στα προηγούμενα μοντέλα. Στα RBF χρησιμοποιούνται συναρτήσεις ακτινικής βάσης, οι οποίες θα αναλυθούν σε επόμενη παράγραφο.

Για ποιο λόγο όμως να προτιμήσουμε τα μοντέλα RBF από οποιονδήποτε άλλο τύπο νευρωνικών δικτύων; Έχουμε αναφέρει πως για την εκπαίδευση των ΤΝΔ ακολουθούμε δύο βήματα. Στο πρώτο βήμα δημιουργείται η δομή του νευρωνικού δικτύου, δηλαδή, ο αριθμός των κρυφών επιπέδων από τα οποία θα αποτελείται, καθώς και ο αριθμός των νευρώνων που θα έχει η κάθε στοιβάδα. Στο δεύτερο βήμα υπολογίζονται οι τιμές των παραμέτρων (των συναπτικών βαρών κατά κύριο λόγο), του ΤΝΔ. Ο υπολογισμός αυτός επιτυγχάνεται με την μέθοδο της ελαχιστοποίησης του σφάλματος μεταξύ της πραγματικής τιμής εξόδου και της τιμής που λαμβάνουμε σαν έξοδο από το νευρωνικό δίκτυο. Οι περισσότεροι αλγόριθμοι μάθησης των ΤΝΔ βασίζονται σε μη γραμμικές μεθόδους βελτιστοποίησης, με αποτέλεσμα να απαιτείται μεγάλος υπολογιστικός χρόνος. Τα RBF νευρωνικά δίκτυα λόγω του ότι χρησιμοποιούν γραμμικές συναρτήσεις κατά την διαδικασία της εκπαίδευσης, παρουσιάζουν συγκριτικό πλεονέκτημα όσον αφορά τις δυνατότητες προσέγγισης, την απλότητα της δομής τους και την μεγάλη ταχύτητα των αλγορίθμων μάθησης. Με βάση τα παραπάνω γίνεται εύκολα κατανοητός ο λόγος που τα συγκεκριμένα μοντέλα έχουν τόσο μεγάλη απήχηση σε πληθώρα συστημάτων, καθώς και σε εφαρμογές που απαιτούν επεξεργασία δεδομένων σε πραγματικό χρόνο.

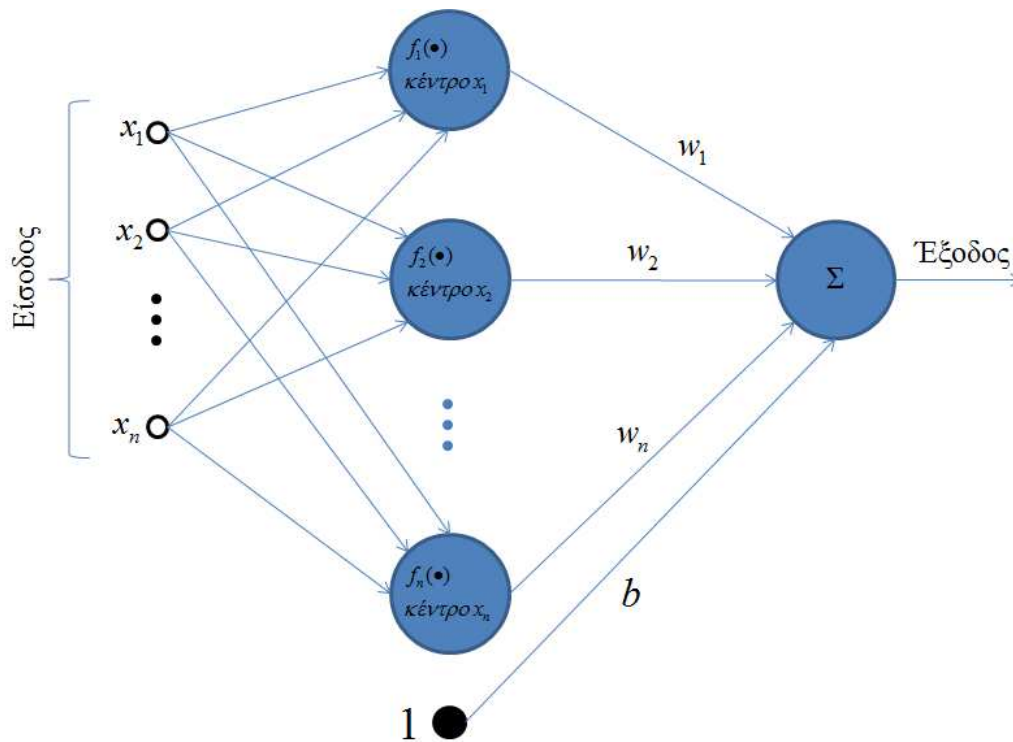
Αρχιτεκτονικές νευρωνικών δικτύων RBF

Η αρχιτεκτονική των νευρωνικών δικτύων RBF παρουσιάζει αρκετές ομοιότητες με τα υπόλοιπα δίκτυα πρόσθιας τροφοδότησης που είδαμε σε προηγούμενο κεφάλαιο. Όπως τονίσαμε, αποτελείται από τρία επίπεδα, το επίπεδο εισόδου, το κρυφό επίπεδο και το εξόδο.

Ας δούμε όμως λίγο αναλυτικότερα για ποιο λόγο τα RBF νευρωνικά δίκτυα διαφοροποιούνται από τα υπόλοιπα. Στο κρυφό επίπεδο εφαρμόζεται μια συνάρτηση ακτινικής βάσης f , με αποτέλεσμα η έξοδος του κάθε νευρώνα στο κρυφό επίπεδο να υπολογίζεται με βάση τον τύπο:

$$y = w_0 + \sum_{i=1}^{n_h} w_i f(\|x - c_i\|) \quad (3.1)$$

όπου f είναι η συνάρτηση ακτινικής βάσης, w_i είναι το βάρος του επιπέδου εξόδου, w_0 είναι η πόλωση (bias), x η είσοδος του δικτύου, c_i το κέντρο που προκύπτει από την συνάρτηση ακτινικής βάσης. Το n_h αποτελεί τον αριθμό της ακτινικής βάσης του δικτύου, και το $\|\cdot\|$ υποδηλώνει την ευκλείδεια απόσταση. Σε αυτό το σημείο πρέπει να τονιστεί πως όσο μικρότερη είναι η απόσταση των δυο διανυσμάτων, τόσο μικρότερη θα είναι η τιμή της εξόδου.



Σχήμα 3.1: Μοντέλο RBF νευρωνικού δικτύου

Συναρτήσεις ενεργοποίησης RBF

Σε αυτό το σημείο θα αναφερθούμε στις συναρτήσεις ακτινικής βάσης, οι οποίες χαρακτηρίζουν τα RBF νευρωνικά δίκτυα. Όπως έχουμε αναφέρει, οι συναρτήσεις αυτές βρίσκονται στους νευρώνες του κρυφού επιπέδου. Είθισται να τις συναντάμε στην βιβλιογραφία και ως συναρτήσεις πυρήνα. Χαρακτηρίζονται από μια μοναδική μέγιστη τιμή που προκύπτει στο κέντρο τους, c_j , και συνήθως οι τιμές τους ελαχιστοποιούνται μόλις απομακρυνθούν λίγο από αυτό. Η έξοδος των νευρώνων υποδεικνύει την κοντινότερη τιμή του διάνυσματος σε σχέση με το διάνυσμα εισόδου, x_i . Στα RBF νευρωνικά δίκτυα, συναντάμε αρκετά συχνά συναρτήσεις οι οποίες χαρακτηρίζονται, εκτός από τις τιμές που έχουν τα κέντρα τους, και από το πλάτος, σ_j .

Οι βασικότερες συναρτήσεις που συναντάμε είναι:

Γραμμική συνάρτηση

$$f(\|x_i - c_j\|_2) = \|x_i - c_j\|_2 \quad (3.2)$$

Κυβική συνάρτηση

$$f(\|x_i - c_j\|_2) = \|x_i - c_j\|_2^3 \quad (3.3)$$

Thin-Plate-Spline

$$f(\|x_i - c_j\|_2) = \|x_i - c_j\|_2^2 \ln \|x_i - c_j\|_2 \quad (3.4)$$

Multiquadratic

$$f(\|x_i - c_j\|_2, \sigma_j) = \sqrt{\|x_i - c_j\|_2^2 + \sigma_j^2} \quad (3.5)$$

Inverse Multiquadratic

$$f(\|x_i - c_j\|_2, \sigma_j) = \frac{1}{\sqrt{\|x_i - c_j\|_2^2 + \sigma_j^2}} \quad (3.6)$$

Gaussian

$$f(\|x_i - c_j\|_2, \sigma_j) = e^{-\frac{\|x_i - c_j\|_2^2}{(2\sigma_j^2)}} \quad (3.7)$$

Logistic

$$f(\|x_i - c_j\|_2, \sigma_j) = \frac{1}{1 + e^{\frac{\|x_i - c_j\|_2^2}{\sigma_j^2 - \theta_j}}} \quad (3.8)$$

Στην συνάρτηση logistic το θ_j , είναι η προκαθορισμένη κλίση της συνάρτησης.

Λαμβάνοντας υπόψη τις παραπάνω συναρτήσεις, η ακρίβεια ενός RBF νευρωνικού δικτύου επηρεάζεται, αρχικά, από τον αριθμό των συναρτήσεων που θα χρησιμοποιήσουμε. Μπορούμε δηλαδή να χρησιμοποιήσουμε περισσότερες από μια συναρτήσεις ακτινικής βάσης στο κρυφό επίπεδο. Αυτό θα μας δώσει την δυνατότητα να έχουμε καλύτερη προσέγγιση του στόχου, πάραυτα ο τρόπος αυτός αποφεύγεται συνήθως γιατί οδηγεί σε αύξηση του υπολογιστικού χρόνου. Στην συνέχεια ιδιαίτερα προσεκτικοί θα πρέπει να ήμαστε στην τοποθέτηση των κέντρων αυτών των συναρτήσεων, προκειμένου να καλυφθεί όλος ο χώρος εισόδου. Για να διευκολυνθούμε βέβαια σε αυτή την περίπτωση οδηγούμαστε σε κανονικοποίηση των τιμών εισόδου, δηλαδή τις μετατρέπουμε στο εύρος $[-1, 1]$.

Εκπαίδευση νευρωνικών δικτύων RBF

Προκειμένου να υπολογιστούν οι τιμές των παραπάνω συναρτήσεων και εν γένει οι τιμές των παραμέτρων του RBF δικτύου, χρησιμοποιούμε την διαδικασία της εκπαίδευσης. Κατά την διάρκεια των ετών έχουν παρουσιαστεί πολλές μέθοδοι εκπαίδευσης RBF δικτύων. Οι βασικές διαφορές των αλγορίθμων αυτών βρίσκονται στον αριθμό των παραμέτρων που μαθαίνουν στο δίκτυο. Απαραίτητη προϋπόθεση για την εφαρμογή των αλγορίθμων αυτών, είναι η διαδικασία εύρεσης των παραμέτρων να γίνεται με μη επιβλεπόμενο τρόπο. Στην ενότητα, λοιπόν, αυτή θα παρουσιάσουμε τους αλγόριθμους μάθησης που έχουν επικρατήσει, καθώς και τον αλγόριθμο που χρησιμοποιήσαμε για την εκπαίδευση του νευρωνικού δικτύου που χρησιμοποιήθηκε στην εφαρμογή αυτής της πτυχιακής. [27]

Αλγόριθμος διόρθωσης κέντρων

Οι Broomhead και Lowe το 1988, παρουσίασαν την μέθοδο της διόρθωσης κέντρων. Στον αλγόριθμο αυτό τα κέντρα επιλέγονται τυχαία στην αρχή της διαδικασίας της εκπαίδευσης με μια βασική προϋπόθεση, να υπάρχει επαρκής αριθμός κέντρων προκειμένου να καλύψουν όλο το φάσμα των τιμών εισόδου. Στο τέλος της εκπαίδευσης, διαγράφονται οι συναρτήσεις που είναι περιττές, δηλαδή αυτές στις οποίες δεν αντιστοιχεί καμία από τις τιμές εισόδου.

Σε περίπτωση που για την υλοποίηση του παραπάνω αλγορίθμου χρησιμοποιήσουμε Gaussian συνάρτηση, τότε για τον υπολογισμό του εύρους χρησιμοποιείται η Εξ. 3.9

$$\sigma_j = \sigma = \frac{d_{\max}}{\sqrt{J}}, j = 1, \dots, J \quad (3.9)$$

όπου J είναι ο αριθμός των κέντρων, και d_{\max} είναι η μέγιστη τιμή των κέντρων με βάση την ευκλείδεια απόστασή τους. Οι τιμές των βαρών μεταξύ του κρυφού επιπέδου και του επιπέδου εξόδου υπολογίζονται με βάση την Εξ. 3.10

$$w_k = (f^T f)^{-1} f^T t_k \quad (3.10)$$

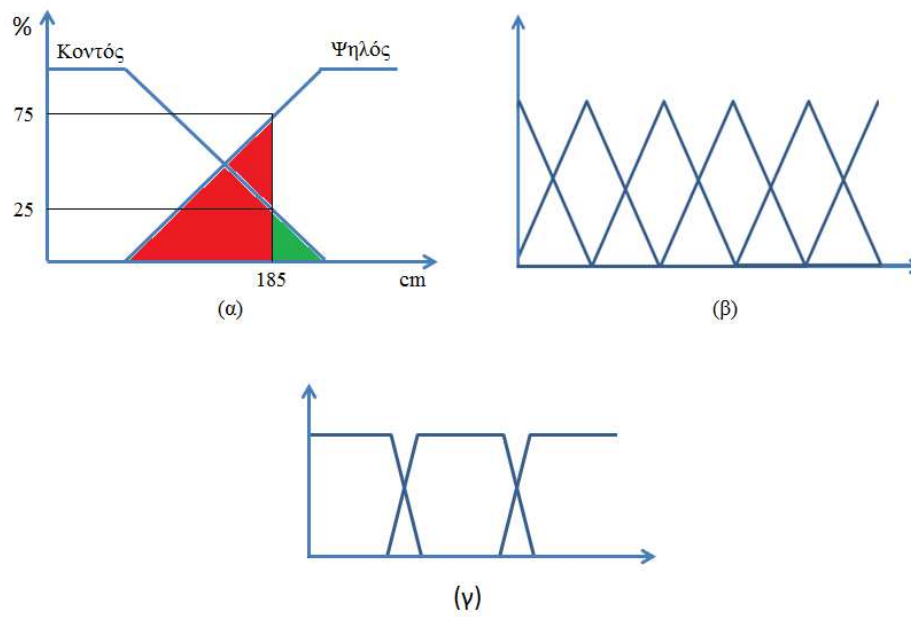
w_k είναι η τιμή που έχουν τα συναπτικά βάρη μέχρι και τον k -οστό νευρώνα, t_k είναι ο πίνακας με τις πραγματικές τιμές εξόδου και τέλος το $f \in \mathbb{R}^{p_T \times J}$ αντιστοιχεί στον πίνακα με τις μη γραμμικές αντιστοιχίες του κρυφού επιπέδου. [18]

Ασαφής λογική (Fuzzy logic)

Τα ασαφή σύνολα χρησιμοποιούνται κατά κόρον στις εφαρμογές σήμερα και η λογική τους συναντάται συχνά στην καθημερινή ζωή των ανθρώπων όταν θέλουν να επιτύχουν ακριβέστερη περιγραφή καταστάσεων [28].

Για να γίνει περισσότερο κατανοητός ο τρόπος λειτουργίας της ασαφούς λογικής, θα δώσουμε ένα παράδειγμα. Ας υποθέσουμε πως θέλουμε να χαρακτηρίσουμε το ύψος ενός ανθρώπου, χρησιμοποιώντας μόνο δύο κατηγορίες ανθρώπων, αυτούς που θεωρούνται ψηλοί και αυτούς που θεωρούνται κοντοί [29]. Για να επιτευχθεί η κατηγοριοποίηση αυτή θα πρέπει να θέσουμε και ένα όριο, έστω λοιπόν πως το όριο αυτό τίθεται στα 175 εκατοστά. Με βάση αυτή την κατηγοριοποίηση ένας άνθρωπος που έχει ύψος 177 εκατοστά θεωρείται ψηλός, και ένας ο οποίος έχει ύψος 150 εκατοστά κοντός. Όμως κοντός θεωρείται και κάποιος ο οποίος έχει ύψος 174 εκατοστά παρότι υπολείπεται μόνο ένα από την κατηγορία των ψηλών. Επίσης δύο άνθρωποι που έχουν ύψος 176 εκατοστά ο πρώτος και 210 εκατοστά ο δεύτερος θεωρούνται και οι δύο ψηλοί. Όπως μπορούμε να καταλάβουμε η μέθοδος της κατηγοριοποίησης αυτής δεν παρέχει ακριβή αποτελέσματα. Για τον λόγο αυτό χρησιμοποιείται η μέθοδος της ασαφούς λογικής, η οποία αναφέρει πως εκτός από το όριο μπορούμε να εισάγουμε και ένα ποσοστό συμμετοχής το οποίο μας δίνει την δυνατότητα να ήμαστε περισσότερο ακριβείς στον χαρακτηρισμό μας. Πιο συγκεκριμένα ένας άνθρωπος ύψος 185 εκατοστών θεωρείται κατά 75% ψηλός και κατά 25% κοντός και κάποιος ο οποίος έχει ύψος 210 θεωρείται κατά 100% ψηλός Σχ. 3.2(α). Με τον τρόπο αυτό υπάρχει μια ορθή διάκριση όσον αφορά στον χαρακτηρισμό του ύψους.

Ανάλογα με το είδος της εφαρμογής μπορούμε να χρησιμοποιήσουμε όσα ασαφή σύνολα κρίνονται απαραίτητα. Συνήθως, για την δημιουργία των ασαφών συνόλων χρησιμοποιούνται τριγωνικά μοντέλα (Σχ. 3.2(β)). Παρολ' αυτά μπορούμε να χρησιμοποιήσουμε όποια μοντέλα κρίνουμε εμείς απαραίτητα όπως αυτά του τραπέζιου Σχ. 3.2(γ), της Gaussian κτλ. Επίσης μπορούμε να χρησιμοποιήσουμε και διαφορετικό εύρος (βάση τριγώνου).



Σχήμα 3.2: (α) Κατηγοριοποίηση με βάση το ύψος με χρήση δύο ασαφών συνόλων, (β) χρήση πέντε τριγωνικών συνόλων, (γ) ασαφή σύνολα τραπεζίου.

Κεφάλαιο 4

Υλοποίηση Μοντέλου

Αλγόριθμος των ασαφών μέσων (Fuzzy means algorithm)

Ο αλγόριθμος των ασαφών μέσων (*fuzzy means*) [30, 31] εισήχθη για πρώτη φορά πριν από μια δεκαετία με σκοπό την αντικατάσταση του πλέον διαδεδομένου μέχρι εκείνη τη στιγμή αλγορίθμου εκπαίδευσης των νευρωνικών δικτύων RBF, γνωστό ως *k-means*, για τον οποίο αναφερθήκαμε σε προηγούμενη παράγραφο. Ο αλγόριθμος *Fuzzy means* παρουσιάζει ιδιαίτερα προτερήματα σε σχέση με τον *k-means*, όπως η αυξημένη ταχύτητα των υπολογισμών και ο αυτόματος προσδιορισμός της δομής του δικτύου. Ο αλγόριθμος αυτός πλέον χρησιμοποιείται σε μια πληθώρα εφαρμογών [32, 33]. Πρόσφατα μάλιστα, παρουσιάστηκε και μια ανανεωμένη έκδοση του με χρήση μη συμμετρικής διαμέρισης του χώρου και βελτιστοποίησης σμήνους σωματιδίων (Particle Source optimization PSO) [34].

Ας υποθέσουμε πως έχουμε ένα σύστημα με N κανονικοποιημένες εισόδους u_i , όπου $i = 1, \dots, N$. Η κάθε μια από τις διαστάσεις εισόδου, χωρίζεται σε ίσα διαστήματα με την χρήση τριγωνικών ασαφών συνόλων (στο Σχ. 4.1(α) ένα παράδειγμα από νευρωνικό δίκτυο RBF δύο εισόδων). Το κάθε ένα από τα ασαφή σύνολα c , μπορεί να πάρει την μορφή που φαίνεται στην Εξ. 4.1.

$$A_{i,j} = \left\{ \alpha_{i,j}, \delta\alpha \right\}, \quad i = 1, \dots, N, \quad j = 1, \dots, c \quad (4.1)$$

όπου $\alpha_{i,j}$ είναι το κέντρο του κάθε τριγωνικού fuzzy set $A_{i,j}$ και $\delta\alpha$ είναι το μισό μέγεθος του αντίστοιχου πλάτους. Εξαιτίας της συμμετρικής διαμέρισης το πλάτος όλων είναι ίσο άρα και το $\delta\alpha$ παραμένει σταθερό. Με τον τρόπο αυτό έχουμε την δυνατότητα να δημιουργήσουμε ένα πλήθος c^N από πολυδιάστατους ασαφείς υπό-χώρους A^l , όπου $l = 1, \dots, c^N$. Κάθε ένα από αυτά τα πολυδιάστατα υπό-διαστήματα, παράγεται από τον συνδυασμό N μονοδιάστατων ασαφών συνόλων, ένα από κάθε δάνυσμα εισόδου. Ο πίνακας που παρουσιάζει τα κέντρα \mathbf{a}^l , καθώς και η τιμή $\delta\alpha$ που έχει το κάθε ένα από τα υπό διαστήματα μπορούν να προσδιοριστούν ως

$$\mathbf{A}^l = \left\{ \mathbf{a}^l, \delta\alpha \right\} = \left\{ [a_{1,j_1}^l, a_{2,j_2}^l, \dots, a_{N,j_N}^l], \underbrace{[\delta\alpha, \delta\alpha, \dots, \delta\alpha]}_N \right\}, \quad l = 1, \dots, c^N \quad (4.2)$$

όπου το α_{i,j_i}^l είναι το κέντρο που έχει επιλεγεί από το μονοδιάστατο ασαφή σύνολο A_{i,j_i} που έχει προσδιοριστεί με βάση την είσοδο i . Κάθε ένα από τα σημεία που δημιουργήθηκαν, ενδέχεται να επιλεγεί ως κέντρο από το RBF νευρωνικό δίκτυο. Η επιλογή γίνεται με βάση την πολυδιάστατη συνάρτηση συμμετοχής $\mu_{\mathbf{A}^l}(\mathbf{u}_k)$.

$$\mu_{\mathbf{A}^l}(\mathbf{u}_k) = \begin{cases} 1 - r_l(\mathbf{u}_k), & \text{Αν } r_l(\mathbf{u}_k) \leq 1 \\ 0 & , \text{ Αλλιώς} \end{cases} \quad (4.3)$$

Στην παραπάνω σχέση το $r_l(\mathbf{u}_k)$ αναπαριστά την σχετική ευκλείδεια απόσταση μεταξύ του εκάστοτε σημείου \mathbf{A}^l και του διανύσματος εισόδου \mathbf{u}_k .

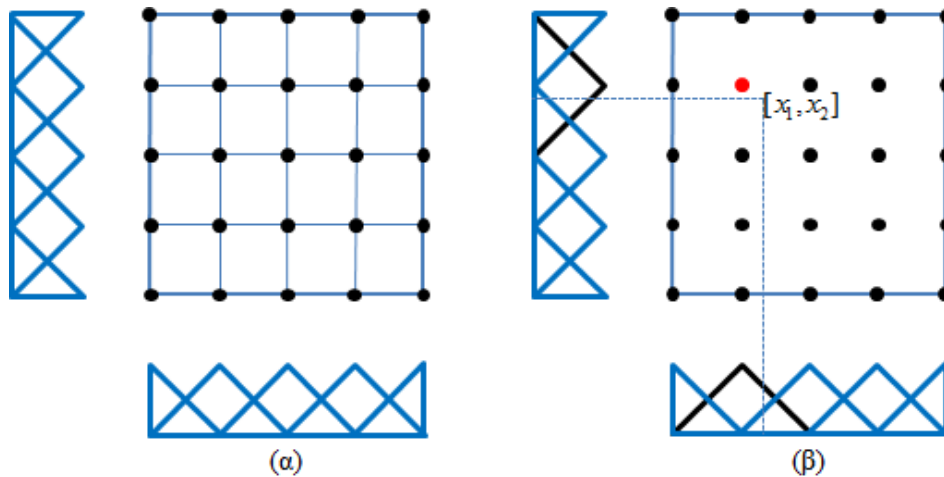
$$r_l(\mathbf{u}_k) = \frac{\sqrt{\sum_{i=1}^N (a_{i,j_i}^l - u_{i,k})^2}}{\delta\alpha\sqrt{N}} \quad (4.4)$$

Για την επιλογή των κέντρων του νευρωνικού δικτύου RBF, ο αλγόριθμος ακολουθεί την παρακάτω διαδικασία. Με την εισαγωγή στο δίκτυο του πρώτου διανύσματος εισόδου, επιλέγεται το κέντρο που ικανοποιεί την Εξ. 4.4 σε μεγαλύτερο βαθμό. Στην συνέχεια όσα δεδομένα ικανοποιούν την πρώτη συνθήκη της Εξ. 4.4 υπάγονται στο ίδιο κέντρο με το πρώτο δεδομένο Σχ. 4.2. Σε διαφορετική περίπτωση επιλέγεται ένα από τα υπόλοιπα (αυτό που είναι πιο κοντά στο καινούριο δεδομένο). Στο τέλος της διαδικασίας εκπαίδευσης, δεν υπάρχει κανένα από τα δεδομένα που να μην ανήκει σε κάποιο από τα κέντρα που έχουν δημιουργηθεί.

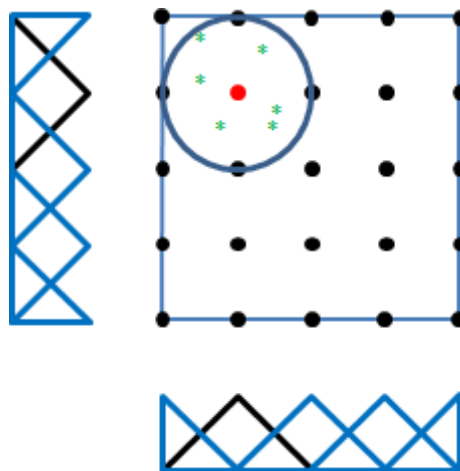
Στην συνέχεια, αφού ο αλγόριθμος ασαφών μέσω έχει δημιουργήσει τα κέντρα του δικτύου RBF, ακολουθεί ο υπολογισμός των συναπτικών βαρών, ο οποίος επιτυγχάνεται χρησιμοποιώντας έναν γραμμικό προσδιορισμό μεταξύ της εξόδου που παράγει το νευρωνικό δίκτυο, και της πραγματικής τιμής που αυτή έχει. Για την σωστή επίλυση του προβλήματος αυτού, χρησιμοποιείται η μέθοδος των ελαχίστων τετραγώνων σε μορφή πινάκων όπως παρουσιάζεται στην Εξ. 4.5.

$$\mathbf{W} = \mathbf{Y}^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \tag{4.5}$$

Το \mathbf{W} αντιπροσωπεύει τα συναπτικά βάρη, το \mathbf{Z} τις τιμές της στοιβάδας εξόδου στο δίκτυο RBF και το \mathbf{Y} , τις πραγματικές τιμές εξόδου.



Σχήμα 4.1: (α) Δημιουργία πιθανών κέντρων για νευρωνικό δίκτυο RBF δύο εισόδων (β) Επιλογή του κατάλληλου κέντρου

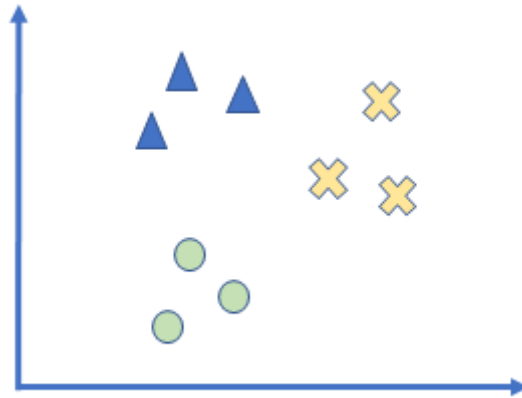


Σχήμα 4.2: Σημεία που υπάγονται στο ίδιο κέντρο

Αλγόριθμος κατηγοριοποίησης one vs all

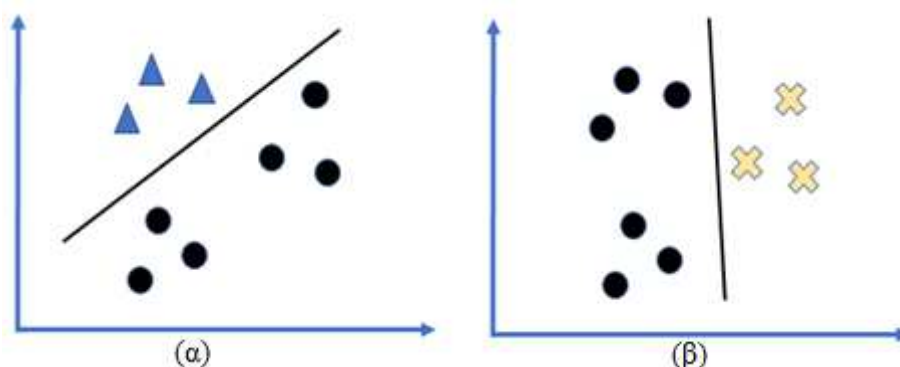
Ένας από τους πλέον διαδεδομένους τρόπους κατηγοριοποίησης είναι ο αλγόριθμος δυαδικής κατηγοριοποίησης (binary classification). Χρησιμοποιείται ευρέως για τον διαχωρισμό συνόλων με τον περιορισμό όμως πως αφορά σύνολα τα οποία διαχωρίζονται σε δύο κλάσεις. Σε εφαρμογές όπως αυτή που θα εξετάσουμε στην παρούσα πτυχιακή εργασία, στις οποίες απαιτείται ο διαχωρισμός περισσότερων συνόλων, η χρήση ενός αλγόριθμου όπως ο one vs all [7] κρίνεται απαραίτητη για την ολοκλήρωση της εφαρμογής.

Ο τρόπος λειτουργίας του αλγορίθμου είναι ο εξής. Έστω ότι έχουμε τρία (X, Δ, O) διαφορετικά σύνολα.



Σχήμα 4.3: Δεδομένα προς διαχωρισμό με την μέθοδο One vs all

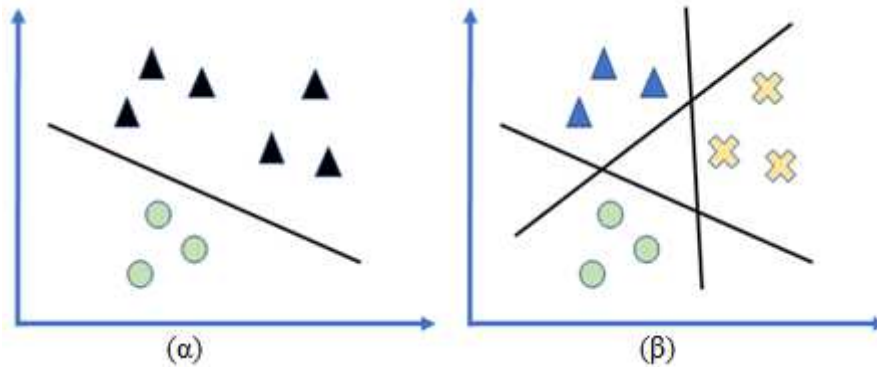
Προκειμένου να μπορέσουμε να κατηγοριοποιήσουμε τα δεδομένα λειτουργούμε σαν να έχουμε να επιλύσουμε ένα πρόβλημα δυαδικής κατηγοριοποίησης. Ορίζουμε σαν τιμή εξόδου την μονάδα (1) για τα δεδομένα με το (Δ) και θέτουμε την τιμή εξόδου ίση με το μηδέν (0) για τα X και O Σχ. 2.4 (α) δεδομένα και υπολογίζουμε την τιμή της ευθείας με την χρήση των συναπτικών βαρών ουσιαστικά υπολογίζουμε την τιμή της πιθανότητας εμφάνισης του $h_{\theta}^{(1)}(x)$. Στο επόμενο βήμα ορίζουμε την τιμή των δεδομένων εξόδου X ίση με το 1 Σχ.2.4(β) και τις τιμές των δεδομένων εξόδου για τα Δ και O ίση με το μηδέν υπολογίζοντας πάλι την πιθανότητα $h_{\theta}^{(2)}(x)$.



Σχήμα 4.4: (α) Διαχωρισμός δεδομένων ως προς Δ (β) Διαχωρισμός δεδομένων ως προς X

Ο τελευταίος συνδυασμός που έχουμε να εξετάσουμε είναι για τα δεδομένα εξόδου X στα οποία θα δώσουμε την τιμή 1 και για τα δεδομένα O και Δ ορίζουμε την τιμή εξόδου ίση με το μηδέν υπολογίζοντας την τιμή της πιθανότητας εμφάνισης $h_{\theta}^{(3)}(x)$. Ο συνδυασμός όλων αυτών μας δίνει το αποτέλεσμα που παρουσιάζεται στην παρακάτω. Έχοντας υπολογίσει λοιπόν τον πίνακα $h_{\theta}(x) = [h_{\theta}^{(1)}(x), h_{\theta}^{(2)}(x), h_{\theta}^{(3)}(x)]$ σαν τιμή πρόβλεψης ορίζουμε την μέγιστη τιμή

$\max(h_{\theta}^{(i)}(x))$. Σε ιδανικές συνθήκες το αποτέλεσμα που θα προκύψει μπορεί να αναπαρασταθεί από το Σχ 2.5 (β).

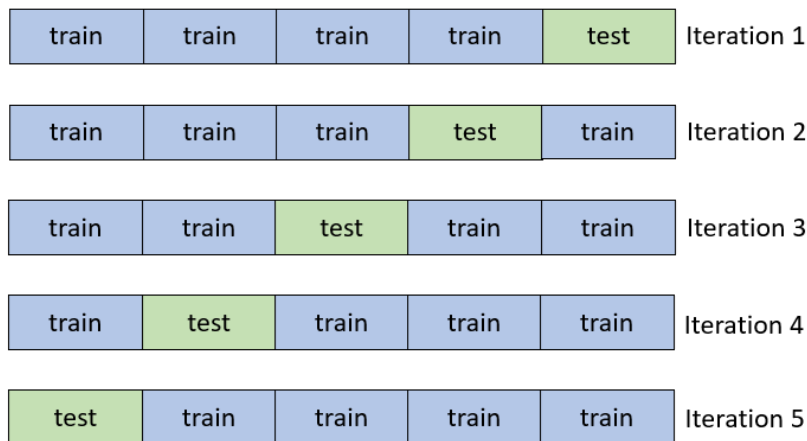


Σχήμα 4.5: (α) Διαχωρισμός δεδομένων ως προς 0 (β) Τελικό αποτέλεσμα διαχωρισμού των δεδομένων

Αλγόριθμος k-folds cross-validation

Η τεχνική του Cross-validation, χρησιμοποιείται για να εξάγουμε συμπεράσματα σχετικά με την αξιοπιστία του μοντέλου που δημιουργούμε. Η ανάγκη χρησιμοποίησης αυτής της τεχνικής έγκειται στο γεγονός πως πολλές φορές τα σωστά αποτελέσματα που παρουσιάζει το δίκτυο ενδεχόμενος να οφείλονται και σε τυχαία γεγονότα, για τον λόγο αυτό θέλουμε να έχουμε ένα δείγμα της συμπεριφοράς του μοντέλου σε όλο το πλήθος των δεδομένων

Συνήθως σε συστήματα πρόβλεψης δίνουμε ένα σύνολο από δεδομένα για την εκπαίδευση του δικτύου και ένα σύνολο δεδομένων το κρατάμε για την αξιολόγηση του συστήματος. Ο στόχος λοιπόν, είναι μπορέσουμε να έχουμε μια αξιολόγηση του συστήματος σε όλο το πλήθος των δεδομένων και να περιορίσουμε προβλήματα που παρουσιάζονται στα συστήματα πρόβλεψης όπως είναι το overfitting[35].



Σχήμα 4.6: Διαχωρισμός δεδομένων εκπαίδευσης και ελέγχου με τον αλγόριθμο k-folds

Επεξεργασία Δεδομένων

Ο συνδυασμός των παραπάνω αλγορίθμων δημιουργεί το μοντέλο πρόβλεψης που θα χρησιμοποιηθεί κατά την διάρκεια της πειραματικής δραστηριότητας.

Το σημαντικότερο βήμα κατά την δημιουργία ενός μοντέλου πρόβλεψης είναι η σωστή δημιουργία των δεδομένων εισόδου και εξόδου. Θα πρέπει τα δεδομένα που δίνονται στο δίκτυο να αντιπροσωπεύουν την συμπεριφορά του αντικείμενου, με αποτέλεσμα η διαμόρφωση των μεταβλητών του δικτύου να αναπαριστούν όσο γίνεται καλύτερα την συμπεριφορά του εκάστοτε αντικείμενου.

Για τις ανάγκες της εργασίας έχουν δοθεί τα δεδομένα που έχει καταγράψει η κεραία του Πανεπιστημίου Πειραιώς και αφορούν τις θέσεις των πλοίων στην ευρύτερη περιοχή του Αργοσαρωνικού, το χρονικό διάστημα λήψης του σήματος καθώς και το αναγνωριστικό mmsi για το κάθε πλοίο.

Για την σωστή εκπαίδευση και δημιουργία του δικτύου, θα πρέπει όπως είπαμε και προηγουμένως, τα δεδομένα που δίνονται σε αυτό να αναπαριστούν όσο πιο ευκρινώς γίνεται την πληροφορία που χαρακτηρίζει το κάθε αντικείμενο. Με αυτόν τον τρόπο επιτυγχάνεται η καλύτερη δυνατή ομαδοποίηση των δεδομένων. Αυτός είναι και ο βασικός λόγος που σε προβλήματα κατηγοριοποίησης επιλέγονται δεδομένα εισόδου που να αναπαριστούν πολλά χαρακτηριστικά του εκάστοτε αντικείμενου, δηλαδή θέλουμε το n να είναι όσο το δυνατόν μεγαλύτερο Εξ 4.6.

$$X = [x_1, x_2, \dots, x_n] \quad (4.6)$$

Το μεγαλύτερο πρόβλημα που καλούμαστε να αντιμετωπίσουμε είναι ο περιορισμένος αριθμός δεδομένων. Όπως έχει ήδη αναφερθεί τα διαθέσιμα δεδομένα είναι το γεωγραφικό μήκος, το γεωγραφικό πλάτος καθώς και χρονική στιγμή λήψης του αντικείμενου. Το γεγονός αυτό μας περιορίζει ως προς την επιλογή των δεδομένων εισόδου.

Το πρώτο βήμα αφορά την συλλογή των δεδομένων. Στον πίνακα 4.1 παρουσιάζεται ο διαχωρισμός των δεδομένων με βάση τον τύπο του κάθε πλοίου. Επίσης, φαίνεται και ο αριθμός των δεδομένων που υπάρχουν σε κάθε κατηγορία, το σύνολο που έχουμε συλλέξει ανέρχεται σε 59.632.

το πρώτο που μπορούμε να υπολογίσουμε είναι η μετατόπιση του κάθε πλοίου με βάση την Εξ. 4,7 που υπολογίζει την μετατόπιση του κάθε πλοίου συνυπολογίζοντας και την παράμετρο για την καμπυλότητα της γης. Ο υπολογισμός έχει γίνει με την χρήση της συνάρτησης *distance* του Matlab δίνοντας επίσης και την τιμή της ακτίνας της γης σε μέτρα.

$$m = \frac{x_1 - x_2}{t_1 - t_2} \quad (4.7)$$

Στην συνέχεια μπορούμε να υπολογίσουμε την έχοντας υπολογίσει την τιμή της μετατόπισης μπορούμε να υπολογίσουμε την ταχύτητα του αντικείμενου με βάση την Εξ. 4.8

$$u_{1,2} = \frac{m_2 - m_1}{t_2 - t_1} \quad (4.8)$$

Το τελευταίο δεδομένο εισόδου που μπορούμε να δημιουργήσουμε αφορά την επιτάχυνση Εξ 4.9.

$$a = \frac{u_{2,3} - u_{1,2}}{\Delta t} \quad (4.9)$$

Με βάση τα παραπάνω θα επιχειρήσουμε μια πρόβλεψη με την χρήση του μοντέλου που δημιουργήσαμε, αρχικά με όλα τα δεδομένα χωρίς καμία επεξεργασία προκειμένου να δούμε την

αποτελεσματικότητα του συστήματος, στην συνέχεια και σε περίπτωση που τα αποτελέσματα δεν είναι ικανοποιητικά, θα επιχειρήσουμε την επεξεργασία των δεδομένων με σκοπό αυτά να εμπεριέχουν όσο το δυνατόν καλύτερη πληροφορία για το μοντέλο. Τα μη επεξεργασμένα δεδομένα ανέρχονται σε 62.630.

Τύπος Πλοίου	Όνομα Πλοίου	mmsi Πλοίου	Κλάση	#Δεδομένων
Ρομουλκό	IFESTOS 2	237070600	1	25818
	ALEXANDER 5	239634000		
	ARMADORES II [GR]	241108000		
Πετρελαιοφόρο	SIROS [GR]	237017200	2	9922
	MENALON [GR]	239741000		
	EKO 5 [GR]	240911000		
Επιβατικό	TERA JET [CY]	212260000	3	18796
	EL.VENIZELOS [GR]	237628000		
	HELLENIC SPIRIT [GR]	239806000		
Flying Dolphins	FLYING DOLPHIN XVII [GR]	237023600	4	3917
	FLYING DOLPHIN XXIX [GR]	237024500		
Μεταφοράς Τσιμέντου	ADAMAS [GR]	240544000	5	2502
	EVIACEMENT IV [GR]	241374000		
Φορτηγό Πλοίο	MAX VALUE [MT]	249095000	6	1675

Πίνακας 4.1:Κατηγοριοποίηση πλοίων χωρίς επεξεργασία

Κεφάλαιο 5

Πειραματική δραστηριότητα

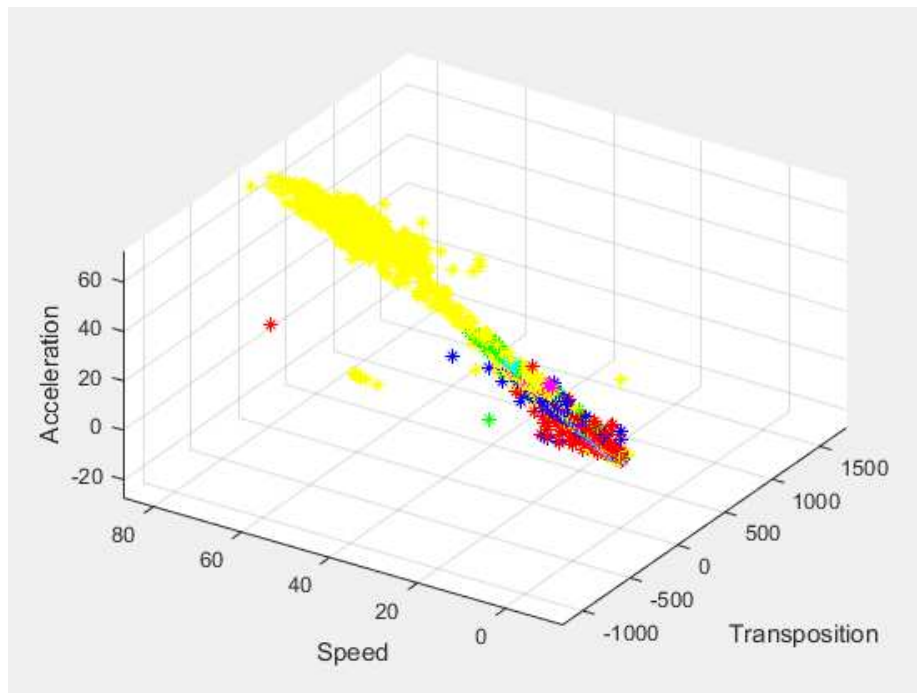
Δεδομένα χωρίς επεξεργασία

Το πρώτο βήμα, αφορά την εκπαίδευση του μοντέλου που έχει δημιουργηθεί με τα 59000 δεδομένα. Τα αποτελέσματα παρουσιάζονται με την χρήση του confusion matrix, έτσι ώστε να παρέχεται μια καλύτερη εικόνα για την αποτελεσματικότητα του μοντέλου.

Output Class	1	2	3	4	5	6	
1	1758 43.1%	662 16.2%	1296 31.7%	283 6.9%	50 1.2%	34 0.8%	43.1% 56.9%
2	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
3	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
4	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	100% 0.0%	0.0% 100%	0.0% 100%	0.0% 100%	0.0% 100%	0.0% 100%	43.1% 56.9%
	1	2	3	4	5	6	
	Target Class						

Σχήμα 5.1: Confusion Matrix για δεδομένα χωρίς επεξεργασία

Με βάση το Σχ 5.1 διακρίνουμε πως το συνολικό ποσοστό επιτυχίας του μοντέλου φτάνει το 43.1%. Εξετάζοντας το παραπάνω αποτέλεσμα παρατηρείται πως το δίκτυο δεν έχει καταφέρει να προσομοιώσει σωστά την συσχέτιση των αποτελεσμάτων. Όλες οι προβλέψεις που δίνει αφορούν την πρώτη κατηγορία δεδομένων. Ο λόγος της αποτυχίας του μοντέλου βρίσκεται στον θόρυβο που έχει εισχωρήσει στα δεδομένα μη επιτρέποντας στο δίκτυο να δημιουργήσει σωστή συσχέτιση μεταξύ των δεδομένων εισόδου και εξόδου. Όπως μπορούμε να δούμε στο Σχ. 5.2 υπάρχουν πολλά δεδομένα που βρίσκονται κοντά στο μηδέν (0) και ενδεχομένως να αναπαριστούν πλοία τα οποία είναι σταματημένα στο λιμάνι, επίσης παρατηρούνται δεδομένα τα οποία έχουν μεγάλες τιμές οι οποίες οφείλονται είτε σε λάθος καταγραφή είτε είναι κάποιο πλοίο το οποίο έχει μπει στο πεδίο λήψης της κεραίας μετά από αρκετό χρονικό διάστημα.



Σχήμα 5.2: Διασπορά δεδομένων χωρίς επεξεργασία

Επεξεργασμένα δεδομένα

Προκειμένου να δημιουργηθεί σωστά το μοντέλο πρέπει τα δεδομένα που αποτελούν τον θόρυβο να αποκοπούν και να μην δοθούν στο δίκτυο. Πιο συγκεκριμένα, δεν θα συμπεριληφθούν στα δεδομένα όσα έχουν τιμές ταχύτητας χαμηλότερες από τα 2km/h, καθώς και όσα δεδομένα έχουν τιμή ταχύτητας μεγαλύτερη από τα 100km/h, από την στιγμή που έχει παρατηρηθεί πως αυτή η τιμή ταχύτητας έχει καταγραφεί σε ελάχιστα πλοία. Το σύνολο των δεδομένων που θα αφαιρεθούν ξεπερνά τις 50000.

Το νέο σύνολο δεδομένα που θα δοθεί εκ νέου στο νευρωνικό δίκτυο αποτυπώνεται στον πίνακα 5.1. Τα αποτελέσματα παρουσιάζονται στο Σχ. 5.3 και είναι προφανώς πολύ καλύτερα σε σχέση με την προηγούμενη περίπτωση, μιας και το ποσοστό επιτυχούς πρόβλεψης φτάνει το 77%. Όπως προκύπτει, λοιπόν, η παραμετροποίηση των δεδομένων που πραγματοποιήθηκε στο προηγούμενο βήμα, βοήθησε στην καλύτερη αναπαράσταση της πληροφορίας μεταξύ των δεδομένων εισόδου και εξόδου. Παρόλη την σαφή βελτίωση του μοντέλου, δεν υπάρχει ακόμα το επιθυμητό αποτέλεσμα που ορίστηκε στις απαιτήσεις του συστήματος.

Εστιάζοντας στα αποτελέσματα του Σχ. 5.3, παρατηρείται μια δυστοκία του μοντέλου στον διαχωρισμό της πρώτης κλάσης με την δεύτερη, καθώς επίσης και της πέμπτης κλάσης με την έκτη. Πιο συγκεκριμένα, πολλά αποτελέσματα (36 δεδομένα) που ανήκουν στην δεύτερη κλάση κατηγοριοποιούνται λανθασμένα στην πρώτη, όπως επίσης και είκοσι (20 δεδομένα) που ανήκουν στην Πέμπτη κλάση κατηγοριοποιούνται λανθασμένα στην έκτη.

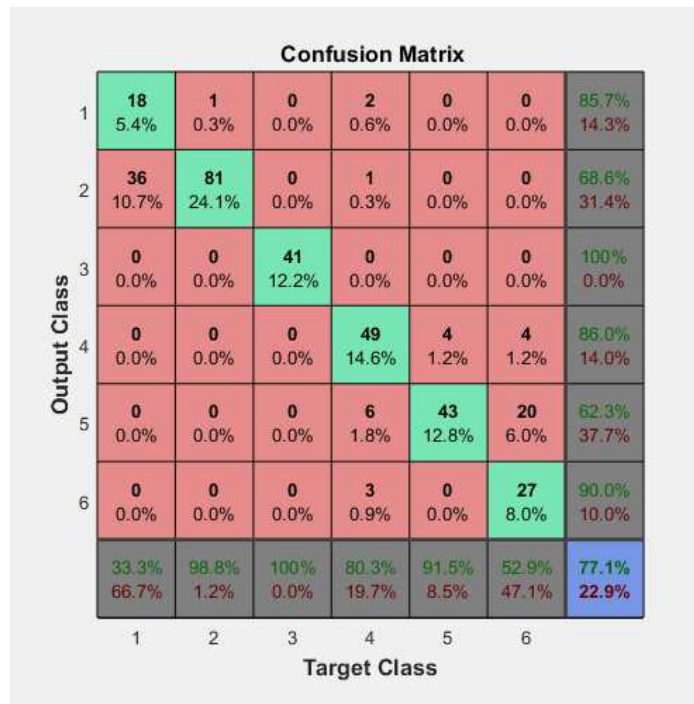
Ο λόγος που παρατηρείται αυτή η συμπεριφορά έχει να κάνει με την παρόμοια συμπεριφορά που παρουσιάζουν στην κίνηση τα πλοία που ανήκουν στην πρώτη κλάση των ρυμουλκών με την δεύτερη κλάση των πετρελαιοφόρων όπως επίσης παρόμοια συμπεριφορά κίνησης παρουσιάζουν και τα πλοία μεταφοράς τσιμέντου που ανήκουν στην πέμπτη κλάση με τα φορτηγά πλοία που ανήκουν στην έκτη. Η παρόμοια συμπεριφορά τους αφορά την ταχύτητα κίνησης τους καθώς επίσης και την επιτάχυνση τους από την στιγμή που πρόκειται για πλοία με παραπλήσιο μέγεθος ο τρόπος που επιταχύνουν ή επιβραδύνουν είναι παρόμοιος.

Τύπος Πλοίου	Όνομα Πλοίου	mmsi Πλοίου	Κλάση	#Δεδομένων
Ρυμουλκό	IFESTOS 2	237070600	1	520
	ALEXANDER 5	239634000		
	ARMADORES II [GR]	241108000		
Πετρελαιοφόρο	SIROS [GR]	237017200	2	693
	MENALON [GR]	239741000		
	EKO 5 [GR]	240911000		
Επιβατικό	TERA JET [CY]	212260000	3	234
	EL.VENIZELOS [GR]	237628000		
	HELLENIC SPIRIT [GR]	239806000		
Flying Dolphins	FLYING DOLPHIN XVII [GR]	237023600	4	367
	FLYING DOLPHIN XXIX [GR]	237024500		
Μεταφοράς Τσιμέντου	ADAMAS [GR]	240544000	5	550
	EVIACEMENT IV [GR]	241374000		
Φορτηγό Πλοίο	MAX VALUE [MT]	249095000	6	500

Πίνακας 4.1:Κατηγοριοποίηση πλοίων μετά από επεξεργασία

Εστιάζοντας στα αποτελέσματα του Σχ. 5.3, παρατηρείται μια δυστοκία του μοντέλου στον διαχωρισμό της πρώτης κλάσης με την δεύτερη, καθώς επίσης και της πέμπτης κλάσης με την έκτη. Πιο συγκεκριμένα, πολλά αποτελέσματα (36 δεδομένα) που ανήκουν στην δεύτερη κλάση κατηγοριοποιούνται λανθασμένα στην πρώτη, όπως επίσης και είκοσι (20 δεδομένα) που ανήκουν στην Πέμπτη κλάση κατηγοριοποιούνται λανθασμένα στην έκτη.

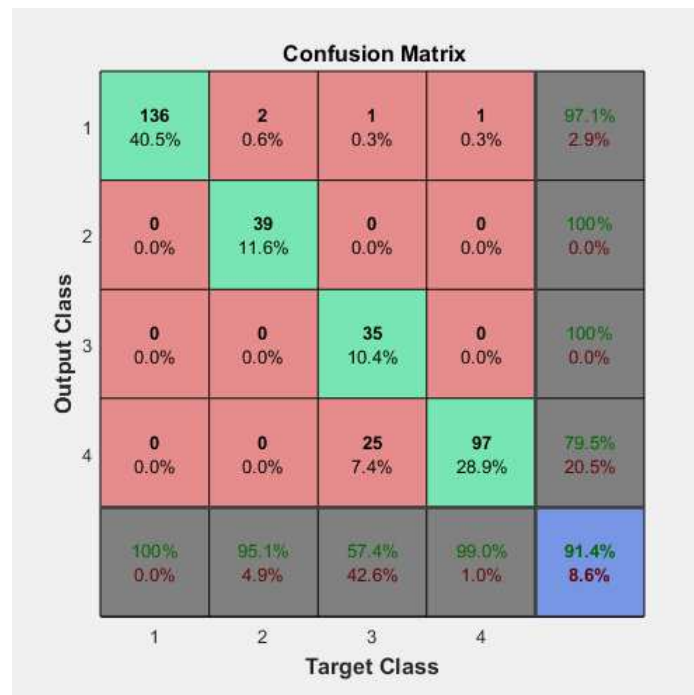
Ο λόγος που παρατηρείται αυτή η συμπεριφορά έχει να κάνει με την παρόμοια συμπεριφορά που παρουσιάζουν στην κίνηση τα πλοία που ανήκουν στην πρώτη κλάση των ρυμουλκών με την δεύτερη κλάση των πετρελαιοφόρων όπως επίσης παρόμοια συμπεριφορά κίνησης παρουσιάζουν και τα πλοία μεταφοράς τσιμέντου που ανήκουν στην πέμπτη κλάση με τα φορτηγά πλοία που ανήκουν στην έκτη. Η παρόμοια συμπεριφορά τους αφορά την ταχύτητα κίνησης τους καθώς επίσης και την επιτάχυνση τους από την στιγμή που πρόκειται για πλοία με παραπλήσιο μέγεθος ο τρόπος που επιταχύνουν ή επιβραδύνουν είναι παρόμοιος.



Σχήμα 5.3: Confusion Matrix μοντέλου για επεξεργασμένα δεδομένα

Συνεπώς, για να βελτιωθεί η συμπεριφορά του μοντέλου θα ενοποιήσουμε την πρώτη κλάση με την δεύτερη και την πέμπτη κλάση με την έκτη. Πλέον τα δεδομένα θα διαμορφωθούν σύμφωνα με τον πίνακα 5.2 σε τέσσερις κλάσεις. Ο αριθμός των δεδομένων σε κάθε κλάση είναι διαφορετικός γεγονός που ενισχύει τον παράγοντα της τυχαιότητας των δεδομένων.

Τα νέα δεδομένα θα δοθούν στο δίκτυο με σκοπό την βελτίωση της απόδοσης του η οποία παρατηρείται στο Σχ. 5.4. Το ποσοστό επιτυχίας του νευρωνικού δικτύου έχει ανέβει στο 91% κάνοντας σωστή πρόβλεψη σε όλο το εύρος των δεδομένων.



Σχήμα 5.4: Confusion Matrix μοντέλου για δεδομένα τεσσάρων κλάσεων

Παραμετροποίηση δεδομένων εισόδου

Η επιλογή των παραμέτρων εισόδου είναι πολύ σημαντική για την δημιουργία του δικτύου. Με βάση την θεωρία, το ιδανικό σενάριο για τις παραμέτρους των δεδομένων εισόδου σε ένα νευρωνικό δίκτυο είναι πως η κάθε μια από τις παραμέτρους θα πρέπει να έχει όσο το δυνατόν μικρότερη συσχέτιση με κάποια από τις υπόλοιπες. Ο λόγος είναι πως θέλουμε η κάθε παράμετρο να έχει όσο το δυνατόν μεγαλύτερη συμμετοχή στην δημιουργία του δικτύου κάτι που αποτελεί ανασταλτικό παράγοντα σε περιπτώσεις που υπάρχει συσχετισμός παραμέτρων.

Υπάρχουν όμως πολλές περιπτώσεις που η επιλογή των παραμέτρων είναι πολύ περιορισμένη, όπως συμβαίνει στην περίπτωση που εξετάζουμε. Όπως έχει αναφερθεί, τα δεδομένα που έχουμε στην διάθεση μας είναι ο προσδιορισμός της θέσης του εκάστοτε αντικειμένου καθώς και το χρονικό διάστημα λήψης του στιγμιότυπου. Επομένως, οι διαθέσιμες παράμετροι εισόδου είναι η μετατόπιση, η ταχύτητα και η επιτάχυνση. Η επιτάχυνση δεν παρουσιάζει μεγάλη συσχέτιση με την ταχύτητα και την μετατόπιση δεδομένου ότι επηρεάζεται ο τρόπος επιτάχυνσης και επιβράδυνσης του πλοίου από το μέγεθος και το βάρος του πλοίου. Η ταχύτητα όμως και η μετατόπιση παρουσιάζουν μεγάλη εξάρτηση. Επομένως, θα επιχειρήσουμε να αφαιρέσουμε την μετατόπιση και να δώσουμε σαν δεδομένα στο δίκτυο μόνο την ταχύτητα και την επιτάχυνση.

	1	2	3	4	
1	119 35.4%	0 0.0%	22 6.5%	43 12.8%	64.7% 35.3%
2	0 0.0%	35 10.4%	10 3.0%	0 0.0%	77.8% 22.2%
3	0 0.0%	0 0.0%	2 0.6%	0 0.0%	100% 0.0%
4	17 5.1%	6 1.8%	27 8.0%	55 16.4%	52.4% 47.6%
	87.5% 12.5%	85.4% 14.6%	3.3% 96.7%	56.1% 43.9%	62.8% 37.2%
	1	2	3	4	

Σχήμα 5.5: Αποτελέσματα μοντέλου με παραμέτρους εισόδου την ταχύτητα και την επιτάχυνση

Τα αποτελέσματα παρουσιάζονται στο Σχ. 5.5 και παρατηρείται πως το μοντέλο επιχειρεί, και σε πολλές περιπτώσεις τα καταφέρνει, πρόβλεψη σε όλες τις διαθέσιμες κατηγοριοποιήσεις εντούτοις όμως το συνολικό αποτέλεσμα έχει μειωθεί σημαντικά φτάνοντας στο 62,8%, τιμή η οποία είναι κατά πολύ χαμηλότερη σε σχέση με το ποσοστό επιτυχίας που είχε το μοντέλο και με τις τρεις εισόδους. Σαν επόμενο βήμα, θα δοθούν σαν παράμετροι εισόδου η μετατόπιση και η επιτάχυνση.

Confusion Matrix

1	128 38.1%	25 7.4%	7 2.1%	2 0.6%	79.0% 21.0%
2	2 0.6%	16 4.8%	8 2.4%	0 0.0%	61.5% 38.5%
3	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
4	6 1.8%	0 0.0%	46 13.7%	96 28.6%	64.9% 35.1%
	94.1% 5.9%	39.0% 61.0%	0.0% 100%	98.0% 2.0%	71.4% 28.6%
	1	2	3	4	
	Target Class				

Σχήμα 5.6: Αποτελέσματα μοντέλου με παραμέτρους εισόδου την μετατόπιση και την επιτάχυνση

Στο Σχ. 5.6 παρουσιάζεται το confusion matrix που προέκυψε δίνοντας στις παραμέτρους εισόδου την μετατόπιση και την επιτάχυνση. Όπως παρατηρήθηκε και στην προηγούμενη περίπτωση έτσι και εδώ τα αποτελέσματα είναι χειρότερα, φτάνοντας το 71%, απ' ότι με το μοντέλο με τις τρεις παραμέτρους εισόδου.

Πρέπει να τονιστεί πως και στις τρεις περιπτώσεις τα δεδομένα εκπαίδευσης και αξιολόγησης είναι τα ίδια και το μόνο που μεταβάλλεται είναι οι παράμετροι εισόδου.

Συμπερασματικά, παρατηρείται πως οι τρεις παράμετροι εισόδου είναι απαραίτητοι. Σε υλοποιήσεις που το πλήθος των παραμέτρων είναι περιορισμένο, όπως στην εφαρμογή που εξετάζουμε, είναι σημαντικές ακόμα και οι παράμετροι που παρουσιάζουν συσχετίσεις μεταξύ τους.

Έλεγχος ευρωστίας (Robustness)

Ο έλεγχος της ευρωστίας του συστήματος είναι ιδιαίτερα σημαντικός στην πορεία ανάπτυξης ενός μοντέλου πρόβλεψης. Τα σωστά αποτελέσματα που παρουσιάζει το εκάστοτε μοντέλο μπορεί να οφείλονται σε τυχαίους παράγοντες και όχι στην σωστή συσχέτιση που έχει δημιουργήσει το δίκτυο μεταξύ των δεδομένων εισόδου και εξόδου. Για να εξαλείψουμε τον παράγοντα της τυχαίας πρόβλεψης πρέπει να παραμετροποιηθούν τα δεδομένα που δίνουμε σαν είσοδο στο δίκτυο και στην συνέχεια να επιχειρήσουμε μια πρόβλεψη εκ νέου.

Στη μέχρι τώρα μελέτη για την εκπαίδευση του μοντέλου δίνονταν το 90% των διαθέσιμων δεδομένων και για τον έλεγχο του το 10%. Πλέον τα διαθέσιμα δεδομένα εκπαίδευσης θα ανακατευτούν εκ νέου και σαν δεδομένα εκπαίδευσης θα χρησιμοποιηθεί το 80% ενώ το ποσοτό του συνόλου αξιολόγησης θα φτάσει το 20%. Στο Σχ 5.7 παρουσιάζονται τα αποτελέσματα προέκυψαν από την παραπάνω διαδικασία και το σύστημα παρατηρούμε ότι αποκρίνεται εξαιρετικά σε αυτή την τροποποίηση των δεδομένων με ποσοστό επιτυχούς πρόβλεψης το 90% στα νέα δεδομένα αξιολόγησης.

Επόμενο βήμα είναι να μειωθεί το σύνολο των δεδομένων εκπαίδευσης τόσο ώστε η αξιολόγηση του συστήματος να πέσει κάτω από το 85%. Αυτό το αποτέλεσμα επιτυγχάνεται για ποσοστό εκπαίδευσης που αγγίζει το 40% και αξιολόγησης στο 60% Σχ 5.8.

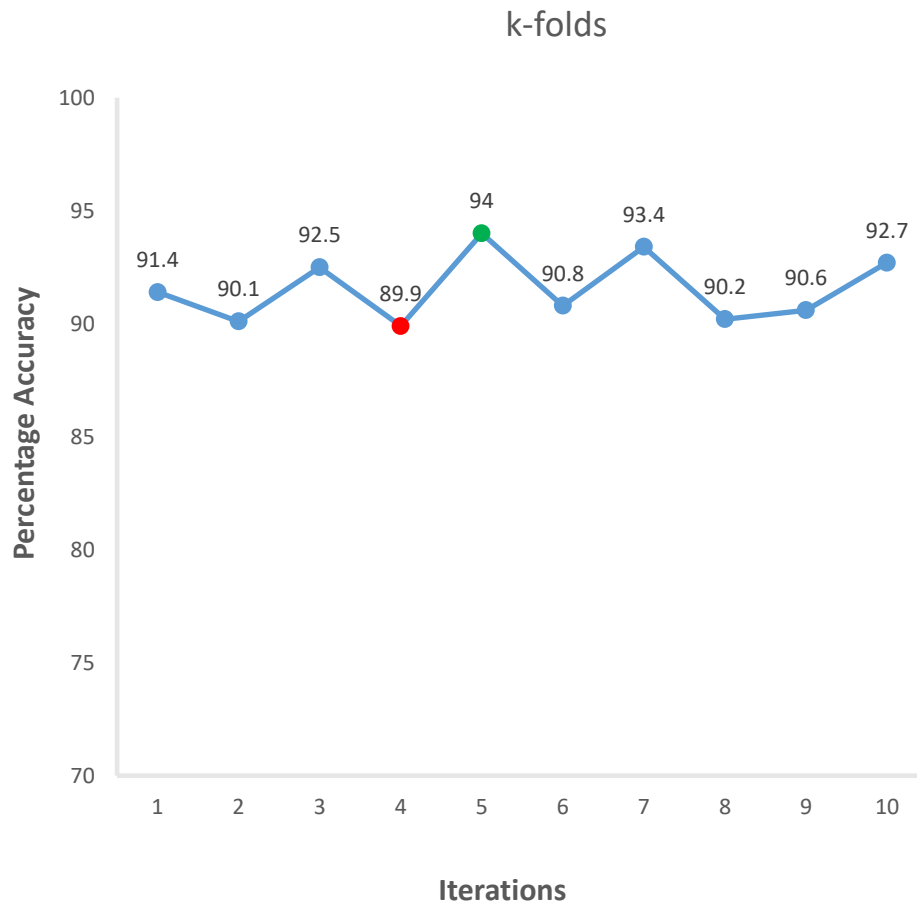
Output Class	1	2	3	4	
1	262 38.9%	2 0.3%	1 0.1%	6 0.9%	96.7% 3.3%
2	0 0.0%	69 10.3%	1 0.1%	0 0.0%	98.6% 1.4%
3	1 0.1%	0 0.0%	61 9.1%	0 0.0%	98.4% 1.6%
4	1 0.1%	0 0.0%	55 8.2%	214 31.8%	79.3% 20.7%
	99.2% 0.8%	97.2% 2.8%	51.7% 48.3%	97.3% 2.7%	90.0% 10.0%
	1	2	3	4	
	Target Class				

Σχήμα 5.7: Αποτελέσματα για δεδομένα εκπαίδευσης 80% και δεδομένα ελέγχου 20%

Output Class	1	2	3	4	
1	914 38.8%	71 3.0%	1 0.0%	23 1.0%	90.6% 9.4%
2	60 2.5%	178 7.6%	3 0.1%	0 0.0%	73.9% 26.1%
3	3 0.1%	0 0.0%	232 9.9%	4 0.2%	97.1% 2.9%
4	32 1.4%	0 0.0%	148 6.3%	686 29.1%	79.2% 20.8%
	90.6% 9.4%	71.5% 28.5%	60.4% 39.6%	96.2% 3.8%	85.4% 14.6%
	1	2	3	4	
	Target Class				

Σχήμα 5.8: Αποτελέσματα για δεδομένα εκπαίδευσης 40% και δεδομένα ελέγχου 60%

Το τελευταίο στάδιο για την αξιολόγηση του δικτύου είναι η υλοποίηση του αλγόριθμου k-folds πάνω στα δεδομένα. Για την υλοποίηση του cross-validation αλγόριθμου θα διαχωρίσουμε τα διαθέσιμα δεδομένα σε δέκα υποκατηγορίες, από τις οποίες θα δίνουμε για την εκπαίδευση του μοντέλου το 90% και για την αξιολόγηση του το υπολειπόμενο 10%. Η διαδικασία αυτή θα επαναληφθεί 10 φορές όσα και τα διαθέσιμα υποσύνολα.

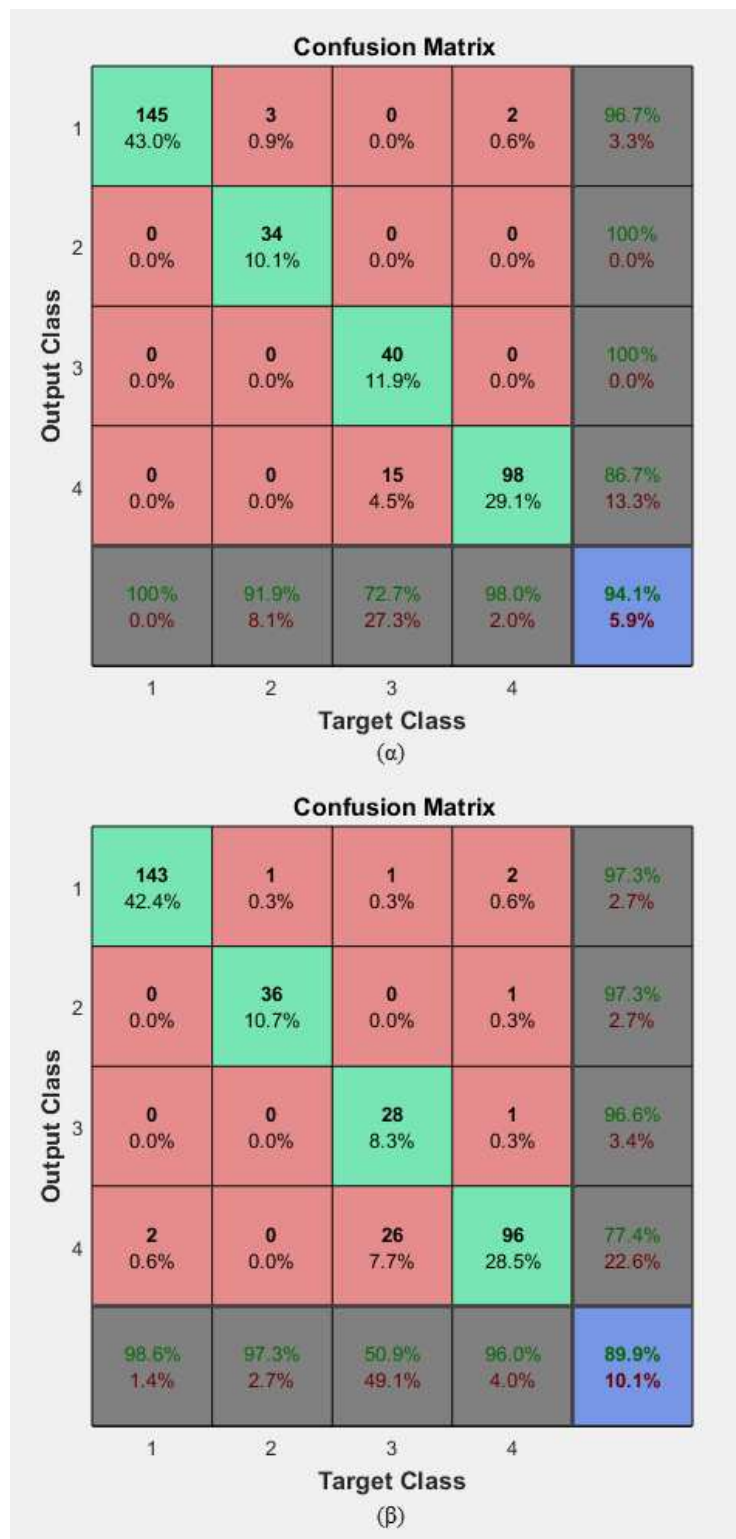


Σχήμα 5.9: Αποτελέσματα από τον k-folds αλγόριθμο

Τα αποτελέσματα που παρουσιάζονται στο Σχ. 5.9 παρουσιάζουν την αποδοτικότητα του συστήματος για όλες τις επαναλήψεις που επιχειρήθηκαν κατά την διάρκεια της αξιολόγησης. Το σύστημα παρουσιάζει πολύ καλή συμπεριφορά σε όλο το εύρος των δεδομένων και πιο συγκεκριμένα το χειρότερο ποσοστό επιτυχίας αφορά την τέταρτη επανάληψη με ποσοστό ακρίβειας 89.8%, ενώ το καλύτερο ποσοστό επιτυγχάνεται στην Πέμπτη επανάληψη και η τιμή ακρίβειας φτάνει το 94%.

Η συμπεριφορά του συστήματος στον αλγόριθμο k-folds και ο συνδυασμός της με τις προηγούμενες τεχνικές ελέγχου της ευρωστίας, φανερώνει πως το νευρωνικό δίκτυο έχει καταφέρει να δημιουργήσει επιτυχώς μια συσχέτιση μεταξύ των δεδομένων εισόδου και εξόδου, επιτυγχάνοντας υψηλό ποσοστό αποτελεσματικότητας. Ιδιαίτερα με την χρήση του k-folds αλγόριθμου εξαλείφεται η πιθανότητα τα αποτελέσματα που προέκυψαν στα προηγούμενα στάδια να οφείλονταν σε τυχαία λήψη δειγμάτων από την στιγμή που το σύστημα παρουσιάζει εξαιρετική συμπεριφορά στην πρόβλεψη του συνόλου των δεδομένων.

Τα αποτελέσματα το βέλτιστου και χείριστου μοντέλου παρουσιάζονται με την χρήση του confusion matrix στο Σχ. 5.10



Σχήμα 5.10: (α)Αποτελέσματα καλύτερου μοντέλου, (β)Αποτελέσματα χειρότερου μοντέλου.

Κεφάλαιο 6

Συμπεράσματα

Στα πλαίσια της διπλωματικής εργασίας δημιουργήθηκε ένα μοντέλο πρόβλεψης βασιζόμενο στην RBF (radial basis Function) αρχιτεκτονική των νευρωνικών δικτύων. Ο αλγόριθμος που χρησιμοποιήθηκε ονομάζεται Fuzzy means (αλγόριθμος ασαφών μέσων). Πρόκειται για ένα μοντέλο το οποίο ανήκει στα δίκτυα με επιβλεπόμενη μάθηση.

Ο στόχος της εργασίας είναι η κατηγοριοποίηση των πλοίων με βάση τον τρόπο κίνησης τους, λαμβάνοντας σαν είσοδο την μετατόπιση, την ταχύτητα και την επιτάχυνση του πλοίου την κάθε χρονική στιγμή και με βάση αυτά θα χρειαστεί να τα κατηγοριοποιήσουμε. Για την κατηγοριοποίηση τους χρησιμοποιήθηκε ο αλγόριθμος one vs all με τον οποίο επιτυγχάνεται η επέκταση του δυαδικού μοντέλου (Binary classification) κατηγοριοποίησης σε περισσότερες κλάσεις.

Αρχικά δίνονται στο δίκτυο τα δεδομένα χωρίς επεξεργασία, χωρίς όπως είναι αναμενόμενο να προκύψουν αξιόλογα αποτελέσματα, μιας και στα δεδομένα υπάρχει αρκετός θόρυβος. Για την βελτίωση των αποτελεσμάτων, θα χρειαστεί να αφαιρέσουμε από τα δεδομένα εκπαίδευσης όσα είναι κάτω από 2km/h και αντιστοιχούν σε δεδομένα πλοίων τα οποία κινούνται με ιδιαίτερα χαμηλή ταχύτητα με αποτέλεσμα να μην είναι διακριτή η συμπεριφορά του κάθε δεδομένου. Τα αποτελέσματα τα οποία προέκυψαν μετά την επεξεργασία των δεδομένων είναι πολύ καλύτερα καταδεικνύοντας πως το μοντέλο έχει καταφέρει να μάθει επιτυχώς την συσχέτιση μεταξύ των δεδομένων εισόδου και εξόδου.

Για την βελτίωση του μοντέλου και με βάση τα αποτελέσματα που προέκυψαν από το confusion matrix του συστήματος επιχειρήθηκε η ένωση τεσσάρων κλάσεων, με αποτέλεσμα από τις 6 κλάσεις που υπήρχαν αρχικά να διατηρηθούν μόνο 4 από αυτές. Η συνένωση δεν προέκυψε αυθαίρετα αλλά με βάση το κατά πόσο το σύστημα παρουσιάζει μια σύγχυση στις κλάσεις που έχουν ομαδοποιηθεί και αντιπροσωπεύουν συναφείς κατηγορίες πλοίων.

Μετά και την ομαδοποίηση των δεδομένων τα αποτελέσματα είναι πολύ καλύτερα και το αποτέλεσμα επιτυχούς πρόβλεψης φτάνει το 91%. Τα αποτελέσματα στα οποία παρουσιάζει μια σύγχυση το σύστημα αφορούν δεδομένα της τέταρτης κλάσης (πετρελαιοφόρα/φορτηγά πλοία) και της τρίτης κλάσης (Flying dolphins). Το γεγονός αυτό οφείλεται σε δεδομένα που έχουν ληφθεί από την περιοχή του λιμανιού στην οποία τα flying dolphins δεν έχουν καταφέρει ακόμα να αναπτύξουν την ταχύτητα που χαρακτηρίζει την συμπεριφορά τους στην θάλασσα.

Για να διαπιστωθεί η ευρωστία του συστήματος, δοκιμάστηκε η εκπαίδευση και αξιολόγησή του σε διαφορετικό κομμάτι δεδομένων. Το σύστημα απέδειξε πως παρά την μεταβολή των χαρακτηριστικών του δεν επηρεάστηκε και τα αποτελέσματά του κυμάνθηκαν σε ιδιαίτερα υψηλό επίπεδο, δίνοντας την εικόνα ενός επιτυχημένου συστήματος.

Για την περαιτέρω βελτίωση των αποτελεσμάτων του συστήματος μπορούν να προστεθούν στο μοντέλο οι καιρικές συνθήκες που επικρατούν την χρονική στιγμή λήψης των δεδομένων. Επίσης, για την βελτίωση του μοντέλου θα πρέπει να δεδομένα που λαμβάνονται να μην επηρεάζονται σε μεγάλο βαθμό από την περιοχή του λιμανιού αλλά να είναι δεδομένα που αφορούν την συμπεριφορά του μοντέλου στα ανοικτά της θάλασσάς.

Βιβλιογραφία

- [1] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2009.
- [2] A. Kusiak, X. Wei, A. P. Verma, and E. Roz, "Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, pp. 2337-2342, 2013.
- [3] K. Ristovski, S. Vucetic, and Z. Obradovic, "Uncertainty Analysis of Neural-Network-Based Aerosol Retrieval," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, pp. 409-414, 2012.
- [4] H. Niska, J. Skon, P. Packalen, T. Tokola, M. Maltamo, and M. Kolehmainen, "Neural Networks for the Prediction of Species-Specific Plot Volumes Using Airborne Laser Scanning and Aerial Photographs," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, pp. 1076-1085, 2010.
- [5] K. S. Ajil, P. K. Thapliyal, M. V. Shukla, P. K. Pal, P. C. Joshi, and R. R. Navalgund, "A New Technique for Temperature and Humidity Profile Retrieval From Infrared-Sounder Observations Using the Adaptive Neuro-Fuzzy Inference System," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, pp. 1650-1659, 2010.
- [6] S. Paloscia, P. Pampaloni, S. Pettinato, and E. Santi, "A Comparison of Algorithms for Retrieving Soil Moisture from ENVISAT/ASAR Images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, pp. 3274-3284, 2008.
- [7] X. Nan, H. Leung, and H. Chan, "A multiple-model prediction approach for sea clutter modeling," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 41, pp. 1491-1502, 2003.
- [8] M. M. Ali, P. S. V. Jagadeesh, I. I. Lin, and H. Je-Yuan, "A Neural Network Approach to Estimate Tropical Cyclone Heat Potential in the Indian Ocean," *Geoscience and Remote Sensing Letters, IEEE*, vol. 9, pp. 1114-1117, 2012.
- [9] F. Külahcı, M. İnceöz, M. Doğru, E. Aksoy, and O. Baykara, "Artificial neural network model for earthquake prediction with radon monitoring," *Applied Radiation and Isotopes*, vol. 67, pp. 212-219, 1// 2009.
- [10] M. Moustra, M. Avraamides, and C. Christodoulou, "Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals," *Expert Systems with Applications*, vol. 38, pp. 15032-15039, 11// 2011.
- [11] B. Bodri, "A neural-network model for earthquake occurrence," *Journal of Geodesy*, vol. 32, pp. 289-310, 2001.
- [12] J. Reyes, A. Morales-Esteban, and F. Martínez-Álvarez, "Neural networks to predict earthquakes in Chile," *Applied Soft Computing*, vol. 13, pp. 1314-1328, 2// 2013.

- [13] A. Morales-Esteban, F. Martínez-Álvarez, and J. Reyes, "Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence," *Tectonophysics*.
- [14] H. Adeli and A. Panakkat, "A probabilistic neural network for earthquake magnitude prediction," *Neural Networks*, vol. 22, pp. 1018-1024, 9// 2009.
- [15] F. Aminzadeh, S. Katz, and K. Aki, "Adaptive neural nets for generation of artificial earthquake precursors," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 32, pp. 1139-1143, 1994.
- [16] Π. Αργυράκης, *Νευρωνικά Δίκτυα και Εφαρμογές*.
- [17] S. Haykin, *Νευρωνικά Δίκτυα και Μηχανική Μάθηση*, Τρίτη Έκδοση ed. Αθήνα: Παπασωτηρίου, 2010.
- [18] A. P. Engelbrecht, *Computational Intelligence: An Introduction*. England: John Wiley & Sons, Ltd, 2007.
- [19] J. Wolberg, *Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments*: Springer, 2005.
- [20] T. Strutz, *Data Fitting and Uncertainty (A practical introduction to weighted least squares and beyond)*: Vieweg+Teubner, 2010.
- [21] N. R. a. S. Draper, H., *Applied Regression Analysis*: Wiley-Interscience., 1998.
- [22] S. A. a. S. Glantz, B.K., *Primer of Applied Regression and Analysis of Variance*: McGraw-Hill, 1990.
- [23] D. O. Hebb, *The Organization of Behavior*: New York: Wiley & Sons, 1949.
- [24] R. Frank, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," 1958.
- [25] J. A. Anderson and E. Rosenfeld, "Talking Nets (book excerpts): an oral history of neural networks," *Ubiquity*, vol. 2000, p. 3, 2000.
- [26] T. Law, H. Itoh, and H. Seki, "A neural network-assisted Japanese-English machine translation system," in *Neural Networks, 1993. IJCNN '93-Nagoya. Proceedings of 1993 International Joint Conference on*, 1993, pp. 2905-2908 vol.3.
- [27] M. D. Buhmann, *Radial Basis Functions: Theory and Implementations*: Cambridge University Press, 2003.
- [28] E. Cox, *The fuzzy systems handbook: a practitioner's guide to building, using, maintaining fuzzy systems*: Boston: AP Professional, 1994.
- [29] V. Novák, *Fuzzy Sets and Their Applications*: 1989, 1989.
- [30] H. Sarimveis, A. Alexandridis, G. Tsekouras, and G. Bafas, "A fast and efficient algorithm for training radial basis function neural networks based on a fuzzy partition of the input space," *Ind. Eng. Chem. Res.*, vol. 41, pp. 751-759, 2002.
- [31] A. Alexandridis, H. Sarimveis, and G. Bafas, "A new algorithm for online structure and parameter adaptation of RBF networks," *Neural Netw.*, vol. 16, pp. 1003-1017, 2003.

- [32] A. Alexandridis, D. Triantis, I. Stavrakas, and C. Stergiopoulos, "A neural network approach for compressive strength prediction in cement-based materials through the study of pressure-stimulated electrical signals," *Construction and Building Materials*, vol. 30, pp. 294–300, 2012.
- [33] K. Ninos, C. Giannakakis, I. Kompogiannis, I. Stavrakas, and A. Alexandridis, "Nonlinear Control of a DC-Motor Based on Radial Basis Function Neural Networks," in *IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, Istanbul, Turkey, 2011.
- [34] A. Alexandridis, E. Chondrodima, and H. Sarimveis, "Radial Basis Function Network Training Using a Nonsymmetric Partition of the Input Space and Particle Swarm Optimization," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 24, pp. 219-230, 2013.
- [35] McLachlan, Geoffrey J.; Do, Kim-Anh; Ambrose, Christophe (2004). Analyzing microarray gene expression data. Wiley.