



Πανεπιστήμιο Αιγαίου
ΤΜΗΜΑ ΠΟΛΙΤΙΣΜΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ
ΕΠΙΚΟΙΝΩΝΙΑΣ

Διδακτορική Διατριβή
ΤΟΥ
Ρήγου Αναστάσιου

Ανάπτυξη Πολυωνυμικών Νευρωνικών
Δικτύων με χρήση Ορθογωνίων
Πολυωνύμων

Τριμελής Συμβουλευτική Επιτροπή:

Γεώργιος Τσεκούρας (Επιβλέπων)
Αναπληρωτής Καθηγητής, Τμήμα Πολιτισμικής Τεχνολογίας και
Επικοινωνίας, Πανεπιστήμιο Αιγαίου

Αριστείδης Λύκας
Καθηγητής, Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και
Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

Σωτήρης Κωτσιαντής
Λέκτορας, Τμήμα Μαθηματικών, Πανεπιστήμιο Πατρών

Φεβρουάριος 2017

Αφιερώνεται στην οικογένειά μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα Καθηγητή Τσεκούρα Γιώργο για την καθοδήγηση και την αμέριστη συμπαράστασή του σε όλη τη διάρκεια της παρούσης διατριβής. Επίσης, ευχαριστώ τα άλλα δύο μέλη της τριμελούς Συμβουλευτικής Επιτροπής, τον κ. Λύκα Αριστείδη και τον κ. Κωτσιαντή Σωτήρη για τις διορθώσεις και παρατηρήσεις τους.

Ευχαριστώ την κόρη μου Ελένη και τη σύζυγό μου Φανή για τη συμπαράστασή τους κατά τη διάρκεια της εκπόνησης της διατριβής αυτής.

Τέλος, ευχαριστώ τους φίλους και συνεργάτες στα ερευνητικά προγράμματα "ΘΑΛΗΣ (ISLA) Physical and Economic Impacts of Sea Level Changes on the beaches of the Aegean Archipelago" και " Beachtour: Συνέργεια για την αειφόρο ανάπτυξη και ασφαλή χρήση των Ελληνικών παραλιών" οι οποίοι βοήθησαν στη συλλογή και επεξεργασία των δεδομένων της παρούσας διατριβής (βλ. παραγράφους 2.6, 3.6, 4.6 και Παράρτημα Β). Κατά αλφαβητική σειρά είναι οι: Ο. Ανδρεάδης, Α. Βελεγράκης, Ι. Μονιούδη, Β. Τριγώνης, Α. Χατζηπαυλής και Φ. Ψαρρός.

Περιεχόμενα

Περίληψη	1
Abstract	3
Εισαγωγή.....	5
1. Βιβλιογραφική Ανασκόπηση	9
1.1 Μέθοδοι βελτιστοποίησης συνάρτησης.....	9
1.1.1 Μέθοδος των ελαχίστων τετραγώνων (Least Squares)	10
1.1.2 Μέθοδος διεύθυνσης καθόδου με τη χρήση της κλίσης της συνάρτησης (Gradient Descent ή Steepest Descent).....	11
1.1.3 Αλγόριθμοι νοημοσύνης σμήνους (swarm intelligence algorithms)	14
1.1.3.1 Ο αλγόριθμος Particle Swarm Optimization (PSO)	15
1.1.3.2 Ο αλγόριθμος Artificial Bee Colony (ABC)	18
1.2 Ορθογώνια Πολυώνυμα.....	24
1.3 Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ)	27
1.3.1 Τεχνικά Νευρωνικά Δίκτυα Συναρτήσεων Ακτινικής Βάσης (RBF ANNs)	32
1.3.1.1 Εκπαίδευση ΤΝΔ συναρτήσεων ακτινικής βάσης.....	34
1.3.2 Εμπροσθοτροφοδοτούμενα ΤΝΔ (FeedForward NN, FFNN).....	45
1.3.3 Functional-Link ΤΝΔ (FLANN)	48
1.3.4 Πολυωνυμικά Νευρωνικά δίκτυα	51
1.4 Ασαφή και Νευροασαφή Συστήματα (Fuzzy and Neuro-Fuzzy Systems)	55
1.4.1 Ασαφή Σύνολα.....	55
1.4.2 Ασαφείς Κανόνες (Fuzzy Rules)	57
1.4.3 Ασαφή συστήματα	58
1.4.3.1 Ασαφή συστήματα Mamdani	58
1.4.3.2 Ασαφή συστήματα Takagi-Sugeno-Kang (TSK)	61
1.4.4 Νευροασαφή δίκτυα	62
2. Ανάπτυξη Πολυωνυμικού Νευρωνικού Δικτύου Με Χρήση Πολυωνύμων Chebyshev..	65
2.1 Συναφείς μέθοδοι και προτεινόμενη μέθοδος.....	66
2.2 Πολυώνυμα Chebyshev.....	69
2.2.1 Ορισμοί.....	69
2.2.2 Ορθογωνιότητα	72
2.2.3 Μέθοδοι παρεμβολής (interpolation)	72
2.3 Το προτεινόμενο ΤΝΔ με χρήση πολυωνύμων Chebyshev.....	75
2.3.1 Το τμήμα πολυωνύμων Chebyshev του ΤΝΔ	76
2.3.2 Το RBF τμήμα του ΤΝΔ	77
2.3.3 Το τμήμα εξαγωγής συμπεράσματος του ΤΝΔ	81
2.4 Εκπαίδευση του ΤΝΔ.....	81

2.5 Αριθμητικές προσομοιώσεις σε σύνολα δεδομένων του UCI Repository	82
2.6 Εξαγωγή Ακτογραμμής από Ψηφιακές Εικόνες	88
3. Ανάπτυξη Νευροασαφούς Δικτύου Με Χρήση Πολυωνύμων Hermite	101
3.1 Συναφείς μέθοδοι και προτεινόμενη μέθοδος	102
3.2 Πολυώνυμα Hermite	105
3.2.1 Ορισμοί και αναδρομικές σχέσεις	105
3.2.2 Ορθογωνιότητα	107
3.2.3 Μέθοδοι παρεμβολής	107
3.3 Το Προτεινόμενο Νευροασαφές δίκτυο με τα πολυώνυμα Hermite	109
3.2.1 Φάση Υπόθεσης : Σύστημα ασαφούς λογικής	109
3.2.2 Φάση Συμπεράσματος: ΤΝΔ με χρήση πολυωνύμων Hermite.....	110
3.2.3 Φάση Μηχανισμού Εξαγωγής Συμπεράσματος.....	112
3.4 Διαδικασία εκπαίδευσης του δικτύου	113
3.5 Αριθμητικές προσομοιώσεις σε σύνολα δεδομένων του UCI Repository	114
3.6 Μοντελοποίηση/Πρόβλεψη μετατόπισης/οπισθοχώρησης ακτογραμμής με χρήση του προτεινόμενου Νευροασαφούς δικτύου	118
4. Ανάπτυξη Πολυωνυμικού Νευρωνικού Δικτύου Με Χρήση Πολυωνύμων Legendre ...	123
4.1 Συναφείς μέθοδοι και η προτεινόμενη μέθοδος	123
4.2 Πολυώνυμα Legendre	126
4.3 Το Πολυωνυμικό δίκτυο με χρήση πολυωνύμων Legendre	127
4.4 Εκπαίδευση του δικτύου	130
4.5 Αριθμητικές προσομοιώσεις σε σύνολα δεδομένων του UCI Repository	132
4.6 Αριθμητικές προσομοιώσεις στο πρόβλημα πρόβλεψης της περιστροφής μιας παραλίας (beach rotation)	139
5. Συγκριτικά Αποτελέσματα Των Πολυωνυμικών Δικτύων Του Παρόντος Κειμένου Και Μελλοντική Εργασία	143
5.1 Σύγκριση των προτεινόμενων μεθόδων	143
5.2 Μελλοντική εργασία	146
Παράρτημα Α	149
Παράρτημα Β	151
B.1 Τα δεδομένα για την Εξαγωγή Ακτογραμμής από Ψηφιακές Εικόνες (Shoreline Extraction)	151
B.2 Τα δεδομένα για την Μοντελοποίηση/Πρόβλεψη μετατόπισης/οπισθοχώρησης ακτογραμμής (Coastal Erosion)	155
B.3 Τα δεδομένα για την Περιστροφή μιας Παραλίας (Beach Rotation)	160
Παράρτημα Γ	165
Βιβλιογραφία	169

Περίληψη

Στην παρούσα διδακτορική διατριβή παρουσιάζεται η ερευνητική προσπάθεια της δημιουργίας, εκπαίδευσης και χρήσης Πολυωνυμικών Τεχνητών Νευρωνικών Δικτύων (ΠΤΝΔ) με χρήση ορθογωνίων πολυωνύμων. Η έρευνα αυτή εστιάζει στην επίδραση που έχουν τα πολυώνυμα αυτά μέσα στη δομή των Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ).

Αναπτύχθηκαν τρεις διαφορετικές καινοτόμες δομές Πολυωνυμικών Τεχνητών Νευρωνικών Δικτύων (ΠΤΝΔ) οι οποίες έχουν δημοσιευθεί σε έγκυρα διεθνή περιοδικά καθώς και σε πρακτικά επιστημονικών συνεδρίων. Σκοπός και των τριών μεθόδων είναι η χρήση τους σε ισχυρά μη-γραμμικά σύνολα δεδομένων με σκοπό την αύξηση της ακριβείας των αποτελεσμάτων τους. Τα σύνολα δεδομένων αφορούν: (α) πραγματικά πειραματικά δεδομένα που συλλέχθηκαν κατά την διάρκεια της παρούσης διδακτορικής διατριβής στο πλαίσιο των ερευνητικών προγραμμάτων "ΘΑΛΗΣ (ISLA) Physical and Economic Impacts of Sea Level Changes on the beaches of the Aegean Archipelago" και "Beachtour: Συνέργεια για την αειφόρο ανάπτυξη και ασφαλή χρήση των Ελληνικών παραλιών" (βλ. Παράρτημα Β), και (β) δεδομένα που επιλέχθηκαν από την βάση δεδομένων UCI Repository.

Στο πρώτο προτεινόμενο δίκτυο που παρουσιάζεται στο κεφάλαιο 2 έγινε χρήση των ορθογωνίων πολυωνύμων Chebychev μέσα στη δομή ενός ΤΝΔ συναρτήσεων ακτινικής βάσης. Το δίκτυο αποτελείται από τα εξής τρία τμήματα: (i) Από ένα πολυωνυμικό νευρωνικό δίκτυο το οποίο χρησιμοποιεί τα πολυώνυμα Chebychev ως συναρτήσεις ενεργοποίησης, (ii) Από ένα ΤΝΔ το οποίο χρησιμοποιεί συναρτήσεις ακτινικής βάσης και μια πρωτότυπη συσταδοποίηση δεδομένων εισόδου-εξόδου. Αυτή συσταδοποιεί τα δεδομένα εισόδου λαμβάνοντας υπόψη της την κατανομή των δεδομένων εξόδου και με αυτόν τον τρόπο δίνει καλύτερες προσεγγιστικές δυνατότητες στο ΤΝΔ. (iii) Από το τμήμα συμπεράσματος το οποίο συνδυάζει τα δύο προηγούμενα τμήματα για να υπολογίσει την έξοδο του δικτύου. Το δίκτυο εκπαιδεύτηκε χρησιμοποιώντας τη μέθοδο Gradient Descent με χρήση του κανόνα του Armijo ώστε να ελαχιστοποιηθεί το τετραγωνικό σφάλμα του δικτύου. Τα πειράματα (αριθμητικές προσομοιώσεις) του ΤΝΔ αυτού πραγματοποιήθηκαν σε σύνολα δεδομένων του UCI Repository καθώς και σε ένα ισχυρά μη-γραμμικό σύνολο δεδομένων το οποίο καθορίζει την εξαγωγή της θέσης μίας ακτογραμμής από εξειδικευμένου

τύπου εικόνες τύπου οι οποίες δείχνουν την τυπική απόκλιση ενός συνόλου διαδοχικών εικόνων.

Η δεύτερη μέθοδος (κεφάλαιο 3) αφορά την ανάπτυξη ενός νευροασαφούς δικτύου το οποίο αποτελείται από (i) ένα ασαφές σύστημα (ii) ένα πολυωνυμικό ΤΝΔ το οποίο χρησιμοποιεί τα ορθογώνια πολώνυμα Hermite και (iii) την διαδικασία εξαγωγής συμπεράσματος, η οποία χρησιμοποιεί τα προηγούμενα δύο μέρη του για να εξάγει το επιθυμητό αποτέλεσμα. Το δίκτυο εκπαιδεύτηκε χρησιμοποιώντας τη μέθοδο Gradient Descent με χρήση του κανόνα του Armijo ώστε να ελαχιστοποιηθεί το ολικό τετραγωνικό σφάλμα του δικτύου. Το δίκτυο αυτό δοκιμάστηκε σε σύνολα δεδομένων του UCI Repository καθώς και σε ένα μη-γραμμικό σύνολο δεδομένων το οποίο προβλέπει την διάβρωση/μετατόπιση μίας ακτογραμμής μέσα από δεδομένα που έχουν προκύψει από χωρικές και μετεωρολογικές μετρήσεις.

Η τρίτη μέθοδος (κεφάλαιο 4) αποτελείται από ένα πολυωνυμικό ΤΝΔ που κάνει χρήση πολυωνύμων Legendre. Στο ΤΝΔ αυτό για να λειτουργεί η ορθογωνιότητα των πολυωνύμων που χρησιμοποιούνται, ήταν απαραίτητη η αλλαγή διαστήματος των δεδομένων (scaling), ώστε τα δεδομένα που διέρχονται μέσα από τους πολυωνυμικούς κόμβους του δικτύου να ανήκουν στο $[-1,1]$ διάστημα. Το προτεινόμενο ΤΝΔ εκπαιδεύτηκε με χρήση αλγορίθμων νοημοσύνης σμήνους, συγκεκριμένα με τον αλγόριθμο GABC η οποία αποτελεί παραμετροποίηση του αλγορίθμου Artificial Bee Colony (ABC) με την καθοδήγηση της θέσης G-best της μεθόδου Particle Swarm Optimization (PSO). Τέλος το δίκτυο αυτό δοκιμάστηκε σε σύνολα δεδομένων του UCI Repository καθώς και σε ένα σύνολο δεδομένων το οποίο προσπαθεί να προβλέψει την περιστροφή του μετώπου μιας παραλίας .

Σε όλες τις παραπάνω μεθόδους τα αποτελέσματα έδειξαν ότι η αναπαράσταση των συναρτήσεων ενεργοποίησης των κόμβων ενός νευρωνικού δικτύου με την χρήση ορθογωνίων πολυωνύμων είναι ικανή να βελτιώσει τόσο την συμπεριφορά όσο και την απόδοση του δικτύου σε προβλήματα παλινδρόμησης. .

Abstract

This PhD thesis presents the research involving the creation, training and use of Polynomial Artificial Neural Networks (pNN's) using orthogonal polynomials. This research focuses on the effect of these polynomials into the structure of Artificial Neural Networks.

Three original structures of Polynomial Neural Networks have been developed and they have been published in peer-reviewed journals and in proceedings of scientific conferences. The purpose of all three methods is their usage on high non-linear data sets in order to increase the accuracy of their results. The datasets involve: (a) experimental data collected from the research programs "THALES (ISLA) Physical and Economic Impacts of Sea Level Changes on the beaches of the Aegean Archipelago" and "Beachtour: Synergy for sustainable development and safe use of the Greek beaches" (see Appendix B), and (b) data selected by the UCI Repository database.

The first proposed network (Chapter 2) uses the orthogonal Chebychev polynomials into the structure of an ANN. The network consists of the following three parts: (i) A polynomial neural network which uses polynomials Chebychev as activation functions. (ii) A part consisting of an ANN which uses radial basis functions (RBF) and uses an original clustering method of the input-output data space. This method clusters input data according to the distribution of the output data space and thus provides better approximation capabilities to the ANN. And (iii) The inference part which combines the two previous parts of the network in order to calculate the network's output. The network was trained using the Gradient Descent method with the Armijo rule, in order to minimize the square error of the network. Experiments (numerical simulations) of the proposed ANN were applied on UCI Repository datasets and on a high non-linear data set which defines the position of a shoreline from digital images that show the standard deviation of a set of consecutive images.

The second method (Chapter 3) presents a neuro-fuzzy network which consists of (i) a fuzzy system (ii) a polynomial ANN that uses the orthogonal Hermite polynomials and (iii) a phase that uses the previous two parts in order to extract the desired output/result. The network was trained using the Gradient Descent method with the Armijo rule, in order to minimize the square error of the network. The network was tested on the UCI Repository data sets and on a

non-linear data set that models/predicts the coastal erosion through data generated from spatial and meteorological measurements.

The third method (Chapter 4) consists of a polynomial ANN using the Legendre polynomials. In order to maintain the orthogonality of the polynomials which operates into the $[-1,1]$ interval, a scaling procedure was applied before the data enter the network's layer that uses the polynomials. The proposed ANN was trained using a swarm intelligence algorithm, the GABC algorithm which is a modification of the Artificial Bee Colony algorithm (ABC) that uses the G-best position of the method Particle Swarm Optimization (PSO). Finally the network was tested on the UCI Repository data sets and on a data set that predicts the beach rotation phenomenon.

In all the above methods the results showed that expressing the activation functions of the nodes of a neural network using orthogonal polynomials are capable of improving the network performance in regression problems. .

Εισαγωγή

Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) είναι το αποτέλεσμα της προσπάθειας μίμησης της δομής του ανθρώπινου νευρικού συστήματος μέσω αλγορίθμων. Η χρησιμότητα των ΤΝΔ έγκειται στο γεγονός ότι μπορούν να χρησιμοποιηθούν για να εξαχθεί μια επιθυμητή έξοδος (αποτέλεσμα) μέσα από κάποιες παρατηρήσεις. Αυτό είναι ιδιαίτερα χρήσιμο σε εφαρμογές όπου η πολυπλοκότητα των δεδομένων είναι μεγάλη. Τα ΤΝΔ εφαρμόζονται σε προσεγγίσεις συναρτήσεων, προβλήματα κατηγοριοποίησης (classification), στη ρομποτική κ.α.

Μια κατηγορία ΤΝΔ είναι τα πολυωνυμικά ΤΝΔ. Τα πολυωνυμικά ΤΝΔ χρησιμοποιούν μέσα στους κόμβους πολυωνυμικές συναρτήσεις ως συναρτήσεις ενεργοποίησης, οι οποίες συνήθως είναι ορθογώνια πολυώνυμα [1]. Θεμελιωτής των πολυωνυμικών ΤΝΔ μπορεί να θεωρηθεί ο Ivakhnenko ο οποίος εισήγαγε τη μέθοδο Group method of data handling (GMDH) [2] σύμφωνα με την οποία η επιθυμητή έξοδος ενός προβλήματος παλινδρόμησης (regression) μπορεί να υπολογιστεί μέσω γραμμικών συνδυασμών των δεδομένων εισόδου και των γινομένων τους. Οι Oh και Predycz [3] δημιούργησαν ένα αυτο-οργανωμένο ΤΝΔ (self-organizing neural network) το οποίο συνδυάζει τη μέθοδο GMDH με αυτο-οργανωμένα ΤΝΔ και γενετικούς αλγορίθμους. Οι Lee και Jeng [4] απέδειξαν ότι αν ένα εμπροσθοτροφοδοτούμενο ΤΝΔ έχει γραμμικές συναρτήσεις ενεργοποίησης οι οποίες είναι ολοκληρώσιμες, τότε αυτό μπορεί να αναπαρασταθεί ισοδύναμα από ένα δίκτυο που έχει ως βάση του ορθογώνια πολυώνυμα. Στηριζόμενοι στην πρόταση αυτή, δημιούργησαν ένα εμπροσθοτροφοδοτούμενο ΤΝΔ το οποίο έχει ως κόμβους πολυώνυμα και πράξεις πολυωνύμων Chebychev. Οι Oh et al. [5] στηριζόμενοι στη μέθοδο GMDH δημιούργησαν ένα πολυωνυμικό ΤΝΔ η έξοδος του οποίου υπολογίζεται ως γραμμικοί συνδυασμοί γινομένων των δεδομένων εισόδου ενός ΤΝΔ. Οι Oh et al. [6] δημιούργησαν ένα πολυωνυμικό ΤΝΔ το οποίο λειτουργεί παράλληλα με ένα ΤΝΔ συναρτήσεων ακτινικής βάσης (RBFNN). Οι Oh και Pedrycz [7] προτείνουν ένα ασαφές πολυωνυμικό ΤΝΔ στο οποίο χρησιμοποιούν πολυώνυμα ως συναρτήσεις ενεργοποίησης στους κόμβους του. Οι Huang et al. [8] δημιούργησαν ένα νευροασαφές σύστημα στο οποίο χρησιμοποιούν ορθογώνια πολυώνυμα. Οι Oh et al. [9] χρησιμοποιούν σε πολυωνυμικά ΤΝΔ συναρτήσεων ακτινικής βάσης αλγορίθμους νοημοσύνης σμήνους.

Οι παραπάνω μέθοδοι όταν χρησιμοποιηθούν σε περίπλοκα συστήματα, καταλήγουν σε μεγάλο πλήθος παραμέτρων. Το πρόβλημα αυτό προσπαθεί να λύσει το προτεινόμενο ΤΝΔ του δευτέρου κεφαλαίου, το οποίο αποτελείται από δύο τμήματα, ένα τμήμα με ορθογώνια πολυώνυμα Chebychev και ένα τμήμα με ΤΝΔ συναρτήσεων ακτινικής βάσης (radial basis function ή RBF). Αυτό σημαίνει ότι ταυτόχρονα χρησιμοποιεί και τις προσεγγιστικές δυνατότητες των ΤΝΔ συναρτήσεων ακτινικής βάσης τα οποία χρησιμοποιούνται ευρέως σε μη-γραμμικά συστήματα (βλ. [10-16]). Το τμήμα των πολυωνύμων Chebychev εφαρμόζει τα πολυώνυμα αυτά στα δεδομένα εισόδου του ΤΝΔ, αφού όμως έχει φροντίσει τα δεδομένα να βρίσκονται μέσα στο διάστημα $[-1,1]$. Το τμήμα συναρτήσεων ακτινικής βάσης είναι ένα κλασσικό ΤΝΔ το οποίο χρησιμοποιεί μια πρωτότυπη συσταδοποίηση των δεδομένων εισόδου-εξόδου. Το προτεινόμενο δίκτυο θα μπορούσαμε να πούμε πως έχει τη μορφή ενός κλασσικού ΤΝΔ συναρτήσεων ακτινικής βάσης τα συναπτικά βάρη του οποίου υπολογίζονται από το πολυωνυμικό τμήμα του δικτύου. Για την εκπαίδευση του δικτύου αυτού χρησιμοποιείται η μέθοδος Gradient Descent με χρήση του Armijo Rule [17-19]. Το προτεινόμενο δίκτυο εφαρμόζεται πάνω σε ένα πολύπλοκο, ισχυρά μη-γραμμικό σύνολο δεδομένων (βλ. παράγραφο 2.6) και καταλήγει σε ικανοποιητικά αποτελέσματα.

Άλλα ορθογώνια πολυώνυμα που χρησιμοποιούνται στη βιβλιογραφία είναι τα πολυώνυμα Hermite. Οι Ma και Khorasani [20, 21] χρησιμοποιούν τα ορθοκανονικοποιημένα πολυώνυμα Hermite ως συναρτήσεις ενεργοποίησης σε ένα Functional Link ΤΝΔ. Ο Beliczynski [22] χρησιμοποιεί δυσδιάστατα πολυώνυμα Hermite, οι Shi et al. [23] δημιούργησαν ένα χαοτικό ΤΝΔ με πολυώνυμα Hermite, οι Linh et al. [24] προσέγγισαν τα δεδομένα εισόδου τους ως εκφράσεις πολυωνύμων Hermite και κατόπιν εφάρμοσαν σε αυτά ένα ασαφές σύστημα Takagi-Sugeno-Kang. Οι Rasiah et al. [25] χρησιμοποιούν τα πολυώνυμα Hermite ως συναρτήσεις βάσης για να μοντελοποιήσουν μονοδιάστατα σήματα.

Όλες οι παραπάνω μέθοδοι χρησιμοποιούν είτε τα ορθογώνια πολυώνυμα Hermite, είτε τα δεδομένα εισόδου απευθείας πάνω στα πολυώνυμα (Shi et al. [23]), είτε χρησιμοποιούν ασαφή συστήματα (Linh et al. [24]). Το προτεινόμενο νευροασαφές σύστημα (κεφάλαιο 3) λειτουργεί εντελώς διαφορετικά από αυτά της βιβλιογραφίας. Καταρχήν, δεν εφαρμόζει τα πολυώνυμα Hermite στα δεδομένα εισόδου αλλά σε γραμμικούς συνδυασμούς τους μέσω συναπτικών βαρών και επίσης χρησιμοποιεί παράλληλα και ένα ασαφές σύστημα (φάση υπόθεσης του προτεινόμενου δικτύου) για να μπορέσει με τη συνεργασία των δύο αυτών τμημάτων να αυξήσει την απόδοση του δικτύου. Το δίκτυο αυτό αποτελείται από τρία

τμήματα, από ένα πολυωνυμικό ΤΝΔ με χρήση πολυωνύμων Hermite, από ένα τμήμα το οποίο είναι ένα ασαφές σύστημα εξαγωγής συμπερασμάτων και τέλος από το τμήμα που συνδέει τα δεδομένα των δύο προηγούμενων τμημάτων με σκοπό την εξαγωγή συμπεράσματος. Για την εκπαίδευση του δικτύου αυτού χρησιμοποιείται η μέθοδος Gradient Descent με χρήση του Armijo Rule [18].

Τα επόμενα πιο διαδεδομένα ορθογώνια πολυώνυμα στη βιβλιογραφία είναι τα Legendre. Οι Patra et al. [26] δημιούργησαν ένα ΤΝΔ με χρήση πολυωνύμων Legendre, το LeNN, στο οποίο εφάρμοσαν απευθείας τα πολυώνυμα Legendre στα δεδομένα εισόδου ως συναρτήσεις ενεργοποίησης του ΤΝΔ. Το ίδιο δίκτυο (LeNN) το χρησιμοποίησαν οι Patra και Bornard [27] για αναγνώριση δυναμικών μη-γραμμικών συστημάτων, ενώ οι Nanda et al. [28] για να προβλέψουν τη ποιότητα του αέρα ως συνάρτηση μετεωρολογικών και ατμοσφαιρικών δεδομένων. Οι Liu και Wang [29] δημιούργησαν ένα τρι-στρωματικό δίκτυο και το χρησιμοποίησαν στην πρόβλεψη των τιμών μετοχών του χρηματιστηρίου. Σε όλες τις παραπάνω αναφορές τα δεδομένα εισόδου εισέρχονται απευθείας μέσα από τους κόμβους του δικτύου που περιέχουν τις συναρτήσεις των πολυωνύμων Legendre εκτός από τους Liu και Wang [29] όπου οι συγγραφείς έχουν παρατηρήσει τα δεδομένα εισόδου του ΤΝΔ να έχουν τιμές στο διάστημα $[0,1]$ αλλά δεν ισχύει το ίδιο και για τα δεδομένα που χρησιμοποίησαν ως μεταβλητές εισόδου στα πολυώνυμα Legendre.

Η χρήση πολυωνύμων Legendre έχει το πρόβλημα ότι τα πολυώνυμα Legendre είναι ορθογώνια μόνο στο διάστημα $[-1,1]$ και ως εκ τούτου πρέπει να γίνει μετασχηματισμός στα δεδομένα που εισέρχονται στα πολυώνυμα αυτά ή θα πρέπει τα δεδομένα που εισέρχονται μέσα στα πολυώνυμα να μην είναι άμεση έκφραση (direct expansion) των δεδομένων εισόδου του ΤΝΔ. Όμως, οποιαδήποτε κανονικοποίηση των δεδομένων εισόδου έχει ως αποτέλεσμα συγκεκριμένα προβλήματα στην υλοποίηση της επαναληπτικής (iterative) εκπαίδευσης του δικτύου, επειδή καθίσταται δύσκολη η αποδοτική διαχείριση των συναπτικών βαρών. Στην παρούσα διατριβή, το πρόβλημα αυτό επιλύεται στην τρίτη προτεινόμενη μέθοδο (ενότητα 4). Συγκεκριμένα, αναπτύχθηκε ένα πολυωνυμικό δίκτυο, το οποίο χρησιμοποιεί γραμμικό αλγόριθμο νοημοσύνης σμήνους, όπου τα συναπτικά βάρη ικανοποιούν συγκεκριμένους περιορισμούς. Το αποτέλεσμα είναι ιδιαίτερα ικανοποιητικό, και οφείλεται στο γεγονός ότι με αυτόν τον τρόπο τα συναπτικά ενσωματώνονται πλήρως στην διαδικασία της εκπαίδευσης του δικτύου. Η εφαρμογή του δικτύου επιφέρει πολύ καλά αποτελέσματα σε προσομοιώσεις σε περίπλοκα συστήματα (βλ. παράγραφο 4.6). Εν

κατακλείδι, το αν λόγω δίκτυο αποτελείται από τρία στρώματα και αυτό που πρέπει να σημειωθεί είναι ότι τα πολώνυμα Legendre δεν εφαρμόζονται στα δεδομένα εισόδου του ΤΝΔ αλλά σε γραμμικούς συνδυασμούς τους (μέσω συναπτικών βαρών) στους οποίους έχει εφαρμοστεί αλλαγή διαστήματος ώστε τα δεδομένα αυτά να έχουν τιμές στο διάστημα $[-1,1]$. Για την εκπαίδευση του δικτύου αυτού χρησιμοποιείται ο αλγόριθμος GABC [30].

Εκτός από τα κεφάλαια 2-4 που το περιεχόμενό τους αναλύθηκε παραπάνω, στο 1^ο κεφάλαιο του παρόντος κειμένου παρουσιάζεται μια βιβλιογραφική ανασκόπηση που δίνει τα προαπαιτούμενα για την ανάλυση που ακολουθεί στα επόμενα κεφάλαια. Πιο συγκεκριμένα, στην παράγραφο 1.1 υπάρχει ο ορισμός της βελτιστοποίησης συνάρτησης και παρουσιάζονται οι μέθοδοι βελτιστοποίησης που χρησιμοποιούνται στα κεφάλαια 2-4. Έπειτα στην παράγραφο 1.2 δίνεται ο ορισμός των ορθογωνίων πολυωνύμων. Κατόπιν στην παράγραφο 1.3 δίνεται ο ορισμός των τεχνητών νευρωνικών δικτύων και παρουσιάζονται οι διάφορες κατηγορίες που χρησιμοποιούνται στο παρόν κείμενο. Τέλος, η ενότητα 1.4 παρουσιάζει τα ασαφή και νευροασαφή συστήματα τα οποία στη συνέχεια χρησιμοποιούνται στο κεφάλαιο 3.

Στο κεφάλαιο 5, δίνονται τα συγκριτικά αποτελέσματα μεταξύ των τριών προτεινόμενων μεθόδων και ακολουθεί η παράγραφος της μελλοντικής εργασίας/έρευνας.

Όλες οι αριθμητικές προσομοιώσεις (πειράματα) πραγματοποιήθηκαν με το λογισμικό MatLab. Τα σχήματα του παρόντος έγγραφου σχεδιάστηκαν με το λογισμικό MatLab, με τη γλώσσα προγραμματισμού Python και με το πρόγραμμα LibreOffice Impress.

1. Βιβλιογραφική Ανασκόπηση

Στο κεφάλαιο αυτό περιγράφονται οι μέθοδοι της διεθνούς βιβλιογραφίας που χρησιμοποιήθηκαν στην παρούσα διδακτορική διατριβή. Στην πρώτη παράγραφο επεξηγείται η έννοια της βελτιστοποίησης συνάρτησης και επεξηγούνται τέσσερις μέθοδοι βελτιστοποίησης. Στη δεύτερη παράγραφο δίνεται ο ορισμός των ορθογωνίων πολυωνύμων. Στη τρίτη παράγραφο δίνεται ο ορισμός των τεχνητών νευρωνικών δικτύων (ΤΝΔ) και κάποια παραδείγματα από ΤΝΔ. Τέλος, στην τέταρτη παράγραφο εξηγούνται τα ασαφή και τα νευροασαφή συστήματα.

1.1 Μέθοδοι βελτιστοποίησης συνάρτησης

Η έννοια της βελτιστοποίησης μίας συνάρτησης είναι συνυφασμένη με τις έννοιες της ελαχιστοποίησης και της μεγιστοποίησης. Στην παρούσα διατριβή θα ασχοληθούμε μόνο με προβλήματα ελαχιστοποίησης. Τα προβλήματα ελαχιστοποίησης μιας συνάρτησης $f: \mathbb{R}^p \rightarrow \mathbb{R}$ χωρίζονται στις επόμενες δύο κατηγορίες [31] :

(α) Ως προβλήματα ελαχιστοποίησης χωρίς περιορισμούς (unconstrained optimization) ορίζονται τα πρόβλημα εύρεσης της τιμής ή των τιμών \mathbf{x}^* έτσι ώστε :

$$\mathbf{x}^* = \{\mathbf{x} \in \mathbb{R}^p / f(\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^p} f(\mathbf{x})\} \quad (1.1)$$

ενώ

(β) Ως προβλήματα ελαχιστοποίησης υπό περιορισμούς (constrained optimization) ορίζονται τα προβλήματα εύρεσης της τιμής ή των τιμών \mathbf{x}^* έτσι ώστε :

$$\mathbf{x}^* = \{\mathbf{x} \in D / f(\mathbf{x}) = \min_{\mathbf{x} \in D} f(\mathbf{x})\} \quad (1.2)$$

όπου $D \subset \mathbb{R}^p$ το σύνολο που ορίζεται από τους περιορισμούς που πρέπει να ικανοποιεί η βέλτιστη λύση \mathbf{x}^* .

Η συνάρτηση f που πρέπει να ελαχιστοποιηθεί ονομάζεται αντικειμενική συνάρτηση κόστους (objective function). Στις ακόλουθες παραγράφους παρουσιάζονται δύο κλασσικές

τεχνικές ελαχιστοποίησης καθώς και δύο τεχνικές ελαχιστοποίησης οι οποίες ανήκουν στην κατηγορία των αλγορίθμων νοημοσύνης σμήνους (swarm intelligence algorithms).

1.1.1 Μέθοδος των ελαχίστων τετραγώνων (Least Squares)

Η μέθοδος των ελαχίστων τετραγώνων είναι από τις κλασικότερες μεθόδους ελαχιστοποίησης συνάρτησης. Σύμφωνα με τη μέθοδο αυτή [32] θεωρούμε ένα γραμμικό σύστημα $\mathbf{Ax} = \mathbf{b}$ με $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} = [b_1, b_2, \dots, b_m]^T$ έναν $m \times 1$ πίνακα-στήλη και $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ έναν $n \times 1$ πίνακα-στήλη με τις άγνωστες παραμέτρους που πρέπει να προσδιοριστούν. Τότε η βέλτιστη λύση \mathbf{x}^* του συστήματος αυτού η οποία ελαχιστοποιεί την ποσότητα $\|\mathbf{b} - \mathbf{Ax}^*\|^2$ είναι η:

$$\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \quad (1.3)$$

Το πλεονέκτημα της μεθόδου είναι ο γρήγορος υπολογισμός του \mathbf{x}^* ακόμη κι αν έχουμε μεγάλων διαστάσεων πίνακες (οι αντιστροφές πινάκων γίνονται με μεθόδους αριθμητικής ανάλυσης). Το μεγάλο πρόβλημα είναι πως ο αντίστροφος του πίνακα της προηγούμενης εξίσωσης υπάρχει μόνο όταν ο πίνακας \mathbf{A} έχει γραμμικώς ανεξάρτητες στήλες, διαφορετικά η επίλυση καταλήγει σε μία λύση η οποία δεν είναι μοναδική. Ένα ακόμη μειονέκτημα είναι ότι τα δεδομένα με τιμές σε πολύ διαφορετικό εύρος από ότι τα υπόλοιπα (outliers) επηρεάζουν πολύ τη λύση του συστήματος.

1.1.2 Μέθοδος διεύθυνσης καθόδου με τη χρήση της κλίσης της συνάρτησης (Gradient Descent ή Steepest Descent)

Η μέθοδος Gradient Descent (ή Steepest Descent) στηρίζεται στο γεγονός ότι η κλίση (ή βαθμίδα) ∇f (gradient) μιας συνάρτησης $f: \mathbb{R}^p \rightarrow \mathbb{R}$ δείχνει τη διεύθυνση της μέγιστης μεταβολής της συνάρτησης αυτής, δηλαδή τη διεύθυνση προς την οποία αυξάνει. Άρα η διεύθυνση $-\nabla f$ δείχνει κατεύθυνση της συνάρτησης προς το ελάχιστό της (τοπικό ή ολικό). Σκοπός της μεθόδου αυτής είναι να βρεθεί ένα σημείο \mathbf{x}^* όπου η συνάρτηση f ελαχιστοποιείται. Η μέθοδος πραγματοποιείται με τον ακόλουθο αλγόριθμο [33]:

Αλγόριθμος Gradient Descent:

Βήμα 1^ο: Αρχικό βήμα επανάληψης ($t = 0$). Επιλέγεται τυχαία ένα σημείο του πεδίου ορισμού της συνάρτησης f :

$$\mathbf{x}_0 = \mathbf{x}_{(t=0)} = [x_{01}, \dots, x_{0p}] \quad (1.4)$$

και ένας μικρός θετικός αριθμός η ο οποίος ονομάζεται ρυθμός μάθησης (learning rate).

Επανάλαβε

Βήμα 2^ο: Βήμα επανάληψης $t = t + 1$. Υπολογισμός του $\nabla f(\mathbf{x}_{t-1})$ και επανυπολογισμός του νέου σημείου $\mathbf{x}_{(t)}$ ως:

$$\mathbf{x}_{(t)} = \mathbf{x}_{(t-1)} - \eta \cdot \nabla f(\mathbf{x}_{(t-1)}) \quad (1.5)$$

Έως ότου: ικανοποιηθεί κάποιο κριτήριο τερματισμού.

Η παράμετρος η στην (1.5) είναι ένας πραγματικός αριθμός και ονομάζεται ρυθμός μάθησης (learning rate).

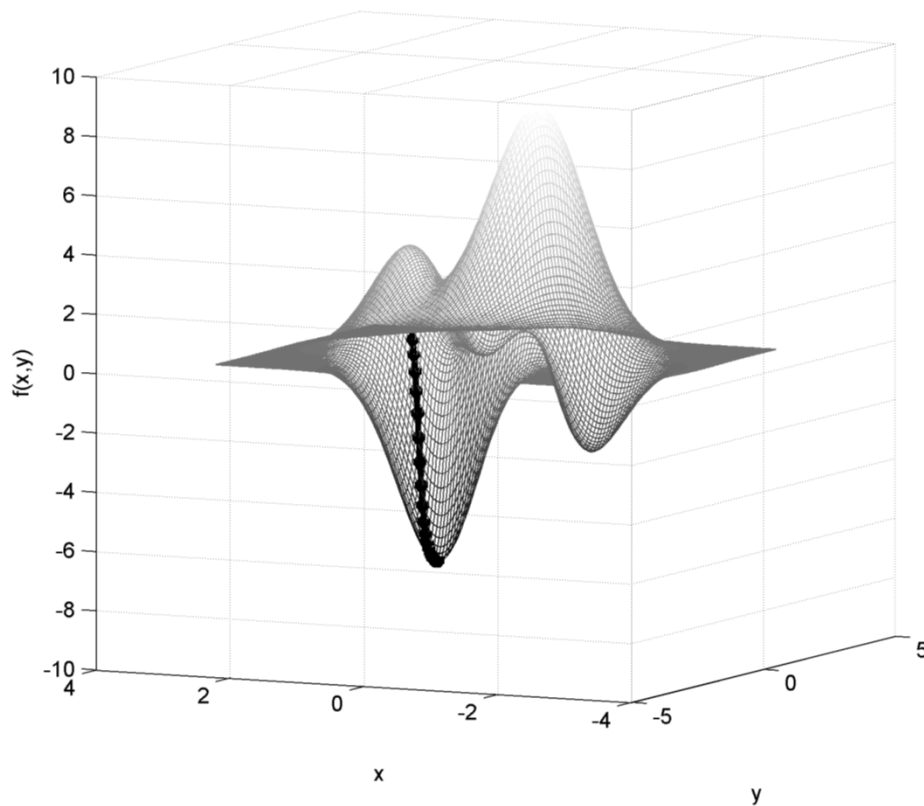
Το κριτήριο τερματισμού μπορεί να είναι κάποιο από τα εξής: τα διαδοχικά σημεία $\mathbf{x}_{(t)}$, $\mathbf{x}_{(t+1)}$ να μην αλλάζουν πολύ από βήμα σε βήμα της επανάληψης ή να φτάσει η επανάληψη σε κάποιο προκαθορισμένο μέγιστο αριθμό επαναλήψεων.

Μειονεκτήματα της μεθόδου είναι τα ακόλουθα.

- Οι μερικές παράγωγοι της εξίσωσης (1.5) μπορεί να καταλήξουν σε πολύ δύσκολους υπολογισμούς.

- Αν ο ρυθμός μάθησης (learning rate) επιλεγεί πολύ μικρός τότε η επανάληψη στον προηγούμενο αλγόριθμο μπορεί να καθυστερήσει να καταλήξει σε ελάχιστο [34].
- Αν ο ρυθμός μάθησης (learning rate) επιλεγεί πιο μεγάλος από ότι πρέπει τότε ο προηγούμενος αλγόριθμος θα αποκλίνει.

Ένα παράδειγμα που δείχνει πώς τα διαδοχικά σημεία $\mathbf{X}_{(t=0)}, \mathbf{X}_{(t=1)}, \mathbf{X}_{(t=2)}, \dots, \mathbf{X}_{(t=t_{\max})}$ της εξίσωσης (1.5) συγκλίνουν στο ελάχιστο μιας συνάρτησης παρουσιάζεται στο Σχήμα 1.1.



Σχήμα 1.1 Το μονοπάτι που ακλούθησε η μέθοδος Gradient Descent για να ελαχιστοποιήσει τη συνάρτηση:

$$f(x, y) = 3(1-x)^2 e^{-x^2-(y+1)^2} - 10 \left(\frac{x}{5} - x^3 - y^5 \right) e^{-x^2-y^2} - \frac{1}{3} e^{-(x+1)^2-y^2}.$$

Για να χρησιμοποιηθεί στη μέθοδο αυτή ένας καλός ρυθμός εκμάθησης η μπορεί να γίνει χρήση του κανόνα του Armijo (βλ. [17-19]) σύμφωνα με τον οποίο ο ρυθμός εκμάθησης είναι μεταβλητός ($\eta(t)$ αντί για η) σε κάθε βήμα της επανάληψης του αλγορίθμου. Η μέθοδος αυτή εγγυάται ότι η συνάρτηση f θα ελαχιστοποιηθεί αποδεκτά. Στον κανόνα

Armijo η εξίσωση (1.5) αλλάζει ώστε στην t -οστή επανάληψη του αλγορίθμου Gradient Descent η διεύθυνση καθόδου είναι η: $-\eta(t) \cdot \nabla f(\mathbf{x}_{(t)})$ και άρα το νέο σημείο $\mathbf{x}_{(t+1)}$ είναι το :

$$\mathbf{x}_{(t+1)} = \mathbf{x}_{(t)} - \eta(t) \cdot \nabla f(\mathbf{x}_{(t)}) \quad (1.6)$$

όπου $\eta(t) = \gamma^\tau$, με γ μια σταθερά $\gamma \in (0,1)$ και τ ο ελάχιστος θετικός ακέραιος τέτοιος ώστε να ισχύει ότι:

$$f(\mathbf{x}_{(t+1)}) - f(\mathbf{x}_{(t)}) < -\varepsilon \eta(t) \|\nabla f(\mathbf{x}_{(t)})\|^2 \quad (1.7)$$

όπου ε ένας μικρός θετικός αριθμός με $\varepsilon \in (0,1)$.

1.1.3 Αλγόριθμοι νοημοσύνης σμήνους (swarm intelligence algorithms)

Οι αλγόριθμοι νοημοσύνης σμήνους είναι υποκατηγορία των εξελικτικών αλγορίθμων (evolutionary algorithms) και είναι η προσομοίωση της συλλογικής συμπεριφοράς πολλών ατόμων/πρακτόρων (agents) ενός πληθυσμού/σμήνους (swarm), τα οποία αλληλεπιδρούν ακολουθώντας μερικούς απλούς κανόνες [35-37]. Κάθε άτομο μπορεί να θεωρηθεί μη-νοήμων (unintelligent), ενώ το όλο σύστημα των πολλών ατόμων μπορεί να παρουσιάζει κάποια αυτοοργανωμένη συμπεριφορά και ως εκ τούτου όλα τα άτομα μαζί μπορούν να επιδεικνύουν ένα είδος συλλογικής νοημοσύνης.

Έχουν αναπτυχθεί πολλοί αλγόριθμοι εμπνευσμένοι από τη συμπεριφορά σμηνών στη φύση. Οι κύριες ιδιότητες για τους αλγόριθμους νοημοσύνης σμήνους είναι εν συντομία οι εξής:

- Κοινή χρήση πληροφοριών μεταξύ των πολλών ατόμων του σμήνους.
- Τα άτομα έχουν ιδιότητες αυτο-οργάνωσης και συν-εξέλιξης.
- Είναι ιδιαίτερα αποτελεσματικοί στη συνεργατική απόκτηση πληροφοριών (συνεργατική εκμάθηση) .
- Μπορούν εύκολα να προσομοιώσουν πρακτικά προβλήματα.

Παραδείγματα τέτοιων αλγορίθμων στη βιβλιογραφία είναι οι: Particle Swarm Optimization (PSO) [38], Artificial Bee Colony (ABC) [39], Ant Colony Optimisation (ACO) [40], Bat Algorithm (BA) [41], Artificial Fish Swarm [42], Wolf Search Algorithm [43] κ.ά.

Οι παραπάνω αλγόριθμοι έχουν ευρέως χρησιμοποιηθεί στην εκπαίδευση τεχνητών νευρωνικών δικτύων [37, 44-46], ακολουθούν ενδεικτικά κάποια παραδείγματα. Οι Socha και Blum [47] χρησιμοποίησαν τον αλγόριθμο ACO για την εκπαίδευση ενός εμπροσθο-τροφοδοτούμενου ΤΝΔ του οποίου στόχος είναι το ταξινόμηση (classification) δεδομένων, το ίδιο πρόβλημα προσέγγισαν και οι Hasan et al. [48] χρησιμοποιώντας τον αλγόριθμο Artificial Fish Swarm. Οι Tsay et al. [49] χρησιμοποιούν τον BA στη βελτιστοποίηση συναρτήσεων. Οι Mirjalili et al. [50] εφαρμόζουν τον αλγόριθμο Grey Wolf Optimizer στην βελτιστοποίηση συναρτήσεων και οι Yamany et al. [51] στην μείωση χαρακτηριστικών (attribute reduction) σε προβλήματα ταξινόμησης (classification). Στη συνέχεια αναλύονται οι αλγόριθμοι PSO και ABC.

1.1.3.1 Ο αλγόριθμος *Particle Swarm Optimization (PSO)*

Ο αλγόριθμος *Particle Swarm Optimization (PSO)* προτάθηκε από τους J. Kennedy και R. Eberhart [38] το 1995 και προσομοιώνει την κοινωνική συμπεριφορά ενός σμήνους πουλιών ή ψαριών.

Το πρόβλημα που επιλύει ο αλγόριθμος είναι η ελαχιστοποίηση της συνάρτησης $f : D \rightarrow \mathbb{R}$, δηλαδή η εύρεση ενός σημείου :

$$\theta^* = \min_{\theta \in D} f(\theta) \quad (1.8)$$

όπου $D \subseteq \mathbb{R}^d$ ο χώρος λύσεων του αλγορίθμου (search space). Αν έπρεπε να μεγιστοποιηθεί μια συνάρτηση, απλά θα χρησιμοποιούσαμε $\theta^* = \min_{\theta \in D} (-f(\theta))$. Ο χώρος λύσεων D μπορεί να θεωρηθεί φραγμένος από τα p -διάστατα διανύσματα $\mathbf{b}^{lo} = [b_1^{lo}, b_2^{lo}, \dots, b_p^{lo}]^T$, $\mathbf{b}^{up} = [b_1^{up}, b_2^{up}, \dots, b_p^{up}]^T$ ως:

$$D = \{ \mathbf{p} = [p_1, p_2, \dots, p_d] \in \mathbb{R}^p / b_i^{lo} \leq p_i \leq b_i^{up}, \forall i = 1, 2, \dots, d \} \quad (1.9)$$

Στον αλγόριθμο PSO το σμήνος (swarm) είναι ένα σύνολο S ατόμων (particles) και η λογική του αλγορίθμου είναι ότι τα άτομα κινούνται μέσα στον χώρο λύσεων καθοδηγούμενα από την προσωπική τους βέλτιστη λύση $\mathbf{p}_i^{best}, i = 1, 2, \dots, S$ και την βέλτιστη λύση ολόκληρου του σμήνους \mathbf{g}^{best} . Όταν ανακαλυφθούν καλύτερες θέσεις/λύσεις τότε αυτές καθοδηγούν την κίνηση του σμήνους και η προηγούμενη διαδικασία επαναλαμβάνεται έως ότου να βρεθεί μια ικανοποιητική λύση.

Συμβολίζοντας ως $\mathbf{p}_i \in \mathbb{R}^d, i = 1, 2, \dots, S$ τη θέση κάθε ατόμου του σμήνους, ως $\mathbf{v}_i \in \mathbb{R}^d, i = 1, 2, \dots, S$ την ταχύτητά του, ως $\mathbf{p}_i^{best} \in \mathbb{R}^d, i = 1, 2, \dots, S$ την προσωπική βέλτιστη λύση (personal best solution) και ως $\mathbf{g}^{best} \in \mathbb{R}^d$ τη βέλτιστη λύση ολόκληρου του σμήνους (global best), ο αλγόριθμος PSO είναι ο ακόλουθος:

Ο Αλγόριθμος PSO

Θέσε ως $f(\mathbf{g}^{best})$ έναν μεγάλο αριθμό, π.χ. $f(\mathbf{g}^{best}) = 10^{100}$

Για κάθε άτομο $i = 1, 2, \dots, S$:

- Όρισε την αρχική θέση του \mathbf{p}_i στην τύχη (ακολουθώντας την ομοιόμορφη κατανομή) εντός του χώρου εύρεσης λύσεων, δηλαδή $\mathbf{p}_i = \text{random}(\mathbf{b}^{lo}, \mathbf{b}^{up}) \sim U(\mathbf{b}^{lo}, \mathbf{b}^{up})$
- Όρισε ως βέλτιστη θέση την $\mathbf{p}_i^{best} = \mathbf{p}_i$
- Αν $f(\mathbf{p}_i^{best}) < f(\mathbf{g}^{best})$ τότε καθολική βέλτιστη λύση είναι η : $\mathbf{g}^{best} = \mathbf{p}_i^{best}$
- Όρισε ως αρχική ταχύτητα του κάθε ατόμου την: $\mathbf{v}_i = [v_{i1}, v_{i2}, \dots, v_{id}]^T$ με συνιστώσες : $v_{ij} = \text{random}(-|b_j^{up} - b_j^{lo}|, |b_j^{up} - b_j^{lo}|) \sim U(-|b_j^{up} - b_j^{lo}|, |b_j^{up} - b_j^{lo}|)$ για κάθε διάσταση $j = 1, 2, \dots, d$.

Επανάλαβε

Για κάθε άτομο $i = 1, 2, \dots, S$:

- Ανανέωσε την ταχύτητα του κάθε ατόμου ως :

$$\mathbf{v}_i = \omega \mathbf{v}_i + c_1 r_p (\mathbf{p}_i^{best} - \mathbf{p}_i) + c_2 r_g (\mathbf{g}_i^{best} - \mathbf{p}_i) \quad (1.10)$$

όπου $r_p, r_g = \text{random}(0,1) \sim U(0,1)$ και ω, c_1, c_2 παράμετροι του αλγορίθμου PSO.

- Ανανέωσε τη θέση κάθε ατόμου ως : $\mathbf{p}_i = \mathbf{p}_i + \mathbf{v}_i$
- Αν $f(\mathbf{x}_i) < f(\mathbf{p}_i^{best})$ τότε:
 - Ανανέωσε την προσωπική βέλτιστη θέση κάθε ατόμου ως : $\mathbf{p}_i^{best} = \mathbf{p}_i$
 - Αν $f(\mathbf{p}_i^{best}) < f(\mathbf{g}^{best})$ τότε καθολική βέλτιστη λύση είναι η : $\mathbf{g}^{best} = \mathbf{p}_i^{best}$

Έως ότου να ικανοποιηθεί κάποιο κριτήριο τερματισμού (π.χ. μέγιστος αριθμός επαναλήψεων ή να μην αλλάζει η βέλτιστη λύση \mathbf{g}^{best}).

Βέλτιστη λύση στο τέλος είναι η \mathbf{g}^{best}

Οι παράμετροι $\omega, \varphi_p, \varphi_g$ του αλγορίθμου PSO ορίζονται ως:

ω : παράμετρος αδράνειας με τιμή $\omega \in [0.5, 2]$

c_1 : γνωστική συνιστώσα (cognitive parameter) με συνήθη τιμή την $c_1 = 1.5$

c_2 : συνιστώσα κοινωνικοποίησης με το υπόλοιπο σμήνος (constant for the social component)

με συνήθη τιμή την $c_2 = 1.5$

Στη βιβλιογραφία υπάρχουν πολλές εφαρμογές του PSO σε ΤΝΔ, κάποια παραδείγματα αναφέρονται στη συνέχεια. Οι Qasem και Shamsuddin [52] χρησιμοποιούν τον PSO για να ελαχιστοποιήσουν το σφάλμα ενός ΤΝΔ ακτινικής βάσης. Οι Oh et al. [6] εκπαίδευσαν ένα ασαφές ΤΝΔ συναρτήσεων ακτινικής βάσης με τη μέθοδο Steepest Descent χρησιμοποιώντας τον PSO ως εξής: έθεσαν ως θέση \mathbf{p}_i του σμήνους τον ρυθμό εκμάθησης η της μεθόδου Gradient Descent και την ασαφή παράμετρο m της ασαφούς συσταδοποίησης. Οι Shao et al. [53] χρησιμοποίησαν τον PSO ορίζοντας ως παραμέτρους ως προς εύρεση όλες τις παραμέτρους του δικτύου. Οι Tsekouras και Tsimikas [54] χρησιμοποίησαν τον PSO για να δημιουργήσουν μία καινούργια μέθοδο συσταδοποίησης με βάση την οποία υπολόγισαν τα κέντρα των συναρτήσεων ενεργοποίησης ενός νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης.

1.1.3.2 Ο αλγόριθμος *Artificial Bee Colony (ABC)*

Ο αλγόριθμος *Artificial Bee Colony* προτάθηκε από τον Karaboga το 2005 [39] και είναι ένας αλγόριθμος βελτιστοποίησης συνάρτησης που στηρίζεται στη συμπεριφορά μιας αποικίας/κυψέλης μελισσών. Ο αλγόριθμος βασίζεται σε τρία δομικά στοιχεία: τις απασχολούμενες μέλισσες (employed bees), τις μέλισσες-θεατές (onlooker bees), τις μέλισσες-ανιχνευτές (scout bees) και τις πηγές τροφής (food sources). Τα είδη μελισσών ψάχνουν για 'πλούσιες' πηγές τροφής κοντά στην κυψέλη τους. Το μοντέλο στηρίζεται στα εξής δύο μοντέλα συμπεριφοράς: (α) θετική ανατροφοδότηση του συστήματος αν βρεθεί πλούσια πηγή τροφής και (β) αρνητική αντίδραση του συστήματος αν βρεθεί τροφή με μικρή ποσότητα φαγητού.

Ο αλγόριθμος επιλύει ζητήματα βελτιστοποίησης υπό περιορισμούς (constrained optimization), μπορεί όμως με μετατροπή να επιλύσει και πρόβλημα χωρίς περιορισμούς (βλ. παράγραφο 4.4). Σύμφωνα με τον Karaboga [39], το πρόβλημα που επιλύει ο αλγόριθμος είναι η μεγιστοποίηση της συνάρτησης $Fitness : \Theta \rightarrow \mathbb{R}$, δηλαδή η εύρεση ενός σημείου :

$$\mathbf{x}^* = \max_{\mathbf{x} \in \Theta} (Fitness(\mathbf{x})) \quad (1.11)$$

όπου $\Theta \subseteq \mathbb{R}^D$ ο χώρος λύσεων του αλγορίθμου (search space). Αν το πρόβλημα ήταν η ελαχιστοποίηση μιας συνάρτησης $f(\mathbf{x})$ τότε η συνάρτηση καταλληλότητας (Fitness Function) ορίζεται ως εξής:

$$Fitness(\mathbf{x}) = \begin{cases} \frac{1}{1+f(\mathbf{x})}, & \text{αν } f(\mathbf{x}) \geq 0 \\ 1+|f(\mathbf{x})|, & \text{αν } f(\mathbf{x}) < 0 \end{cases} \quad (1.12)$$

Ο χώρος λύσεων Θ μπορεί να θεωρηθεί φραγμένος από τα D -διάστατα διανύσματα: $\mathbf{x}^{\min} = [x_1^{\min}, \dots, x_D^{\min}]^T$, $\mathbf{x}^{\max} = [x_1^{\max}, \dots, x_D^{\max}]^T$ ως:

$$\Theta = \{ \mathbf{x} = [x_1, x_2, \dots, x_D] \in \mathbb{R}^D / x_j^{\min} \leq x_j \leq x_j^{\max}, \forall j = 1, 2, \dots, D \} \quad (1.13)$$

Στον αλγόριθμο ABC μια αποικία από μέλισσες-εξερευνητές ψάχνουν για πλούσιες πηγές τροφής δηλαδή καλές λύσεις σε ένα πρόβλημα βελτιστοποίησης συνάρτησης. Αφού πρώτα αρχικοποιηθούν κάποιες τροφές-λύσεις στην τύχη, μετά το τεχνητό σμήνος μελισσών βελτιώνει επαναληπτικά τις θέσεις που έχουν βρεθεί ακολουθώντας τις δύο εξής

στρατηγικές: (α) κίνηση προς καλύτερες γειτονικές λύσεις και (β) εγκαταλείποντας τροφές/λύσεις που δεν είναι τόσο καλές.

Η δομή του αλγορίθμου είναι η ακόλουθη και ακολουθεί η ανάλυσή του.

Αλγόριθμος ABC:

Αρχικοποίηση

Επανάλαβε:

 Φάση των απασχολούμενων-μελισσών (Employed Bees Phase)

 Φάση των θεατών-μελισσών (Onlooker Bees Phase)

 Φάση των ανιχνευτών-μελισσών (Scout Bees Phase)

 Αποθήκευσε την καλύτερη ως τώρα λύση

Έως ότου: να πραγματοποιηθεί ένας μέγιστος αριθμός επαναλήψεων ή να ικανοποιηθεί κάποιο άλλο κριτήριο τερματισμού.

Το πλήθος των πηγών τροφής συμβολίζεται με SN (Solutions Number), το μέγεθος της αποικίας με CS (Colony Size) και ισχύει ότι το πλήθος της αποικίας είναι διπλάσιο από τις πηγές τροφής. Το πλήθος της αποικίας των μελισσών χωρίζεται σε δύο ισοπληθείς κατηγορίες, τις απασχολούμενες-μέλισσες και τις θεατές-μέλισσες. Δηλαδή ισχύει ότι:

$$SN = \frac{CS}{2} = (\text{πλήθος απασχολούμενων μελισσών}) = (\text{πλήθος θεατών μελισσών}) \quad (1.14)$$

Αρχικοποίηση

Έχοντας επιλέξει το μέγεθος της αποικίας CS , δημιουργούνται οι αρχικές πηγές τροφής θεωρώντας τυχαία σημεία ομοιόμορφα κατανεμημένα στο χώρο εύρεσης λύσεων ως [55, 56]:

$$x_{ij} = x_j^{\min} + rand(0,1)(x_j^{\max} - x_j^{\min}) \quad (1.15)$$

όπου $1 \leq i \leq SN, 1 \leq j \leq D$, $SN = \frac{CS}{2}$ το πλήθος των πηγών τροφής (Solutions Number), CS το πλήθος όλων των μελισσών της κυνέλης (Colony Size) και $rand(0,1)$ ένας τυχαίος αριθμός που ακολουθεί ομοιόμορφη κατανομή $\sim U(0,1)$.

Σε αυτό το στάδιο οι μετρητές οι οποίοι μας δείχνουν πόσες φορές κάποια πηγή τροφής έχει επισκεφτεί από τις μέλισσες παίρνουν την τιμή μηδέν ($trial_i = 0, i = 1, 2, \dots, SN$). Σκοπός αυτών των μετρητών είναι να δείξουν κάποια στιγμή μέσα στις επαναλήψεις ότι κάποια συγκεκριμένη πηγή τροφής έχει εξαντληθεί.

Φάση των απασχολούμενων-μελισσών (Employed Bees Phase)

Μια απασχολούμενη μέλισσα παράγει μια διαφοροποίηση της τροφής/λύσης και ανακαλύπτει έτσι μια κοντινή θέση τροφής υπολογίζοντας το νέκταρ (την τιμή της συνάρτησης f) στη νέα αυτή θέση. Έτσι παράγει μια κοντινή θέση \mathbf{v}_i αλλάζοντας την j -στή συνιστώσα της, χρησιμοποιώντας τον τύπο [55]:

$$v_{ij} = x_{ij} + \varphi_{ij} (x_{ij} - x_{kj}) \quad (1.16)$$

όπου j ένας τυχαίος αριθμός από τους $1, 2, \dots, D$, ο k είναι ένας τυχαίος αριθμός από τους $1, 2, \dots, SN$ ο οποίος πρέπει να είναι διαφορετικός του i (γιατί διαφορετικά δεν θα άλλαζε η θέση) και φ_{ij} ένας τυχαίος αριθμός $\varphi_{ij} = rand(-1, 1) \sim U(-1, 1)$.

Στην προηγούμενη εξίσωση αν η θέση \mathbf{v}_i που προκύπτει βρίσκεται εκτός των ορίων του χώρου εύρεσης λύσεων (1.13) τότε αυτή αντικαθίσταται από τα όρια αυτά, δηλαδή αν $v_{ij} > x_j^{\max}$ τότε $v_{ij} = x_j^{\max}$ και αν $v_{ij} < x_j^{\min}$ τότε $v_{ij} = x_j^{\min}$.

Αφού έχει υλοποιηθεί η θέση \mathbf{v}_i , στη συνέχεια υπολογίζεται η τιμή της αντικειμενικής συνάρτησης $f(\mathbf{v}_i)$ σε αυτή τη θέση και αν αυτή είναι καλύτερη της ήδη υπάρχουσας $f(\mathbf{x}_i)$ τότε η i -στή μέλισσα θυμάται τη θέση \mathbf{v}_i και ξεχνάει την προηγούμενη θέση της \mathbf{x}_i . Αλλιώς η μέλισσα αυτή παραμένει στη θέση που ήταν, δηλαδή θυμάται ως καλύτερη θέση την \mathbf{x}_i και ο μετρητής αυτής της τροφής αυξάνεται κατά μια μονάδα ($trial_i = trial_i + 1$).

Αυτό το στάδιο του αλγόριθμου πραγματοποιείται για όλες τις απασχολούμενες μέλισσες.

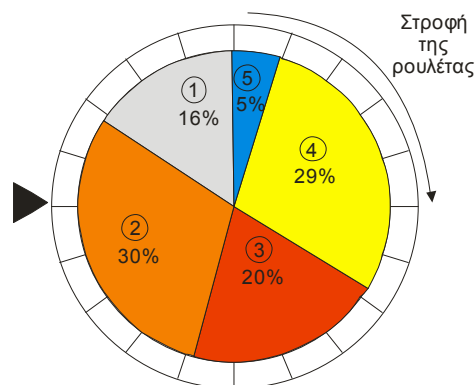
Φάση των θεατών-μελισσών (Onlooker Bees Phase)

Αφού πραγματοποιηθεί η προηγούμενη φάση του αλγόριθμου, οι απασχολούμενες μέλισσες μοιράζονται τις πληροφορίες για τις θέσεις των τροφών τους και για την ποσότητα τροφής της κάθε θέσης με τις μέλισσες-θεατές. Τότε η κάθε μέλισσα-θεατής υπολογίζει τις ποσότητες τροφής για όλες τις απασχολούμενες μέλισσες υπολογίζοντας την πιθανότητα της κάθε $i = 1, \dots, SN$ θέσης τροφής ως:

$$p_i = \frac{Fitness_i}{\sum_{i=1}^{SN} Fitness_i} \quad (1.17)$$

και μετά εφαρμόζεται η μέθοδος του τροχού της τύχης (roulette wheel selection) στηριζόμενη στις πιθανότητες p_i ώστε να υπολογιστεί πόσες μέλισσες-θεατές θα επισκεφτούν την i -στή θέση τροφής. Με αυτόν τον τρόπο εξασφαλίζεται ότι όσο περισσότερη είναι η ποιότητα τροφής τόσο περισσότερες θεατές-μέλισσες θα την επισκεφτούν.

Η μέθοδος του τροχού της τύχης (roulette wheel selection) είναι μια διαδικασία η οποία επιλέγει τυχαίους αριθμούς που δεν έχουν όλοι την ίδια πιθανότητα να επιλεγθούν αλλά κάποια πιθανότητα επιλογής. Για να γίνει πιο κατανοητή η μέθοδος ακολουθεί το εξής παράδειγμα: έστω οι κατηγορίες ①, ②, ③, ④, ⑤ με πιθανότητες να επιλεγούν 16%, 30%, 20%, 29% και 5% αντίστοιχα. Τότε ο τρόπος επιλογής τους είναι σαν να γυρίζουμε μια ρουλέτα όπου ο καθένας έχει πιθανότητα να επιλεγεί ίση με το αντίστοιχο ποσοστό του (Βλέπε Σχήμα 1.2).



Σχήμα 1.2 Ο κανόνας επιλογής του τροχού της τύχης (roulette wheel selection)

Στον αλγόριθμο ABC αποθηκεύεται στους υπολογισμούς μια παράμετρος $r_i \in [0,1]$ για κάθε πηγή τροφής, έτσι ώστε αν η πιθανότητα της i -στής θέσης τροφής που παράχθηκε από την εξίσωση (1.17) είναι μεγαλύτερη του r_i ($p_i > r_i$) τότε η θεατής μέλισσα επιλέγει μια κοντινή θέση της τροφής \mathbf{v}_i χρησιμοποιώντας την εξίσωση (1.16). Όπως και στο στάδιο των απασχολούμενων μελισσών έτσι κι εδώ, αν η θεατής-μέλισσα ανακαλύψει πως η νέα θέση \mathbf{v}_i είναι χειρότερη από τη ήδη υπάρχουσα, τότε παραμένει στην προηγούμενη θέση της. Αν η θέση δεν βελτιώνεται από τη νέα \mathbf{v}_i τότε ο μετρητής που δείχνει πόσες φορές επισκέφτηκε η i -στή θέση τροφής αυξάνει κατά μία μονάδα ($trial_i = trial_i + 1$).

Φάση των ανιχνευτών-μελισσών (Scout Bees Phase)

Αφού οι απασχολούμενες-μέλισσες κι οι μέλισσες-θεατές έχουν πραγματοποιήσει τα στάδια εύρεσης τροφής, ο αλγόριθμος ελέγχει να δει αν υπάρχει κάποια πηγή τροφής που να έχει εξαντληθεί με σκοπό να εγκαταλειφτεί η τροφή αυτή. Αν μια τροφή/λύση δεν βελτιώνεται περαιτέρω μετά από έναν προκαθορισμένο αριθμό επαναλήψεων L που ονομάζεται *όριο του ABC* (ABC limit), τότε αυτή η πηγή εγκαταλείπεται από την απασχολούμενη-μέλισσα που αντιστοιχεί στην τροφή αυτή και η μέλισσα αυτή γίνεται μέλισσα-ανιχνευτής. Το πότε η τροφή αυτή εξαντλείται για να γίνει η αλλαγή αυτή καθορίζεται από το αν ο αριθμός $trial_i$ έχει φτάσει το όριο του ABC. Αν \mathbf{x}_i η τροφή που θα εγκαταλειφθεί τότε η i -στή μέλισσα που εγκαταλείπει την πηγή αυτή παράγει μια νέα θέση τροφής ακολουθώντας την εξίσωση (1.15). Η τιμή του ορίου του ABC εξαρτάται από το πλήθος των πηγών τροφής και τη διάσταση του χώρου λύσεων και είναι :

$$(\text{όριο ABC}) = L = SN \cdot D \quad (1.18)$$

σύμφωνα με το [55]. Δηλαδή κατά τη διαδικασία αυτού του σταδίου αντί να βελτιώνεται η θέση όπως έγινε με την εξίσωση (1.16), ξαναδημιουργείται μια τελείως νέα λύση.

Η Gbest διαφοροποίηση του αλγορίθμου ABC (GABC)

Στη βιβλιογραφία έχουν υπάρξει πάρα πολλές διαφοροποιήσεις/μετατροπές του αλγορίθμου ABC. Μια μετατροπή που αξίζει να σημειωθεί είναι η παραμετροποίηση *Gbest* η οποία είναι εμπνευσμένη από τον Αλγόριθμο PSO. Σύμφωνα με την παραμετροποίηση αυτή των Zhu και Kwong [30] η εξίσωση (1.16) μετατρέπεται στην:

$$v_{ij} = x_{ij} + \varphi_{ij} (x_{ij} - x_{kj}) + \psi_{ij} (y_i - x_{ij}) \quad (1.19)$$

δηλαδή προστίθεται ο όρος $\psi_{ij} (y_i - x_{ij})$ όπου ψ_{ij} ένας τυχαίος αριθμός στο διάστημα $[0, C]$ με C ένα θετικό αριθμό και y_i η i -στή συνιστώσα της καλύτερης λύσης του αλγορίθμου. Αυτός ο νέος όρος παράγει μια θέση v_i η οποία δεν είναι απλά κοντινή θέση της προηγούμενης (όπως στην (1.16)) αλλά συγχρόνως κατευθύνεται και προς την διεύθυνση της καλύτερης λύσης y (ή \mathbf{g}^{best} στον αλγόριθμο PSO) από όλες όσες έχει βρει ο αλγόριθμος.

Ο αλγόριθμος ABC έχει πάρα πολλές εφαρμογές, ενδεικτικά αναφέρονται κάποιες από αυτές. Οι Song et al. στην [56] προσομοιώνουν τα επιφανειακά κύματα της γης (τα οποία προκαλούνται από σεισμούς εκρήξεις κ.ά.) με χρήση του αλγορίθμου ABC. Οι Cruz et al. στην [57] προτείνουν τον αλγόριθμο *cOptBees* ο οποίος συσταδοποιεί δεδομένα με παρόμοιο τρόπο με τον αλγόριθμο K-Means και στην [58] χρησιμοποιείται ο ίδιος αλγόριθμος από τους ίδιους συγγραφείς για να δημιουργήσει ένα ΤΝΔ ακτινικής βάσης το οποίο οι συγγραφείς το ονομάζουν ως *BeeRBF*. Οι Karaboga et al. στην [59] δημιουργούν ένα εμπροσθοτροφοδοτούμενο ΤΝΔ το οποίο εκπαιδεύεται με τη χρήση του ABC και οι Karaboga και Akay στην [60] εκπαιδεύουν ΤΝΔ με σκοπό τη χρήση του σε εφαρμογές επεξεργασίας σημάτων (signal processing). Τέλος οι Karaboga et al. αναλύουν διάφορες εφαρμογές του αλγορίθμου ABC στην [61].

1.2 Ορθογώνια Πολυώνυμα

Το αντικείμενο των ορθογωνίων πολυωνύμων είναι ένα κλασικό μαθηματικό αντικείμενο η προέλευση του οποίου μπορεί να αποδοθεί στο έργο Legendre για τις κινήσεις των πλανητών [62]. Έχουν σημαντικές εφαρμογές στη φυσική, στη θεωρία πιθανοτήτων, στη στατιστική και σε άλλους κλάδους των μαθηματικών. Η χρήση μεθόδων αριθμητικής ανάλυσης μέσω ηλεκτρονικών υπολογιστών στη θεωρία βελτιστοποίησης έχει ως συνέπεια την αναβίωση του ενδιαφέροντος για τα ορθογώνια πολυώνυμα [63].

Με τον όρο ορθογώνια πολυώνυμα εννοούμε μία οικογένεια πολυωνύμων $P_0(x), P_1(x), \dots, P_n(x), \dots$ που να ικανοποιεί τις ακόλουθες ιδιότητες [64] :

(α) κάθε πολυώνυμο $P_n(x)$ να είναι βαθμού n

(β) να ικανοποιείται η επόμενη σχέση:

$$\langle P_n(x), P_m(x) \rangle = \int_a^b P_n(x) P_m(x) w(x) dx = \begin{cases} 0, & \text{αν } n \neq m \\ c, & \text{αν } n = m \end{cases} \quad (1.20)$$

όπου c ένας θετικός πραγματικός αριθμός. Το σύμβολο \langle , \rangle αναπαριστά το εσωτερικό γινόμενο, όπως αυτό ορίζεται στην εξίσωση (1.20) και η $w(x)$ ονομάζεται συνάρτηση βαρύτητας (weighting function).

Τα πιο γνωστά ορθογώνια πολυώνυμα είναι τα Chebychev, Hermite, Legendre, Jacobi, Laguerre, Gegenbauer, Wilson, Askey-Wilson κ.ά. (Πίνακας 1.1) [63, 65-73].

Πίνακας 1.1 Τα πιο γνωστά ορθογώνια πολυώνυμα, τα διαστήματα στα οποία είναι ορθογώνια και οι συναρτήσεις βαρύτητας $w(x)$ της εξίσωσης (1.20)

Ορθογώνια πολυώνυμα	Διάστημα ορθογωνιότητας $[\alpha, \beta]$ ή (α, β) της εξίσωσης (1.20)	Συνάρτηση βαρύτητας $w(x)$
Chebyshev 1 ^{ου} είδους	$[-1, 1]$	$\frac{1}{\sqrt{1-x^2}}$
Chebyshev 2 ^{ου} είδους	$[-1, 1]$	$\sqrt{1-x^2}$
Chebyshev 3 ^{ου} είδους	$[-1, 1]$	$\sqrt{\frac{1+x}{1-x}}$
Chebyshev 4 ^{ου} είδους	$[-1, 1]$	$\sqrt{\frac{1-x}{1+x}}$
Hermite	\mathbb{R}	e^{-x^2}
Legendre	$[-1, 1]$	1
Jacobi	$(-1, 1)$	$(1-x)^\alpha (1+x)^\beta$
Laguerre	$[0, +\infty)$	e^{-x}
Gegenbauer	$[-1, 1]$	$(1-x^2)^{\alpha-\frac{1}{2}}$

Τα ορθογώνια πολυώνυμα έχουν πλήθος εφαρμογών σε προσεγγίσεις συναρτήσεων (βλέπε παράγραφο 2.2.3), μεθόδους παρεμβολής (βλ. παραγράφους 2.2.3, 3.2.3), στις σειρές Fourier, στην κατασκευή Τεχνητών Νευρωνικών δικτύων (βλ. παράγραφο 1.3.4), κ.ά. Ακολουθεί παράδειγμα της χρήσης των πολυωνύμων Chebyshev στις σειρές Fourier, διότι οι υπόλοιπες εφαρμογές αναλύονται σε άλλες παραγράφους του παρόν κειμένου.

Από τη θεωρία Fourier είναι γνωστό ότι μια προσέγγιση μιας συνεχούς συνάρτησης f στο διάστημα $[-\pi, \pi]$ μπορεί να δοθεί από το 'τριγωνομετρικό πολυώνυμο' [74, 75] :

$$p_n(x) = \frac{1}{2}a_0 + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx) \quad (1.21)$$

όπου οι συντελεστές είναι οι:

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx, \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \, dx \quad (1.22)$$

Στην εξίσωση (1.21) για μεγάλα n το $p_n(x)$ προσεγγίζει την συνάρτηση f , δηλαδή:

$$f(x) \approx \frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \quad (1.23)$$

Για να αποφευχθεί η προϋπόθεση της συνέχειας της συνάρτησης f αντί για τις σειρές Fourier χρησιμοποιούνται οι σειρές Chebyshev [76, 77] όπως περιγράφεται παρακάτω. Αν η συνάρτηση f ορίζεται στο διάστημα $[-1,1]$ τότε η σειρά Chebyshev αυτής είναι η :

$$f(x) \approx \frac{1}{2}a_0 + \sum_{k=1}^{\infty} a_k T_k(x) \quad (1.24)$$

όπου :

$$a_k = \frac{2}{\pi} \int_{-1}^1 \frac{1}{\sqrt{1-x^2}} f(x) T_k(x) dx, \quad k = 0, 1, 2, \dots \quad (1.25)$$

όπου $T_k(x)$ τα πολυώνυμα Chebyshev 1^{ου} είδους (βλ. παράγραφο 2.2.1).

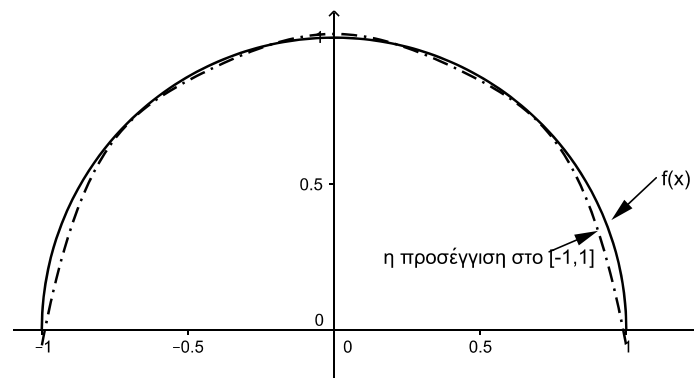
Αν για παράδειγμα πρέπει να προσεγγιστεί η συνάρτηση $f(x) = \sqrt{1-x^2}$ στο διάστημα $[-1,1]$ τότε οι συντελεστές της (1.25) είναι οι:

$$\left\{ \begin{array}{l} a_0 = \frac{4}{\pi} \\ a_{2k} = -\frac{4}{\pi(4k^2-1)}, \quad a_{2k-1} = 0, \quad k = 1, 2, \dots \end{array} \right\} \quad (1.26)$$

και τότε προκύπτει η προσέγγιση:

$$\sqrt{1-x^2} \approx \frac{2}{\pi} - \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{T_{2k}(x)}{4k^2-1} = \frac{2}{\pi} - \frac{4}{\pi} \left(\frac{1}{3} T_2(x) + \frac{1}{15} T_4(x) + \frac{1}{35} T_6(x) + \dots \right) \quad (1.27)$$

Αν στη σχέση (1.27) κρατηθούν οι όροι μέχρι και τον $T_6(x)$ τότε καταλήγουμε στην προσέγγιση που παρουσιάζεται στο Σχήμα 1.3.



Σχήμα 1.3 Η γραφική παράσταση της $f(x) = \sqrt{1-x^2}$ και της προσέγγισής της που παρουσιάζεται στην εξίσωση (1.27)

1.3 Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ)

Τα Τεχνητά Νευρωνικά Δίκτυα ΤΝΔ (Artificial Neural Networks- ANNs) είναι το αποτέλεσμα της προσπάθειας μίμησης της δομής του ανθρώπινου νευρικού συστήματος. Συνεπώς η φιλοσοφία των τεχνητών νευρωνικών δικτύων βασίζεται στην ανθρώπινη φύση και οι ομοιότητες τους με το νευρικό σύστημα του ανθρώπου δεν περιορίζονται μόνο στο όνομα. Εκτιμάται ότι υπάρχουν γύρω στα δέκα δισεκατομμύρια νεύρα στον ανθρώπινο εγκεφαλικό φλοιό και 100 τρισεκατομμύρια συνδέσεις (συνάψεις). Κατά την γέννηση του οργανισμού ο εγκέφαλος έχει ευρεία δομή και την ικανότητα να δημιουργεί τους δικούς του κανόνες, τους οποίους εμείς αναφέρουμε συχνά ως εμπειρία. Πράγματι, η εμπειρία είναι μια μακροχρόνια διαδικασία με την πιο σημαντική ανάπτυξη να λαμβάνει χώρα στα δύο πρώτα χρόνια από την γέννηση. Κατά τη διάρκεια του σταδίου αυτού, διαμορφώνονται γύρω στις ένα εκατομμύριο συνάψεις το δευτερόλεπτο. Οι συνάψεις είναι δομικά και λειτουργικά στοιχεία του εγκεφάλου που μεσολαβούν μεταξύ των αντιδράσεων των νευρών. Η πιο κοινή σύναψη είναι η χημική η οποία μετατρέπει τα ηλεκτρικά σήματα που δέχεται από ένα νεύρο σε χημικά για να τα μετατρέψει εκ νέου σε ηλεκτρικά και να τα μεταδώσει σε ένα άλλο νεύρο. Θεωρείται ότι οι συνάψεις εξάπτουν ή διεγείρουν τα νεύρα.

Στην προσπάθεια προσομοίωσης της λειτουργίας αυτής του ανθρώπινου εγκεφάλου δημιουργήθηκαν τα ΤΝΔ. Για ορισμό του ΤΝΔ μπορούμε να δώσουμε τον παρακάτω [78] :

Ένα ΤΝΔ είναι ένας μαζικός παράλληλος επεξεργαστής που αποτελείται από απλές μονάδες επεξεργασίας, ο οποίος έχει ιδιότητα της αποθήκευσης βιωματικής γνώσης και τη διάθεσή της για χρήση. Μοιάζει με τον εγκέφαλο σε δύο σημεία:

- 1. Η γνώση αποκτάται από το δίκτυο μέσω του περιβάλλοντος του, μέσα από μια διαδικασία μάθησης.*
- 2. Οι συνδέσεις εντός των νευρώνων, γνωστές και ως συνοπτικά βάρη, χρησιμοποιούνται για την αποθήκευση της αποκτηθείσας γνώσης.*

Ένα ΤΝΔ είναι ένας μηχανισμός, ο οποίος έχει τη δυνατότητα να ‘μαθαίνει’ και να χρησιμοποιεί την πληροφορία αυτή. Η διαδικασία κατά την οποία το δίκτυο ‘μαθαίνει’ ονομάζεται διαδικασία εκπαίδευσης (training) και είναι στην πραγματικότητα ο υπολογισμός των συναπτικών βαρών. Η διαδικασία κατά την οποία χρησιμοποιείται η γνώση που έχει

αποκτήσει το ΤΝΔ για να πάρουμε μια τιμή εξόδου ονομάζεται ανάκληση. Τέλος η διαδικασία που εφαρμόζουμε για να δούμε αν ένα δίκτυο δίνει τα επιθυμητά αποτελέσματα σε νέες καταστάσεις (νέα δεδομένα) ονομάζεται έλεγχος ή επαλήθευση (testing) του ΤΝΔ.

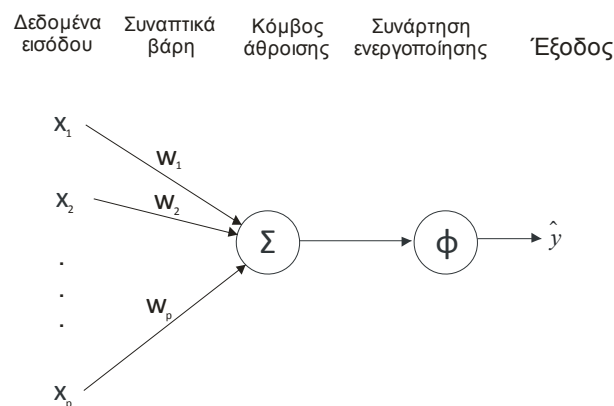
Η βάση ενός ΤΝΔ είναι ο νευρώνας, η βασική δομή του οποίου φαίνεται στο Σχήμα 1.4. Στο σχήμα αυτό ξεχωρίζουμε τα εξής :

- τις p στο πλήθος εισόδους (δεδομένα $\mathbf{x}=[x_1, x_2, \dots, x_p]^T$), οι οποίοι είναι παρομοίωση των ερεθισμάτων που δέχεται ο άνθρωπος από το περιβάλλον του
- τις p στο πλήθος συνάψεις με τα συναπτικά βάρη τους (w_1, w_2, \dots, w_p) που συνδέουν το νευρώνα με τα δεδομένα εξόδου. Είναι η προσομοίωση των συνάψεων των κυττάρων του ανθρώπινου εγκεφάλου.
- τον κόμβο άθροισης που αθροίζει τα γινόμενα $x_j w_j$
- τη συνάρτηση ενεργοποίησης φ .
- το δεδομένο εξόδου ή εκτιμώμενη έξοδος \hat{y} .

Η έξοδος του νευρώνα αυτού του σχήματος είναι :

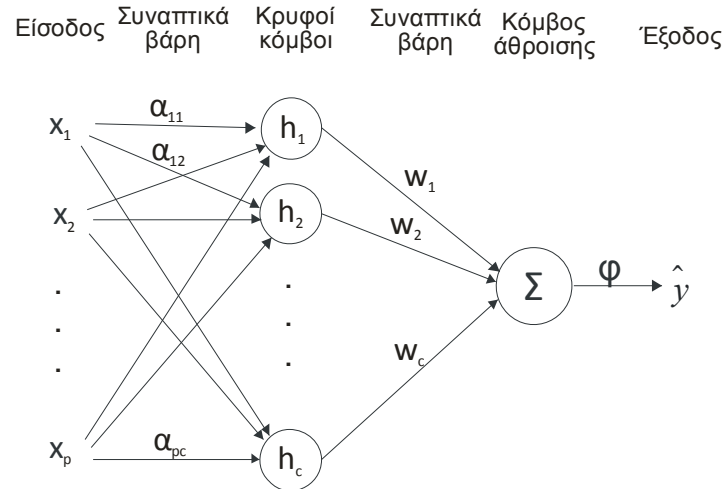
$$\hat{y} = \varphi \left(\sum_{j=1}^p x_j w_j \right) \quad (1.28)$$

Όπου η συνάρτηση φ ονομάζεται συνάρτηση ενεργοποίησης (activation function) του νευρώνα.



Σχήμα 1.4 Η δομή ενός νευρώνα με p εισόδους (τις x_1, x_2, \dots, x_p), και μία έξοδο (\hat{y}). Μεταξύ τους βλέπουμε τις p στο πλήθος συνάψεις μεταξύ των δεδομένων εισόδου και της εξόδου του νευρώνα.

Ένα ΤΝΔ μονοδιάστατης εξόδου με ένα κρυφό στρώμα αποτελείται από c στο πλήθος νευρώνες (όπως αυτοί που περιγράφηκαν παραπάνω) και η τοπολογία του είναι αυτή του Σχήματος 1.5.



Σχήμα 1.5 Η τοπολογία ενός ΤΝΔ με p εισόδους (τις x_1, x_2, \dots, x_p), c κρυφούς κόμβους και μία έξοδο \hat{y} . Μεταξύ τους βλέπουμε τις $c \cdot p$ στο πλήθος συνάψεις μεταξύ των δεδομένων εισόδου και των κρυφών κόμβων του δικτύου και τις c στο πλήθος συνάψεις μεταξύ των κρυφών κόμβων του δικτύου και των δεδομένων εξόδου.

Στο ΤΝΔ αυτό τα δεδομένα εισόδου αθροίζονται μέσω των συναπτικών βαρών $\alpha_{ji}, 1 \leq j \leq p, 1 \leq i \leq c$ και εισέρχονται μέσα στους c κρυφούς κόμβους οι οποίοι υπολογίζουν τις ποσότητες $h_i, i = 1, 2, \dots, c$. Για την δημιουργία των δεδομένων h_i δε θα δοθεί περαιτέρω ανάλυση μιας και αυτές έχουν διαφορετικό τρόπο δημιουργίας αν το ΤΝΔ είναι συναρτήσεων ακτινικής βάσης, εμπροσθοτροφοδοτούμενο, functional link, πολυωνυμικού ή άλλου τύπου. Κατόπιν τα δεδομένα h_i αθροίζονται μέσω των συναπτικών βαρών $w_i, i = 1, 2, \dots, c$ και τότε η εκτιμώμενη έξοδος του ΤΝΔ είναι:

$$\hat{y} = \phi \left(\sum_{i=1}^c w_i \cdot h_i \right) \quad (1.29)$$

όπου η συνάρτηση ϕ ονομάζεται συνάρτηση ενεργοποίησης του ΤΝΔ.

Κατά τη διαδικασία εκπαίδευσης έχουμε ένα σύνολο N δεδομένων της μορφής:

$$S = \left\{ (\mathbf{x}_k, y_k) / \mathbf{x}_k \in \mathbb{R}^p, y_k \in \mathbb{R}, k = 1, 2, \dots, N \right\} \quad (1.30)$$

και η εκπαίδευση του ΤΝΔ βασίζεται στην ελαχιστοποίηση των παρακάτω συναρτήσεων ως προς τα συναπτικά βάρη:

- Το ολικό τετραγωνικό σφάλμα του ΤΝΔ:

$$J_{SE} = \sum_{k=1}^N (y_k - \hat{y}_k)^2 \quad (1.31)$$

όπου y_k η μετρούμενη έξοδος και \hat{y}_k αυτή που εκτίμησε το ΤΝΔ.

- Το μέσο τετραγωνικό σφάλμα:

$$MSE = \frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{N} \quad (1.32)$$

το οποίο μετράει το σφάλμα ανά δεδομένο εξόδο.

- Η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος:

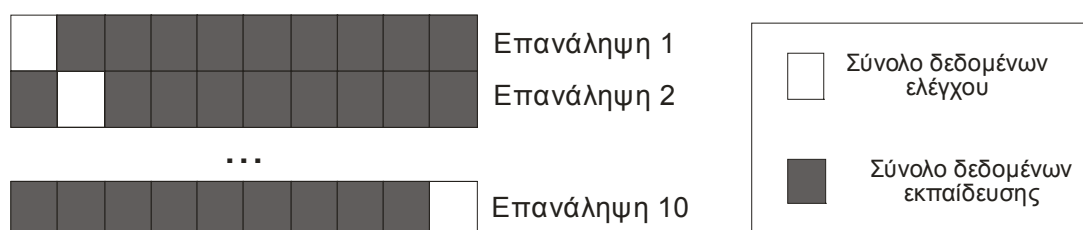
$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{N}} \quad (1.33)$$

η οποία μετράει το σφάλμα ανά δεδομένο εξόδο και το πλεονέκτημά της είναι πως έχει τις ίδιες μονάδες μέτρησης με αυτές της εξόδου y_k .

Προκειμένου να εκπαιδευτεί και να ελεγχθεί ένα ΤΝΔ υπάρχουν δύο μέθοδοι. Η πρώτη μέθοδος είναι ο διαχωρισμός του συνόλου δεδομένων σε σύνολο δεδομένων εκπαίδευσης (training dataset) και σύνολο δεδομένων ελέγχου (testing dataset), μπορούμε να διαλέξουμε ένα ποσοστό (π.χ. το 60%) του συνόλου δεδομένων ως δεδομένα εκπαίδευσης και τα υπόλοιπα δεδομένα ως σύνολο δεδομένων ελέγχου. Κατά τη διαδικασία αυτή εκπαιδεύεται το ΤΝΔ με τα δεδομένα εκπαίδευσης, κατόπιν εφαρμόζεται στα δεδομένα ελέγχου και υπολογίζεται η συνάρτηση του σφάλματός του στα δεδομένα εξόδο προκειμένου να αποφανθούμε αν η εκπαίδευσή του ήταν ικανοποιητική.

Δεύτερη μέθοδος είναι η ακόλουθη. Επειδή στην πρώτη μέθοδο που περιγράφηκε δεν υπάρχει κάποιος τρόπος ώστε να επιλεγθεί ποιο τμήμα των δεδομένων θα πρέπει να αποτελέσει το σύνολο των δεδομένων εκπαίδευσης και ποιο το σύνολο των δεδομένων ελέγχου, προτιμάται η μέθοδος k-fold cross validation [79, 80]. Στη μέθοδο k-fold cross validation (συνήθως $k = 10$ και η μέθοδος ονομάζεται 10-fold cross validation) χωρίζονται τα δεδομένα σε k υποσύνολα ίσα ή σχεδόν ίσα σε μέγεθος μεταξύ τους. Στη συνέχεια

εκτελούνται k επαναλήψεις εκπαίδευσης και ελέγχου ώστε σε κάθε επανάληψη ένα διαφορετικό υποσύνολο δεδομένων θα χρησιμοποιείται για τον έλεγχο ενώ τα υπόλοιπα $k-1$ υποσύνολα θα έχουν χρησιμοποιηθεί για την εκπαίδευση. Αφού εκτελεστεί η διαδικασία αυτή, καταλήγουμε στο μέσο όρο των σφαλμάτων εκπαίδευσης και ελέγχου. Το πλεονέκτημα της διαδικασίας αυτής είναι ότι όλα τα δεδομένα έχουν παίξει το ρόλο δεδομένων εκπαίδευσης και το ρόλο δεδομένων ελέγχου. Οι παραπάνω διαδικασίες ονομάζονται διαδικασίες ελέγχου ή επικύρωσης του ΤΝΔ (testing ή validation). Στο Σχήμα 1.6 παρουσιάζεται η επαναληπτική διαμέριση του συνόλου δεδομένων σε σύνολο δεδομένων εκπαίδευσης και ελέγχου με τη διαδικασία 10- fold cross validation.



Σχήμα 1.6 Ο αλγόριθμος 10- fold cross validation. Στην 1^η επανάληψη το πρώτο 10% των δεδομένων είναι το σύνολο ελέγχου, στην 2^η επανάληψη το σύνολο ελέγχου είναι το δεύτερο 10% των δεδομένων, κ.ο.κ.

Εφαρμογές των ΤΝΔ

Οι εφαρμογές των ΤΝΔ μπορούν να βρεθούν σε πολλούς και διαφορετικούς τομείς. Τα ΤΝΔ, εφαρμόζονται άριστα σε προβλήματα όπου συγκεκριμένοι κανόνες δεν υπάρχουν ή είναι δύσκολο να βρεθούν ή να εφαρμοστούν. Κάποια παραδείγματα τομέων εφαρμογής είναι στα ακόλουθα [81] :

- Ελέγχου συστημάτων για τον έλεγχο κινητήρων [82], τεχνητών βραχιόνων [83] αλλά και ολόκληρων διεργασιών.
- Επίλυσης προβλημάτων μαθηματικής βελτιστοποίησης [84].
- Οικονομικών συναλλαγών [85] και ιατρικής [86] για τη δημιουργία συστημάτων υποστήριξης αποφάσεων, μοντελοποίηση και εξομοίωση αυτών.
- Επεξεργασίας εικόνων: αναγνώριση χαρακτηριστικών, μετασχηματισμοί, συμπίεση και αποθήκευση τους [87-89].
- Αναγνώρισης προτύπων (ομιλίας, χαρακτήρων, εικόνας κ.ο.κ.), για την ανεύρεση συγκεκριμένων χαρακτηριστικών με ευρύ φάσμα εφαρμογών [87, 90, 91].
- Σύνθεσης και αναγνώρισης φυσικής ομιλίας [92, 93].
- Εξαγωγής πληροφοριών από βάσεις δεδομένων [94].

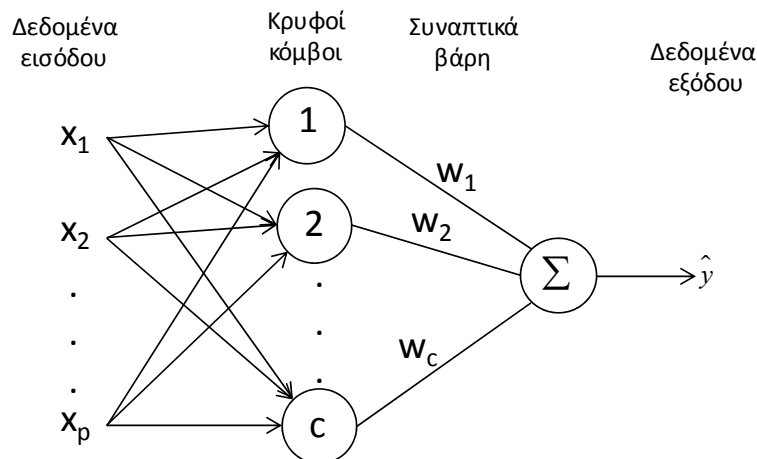
1.3.1 Τεχνικά Νευρωνικά Δίκτυα Συναρτήσεων Ακτινικής Βάσης (RBF ANNs)

Στα ΤΝΔ συναρτήσεων ακτινικής βάσης η συνάρτηση ενεργοποίησης φ της εξίσωσης (1.28) παίρνει τη μορφή :

$$\varphi = \varphi(\|\mathbf{x} - \mathbf{v}\|) \quad (1.34)$$

όπου το \mathbf{X} αναπαριστά τα δεδομένα εισόδου, το \mathbf{v} κάποια σημεία που έχουν οριστεί ως κέντρα και το σύμβολο $\|\dots\|$ κάποια νόρμα, συνήθως την Ευκλείδεια απόσταση. Άρα η διαφορά σε σχέση με το ΤΝΔ της προηγούμενης παραγράφου είναι ότι η συνάρτηση ενεργοποίησης εξαρτάται από τα κέντρα \mathbf{v} . Για μονοδιάστατη έξοδο $\hat{y} \in \mathbb{R}$ η τοπολογία αυτού του δικτύου είναι αυτή του Σχήματος 1.6 και η έξοδος του δικτύου αυτού είναι η:

$$\hat{y} = \sum_{i=1}^c w_i \varphi(\|\mathbf{x} - \mathbf{v}_i\|) \quad (1.35)$$



Σχήμα 1.7 Η τοπολογία ενός ΤΝΔ με p εισόδους (τις x_1, x_2, \dots, x_p), c κρυφούς κόμβους και μία έξοδο \hat{y} . Μεταξύ τους βλέπουμε τις $c \cdot p$ στο πλήθος συνάψεις μεταξύ των δεδομένων εισόδου και των κρυφών κόμβων του δικτύου και τις c στο πλήθος συνάψεις μεταξύ των κρυφών κόμβων του δικτύου και των δεδομένων εξόδου.

Η συνάρτηση ενεργοποίησης φ ονομάζεται συνάρτηση ακτινικής βάσης (radial basis function). Οι πιο συχνά χρησιμοποιούμενοι τύποι της συνάρτησης φ είναι οι [78] :

- Γκαουσιανή :

$$\varphi(d) = e^{-\alpha d^2}, \alpha > 0 \quad (1.36)$$

- Multiquadratic:

$$\varphi(d) = \sqrt{\alpha^2 + d^2}, \alpha > 0 \quad (1.37)$$

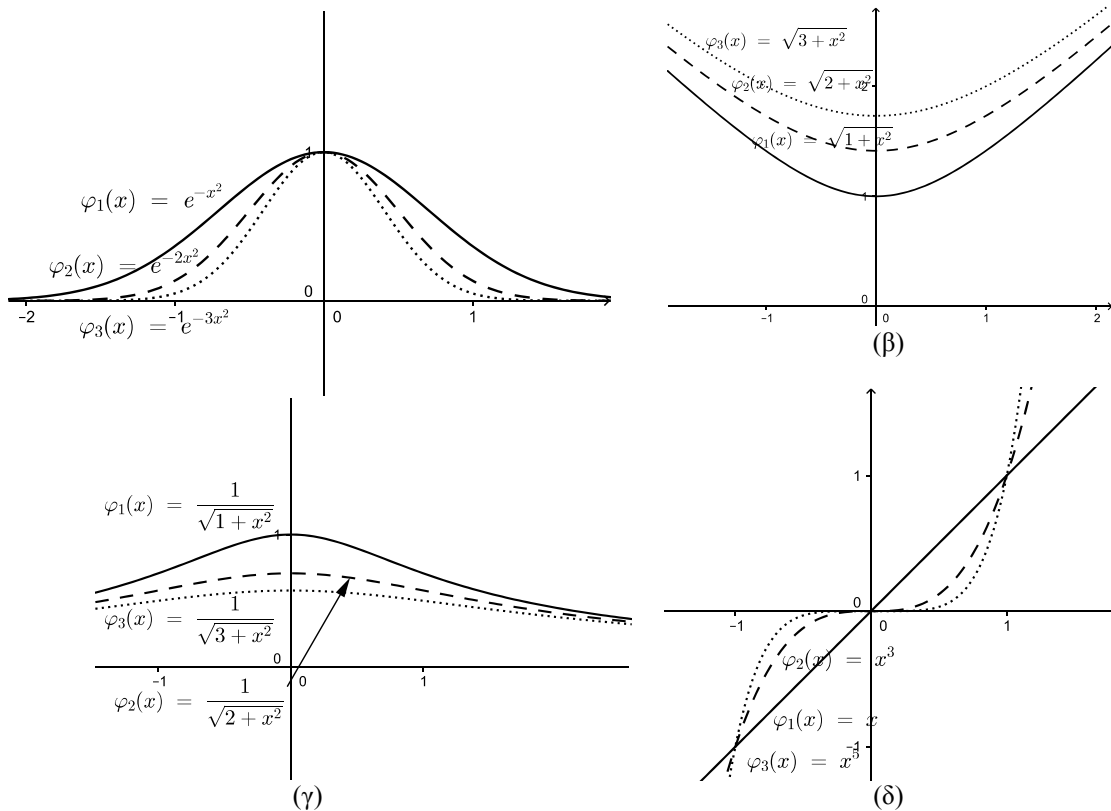
- Inverse Multiquadratic:

$$\varphi(d) = \frac{1}{\sqrt{\alpha^2 + d^2}}, \alpha > 0 \tag{1.38}$$

- Polyharmonic Spline

$$\varphi(d) = d^\alpha, \alpha = 1, 3, 5, \dots \tag{1.39}$$

όπου d έχουμε ορίσει την απόσταση $d = \|\mathbf{x} - \mathbf{v}\|$. Για διάφορες τιμές τις παραμέτρου α οι γραφικές παραστάσεις των συναρτήσεων $\varphi = \varphi(d)$ παρουσιάζονται στο Σχήμα 1.7.



Σχήμα 1.8 Οι γραφικές παραστάσεις από κάποιες συναρτήσεις ακτινικής βάσης. Στο (α) η γκαουσιανή, στο (β) η Multiquadric, στο (γ) η Inverse Multiquadric και στο (δ) η Polyharmonic Spline

Τα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης είναι νευρωνικά δίκτυα τριών στρώματων. Τα τρία αυτά στρώματα είναι τα επόμενα:

α) Στρώμα εισόδου: αποτελείται από νευρώνες οι οποίοι συνδέουν τα δεδομένα εισόδου $\mathbf{x} = (x_1, x_2, \dots, x_p)$ με το κρυφό στρώμα

β) Κρυφό στρώμα: είναι μια συνάρτηση ακτινικής βάσης, έχει c στο πλήθος στοιχεία (τους κρυφούς κόμβους) τα οποία είναι ίσα με το πλήθος c των κέντρων \mathbf{v} των εξισώσεων (1.34) και (1.35).

γ) Στρώμα εξόδου: εδώ γίνεται η άθροιση της εξίσωσης (1.35) και προκύπτει η έξοδος του δικτύου.

Παρακάτω ακολουθεί η ανάλυση της δομής ενός τυπικού ΤΝΔ ακτινικής βάσης. Έστω πως έχουμε τα δεδομένα εκπαίδευσης:

$$S = \{(\mathbf{x}_k, y_k) / \mathbf{x}_k \in \mathbb{R}^p, y_k \in \mathbb{R}, k = 1, 2, \dots, N\} \quad (1.40)$$

όπως περιγράφηκε και στη σχέση (1.30). Τότε αν επιλεγεί ως συνάρτηση ενεργοποίησης τη γκαουσιανή (ή αντίστοιχα οποιαδήποτε από αυτές που παρουσιάζονται στις εξισώσεις (1.37)-(1.39)) :

$$g_i(\mathbf{x}_k) = e^{-\left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\sigma_i}\right)^2} \quad (1.41)$$

η οποία έχει τη μορφή της εξίσωσης (1.36), με το $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kp}]$ να είναι ένα από τα δεδομένα εισόδου, τα $\mathbf{v}_i \in \mathbb{R}^p, i = 1, 2, \dots, c$ είναι τα κέντρα των πυρήνων και σ_i τα πλάτη των πυρήνων, τότε η έξοδος του δικτύου είναι η :

$$\hat{y}_k = \sum_{i=1}^c w_i g_i(\mathbf{x}_k) \quad (1.42)$$

όπου $w_i, i = 1, 2, \dots, c$ το συναπτικό βάρος (ή συντελεστής βαρύτητας) που συνδέει τον κάθε κρυφό κόμβο με την έξοδο του δικτύου.

1.3.1.1 Εκπαίδευση ΤΝΔ συναρτήσεων ακτινικής βάσης

Το μοναδικό που απομένει στο ΤΝΔ που περιγράφηκε παραπάνω είναι αυτό να εκπαιδευτεί, και με την έννοια εκπαίδευση εννοούμε των υπολογισμό των παραμέτρων $\{\mathbf{v}_i, \sigma_i, w_i\}, i = 1, 2, \dots, c$ έτσι ώστε το σφάλμα του δικτύου (βλ. εξισώσεις (1.31)-(1.33)) να γίνει όσο το δυνατόν μικρότερο.

Το πλήθος κρυφών κόμβων c θεωρούμε ότι έχει επιλεγεί εξ' αρχής. Γενικά η τάση των ΤΝΔ είναι πως όσο περισσότεροι είναι οι κρυφοί κόμβοι, τόσο μικρότερο αναμένεται να είναι το σφάλμα που θα προκύψει.

(α) Υπολογισμός των κέντρων των πυρήνων

Στα ΤΝΔ συναρτήσεων ακτινικής βάσης τα **κέντρα των πυρήνων** \mathbf{v}_i επιλέγονται από κάποιο αλγόριθμο συσταδοποίησης. Η συσταδοποίηση δεδομένων είναι η διαδικασία κατά την οποία ένα σύνολο δεδομένων χωρίζεται/διαμερίζεται σε συστάδες (clusters). Η συσταδοποίηση συνήθως επιτυγχάνεται με τον καθορισμό της ομοιότητας, ως προς προκαθορισμένα γνωρίσματα (συνήθως η θέση στον χώρο), μεταξύ των δεδομένων. Τα πιο όμοια μεταξύ τους δεδομένα ομαδοποιούνται στις ίδιες ομάδες. Το μοναδικό μειονέκτημα της συσταδοποίησης είναι ότι συνήθως δεν είναι αυτοματοποιημένο το πλήθος των ομάδων που θα προκύψει αλλά είναι κάτι που θα πρέπει να έχει επιλέξει εξ' αρχής ο χρήστης της μεθόδου. Η συσταδοποίηση χρησιμεύει στα ΤΝΔ ακτινικής βάσης, στην εξόρυξη δεδομένων [95, 96], στην επεξεργασία εικόνας [97, 98] κ.ά. Ακολουθεί α ανάλυση από τις τρεις πιο διαδεδομένες μεθόδους συσταδοποίησης, ο αλγόριθμος συσταδοποίησης k-means, ο fuzzy c-means (FCM) και ο Conditional Fuzzy C Means (CFCM).

Η συσταδοποίηση K-μέσων (K-Means clustering)

Η συσταδοποίηση K-μέσων ή αλγόριθμος του Lloyd [99] διαμερίζει -όπως κι οι υπόλοιποι αλγόριθμοι συσταδοποίησης- τα δεδομένα σε K το πλήθος ομάδες. Δεδομένου ενός συνόλου δεδομένων της μορφής:

$$(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N), \text{ με } \mathbf{x}_i \in \mathbb{R}^p \quad (1.43)$$

στόχος της συσταδοποίησης είναι να δημιουργηθεί ένα σύνολο K συστάδων με $K \ll N$, των:

$$S = \{S_1, S_2, \dots, S_K\} \quad (1.44)$$

οι οποίες χαρακτηρίζονται από τα ποια δεδομένα \mathbf{x}_i ανήκουν στην καθεμία από αυτές και από τα κέντρα της κάθε συστάδας $\mathbf{v}_j \in \mathbb{R}^p, j = 1, 2, \dots, K$ τα οποία είναι τα κέντρα βάρους των δεδομένων \mathbf{x}_i που ανήκουν στην καθεμία. Για να επιτευχθεί η συσταδοποίηση πρέπει να ελαχιστοποιηθεί το άθροισμα των αποστάσεων μεταξύ των κέντρων \mathbf{v}_j και των δεδομένων \mathbf{x}_i που ανήκουν στη συστάδα S_j :

$$J_{KMC} = \sum_{i=1}^N \sum_{\mathbf{x}_i \in S_j} \|\mathbf{x}_i - \mathbf{v}_j\|^2 \quad (1.45)$$

Χρησιμοποιώντας τη συνάρτηση συμμετοχής:

$$\mu_{ij} = \begin{cases} 1, & \mathbf{x}_i \in S_j \\ 0, & \text{αλλιώς} \end{cases} = \begin{cases} 1, & \text{αν } \|\mathbf{x}_i - \mathbf{v}_j\| = \min_{l=1,\dots,k} \{\|\mathbf{x}_i - \mathbf{v}_l\|\} \\ 0, & \text{αλλιώς} \end{cases} \quad (1.46)$$

η αντικειμενική συνάρτηση (1.45) παίρνει τη μορφή:

$$J_{KMC} = \sum_{i=1}^N \sum_{j=1}^K \mu_{ij} \|\mathbf{x}_i - \mathbf{v}_j\|^2 \quad (1.47)$$

η ελαχιστοποίηση της οποίας επιτυγχάνεται θεωρώντας ως κέντρα τα κέντρα βάρους (μέσος όρος διανυσμάτων) των σημείων \mathbf{x}_i της κάθε συστάδας, δηλαδή:

$$\mathbf{v}_j = \frac{\sum_{i=1}^N \mu_{ij} \mathbf{x}_i}{\sum_{i=1}^N \mu_{ij}} \quad (1.48)$$

Για να επιτευχθεί η συσταδοποίηση που περιγράφηκε, χρησιμοποιείται ο επόμενος αλγόριθμος.

Αλγόριθμος συσταδοποίησης K-means:

Βήμα 1^ο: Τυχαία επιλογή των αρχικών κέντρων συσταδοποίησης $\mathbf{v}_j, j=1,2,\dots,K$ (οι τυχαίες αρχικές τιμές των $\mathbf{v}_1,\dots,\mathbf{v}_c$ πρέπει να είναι εντός του εύρους των δεδομένων \mathbf{x}_i).

Βήμα 2^ο: Προσδιορισμός των συναρτήσεων συμμετοχής με βάση την εξίσωση (1.46) και κατάταξη των διανυσματικών δεδομένων στις συστάδες.

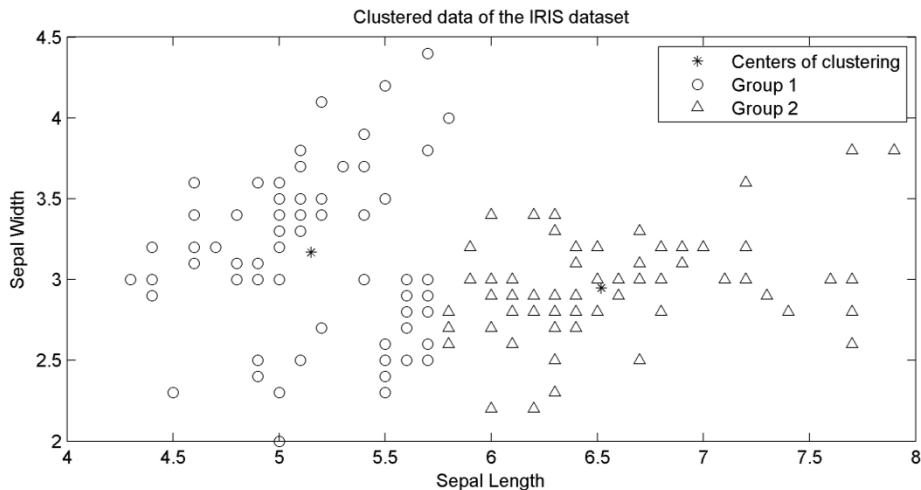
Επανάλαβε:

Βήμα 3^ο: Υπολογισμός των νέων κέντρων των συστάδων με βάση την εξίσωση (1.48).

Βήμα 4^ο: Προσδιορισμός των νέων συναρτήσεων συμμετοχής μ_{ij} με βάση την εξίσωση (1.46) και ανακατάταξη των διανυσματικών δεδομένων στις συστάδες.

Έως ότου τα νέα κέντρα συσταδοποίησης που προκύπτουν στο βήμα 3 να συγκλίνουν ή η επανάληψη να φτάσει κάποιο προκαθορισμένο μέγιστο αριθμό επαναλήψεων.

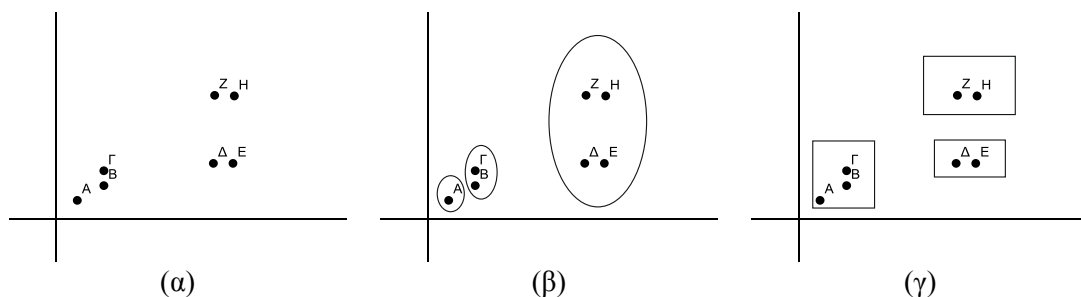
Ένα αποτέλεσμα της συσταδοποίησης για τυχαία δυσδιάστατα δεδομένα $\mathbf{x}_i \in \mathbb{R}^2$ παρουσιάζεται στο Σχήμα 1.8 όπου φαίνονται $N = 150$ δεδομένα να έχουν συσταδοποιηθεί σε $K = 2$ συστάδες.



Σχήμα 1.9 Αποτέλεσμα της συσταδοποίησης των $N = 150$ δεδομένων Fisher Iris του UCI repository [100] σε $K = 2$ συστάδες. Τα δεδομένα έχουν χωριστεί σε δύο συστάδες, στη μία συστάδα τα δεδομένα συμβολίζονται με το \circ και στην άλλη με Δ . Τα κέντρα της κάθε συστάδας συμβολίζονται ως: *

Ο προηγούμενος αλγόριθμος δεν απαιτεί μεγάλη υπολογιστική ισχύ και δεν χρειάζεται να κρατάει στη μνήμη του H/Y μεγάλο όγκο δεδομένων, το οποίο ισχύει και για μεγάλο πλήθος δεδομένων N .

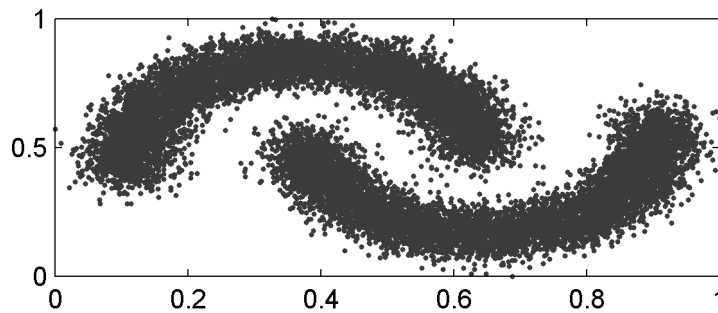
Το μεγάλο μειονέκτημα του είναι πως δεν εξάγει μοναδική λύση, αλλά εξαρτάται πολύ από την επιλογή των αρχικών κέντρων συσταδοποίησης (Βήμα 1^ο του παραπάνω αλγόριθμου), όπως φαίνεται και στο Σχήμα 1.9 στο οποίο υπάρχουν $N = 7$ σημεία που πρέπει να συσταδοποιηθούν σε $K = 3$ συστάδες. Αν επιλεγθούν τρία αρχικά κέντρα κοντά στα σημεία A, Γ, Z τότε θα καταλήξουμε στη συσταδοποίηση που φαίνεται με τις ελλείψεις στο (β) σχήμα ενώ αν επιλεγθούν τρία αρχικά κέντρα κοντά στα σημεία A, Z, Δ καταλήγει ο αλγόριθμος στις συστάδες του (γ) σχήματος με τα ορθογώνια. Άρα ο αλγόριθμος συσταδοποίησης δεν έχει μοναδική λύση και εξαρτάται από την αρχική επιλογή των κέντρων.



Σχήμα 1.10 Ευαισθησία του αλγόριθμου K-means στην αρχική επιλογή κέντρων συσταδοποίησης

Δεύτερο μειονέκτημα του αλγόριθμου είναι ότι το πλήθος K των κέντρων συσταδοποίησης δεν υπάρχει κάποιος τρόπος να επιλεγθεί εκ των προτέρων με κάποιο ‘σωστό τρόπο’ και άρα θα πρέπει να δοκιμάζονται διαφορετικές επιλογές της τιμής K .

Ένα τρίτο μειονέκτημα είναι ότι η απόδοση του αλγορίθμου δεν είναι ικανοποιητική όταν οι συστάδες έχουν πολύπλοκη μορφή, όπως για παράδειγμα αυτά που παρουσιάζονται στο Σχήμα 1.10.



Σχήμα 1.11 Δεδομένα τα οποία δε θα συσταδοποιούσε σωστά η συσταδοποίηση K -μέσων

Η ασαφής συσταδοποίηση c -μέσων (Fuzzy c -Means clustering - FCM)

Την ασαφή λογική την εισήγαγε ο Lotfi Zadeh το 1965 [40] ο οποίος στην προσπάθεια του να προσομοιώσει την ανθρώπινη γλώσσα με χρήση υπολογιστών εισήγαγε τα ασαφή σύνολα. Ενώ στην διακριτή λογική (crisp logic) η τιμή αλήθειας μιας πρότασης παίρνει τιμές 0 (Ψευδής πρόταση) και 1 (Αληθής), στην ασαφή λογική (fuzzy logic) μια πρόταση λογικής μπορεί να πάρει όλες τιμές στο διάστημα $[0,1]$. Πιο εκτενής αναφορά στην ασαφή λογική και τα ασαφή συστήματα δίνεται στην ενότητα 1.4.

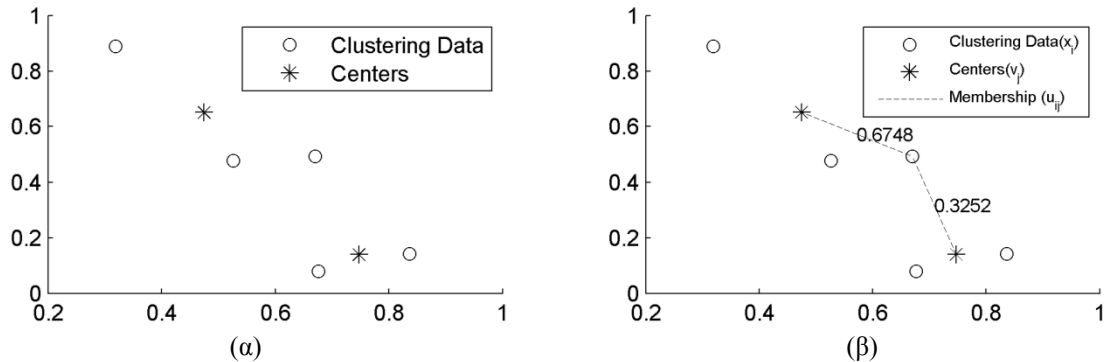
Η ασαφής συσταδοποίηση c κέντρων (fuzzy c means Clustering ή FCM) [101] είναι αντίστοιχη με τη συσταδοποίηση K -μέσων με την εξής διαφορά: στη συσταδοποίηση K -μέσων, κάθε δεδομένο $\mathbf{x}_i, i=1, \dots, N$ ανήκε μόνο σε ένα κέντρο συσταδοποίησης και αυτό φαίνεται από τη συνάρτηση συμμετοχής μ_{ij} της σχέσης (1.46) η οποία έπαιρνε τις τιμές 1 αν το δεδομένο \mathbf{x}_i ανήκε στη συστάδα κέντρου \mathbf{v}_j ή 0 αλλιώς, ενώ στην ασαφή συσταδοποίηση c κέντρων κάθε δεδομένο ανήκει σε όλα τα κέντρα συσταδοποίησης κατά ένα ποσοστό, οπότε η νέα συνάρτηση συμμετοχής είναι η:

$$u_{ij} \in [0,1], \quad i=1, \dots, N, \quad j=1, \dots, c \quad (1.49)$$

υπό τον περιορισμό :

$$\sum_{j=1}^c u_{ij} = 1, \forall i = 1, 2, \dots, N \quad (1.50)$$

του οποίου η ερμηνεία είναι ότι αν για το κάθε δεδομένο \mathbf{x}_i αθροιστούν τα ποσοστά που δείχνουν κατά πόσο ανήκει στη κάθε συστάδα κέντρου $\mathbf{v}_j, j = 1, \dots, c$ τότε το αποτέλεσμα θα έχει άθροισμα 1 (Σχήμα 1.12).



Σχήμα 1.12 Στην ασαφή συσταδοποίηση όλα τα δεδομένα ανήκουν σε όλες τις συστάδες κατά ένα ποσοστό. Στο (α) παρουσιάζονται τα δεδομένα και στο (β) το αποτέλεσμα της FCM συσταδοποίησής τους. Στο (β) παρουσιάζεται ένα δεδομένο το οποίο ανήκει κατά 67.48% στην μία συστάδα και κατά 32.52% στην άλλη.

Σκοπός της ασαφούς συσταδοποίησης c κέντρων είναι η ελαχιστοποίηση της αντικειμενικής συνάρτησης :

$$J_{FCM} = \sum_{j=1}^c \sum_{i=1}^N (u_{ij})^m \|\mathbf{x}_i - \mathbf{v}_j\|^2 \quad (1.51)$$

υπό τον περιορισμό:

$$\sum_{j=1}^c u_{ij} = 1, \forall i = 1, 2, \dots, N \quad (1.52)$$

όπου η m ονομάζεται παράμετρος ασάφειας με $m \geq 1$ και συνήθη τιμή την $m = 2$, $\mathbf{x}_i \in \mathbb{R}^p, i = 1, 2, \dots, N$ τα δεδομένα που πρέπει να συσταδοποιηθούν, u_{ij} η συνάρτηση συμμετοχής που περιγράφηκε παραπάνω και $\mathbf{v}_j, j = 1, 2, \dots, c$ τα κέντρα που θα πρέπει να προκύψουν από τη συσταδοποίηση. Αν υποθέσουμε ότι τα u_{ij} είναι σταθερά τότε τα κέντρα της συσταδοποίησης θα δίνονται από την εξίσωση:

$$\mathbf{v}_j = \frac{\sum_{i=1}^N (u_{ij})^m \mathbf{x}_i}{\sum_{i=1}^N (u_{ij})^m} \quad (1.53)$$

ενώ υποθέτοντας ότι οι θέσεις των κέντρων \mathbf{v}_j είναι σταθερές τότε προκύπτει (με χρήση των πολλαπλασιαστών Lagrange) ότι οι τιμές των u_{ij} είναι:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|\mathbf{x}_i - \mathbf{v}_j\|}{\|\mathbf{x}_i - \mathbf{v}_k\|} \right)^{\frac{2}{m-1}}} \quad (1.54)$$

Με βάση τα παραπάνω ο αλγόριθμος της διαδικασίας συσταδοποίησης ασαφών c κέντρων είναι ο εξής [101, 102]:

Αλγόριθμος ασαφούς συσταδοποίησης c -κέντρων (FCM):

Βήμα 1^ο: Επιλογή του αριθμού c των ασαφών συστάδων, της παραμέτρου m και τυχαίων αρχικών τιμών για τα διανύσματα $\mathbf{v}_1, \dots, \mathbf{v}_c$ (οι τυχαίες αρχικές τιμές των $\mathbf{v}_1, \dots, \mathbf{v}_c$ πρέπει να είναι εντός των ορίων των δεδομένων \mathbf{x}_i).

Επανάλαβε:

Βήμα 2^ο: Υπολογισμός των συναρτήσεων συμμετοχής: $u_{ij}, i=1,2,\dots,N, j=1,2,\dots,c$ με χρήση της εξίσωσης (1.54)

Βήμα 3^ο: Προσδιορισμός των νέων κέντρων ασαφούς συσταδοποίησης με την εξίσωση (1.53).

Έως ότου: Να ικανοποιηθεί κάποιο κριτήριο τερματισμού, όπως τα νέα κέντρα που δημιουργούνται να μην διαφέρουν πολύ από τα προηγούμενα (να συγκλίνουν) ή να φτάσει ο αλγόριθμος σε κάποιον προκαθορισμένο μέγιστο αριθμό επαναλήψεων.

Τέλος, αξίζει να σημειωθεί ότι η ασαφής συσταδοποίηση c μέσων είναι γενικότερη περίπτωση της συσταδοποίησης K -μέσων.

Η υπο-συνθήκη ασαφής συσταδοποίησης C -μέσων (Conditional Fuzzy C-Means clustering - CFCM)

Ο Pedrycz εισήγαγε το 1996 τη μέθοδο του Conditional Fuzzy C-Means Clustering (CFCM) [103] η οποία είναι παραμετροποίηση της ασαφούς συσταδοποίησης C -μέσων (FCM) έτσι ώστε να συσταδοποιούνται τα δεδομένα εισόδου υπό κάποια συνθήκη (condition). Η

συνθήκη αυτή μπορεί να εξαχθεί από συσταδοποίηση των δεδομένων εξόδου [104]. Η διαδικασία της συσταδοποίησης αυτής είναι η ακόλουθη:

Αλγόριθμος CFCM

Βήμα 1^ο: Επιλέγονται c^{in} στο πλήθος κέντρα συσταδοποίησης για τα δεδομένα εισόδου \mathbf{X} και c^{out} στο πλήθος κέντρα για τα δεδομένα εξόδου y . Ορίζουμε ως συνολικό πλήθος συστάδων το γινόμενο $c = c^{in} \cdot c^{out}$. Ορίζουμε την παράμετρο ασάφειας m (συνήθως $m = 2$) και τη νόρμα $\|\cdot\|$ που θα χρησιμοποιηθεί (συνήθως η ευκλείδεια απόσταση $\|\cdot\|_2$).

Βήμα 2^ο: Χωρίζονται τα δεδομένα εξόδου $(y_i, i = 1, 2, \dots, N)$ σε c^{out} συστάδες κέντρων $v_j, j = 1, \dots, c^{out}$ με τον αλγόριθμο συσταδοποίησης fuzzy c-means (FCM) και υπολογίζεται ο πίνακας συμμετοχής τους (partition matrix) :

$$f_{ik} = \frac{1}{\sum_{j=1}^{c^{out}} \left(\frac{\|y_k - v_j\|}{\|y_k - v_i\|} \right)^{\frac{2}{m-1}}}, i = 1, 2, \dots, N, k = 1, 2, \dots, c^{out} \quad (1.55)$$

όπου m η παράμετρος ασαφούς συσταδοποίησης.

Βήμα 3^ο: Χωρίζονται τα δεδομένα εισόδου σε $c = c^{in} \cdot c^{out}$ συστάδες κέντρων v_i , παίρνοντας υπόψη τη συσταδοποίηση του χώρου εξόδου. Αυτό γίνεται χρησιμοποιώντας ως πίνακα συμμετοχής της συσταδοποίησης τον:

$$u_{ik} = \frac{f_{ik}}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{m-1}}} \quad (1.56)$$

Τα κέντρα συσταδοποίησης της εισόδου που προκύπτουν τότε είναι τα:

$$\mathbf{v}_i = \frac{\sum_{k=1}^N u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^N u_{ik}^m}, i = 1, 2, \dots, c \quad (1.57)$$

Βήμα 4^ο: Επαναληπτικά επαναυπολογίζονται οι συναρτήσεις συμμετοχής u_{ik} και τα κέντρα συσταδοποίησης V_i από τις δύο προηγούμενες εξισώσεις ώστε τα κέντρα συσταδοποίησης που προκύπτουν να συγκλίνουν (να μην αλλάζουν πολύ από τον προηγούμενο υπολογισμό τους).

(β) Υπολογισμός των τυπικών αποκλίσεων των πυρήνων

Για να επιλεγθούν τα πλάτη των πυρήνων (δηλ. οι τυπικές αποκλίσεις) σ_i υπάρχουν πολλές διαφορετικές τεχνικές στην βιβλιογραφία. Αναφέρονται παρακάτω κάποιες από αυτές:

- Οι Moody και Darken [105] προτείνουν ως πλάτος να χρησιμοποιηθεί η μέση απόσταση του κάθε κέντρου \mathbf{v}_i από τους τ κοντινότερους γείτονές του $\mathbf{v}_l, l \neq i$, δηλαδή:

$$\sigma_i = \frac{1}{\tau} \sqrt{\sum_{l=1}^{\tau} d_{il}^2} \quad (1.58)$$

όπου $d_{il} = \|\mathbf{v}_l - \mathbf{v}_i\|$, $l \neq i$ με τυπικές τιμές του τ τις $\tau=2$ και $\tau=3$.

- Οι Saha και Keeler [106] προτείνουν ως πλάτη να χρησιμοποιούνται η απόσταση του αντίστοιχου κέντρου \mathbf{v}_i από το κοντινότερό του κέντρο πολλαπλασιασμένο με μια θετική σταθερά m , δηλαδή:

$$\sigma_i = m \cdot \min_{l \neq i} \{\|\mathbf{v}_i - \mathbf{v}_l\|\} \quad (1.59)$$

με τυπική τιμή του m την $m = 2$.

- Οι Niros και Tsekouras [107] προτείνουν αφού γίνει ασαφής συσταδοποίηση των δεδομένων εισόδου με τον αλγόριθμο fuzzy c-means, και προκύψουν τα κέντρα $C_i, i=1,2,\dots,c$ και η συνάρτηση/πίνακας συμμετοχής u_{ik} της συσταδοποίησης, να κρατηθούν μόνο οι τιμές των u_{ik} πάνω από κάποιο κατώφλι (threshold), δηλαδή τα στοιχεία για τα οποία $u_{ik} \geq \xi$, με $\xi \in (0,1)$ ένα θετικό αριθμό (π.χ. $\xi = 0.05$). Τότε δημιουργείται μια υβριδική συσταδοποίηση με νέες συστάδες τις: $G_i = \{x_k \in C_i / u_{ik} \geq \xi\}$ τότε ως πλάτη μπορούν να χρησιμοποιηθούν τα:

$$\sigma_i = \frac{2}{3} \cdot \beta \cdot \max_{(x_k, v_i) \in G_i} \{\|x_k - v_i\|\}, i=1,2,\dots,c \quad (1.60)$$

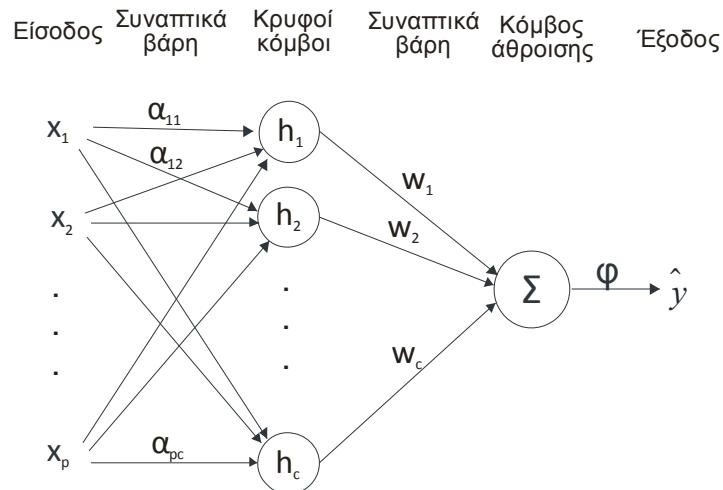
όπου $\beta > 0$ με συνήθη τιμή του β το $\beta=1$.

(γ) Υπολογισμός των συναπτικών βαρών του δικτύου

Τα συναπτικά βάρη w_i πρέπει να επιλεγούν με τρόπο ώστε να ελαχιστοποιείται το σφάλμα του δικτύου (βλ. εξισώσεις (1.31)-(1.33)). Για να γίνει αυτό χρησιμοποιείται μια μέθοδος βελτιστοποίησης (optimization) με συνηθέστερες να είναι η μέθοδος των ελαχίστων τετραγώνων (least squares) μιας και το σφάλμα είναι γραμμικό ως προς τα βάρη w_i και η μέθοδος Gradient Descent που περιγράφηκαν σε προηγούμενη παράγραφο. Στη βιβλιογραφία επίσης χρησιμοποιούνται οι αλγόριθμοι νοημοσύνης σμήνους (βλ. παράγραφο 1.1.3) για τον προσδιορισμό των συναπτικών βαρών καθώς και εξελικτικοί αλγόριθμοι (evolutionary algorithms) [108-110].

1.3.2 Εμπροσθοτροφοδοτούμενα ΤΝΔ (FeedForward NN, FFNN)

Μια άλλη κατηγορία ΤΝΔ είναι τα Εμπροσθοτροφοδοτούμενα ΤΝΔ (FeedForward ANN ή FFNN). Η τοπολογία ενός τέτοιου δικτύου με μονοδιάστατη έξοδο παρουσιάζεται στο Σχήμα 1.13.



Σχήμα 1.13 Η μορφή ενός FeedForward ΤΝΔ με p εισόδους (τις x_1, x_2, \dots, x_p), c κρυφούς κόμβους και μία έξοδο (\hat{y}). Μεταξύ τους βλέπουμε τις $c \cdot p$ στο πλήθος συνάψεις μεταξύ των δεδομένων εισόδου και των κρυφών κόμβων του δικτύου.

Το ΤΝΔ αυτό αποτελείται τοπολογικά από τα εξής επίπεδα [34, 78]:

- **Επίπεδο εισόδου:** εδώ εισάγονται τα δεδομένα μέσα στο ΤΝΔ και οι συνάψεις τους τα συνδέουν με το επόμενο στρώμα του δικτύου
- **Επίπεδο κρυφών κόμβων:** είναι c στο πλήθος και αθροίζουν τα δεδομένα εισόδου μέσω των συναπτικών βαρών α_{ji} με $j=1,2,\dots,p, i=1,2,\dots,c$ όπου p η διάσταση του χώρου εισόδου. Τότε ο i -στος κρυφός κόμβος υπολογίζει τις ποσότητες:

$$h_i = \sum_{j=1}^p x_j \alpha_{ji}, \quad i=1,2,\dots,c \quad (1.61)$$

- **Επίπεδο εξόδου:** εδώ προκύπτει η εκτιμώμενη έξοδος του δικτύου. Η εκτιμώμενη έξοδος \hat{y} προκύπτει αφού οι ποσότητες h_i αθροιστούν μέσω των συναπτικών βαρών w_i ,

$i = 1, 2, \dots, c$ και κατόπιν στο άθροισμα αυτό εφαρμοστεί η συνάρτηση φ . Δηλαδή η εκτιμώμενη έξοδος είναι η :

$$\hat{y} = \varphi \left(\sum_{i=1}^c w_i h_i \right) \quad (1.62)$$

Η συνάρτηση φ ονομάζεται συνάρτηση ενεργοποίησης και έχει συνήθως τις εξής μορφές :

- Λογιστική συνάρτηση:

$$\varphi(x) = \frac{1}{1 + e^{-\alpha x}}, \quad \alpha > 0 \quad (1.63)$$

- Σιγμοειδής (η λογιστική με παράμετρο $\alpha = 1$, βλ. Σχήμα 1.14) :

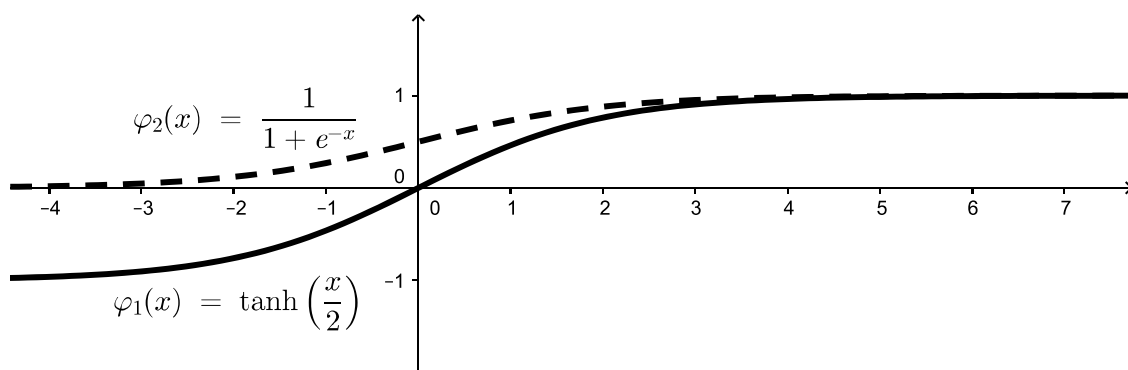
$$\varphi(x) = \frac{1}{1 + e^{-x}} \quad (1.64)$$

- Υπερβολική εφαπτομένη:

$$\varphi(x) = \alpha \tanh(\beta x) \quad (1.65)$$

η οποία με τυπικές τιμές τις $\alpha = 1$ και $\beta = \frac{1}{2}$ γίνεται η (βλ. Σχήμα 1.14) :

$$\varphi(x) = \tanh \frac{x}{2} = \frac{e^{x/2} - e^{-x/2}}{e^{x/2} + e^{-x/2}} \quad (1.66)$$



Σχήμα 1.14 Οι γραφικές παραστάσεις από δύο συνηθισμένες συναρτήσεις ενεργοποίησης ενός FeedForward ΤΔΝ.

Εκπαίδευση ενός Εμπροσθοτροφοδοτούμενου ΤΝΔ

Κατά την διαδικασία εκπαίδευσης του FFNN σκοπός είναι να βρεθούν τα συναπτικά βάρη $\alpha_{ji}, j=1,2,\dots,p, i=1,2,\dots,c$ και w_1,\dots,w_c ώστε το σφάλμα που προκύπτει από το FFNN (εξισώσεις (1.31)-(1.33)) να ελαχιστοποιείται. Η διαδικασία αυτή επιτυγχάνεται χρησιμοποιώντας μια οποιαδήποτε μέθοδο βελτιστοποίησης (optimization), συνήθως χρησιμοποιώντας τη μέθοδο Gradient Descent όπως αυτή αναλύεται παρακάτω.

Για να χρησιμοποιηθεί η μέθοδος Gradient Descent για την εκπαίδευση του FFNN, τότε αρκεί στη μέθοδο που περιγράφηκε στην παράγραφο 1.1.2 να χρησιμοποιηθεί ως συνάρτηση f της εξίσωσης (1.5) η συνάρτηση των συναπτικών βαρών:

$$f(\mathbf{a}; \mathbf{w}) = f(\alpha_{11}, \alpha_{12}, \dots, \alpha_{pc}, w_1, \dots, w_c) = J_{SE} = \sum_{k=1}^N (y_k - \hat{y}_k)^2 \quad (1.67)$$

η οποία προσδιορίζει το ολικό τετραγωνικό σφάλμα του ΤΝΔ και η βαθμίδα της (gradient) τότε είναι η:

$$\nabla f = \left(\frac{\partial f}{\partial \alpha_{11}}, \dots, \frac{\partial f}{\partial \alpha_{pc}}, \frac{\partial f}{\partial w_1}, \dots, \frac{\partial f}{\partial w_c} \right) \quad (1.68)$$

1.3.3 Functional-Link ΤΝΔ (FLANN)

Μια άλλη δομή ΤΝΔ είναι τα Functional Link (FLANN), αυτά είναι παρόμοια με πολυστρωματικά ΤΝΔ (MLP) αλλά συνδέουν τα δεδομένα εξόδου με τα δεδομένα που προκύπτουν όταν βαθμωτές συναρτήσεις εφαρμοστούν στα δεδομένα εισόδου [111]. Τα FLANN πρωτοεισήχθησαν από τον Ραο [112] και έχουν εφαρμογές σε προσεγγίσεις συναρτήσεων [113], αναγνώριση προτύπων [114] και στην εξισορρόπηση μη-γραμμικών καναλιών (nonlinear channel equalization) [115, 116].

Οι Ραο και Ραλ [115] προτείνουν ως συναρτήσεις στα FLANN να χρησιμοποιούνται οι τριγωνομετρικές $\sin(\cdot)$ και $\cos(\cdot)$. Πιο συγκεκριμένα αν τα δεδομένα εισόδου ενός ΤΝΔ είναι τα $\mathbf{x} = [x_1 \ x_2]$ τότε ως συναρτησιακή έκφρασή τους (Functional Expansion ή F.E.) μπορεί να χρησιμοποιείται η:

$$\mathbf{X} = F.E. = [x_1, \cos(\pi x_1), \sin(\pi x_1), \cos(2\pi x_1), \sin(2\pi x_1), x_2, \cos(\pi x_2), \sin(\pi x_2), \cos(2\pi x_2), \sin(2\pi x_2), x_1 \cdot x_2] \quad (1.69)$$

κατόπιν στο FLANN τα δεδομένα της συναρτησιακής έκφρασης \mathbf{X} αθροίζονται μέσω των συντελεστών βαρύτητας w_0, w_1, w_2, \dots και η προτεινόμενη έξοδος του δικτύου είναι η:

$$\hat{y} = g \left(\sum_{i=0}^n w_i X_i \right) \quad (1.70)$$

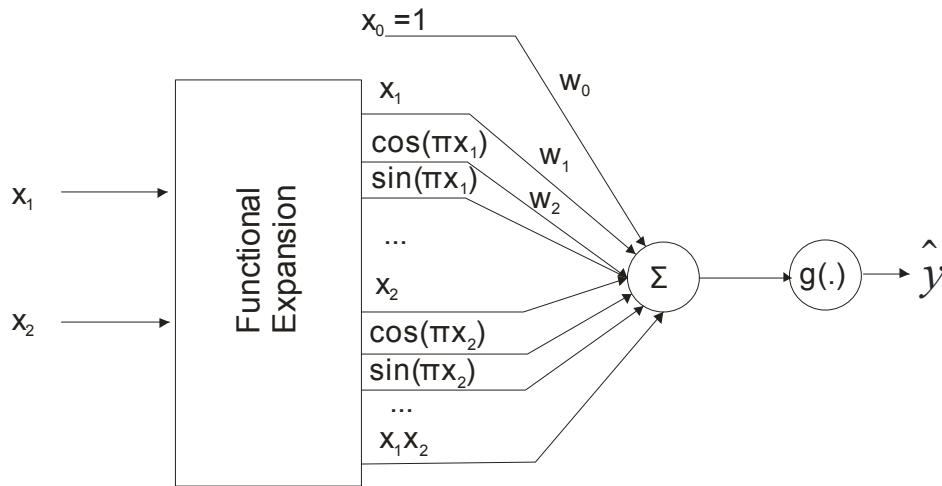
όπου $X_i, i = 1, 2, \dots, n$ οι συνιστώσες του διανύσματος \mathbf{X} , $X_0 = 1$ το bias του δικτύου και g η συνάρτηση ενεργοποίησης η οποία είναι η :

$$g(x) = \frac{1}{1 + e^{-x}} \quad (1.71)$$

ή η:

$$g(x) = \tanh x \quad (1.72)$$

Η δομή του δικτύου αυτή παρουσιάζεται στο Σχήμα 1.15.



Σχήμα 1.15 Η τοπολογία ενός Functional-Link TNA που χρησιμοποιεί τριγωνομετρικές εκφράσεις των δεδομένων εισόδου.

Για την εκπαίδευση του FLANN απομένει η εύρεση των κατάλληλων συναπτικών βαρών $w_i, i=0,1,\dots,n$ με οποιαδήποτε μέθοδο βελτιστοποίησης. Οι Pao και Pal [115] χρησιμοποίησαν τη μέθοδο BackPropagation.

Σύμφωνα με τους Patra et al. [116] οι συναρτήσεις $\varphi_i(x), i=1,\dots,c$ που χρησιμοποιούνται για τη συναρτησιακή έκφραση *F.E.* στα functional-link TNA θα πρέπει να ικανοποιούν τις ακόλουθες ιδιότητες για κάθε $x \in A$ (το $A \subseteq \mathbb{R}$):

- $\varphi_1(x) = 1$, δηλαδή το δίκτυο να έχει bias.
- οι $\varphi_i(x), i=1,\dots,c$ θα πρέπει να είναι γραμμικώς ανεξάρτητες συναρτήσεις
- $\sup_{x \in A} \sqrt{\sum_{i=1}^c \|\varphi_i(x)\|^2} < \infty$, δηλαδή στο διάστημα A το άθροισμα των μέτρων συναρτήσεων $\varphi_i(x), i=1,\dots,c$ να είναι πεπερασμένο.

Οι τρεις αυτές ιδιότητες ισχύουν αν οι $\varphi_i(x), i=1,\dots,c$ είναι τα ορθογώνια πολυώνυμα που αναφέρονται στην επόμενη παράγραφο (Πίνακας 1.1).

Στη βιβλιογραφία υπάρχουν και FLANN τα οποία χρησιμοποιούν διαφορετικές συναρτησιακές εκφράσεις *F.E.* από αυτή της εξίσωσης (1.69). Ο Pao [112] στα FLANN που χρησιμοποιεί ταυστές των δεδομένων εισόδου και η *F.E.* παίρνει τη μορφή:

$$F.E. = [x_1 \ x_2 \ x_1 x_2 \ x_1^2 \ x_2^2] \quad (1.73)$$

Υπάρχουν επίσης και τα FLANN τυχαίας μεταβλητής (Random variable functional link neural networks ή RVFLN) στα οποία σύμφωνα με τους Pao et al.[117] η η $F.E.$ έχει τη μορφή:

$$F.E. = [x_1 \ x_2 \ A_1x_1 + b_1 \ A_1x_2 + b_1 \ A_2x_1 + b_2 \ A_2x_2 + b_2 \ \dots] \quad (1.74)$$

Στη βιβλιογραφία χρησιμοποιούνται ως $F.E.$ και σιγμοειδής συναρτήσεις και πολυωνυμικές. Μια ανασκόπηση των μεθόδων δημιουργίας, εκπαίδευσης και χρήσης των FLANN έχει γίνει από τους Dehuri και Cho [118].

1.3.4 Πολυωνυμικά Νευρωνικά δίκτυα

Τα πολυωνυμικά ΤΝΔ (pNN) είναι ΤΝΔ τα οποία μπλέκουν πολυωνυμικές συναρτήσεις μέσα σε ΤΝΔ και δημιουργήθηκαν με σκοπό να αυξηθεί η ακρίβεια των ΤΝΔ όταν τα δεδομένα έχουν μη-γραμμικότητα [3-7]. Τα ΤΝΔ αυτά είναι δύο κατηγοριών: (α) αυτά που χρησιμοποιούν πολώνυμα στους κρυφούς κόμβους του δικτύου και (β) αυτά που χρησιμοποιούν πολώνυμα στη διαδικασία εκπαίδευσής τους (regression parameters).

Οι Oh et al. [5] δημιούργησαν ένα πολυωνυμικό ΤΝΔ η έξοδος του οποίου για ένα σύνολο δεδομένων εισόδου-εξόδου της μορφής :

$$S = \{(\mathbf{x}_i, y_i) / \mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi}) \in \mathbb{R}^p, y_i \in \mathbb{R}, i = 1, 2, \dots, N\} \quad (1.75)$$

υπολογίζεται ως εξής :

$$\hat{y}_k = c_0 + \sum_{k_1} c_{k_1} x_{k_1} + \sum_{k_1 k_2} c_{k_1 k_2} x_{k_1} x_{k_2} + \sum_{k_1 k_2 k_3} c_{k_1 k_2 k_3} x_{k_1} x_{k_2} x_{k_3} + \dots \quad (1.76)$$

όπου τα c_k είναι οι παράμετροι του ΤΝΔ. Η έξοδος του δικτύου αυτή μπορεί να πάρει εξής μορφές :

- Γραμμικό μοντέλο (linear):

$$\hat{y}_k = c_0 + \sum_{j=1}^p a_k x_{kj} \quad (1.77)$$

- Δευτέρου βαθμού μοντέλο (quadratic):

$$\hat{y}_k = c_0 + \sum_{j=1}^p a_k x_{kj} + \sum_{\substack{j=1 \\ j \neq m}}^p \sum_{m=1}^p b_k x_{kj} x_{km} + \sum_{j=1}^p c_k x_{kj}^2 \quad (1.78)$$

- Μειωμένο Δευτέρου βαθμού μοντέλο (Reduced quadratic):

$$\hat{y}_k = c_0 + \sum_{j=1}^p a_k x_{kj} + \sum_{j=1}^p c_k x_{kj}^2 \quad (1.79)$$

- Τρίτου βαθμού μοντέλο:

$$\hat{y}_k = (\text{Δευτέρου βαθμού μοντέλο}) + \sum_{j=1}^p \sum_{m=1}^p \sum_{q=1}^p d_k x_{kj} x_{km} x_{kq} \quad (1.80)$$

Για να εκπαιδευτεί το ΤΝΔ πρέπει να προσδιοριστούν οι βέλτιστες παράμετροι $c_0, a_k, \beta_k, c_k, d_k, \dots$ (λύνοντας το πρόβλημα βελτιστοποίησης που προκύπτει) το οποίο όπως γίνεται κατανοητό κι από τον επόμενο πίνακα, όσο αυξάνει η διάσταση του χώρου εισόδου p και ο βαθμός του μοντέλου που χρησιμοποιείται, αυξάνεται η υπολογιστική ισχύς που χρειάζεται για την εκπαίδευση του ΤΝΔ. Το πλεονέκτημά του είναι πως η εκπαίδευση του

δικτύου μπορεί να γίνει με τη μέθοδο των ελαχίστων τετραγώνων μιας και η έξοδος του ΤΝΔ είναι γραμμική ως προς τις παραμέτρους $c_0, a_k, \beta_k, c_k, d_k, \dots$.

Πίνακας 1.2 Το πλήθος των παραμέτρων που προκύπτουν από τις εξισώσεις (1.77)-(1.80) κατά την εκπαίδευση ενός πολυωνυμικού ΤΝΔ

Μέγεθος δείγματος	Χώρος δεδομένων εισόδου	Μοντέλο του ΤΝΔ	Πλήθος παραμέτρων που πρέπει να υπολογιστούν	Πλήθος εξισώσεων που εισέρχονται στο πρόβλημα βελτιστοποίησης
N	p	Γραμμικό	$p + 1$	$N(p + 1)$
		Δευτέρου βαθμού	$p^2 + p + 1$	$N(p^2 + p + 1)$
		Τρίτου βαθμού	$p^3 + p^2 + p + 1$	$N(p^3 + p^2 + p + 1)$

Οι Oh et al. [6] δημιούργησαν ένα ΤΝΔ συναρτήσεων ακτινικής βάσης με χρήση πολυωνύμων (P-RBF NN). Το ΤΝΔ αυτό αποτελείται από τις τρεις εξής φάσεις:

- (i) Φάση υπόθεσης (condition phase): η φάση αυτή αποτελείται από ένα ΤΝΔ ακτινικής βάσης του οποίου οι κρυφοί κόμβοι προσδιορίζονται από την ασαφή συσταδοποίηση C-μέσων (FCM)
- (ii) Φάση συμπεράσματος (conclusion phase): η φάση αυτή περιέχει το πολυωνυμικό ΤΝΔ των Oh et al. [5] που περιγράφηκε προηγουμένως με χρήση μίας εκ των συναρτήσεων (1.77)-(1.79).
- (iii) Φάση άθροισης (aggregation phase): στη φάση αυτή προκύπτει η έξοδος του δικτύου \hat{y} . Η έξοδος που προκύπτει είναι το άθροισμα των γινομένων των αντίστοιχων εξόδων των δύο προηγούμενων φάσεων.

Το ΤΝΔ που μόλις περιγράφηκε έχει τη μορφή ενός ΤΝΔ ακτινικής βάσης όπως αυτό που περιγράφηκε στην παράγραφο 1.3.1 με τη διαφορά ότι αντί για τα συναπτικά βάρη w_i της σχέσης (1.42) χρησιμοποιούνται τα αποτελέσματα της φάσης συμπεράσματος.

Οι Lee και Jeng [4] δημιούργησαν ένα εμπροσθοτροφοδοτούμενο ΤΝΔ το οποίο έχει ως κόμβους πολυώνυμα και τανυστές γινομένων πολυωνύμων Chebychev, πράγμα το οποίο

όμως αυξάνει κατά πολύ το πλήθος των συνάψεων του ΤΝΔ. Οι Oh και Pedrycz [3, 7] προτείνουν ένα ασαφές πολυωνυμικό ΤΝΔ το οποίο βασίζεται στη μέθοδο group method of data handling [2, 119], η δομή του δικτύου δημιουργείται μέσω μιας δυναμικής διαδικασίας και εμπεριέχει αρκετά στρώματα το καθένα από τα οποία θεωρείται ως ένα σύνολο πολλαπλής εισόδου και μονής εξόδου. Στο ΤΝΔ αυτό χρησιμοποιούνται στους κόμβους πολυώνυμα ως συναρτήσεις ενεργοποίησης.

Στα πολυωνυμικά ΤΝΔ που χρησιμοποιούν πολυώνυμα στους κόμβους τους δικτύου έχει μεγάλη σημασία να χρησιμοποιηθούν **ορθογώνια πολυώνυμα**. Ο λόγος της χρησιμότητας των ορθογωνίων πολυωνύμων είναι ότι κάθε πεπερασμένη δυναμοσειρά μπορεί να προσεγγιστεί από ορθογώνια πολυώνυμα. Αν για παράδειγμα χρησιμοποιηθούν τα πολυώνυμα Chebychev 1^{ου} είδους $T_n(x)$, $n=0,1,\dots$ (βλ. ενότητα 2) τότε κάθε δυναμοσειρά

$\Delta(x) = \sum_{n=0}^N \alpha_n x^n$ μπορεί να γραφεί ως :

$$\Delta(x) = \sum_{n=0}^N \alpha_n x^n = \sum_{n=0}^N c_n T_n(x) \quad (1.81)$$

όπου οι συντελεστές c_n μπορούν να προσδιοριστούν από τον αλγόριθμο του Schur όπως αναλύεται από τους Charman και Rahman [120]. Συνέπεια του ιδιότητας αυτής σύμφωνα με τους Lee and Jeng [4] είναι τα εξής θεωρήματα (οι αποδείξεις τους βρίσκονται στο άρθρο [4]) :

Θεώρημα 1.1: Αν κάθε συνάρτηση ενεργοποίησης ενός ΤΝΔ είναι ολοκληρώσιμη στο $[-1,1]$ τότε η συνάρτηση ενεργοποίησης του ΤΝΔ μπορεί να προσεγγιστεί από σειρές πολυωνύμων Chebychev στο $[-1,1]$. ■

Θεώρημα 1.2: Έστω ένα εμπροσθοτροφοδοτούμενο ΤΝΔ με ένα κρυφό στρώμα και τις συναρτήσεις ενεργοποίησης να είναι όλες γραμμικές. Αν όλες οι συναρτήσεις ενεργοποίησης είναι κατά Riemann ολοκληρώσιμες, τότε το εμπροσθοτροφοδοτούμενο ΤΝΔ μπορεί να αναπαρασταθεί από ένα πολυωνυμικό ΤΝΔ με βάση τα πολυώνυμα Chebychev ■

Παρόμοια συμπεράσματα ισχύουν και για άλλα ορθογώνια πολυώνυμα.

Μια διαφορετική προσέγγιση ενός πολυωνυμικού ΤΝΔ προτείνεται από τους Huang et al. στο [8] όπου έχει δημιουργηθεί ένα νευροασαφές σύστημα με υψηλές αποδόσεις το οποίο

μπλέκει πολυώνυμα στη διαδικασία της εκπαίδευσης του ΤΝΔ. Η απόδοση του ΤΝΔ αυτού είναι αρκετά καλή (αποδίδει μικρά σφάλματα).

Στη βιβλιογραφία επίσης υπάρχουν ΤΝΔ τα οποία εμπλέκουν πολυώνυμα στη διαδικασία εκπαίδευσής τους, με τους συντελεστές των πολυωνύμων να προκύπτουν με χρήση εξελικτικών αλγορίθμων (evolutionary algorithms) και αλγορίθμους νοημοσύνης σμήνους (βλ. [6, 9]). Ένα βασικό μειονέκτημα της μεθόδου αυτής είναι το μεγάλο πλήθος των (ως προς προσδιορισμό) παραμέτρων του ΤΝΔ που προκύπτει.

1.4 Ασαφή και Νευροασαφή Συστήματα (Fuzzy and Neuro-Fuzzy Systems)

1.4.1 Ασαφή Σύνολα

Στην διακριτή θεωρία συνόλων τα στοιχεία x ενός χώρου U ανήκουν ή δεν ανήκουν σε ένα σύνολο A (διακριτή λογική), δηλαδή μπορούμε να γράψουμε ότι:

$$A = \{x \in U \mid \mu_A(x) = 1\} \quad (1.82)$$

ή ισοδύναμα το σύνολο αυτό μπορεί να περιγραφεί από τη συνάρτηση συμμετοχής των στοιχείων x του συνόλου U στο σύνολο A :

$$\mu_A(x) = \begin{cases} 0, & x \notin A \\ 1, & x \in A \end{cases} \quad (1.83)$$

Στην ασαφή λογική όπως αναφέρθηκε και στην παράγραφο 1.3.1.1 ένα το σύνολο A περιγράφεται από την ασαφή συνάρτηση συμμετοχής:

$$\mu_A(x) \in [0,1] \quad (1.84)$$

δηλαδή όλα τα στοιχεία x του χώρου U ανήκουν στο ασαφές σύνολο A κατά ένα ποσοστό $\mu_A(x)$.

Στην ασαφή λογική οι συναρτήσεις συμμετοχής δίνονται από συναρτήσεις, οι πιο γνωστές από τις οποίες είναι οι:

- Τριγωνική συνάρτηση συμμετοχής:

$$\mu(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ -\frac{x-c}{c-b}, & b \leq x \leq c \\ 0, & x \geq c \end{cases} \quad (1.85)$$

- Τραπεζοειδής συνάρτηση συμμετοχής:

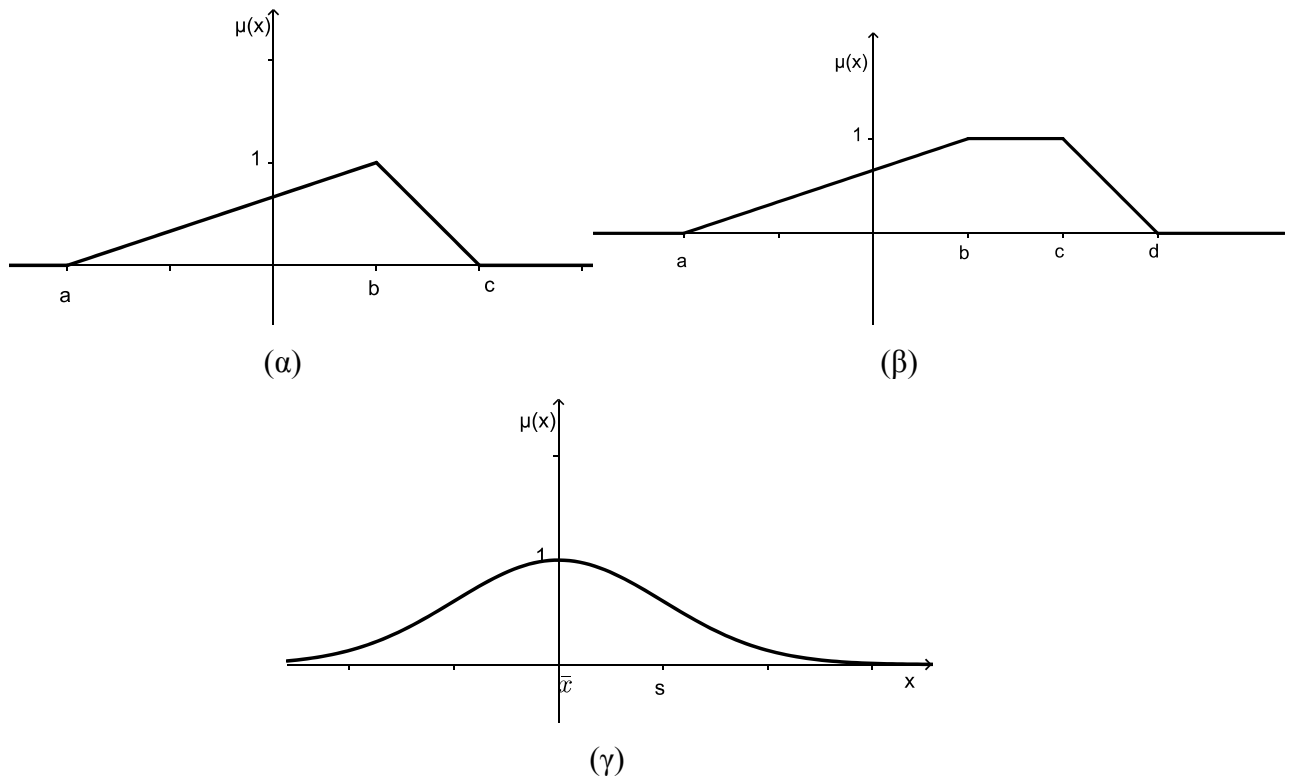
$$\mu(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ -\frac{x-d}{d-c}, & c \leq x \leq d \\ 0, & x \geq d \end{cases} \quad (1.86)$$

- Γκαουσιανή (gaussian) συνάρτηση συμμετοχής:

$$\mu(x) = e^{-\left(\frac{x-\bar{x}}{\sqrt{2}s}\right)^2} \quad (1.87)$$

όπου \bar{x} η μέση τιμή και s η τυπική απόκλιση.

Οι γραφικές παραστάσεις των τριών συναρτήσεων συμμετοχής παρουσιάζονται στο Σχήμα 1.16.



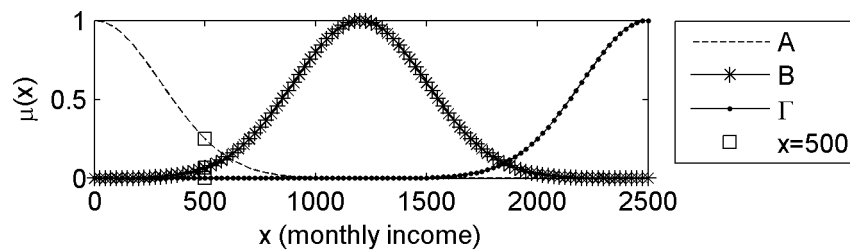
Σχήμα 1.16 Οι γραφικές παραστάσεις τριών συναρτήσεων συμμετοχής. (α) Η τριγωνική (β) η τραπεζοειδής και (γ) η γκαουσιανή συνάρτηση συμμετοχής.

Ένα παράδειγμα να εξηγεί τα προηγούμενα είναι το ακόλουθο. Αν υποθέσουμε ότι θέλουμε να περιγράψουμε τη μεταβλητή $x =$ 'μηνιαίο εισόδημα' η οποία να ορίζεται στο σύνολο $U = [0, 2500]$. Σύμφωνα με την κλασική θεωρία συνόλων θα μπορούσαμε να κατατάξουμε ένα μηνιαίο εισόδημα σε 3 κατηγορίες $A =$ 'χαμηλό', $B =$ 'μεσαίο' και $\Gamma =$ 'υψηλό' και αν επιλεγθεί το όριο 500€ να διαχωρίζει το εισόδημα από την κατηγορία A στην B τότε το σύνολο A θα είχε την εξής μορφή:

$$A = \{x \in U \mid \mu_A(x) = 1\} \text{ όπου } \mu_A(x) = \begin{cases} 1, x \leq 500€ \\ 0, x > 500€ \end{cases} \quad (1.88)$$

Ο ορισμός αυτός δεν είναι ρεαλιστικός αφού κάποιος με μηνιαίο εισόδημα 500€ θα κατατασσόταν στο σύνολο Α ενώ κάποιος με 501€ στον σύνολο Β, οπότε εδώ φαίνεται η ανάγκη της χρήσης της ασαφούς λογικής.

Στην ασαφή λογική θα μπορούσαμε να διαλέξουμε ως σύνολα Α,Β,Γ τα σύνολα με συναρτήσεις συμμετοχής τις γκαουσιανές συναρτήσεις της εξίσωσης (1.87), και χρησιμοποιώντας για τα σύνολα Α,Β,Γ ως \bar{x} τις τιμές 0,1200,2500€ αντίστοιχα και ως τυπική απόκλιση την $s = 300€$ προκύπτει το Σχήμα 1.17.



Σχήμα 1.17 Οι γκαουσιανές συναρτήσεις συμμετοχής τριών ασαφών συνόλων Α,Β,Γ.

Στο παράδειγμα αυτό ένας άνθρωπος με εισόδημα $x = 500€$ θα έχει συναρτήσεις συμμετοχής $\mu_A(x) = 0.2494$, $\mu_B(x) = 0.0657$ και $\mu_\Gamma(x) = 2.23 \cdot 10^{-10}$ (οι τιμές αυτές παριστάνονται με το σύμβολο \square στο Σχήμα 1.17). Η φυσική ερμηνεία των τιμών των συναρτήσεων συμμετοχής είναι ότι ο μισθός $x = 500€$ ανήκει στο σύνολο Α = 'χαμηλός μισθός' κατά 25% στο σύνολο Β = 'μέσος μισθός' κατά 6.6% και στο σύνολο Γ = 'υψηλός μισθός' κατά ένα πολύ μικρό ποσοστό.

1.4.2 Ασαφείς Κανόνες (Fuzzy Rules)

Οι ασαφείς κανόνες είναι εκφράσεις της μορφής :

$$\text{IF } x \text{ is } A \text{ THEN } y \text{ is } B \quad (1.89)$$

αν x ένα μονοδιάστατο δεδομένο εισόδου ($x \in \mathbb{R}$) ενώ για πολυδιάστατη είσοδο $\mathbf{x} = [x_1, \dots, x_p]$ ο παραπάνω κανόνας γίνεται:

$$\text{IF } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \dots \text{ and } x_p \text{ is } A_p \text{ THEN } y \text{ is } B \quad (1.90)$$

όπου τα Α,Β στην 1^η περίπτωση και τα A_1, \dots, A_p, B στην 2^η είναι ασαφή σύνολα που χαρακτηρίζονται από τις συναρτήσεις συμμετοχής τους, x ή $\mathbf{x} = [x_1, \dots, x_p]$ είναι η μεταβλητή εισόδου η οποία παίρνει τιμές συμμετοχής στο σύνολο Α ή στα σύνολα A_1, \dots, A_p και y

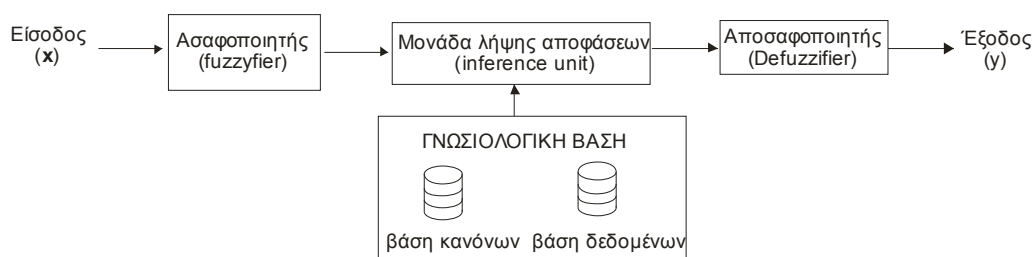
είναι η έξοδος του συστήματος που υπολογίζεται από μηχανισμό εξαγωγής συμπεράσματος (inference engine) σε ασαφή μορφή και δηλώνει την απόφαση του κανόνα. Το ασαφές συμπέρασμα μετά από-ασαφοποιείται με τον μηχανισμό της αποασαφοποίησης (defuzzification) ώστε στο τέλος να προκύψει μία σαφής τιμή.

1.4.3 Ασαφή συστήματα

Γενικά τα ασαφή συστήματα αποτελούνται από τρεις κατηγορίες: τα ασαφή συστήματα Mamdani, τα ασαφή συστήματα Takagi-Sugeno-Kang (TSK) και τα δυναμικά ασαφή συστήματα [121]. Στις παρακάτω ενότητες ακολουθεί η δομή των δύο πρώτων κατηγοριών από αυτές.

1.4.3.1 Ασαφή συστήματα Mamdani

Η δομή ενός ασαφούς συστήματος Mamdani είναι αυτή που παρουσιάζεται στο Σχήμα 1.18 και αποτελείται από: (α) τον ασαφοποιητή (β) τη γνωσιολογική βάση, η οποία αποτελείται από την βάση δεδομένων και την βάση κανόνων (γ) τη μονάδα λήψης αποφάσεων η οποία λαμβάνει υπόψη της τη βάση κανόνων και τη βάση δεδομένων (δ) τον αποασαφοποιητή. Η λειτουργία του κάθε τμήματος αναλύεται ακολούθως.



Σχήμα 1.18 Η δομή ενός κλασσικού ασαφούς συστήματος.

(α) Ασαφοποιητής

Στο τμήμα αυτό τα δεδομένα εισόδου μετατρέπονται από πραγματικούς αριθμούς σε ασαφείς αριθμούς ή ασαφή σύνολα. Οι μέθοδοι ασαφοποίησης είναι οι παρακάτω ([121]). :

(i) Μετατροπή σε ασαφή σύνολα του ενός στοιχείου

Με βάση αυτή την μέθοδο, μία πραγματική τιμή μετασχηματίζεται σε ένα ασαφές σύνολο το οποίο περιέχει μόνο ένα στοιχείο (fuzzy singleton). Το στοιχείο αυτό μετέχει στο ασαφές σύνολο με βαθμό συμμετοχής ίσο με την μονάδα. Βασικά η μέθοδος αυτή δεν εισάγει κανένος είδους ασάφειας για την πραγματική τιμή, αλλά χρησιμοποιείται στον σχεδιασμό ασαφών συστημάτων επειδή μπορεί να εφαρμοστεί πολύ εύκολα.

(ii) Μετατροπή σε ασαφείς αριθμούς

Η συγκεκριμένη μέθοδος ασαφοποίησης μετατρέπει πραγματικούς αριθμούς σε ασαφείς αριθμούς, οι οποίοι χρησιμοποιούνται ευρέως στον σχεδιασμό ασαφών συστημάτων. Η μέθοδος αυτή εφαρμόζεται κυρίως όταν τα μετρήσιμα μεγέθη που θέλουμε να ασαφοποιήσουμε εμπεριέχουν θόρυβο.

(β) Γνωσιολογική Βάση.

Η γνωσιολογική βάση αποτελείται από δύο τμήματα, τη βάση δεδομένων και τη βάση κανόνων. Η **βάση δεδομένων** είναι η μονάδα του ασαφούς συστήματος στην οποία ορίζονται τα χαρακτηριστικά των γλωσσολογικών μεταβλητών του συστήματος δηλ. τα πεδία ορισμού και τα ασαφή σύνολα στα οποία αυτά διαιρούνται. Οι παράμετροι των ασαφών συνόλων ονομάζονται παράμετροι ασαφούς συστήματος. Το κύριο μέλημα στον σχεδιασμό της βάσης δεδομένων είναι ο σωστός ορισμός αυτών των παραμέτρων.

Τα πεδία ορισμού, για λόγους ευκολίας, κανονικοποιούνται σ' ένα διάστημα $[-\alpha, +\alpha]$. Αυτό έχει ως αποτέλεσμα οι διακριτές τιμές των εισόδων πριν την μονάδα ασαφοποίησης να πολλαπλασιάζονται με κάποιους συντελεστές, οι οποίοι ονομάζονται συντελεστές μετατροπής κλίμακας εισόδων (input scaling factors), ούτως ώστε οι τιμές αυτές να απεικονιστούν από το πραγματικό πεδίο ορισμού στο παραπάνω διάστημα και στην συνέχεια να ασαφοποιηθούν. Η **βάση κανόνων** (Fuzzy Rule Base) αποτελείται από ένα αριθμό L ασαφών "IF-THEN" κανόνων οι οποίοι περιγράφουν την γλωσσολογική σχέση μεταξύ των εισόδων και των εξόδων του ασαφούς συστήματος. Για m εισόδους και n εξόδους ο l κανόνας ($1 \leq l \leq L$), εκφράζεται ως εξής:

$$\begin{aligned} R_j: & \text{ IF } x_1 \text{ is } A_1^l \text{ and } x_2 \text{ is } A_2^l \text{ and } \dots \text{ and } x_m \text{ is } A_m^l \\ & \text{ THEN } y_1 \text{ is } B_1^l \text{ and } y_2 \text{ is } B_2^l \text{ and } \dots \text{ and } y_n \text{ is } B_n^l \end{aligned} \quad (1.91)$$

όπου x_i ($i = 1, 2, \dots, m$) και y_j ($j = 1, 2, \dots, n$) είναι αντίστοιχα οι γλωσσολογικές μεταβλητές εισόδου και εξόδου, ενώ A_i^l, B_j^l είναι ασαφή σύνολα.

(γ) Μονάδα λήψης αποφάσεων

Στο στάδιο αυτό συνδυάζονται οι ασαφείς κανόνες που πυροδοτούνται κάθε φορά με σκοπό να εξαχθεί κάποιο συμπέρασμα ή να εκτελεστεί κάποια ενέργεια. Αυτό επιτυγχάνεται με τις παρακάτω συνεπαγωγές:

- Συνεπαγωγή Mamdani:

$$\mu_B(x_1, \dots, x_p) = \mu_{A_1}(x_1) \cap \mu_{A_2}(x_2) \cap \dots \cap \mu_{A_p}(x_p) = \min_{j=1, \dots, p} \{ \mu_{A_j}(x_j) \} \quad (1.92)$$

- Συνεπαγωγή Larsen:

$$\mu_B(x_1, \dots, x_p) = \mu_{A_1}(x_1) \cdot \mu_{A_2}(x_2) \cdot \dots \cdot \mu_{A_p}(x_p) \quad (1.93)$$

(δ) Αποασαφοποιητής.

Η μονάδα αποασαφοποίησης απεικονίζει τις ασαφείς τιμές της εξόδου σε διακριτές. Οι δύο βασικές μέθοδοι αποασαφοποίησης είναι οι παρακάτω:

(i) Μέση τιμή των μεγίστων

Η μέθοδος της “μέσης τιμής των μεγίστων” χρησιμοποιεί την εξίσωση (1.92) ή την (1.93) και υπολογίζει τον βαθμό πυροδότησης του κάθε κανόνα. Στην συνέχεια από τους L κανόνες διαλέγει τους L' ($1 \leq L' \leq L$), οι οποίοι παρουσιάζουν βαθμούς πυροδότησης διάφορους του μηδενός. Τέλος, λαμβάνοντας υπόψη τα ασαφή σύνολα των συμπερασμάτων των L' κανόνων, δίνει ως διακριτή τιμή εξόδου τον μέσο όρο των κέντρων τους b^l ($l=1, 2, \dots, L'$), την:

$$\hat{y} = \frac{\sum_{l=1}^{L'} b^l}{L'} \quad (1.94)$$

(ii) Κέντρο βάρους

Όπως και προηγουμένως, η μέθοδος του “κέντρου βάρους” προσδιορίζει το βαθμό πυροδότησης του κάθε κανόνα ξεχωριστά. Στην συνέχεια η διακριτή τιμή της εξόδου δίνεται από την σχέση:

$$\hat{y} = \frac{\sum_{l=1}^L b^l \mu_{B_l}(y)}{\sum_{l=1}^L \mu_{B_l}(y)} \quad (1.95)$$

όπου b^l είναι το κέντρο του ασαφούς συνόλου της μεταβλητής εξόδου που συμμετέχει στο l κανόνα και $\mu_{B_l}(y)$ είναι ο βαθμός πυροδότησης του l κανόνα.

Οι δύο μέθοδοι αποασαφοποίησης που παρουσιάστηκαν παραπάνω χρησιμοποιούνται πολύ συχνά στον σχεδιασμό ασαφών συστημάτων γιατί έχουν αποδειχθεί πολύ αποδοτικές σε διάφορες περιοχές εφαρμογών. Πρέπει να σημειωθεί ότι, όταν τα πεδία ορισμού των γλωσσολογικών μεταβλητών κανονικοποιούνται σ' ένα διάστημα $[-a, +a]$ (όπως π.χ. συμβαίνει στην περιοχή της ασαφούς αυτόματης ρύθμισης) είναι απαραίτητο η διακριτή τιμή της εξόδου που προκύπτει από την μέθοδο αποασαφοποίησης να πολλαπλασιαστεί με έναν συντελεστή, ο οποίος ονομάζεται συντελεστής αλλαγής κλίμακας εξόδου (output scaling factor), ώστε να απεικονιστεί η τιμή αυτή στο πραγματικό πεδίο ορισμού της μεταβλητής.

1.4.3.2 Ασαφή συστήματα Takagi-Sugeno-Kang (TSK)

Οι κύριες διαφορές μεταξύ του ασαφούς συστήματος Takagi-Sugeno και του κλασσικού ασαφούς συστήματος είναι: (α) η διαφορετική δομή των ασαφών κανόνων και (β) το ασαφές σύστημα Takagi-Sugeno-Kang δεν έχει μονάδα αποασαφοποίησης. Πιο συγκεκριμένα, η βάση κανόνων στο ασαφές σύστημα Takagi-Sugeno-Kang (TSK) περιέχει κανόνες της μορφής:

$$R_l : \text{IF } x_1 \text{ is } A_1^l \text{ and } \dots \text{ and } x_m \text{ is } A_m^l \text{ THEN } y^l = b_0^l + b_1^l x_1 + \dots + b_m^l x_m \quad (1.96)$$

όπου b_0^l, \dots, b_m^l είναι πραγματικοί αριθμοί και y^l είναι η έξοδος του συστήματος που οφείλεται στον παραπάνω κανόνα. Όπως δείχνει η (1.96) το μέρος της υπόθεσης του κανόνα είναι ασαφές, ενώ το συμπέρασμα του κανόνα είναι γραμμικός συνδυασμός διακριτών αριθμών και συνεπώς είναι διακριτός αριθμός. Ως εκ τούτου, επειδή η έξοδος είναι ήδη

διακριτός αριθμός, δεν είναι απαραίτητη η εφαρμογή καμιάς μεθόδου αποασαφοποίησης. Τέλος, για ένα σύνολο L κανόνων η έξοδος του συστήματος δίνεται από την σχέση:

$$\hat{y} = \frac{\sum_{l=1}^L y^l \mu_l}{\sum_{l=1}^L \mu_l} \quad (1.97)$$

όπου μ_l είναι ο βαθμός πυροδότησης του l κανόνα και υπολογίζεται ως:

$$\mu_l = \mu_{\bigcap_{i=1}^m A_i^l}(x_1, x_2, \dots, x_m) = \min \{ \mu_{A_1^l}(x_1), \mu_{A_2^l}(x_2), \dots, \mu_{A_m^l}(x_m) \} = \mu_{A_1^l}(x_1) \mu_{A_2^l}(x_2) \dots \mu_{A_m^l}(x_m) \quad (1.98)$$

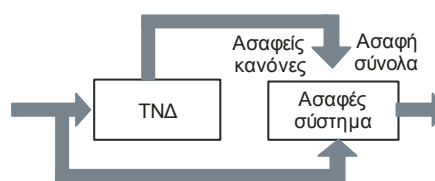
Το κύρια πλεονέκτημα του ασαφούς συστήματος Takagi-Sugeno-Kang είναι ότι η εξίσωση (1.97) δίνει ένα συμπαγές σύστημα και με βάση την γραμμική μορφή, με την οποία εκφράζονται οι παράμετροι του συστήματος, καθίσταται εφικτή η χρήση κλασικών μεθόδων για τον προσδιορισμό τους.

1.4.4 Νευροασαφή δίκτυα

Ως νευροασαφή δίκτυα (Neuro-Fuzzy Networks) ορίζονται τα δίκτυα που προκύπτουν από τη σύνθεση συστημάτων ασαφούς λογικής με ΤΝΔ [122] και οι κύριες μορφές τους είναι οι εξής:

(α) Συνεργατικό Νευροασαφές δίκτυο (Cooperative Neuro-Fuzzy Network)

Στο δίκτυο αυτό (Σχήμα 1.19) υπάρχει μια προεργασία όπου το ΤΝΔ ορίζει κάποια στοιχεία του ασαφούς συστήματος, όπως για παράδειγμα τα ασαφή σύνολα ή τους ασαφείς κανόνες. Αφού οι ασαφείς παράμετροι υπολογιστούν από το ΤΝΔ τότε εκτελείται μόνο το ασαφές σύστημα.



Σχήμα 1.19 Ένα τυπικό συνεργατικό νευροασαφές σύστημα

Ένα παράδειγμα τέτοιου δικτύου δίνεται από τους Linh et al. στο [24] οι οποίοι έχοντας δεδομένα από τους χτύπους της καρδιάς, προσέγγισαν τα δεδομένα αυτά ως εκφράσεις πολυωνύμων Hermite και κατόπιν εφάρμοσαν ένα ασαφές σύστημα Takagi-Sugeno-Kang στα δεδομένα αυτά.

(β) Ταυτόχρονο Νευροασαφές δίκτυο (Concurrent Neuro-Fuzzy Network)

Στα δίκτυα αυτά (Σχήμα 1.20) το ΤΝΔ κι το ασαφές σύστημα συνεργάζονται συνεχόμενα.

Γενικά το ΤΝΔ επεξεργάζεται την είσοδο ή την έξοδο του ασαφούς συστήματος.



Σχήμα 1.20 Ένα τυπικό ταυτόχρονο νευροασαφές σύστημα

Οι Chen και Lin στο [123] δημιούργησαν ένα νευροασαφές δίκτυο το οποίο αποτελείται από τις εξής δύο φάσεις: (α) ένα functional-link ΤΝΔ το οποίο χρησιμοποιεί τριγωνομετρικές συναρτήσεις (β) ένα ασαφές σύστημα Takagi-Sugeno-Kang. Οι δύο αυτές φάσεις επεξεργάζονται παράλληλα τα δεδομένα εισόδου και η έξοδος είναι το άθροισμα των αντίστοιχων γινομένων των δύο αυτών φάσεων. Οι Wang et al. [124] χρησιμοποιούν ένα ταυτόχρονο νευροασαφές σύστημα για να εξορύξουν δεδομένα που αφορούν την κίνηση του διαδικτύου ενώ οι Ocampo-Duque et al.[125] για να εξετάσουν τον οικολογικό κίνδυνο στα ποτάμια.

(γ) Υβριδικό νευροασαφές δίκτυο (Hybrid Neuro-Fuzzy Network)

Σε αυτή την κατηγορία το ΤΝΔ χρησιμοποιείται για να καθοριστούν κάποιες παράμετροι του ασαφούς συστήματος με κάποια επαναληπτική διαδικασία. Η πλειονότητα των επιστημόνων όταν χρησιμοποιεί τον όρο νευροασαφές δίκτυα, αναφέρεται στα υβριδικά. Γνωστότερα από αυτά τα συστήματα είναι τα ANFIS[126], FALCON[127], GARIC[128], NEFCON[129], SONFIN[130] και dmEfuNN[131] των οποίων η δομή και λειτουργία μπορεί να βρεθεί και στα άρθρα-ανασκόπησης (survey papers) των Vieira et al. [122] και των Viharos και Kis [132].

2. Ανάπτυξη Πολυωνυμικού Νευρωνικού Δικτύου Με Χρήση Πολυωνύμων Chebyshev

Στα πλαίσια της εκπόνησης της παρούσας διδακτορικής διατριβής αναπτύχθηκε ένα ΤΝΔ με τη χρήση ορθογωνίων πολυωνύμων Chebyshev και δημοσιεύτηκε στην [1]. Το ΤΝΔ αυτό αποτελείται από τα εξής τρία τμήματα: (α) ένα πολυωνυμικό ΤΝΔ το οποίο έχει πολώνυμα Chebyshev ως συναρτήσεις ενεργοποίησης (β) ένα ΤΝΔ συναρτήσεων ακτινικής βάσης το οποίο χρησιμοποιεί μια πρωτότυπη ασαφή συσταδοποίηση χώρου εισόδου-εξόδου και (γ) το τμήμα εξαγωγής συμπεράσματος το οποίο συνδυάζει τα συμπεράσματα των δύο προηγούμενων τμημάτων.

Στο πολυωνυμικό τμήμα του προτεινόμενου δικτύου δημιουργεί σειρές Chebyshev των δεδομένων εισόδου, αφού όμως στα δεδομένα εισόδου έχουν εφαρμοστεί κατάλληλοι μετασχηματισμοί ώστε να παίρνουν τιμές στο διάστημα $[-1,1]$ στο οποίο λειτουργούν τα πολώνυμα Chebyshev. Το τμήμα του προτεινόμενου ΤΝΔ το οποίο χρησιμοποιεί συναρτήσεις ακτινικής βάσης χρησιμοποιεί μια πρωτότυπη συσταδοποίηση εισόδου-εξόδου. Αυτή συσταδοποιεί τα δεδομένα εισόδου λαμβάνοντας υπόψη της την κατανομή των δεδομένων εξόδου και με αυτόν τον τρόπο δίνει καλύτερες προσεγγιστικές δυνατότητες στο ΤΝΔ [54].

Στόχος του δικτύου αυτού είναι η χρήση του σε ισχυρά μη-γραμμικά συστήματα ώστε να αυξηθούν οι προσεγγιστικές δυνατότητες που αποδίδει. Για το λόγο αυτό το προτεινόμενο δίκτυο συνδέει τις προσεγγιστικές δυνατότητες των ορθογωνίων πολυωνύμων Chebyshev τα οποία χρησιμοποιούνται στην προσέγγιση μη-γραμμικών συστημάτων (βλ. [4, 133-137]) και των ΤΝΔ ακτινικής βάσης τα οποία επίσης χρησιμοποιούνται σε μη-γραμμικά δεδομένα (βλ. [10-16]). Το δίκτυο που δημιουργήθηκε έχει δοκιμαστεί και σε ένα πρωτότυπο μη-γραμμικό σύνολο δεδομένων το οποίο προέκυψε από επεξεργασία ψηφιακών εικόνων και έχει ως στόχο την αναγνώριση της θέση μιας ακτογραμμής μέσα σε εικόνες (βλ. παράγραφο 2.6 και [1]).

Η δομή των ακολούθων παραγράφων είναι η εξής: πρώτα δίνονται συναφείς μέθοδοι της βιβλιογραφίας και γίνεται μια εισαγωγή στην προτεινόμενη μέθοδο, ακολουθεί η παράγραφος με τα πολώνυμα Chebyshev όπου παρουσιάζονται οι ορισμοί τους και χρήσεις τους, έπειτα δίνεται η δομή του προτεινόμενου ΤΝΔ και η διαδικασία εκπαίδευσης του. Τέλος ακολουθεί η παράγραφος των αριθμητικών προσομοιώσεων σε δεδομένα του UCI Machine Learning Repository [100] καθώς και η παράγραφος που εξηγεί τη χρήση του προτεινόμενου ΤΝΔ στο πρόβλημα επεξεργασίας εικόνας του άρθρου [1].

2.1 Συναφείς μέθοδοι και προτεινόμενη μέθοδος

Τα πολώνυμα Chebyshev έχουν χρησιμοποιηθεί από πολλούς ερευνητές στη δομή των ΤΝΔ. Οι Lee and Jeng [4] απέδειξαν ότι κάθε εμπροσθοτροφοδοτούμενο ΤΝΔ μπορεί να προσεγγιστεί από δυναμοσειρές Chebyshev αρκεί η συνάρτηση ενεργοποίησης να είναι ολοκληρώσιμη κατά Riemann στο διάστημα $[-1,1]$, κατόπιν δημιούργησαν ένα πολυωνυμικό ΤΝΔ (pNN) χρησιμοποιώντας τα πολώνυμα Chebyshev ως νευρώνες του ΤΝΔ (βλ. και παράγραφο 1.3.4), το δίκτυο αυτό το ονόμασαν Chebyshev-Polynomials-Based (CPB) και κατέληξαν ότι αυτό έχει μεγαλύτερη ακρίβεια σε σχέση με εμπροσθοτροφοδοτούμενα ΤΝΔ.

Οι Purwar et al. [138] χρησιμοποίησαν ένα functional link ΤΝΔ (FLANN) το οποίο δημιουργούσε προσέγγιση των μεταβλητών εισόδου με μια ορθογώνια βάση πεπερασμένου πλήθους πολυωνύμων Chebyshev. Συγκεκριμένα χρησιμοποιώντας το θεώρημα των Lee and Jeng δημιούργησαν ένα πολυωνυμικό ΤΝΔ με συναρτήσεις ενεργοποίησης τις :

$$\varphi = [1 \ T_1(x_1) \ T_2(x_1) \ \dots \ T_n(x_p)]^T \quad (2.1)$$

όπου $T_i(x_j)$ τα πολώνυμα Chebyshev $1^{ov}, 2^{ov}, 3^{ov}$ ή 4^{ov} είδους. Το ίδιο ΤΝΔ δημιούργησαν και οι Basios et al. [139] το οποίο δίνει έξοδο την:

$$\hat{y}(x_k) = \sum_{i=0}^c w_i T_i(x_k), \quad k = 1, 2, \dots, N_{train} \quad (2.2)$$

και για την εκπαίδευσή του χρησιμοποίησαν την μέθοδο back-propagation [140] έτσι ώστε να ελαχιστοποιηθεί το σφάλμα J_{SE} (βλ. εξίσωση (1.31)) του ΤΝΔ . Οι Akritas et al. [133] σύγκριναν το τελευταίο ΤΝΔ σε σχέση με ένα εμπροσθοτροφοδοτούμενο σε ένα χαστικό σύνολο δεδομένων που προέκυψαν από τη χρονοσειρά :

$$X_{n+1} = \alpha X_n (1 - X_n), \quad n = 0, 1, \dots \quad (2.3)$$

η οποία ονομάζεται λογιστική απεικόνιση (Logistic map) [141, 142] χρησιμοποιώντας παράμετρο χάους την $\alpha = 4$. Κατέληξαν στο συμπέρασμα ότι το πολυωνυμικό ΤΝΔ προσδίδει πολύ καλύτερα προσεγγιστικά αποτελέσματα σε σχέση με πολυστρωματικά ΤΝΔ (Multi-Layer Preceptron ή MLP). Οι Zou και Kumar [134] εφάρμοσαν το προηγούμενο ΤΝΔ στον έλεγχο της στάσης διαστημικών σκαφών, οι Shaik et al. [135] το χρησιμοποίησαν σε δεδομένα ενός πειραματικού έλικα, ο Patra [136] σε δεδομένα που προέκυψαν από ηλιακούς συλλέκτες και οι Patra και Kot [137] για την αναγνώριση μη-γραμμικών δυναμικών συστημάτων. Όλοι οι παραπάνω καταλήγουν σε καλές προσεγγιστικές δυνατότητες με τη χρήση πολυωνύμων Chebychev σε ΤΝΔ.

Οι Huang et al. [8] δημιούργησαν ένα υβριδικό ΤΝΔ συναρτήσεων ακτινικής βάσης, το HRBFNN, το οποίο χρησιμοποιεί ασαφή συσταδοποίηση (με τη μέθοδο FCM) και πολυωνυμικό ΤΝΔ (pNN). Το HRBFNN στηρίχθηκε στο ασαφές πολυωνυμικό ΤΝΔ (FPNN) των Oh και Pedrycz [143] το οποίο χρησιμοποιεί ασαφείς κανόνες της μορφής :

$$R^1: \text{IF } x \text{ is in cluster } A_i \text{ THEN } y_i = [\text{Polynomial Model}] \quad (2.4)$$

όπου [Polynomial Model] είναι ένα από αυτά των εξισώσεων (1.77)-(1.80). Το HRBFNN μπορεί να θεωρηθεί και ως εκτεταμένη έκδοση των pRBFNN των Oh et al. [6]. Το HRBFNN αποτελείται από πέντε τμήματα, το τμήμα εισόδου, το τμήμα της παραδοχής (premise part), το τμήμα συμπεράσματος (conclusion part), το τμήμα συνάθροισης (aggregation part) και αυτό της εξόδου. Τα τμήματα παραδοχής και συμπεράσματος σχετίζονται με κανόνες ασαφούς λογικής και το τμήμα συνάθροισης με κανόνες ασαφών συμπερασμάτων. Στο HRBFNN πρέπει να σημειωθεί ότι ως είσοδος του δικτύου χρησιμοποιήθηκαν τα αποτελέσματα (οι συνιστώσες) που προκύπτουν έπειτα από εφαρμογή της μεθόδου ανάλυσης σε κύριες συνιστώσες (Principal Component Analysis ή PCA, [144, 145]) στα δεδομένα εισόδου. Για την εκπαίδευση του HRBFNN έγινε η χρήση γενετικών αλγορίθμων [146, 147].

Οι παραπάνω μέθοδοι χρησιμοποιούν τα πολυώνυμα Chebychev ως συναρτήσεις ενεργοποίησης μέσα σε ΤΝΔ ή σε ασαφή συστήματα. Στο παρόν κεφάλαιο γίνεται μια προσπάθεια ώστε ένα πολυωνυμικό ΤΝΔ να μπορέσει να αυξήσει την απόδοση των ήδη υπαρχόντων, για αυτό το λόγο το προτεινόμενο ΤΝΔ χρησιμοποιεί τα πολυώνυμα Chebychev σε συνδυασμό με ΤΝΔ συναρτήσεων ακτινικής βάσης. Τα ΤΝΔ συναρτήσεων ακτινικής βάσης (RBFNN) βελτιώνουν αρκετά την απόδοση των ΤΝΔ και βρίσκουν εφαρμογή σε πολλούς διαφορετικούς επιστημονικούς τομείς [10-16].

Στο παρόν κεφάλαιο, το προτεινόμενο δίκτυο είναι διαφορετικό με τα προηγούμενα της βιβλιογραφίας, δεν χρησιμοποιεί τις πολυωνυμικές συναρτήσεις απευθείας στα δεδομένα εισόδου, ούτε προσεγγίζονται τα δεδομένα εισόδου από πολυώνυμα Chebyshev. Το προτεινόμενο δίκτυο χρησιμοποιεί εκφράσεις των δεδομένων εισόδου ως σειρές πολυωνύμων Chebyshev (Chebyshev expansion series). Το προτεινόμενο δίκτυο συνδυάζει τη χρήση των RBFNN με πολυωνυμικό ΤΝΔ (pNN) το οποίο χρησιμοποιεί τα πολυώνυμα Chebyshev. Η χρήση των πολυωνύμων αυτών δεν γίνεται απευθείας στα δεδομένα εισόδου λόγω του περιορισμού της ορθογωνιότητας των πολυωνύμων στο διάστημα $[-1,1]$ (βλ. παράγραφο 2.2.2) αλλά στα δεδομένα εισόδου εφαρμόζεται πρώτα αλλαγή διαστήματος (βλ. εξίσωση (2.30)). Το προτεινόμενο ΤΝΔ περιέχει κι ένα τμήμα RBF ΤΝΔ, στο οποίο έγινε χρήση μια πρωτότυπης συσταδοποίησης η οποία δίνει καλύτερα αποτελέσματα σε σχέση με άλλες συσταδοποιήσεις όταν χρησιμοποιηθεί σε RBF ΤΝΔ όπως για παράδειγμα αυτό των Uykan et al. [148], το CFCM-RBF του Pedrycz [104] κ.ά.

2.2 Πολυώνυμα Chebychev

2.2.1 Ορισμοί

Τα πολυώνυμα Chebychev (ή Chebyshev) [76, 77] εισήχθησαν από τον *Pafnuty Chebyshev*^{*}, και έχουν τόσες εφαρμογές που μάλιστα ο Forsythe έχει πει ότι βρίσκονται παντού στην αριθμητική ανάλυση [77]. Τα πολυώνυμα αυτά είναι τεσσάρων ειδών, τα 1^ο, 2^ο, 3^ο και 4^ο είδους και συμβολίζονται ως $T_n(x), U_n(x), V_n(x), W_n(x)$, $n = 0, 1, 2, \dots$ αντίστοιχα[†]. Τα πολυώνυμα του 1^ο είδους ορίζονται ως :

$$T_n(x) = \cos n\theta, \text{ όπου } x = \cos \theta \quad (2.5)$$

και είναι οι πολυωνυμικές λύσεις της διαφορικής εξίσωσης:

$$(1-x^2)y'' - xy' + n^2y = 0 \quad (2.6)$$

δηλαδή οι λύσεις της μορφής $y = P_n(x) = \alpha_n x^n + \alpha_{n-1} x^{n-1} + \dots + \alpha_1 x + \alpha_0$ όπου $n = 0, 1, 2, \dots$ ο βαθμός του πολυωνύμου. Αποδεικνύεται (βλ. [149]) ότι ένας ισοδύναμος ορισμός των πολυωνύμων αυτών είναι ο :

$$T_n(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k 2^{2n-k-1} \frac{n}{n-k} \binom{n-k}{k} x^{n-2k} \quad (2.7)$$

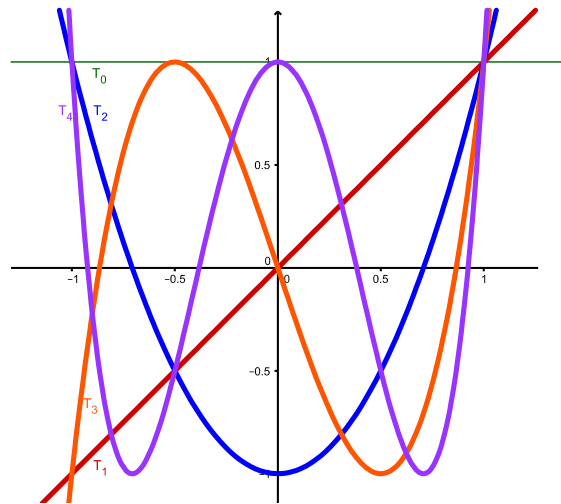
όπου $\lfloor \cdot \rfloor$ η στρογγυλοποίηση στον μικρότερο ακέραιο (ή ακέραιο μέρος) (floor function). Για $n = 0, 1, 2, 3, 4$ προκύπτουν τα πολυώνυμα:

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_4(x) &= 8x^4 - 8x^2 + 1 \end{aligned} \quad (2.8)$$

οι γραφικές παραστάσεις των οποίων παρουσιάζονται στο Σχήμα 2.1.

^{*} Ρώσος Μαθηματικός, 1821-1894, γνωστός για τη συνεισφορά του στις πιθανότητες, στη στατιστική, στη μηχανική και στην αναλυτική γεωμετρία.

[†] Παρόλο που το όνομα του Chebychev ξεκινάει με το γράμμα "C", τα πολυώνυμα 1^ο είδους ξεκινούν με το γράμμα "T" λόγω των μεταφράσεων του ονόματός του ως Tchebyshev στα Γαλλικά και ως Tschebyschow στα Γερμανικά.



Σχήμα 2.1 Τα πρώτα 5 πολυώνυμα Chebyshev πρώτου είδους στο διάστημα $[-1,1]$

Τα πολυώνυμα μπορούν να δοθούν από την αναδρομική σχέση:

$$\left\{ \begin{array}{l} T_0(x) = 1 \\ T_1(x) = x \\ T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), \quad n \geq 2 \end{array} \right\} \quad (2.9)$$

για προγραμματιστικούς σκοπούς. Τα πολυώνυμα 2^{ου}, 3^{ου} και 4^{ου} είδους προκύπτουν αν απλά αλλαχτεί στην αναδρομή σχέση (2.9) το πολυώνυμο πρώτου βαθμού σε $2x, 2x-1, 2x+1$ αντίστοιχα. Δηλαδή, τα πολυώνυμα 2^{ου} είδους είναι τα:

$$\left\{ \begin{array}{l} U_0(x) = 1 \\ U_1(x) = 2x \\ U_n(x) = 2xU_{n-1}(x) - U_{n-2}(x), \quad n \geq 2 \end{array} \right\} \quad (2.10)$$

τα 3^{ου} είδους είναι τα:

$$\left\{ \begin{array}{l} V_0(x) = 1 \\ V_1(x) = 2x - 1 \\ V_n(x) = 2xV_{n-1}(x) - V_{n-2}(x), \quad n \geq 2 \end{array} \right\} \quad (2.11)$$

και 4^{ου} είδους τα:

$$\left\{ \begin{array}{l} W_0(x) = 1 \\ W_1(x) = 2x + 1 \\ W_n(x) = 2xW_{n-1}(x) - W_{n-2}(x), \quad n \geq 2 \end{array} \right\} \quad (2.12)$$

Οι ιδιότητες και των τεσσάρων ειδών πολυωνύμων είναι πάρα πολλές. Ενδεικτικά παρουσιάζονται κάποιες παρακάτω για τα πολυώνυμα 1^{ου} είδους. Για $n > 0$ το κάθε πολυώνυμο $T_n(x)$ έχει n ρίζες για τις τιμές:

$$x = \cos \frac{\left(k - \frac{1}{2}\right)\pi}{n}, \quad k = 1, 2, \dots, n \quad (2.13)$$

και $n+1$ τοπικά ακρότατα στο διάστημα $[-1, 1]$ (Σχήμα 2.1) στα σημεία:

$$x = \cos \frac{(n-k)\pi}{n}, \quad k = 0, 1, \dots, n \quad (2.14)$$

Αν ο n είναι άρτιος, η $T_n(x)$ είναι άρτια συνάρτηση ενώ είναι περιττή για n περιττό (Σχήμα 2.1). Για $x=1$ όλα τα πολυώνυμα έχουν την τιμή 1 και για $x=-1$ δίνουν την τιμή +1 ή -1 ανάλογα με το αν είναι άρτιου ή περιττού βαθμού αντίστοιχα.

Οι δυνάμεις του x μπορούν να γραφούν ως γραμμικός συνδυασμός πολυωνύμων Chebychev ως εξής [150]:

$$x^n = 2^{1-n} \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{k} T_{n-2k}(x) \quad (2.15)$$

όπου το σύμβολο \sum' δηλώνει ότι αντί του πρώτου όρου του αθροίσματος χρησιμοποιείται ο μισός του.

Επειδή σε πολλά προβλήματα χρειάζεται να γίνει χρήση του διαστήματος $[0, 1]$ αντί του $[-1, 1]$ δημιουργήθηκαν τα μετατοπισμένα (shifted) πολυώνυμα Chebychev τα οποία συμβολίζονται ως $T_n^*(x), U_n^*(x), V_n^*(x), W_n^*(x)$. Ο μετασχηματισμός που κάνει την αλλαγή διαστήματος (min-max normalization ή min-max scaling) των δεδομένων $x \in [a, b]$ σε δεδομένα $x' \in [c, d]$ είναι ο [151, 152]:

$$x' = c + (d - c) \cdot \frac{x - a}{b - a} \quad (2.16)$$

άρα για να γίνει η αλλαγή διαστήματος $s \in [-1, 1] \rightarrow x \in [0, 1]$ πρέπει να χρησιμοποιηθεί ο μετασχηματισμός:

$$s = \frac{x+1}{2} \Leftrightarrow x = 2s - 1 \quad (2.17)$$

και άρα τα μετατοπισμένα πολυώνυμα Chebychev $1^{00}-4^{00}$ είδους είναι τα:

$$\left\{ T_n^*(x) = T_n(2x-1); \quad U_n^*(x) = U_n(2x-1); \quad V_n^*(x) = V_n(2x-1); \quad W_n^*(x) = W_n(2x-1) \right\} \quad (2.18)$$

2.2.2 Ορθογωνιότητα

Τα πολώνυμα πρώτου είδους είναι ορθογώνια στο διάστημα $[-1,1]$ χρησιμοποιώντας τη συνάρτηση βαρύτητας: $w(x) = \frac{1}{\sqrt{1-x^2}}$, δηλαδή για το εσωτερικό τους γινόμενο \langle, \rangle ισχύει

ότι:

$$\langle T_i(x), T_j(x) \rangle = \int_{-1}^1 w(x) T_i(x) T_j(x) dx = \int_{-1}^1 \frac{T_i(x) T_j(x)}{\sqrt{1-x^2}} dx = \begin{cases} 0, & i \neq j \\ \frac{\pi}{2}, & i = j \neq 0 \\ \pi, & i = j = 0 \end{cases} \quad (2.19)$$

Τα πολώνυμα 2^ο, 3^ο και 4^ο είδους είναι και αυτά ορθογώνια στο $[-1,1]$ μέσω των συναρτήσεων βαρύτητας που παρουσιάζονται στον πίνακα 1.1. Πιο συγκεκριμένα, ισχύει ότι:

$$\langle U_i(x), U_j(x) \rangle = \int_{-1}^1 w(x) U_i(x) U_j(x) dx = \int_{-1}^1 \sqrt{1-x^2} U_i(x) U_j(x) dx = \begin{cases} 0, & i \neq j \\ \frac{\pi}{2}, & i = j \end{cases} \quad (2.20)$$

$$\langle V_i(x), V_j(x) \rangle = \int_{-1}^1 w(x) V_i(x) V_j(x) dx = \int_{-1}^1 \sqrt{\frac{1+x}{1-x}} V_i(x) V_j(x) dx = \begin{cases} 0, & i \neq j \\ \pi, & i = j \end{cases} \quad (2.21)$$

$$\langle W_i(x), W_j(x) \rangle = \int_{-1}^1 w(x) W_i(x) W_j(x) dx = \int_{-1}^1 \sqrt{\frac{1-x}{1+x}} W_i(x) W_j(x) dx = \begin{cases} 0, & i \neq j \\ \pi, & i = j \end{cases} \quad (2.22)$$

Η ορθογωνιότητα αυτή δίνει στα πολώνυμα αυτά μία ευρεία γκάμα ιδιοτήτων, κάποιες από τις οποίες παρουσιάζονται αμέσως μετά.

2.2.3 Μέθοδοι παρεμβολής (interpolation)

Για να αναλυθούν οι μέθοδοι παρεμβολής με χρήση πολυωνύμων Chebychev, είναι απαραίτητο να δοθεί πρώτα το επόμενο θεώρημα.

Θεώρημα 2.1 Το μοναδικό σύστημα των κανονικών* πολυωνύμων $\{\varphi_i\}$ ώστε κάθε πολώνυμο από αυτά να είναι βαθμού $\deg(\varphi_i) = i$ και τα πολώνυμα αυτά να είναι ορθογώνια στο $[\alpha, \beta]$ μέσω της συνάρτησης βαρύτητας $w(x)$ είναι τα:

* Κανονικό (monic) ονομάζεται το πολώνυμο που ο συντελεστής του μεγιστοβάθμιου όρου του είναι η μονάδα.

$$\left\{ \begin{array}{l} \varphi_0(x) = 1 \\ \varphi_1(x) = x - \alpha_1 \\ \varphi_n(x) = (x - \alpha_n)\varphi_{n-1}(x) - \beta_n\varphi_{n-2}(x) \end{array} \right\} \quad (2.23)$$

όπου οι συντελεστές α_n, β_n προσδιορίζονται από τις εξισώσεις:

$$\alpha_n = \frac{\langle x\varphi_{n-1}, \varphi_{n-1} \rangle}{\langle \varphi_{n-1}, \varphi_{n-1} \rangle}, \quad \beta_n = \frac{\langle \varphi_{n-1}, \varphi_{n-1} \rangle}{\langle \varphi_{n-2}, \varphi_{n-2} \rangle} \quad (2.24) \quad \blacksquare$$

Απόδειξη του θεωρήματος αυτού μπορεί να βρεθεί στην παράγραφο 4.4 του βιβλίου [77]. Πρέπει να σημειωθεί πως αναλόγως τη συνάρτηση βαρύτητας $w(x)$ και το διάστημα $[\alpha, \beta]$ που θα επιλεγθεί (η $w(x)$ και το $[\alpha, \beta]$ βρίσκονται 'κρυμμένα' στα εσωτερικά γινόμενα της (2.24)), αλλάζουν τα ορθογώνια πολυώνυμα στα οποία καταλήγει η (2.23). Αν δηλαδή επιλεγεί ως διάστημα το $[-1, 1]$ και $w(x) = \frac{1}{\sqrt{1-x^2}}$ καταλήγουμε στα $T_n(x)$, αν επιλεγεί η $w(x) = \sqrt{1-x^2}$ καταλήγουμε στα $U_n(x)$ κ.ο.κ. Επίσης το παραπάνω θεώρημα αποδεικνύει τις αναδρομικές εξισώσεις (2.9)-(2.12).

Οι μέθοδοι παρεμβολής (interpolation) των Forsythe–Clenshaw προσεγγίζουν ένα σύνολο σημείων $(x_k, y_k), k = 1, 2, \dots, m$ μέσω ενός πολυωνύμου $P(x)$ βαθμού n , με $n + 1 \leq m$. Η προσέγγιση γίνεται ώστε να ελαχιστοποιηθεί το τετραγωνικό σφάλμα: $\sum_{k=1}^m (y_k - P(x_k))^2$, όπως και στη μέθοδο των ελαχίστων τετραγώνων.

Σύμφωνα με τη μέθοδο του Forsythe [153] το πολυώνυμο αυτό είναι το:

$$P(x) = \sum_{i=0}^n c_i \varphi_i(x) \quad (2.25)$$

όπου τα πολυώνυμα φ_i είναι αυτά της εξίσωσης (2.23) και οι συντελεστές c_i προσδιορίζονται από την σχέση:

$$c_i = \frac{\langle y, \varphi_i \rangle}{\langle \varphi_i, \varphi_i \rangle} = \frac{\sum_{k=1}^m w_k y_k \varphi_i(x_k)}{\sum_{k=1}^m w_k [\varphi_i(x_k)]^2} \quad (2.26)$$

στην οποία έγινε χρήση του διακριτού εσωτερικού γινομένου:

$$\langle u, v \rangle = \sum_{k=1}^m w_k u(x_k) v(x_k) \quad (2.27)$$

Η μέθοδος μέχρι εδώ δεν εμπεριέχει πολυώνυμα Chebyshev, αν όμως τα σημεία έχουν επιλεγεί ομοιόμορφα κατανεμημένα στο διάστημα $[-1,1]$ και η συνάρτηση βαρύτητας είναι η $w_k = \frac{1}{\sqrt{1-x_k^2}}$ δηλαδή αυτή των πολυωνύμων Chebyshev πρώτου είδους $T_n(x)$, τότε τα πολυώνυμα $\varphi_i(x)$ θα θυμίζουν τα $T_n(x)$ [77].

Ο Clenshaw [154] χρησιμοποίησε αντί για τα πολυώνυμα φ_i της σχέσης (2.25) κατευθείαν τα πολυώνυμα $T_n(x)$ και άρα το πολυώνυμο που κατέληξε είναι της μορφής:

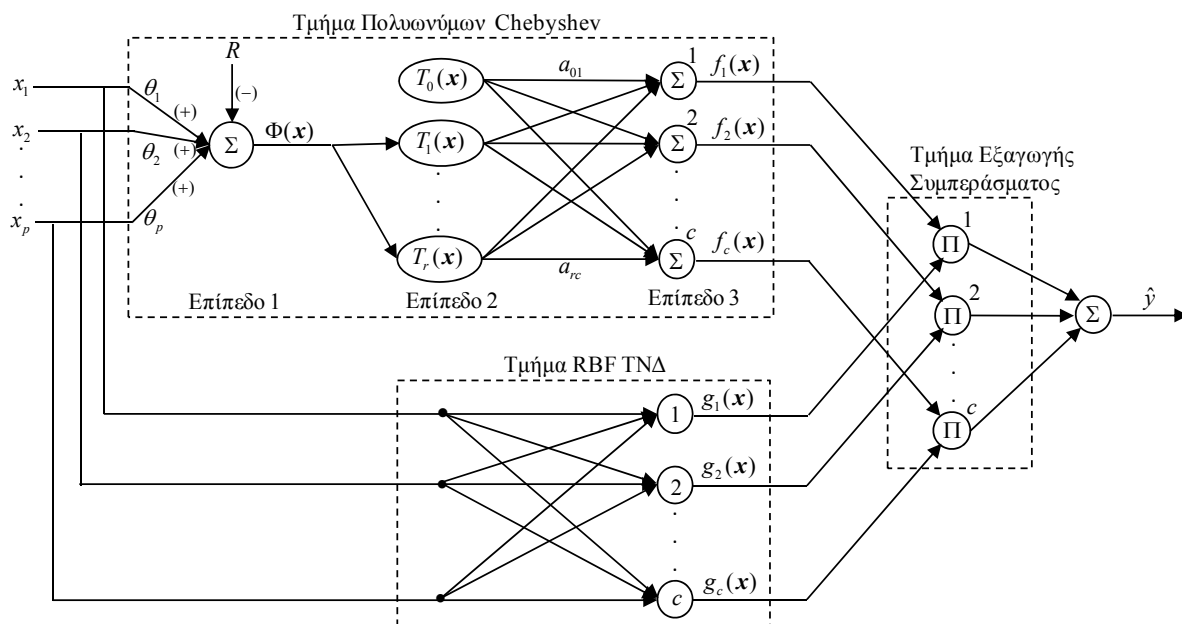
$$P(x) = \sum_{i=0}^n d_i T_i(x) \quad (2.28)$$

και τους συντελεστές d_i τους επέλεξε ώστε η εξίσωση (2.28) να καταλήγει στο ίδιο αποτέλεσμα με την εξίσωση (2.25).

Μία εφαρμογή προσέγγισης δεδομένων μέσω πολυωνύμων Chebyshev έχει γίνει από τους Rigos et al. [155] οι οποίοι στηρίχθηκαν στη μέθοδο παρεμβολής του Forsythe και χρησιμοποίησαν τους συντελεστές c_i της εξίσωσης (2.26) ως δεδομένα εισόδου ενός ΤΝΔ.

2.3 Το προτεινόμενο ΤΝΔ με χρήση πολυωνύμων Chebyshev

Το προτεινόμενο ΤΝΔ [1] αποτελείται από τρία τμήματα και η τοπολογία του παρουσιάζεται στο Σχήμα 2.4. Το πρώτο τμήμα του ΤΝΔ (*Τμήμα Πολυωνύμων Chebyshev*) αποτελείται από ένα πολυωνυμικό ΤΝΔ το οποίο χρησιμοποιεί $r + 1$ στο πλήθος πολυώνυμα Chebyshev 1^{ου} είδους ως συναρτήσεις ενεργοποίησης. Το Τμήμα Πολυωνύμων Chebyshev αποτελείται από τρία επίπεδα, στο 1^ο επίπεδο γίνονται μετασχηματισμοί στα δεδομένα εισόδου ώστε να κανονικοποιηθούν στο διάστημα $[-1,1]$, στο 2^ο επίπεδο τα δεδομένα εισάγονται στους κόμβους που αποτελούνται από πολυώνυμα Chebyshev και το 3^ο επίπεδο δίνει την έξοδο αυτού του τμήματος του δικτύου. Το δεύτερο τμήμα είναι ένα κλασικό ΤΝΔ συναρτήσεων ακτινικής βάσης (RBFNN) με c κρυφούς κόμβους όπως αυτό παρουσιάστηκε στην παράγραφο 1.3.1 το οποίο χρησιμοποιεί μια πρωτότυπη συσταδοποίηση δεδομένων χώρου εισόδου-εξόδου. Τέλος στο τμήμα εξαγωγής συμπεράσματος του δικτύου γίνεται η συνένωση των αποτελεσμάτων των δύο προηγούμενων ΤΝΔ και προκύπτει η έξοδος του δικτύου.



Σχήμα 2.2 Το προτεινόμενο ΤΝΔ με τα πολυώνυμα Chebyshev

Στόχος του προτεινόμενου ΤΝΔ είναι η προσέγγιση ισχυρά μη-γραμμικών συνόλων δεδομένων. Για το λόγο αυτό έγινε η συνένωση δύο ΤΝΔ που εφαρμόζονται σε μη-γραμμικά συστήματα, ενός πολυωνυμικού ΤΝΔ που και ενός ΤΝΔ συναρτήσεων ακτινικής βάσεως (RBFNN). Και τα πολυωνυμικά ΤΝΔ χρησιμοποιούνται στη βιβλιογραφία σε μη-γραμμικά

συστήματα (βλ. παραγράφους 1.2.4 και 2.1) αλλά και για τα RBFNN έχει αποδειχθεί στη βιβλιογραφία ότι δίνουν ικανοποιητικά αποτελέσματα σε μη-γραμμικά δεδομένα [10-16]. Ακολουθεί η ανάλυση του προτεινόμενου ΤΝΔ.

2.3.1 Το τμήμα πολυωνύμων Chebyshev του ΤΝΔ

Υποθέτουμε ότι τα δεδομένα εισόδου-εξόδου του ΤΝΔ έχουν τη μορφή που έχει περιγραφεί και στο πρώτο κεφάλαιο:

$$S = \{(\mathbf{x}_i, y_i) / \mathbf{x}_i \in \mathbb{R}^p, y_i \in \mathbb{R}, i = 1, 2, \dots, N\} \quad (2.29)$$

για την ανάλυση που ακολουθεί. Το τμήμα του ΤΝΔ με τα πολυώνυμα Chebyshev αποτελείται από τρία επίπεδα. Στο πρώτο επίπεδο γίνεται η αλλαγή διαστήματος των δεδομένων εισόδου $\mathbf{x} = [x_1, x_2, \dots, x_p]^T$ ώστε να προκύψει από αυτά η συνάρτηση $\Phi(\mathbf{x})$, η οποία να ανήκει στο διάστημα $[-1, 1]$. Όπως αναφέρθηκε προηγουμένως, τα πολυώνυμα Chebyshev είναι ορθογώνια μόνο στο διάστημα $[-1, 1]$, για το λόγο αυτό στα δεδομένα εισόδου πρέπει να εφαρμοστεί διαδικασία αλλαγής διαστήματος. Υποθέτοντας ότι οι συνιστώσες x_1, x_2, \dots, x_p των δεδομένων εισόδου ανήκουν στο διάστημα $[L_j, U_j] = [\min_{1 \leq k \leq N} (x_{kj}), \max_{1 \leq k \leq N} (x_{kj})]$, $1 \leq j \leq p$ τότε για να εφαρμοστεί σε αυτές η κλασσική αλλαγή διαστήματος (min-max data scaling) και να μετατραπούν από δεδομένα $x_j \in [L_j, U_j]$ σε δεδομένα $\hat{x}_j \in [-1, 1]$ ο μετασχηματισμός που πρέπει να εφαρμοστεί είναι (βλ. εξίσωση (2.16)) :

$$\hat{x}_j = -1 + \frac{1 - (-1)}{U_j - L_j} (x_j - L_j) = \frac{2}{U_j - L_j} x_j - \frac{U_j + L_j}{U_j - L_j} \quad (2.30)$$

Τότε η συνάρτηση του 1^{ου} επιπέδου είναι η:

$$\Phi(\mathbf{x}) = \frac{1}{p} \sum_{j=1}^p \hat{x}_j = \frac{1}{p} \sum_{j=1}^p \left(\frac{2}{U_j - L_j} x_j - \frac{U_j + L_j}{U_j - L_j} \right) \quad (2.31)$$

Η συνάρτηση $\Phi(\mathbf{x})$ πράγματι παίρνει τιμές στο διάστημα $[-1, 1]$, αφού:

$$-1 \leq \hat{x}_j \leq 1 \Rightarrow -p \leq \sum_{j=1}^p \hat{x}_j \leq p \Rightarrow -1 \leq \frac{1}{p} \sum_{j=1}^p \hat{x}_j \leq 1 \Rightarrow -1 \leq \Phi(\mathbf{x}) \leq 1 \quad (2.32)$$

Ακολουθώς, θέτοντας τις σταθερές (σ.σ. εξαρτώνται μόνο από τα δεδομένα εισόδου) της εξίσωσης (2.31) ως: $\theta_j = \frac{2}{p(U_j - L_j)}$ και $R = \frac{1}{p} \sum_{j=1}^p \frac{U_j + L_j}{U_j - L_j}$ η εξίσωση (2.31) παίρνει τη μορφή :

$$\Phi(\mathbf{x}) = \sum_{j=1}^p \theta_j x_j - R \quad (2.33)$$

η οποία είναι γραμμική ως προς τα δεδομένα εισόδου, το οποίο είναι πολύ σημαντικό για την ανάλυση που ακολουθεί. Στο δεύτερο επίπεδο τα δεδομένα $\Phi(\mathbf{x})$ περνούν μέσα από $r+1$ στο πλήθος πολυώνυμα Chebyshev, τα πολυώνυμα μηδενικού ως και r -στού βαθμού. Η έξοδος του 2^{ου} επιπέδου αυτού του τμήματος του δικτύου είναι τότε η:

$$T_\ell(\Phi(\mathbf{x})) = T_\ell\left(\sum_{j=1}^p \theta_j x_j - R\right), \quad 0 \leq \ell \leq r \quad (2.34)$$

όπου T_ℓ το πολυώνυμο Chebyshev 1^{ου} είδους βαθμού ℓ . Επειδή το πολυώνυμο μηδενικού βαθμού είναι σταθερό ($T_0(x) = 1$), αυτό δεν επηρεάζεται από τα δεδομένα εισόδου του δικτύου και μπορεί να θεωρηθεί ως bias του δικτύου.

Τέλος, στο 3^ο επίπεδο του δικτύου γίνεται άθροιση των δεδομένων του προηγούμενου κόμβου μέσω των συντελεστών βαρύτητας $a_{\ell i}$ ($0 \leq \ell \leq r, 1 \leq i \leq c$) όπου c το πλήθος των κρυφών κόμβων που χρησιμοποιεί το RBF τμήμα του προτεινόμενου δικτύου. Άρα η έξοδος του Chebyshev τμήματος του δικτύου είναι η:

$$f_i(\mathbf{x}) = \sum_{\ell=0}^r a_{\ell i} T_\ell(\mathbf{x}) \quad (2.35)$$

Το πρώτο τμήμα του ΤΝΔ που μόλις περιγράφηκε είναι ένα πολυωνυμικό ΤΝΔ.

2.3.2 Το RBF τμήμα του ΤΝΔ

Το δεύτερο τμήμα του προτεινόμενου ΤΝΔ, αποτελείται από ένα κλασσικό ΤΝΔ ακτινικής βάσης το οποίο χρησιμοποιεί c στο πλήθος γκαουσιανές συναρτήσεις ακτινικής βάσεις, τις:

$$g_i(\mathbf{x}_k) = \exp\left(-\left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\sigma_i}\right)^2\right), \quad 1 \leq i \leq c \quad (2.36)$$

όπου \mathbf{v}_i το κέντρο της i -στης συνάρτησης ακτινικής βάσης και σ_i το αντίστοιχο πλάτος. Η επιλογή των κέντρων γίνεται με συσταδοποίηση του χώρου εισόδου-εξόδου η οποία παίρνει υπόψη της την κατανομή των δεδομένων εξόδου και με αυτόν τον τρόπο δίνει καλύτερες προσεγγιστικές δυνατότητες στο ΤΝΔ [54]. Για την συσταδοποίηση αυτή ακολουθήθηκαν δύο στάδια, στο πρώτο στάδιο γίνεται ασαφής συσταδοποίηση του χώρου εξόδου για την επίτευξη της οποίας, γίνεται ελαχιστοποίηση της αντικειμενικής συνάρτησης συσταδοποίησης :

$$J_{OS} = \sum_{k=1}^N \sum_{i=1}^c (\mu_{ik})^m (y_k - o_i)^2 \quad (2.37)$$

υπό τον περιορισμό :

$$\sum_{i=1}^c \mu_{ik} = 1, \quad k = 1, 2, \dots, N \quad (2.38)$$

όπου $o_i \in \mathbb{R}$ το κέντρο συσταδοποίησης που προκύπτει για τον χώρο εξόδου, μ_{ik} η συνάρτηση συμμετοχής του δεδομένου y_k στην ασαφή συστάδα κέντρου o_i και $m > 1$ η παράμετρος της ασαφούς συσταδοποίησης (συνήθως χρησιμοποιείται η τιμή $m = 2$). Το παραπάνω πρόβλημα έχει ως λύση τις συναρτήσεις συμμετοχής:

$$\mu_{ik} = 1 / \sum_{j=1}^c \left(\frac{|y_k - o_i|}{|y_k - o_j|} \right)^{\frac{2}{m-1}} \quad (2.39)$$

και τα κέντρα συσταδοποίησης :

$$o_i = \frac{\sum_{k=1}^N (\mu_{ik})^m y_k}{\sum_{k=1}^N (\mu_{ik})^m} \quad (2.40)$$

Αφού γίνει η συσταδοποίηση του χώρου εξόδου, τότε αυτή καθοδηγεί τη συσταδοποίηση του χώρου εισόδου λαμβάνοντας ως αντικειμενική συνάρτηση της συσταδοποίησης τη:

$$J_{IS} = \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \left(\|\mathbf{x}_k - \mathbf{v}_i\|^2 + \gamma |y_k - o_i|^2 \right) \quad (2.41)$$

όπου γ μία παράμετρος η οποία καθοδηγεί τη συσταδοποίηση των δεδομένων εισόδου σύμφωνα με την πληροφορία που προέκυψε από τη συσταδοποίηση των δεδομένων εξόδου και μεταβάλλεται ως εξής :

$$\gamma = \frac{\gamma_0}{1 + t \tau} \quad (2.42)$$

με γ_0 την αρχική τιμή της παραμέτρου γ , t το βήμα της επανάληψης του αλγόριθμου της συσταδοποίησης και τ μια παράμετρος. Τότε η λύση του παραπάνω προβλήματος είναι οι συναρτήσεις συμμετοχής:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|^2 + \gamma|y_k - o_i|}{\|\mathbf{x}_k - \mathbf{v}_j\|^2 + \gamma|y_k - o_j|} \right)^{\frac{1}{m-1}}} \quad (2.43)$$

και τα κέντρα συσταδοποίησης :

$$\mathbf{v}_i = \frac{\sum_{k=1}^N (u_{ik})^m \mathbf{x}_k}{\sum_{k=1}^N (u_{ik})^m} \quad (2.44)$$

Ακολουθεί η διαδικασία της συσταδοποίησης αυτής ως αλγόριθμος.

Ο Αλγόριθμος συσταδοποίησης εισόδου-εξόδου.

Βήμα 1^ο: (Συσταδοποίηση χώρου εξόδου)

Επέλεξε την τιμή την παραμέτρου ασάφειας m . Εφάρμοσε τη συσταδοποίηση του χώρου εξόδου με τη μέθοδο FCM (βλ. παράγραφο 1.3.1.1) σε c συστάδες. Τότε προκύπτουν τα κέντρα o_i των συστάδων και οι συναρτήσεις συμμετοχής μ_{ik} που υπολογίζονται από την επαναληπτική εφαρμογή των εξισώσεων (2.40) και (2.39) αντίστοιχα.

Βήμα 2^ο: (Συσταδοποίηση του χώρου εισόδου που καθοδηγείται από τη συσταδοποίηση των δεδομένων εξόδου).

- Επέλεξε τιμές για τα γ_0, τ και θέσε $t = 0$.
- Επέλεξε τυχαίες τιμές για τα c αρχικά κέντρα συσταδοποίησης $\mathbf{v}_i \in \mathbb{R}^p, i = 1, \dots, c$ των δεδομένων εισόδου $\mathbf{x}_k, k = 1, \dots, N$.
- **Επανάλαβε:**
 - $t = t + 1$
 - Υπολόγισε την παράμετρο γ από την εξίσωση (2.42)
 - Υπολόγισε τις συναρτήσεις συμμετοχής u_{ik} κάθε δεδομένου \mathbf{x}_k σε κάθε συστάδα κέντρου \mathbf{v}_i από την εξίσωση (2.43)
 - Υπολόγισε τα νέα κέντρα \mathbf{v}_i από την εξίσωση (2.44)

Έως ότου ή η μέθοδος να συγκλίνει, δηλαδή τα κέντρα συσταδοποίησης \mathbf{v}_i που προκύπτουν μέσα στην επανάληψη να συγκλίνουν ή να φτάσει η μέθοδος σε κάποιο προκαθορισμένο μέγιστο πλήθος επαναλήψεων.

Η συσταδοποίηση χώρου εισόδου-εξόδου που περιγράφηκε δημιουργεί c στο πλήθος συστάδες \mathbf{C}_i κέντρων \mathbf{v}_i , $i=1, \dots, c$ τα οποία είναι τα κέντρα των συναρτήσεων ακτινικής βάσης αυτού του μέρους του ΤΝΔ (Τμήμα RBF ΤΝΔ). Προκειμένου να υπολογιστούν τα πλάτη των συναρτήσεων ακτινικής βάσης, ακολουθήθηκε η διαδικασία που περιγράφεται στο άρθρο [156] σύμφωνα με την οποία για κάθε συστάδα \mathbf{C}_i ορίζουμε το σύνολο G_i ως:

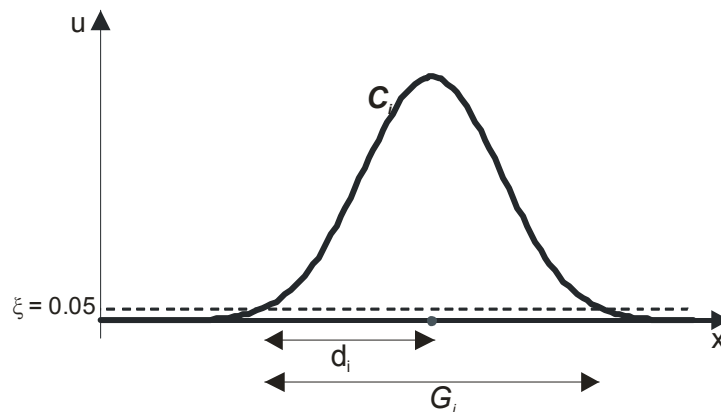
$$G_i = \{\mathbf{x}_k \in \mathbf{C}_i / u_{ik} \geq \xi \text{ με } \xi \in (0,1) \text{ και } 1 \leq k \leq N\} \quad (2.45)$$

δηλαδή κρατιούνται μόνο τα δεδομένα \mathbf{x}_k από τα οποία η συνάρτηση συμμετοχής τους u_{ik} είναι μεγαλύτερη ή ίση από κάποια τιμή-κατώφλι $\xi \in (0, 1)$ η οποία έχει επιλεγεί εκ των προτέρων (Σχήμα 2.3). Τότε για τα σύνολα G_i ορίζεται η τιμή:

$$d_i = \max_{\mathbf{x}_k \in G_i} \{\|\mathbf{x}_k - \mathbf{v}_i\|\} \quad (2.46)$$

και ως πλάτη στις γκαουσιανές συναρτήσεις (2.36) χρησιμοποιούνται οι ποσότητες:

$$\sigma_i = \frac{2}{3} d_i \quad (2.47)$$



Σχήμα 2.3 Η συσταδοποίηση που χρησιμοποιείται στην παράγραφο αυτή διατηρεί από ένα ασαφές σύνολο \mathbf{C}_i μόνο τα δεδομένα με συναρτήσεις συμμετοχής πάνω από κάποια τιμή-κατώφλι ξ , π.χ. την $\xi = 0.05$.

Τότε προκύπτει το σύνολο G_i και η απόσταση d_i που ορίζεται ως $d_i = \max_{\mathbf{x}_k \in G_i} \{\|\mathbf{x}_k - \mathbf{v}_i\|\}$. Το παράδειγμα του παρόντος σχήματος αναφέρεται σε μονοδιάστατο ασαφές σύνολο.

2.3.3 Το τμήμα εξαγωγής συμπεράσματος του ΤΝΔ

Το τρίτο τμήμα του ΤΝΔ είναι το τμήμα εξαγωγής συμπεράσματος χρησιμοποιώντας τα δεδομένα από τα δύο προηγούμενα τμήματα του ΤΝΔ. Όπως φαίνεται και στο Σχήμα 2.4 για κάθε μία έξοδο $f_i(\mathbf{x})$ της εξίσωσης (2.35) του πρώτου τμήματος (τμήμα πολυωνύμων Chebyshev) του ΤΝΔ αντιστοιχεί μία έξοδος $g_i(\mathbf{x})$ της εξίσωσης (2.36) από το δεύτερο τμήμα (RBF τμήμα) του ΤΝΔ. Η έξοδος τότε του ΤΝΔ είναι το άθροισμα των γινομένων των c στο πλήθος αντίστοιχων εξόδων των δύο προηγούμενων τμημάτων του ΤΝΔ, δηλαδή είναι η:

$$\hat{y} = \sum_{i=1}^c f_i(\mathbf{x})g_i(\mathbf{x}) = \sum_{i=1}^c \left(\sum_{\ell=0}^r \alpha_{\ell i} T_{\ell}(\mathbf{x}) \right) g_i(\mathbf{x}) \quad (2.48)$$

2.4 Εκπαίδευση του ΤΝΔ

Εφόσον στο δεύτερο μέρος του ΤΝΔ προσδιορίστηκαν οι παράμετροι \mathbf{v}_i και σ_i , οι παράμετροι του ΤΝΔ που απομένει να προσδιοριστούν είναι τα συναπτικά βάρη $a_{\ell i}$ με $0 \leq \ell \leq r$ και $1 \leq i \leq c$ έτσι ώστε να ελαχιστοποιείται το ολικό τετραγωνικό σφάλμα του δικτύου:

$$J_{SE} = \sum_{k=1}^N (y_k - \hat{y}_k)^2 = \sum_{k=1}^N \left[y_k - \left(\sum_{i=1}^c \left(\sum_{\ell=0}^r \alpha_{\ell i} T_{\ell}(\mathbf{x}) \right) g_i(\mathbf{x}) \right) \right]^2 \quad (2.49)$$

για αυτό το σκοπό χρησιμοποιήθηκε ο αλγόριθμος Gradient Descent (βλ. παράγραφο 1.1.2) σύμφωνα με τον οποίο τα αρχικά συναπτικά βάρη $a_{\ell i}(0)$ (αρχική επανάληψη αλγορίθμου: $t=0$) επιλέγονται στη τύχη μέσα σε κάποιο διάστημα (εδώ χρησιμοποιήθηκε το διάστημα $[-1,1]$) και ανανεώνονται ως :

$$\alpha_{\ell i}(t+1) = \alpha_{\ell i}(t) - \eta(t) \cdot \nabla J_{SE}(t) \quad (2.50)$$

όπου η βαθμίδα (gradient) $\nabla J_{SE}(t)$ περιέχει τις μερικές παραγώγους της εξίσωσης (2.49) ως προς τα συναπτικά βάρη $\alpha_{\ell i}$ οι οποίες είναι οι:

$$\frac{\partial J_{SE}}{\partial \alpha_{\ell i}}(t) = -2 \sum_{k=1}^N (y_k - \hat{y}_k) T_{\ell}(\mathbf{x}_k) g_i(\mathbf{x}_k) \quad (2.51)$$

και $\eta(t)$ ο ρυθμός εκμάθησης (learning rate) της μεθόδου Gradient Descent ο οποίος καθορίζεται από τον κανόνα του Armijo (βλ. [17-19] και παράγραφο 1.1.2). Ο $\eta(t)$ ορίζεται ως: $\eta(t) = c^\lambda$ όπου c μια σταθερά $c \in (0,1)$ και λ ο ελάχιστος θετικός ακέραιος τέτοιος ώστε να ισχύει ότι:

$$J_{SE}(t+1) - J_{SE}(t) < -\varepsilon \eta(t) \|\nabla J_{SE}(t)\|^2 \quad (2.52)$$

με ε έναν μικρό θετικό αριθμό, τέτοιον ώστε $\varepsilon \in (0,1)$. Στη διαδικασία Gradient Descent επιλέχθηκε ως κριτήριο τερματισμού η σύγκλιση της μεθόδου, δηλαδή ο αλγόριθμος τερματίζει όταν σε κάποια επανάληψή του ισχύει ότι:

$$|J_{SE}(t+1) - J_{SE}(t)| < \varepsilon' \quad (2.53)$$

όπου ε' ένας μικρός θετικός αριθμός.

2.5 Αριθμητικές προσομοιώσεις σε σύνολα δεδομένων του UCI Repository

Προκειμένου να ελεγχθεί η απόδοση του προτεινόμενου ΤΝΔ επιλέχθηκαν από το UCI repository (βλ. [100] και παράρτημα Α) τα σύνολα δεδομένων AutoMPG, ForestFires, Housing και Servo για να γίνουν αριθμητικές προσομοιώσεις σε αυτά. Οι προσομοιώσεις έγιναν χρησιμοποιώντας $r = 1, 2, 3, 4$ ως μέγιστο βαθμό πολυωνύμων Chebychev των εξισώσεων (2.34)-(2.35) και $c = 4, 6, 8, 10, 12, 15$ για πλήθος των κρυφών κόμβων του 2^{ου} τμήματος του ΤΝΔ (τμήμα RBF ΤΝΔ). Κατά την διαδικασία της εκπαίδευσης, τα αρχικά συναπτικά βάρη $\alpha_{\ell i}$ ($0 \leq \ell \leq r; 1 \leq i \leq c$) του 1^{ου} τμήματος του ΤΝΔ, επιλέχθηκαν τυχαία στο διάστημα $[-1,1]$. Στο 2^ο τμήμα του ΤΝΔ ως παράμετρος ασάφειας των (2.37)-(2.44) επιλέχθηκε η $m = 2$, στην (2.42) χρησιμοποιήθηκαν οι παράμετροι $\gamma_0 = 1$ και $\tau = 2$ και ως τιμή-κατώφλι της εξίσωσης (2.45) χρησιμοποιήθηκε η τιμή $\xi = 0.05$. Τέλος, στη διαδικασία της εκπαίδευσης του ΤΝΔ χρησιμοποιήθηκε στον ρυθμό εκμάθησης $\eta(t) = c^\lambda$ του κανόνα του Armijo της εξίσωσης (2.52) η τιμή $c = 0.01$, στην εξίσωση (2.52) χρησιμοποιήθηκε η τιμή $\varepsilon = 10^{-3}$ και για το κριτήριο τερματισμού της μεθόδου Gradient Descent στην εξίσωση (2.53) η τιμή $\varepsilon' = 10^{-5}$.

Για να συγκριθεί η απόδοση του προτεινόμενου ΤΝΔ με άλλα τις βιβλιογραφίας, χρησιμοποιήθηκαν δύο ακόμη ΤΝΔ, τα οποία περιγράφονται ακολούθως:

(α) Ένα εμπροσθοτροφοδοτούμενο νευρωνικό δίκτυο (FFNN) το οποίο χρησιμοποιεί ως συνάρτηση ενεργοποίησης την $f(x) = \tanh \frac{x}{2}$ και έχει $(p+1) \cdot c$ στο πλήθος παραμέτρους (Σχήμα 1.13) όπου p η διάσταση χώρου εισόδου και c το πλήθος των κρυφών κόμβων. Η έξοδος του i -στού κρυφού κόμβου του FFNN είναι :

$$f_i(\mathbf{x}) = f\left(\sum_{j=1}^p \alpha_{ji} x_j\right) = \tanh\left(\frac{1}{2} \sum_{j=1}^p \alpha_{ji} x_j\right), \quad 1 \leq i \leq c \quad (2.54)$$

όπου α_{ji} με $1 \leq j \leq p, 1 \leq i \leq c$ οι συντελεστές βαρύτητας μεταξύ των δεδομένων εισόδου και του πρώτου κόμβου. Η έξοδος του FFNN είναι η:

$$\hat{y} = \sum_{i=1}^c w_i f_i(\mathbf{x}) = \sum_{i=1}^c w_i \tanh\left(\frac{1}{2} \sum_{j=1}^p \alpha_{ji} x_j\right) \quad (2.55)$$

όπου w_i τα συναπτικά βάρη μεταξύ των δεδομένων του κρυφού κόμβου και τις εξόδου. Για να εκπαιδευτεί το FFNN χρησιμοποιήθηκε η μέθοδος Gradient Descent με τις ίδιες παραμέτρους όπως και στο προτεινόμενο ΤΝΔ ενώ για την αρχικοποίηση των συναπτικών βαρών α_{ji} και w_i χρησιμοποιήθηκαν τυχαίες τιμές στο διάστημα $[-5, 5]$.

(β) Δημιουργήθηκε επίσης ένα πολυωνυμικό ΤΝΔ ακτινικής βάσης που θα συμβολίζεται εδώ ως pRBFN. Το pRBFN είναι το πολυωνυμικό ΤΝΔ που δημιουργήθηκε από τους Oh et al. [6] και αναλύθηκε στην παράγραφο 1.3.4. Ως πολυωνυμικό μοντέλο στο pRBFN χρησιμοποιήθηκε το γραμμικό (linear) της εξίσωσης (1.77). Στο 1^ο τμήμα (RBF τμήμα) του δικτύου αυτού (αναφέρεται ως Condition phase στο [6]) χρησιμοποιήθηκαν οι γκαουσιανές συναρτήσεις ακτινικής βάσης της εξίσωσης (2.36), των οποίων τα κέντρα υπολογίστηκαν με τη μέθοδο συσταδοποίησης CFCM (βλ. παράγραφο 1.3.1.1) και ως πλάτη των συναρτήσεων αυτών χρησιμοποιήθηκαν τα:

$$\sigma_i = \min_{\substack{1 \leq j \leq c \\ j \neq i}} \{\|\mathbf{v}_i - \mathbf{v}_j\|\} \quad (2.56)$$

όπως προτείνουν οι Saha και Keeleer [106]. Η έξοδος του δικτύου αυτού είναι η :

$$\hat{y} = \sum_{i=1}^c g_i(\mathbf{x}) h_i(\mathbf{x}) \quad (2.57)$$

όπου $h_i(\mathbf{x})$ είναι το γραμμικό μοντέλο:

$$h_i(\mathbf{x}) = \beta_{i0} + \sum_{j=1}^p \beta_{ij} x_j \quad (2.58)$$

Ο προσδιορισμός των παραμέτρων β_{ij} , $1 \leq i \leq c, 1 \leq j \leq p$ έγινε χρησιμοποιώντας τη μέθοδο Gradient Descent ώστε να ελαχιστοποιηθεί η συνάρτηση σφάλματος του pRBFN:

$$J_{SE} = \sum_{k=1}^N (y_k - \hat{y}_k)^2 = \sum_{k=1}^N \left(y_k - \sum_{i=1}^c g_i(\mathbf{x}_k) h_i(\mathbf{x}_k) \right)^2 \quad (2.59)$$

Σε όλες τις αριθμητικές προσομοιώσεις τα σύνολα δεδομένων εισόδου-εξόδου χωρίστηκαν σε δεδομένα εκπαίδευσης και δεδομένα ελέγχου με τη διαδικασία 10-fold cross validation (βλ. παράγραφο 1.3) και όλες οι προσομοιώσεις έγιναν με 10 διαφορετικές αρχικοποιήσεις (initializations). Τα αποτελέσματα των ΤΝΔ παρουσιάζονται στους πίνακες 2.1-2.12 όπου παρουσιάζεται το σφάλμα $RMSE$ (βλ. εξίσωση (1.33)) στη μορφή: (μέση τιμή) \pm (τυπική απόκλιση).

Πίνακας 2.1 Αποτελέσματα (τιμές των RMSEs) του προτεινόμενου ΤΝΔ για $r = 1$ και $r = 2$ στο σύνολο δεδομένων **AutoMPG** του UCI Repository

c	Προτεινόμενο, $r = 1$		Προτεινόμενο, $r = 2$	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα ελέγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα ελέγχου
4	6.2959 \pm 0.1768	6.4264 \pm 1.5314	5.4238 \pm 0.2679	7.2315 \pm 1.3247
6	6.2958 \pm 0.1768	6.4273 \pm 1.5271	5.4208 \pm 0.2679	7.2597 \pm 1.3210
8	6.2958 \pm 0.1768	6.4279 \pm 1.5245	5.4194 \pm 0.2680	7.2767 \pm 1.3188
10	6.2957 \pm 0.1768	6.4283 \pm 1.5228	5.4186 \pm 0.2680	7.2884 \pm 1.3174
12	6.2957 \pm 0.1768	6.4286 \pm 1.5215	5.4180 \pm 0.2680	7.2970 \pm 1.3163
15	6.2957 \pm 0.1768	6.4290 \pm 1.5200	5.4174 \pm 0.2681	7.3067 \pm 1.3150

Πίνακας 2.2 Αποτελέσματα (τιμές των RMSEs) του προτεινόμενου ΤΝΔ για $r = 3$ και $r = 4$ στο σύνολο δεδομένων **AutoMPG** του UCI Repository

c	Προτεινόμενο, $r = 3$		Προτεινόμενο, $r = 4$	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα ελέγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα ελέγχου
4	5.3518 \pm 0.2491	7.2243 \pm 1.3114	5.2401 \pm 0.2294	7.3170 \pm 1.3482
6	5.3262 \pm 0.2395	7.2527 \pm 1.3039	5.2392 \pm 0.2302	7.3338 \pm 1.3447
8	5.3089 \pm 0.2347	7.2608 \pm 1.3041	5.2392 \pm 0.2316	7.3391 \pm 1.3446
10	5.2974 \pm 0.2333	7.2651 \pm 1.3055	5.2390 \pm 0.2324	7.3422 \pm 1.3447
12	5.2899 \pm 0.2316	7.2678 \pm 1.3067	5.2381 \pm 0.2338	7.3434 \pm 1.3369
15	5.2814 \pm 0.2321	7.2708 \pm 1.3069	5.2376 \pm 0.2344	7.3473 \pm 1.3323

Πίνακας 2.3 Αποτελέσματα (τιμές των RMSEs) των δικτύων FFNN και pRBFN στο σύνολο δεδομένων **AutoMPG** του UCI Repository

<i>c</i>	FFNN		pRBFN	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	8.1626 ± 1.7007	8.2639 ± 2.6102	5.3582 ± 0.5054	5.7440 ± 1.5589
6	8.1684 ± 2.1697	8.2219 ± 2.6494	5.9892 ± 0.4395	6.8501 ± 1.8389
8	8.0340 ± 2.2476	8.2083 ± 2.9286	6.1917 ± 0.4515	6.9483 ± 1.9909
10	7.7746 ± 0.2206	7.9368 ± 1.8641	6.2738 ± 0.4143	6.9314 ± 1.8849
12	7.7724 ± 0.2183	7.9369 ± 1.8625	6.5413 ± 0.4566	7.4591 ± 2.0949
15	7.7724 ± 0.2183	7.9405 ± 1.8620	5.3271 ± 0.5231	5.6834 ± 1.5313

Πίνακας 2.4 Αποτελέσματα (τιμές των RMSEs) του προτεινόμενου ΤΝΔ για $r = 1$ και $r = 2$ στο σύνολο δεδομένων **ForestFires** του UCI Repository

<i>c</i>	Προτεινόμενο, $r = 1$		Προτεινόμενο, $r = 2$	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	52.9057 ± 6.9665	41.7628 ± 49.4679	52.6929 ± 6.9338	62.9830 ± 8.2934
6	52.9057 ± 6.9665	41.7684 ± 49.4630	52.6925 ± 6.9339	62.9830 ± 8.2934
8	52.9057 ± 6.9665	41.7718 ± 49.4602	52.6924 ± 6.9339	62.9830 ± 8.2934
10	52.6618 ± 6.9320	42.3139 ± 49.1990	52.6923 ± 6.9339	62.6926 ± 8.2524
12	52.9057 ± 6.9665	41.7758 ± 49.4567	52.6922 ± 6.9339	62.9830 ± 8.2934
15	52.9057 ± 6.9665	41.7776 ± 49.4551	52.6921 ± 6.9339	62.9830 ± 8.2934

Πίνακας 2.5 Αποτελέσματα (τιμές των RMSEs) του προτεινόμενου ΤΝΔ για $r = 3$ και $r = 4$ στο σύνολο δεδομένων **ForestFires** του UCI Repository

<i>c</i>	Προτεινόμενο, $r = 3$		Προτεινόμενο, $r = 4$	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	52.6729 ± 6.9340	42.2306 ± 49.2115	52.6651 ± 6.9315	42.3238 ± 49.1797
6	52.6686 ± 6.9347	42.2217 ± 49.2287	52.6629 ± 6.9318	42.3167 ± 49.1924
8	52.6666 ± 6.9350	42.2175 ± 49.2380	52.6618 ± 6.9320	42.3139 ± 49.1990
10	52.6653 ± 6.9352	42.2153 ± 49.2442	52.6612 ± 6.9321	42.3117 ± 49.2034
12	52.6645 ± 6.9353	42.2140 ± 49.2487	52.6607 ± 6.9321	42.3116 ± 49.2064
15	52.6636 ± 6.9355	42.2128 ± 49.2535	52.6603 ± 6.9322	42.3101 ± 49.2104

Πίνακας 2.6 Αποτελέσματα (τιμές των RMSEs) των δικτύων FFNN και pRBFN στο σύνολο δεδομένων **ForestFires** του UCI Repository

<i>c</i>	FFNN		pRBFN	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	62.9830 ± 8.2935	41.7930 ± 49.4416	59.4208 ± 7.8036	47.2391 ± 47.4135
6	62.9830 ± 8.2934	41.7932 ± 49.4416	57.5422 ± 7.4488	50.0804 ± 45.9863
8	62.9829 ± 8.2934	41.7933 ± 49.4417	55.4271 ± 7.5860	52.0322 ± 45.3252
10	62.9824 ± 8.2934	41.7968 ± 49.4392	54.7325 ± 7.3755	52.4954 ± 45.0881
12	62.9825 ± 8.2932	41.7927 ± 49.4409	51.8516 ± 7.4042	53.7711 ± 45.5390
15	62.9770 ± 8.2950	41.7964 ± 49.4203	53.1630 ± 7.0121	53.6392 ± 44.9500

Πίνακας 2.7 Αποτελέσματα (τιμές των RMSEs) του προτεινόμενου ΤΝΔ για $r = 1$ και $r = 2$ στο σύνολο δεδομένων **Housing** του UCI Repository

c	Προτεινόμενο, $r = 1$		Προτεινόμενο, $r = 2$	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	8.7792 ± 0.3348	8.4989 ± 3.1391	8.2245 ± 0.2755	8.8919 ± 2.5868
6	8.7791 ± 0.3348	8.5002 ± 3.1349	8.2217 ± 0.2758	8.9038 ± 2.5765
8	8.7790 ± 0.3348	8.5011 ± 3.1324	8.2203 ± 0.2760	8.9110 ± 2.5704
10	8.7790 ± 0.3348	8.5016 ± 3.1307	8.2194 ± 0.2761	8.9159 ± 2.5662
12	8.7790 ± 0.3348	8.5021 ± 3.1294	8.2189 ± 0.2761	8.9195 ± 2.5632
15	8.7789 ± 0.3348	8.5025 ± 3.1280	8.2183 ± 0.2762	8.9235 ± 2.5598

Πίνακας 2.8 Αποτελέσματα (τιμές των RMSEs) του προτεινόμενου ΤΝΔ για $r = 3$ και $r = 4$ στο σύνολο δεδομένων **Housing** του UCI Repository

c	Προτεινόμενο, $r = 3$		Προτεινόμενο, $r = 4$	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	8.1884 ± 0.2752	9.0073 ± 2.4764	8.1625 ± 0.2689	9.0086 ± 2.4936
6	8.2217 ± 0.2758	8.9038 ± 2.5765	8.1406 ± 0.2696	9.0613 ± 2.4719
8	8.1440 ± 0.2712	9.0941 ± 2.4440	8.1293 ± 0.2698	9.0946 ± 2.4590
10	8.1348 ± 0.2704	9.1171 ± 2.4361	8.1228 ± 0.2700	9.1188 ± 2.4495
12	8.1287 ± 0.2697	9.1346 ± 2.4303	8.1182 ± 0.2703	9.1367 ± 2.4421
15	8.1228 ± 0.2689	9.1547 ± 2.4241	8.1137 ± 0.2702	9.1569 ± 2.4358

Πίνακας 2.9 Αποτελέσματα (τιμές των RMSEs) των δικτύων FFNN και pRBFN στο σύνολο δεδομένων **Housing** του UCI Repository

c	FFNN		pRBFN	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	9.2806 ± 0.4822	8.9948 ± 3.5977	8.5928 ± 0.6375	9.2084 ± 3.2020
6	9.2698 ± 0.5289	8.9326 ± 3.5224	8.8752 ± 0.6641	9.7386 ± 3.1008
8	8.9886 ± 0.6092	8.8275 ± 3.3196	8.9915 ± 0.6606	10.1241 ± 3.0831
10	8.7668 ± 0.6165	8.4775 ± 3.2590	9.0611 ± 0.6457	10.6916 ± 3.5559
12	8.4847 ± 0.5643	8.3558 ± 2.9744	8.9824 ± 0.6378	10.5131 ± 3.4897
15	8.1956 ± 0.5352	8.2093 ± 2.8976	8.8331 ± 0.5831	11.2335 ± 3.4369

Πίνακας 2.10 Αποτελέσματα (τιμές των RMSEs) του προτεινόμενου ΤΝΔ για $r = 1$ και $r = 2$ στο σύνολο δεδομένων **Servo** του UCI Repository

c	Προτεινόμενο, $r = 1$		Προτεινόμενο, $r = 2$	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	1.5620 ± 0.0361	1.4999 ± 0.3880	1.3320 ± 0.0298	1.7464 ± 0.4386
6	1.5595 ± 0.0366	1.4973 ± 0.3736	1.3214 ± 0.0292	1.8124 ± 0.4389
8	1.5587 ± 0.0366	1.4983 ± 0.3759	1.3156 ± 0.0298	1.8267 ± 0.4496
10	1.5567 ± 0.0360	1.4970 ± 0.3753	1.3132 ± 0.0292	1.8554 ± 0.4437
12	1.5576 ± 0.0377	1.4991 ± 0.3664	1.3115 ± 0.0290	1.8760 ± 0.4442
15	1.5577 ± 0.0360	1.4989 ± 0.3654	1.3090 ± 0.0294	1.8698 ± 0.4545

Πίνακας 2.11 Αποτελέσματα (τιμές των RMSEs) του προτεινόμενου ΤΝΔ για $r = 3$ και $r = 4$ στο σύνολο δεδομένων **Servo** του UCI Repository

c	Προτεινόμενο, $r = 3$		Προτεινόμενο, $r = 4$	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	1.2505 ± 0.0432	1.8614 ± 0.4053	1.2640 ± 0.0491	1.8533 ± 0.3947
6	1.2364 ± 0.0436	1.8942 ± 0.3793	1.2484 ± 0.0494	1.8942 ± 0.3851
8	1.2326 ± 0.0424	1.9510 ± 0.3811	1.2384 ± 0.0456	1.9151 ± 0.3594
10	1.2275 ± 0.0418	1.9360 ± 0.3682	1.2333 ± 0.0478	1.9264 ± 0.3726
12	1.2254 ± 0.0428	1.9602 ± 0.3705	1.2269 ± 0.0466	1.9590 ± 0.3625
15	1.2229 ± 0.0425	1.9806 ± 0.3693	1.2246 ± 0.0461	1.9611 ± 0.3645

Πίνακας 2.12 Αποτελέσματα (τιμές των RMSEs) των δικτύων FFNN και pRBFN στο σύνολο δεδομένων **Servo** του UCI Repository

c	FFNN		pRBFN	
	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου	Σφάλμα στα δεδομένα εκπαίδευσης	Σφάλμα στα δεδομένα έλεγχου
4	1.6642 ± 0.1298	1.6619 ± 0.4236	1.6775 ± 0.1268	1.6054 ± 0.4480
6	1.6627 ± 0.1340	1.6870 ± 0.4053	1.7931 ± 0.1568	1.7247 ± 0.5363
8	1.7451 ± 0.1871	1.7688 ± 0.4290	1.8413 ± 0.1530	1.7951 ± 0.5114
10	1.7919 ± 0.1988	1.8418 ± 0.4683	1.8831 ± 0.1522	1.8172 ± 0.4880
12	1.8150 ± 0.2456	1.8718 ± 0.5156	1.9315 ± 0.1531	1.8689 ± 0.5283
15	1.8975 ± 0.2824	1.9092 ± 0.4836	1.9808 ± 0.1448	1.8080 ± 0.5155

Όπως παρουσιάζεται στους παραπάνω πίνακες, στο σύνολο δεδομένων AutoMPG καλύτερα αποτελέσματα στα δεδομένα εκπαίδευσης δίνει το προτεινόμενο ΤΝΔ για $c = 15$ και $r = 4$ και ακολουθεί το προτεινόμενο για $r = 3$, το pRBFN δίνει αποτελέσματα λίγο καλύτερα από το προτεινόμενο για $r = 2$ ενώ στα δεδομένα ελέγχου το pRBFN για $c = 15$ δίνει τα καλύτερα αποτελέσματα. Στο σύνολο ForestFires καλύτερα αποτελέσματα δίνει το pRBFN στα δεδομένα εκπαίδευσης για $c = 12$ και ακολουθεί το προτεινόμενο για $r = 4$ και $c = 15$ και για $r = 1$ και $c = 4$. Στα δεδομένα έλεγχου καλύτερα αποτελέσματα δίνει το προτεινόμενο δίκτυο για $r = 1$, ακολουθεί το FFNN ενώ το pRBFN και το προτεινόμενο για $r = 2$ δίνουν τα χειρότερα αποτελέσματα. Στο σύνολο Housing στην εκπαίδευση παρατηρείται να δίνει καλύτερα αποτελέσματα τα προτεινόμενα για $r = 4, c = 15$, το FFNN για $c = 15$ δίνει αποτελέσματα ενδιάμεσα σε αυτά του προτεινόμενου για $r = 2$ και $r = 3$ ενώ το pRBFN δίνει τα χειρότερα αποτελέσματα από όλα τα δίκτυα. Στα δεδομένα έλεγχου του ίδιου συνόλου τα καλύτερα αποτελέσματα τα δίνει το FFNN και ακολουθεί το προτεινόμενο για $r = 1$. Τέλος στο σύνολο Servo τα καλύτερα αποτελέσματα εκπαίδευσης τα δίνει το προτεινόμενο με $r = 3, c = 15$ ενώ στα δεδομένα έλεγχου το προτεινόμενο με $r = 1, c = 10$, τα FFNN και pRBFN δεν δίνουν τόσο καλά αποτελέσματα στο σύνολο δεδομένων Servo.

2.6 Εξαγωγή Ακτογραμμής από Ψηφιακές Εικόνες

Το πρόβλημα της παρούσας ενότητας είναι η αναγνώριση της θέσης της ακτογραμμής μιας παραλίας από εικόνες τύπου SIGMA. Ο ορισμός και ο τρόπος δημιουργίας των εικόνων αυτών βρίσκεται αναλυτικά στο παράρτημα Β.1. Επίσης στο παράρτημα Β.1 βρίσκεται η διαδικασία συλλογής των δεδομένων που χρησιμοποιούνται στην παρούσα παράγραφο και θα ονομάζονται δεδομένα Shoreline Extraction. Οι εικόνες αυτές παρουσιάζουν την τυπική απόκλιση ενός συνόλου διαδοχικών εικόνων (εικόνες τύπου snapshot) ή διαδοχικών frames ενός βίντεο. Επειδή η τυπική απόκλιση είναι ένα μέτρο διασποράς των εικόνων, η εικόνα SIGMA είναι ευνόητο ότι θα αποτελείται από μαύρα εικονοστοιχεία (pixels) σε μια περιοχή χωρίς κίνηση, ενώ εικονοστοιχεία με τιμές κοντά στο άσπρο θα δείχνουν μεγάλη κίνηση (Σχήμα 2.4). Η ιδέα της αναγνώρισης της ακτογραμμής είναι να βρεθεί μια τιμή-κατώφλι (threshold) που να καθορίζει το σημείο της εικόνας που υπάρχει η επιθυμητή κίνηση. Η τιμή αυτή τότε θα καθορίζει τη θέση της ακτογραμμής στην κάθε εικόνα SIGMA.

Προκειμένου να προκύψουν από τις εικόνες SIGMA αριθμητικές τιμές για να χρησιμοποιηθούν ως δεδομένα εισόδου-εξόδου στο ΤΝΔ ακολουθήθηκε η διαδικασία που περιγράφεται στα εξής βήματα:

(α) Από τις εικόνες SIGMA εξάχθηκε το ιστόγραμμα τους χωρίζοντας τις τιμές των εικονοστοιχείων σε $q = 200$ κλάσεις ίσου πλάτους. Πρέπει να σημειωθεί ότι οι εικόνες δεν είχαν 256 διακριτές τιμές (όπως γίνεται με τις συνήθεις εικόνες τύπου grayscale) μιας και είναι τύπου intensity-images (βλ. Παράρτημα Β.1 και [157]).

(β) Στις τιμές του ιστογράμματος έγινε κανονικοποίηση (min-max data scaling) ώστε να έχει τιμές στο διάστημα $[0,1]$. Για την κανονικοποίηση αυτή αν υποθέσουμε πως οι τιμές της συχνότητας του ιστογράμματος ανήκαν στο διάστημα $[\min(I_k^{old}), \max(I_k^{old})]$ τότε οι τιμές μετατρέπονται (βλ. και εξίσωση (2.16)) στις:

$$I_k = \frac{I_k^{old} - \min(I_k^{old})}{\max(I_k^{old}) - \min(I_k^{old})} \quad (2.60)$$

οι οποίες παίρνουν τιμές στο διάστημα $[0,1]$ με την τιμή 0 να δείχνει εικονοστοιχείο μαύρου χρώματος και την 1 να δείχνει άσπρο χρώμα.

(γ) Η συχνότητα της μικρότερης κλάσης του ιστογράμματος αλλάχθηκε σε μηδέν. Με τον τρόπο αυτό αποφεύγονται οι μεγάλες τιμές που θα είχαν οι συχνότητες του μαύρου χρώματος μέσα στις εικόνες (Σχήμα 2.4) και θα οδηγούσε σε μεγάλα προσεγγιστικά σφάλματα την εξίσωση (2.65). Οι τιμές που έχουν ως τώρα τα ιστογράμματα των εικόνων περιγράφονται από τα ζεύγη $(I_k, fr_k), k=1,2,\dots,q$ όπου I_k η τιμή του κάθε εικονοστοιχείου και fr_k η συχνότητα της τιμής αυτής.

(δ) Τα ιστογράμματα μετατράπηκαν ώστε να έχουν μικρότερες τιμές I_k , έτσι ώστε το νέο ιστόγραμμα που θα προκύψει να έχει εμβαδόν ίσο με 1. Για να γίνει αυτό αρκεί να σκεφτούμε ότι το ιστόγραμμα σχεδιάζεται από $q=200$ διαδοχικά ορθογώνια παραλληλόγραμμα τα οποία έχουν για ύψος την τιμή fr_k και ως βάση το διάστημα $[\hat{I}_k, \hat{I}_{k+1}]$ με το I_k να είναι το μέσον του διαστήματος $[\hat{I}_k, \hat{I}_{k+1}]$. Τότε για το συνολικό εμβαδό των ορθογωνίων αυτών ισχύει ότι:

$$\text{Εμβαδόν} = \sum_{k=1}^q (\hat{I}_{k+1} - \hat{I}_k) \cdot fr_k \quad (2.61)$$

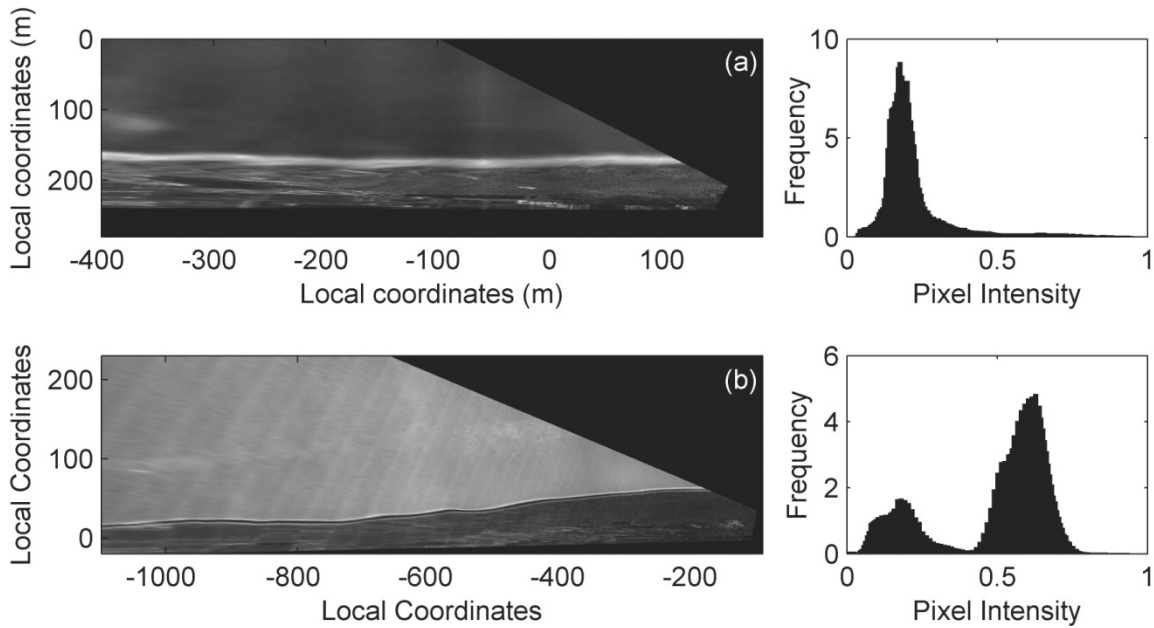
Επειδή τα διαστήματα είναι 200 στο πλήθος και ίσου πλάτους, άρα καθένα έχει πλάτος $\frac{1}{q} = \frac{1}{200}$ και τότε το Εμβαδόν γίνεται:

$$\text{Εμβαδόν} = \sum_{k=1}^q \frac{1}{q} \cdot fr_k = \sum_{k=1}^q \frac{1}{200} \cdot fr_k \quad (2.62)$$

τότε το νέο ιστόγραμμα εμβαδού ίσου με ένα προκύπτει από τα ζεύγη (I_k, fr_k^{New}) , $k=1,2,\dots,q$ με:

$$fr_k^{New} = \frac{fr_k}{\text{Εμβαδόν}} \quad (2.63)$$

Τα νέα ιστογράμματα που προκύπτουν τότε είναι όπως αυτά που παρουσιάζονται στο Σχήμα 2.6.



Σχήμα 2.4 Γεωαναφερμένες εικόνες από την περιοχή (α) της παραλίας Faro της Πορτογαλίας και (β) της Αμμουδάρας της Κρήτης. Δίπλα στην κάθε εικόνα βρίσκεται το ιστόγραμμα της εμβადού ίσου με 1 που προέκυψε έπειτα από επεξεργασία. Οι συντεταγμένες των εικόνων είναι σε m σε τοπικό σύστημα αναφοράς (βλ. Παράρτημα Β.1 για τη μετατροπή αυτή)

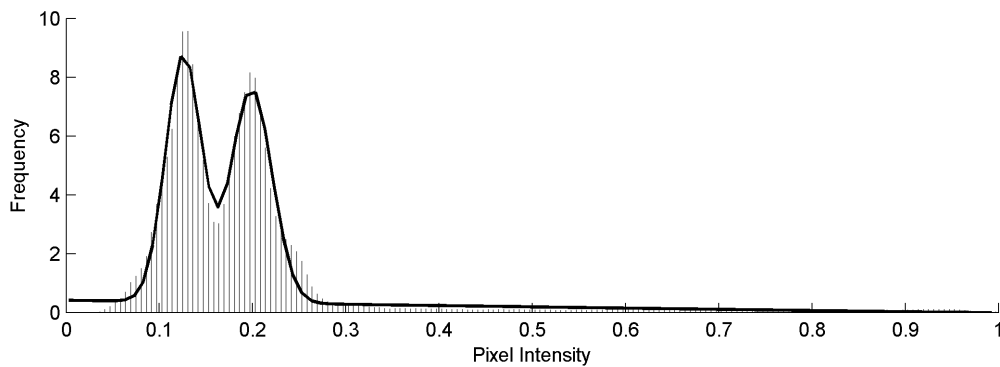
Αφού έγινε η επεξεργασία που αναφέρθηκε στα ιστογράμματα των εικόνων, κάθε ένα ιστόγραμμα από αυτά 'συμπιέστηκε' σε 9 παραμέτρους. Αφού παρατηρήθηκε ότι το κάθε ιστόγραμμα από τις διαθέσιμες εικόνες συνήθως αποτελούταν από 2 κορυφές οι οποίες σχημάτιζαν 2 σχήματα σαν 'καμπάνες' (Σχήμα 2.4 και Σχήμα 2.5), το κάθε ιστόγραμμα εικόνας συμπιέστηκε ώστε να μπορέσει μια εξίσωση που αποτελείται από δύο γκαουσιανές συναρτήσεις και ένα πολυώνυμο 2^{ου} βαθμού, δηλαδή, η :

$$F(I) = x_1 + x_2 I + x_3 I^2 + x_4 \exp\left(-\left(\frac{I - x_5}{x_6}\right)^2\right) + x_7 \exp\left(-\left(\frac{I - x_8}{x_9}\right)^2\right) \quad (2.64)$$

να περιγράψει όσο το δυνατόν καλύτερα τα σημεία (I_k, f_k^{New}) μέσω των τιμών $(I, F(I))$ για $I = I_k, k = 1, 2, \dots, q$. Προκειμένου να γίνει η εύρεση των παραμέτρων $[x_1, x_2, \dots, x_9]$ ώστε τα σημεία (I_k, f_k^{New}) να περιγράφουν όσο το δυνατόν καλύτερα τα σημεία $(I, F(I))$ έγινε ελαχιστοποίηση της συνάρτησης σφάλματος :

$$J_F = \sum_{k=1}^q (f_k^{New} - F(I_k))^2 \quad (2.65)$$

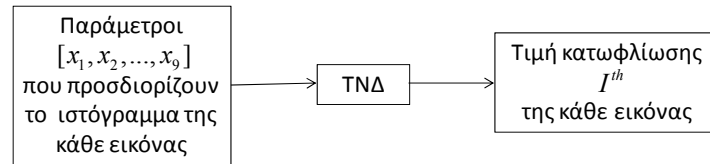
Η ελαχιστοποίηση αυτή έγινε χρησιμοποιώντας τη μέθοδο Gradient Descent. Παράδειγμα της προσέγγισης αυτής φαίνεται στο Σχήμα 2.7.



Σχήμα 2.5 Το ιστόγραμμα μιας εικόνας και η προσέγγισή του μέσω της συνάρτησης (2.64)

Τα δεδομένα $[x_1, x_2, \dots, x_9]$ που εξήχθησαν ως περιγραφή των ιστογραμμάτων χρησιμοποιήθηκαν ως δεδομένα εισόδου στο ΤΝΔ (Σχήμα 2.6). Ως δεδομένα εξόδου χρησιμοποιήθηκαν οι τιμές-κατώφλι I^{th} των εικονοστοιχείων οι οποίες προσδιορίζουν την ακτογραμμή την εικόνας. Για να υπολογιστούν οι τιμές I^{th} για την κάθε εικόνα χρησιμοποιήθηκε ο αλγόριθμος κατωφλίωσης που δημιουργήθηκε από τους Vousdoukas et al. [158]. Ο αλγόριθμος αυτός επιτρέπει σε ένα χρήστη να διαλέξει διαφορετικές τιμές για την τιμή-κατώφλι I^{th} για κάθε εικόνα μέχρι να βρει αυτή που του ταιριάζει, δηλαδή κάποιος τρέχοντας με αυτόν τον αλγόριθμο όλες τις διαθέσιμες εικόνες δοκιμάζει για κάθε εικόνα διαφορετικές τιμές κατωφλίου I^{th} μέχρι να βρει αυτή που προσδιορίζει καλύτερα αυτό που βλέπει στην οθόνη του ως ακτογραμμή. Ο αλγόριθμος στα συγκεκριμένα δεδομένα έψαχνε την κάθε εικόνα από κάτω προς τα πάνω μέχρι να βρει μέσα στην κάθε εικόνα την τιμή I^{th} που έχει επιλέξει ο χρήστης: αν στον χρήστη δεν άρεσε η ακτογραμμή που εξάγει και σχεδιάζει ο αλγόριθμος, τότε δοκιμάζει άλλες τιμές για το I^{th} μέχρι αυτό να περιγράψει καλύτερα την ακτογραμμή μίας εικόνας.

Με την διαδικασία που περιγράφηκε παραπάνω έχουμε καταλήξει σε N το πλήθος 9-διάστατα δεδομένα εισόδου ($p = 9$) και μονοδιάστατο δεδομένο εξόδου (την τιμή I^{th}) για την κάθε διαθέσιμη εικόνα. Η δομή δεδομένων εισόδου-εξόδου περιγράφεται στο Σχήμα 2.6 και



Σχήμα 2.6 Τα δεδομένα εισόδου-εξόδου τα οποία χρησιμοποιεί το ΤΝΔ του άρθρου [1]

τα δεδομένα εισόδου-εξόδου του ΤΝΔ πλέον μπορούν να γραφούν ως :

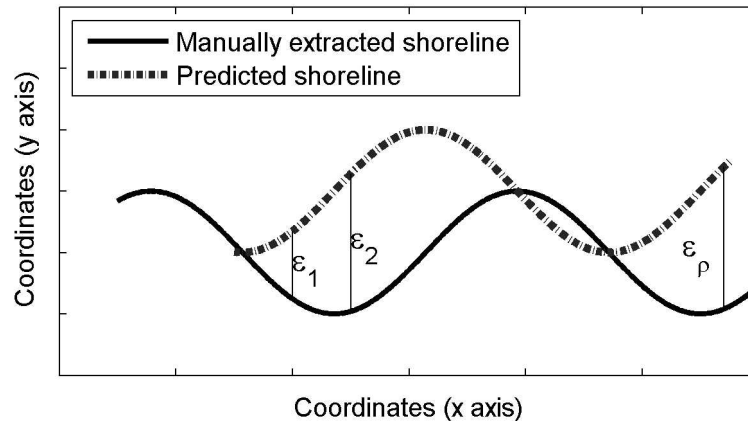
$$S = \left\{ (\mathbf{x}_k, y_k = I^{th}) / \mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{k9}] \in \mathbb{R}^9, y_k = I^{th} \in \mathbb{R}, k = 1, 2, \dots, N \right\} \quad (2.66)$$

Πρέπει να σημειωθεί ότι τα δεδομένα από την παραλία Faro είναι μεγέθους $N = 728$ και $N = 843$ για την παραλία της Αμμουδάρας. Αναλυτική περιγραφή για τις περιοχές που συλλέχθηκαν τα δεδομένα υπάρχει στο παράρτημα Β.1. Το σύνολο δεδομένων που δημιουργήθηκε θα ονομάζεται σύνολο Shoreline Extraction και υπάρχει διαθέσιμο στο διαδίκτυο στο σύνδεσμο: <http://isla.aegean.gr/news.php?lg=2>.

Στα δεδομένα αυτά εφαρμόστηκαν δύο πειράματα, στο Πείραμα 1 (Πίνακας 2.13) εφαρμόστηκε η μέθοδος 10-fold cross validation προκειμένου να εκπαιδευθεί και να ελεγχθεί το προτεινόμενο ΤΝΔ και τα δύο με τα οποία γίνεται η σύγκριση (βλ. παράγραφο 2.5) στα δεδομένα που έχουν συλλεχθεί από την παραλία της Αμμουδάρας. Τα αποτελέσματα του πειράματος αυτού έχουν μετρηθεί με χρησιμοποιώντας το μέσο τετραγωνικό σφάλμα (MSE , βλ. εξίσωση (1.32)) του κάθε ΤΝΔ στα δεδομένα εκπαίδευσης και στα δεδομένα ελέγχου. Στο Πείραμα 2 χρησιμοποιήθηκαν τα δεδομένα που συλλέχθηκαν από την περιοχή της Αμμουδάρας ως σύνολο εκπαίδευσης (training) του ΤΝΔ και τα δεδομένα της παραλίας Faro ως δεδομένα ελέγχου (testing). Για το πείραμα αυτό υπολογίστηκε το μέσο τετραγωνικό σφάλμα (MSE) όπως και στο προηγούμενο πείραμα. Το πιο σημαντικό ως προς τη φυσική σημασία του προβλήματος ήταν όμως να υπολογιστεί το σφάλμα RMSE σε μέτρα για την ακτογραμμή που έχει εξαχθεί από το ΤΝΔ και τα ποσοστά αποδεκτών αποτελεσμάτων (data return rates). Τα σφάλματα RMSE της ακτογραμμής υπολογίστηκαν από την εξίσωση:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{\rho} \varepsilon_i^2}{\rho}} \quad (2.67)$$

όπου το σφάλμα ε_i είναι το σφάλμα μεταξύ των δύο ακτογραμμών, αυτής που έχει υπολογιστεί από ένα χρήστη και αυτής που εξάγει το ΤΝΔ σε ρ το πλήθος σημεία του οριζόντιου άξονα της εικόνας (Σχήμα 2.7).



Σχήμα 2.7 Η μέθοδος υπολογισμού του σφάλματος RMSE σε m για δύο διαφορετικές ακτογραμμές.

Για να υπολογιστούν τα ποσοστά αποδεκτών αποτελεσμάτων (data return rates) θεωρήθηκε ως αποδεκτή μια ακτογραμμή αν αυτή ανήκε στην περιοχή ενδιαφέροντος (ROI, βλ. Παράρτημα Β.1) σε ποσοστό μεγαλύτερο του 50%. Στα δύο πειράματα έγινε επίσης σύγκριση με τα αποτελέσματα του ΤΝΔ που δημοσιεύτηκαν από τους Rigos et al. [159] και με τα αποτελέσματα που δίνει η μέθοδος SLIM (Shoreline Intensity Maximum) των Plant και Holman [160]. Σύμφωνα με τη μέθοδο SLIM η ακτογραμμή εξάγεται από εικόνες τύπου TIMEX (βλ. Παράρτημα Β.1) αν από κάθε κατακόρυφη στήλη από εικονοστοιχεία της εικόνας, επιλεγθεί αυτό με τη μέγιστη τιμή, διότι αυτή δείχνει δημιουργία αφρού (άσπρο χρώμα) λόγω μεγάλης κίνησης των κυμάτων. Για το πείραμα 2 υπήρχαν επίσης διαθέσιμα τα αποτελέσματα των Vousdoukas et al. [158] τα οποία παρουσιάζονται στον πίνακα 2.20 για σκοπούς σύγκρισης. Ο πίνακας 2.13 δείχνει εν συντομία τα χαρακτηριστικά των δύο πειραμάτων.

Πίνακας 2.13 Τα χαρακτηριστικά των δύο πειραμάτων.

	Πείραμα 1	Πείραμα 2
Σύνολο/α δεδομένων	Αμμουδάρα	Αμμουδάρα και Faro
Συναρτήσεις μέτρησης απόδοσης	MSE (εκπαίδευσης και ελέγχου) για τα ΤΝΔ	MSE (στα δεδομένα ελέγχου) για τα ΤΝΔ, RMSE (σε m) για τις ακτογραμμές που εξάγονται, και ποσοστά αποδεκτών αποτελεσμάτων
Τρόπος υπολογισμού της απόδοσης	10-fold cross-validation	Τα δεδομένα της Αμμουδάρας ως δεδομένα εκπαίδευσης και του Faro ως δεδομένα ελέγχου. 10 διαφορετικές αρχικοποιήσεις σε κάθε περίπτωση.

Οι πίνακες 2.13-2.16 παρουσιάζουν τα αποτελέσματα του πειράματος 1, χρησιμοποιώντας το προτεινόμενο ΤΝΔ και τα δύο άλλα ΤΝΔ με τα οποία γίνεται σύγκριση (βλ. παράγραφο 2.5) όπως αυτά έχουν δημοσιευθεί και στο [1]. Στο Σχήμα 2.8 παρουσιάζονται οι μέσες τιμές των αποτελεσμάτων αυτών. Από τους πίνακες παρατηρείται ότι το προτεινόμενο ΤΝΔ για $r = 1$ δίνει τα χειρότερα αποτελέσματα ενώ για $r = 2$ δίνει καλύτερα αποτελέσματα από το pRBFN. Το προτεινόμενο ΤΝΔ, για μεγαλύτερους βαθμούς πολυωνύμων ($r = 3, r = 4$) δίνει καλύτερα αποτελέσματα σε σχέση με μικρότερους βαθμούς πολυωνύμων αλλά και σε σχέση με τα pRBFN και FFNN (Σχήμα 2.8). Τα καλύτερα αποτελέσματα προέκυψαν και στη διαδικασία εκπαίδευσης και στη διαδικασία ελέγχου, χρησιμοποιώντας το προτεινόμενο ΤΝΔ με πολυώνυμο 3^{ου} βαθμού ($r = 3, c = 12$). Αυτό όμως δε σημαίνει ότι αναγκαστικά τα 3^{ου} βαθμού πολυώνυμα θα δίνουν καλύτερα αποτελέσματα από αυτά του 4^{ου} βαθμού σε διαφορετικές προσομοιώσεις, βλ. για παράδειγμα τα δεδομένα ελέγχου (Σχήμα 2.8(β)). Γενικά το προτεινόμενο ΤΝΔ έχει την τάση να αποδίδει μικρότερα σφάλματα όσο μεγαλύτερος είναι ο βαθμός r των πολυωνύμων που χρησιμοποιούνται, όταν δε συμβαίνει αυτό, τότε αυτό οφείλεται σε καταστάσεις overfitting, ο οποίος είναι και ο λόγος που σε όλες τις προσομοιώσεις διατηρήθηκε ο βαθμός πολυωνύμων μέχρι και $r = 4$.

Πίνακας 2.14 Αποτελέσματα (τιμές MSEs) των ΤΝΔ του Πειράματος 1.

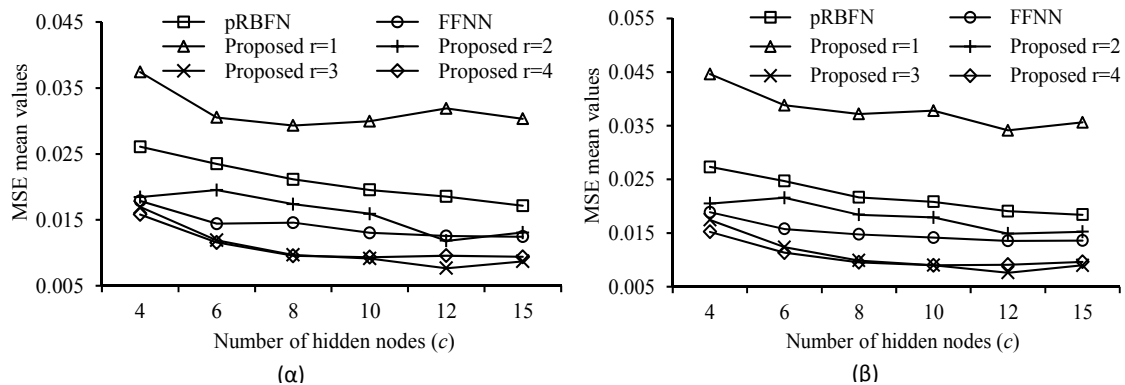
c	pRBFN		FFNN	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
4	0.02609 ± 0.00266	0.02732 ± 0.00566	0.01786 ± 0.00405	0.01889 ± 0.00583
6	0.02348 ± 0.00215	0.02471 ± 0.00497	0.01441 ± 0.00385	0.01578 ± 0.00562
8	0.02113 ± 0.00194	0.02167 ± 0.00503	0.01456 ± 0.00326	0.01477 ± 0.00453
10	0.01953 ± 0.00205	0.02084 ± 0.00507	0.01302 ± 0.00265	0.01415 ± 0.00568
12	0.01855 ± 0.00146	0.01908 ± 0.00409	0.01255 ± 0.00432	0.01355 ± 0.00595
15	0.01714 ± 0.00213	0.01843 ± 0.00455	0.01244 ± 0.00326	0.01360 ± 0.00400

Πίνακας 2.15 Αποτελέσματα (τιμές MSEs) των ΤΝΔ του Πειράματος 1.

c	Προτεινόμενο, $r = 1$		Προτεινόμενο, $r = 2$	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
4	0.03742 ± 0.00877	0.04462 ± 0.01434	0.01845 ± 0.01414	0.02051 ± 0.01336
6	0.03054 ± 0.01145	0.03881 ± 0.01248	0.01952 ± 0.01381	0.02159 ± 0.01393
8	0.02931 ± 0.01504	0.03719 ± 0.01727	0.01737 ± 0.00814	0.01842 ± 0.00814
10	0.02996 ± 0.01387	0.03780 ± 0.01792	0.01592 ± 0.00582	0.01793 ± 0.00674
12	0.03189 ± 0.01739	0.03413 ± 0.01941	0.01182 ± 0.00214	0.01490 ± 0.00333
15	0.03033 ± 0.01860	0.03563 ± 0.02006	0.01306 ± 0.00118	0.01525 ± 0.00415

Πίνακας 2.16 Αποτελέσματα (τιμές MSEs) των ΤΝΔ του Πειράματος 1.

c	Προτεινόμενο, $r = 3$		Προτεινόμενο, $r = 4$	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
4	0.01694 ± 0.00587	0.01747 ± 0.00712	0.01581 ± 0.00338	0.01525 ± 0.00468
6	0.01189 ± 0.00148	0.01238 ± 0.00335	0.01155 ± 0.00182	0.01136 ± 0.00319
8	0.00967 ± 0.00105	0.00988 ± 0.00297	0.00953 ± 0.00160	0.00955 ± 0.00280
10	0.00913 ± 0.00111	0.00903 ± 0.00272	0.00933 ± 0.00253	0.00900 ± 0.00322
12	0.00766 ± 0.00061	0.00763 ± 0.00246	0.00956 ± 0.00310	0.00913 ± 0.00328
15	0.00869 ± 0.00072	0.00901 ± 0.00320	0.00938 ± 0.00267	0.00966 ± 0.00315

**Σχήμα 2.8** Οι μέσες τιμές των αριθμητικών αποτελεσμάτων που παρουσιάζονται στους πίνακες 2.14-2.16 (πείραμα 1)

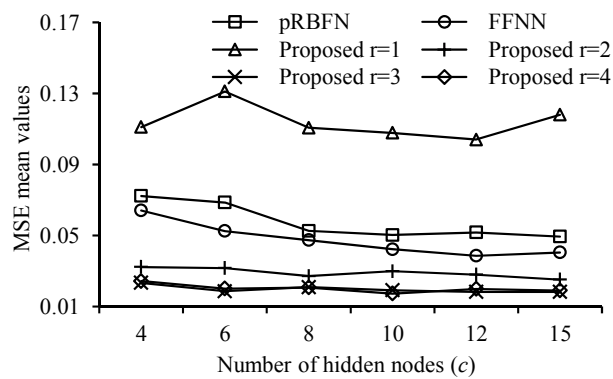
Στους πίνακες 2.17-2.18 δίνονται τα αποτελέσματα (σφάλματα MSE) των ΤΝΔ όταν αυτά εφαρμόστηκαν στα δεδομένων ελέγχου (δεδομένα από την περιοχή Faro). Τα αποτελέσματα δείχνουν ότι το προτεινόμενο ΤΝΔ για μεγαλύτερους βαθμούς πολυωνύμων ($r = 3, r = 4$) δίνει καλύτερα αποτελέσματα σε σχέση με μικρότερους βαθμούς ($r = 1, r = 2$) και σε σχέση με τα pRBFN και FFNN. Στο Σχήμα 2.9 παρουσιάζονται οι μέσες τιμές των αποτελεσμάτων αυτών. Στα Σχήματα 2.10-2.11 παρουσιάζονται οι εξαχθείσες ακτογραμμές από τις περιοχές Αμμουδάρα και Faro αντίστοιχα. Στα δεδομένα της Αμμουδάρας (Σχήμα 2.10) το προτεινόμενο ΤΝΔ με χρήση πολυωνύμων πρώτου βαθμού ($r=1$) δίνει χειρότερες εκτιμήσεις της ακτογραμμής από τα pRBFN και FFNN, ενώ οι καλύτερες προβλέψεις προκύπτουν από το προτεινόμενο ΤΝΔ 3^{ου} βαθμού ($r=3$). Παρόμοια αποτελέσματα προκύπτουν κι από την περιοχή Faro (Σχήμα 2.11) με το pRBFN να δίνει τη χειρότερη ακτογραμμή και το προτεινόμενο ΤΝΔ 4^{ου} βαθμού ($r = 4$) την καλύτερη.

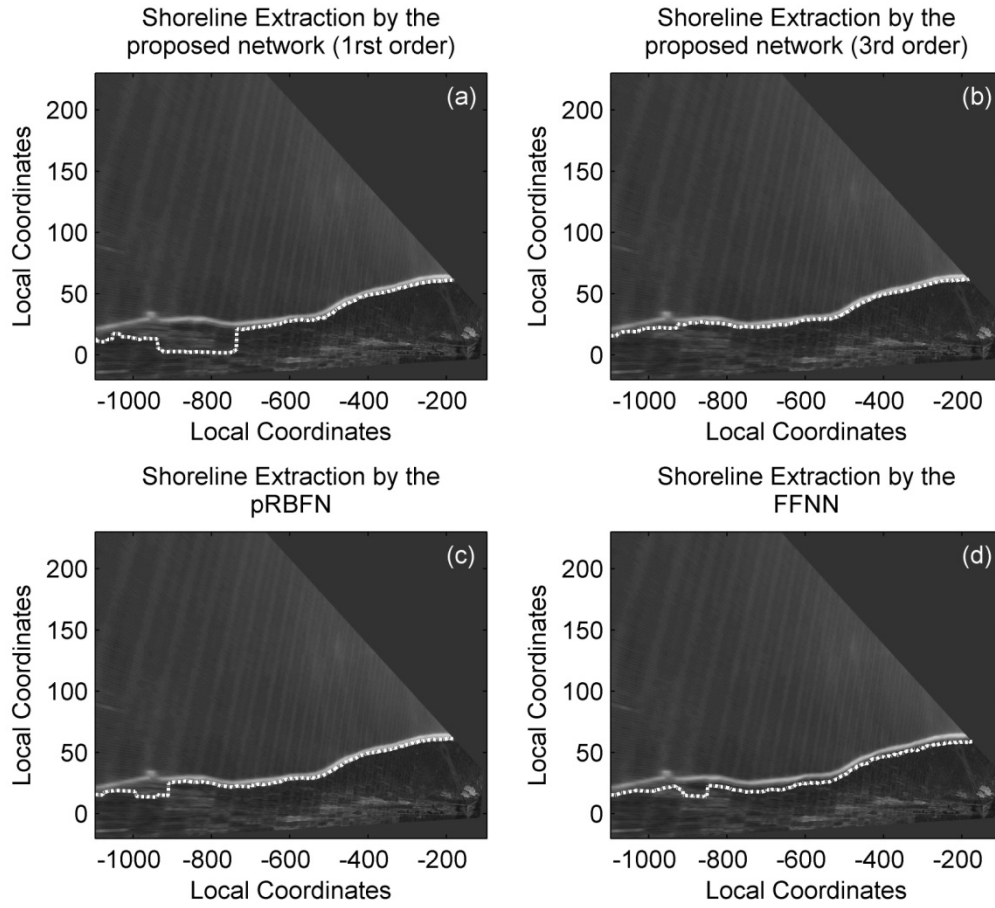
Πίνακας 2.17 Αποτελέσματα ελέγχου (MSE των ΤΝΔ) του Πειράματος 2.

c	pRBFNN	FFNN
4	0.07229 ± 0.01341	0.06407 ± 0.03197
6	0.06858 ± 0.01188	0.05249 ± 0.02863
8	0.05257 ± 0.00756	0.04739 ± 0.02664
10	0.05027 ± 0.00914	0.04221 ± 0.02398
12	0.05175 ± 0.01330	0.04663 ± 0.03452
15	0.04942 ± 0.00950	0.04044 ± 0.02544

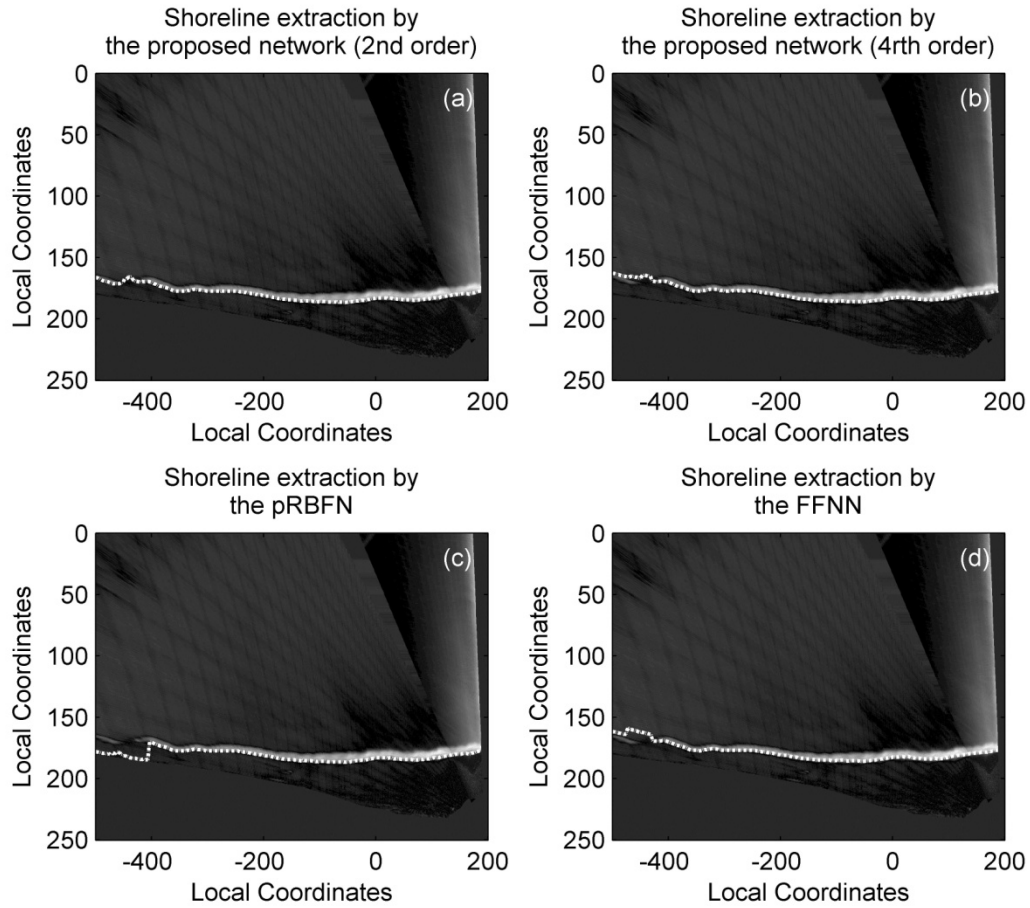
Πίνακας 2.18 Αποτελέσματα ελέγχου (MSE) από το προτεινόμενο ΤΝΔ για το Πείραμα 2.

c	Προτεινόμενο, $r = 1$	Προτεινόμενο, $r = 2$	Προτεινόμενο, $r = 3$	Προτεινόμενο, $r = 4$
4	0.11103 ± 0.14505	0.03226 ± 0.02504	0.02331 ± 0.01319	0.02437 ± 0.01082
6	0.13104 ± 0.08830	0.03165 ± 0.02524	0.01865 ± 0.00773	0.02013 ± 0.00894
8	0.11065 ± 0.12570	0.02713 ± 0.01877	0.02098 ± 0.01153	0.02074 ± 0.01138
10	0.10774 ± 0.10501	0.02987 ± 0.01563	0.01916 ± 0.00297	0.01728 ± 0.00626
12	0.10397 ± 0.14599	0.02797 ± 0.00899	0.01824 ± 0.00383	0.01992 ± 0.00571
15	0.11798 ± 0.07766	0.02522 ± 0.01034	0.01826 ± 0.00405	0.01901 ± 0.00981

**Σχήμα 2.9** Οι μέσες τιμές των πινάκων 2.17-2.18



Σχήμα 2.10 Οι ακτογραμμές που εξάγονται (διακεκομμένη γραμμή) από εικόνες τύπου SIGMA από την περιοχή της Αμμουδάρας με τη χρήση διαφορετικών ΤΝΔ. Η εικόνα προέκυψε από το σύνολο δεδομένων εκπαίδευσης του πειράματος 2. Τα ΤΝΔ που επιλέχθηκαν είναι όλα με 10 κρυφούς κόμβους c . Οι άξονες αναφέρονται σε τοπικό σύστημα συντεταγμένων σε μέτρα (βλ. παράρτημα Β.1)



Σχήμα 2.11 Οι ακτογραμμές που εξάγονται (διακεκομμένη γραμμή) από εικόνες τύπου SIGMA από την περιοχή της Faro με τη χρήση διαφορετικών ΤΝΔ. Η εικόνα προέκυψε από το σύνολο δεδομένων εκπαίδευσης του πειράματος 2. Τα ΤΝΔ που επιλέχθηκαν είναι όλα με 10 κρυφούς κόμβους c . Οι άξονες αναφέρονται σε τοπικό σύστημα συντεταγμένων σε μέτρα (βλ. παράρτημα Β.1)

Στον πίνακα 2.19 δίνονται τα αποτελέσματα που αφορούν τη φυσική σημασία του προβλήματος. Δίνονται τα σφάλματα $RMSE$ της θέσης της ακτογραμμής (εξίσωση (2.67)) τα οποία κυμαίνονται από 1.8 ως 2.9m αναλόγως το ΤΝΔ καθώς και τα ποσοστά αποδεκτών αποτελεσμάτων τα οποία προκύπτουν έχουν τιμές από 66 ως και 77%. Το προτεινόμενο ΤΝΔ 1^{ου} βαθμού δίνει το μικρότερο ποσοστό αποδεκτών αποτελεσμάτων (63.42%) και το μεγαλύτερο $RMSE$ σφάλμα: η απόδοσή του είναι χειρότερη από τα pRBFN, FFNN, τη μέθοδο των Rigos et al.[159] και από τη μέθοδο SLIM. Την καλύτερη απόδοση την δίνει το 3^{ου} βαθμού προτεινόμενο δίκτυο (ποσοστό αποδεκτών αποτελεσμάτων μεγαλύτερο του 76% και $RMSE$ 1.81m)

Πίνακας 2.19 Μέσες τιμές σφαλμάτων RMSE θέσεων ακτογραμμής σε μέτρα και ποσοστά αποδεκτών αποτελεσμάτων για την παραλία της Αμμουδάρας (δεδομένα εκπαίδευσης ΤΝΔ του πειράματος 2)

	Προτεινόμενο ΤΝΔ				pRBFNN	FFNN	Rigos et al [159]	SLIM
	$r = 1$	$r = 2$	$r = 3$	$r = 4$				
Μέσο RMSE θέσης ακτογραμμής (σε μέτρα)	2.93798	1.92634	1.80809	1.91493	2.10740	2.08603	2.29179	4.40357
Ποσοστά αποδεκτών αποτελεσμάτων	63.42%	75.65%	76.25%	75.42%	69.48%	71.94%	67.05%	71.32%

Ο πίνακας 2.20 παρουσιάζει τα αποτελέσματα που αφορούν τη φυσική σημασία του προβλήματος για την περιοχή Faro και περιέχει επίσης τα αποτελέσματα των Vousdoukas et al. [158]. Εδώ τα ποσοστά αποδεκτών αποτελεσμάτων κυμαίνονται από 26 ως 87% και τα μέσα *RMSE* σφάλματα από 1.5 ως 2.7m. Τη χειρότερη απόδοση και εδώ έχει το προτεινόμενο ΤΝΔ 1^ο βαθμού ενώ την καλύτερη από όλα τα δίκτυα το προτεινόμενο 4^ο βαθμού. Αξίζει να σημειωθεί ότι τα αποτελέσματα εδώ μπορούν να συγκριθούν με αυτά των Vousdoukas et al. [158] παρότι τα αποτελέσματα του τελευταίου είναι πάνω σε σύνολο δεδομένων εκπαίδευσης ενώ τα αποτελέσματα του πίνακα 2.20 προκύπτουν από δεδομένα ελέγχου. Το προτεινόμενο ΤΝΔ είναι ένα γρήγορο και πολύ αποδοτικό μοντέλο αν συγκριθεί με τις προϋπάρχουσες μεθόδους

Πίνακας 2.20 Μέσες τιμές σφαλμάτων RMSE θέσεων ακτογραμμής σε μέτρα και ποσοστά αποδεκτών αποτελεσμάτων για την παραλία της Faro (δεδομένα ελέγχου των ΤΝΔ του πειράματος 2)

	Προτεινόμενο ΤΝΔ				pRBFNN	FFNN	Rigos et al [159]	Vousdoukas et al [158]	SLIM
	$r = 1$	$r = 2$	$r = 3$	$r = 4$					
Μέσο RMSE θέσης ακτογραμμής (σε μέτρα)	2.67432	1.62635	1.52280	1.50591	1.57544	1.63576	2.25292	1.65	4.28909
Ποσοστά αποδεκτών αποτελεσμάτων	25.93%	84.07%	85.03%	86.95%	76.65%	75.93%	53.48%	65%	55.96%

3. Ανάπτυξη Νευροασαφούς Δικτύου Με Χρήση Πολυωνύμων Hermite

Στα πλαίσια της εκπόνησης της παρούσας διδακτορικής διατριβής αναπτύχθηκε ένα νευροασαφές δίκτυο [161] το οποίο αποτελείται από τις εξής τρεις φάσεις: (α) ένα πολυωνυμικό ΤΝΔ με τη χρήση ορθογωνίων πολυωνύμων Hermite, (β) ένα ασαφές σύστημα και (γ) τη φάση εξαγωγής συμπεράσματος το οποίο χρησιμοποιεί τα δεδομένα των δύο προηγούμενων φάσεων.

Στόχος του δικτύου αυτού είναι η χρήση του η χρήση του σε γραμμικά δεδομένα αλλά και σε δεδομένα τα οποία είναι ισχυρά μη-γραμμικά (όπως αυτά της ενότητας 3.6) ώστε να αυξηθούν οι προσεγγιστικές δυνατότητες που μπορεί να προσφέρει. Για το λόγο αυτό έγινε η συνένωση ενός ασαφούς συστήματος τα οποία χρησιμοποιούνται σε μη-γραμμικά δεδομένα (βλ. [121, 122]) μαζί και με ένα πολυωνυμικό ΤΝΔ τα οποία και αυτά χρησιμοποιούνται σε μη-γραμμικά σύνολα δεδομένων ώστε να αυξηθεί η απόδοση των δικτύων (βλ. [20, 21, 23-25]). Το προτεινόμενο νευροασαφές δίκτυο εφαρμόστηκε και σε ένα πρωτότυπο σύνολο δεδομένων που έχει στόχο να προβλέψει την μετατόπιση των ακτογραμμών και είναι ισχυρά μη-γραμμικό (βλ. [161] και παράγραφο 3.6). Όμως πραγματοποιήθηκαν και επιπλέον πειράματα (αριθμητικές προσομοιώσεις) του δικτύου αυτού σε σύνολα δεδομένων του UCI repository (βλ. Παράρτημα Α και [100]) ώστε να παρουσιαστεί η απόδοσή του και σε ευρέως γνωστά σύνολα δεδομένων.

Η δομή των ακολούθων παραγράφων είναι η εξής: πρώτα δίνεται μια βιβλιογραφική ανασκόπηση της χρήσης των πολυωνύμων Hermite μέσα σε ΤΝΔ, έπειτα δίνονται οι ορισμοί και οι ιδιότητες των πολυωνύμων Hermite τα οποία χρησιμοποιούνται στο νευροασαφές δίκτυο, ακολουθεί η δομή του νευροασαφούς δικτύου και η διαδικασία εκπαίδευσής του. Τέλος ακολουθούν οι παράγραφοι των αριθμητικών προσομοιώσεων στα σύνολα δεδομένων του UCI repository και στο φυσικό πρόβλημα που εφαρμόστηκε το προτεινόμενο νευροασαφές σύστημα στην [161].

3.1 Συναφείς μέθοδοι και προτεινόμενη μέθοδος

Στη βιβλιογραφία τα πολυώνυμα Hermite έχουν χρησιμοποιηθεί από πολλούς μέσα σε ΤΝΔ, ακολουθεί μια ανασκόπηση των κυριοτέρων. Οι Ma και Khorasani [20] χρησιμοποιούν τα Hermite ως συναρτήσεις ενεργοποίησης σε ένα πολυωνυμικό ΤΝΔ αφού πρώτα όμως τα ορθοκανονικοποιούν. Το πολυωνυμικό ΤΝΔ το οποίο χρησιμοποίησαν έχει τη μορφή τύπου Functional Link ΤΝΔ με τα ορθοκανονικοποιημένα πολυώνυμα Hermite να αποτελούν τις συναρτήσεις ενεργοποίησης του. Η ορθοκανονικοποίηση των πολυωνύμων Hermite $H_i(x)$, δημιουργεί τα πολυώνυμα:

$$h_n(x) = \alpha_n H_n(x) \varphi(x) \quad (3.1)$$

όπου:

$$\alpha_n = (n!)^{-\frac{1}{2}} \pi^{\frac{1}{4}} 2^{\frac{n-1}{2}} \quad (3.2)$$

και

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (3.3)$$

τα οποία είναι ορθοκανονικά στο \mathbb{R} , χωρίς τη χρήση συνάρτησης βαρύτητας, δηλαδή ισχύει ότι:

$$\int_{-\infty}^{+\infty} h_i(x) h_j(x) dx = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (3.4)$$

ενώ τα (μη-ορθοκανονικοποιημένα) πολυώνυμα Hermite χρειάζεται η συνάρτηση βαρύτητας e^{-x^2} (βλ. εξίσωση (3.19)).

Για τις παραγώγους των $h_n(x)$ ισχύει ότι:

$$\frac{dh_n(x)}{dx} = \begin{cases} \sqrt{2n} h_{n-1}(x) - x h_n(x), & n \geq 1 \\ -x h_0(x), & n = 0 \end{cases} \quad (3.5)$$

Στην [20] έγιναν αριθμητικές προσομοιώσεις σε δεδομένα που παρήχθησαν από τρεις συναρτήσεις παλινδρόμησης (regression functions) και δύο συναρτήσεις ταξινόμησης (classification functions) και οι συγγραφείς κατέληξαν σε καλύτερα αποτελέσματα με το ΤΝΔ που προτείνουν σε σχέση με συνήθη ΤΝΔ της βιβλιογραφίας με τα οποία συγκρίθηκαν.

Οι ίδιοι συγγραφείς στην [21] χρησιμοποιούν το ίδιο ΤΝΔ στη συμπίεση ψηφιακών εικόνων. Για να επιτευχθεί η συμπίεση αυτή κάθε ψηφιακή εικόνα μεγέθους $L \times L$ διαμερίζεται σε P επιμέρους τετράγωνα υποεικόνες (blocks) μεγέθους $M \times M$, κατόπιν κάθε block

μετατρέπεται σε ένα διάνυσμα I μεγέθους $1 \times M^2$. Ως δεδομένα εισόδου X του ΤΝΔ τότε είναι τα P στο πλήθος διανύσματα I . Η έξοδος Y του ΤΝΔ είναι πάλι τα δεδομένα X μιας και στόχος της συμπίεσης εικόνων είναι η ελαχιστοποίηση της απώλειας δεδομένων. Ως συνάρτηση σφάλματος του ΤΝΔ χρησιμοποιήθηκε η συνάρτηση Peak Signal-to-Noise Ratio (PSNR) η οποία ορίζεται ως :

$$PSNR = 10 \cdot \log_{10} \left(\frac{P \cdot I \cdot 255^2 \text{MAX}_I^2}{MSE} \right) \quad (3.6)$$

όπου το MSE σφάλμα είναι αυτό που προκύπτει μεταξύ της εξόδου που προκύπτει από το ΤΝΔ και της επιθυμητής εξόδου. Οι συγγραφείς εφάρμοσαν τη μέθοδο συμπίεσης εκπαιδεύοντας το ΤΝΔ σε δύο γνωστές (στον τομέα της επεξεργασίας εικόνας) εικόνες, τις Lena και Lake, και τα δεδομένα που προέκυψαν από την εκπαίδευση, τα χρησιμοποίησαν ώστε να συμπεστεί η εικόνα Girl. Κατέληξαν σε επιτυχή συμπίεση των τριών παραπάνω εικόνων η οποία (όπως αναφέρουν) στηρίζεται στην μη-γραμμικότητα των πολυωνύμων Hermite.

Ο Beliczynski [22] προσέγγισε συναρτήσεις πολλών μεταβλητών με δυσδιάστατα πολυώνυμα Hermite τα οποία τα όρισε ως :

$$h_n^e(t, \sigma) = \frac{1}{\sqrt{\sigma 2^n n! \sqrt{\pi}}} e^{-\frac{t^2}{2\sigma^2}} H_n \left(\frac{t}{\sigma} \right) \quad (3.7)$$

Ως εφαρμογή των πολυωνύμων αυτών προσέγγισε τις πολυδιάστατες συναρτήσεις :

$$\begin{aligned} f(x_1, x_2) &= x_1 x_2 e^{-x_1^2 - x_2^2} \\ g(x_1, x_2, x_3) &= x_1 e^{-x_1^2 - x_2^2} \sin(x_1 + x_2 + x_3) \\ \Gamma(z) &= \ln \left(\int_0^{+\infty} t^{z-1} e^{-t} dt \right) \end{aligned} \quad (3.8)$$

Για την πρώτη συνάρτηση κατέληξε σε ικανοποιητική προσέγγισή της ($MSE = 9 \cdot 10^{-12}$) μέσα σε 2-3sec (με χρήση H/Y) ακόμη και με χρήση θορύβου στα δεδομένα. Για τη συνάρτηση g επίσης βρέθηκε ικανοποιητική προσέγγισή της μέσα σε 5sec. Για να επιτευχθεί μικρό σφάλμα όμως ($MSE = 1.3$) στην προσέγγιση της τρίτης συνάρτησης χρειάζονται πολλές (180) συναρτήσεις βάσης Hermite.

Οι Shi et al. [23] δημιούργησαν ένα χαοτικό ΤΝΔ με πολώνυμα Hermite και κατόπιν το χρησιμοποίησαν σε ηλεκτρονικά κυκλώματα (με χρήση memristors^{*}) για να κρυπτογραφήσουν δεδομένα.

Οι Linh et al. [24] έχοντας δεδομένα από τους χτύπους της καρδιάς, προσέγγισαν τα δεδομένα αυτά ως εκφράσεις πολυωνύμων Hermite και κατόπιν εφάρμοσαν ένα ασαφές σύστημα Takagi-Sugeno-Kang (βλ. παράγραφο 1.4) στα δεδομένα αυτά.

Οι Rasiah et al. [25] χρησιμοποιούν τα ορθοκανονικά πολώνυμα Hermite ως συναρτήσεις βάσης για να μοντελοποιήσουν μονοδιάστατα σήματα. Πιο συγκεκριμένα χρησιμοποιούν τα πολώνυμα αυτά στον τομέα της βιοϊατρικής με σκοπό την ανίχνευση χαρακτηριστικών (feature extraction) μέσω μεθόδων ανάλυσης σήματος (signal analysis).

Όλοι οι παραπάνω συγγραφείς χρησιμοποιούν είτε τα ορθοκανονικοποιημένα πολώνυμα Hermite, είτε τα δεδομένα εισόδου απευθείας πάνω στα πολώνυμα (Shi et al. [23]), είτε χρησιμοποιούν ασαφή συστήματα (Linh et al. [24]). Όμως το προτεινόμενο νευροασαφές σύστημα λειτουργεί εντελώς διαφορετικά, καταρχήν δεν εφαρμόζει τα πολώνυμα Hermite στα δεδομένα εισόδου αλλά σε γραμμικούς συνδυασμούς τους μέσω συναπτικών βαρών με σκοπό να αυξήσει τη απόδοση του δικτύου (αφού γίνει η εύρεση των κατάλληλων συναπτικών βαρών), επίσης χρησιμοποιεί παράλληλα και ένα ασαφές σύστημα (φάση υπόθεσης του προτεινόμενου δικτύου) για να μπορέσει με τη συνεργασία των δύο αυτών τμημάτων να αυξήσει παραπάνω την απόδοση του δικτύου.

* Προέρχεται από τη λέξη memory resistor. Είναι ένα υποθετικό ηλεκτρονικό εξάρτημα που σχετίζει το ηλεκτρικό φορτίο με τη μαγνητική ροή.

3.2 Πολυώνυμα Hermite

3.2.1 Ορισμοί και αναδρομικές σχέσεις

Εξίσωση του Hermite ονομάζεται η διαφορική εξίσωση:

$$\frac{d^2 y}{dx^2} - 2x \frac{dy}{dx} + 2ny = 0 \quad (3.9)$$

Η προσπάθεια εύρεσης πεπερασμένης λύσης $y = y(x)$ της διαφορικής εξίσωσης αυτής για την οποία να ισχύει ότι:

$$\lim_{x \rightarrow +\infty} \left(e^{\frac{1}{2}x^2} y(x) \right) = 0 \quad (3.10)$$

καταλήγει στα πολυώνυμα [162] :

$$H_n(x) = \sum_{r=0}^{\lfloor n/2 \rfloor} (-1)^r \frac{n!}{r!(n-2r)!} (2x)^{n-2r}, \text{ με } n = 0, 1, 2, \dots \quad (3.11)$$

τα οποία ονομάζονται πολυώνυμα Hermite (στη βιβλιογραφία πολλές φορές αναφέρονται και ως 'physicists' Hermite polynomials^{*}). Ένας απλούστερος τύπος ο οποίος δίνει τα πολυώνυμα αυτά είναι ο :

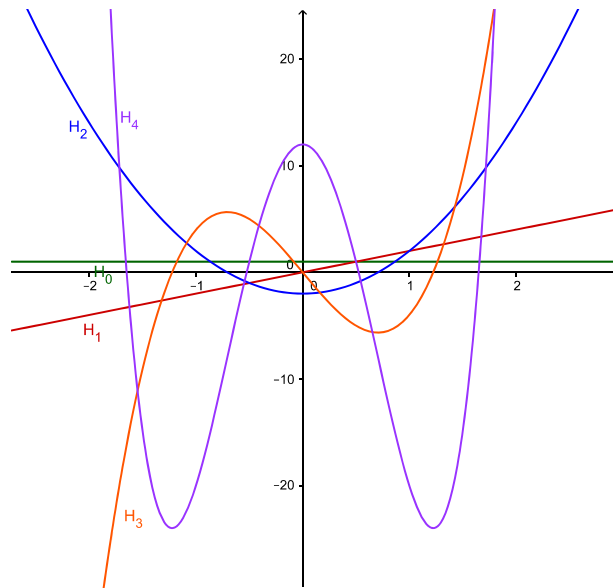
$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \quad (3.12)$$

Από την εξίσωση (3.12) προκύπτει ότι τα πρώτα πέντε πολυώνυμα είναι τα :

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= 2x \\ H_2(x) &= 4x^2 - 2 \\ H_3(x) &= 8x^3 - 12x \\ H_4(x) &= 16x^4 - 48x^2 + 12 \end{aligned} \quad (3.13)$$

οι γραφικές παραστάσεις των οποίων παρουσιάζονται στο Σχήμα 3.1.

^{*} Στην βιβλιογραφία υπάρχουν και τα 'probabilists' Hermite polynomials' τα οποία συμβολίζονται ως $He_n(x)$, ορίζονται ως $He_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}$ αλλά είναι παρόμοια με τα πολυώνυμα Hermite που ασχολούμαστε στο παρόν κείμενο μιας και ισχύουν οι σχέσεις: $H_n(x) = 2^{\frac{n}{2}} He_n(\sqrt{2}x)$, $He_n(x) = 2^{-\frac{n}{2}} H_n\left(\frac{x}{\sqrt{2}}\right)$ που μετρατρέπουν τα physicists' Hermite στα probabilists' Hermite κι αντίστροφα.



Σχήμα 3.1 Τα πολυώνυμα Hermite $H_n(x)$ βαθμών $n = 0,1,2,3,4$.

Για προγραμματιστικούς σκοπούς, τα πολυώνυμα Hermite μπορούν να οριστούν και αναδρομικά από τη σχέση:

$$\left\{ \begin{array}{l} H_0(x) = 1 \\ H_1(x) = 2x \\ H_n(x) = 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x), \quad n = 2, 3, \dots \end{array} \right\} \quad (3.14)$$

Ενώ για τις παραγώγους των πολυωνύμων αυτών ισχύει ότι [149]:

$$\left\{ \begin{array}{l} \frac{d}{dx} H_0(x) = 0 \\ \frac{d}{dx} H_n(x) = 2nH_{n-1}(x), \quad n \geq 1 \end{array} \right\} \quad (3.15)$$

Αντίστροφα, αν θέλουμε να γράψουμε μια δύναμη του x μέσω πολυωνύμων Hermite, τότε χρησιμοποιούμε την εξίσωση*:

$$x^n = \frac{n!}{2^n} \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{1}{m!(n-2m)!} H_{n-2m}(x) \quad (3.16)$$

Οι τιμές των πολυωνύμων αυτών για $x = 0$ ονομάζονται αριθμοί Hermite και έχουν την τιμή:

$$H_n(0) = \begin{cases} 0, & n : \text{περιττός} \\ (-1)^{\frac{n}{2}} 2^{\frac{n}{2}} (n-1)!!, & n : \text{άρτιος} \end{cases} \quad (3.17)$$

* "18 Orthogonal Polynomials, Classical Orthogonal Polynomials, Sums". Digital Library of Mathematical Functions. National Institute of Standards and Technology. <http://dlmf.nist.gov/18.18.E20>

όπου το σύμβολο $!!$ ορίζεται ως το γινόμενο των περιττών ακεραίων από το 1 ως και τον περιττό $n-1$, δηλαδή για n άρτιο ορίζεται ως: $(n-1)!! = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (n-1)$.

Τα πολυώνυμα Hermite μπορούν να προκύψουν και από τη σειρά Taylor της e^{2xt-t^2} ως [64]:

$$e^{2xt-t^2} = H_0(x) + \frac{H_1(x)}{1!}t + \frac{H_2(x)}{2!}t^2 + \dots \quad (3.18)$$

3.2.1 Ορθογωνιότητα

Τα πολυώνυμα Hermite είναι ορθογώνια στο \mathbb{R} αν χρησιμοποιηθεί ως συνάρτηση βαρύτητας η $w(x) = e^{-x^2}$, δηλαδή ικανοποιούν τη σχέση εσωτερικού γινομένου:

$$\langle H_m(x), H_n(x) \rangle = \int_{-\infty}^{\infty} H_m(x)H_n(x)e^{-x^2} dx = \sqrt{\pi} 2^n n! \delta_{mn} = \begin{cases} \sqrt{\pi} 2^n n!, & m = n \\ 0, & m \neq n \end{cases} \quad (3.19)$$

Η ορθογωνιότητα αυτή σε ολόκληρο το \mathbb{R} θεωρείται πλεονέκτημα, μιας και όπως παρουσιάστηκε στο προηγούμενο κεφάλαιο, σε πολυώνυμα που είναι ορθογώνια μόνο σε συγκεκριμένο διάστημα, πρέπει να γίνουν αλλαγές διαστημάτων σε μεταβλητές για να καθιστεί εφικτή η χρήση τους μέσα σε ΤΝΔ.

3.2.3 Μέθοδοι παρεμβολής

Οι χρησιμότητες των πολυωνύμων αυτές είναι πάρα πολλές, ενδεικτικά αναφέρονται κάποιες. Κάθε πολυώνυμο $f(x)$ βαθμού m μπορεί αν γραφεί ως πεπερασμένη σειρά πολυωνύμων Hermite ως:

$$f(x) = \sum_{r=0}^m c_r H_r(x) \quad (3.20)$$

όπου οι συντελεστές c_r υπολογίζονται ως :

$$c_r = \frac{1}{2^r r! \sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-x^2} f(x) H_r(x) dx \quad (3.21)$$

Επίσης και κάθε συνάρτηση $f(x)$ η οποία ορίζεται στο \mathbb{R} , είναι λεία* σε κάθε διάστημα $[-a, a]$, και για την οποία ισχύει ότι το ολοκλήρωμα:

* Λεία (smooth) ονομάζεται μια συνάρτηση η οποία είναι άπειρες φορές παραγωγίσιμη στο πεδίο ορισμού της.

$$\int_{-\infty}^{+\infty} e^{-x^2} f^2(x) dx \quad (3.22)$$

είναι πεπερασμένο, μπορεί να προσεγγιστεί από μη-πεπερασμένες σειρές πολυωνύμων Hermite ως [163] :

$$f(x) = \sum_{r=0}^{\infty} c_r H_r(x) \quad (3.23)$$

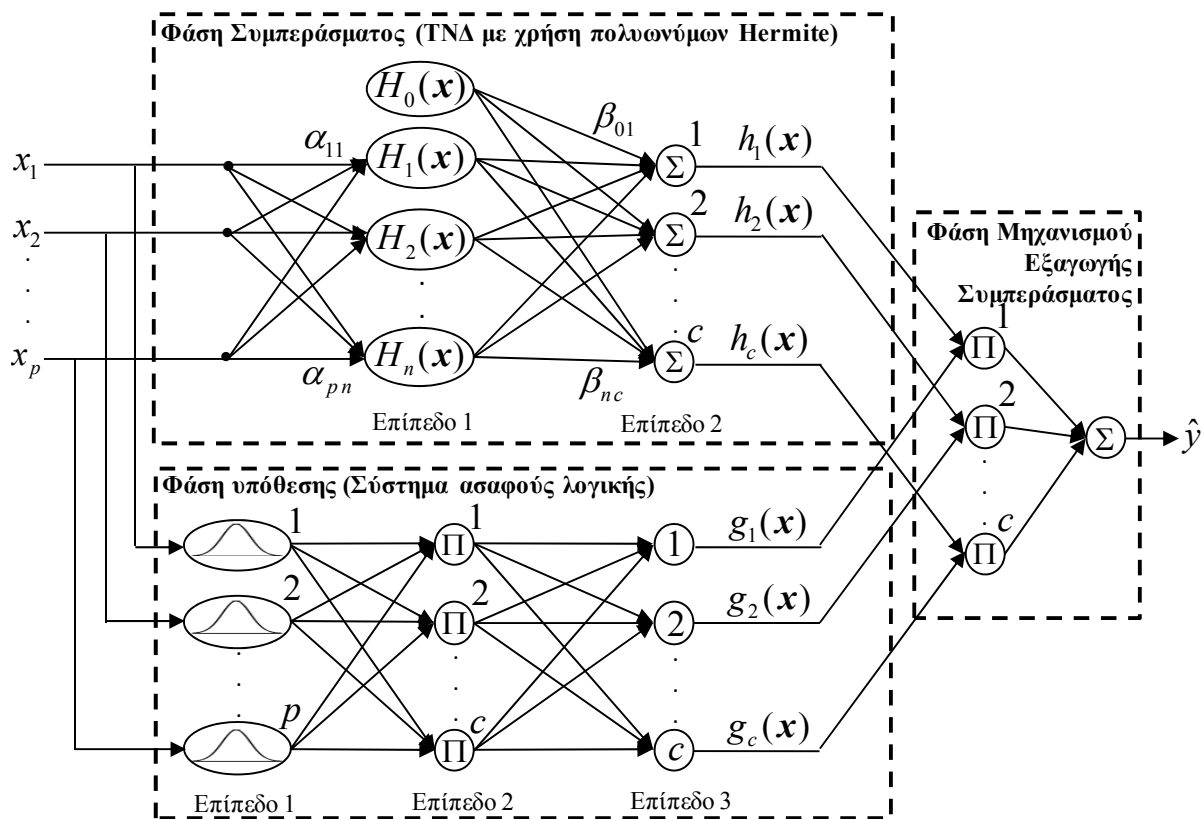
με συντελεστές c_r αυτούς της εξίσωσης (3.21). Πρέπει να σημειωθεί ότι η προσέγγιση της (3.23) ισχύει μόνο για τα σημεία x στα οποία η $f(x)$ είναι συνεχής συνάρτηση.

3.3 Το Προτεινόμενο Νευροασαφές δίκτυο με τα πολυώνυμα Hermite

Το νευροασαφές δίκτυο που δημιουργήθηκε, αποτελείται από 3 μέρη (Σχήμα 3.2):

- (α) την φάση υπόθεσης η οποία περιέχει ένα σύστημα ασαφούς λογικής
- (β) την φάση συμπεράσματος η οποία περιέχει ένα πολυωνυμικό ΤΝΔ που χρησιμοποιεί τα ορθογώνια πολυώνυμα Hermite και αποτελείται από δύο επίπεδα.
- (γ) την φάση εξαγωγής συμπεράσματος.

Ακολουθεί η ανάλυσή του.



Σχήμα 3.2 Η τοπολογία του Νευροασαφούς συστήματος με χρήση πολυωνύμων Hermite.

3.2.1 Φάση Υπόθεσης : Σύστημα ασαφούς λογικής

Η φάση υπόθεσης του δικτύου αποτελείται από τρία επίπεδα (Σχήμα 3.2). Στο πρώτο επίπεδο τα δεδομένα εισόδου διαμερίζονται σε c ασαφή σύνολα της μορφής:

$$A_{ij}(x_{kj}) = \exp\left(-\left(\frac{x_{kj} - v_{ij}}{\sigma_{ij}}\right)^2\right) \quad (3.24)$$

όπου v_{ij} είναι το κέντρο του ασαφές συνόλου και σ_{ij} το αντίστοιχο πλάτος, με $1 \leq k \leq N$, $1 \leq i \leq c$, $1 \leq j \leq p$. Άρα υπάρχουν c στο πλήθος ασαφείς κανόνες και κάθε ασαφές σύνολο A_{ij} συνδέεται μόνο με έναν ασαφή κανόνα.

Το δεύτερο επίπεδο αποτελείται από c ασαφείς κόμβους οι οποίοι χρησιμοποιούν τα δεδομένα του πρώτου επιπέδου και δίνουν ως έξοδο την:

$$f_i(\mathbf{x}_k) = \prod_{j=1}^p A_{ij}(x_{kj}) \quad (3.25)$$

όπου $1 \leq k \leq N$ και $1 \leq i \leq c$. Πρέπει να σημειωθεί ότι το $f_i(\mathbf{x}_k)$ είναι ένα πολυδιάστατο ασαφές σύνολο και η συνεισφορά του στο σύστημα είναι να υπολογίζει τις συνάψεις πυροδότησης των αντίστοιχων ασαφών κανόνων.

Το τρίτο επίπεδο αποτελείται από c κόμβους καθένας από τους οποίους συσσωματώνει τους κανόνες του προηγούμενου στρώματος σύμφωνα με την ακόλουθη εξίσωση:

$$g_i(\mathbf{x}_k) = \frac{f_i(\mathbf{x}_k)}{\sum_{s=1}^c f_s(\mathbf{x}_k)} \quad (3.26)$$

3.2.2 Φάση Συμπεράσματος: ΤΝΔ με χρήση πολυωνύμων Hermite

Θα γίνει η υπόθεση ότι τα δεδομένα εκπαίδευσης του νευροασαφούς συστήματος έχουν p -διάστατη είσοδο και μονοδιάστατη έξοδο, δηλαδή έχουν τη μορφή που έχει αναφερθεί και στις προηγούμενες ενότητες:

$$S = \{(\mathbf{x}_k, y_k) / \mathbf{x}_k \in \mathbb{R}^p, y_k \in \mathbb{R}, k = 1, 2, \dots, N\} \quad (3.27)$$

για την ανάλυση που ακολουθεί.

Η φάση συμπεράσματος του νευροασαφούς συστήματος είναι ένα πολυωνυμικό ΤΝΔ που αποτελείται από δύο επίπεδα (Layers) και εμπεριέχει δύο σύνολα συναπτικών βαρών, τα

συναπτικά βάρη μεταξύ των δεδομένων εισόδου και των πολυωνυμικών συναρτήσεων που αποτελούν το πρώτο επίπεδο και αυτά μεταξύ του πρώτου και του δευτέρου επιπέδου.

Το πρώτο επίπεδο (επίπεδο 1 στο Σχήμα 3.2) αποτελείται από τα πολυώνυμα Hermite μηδενικού ως και n -στού βαθμού τα οποία έχουν το ρόλο των συναρτήσεων ενεργοποίησης σε αυτό το τμήμα του δικτύου. Επειδή το πολυώνυμο μηδενικού βαθμού $H_0(x)$ είναι σταθερό (βλ. εξισώσεις (3.13)), άρα δεν εξαρτάται από τα δεδομένα εισόδου και μπορεί να θεωρηθεί ως bias του δικτύου. Δηλαδή στο ℓ -στό κόμβο του πρώτου επιπέδου του δικτύου εισέρχεται ο αριθμός $\sum_{j=1}^p \alpha_{j\ell} x_{kj}$ και η έξοδος του ℓ -στού κόμβου είναι:

$$\begin{aligned} H_0(\mathbf{x}_k) &= 1 \\ H_\ell(\mathbf{x}_k) &= H_\ell\left(\sum_{j=1}^p \alpha_{j\ell} x_{kj}\right), \quad 1 \leq \ell \leq n \end{aligned} \quad (3.28)$$

όπου $\alpha_{j\ell}$, $1 \leq j \leq p, 1 \leq \ell \leq n$ τα συναπτικά βάρη του δικτύου που συνδέουν τα δεδομένα εισόδου με το πρώτο επίπεδο. Το πρώτο και το δεύτερο επίπεδο του δικτύου συνδέονται μέσω των συναπτικών βαρών $\beta_{\ell i}$, $0 \leq \ell \leq n, 1 \leq i \leq c$. Το δεύτερο επίπεδο αποτελείται από c το πλήθος κόμβους άθροισης και η έξοδος τότε του i -στού κόμβου από αυτούς είναι:

$$h_i(\mathbf{x}_k) = \sum_{\ell=0}^n \beta_{\ell i} H_\ell\left(\sum_{j=1}^p \alpha_{j\ell} x_{kj}\right) \quad (3.29)$$

όπου $i = 1, 2, \dots, c$ και $k = 1, 2, \dots, N$.

Θα πρέπει να σημειωθεί πως μόνο στην περίπτωση που όλα τα δεδομένα $\sum_{j=1}^p \alpha_{j\ell} x_{kj}$ είναι μεταξύ τους ίσα, η εξίσωση (3.29) θα παρίστανε μια ακριβή έκφραση σειράς πολυωνύμων Hermite. Στο παρόν δίκτυο όμως η παράσταση $h_i(\mathbf{x}_k)$ αποτελεί προσέγγιση εφόσον τα βάρη $\alpha_{j\ell}$ που συνδέουν τους νευρώνες υπολογίζονται ξεχωριστά και έχουν ως αποτέλεσμα διαφορετικές εισόδους $\sum_{j=1}^p \alpha_{j\ell} x_{kj}$ στις συναρτήσεις ενεργοποίησης H_ℓ .

3.2.3 Φάση Μηχανισμού Εξαγωγής Συμπεράσματος

Βασιζόμενοι στην ανάλυση που προηγήθηκε, ο i -στός ασαφής κανόνας του προτεινόμενου νευροασαφούς συστήματος είναι ο εξής:

$$R^i : IF x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \text{ and } \dots \text{ and } x_p \text{ is } A_{ip} \text{ THEN } y \text{ is } h_i(\mathbf{x}) = \sum_{\ell=1}^n \beta_{\ell i} H_{\ell}(\mathbf{x}) \quad (3.30)$$

Χρησιμοποιώντας τις εξισώσεις (3.29) και (3.26) αντιστοιχίζεται κάθε έξοδος $h_i(\mathbf{x}_k)$ του πολυωνυμικού ΤΝΔ με κάθε έξοδο $g_i(\mathbf{x}_k)$ του ασαφούς δικτύου χρησιμοποιώντας τον κανόνα (3.30) και προκύπτει ως έξοδος του νευροασαφούς δικτύου η :

$$\tilde{y}_k = \sum_{i=1}^c h_i(\mathbf{x}_k) g_i(\mathbf{x}_k) = \sum_{i=1}^c \left[\left(\sum_{\ell=0}^n \beta_{\ell i} H_{\ell}(\mathbf{x}_k) \right) g_i(\mathbf{x}_k) \right] \quad (3.31)$$

3.4 Διαδικασία εκπαίδευσης του δικτύου

Για την εκπαίδευση του νευροασαφούς δικτύου, υπάρχουν τριών ειδών παράμετροι οι οποίες πρέπει να προσδιοριστούν. Για την φάση συμπεράσματος (ΤΝΔ με πολυώνυμα Hermite) πρέπει να προσδιοριστούν τα συναπτικά βάρη ($\alpha_{j\ell}$ και $\beta_{\ell i}$) ενώ για την φάση υπόθεσης (ασαφές σύστημα) πρέπει να προσδιοριστούν τα ασαφή κέντρα και τα ασαφή πλάτη. Για να υπολογιστούν οι παραπάνω παράμετροι, πρώτα υπολογίστηκαν τα ασαφή κέντρα και τα ασαφή πλάτη του ασαφούς συστήματος με τη χρήση του ασαφούς πίνακα συνδιασποράς (fuzzy covariance matrix) [164, 165]. Προκειμένου να προσδιοριστούν οι παράμετροι $\alpha_{j\ell}$ και $\beta_{\ell i}$ έγινε χρήση της μεθόδου Gradient Descent χρησιμοποιώντας τον κανόνα του Armijo (βλ. [18] και παράγραφο 1.1.2) έτσι ώστε να ελαχιστοποιηθεί η συνάρτηση του ολικού τετραγωνικού σφάλματος του δικτύου:

$$J_{SE} = \sum_{k=1}^N (y_k - \tilde{y}_k)^2 = \sum_{k=1}^N \left(y_k - \sum_{i=1}^c \left[\left(\sum_{\ell=0}^n \beta_{\ell i} H_{\ell}(\mathbf{x}_k) \right) g_i(\mathbf{x}_k) \right] \right)^2 \quad (3.32)$$

Για να χρησιμοποιηθεί ο κανόνας του Armijo ορίστηκε το διάνυσμα που περιέχει τα συναπτικά βάρη που πρέπει να προσδιοριστούν :

$$\mathbf{z} = [z_1, \dots, z_{pn}, z_{pn+1}, \dots, z_{(p+c)n+c}]^T = [\alpha_{11}, \dots, \alpha_{pn}, \beta_{01}, \dots, \beta_{nc}]^T \quad (3.33)$$

και τότε στην $(t+1)$ -στή επανάληψη του αλγορίθμου Gradient Descent το νέο διάνυσμα που περιέχει τις μεταβλητές είναι:

$$\mathbf{z}(t+1) = \mathbf{z}(t) - \eta(t) \nabla J_{SE}(\mathbf{z}(t)) \quad (3.34)$$

όπου (σύμφωνα με τον κανόνα του Armijo) η παράμετρος $\eta(t)$ είναι: $\eta(t) = \gamma^t$, με γ μια σταθερά $\gamma \in (0,1)$ και t ο ελάχιστος θετικός ακέραιος τέτοιος ώστε να ισχύει ότι:

$$J_{SE}(\mathbf{z}(t) - \eta(t) \nabla J_{SE}(\mathbf{z}(t))) - J_{SE}(\mathbf{z}(t)) < -\varepsilon \eta(t) \|\nabla J_{SE}(\mathbf{z}(t))\|^2 \quad (3.35)$$

όπου ε ένας μικρός θετικός αριθμός με $\varepsilon \in (0,1)$.

3.5 Αριθμητικές προσομοιώσεις σε σύνολα δεδομένων του UCI Repository

Προκειμένου να γίνουν αριθμητικά πειράματα/προσομοιώσεις στο προτεινόμενο νευροασαφές δίκτυο επιλέχθηκαν από το UCI Repository (βλ. Παράρτημα Α και [100]) τα σύνολα δεδομένων AutoMPG, ForestFires, Housing και Servo. Το προτεινόμενο νευροασαφές σύστημα εκπαιδεύθηκε και ελέγχθηκε για $c = 3,5,6,8,9,10$ κόμβους της εξίσωσης (3.26) και $n = 2,3,4,5$ μέγιστους βαθμούς πολυωνύμων Hermite των εξισώσεων (3.28)-(3.29).

Παράλληλα για να γίνει σύγκριση και με άλλα ΤΝΔ δημιουργήθηκαν άλλα δύο συστήματα, (α) ένα ασαφές σύστημα και (β) ένα ΤΝΔ ακτινικής βάσης, η ανάλυση των οποίων ακολουθεί.

(α) Το ασαφές σύστημα Takagi-Sugeno-Kang (θα αναφέρεται ως TSKFS στην παρούσα ενότητα) δημιουργήθηκε και εκπαιδεύτηκε εφαρμόζοντας τον αλγόριθμο Fuzzy C Means στα δεδομένα εισόδου για να προσδιοριστούν τα ασαφή σύνολα και για τα πλάτη των πυρήνων χρησιμοποιήθηκε ο ασαφής πίνακας διασποράς/συνδιασποράς [164, 165]. Τέλος, οι συντελεστές της διαδικασίας εξαγωγής συμπεράσματος (consequent coefficients) υπολογίστηκαν με τη μέθοδο Gradient Descent.

(β) Δημιουργήθηκε και ένα ΤΝΔ συναρτήσεων ακτινικής βάσης, το οποίο θα αναφέρεται ως RBFNN στην παράγραφο αυτή. Στο RBFNN χρησιμοποιήθηκε ο αλγόριθμος Conditional Fuzzy C Means (βλ. παράγραφο 1.2.1.1) για τη συσταδοποίηση δεδομένων εισόδου-εξόδου του ΤΝΔ. Τα συναπτικά βάρη του δικτύου αυτού υπολογίστηκαν με τη μέθοδο των ελαχίστων τετραγώνων (βλ. παράγραφο 1.1.1).

Σε όλες τις προσομοιώσεις εφαρμόστηκε η μέθοδος 10-cross fold validation με 10 αρχικοποιήσεις (initializations). Οι αρχικοποιήσεις είναι απαραίτητο να χρησιμοποιούνται εφόσον στην εκπαίδευση υπάρχουν διαδικασίες που δεν δίνουν μονοσήμαντα αποτελέσματα (π.χ. οι αρχικοποιήσεις των κέντρων στις συσταδοποιήσεις των δεδομένων). Για τα αποτελέσματα των προσομοιώσεων του προτεινόμενου νευροασαφούς δικτύου και των δύο συστημάτων με τα οποία γίνεται σύγκριση παρουσιάζονται στους πίνακες 3.1-3.6. Ως συνάρτηση σφάλματος επιλέχθηκε η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος

$RMSE$ (βλ. εξίσωση (1.33)) και τα αποτελέσματα όλα παρουσιάζονται στη μορφή (μέση τιμή) \pm (τυπική απόκλιση) στους πίνακες των αποτελεσμάτων.

Πίνακας 3.1 Αποτελέσματα (σφάλματα RMSEs) στο σύνολο δεδομένων **AutoMPG** του UCI repository

		Προτεινόμενο, $n = 2$		Προτεινόμενο, $n = 3$	
c	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	
3	14.3932 \pm 8.7548	14.6474 \pm 8.8552	13.7056 \pm 6.8282	13.7568 \pm 7.2471	
5	14.1115 \pm 7.3065	14.1585 \pm 7.8396	13.6235 \pm 8.1949	13.4697 \pm 8.7029	
6	14.3391 \pm 8.2535	14.1942 \pm 8.4825	6.2165 \pm 2.9847	6.6572 \pm 3.6820	
8	13.0714 \pm 7.4485	13.2442 \pm 8.3718	6.3500 \pm 4.6442	6.6778 \pm 5.0487	
9	8.8316 \pm 1.9680	9.1118 \pm 2.9754	6.2573 \pm 4.6285	6.4832 \pm 4.6043	
10	8.2367 \pm 4.5722	8.5119 \pm 4.6902	5.9019 \pm 4.2250	6.0384 \pm 3.8971	

		Προτεινόμενο, $n = 4$		Προτεινόμενο, $n = 5$	
c	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	
3	10.2818 \pm 6.3181	10.3046 \pm 6.1970	6.0472 \pm 0.1835	6.3275 \pm 1.4261	
5	8.8370 \pm 7.2401	8.8182 \pm 6.8931	4.7187 \pm 0.1429	4.8710 \pm 1.3561	
6	5.6607 \pm 4.2242	6.1485 \pm 5.6965	4.2533 \pm 0.1406	4.4431 \pm 1.3865	
8	4.8458 \pm 3.4320	5.1041 \pm 4.2365	3.9313 \pm 0.1820	4.0746 \pm 1.2903	
9	4.8265 \pm 3.4840	4.9603 \pm 3.8793	3.9542 \pm 0.1782	4.1049 \pm 1.2865	
10	4.7526 \pm 3.4991	4.9241 \pm 3.8725	3.9554 \pm 0.1928	4.0953 \pm 1.3166	

		TSKFS		RBFNN	
c	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	
3	5.9538 \pm 0.1357	6.2463 \pm 1.4172	4.4711 \pm 0.1799	4.5663 \pm 1.5186	
5	4.6998 \pm 0.1407	4.8575 \pm 1.3536	4.2114 \pm 0.1773	4.2755 \pm 1.4990	
6	4.2374 \pm 0.1382	4.4415 \pm 1.3816	4.1951 \pm 0.1806	4.3103 \pm 1.5235	
8	3.9753 \pm 0.1773	4.1308 \pm 1.2855	4.1508 \pm 0.1955	4.2486 \pm 1.5816	
9	3.9799 \pm 0.1819	4.1279 \pm 1.2743	4.1430 \pm 0.1929	4.2267 \pm 1.5672	
10	3.9590 \pm 0.1918	4.1532 \pm 1.3323	4.0981 \pm 0.1968	4.2465 \pm 1.6056	

Πίνακας 3.2 Αποτελέσματα (σφάλματα RMSEs) του προτεινόμενου δικτύου στο σύνολο δεδομένων **ForestFires** του UCI repository

		Προτεινόμενο, $n = 2$		Προτεινόμενο, $n = 3$	
c	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	
3	62.9158 \pm 8.4116	41.9604 \pm 49.5644	62.8138 \pm 8.2536	41.9203 \pm 49.9570	
5	62.7929 \pm 8.3962	42.2861 \pm 49.6400	62.7737 \pm 8.4403	41.6635 \pm 50.0213	
6	62.7253 \pm 8.2246	44.0071 \pm 49.3175	62.5207 \pm 8.1675	43.9207 \pm 48.7803	
8	62.8751 \pm 8.3708	43.1581 \pm 49.9836	62.4594 \pm 8.2791	44.2328 \pm 48.8216	
9	62.5304 \pm 8.3807	43.9704 \pm 48.5767	62.4424 \pm 8.1182	44.0674 \pm 48.9258	
10	62.4654 \pm 8.3411	44.3074 \pm 48.4609	62.3084 \pm 8.2352	44.8798 \pm 48.4637	

		Προτεινόμενο, $n = 4$		Προτεινόμενο, $n = 5$	
c	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	
3	63.0582 \pm 8.3864	41.7990 \pm 50.1320	63.3130 \pm 8.4641	40.7878 \pm 50.6808	
5	63.0561 \pm 8.3405	41.5572 \pm 50.0436	63.2815 \pm 8.3450	41.0868 \pm 50.8525	
6	62.7232 \pm 8.1803	43.6186 \pm 49.1073	63.0346 \pm 8.3485	42.0534 \pm 50.3233	
8	62.6554 \pm 8.2648	43.8354 \pm 49.2546	63.0196 \pm 8.2486	42.5127 \pm 49.9552	
9	62.6525 \pm 8.2345	44.0555 \pm 48.9262	62.3400 \pm 8.3311	44.9383 \pm 48.2779	
10	62.6146 \pm 8.2431	44.7345 \pm 48.7056	62.1555 \pm 8.2598	45.9329 \pm 47.7153	

Πίνακας 3.3 Αποτελέσματα (σφάλματα RMSEs) των δικτύων TSKFS και RBFNN στο σύνολο δεδομένων **ForestFires** του UCI repository

c	TSKFS		RBFNN	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
3	62.7450 ± 8.3486	41.6059 ± 49.6217	62.8476 ± 8.2535	42.2969 ± 49.1077
5	62.6236 ± 8.3324	42.0984 ± 49.6023	62.6844 ± 8.3778	42.3014 ± 49.4619
6	62.5912 ± 8.3318	43.1384 ± 49.1288	62.6179 ± 8.3361	43.2477 ± 48.8970
8	62.3195 ± 8.2169	44.1354 ± 48.4868	62.3841 ± 8.1808	44.8823 ± 47.9280
9	62.2649 ± 8.2548	44.2061 ± 48.6901	62.3823 ± 8.2297	49.1329 ± 45.8095
10	62.2239 ± 8.2598	44.8817 ± 48.4057	62.3616 ± 8.2276	44.1953 ± 48.4961

Πίνακας 3.4 Αποτελέσματα (σφάλματα RMSEs) στο σύνολο δεδομένων **Housing** του UCI repository

c	Προτεινόμενο, $n = 2$		Προτεινόμενο, $n = 3$	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
3	10.2937 ± 5.1121	11.0129 ± 5.6321	9.5529 ± 4.3285	10.0551 ± 4.6107
5	9.6747 ± 4.2440	10.2990 ± 5.0893	8.8836 ± 1.0740	9.0798 ± 3.2037
6	8.8520 ± 2.1947	9.4221 ± 4.2048	8.8449 ± 2.7285	9.0435 ± 3.9100
8	8.5900 ± 0.8523	8.9094 ± 2.8034	8.5988 ± 0.7069	9.1740 ± 2.8883
9	8.4779 ± 0.8148	8.9213 ± 2.9339	8.4672 ± 1.0668	8.8289 ± 3.2125
10	8.4025 ± 0.7745	9.0787 ± 2.7530	8.2417 ± 1.1160	8.8064 ± 2.8442

c	Προτεινόμενο, $n = 4$		Προτεινόμενο, $n = 5$	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
3	7.3850 ± 3.0320	7.6657 ± 3.9452	8.7025 ± 3.7121	9.3882 ± 4.8544
5	7.3417 ± 3.1792	7.2914 ± 3.9992	8.9171 ± 3.4937	8.9008 ± 4.1879
6	7.3547 ± 3.2887	7.4208 ± 3.7276	9.4788 ± 3.5409	9.7133 ± 4.0239
8	8.0436 ± 3.1925	8.7651 ± 3.3952	10.1996 ± 3.4081	10.1586 ± 3.9417
9	8.1639 ± 3.0993	8.7569 ± 3.9725	9.4274 ± 3.5077	9.5285 ± 4.5176
10	7.9994 ± 3.4749	8.3535 ± 4.2536	9.4491 ± 3.5393	9.7125 ± 4.3183

c	TSKFS		RBFNN	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
3	8.1231 ± 0.3322	8.4184 ± 2.6651	8.3706 ± 0.3049	8.4206 ± 2.7753
5	8.1124 ± 0.3402	8.6577 ± 2.6785	8.1810 ± 0.3137	8.3061 ± 2.7712
6	8.0575 ± 0.3296	8.7408 ± 2.6911	8.1810 ± 0.3137	8.3061 ± 2.7712
8	8.0416 ± 0.3427	8.2624 ± 2.7829	7.8133 ± 0.3765	8.3805 ± 2.8092
9	7.9661 ± 0.3522	8.2388 ± 2.7474	7.7361 ± 0.4974	8.6563 ± 2.7140
10	7.9622 ± 0.3531	8.6625 ± 2.8065	7.4681 ± 0.4106	8.2533 ± 2.7212

Πίνακας 3.5 Αποτελέσματα (σφάλματα RMSEs) του προτεινόμενου δικτύου στο σύνολο δεδομένων **Servo** του UCI repository

c	Προτεινόμενο, $n = 2$		Προτεινόμενο, $n = 3$	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
3	1.4418 ± 0.0427	1.4383 ± 0.3189	1.4166 ± 0.1172	1.3856 ± 0.3602
5	1.4005 ± 0.0616	1.4608 ± 0.2822	1.2913 ± 0.0383	1.3358 ± 0.2621
6	1.3857 ± 0.0490	1.4761 ± 0.3148	1.2537 ± 0.0483	1.2974 ± 0.3062
8	1.3747 ± 0.0664	1.4697 ± 0.2918	1.1832 ± 0.0648	1.3110 ± 0.3281
9	1.3625 ± 0.0626	1.4962 ± 0.2921	1.1808 ± 0.0639	1.3111 ± 0.3258
10	1.3592 ± 0.0643	1.4691 ± 0.2772	1.1668 ± 0.0842	1.2984 ± 0.3105

c	Προτεινόμενο, $n = 4$		Προτεινόμενο, $n = 5$	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
3	1.2293 ± 0.1102	1.2840 ± 0.3230	1.3161 ± 0.0733	1.3176 ± 0.3214
5	1.1856 ± 0.0408	1.1801 ± 0.3343	1.2347 ± 0.0814	1.2752 ± 0.3004
6	1.1630 ± 0.1079	1.2106 ± 0.3183	1.2104 ± 0.0784	1.2734 ± 0.3081
8	1.1070 ± 0.1375	1.2092 ± 0.3482	1.1414 ± 0.0898	1.1976 ± 0.3262
9	1.0985 ± 0.1637	1.1969 ± 0.3651	1.1170 ± 0.0827	1.1856 ± 0.3256
10	1.0828 ± 0.1198	1.1499 ± 0.3309	1.0546 ± 0.1012	1.1480 ± 0.2937

Πίνακας 3.6 Αποτελέσματα (σφάλματα RMSEs) των δικτύων TSKFS και RBFNN στο σύνολο δεδομένων **Servo** του UCI repository

c	TSKFS		RBFNN	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
3	2.0891 ± 0.0539	1.9537 ± 0.5996	1.4403 ± 0.0294	1.4436 ± 0.2898
5	2.0873 ± 0.0514	1.9518 ± 0.5719	1.3924 ± 0.0284	1.4454 ± 0.2650
6	2.0883 ± 0.0514	1.9528 ± 0.5719	1.3695 ± 0.0356	1.4409 ± 0.2810
8	2.0886 ± 0.0514	1.9531 ± 0.5718	1.3461 ± 0.0380	1.4175 ± 0.2751
9	2.0889 ± 0.0514	1.9535 ± 0.5717	1.0673 ± 0.0513	1.2967 ± 0.3145
10	2.0890 ± 0.0514	1.9536 ± 0.5718	1.0620 ± 0.0468	1.2243 ± 0.2663

Όπως φαίνεται από τα παραπάνω αποτελέσματα, στο σύνολο AutoMPG και στα δεδομένα εκπαίδευσης και στα δεδομένα ελέγχου, το προτεινόμενο νευροασαφές σύστημα δίνει τα καλύτερα αποτελέσματα για $n = 5$ και $c = 8$ και ακολουθεί το TSKFS για $c = 10$. Στο σύνολο ForestFires στα δεδομένα εκπαίδευσης το προτεινόμενο νευροασαφές σύστημα δίνει τα καλύτερα αποτελέσματα για $n = 5$ και $c = 10$ και ακολουθεί το TSKFS για $c = 10$. Στα δεδομένα ελέγχου τα καλύτερα αποτελέσματα προκύπτουν από το προτεινόμενο για $n = 5$ και $c = 3$. Στο σύνολο Housing στα δεδομένα εκπαίδευσης και στα δεδομένα ελέγχου, το προτεινόμενο νευροασαφές σύστημα δίνει τα καλύτερα αποτελέσματα για $n = 4$ και $c = 5$ και ακολουθεί το RBFNN για $c = 10$. Στο σύνολο Servo στα δεδομένα εκπαίδευσης και στα δεδομένα ελέγχου, το προτεινόμενο νευροασαφές σύστημα δίνει τα καλύτερα αποτελέσματα για $n = 5$ και $c = 10$ και ακολουθεί το RBFNN για $c = 10$.

Όπως παρατηρούμε στα περισσότερα σύνολα δεδομένων όσο αυξάνουν οι παράμετροι n και c του προτεινόμενου δικτύου, τόσο πιο καλά είναι τα αποτελέσματα που προκύπτουν. Εξαιρέση αποτελεί το σύνολο Housing στο οποίο παρατηρείται ότι για μεγάλους βαθμούς πολυωνύμων και κρυφών κόμβων ($c \geq 6$ και $n \geq 4$) τα σφάλματα μεγαλώνουν. Σε αυτήν την περίπτωση μπορούμε να ισχυριστούμε ότι αυτά είναι αποτελέσματα λόγω συνθηκών overfitting.

3.6 Μοντελοποίηση/Πρόβλεψη μετατόπισης/οπισθοχώρησης ακτογραμμής με χρήση του προτεινόμενου Νευροασαφούς δικτύου

Στην παρούσα ενότητα γίνονται προσομοιώσεις με χρήση του νευροασαφούς δικτύου στο σύνολο δεδομένων Coastal Erosion (βλ. Παράρτημα Β.2 και [161]). Τα δεδομένα Coastal Erosion έχουν χρησιμοποιηθεί στο άρθρο [161] αποτελούνται από $N=588$ δεδομένα με διάσταση εισόδου $p=9$ και μονοδιάσταση έξοδο y και είναι διαθέσιμα στην ιστοσελίδα <http://isla.aegean.gr/news.php?>. Τα δεδομένα αυτά προέκυψαν στην προσπάθεια δημιουργίας ενός δικτύου που θα προβλέπει την μετατόπιση μιας ακτογραμμής γρηγορότερα από το ευρέως χρησιμοποιούμενο μοντέλο XBeach [166]. Το μοντέλο XBeach αποτελείται από ένα πακέτο αλγορίθμων ανοιχτού λογισμικού που προσομοιώνει την συμπεριφορά της κίνησης των παραλιών σε ακραίες μετεωρολογικές συνθήκες. Τα δεδομένα εισόδου του ΤΝΔ είναι τα ίδια με αυτά που χρησιμοποιήθηκαν στο μοντέλο XBeach και δημιουργήθηκαν από 8 βαθυμετρικά προφίλ της παραλίας της Ερεσού της Λέσβου. Η διαδικασία συλλογής των πρωταρχικών δεδομένων, στοιχεία της περιοχής μελέτης όπως και η διαδικασία της δημιουργίας του συνόλου δεδομένων Coastal Erosion βρίσκονται αναλυτικά στα παράρτημα Β.2.

Κάθε αριθμητικό δεδομένο $\mathbf{x}=[x_1, x_2, \dots, x_9]^T$ που χρησιμοποιήθηκε στο νευροασαφές δίκτυο, δημιουργήθηκε ώστε οι πρώτες πέντε συνιστώσες του να περιγράφουν μια ιζηματική μπάρα που ανακαλύφθηκε στα βαθυμετρικά προφίλ της περιοχής μελέτης (Σχήμα 3.3 (α)), οι επόμενες τρεις συνιστώσες αποτελούν τα χαρακτηριστικά μιας μετεωρολογικής καταιγίδας και η τελευταία συνιστώσα είναι μια ιζηματική παράμετρος της παραλίας. Οι συνιστώσες αυτές είναι οι εξής (Σχήμα 3.3) :

$$x_1 = h = \text{το ύψος αυτής της μπάρας,}$$

$$x_2 = \frac{w}{2} = \text{το μισό πλάτος της μπάρας}$$

$$x_3 = d = \text{η απόσταση της κορυφής της μπάρας από το σημείο που δείχνει το όριο θάλασσας-στεριάς}$$

$$x_4 = \tan \theta \text{ όπου } \theta \text{ η γωνία που δείχνει την κλίση αυτής της μπάρας}$$

$$x_5 = \tan \omega \text{ όπου } \omega \text{ η γωνία που δείχνει την κλίση του βυθού μεταξύ του ορίου της ακτογραμμής και της βάσης της ιζηματικής μπάρας}$$

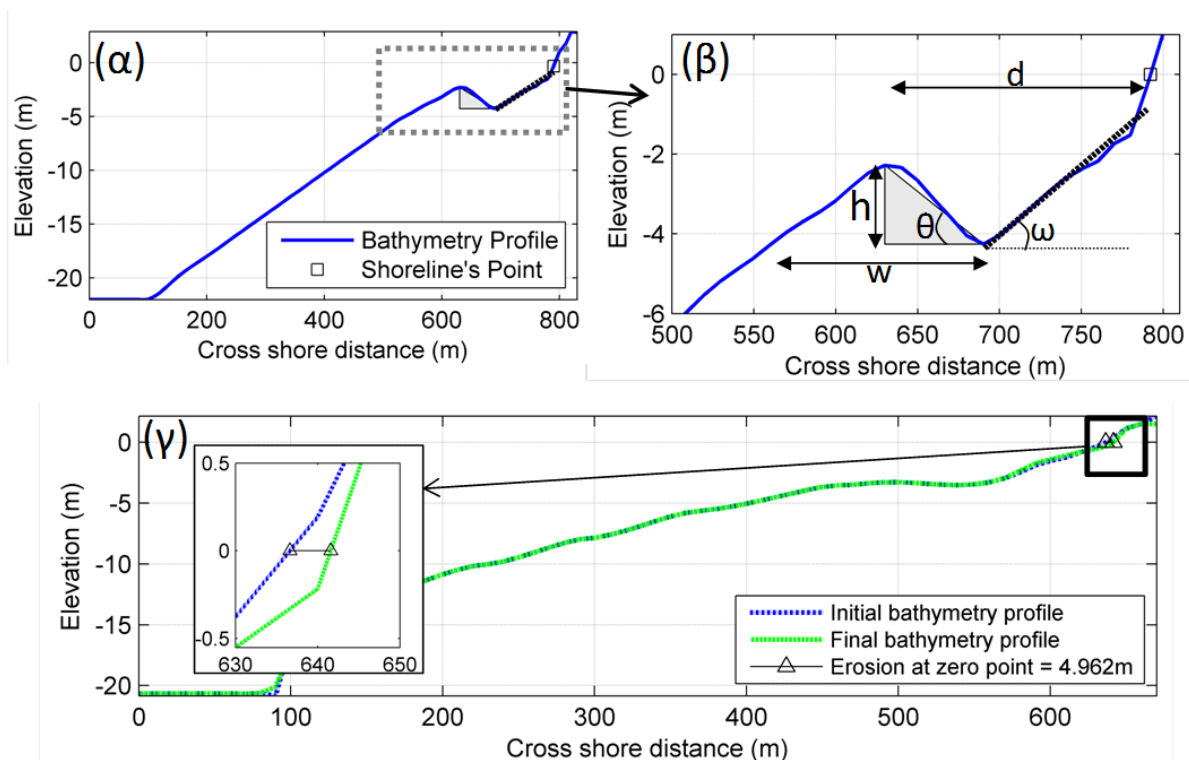
x_6 = η διάρκεια της καταιγίδας σε ώρες

x_7 = το σημαντικό ύψος κύματος

x_8 = η μέγιστη περίοδος κυμάτων που έχει προκύψει μέσω φασματικής ανάλυσης

x_9 = το μέγιστο πάχος ιζήματος της παραλίας

ως έξοδος του νευροασαφούς δικτύου χρησιμοποιήθηκε η διάβρωση της παραλίας σε μέτρα (Σχήμα 3.3 (γ)) η οποία προέκυψε από χρήση του μοντέλου XBeach [166].



Σχήμα 3.3 (α) Ένα βαθυμετρικό προφίλ όπου φαίνεται μια ιζηματική μπάρα στο σχήμα (β) Η ιζηματική μπάρα σε μεγέθυνση και πέντε δεδομένα τα οποία χρησιμοποιήθηκαν ως είσοδοι του δικτύου (γ) η διάβρωση της παραλίας που χρησιμοποιήθηκε ως έξοδος του δικτύου.

Στις προσομοιώσεις του προτεινόμενου νευροασαφούς συστήματος χρησιμοποιήθηκαν ως βαθμός πολωνύμου Hermite οι τιμές $n = 2,3,4,5$ (βλ. εξισώσεις (3.28)-(3.29)) και $c = 3,5,6,8,9,10$ ως πλήθος κρυφών κόμβων του δικτύου (βλ. εξίσωση (3.26)). Το 60% των δεδομένων χρησιμοποιήθηκε ως δεδομένα εκπαίδευσης και τα υπόλοιπα 40% ως δεδομένα ελέγχου και πραγματοποιήθηκαν δέκα αρχικοποιήσεις. Τα αποτελέσματα παρουσιάζονται στους πίνακες 3.7-3.9 ως σφάλματα $RMSE$ στη μορφή $\bar{x} \pm s$. Για σκοπούς σύγκρισης δημιουργήθηκε το νευροασαφές δίκτυο Takagi-Sugeno-Kang και το ακτινικής βάσης ΤΝΔ Conditional Fuzzy C Means NN όπως και στην προηγούμενη παράγραφο. Αυτά θα

συμβολίζονται ως TSKFS και RBFNN αντίστοιχα και τα αποτελέσματά τους φαίνονται στους πίνακες 3.7-3.9.

Πίνακας 3.7 Αποτελέσματα RMSE του προτεινόμενου νευροασαφούς δικτύου στο σύνολο δεδομένων Coastal Erosion για βαθμούς πολυωνύμων Hermite $n = 2$ και $n = 3$.

	Προτεινόμενο, $n = 2$		Προτεινόμενο, $n = 3$	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
$c = 3$	2.3642 ± 0.1141	2.5390 ± 0.1222	2.4055 ± 0.1498	2.6464 ± 0.2272
$c = 5$	2.3569 ± 0.0952	2.5450 ± 0.1473	2.2110 ± 0.0771	2.2937 ± 0.1196
$c = 6$	2.2560 ± 0.0614	2.3940 ± 0.1049	2.1906 ± 0.0948	2.2595 ± 0.1450
$c = 8$	2.2088 ± 0.0380	2.3886 ± 0.0634	2.2188 ± 0.0516	2.3490 ± 0.0779
$c = 9$	2.1992 ± 0.1241	2.3876 ± 0.1986	2.1676 ± 0.0602	2.2895 ± 0.0774
$c = 10$	2.1999 ± 0.0406	2.3633 ± 0.0606	2.1681 ± 0.0705	2.2812 ± 0.1300

Πίνακας 3.8 Αποτελέσματα RMSE του προτεινόμενου νευροασαφούς δικτύου στο σύνολο δεδομένων Coastal Erosion για βαθμούς πολυωνύμων Hermite $n = 4$ και $n = 5$.

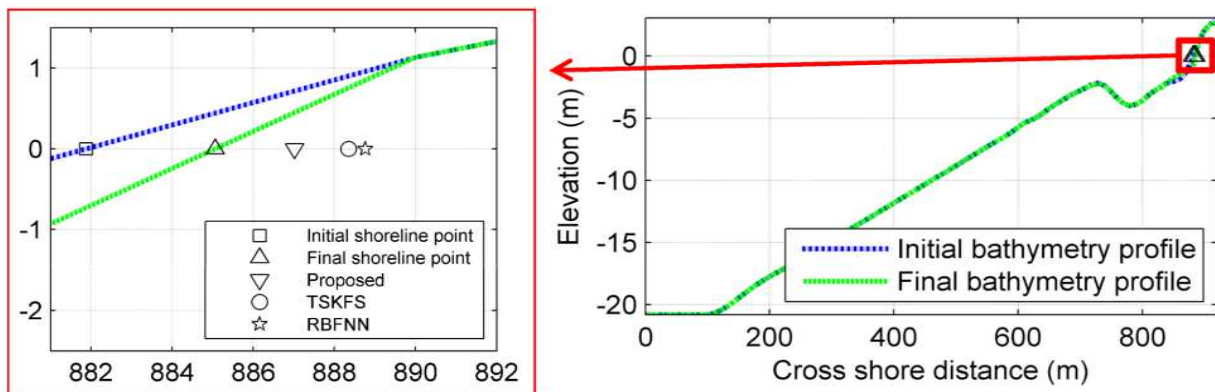
	Προτεινόμενο, $n = 4$		Προτεινόμενο, $n = 5$	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
$c = 3$	2.5407 ± 0.2871	2.8181 ± 0.4150	2.4977 ± 0.2786	2.6769 ± 0.5780
$c = 5$	2.3784 ± 0.1141	2.5045 ± 0.1638	2.4321 ± 0.3322	2.6002 ± 0.4865
$c = 6$	2.4066 ± 0.0755	2.5631 ± 0.1721	2.4623 ± 0.3098	2.6967 ± 0.3808
$c = 8$	2.3924 ± 0.1704	2.6002 ± 0.2645	2.4619 ± 0.2987	2.5734 ± 0.4234
$c = 9$	2.3379 ± 0.1781	2.5132 ± 0.2734	2.5218 ± 0.2865	2.5552 ± 0.3234
$c = 10$	2.3429 ± 0.1362	2.5372 ± 0.2066	2.5955 ± 0.4252	2.5624 ± 0.5066

Πίνακας 3.9 Αποτελέσματα RMSE των δύο δικτύων που δημιουργήθηκαν για σκοπούς σύγκρισης.

	TSKFS		RBFNN	
	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
$c = 3$	3.2111 ± 0.0197	3.7829 ± 0.0255	3.2670 ± 0.0211	3.8192 ± 0.0290
$c = 5$	3.1011 ± 0.0153	3.4837 ± 0.0155	3.1877 ± 0.0108	3.6159 ± 0.0250
$c = 6$	3.0307 ± 0.0064	3.2549 ± 0.0095	3.1389 ± 0.0093	3.5862 ± 0.0240
$c = 8$	3.0584 ± 0.0121	3.1859 ± 0.0295	3.1249 ± 0.0203	3.4186 ± 0.0164
$c = 9$	2.9023 ± 0.0064	3.1030 ± 0.0184	3.1154 ± 0.0184	3.3878 ± 0.0186
$c = 10$	2.9546 ± 0.0129	3.1200 ± 0.0284	3.1057 ± 0.0175	3.3083 ± 0.0174

Από τους προηγούμενους πίνακες φαίνεται ότι το προτεινόμενο νευροασαφές σύστημα δίνει καλύτερα αποτελέσματα σε σχέση με τα δύο με τα οποία έγινε η σύγκριση. Το καλύτερο σφάλμα το δίνει στα δεδομένα εκπαίδευσης το προτεινόμενο για $c = 9$ και $n = 3$ ενώ στα δεδομένα ελέγχου για $c = 6$ και $n = 3$. Μπορούμε να δούμε επίσης ότι για τους βαθμούς πολυωνύμων $n = 2$ και $n = 3$ το προτεινόμενο δίκτυο μοντελοποιεί σε ικανοποιητικό βαθμό τα δεδομένα Coastal Erosion

Για να φανεί η φυσική ερμηνεία των αποτελεσμάτων αυτών στα δεδομένα του προβλήματος Coastal Erosion, δίνεται το Σχήμα 3.4. Το σχήμα αυτό σχεδιάστηκε χρησιμοποιώντας τα σφάλματα των παραπάνω δικτύων για $c = 8$ και για το προτεινόμενο επιπλέον $n = 3$ ως βαθμός πολυωνύμων Hermite. Αυτό παρουσιάζει δεξιά τα δύο βαθυμετρικά προφίλ, ένα που προκύπτει από βαθυμετρικές μετρήσεις (Initial Bathymetry Profile στο Σχήμα 3.4) και το αποτέλεσμα του όταν σε αυτό εφαρμόστηκε το μοντέλο XBeach [166] (Final Bathymetry Profile στο Σχήμα 3.4). Στο ίδιο σχήμα στα αριστερά φαίνονται εκτός από τα δύο προφίλ, οι προβλέψεις που δίνουν τα τρία διαφορετικά δίκτυα για την μετατόπιση της θέσης της ακτογραμμής και από το σχήμα φαίνεται ότι το προτεινόμενο δίκτυο δίνει αποτελέσματα πιο κοντά σε αυτά του επιθυμητού αποτελέσματος. Τα σφάλματα RMSE που έχουν σχεδιαστεί στο σχήμα αυτό είναι 2.3867m για το προτεινόμενο δίκτυο, 3.1678 για το TSKFS, και 3.4007 για το RBFNN.



Σχήμα 3.4 Η φυσική ερμηνεία των σφαλμάτων *RMSE* στο πρόβλημα Coastal Erosion

4. Ανάπτυξη Πολυωνυμικού Νευρωνικού Δικτύου Με Χρήση Πολυωνύμων Legendre

Στο παρόν κεφάλαιο γίνεται χρήση των ορθογωνίων πολυωνύμων Legendre. Σκοπός χρήσης των πολυωνύμων αυτών είναι η χρήση τους μέσα σε ΤΝΔ και η εφαρμογή τους σε μη-γραμμικά σύνολα δεδομένων με σκοπό να αυξηθεί η απόδοση των ΤΝΔ. Επιλέγησαν τα πολυώνυμα Legendre επειδή (i) είναι ορθογώνια, (ii) προκύπτουν σε πολυάριθμα προβλήματα, ιδίως σε αυτά που αφορούν σφαιρικές συντεταγμένες ή παρουσιάζουν σφαιρική συμμετρία και (iii) επειδή σε σφαιρικές πολικές συντεταγμένες, η γωνιακή εξάρτηση αντιμετωπίζεται καλύτερα με τη χρήση συναρτήσεων Legendre [167]. Στο παρόν κεφάλαιο δίνεται η ανάλυση του ΤΝΔ με τοπολογία αυτή που δημοσιεύτηκε στο άρθρο [168]. Αυτό είναι ένα πολυωνυμικό ΤΝΔ στο οποίο για να λειτουργεί η ορθογωνιότητα των πολυωνύμων που χρησιμοποιούνται, είναι απαραίτητη η αλλαγή διαστήματος των δεδομένων (scaling). Το προτεινόμενο ΤΝΔ εκπαιδεύεται με χρήση αλγορίθμων νοημοσύνης σμήνους, συγκεκριμένα με τον αλγόριθμο GABC (βλ. ενότητα 1.1.3.2) ώστε να μειωθεί το ολικό τετραγωνικό σφάλμα του δικτύου. Οι προσομοιώσεις του δικτύου έγιναν πάνω σε σύνολα του UCI repository και σε ένα πρωτότυπο σύνολο δεδομένων το οποίο έχει ως στόχο να προβλέψει την περιστροφή μιας παραλίας. Οι προσομοιώσεις κατέληξαν σε ικανοποιητικά αποτελέσματα. Η δομή των ακολούθων παραγράφων είναι η εξής: στην πρώτη παράγραφο δίνεται μια βιβλιογραφική ανασκόπηση της χρήσης των πολυωνύμων Legendre μέσα σε ΤΝΔ, στην δεύτερη δίνονται ορισμοί και ιδιότητες των πολυωνύμων αυτών, ακολουθεί η παράγραφος με τη δομή του προτεινόμενου ΤΝΔ και η μέθοδος εκπαίδευσής του και τέλος ακολουθούν οι παράγραφοι των αποτελεσμάτων (αριθμητικές προσομοιώσεις).

4.1 Συναφείς μέθοδοι και η προτεινόμενη μέθοδος

Τα πολυώνυμα Legendre έχουν πολλές εφαρμογές στη βιβλιογραφία, ακολουθούν μερικά παραδείγματα χρήσης τους σε ΤΝΔ. Οι Patra et al. [26] δημιούργησαν ένα ΤΝΔ με χρήση πολυωνύμων Legendre, το LeNN, το οποίο χρησιμοποιήθηκε σε εξισορρόπηση (equalization) σε ασύρματα συστήματα επικοινωνίας, Το LeNN χρησιμοποιήθηκε επειδή χρειάζεται μικρότερο χρόνο υπολογισμών (computational costs) σε σχέση με Functional Link

ΤΝΔ (FLANN) τα οποία χρησιμοποιούν τριγωνομετρικές εκφράσεις των δεδομένων εισόδου. Για να λειτουργήσει το LeNN εφάρμοσαν απευθείας στα δεδομένα εισόδου του ΤΝΔ ως συναρτήσεις ενεργοποίησης τα πολώνυμα Legendre, δηλαδή αν π.χ. ένα δεδομένο εισόδου είναι δυσδιάστατο κι έχει τη μορφή $\mathbf{x} = [x_1, x_2]^T$ τότε αυτό εκφράστηκε μέσω των συναρτήσεων Legendre ως $\mathbf{x}^e = [1, P_1(x_1), P_2(x_1), P_3(x_1), P_1(x_2), P_2(x_2), P_3(x_2)]^T$ δηλαδή με τον τρόπο που εφαρμόζονται στα FLANN. Κατόπιν αυτά αθροίζονται μέσω των συναπτικών βαρών $\mathbf{w} = [w_1, w_2, \dots, w_i]^T, i = 1, 2, \dots, \dim(\mathbf{x}^e)$ και μέσω της συνάρτησης ενεργοποίησης

$\tanh(\cdot)$ δίνουν ως έξοδο του LeNN την: $\hat{y} = \tanh\left(\sum_{i=1}^{\dim(\mathbf{x}^e)} \mathbf{x}_i^e w_i\right)$ όπου \mathbf{x}_i^e η i -στή συνιστώσα

του \mathbf{x}^e . Στις αριθμητικές προσομοιώσεις οι συγγραφείς σύγκριναν το LeNN με ένα πολυστρωματικό ΤΝΔ (multilayer perceptron ή MLP) και ένα FLANN που χρησιμοποιεί τριγωνομετρικές εκφράσεις των δεδομένων εισόδου (Σχήμα 1.15) και κατέληξαν ότι το LeNN χρειάζεται λιγότερο χρόνο εκπαίδευσης και αποδίδει μικρότερα σφάλματα MSE.

Το ίδιο δίκτυο (LeNN) το χρησιμοποίησαν οι Patra και Bornard [27] για αναγνώριση δυναμικών μη-γραμμικών συστημάτων [169, 170]. Στο άρθρο αυτό χρησιμοποιούν το LeNN όπως ακριβώς και οι Patra et al. [26] και καταλήγουν σε καλύτερες αριθμητικές προσεγγίσεις σε τέσσερα δυναμικά συστήματα συγκρινόμενοι με MLP δίκτυα.

Οι Liu και Wang [29] δημιούργησαν ένα τρι-στρωματικό ΤΝΔ και το χρησιμοποίησαν στην πρόβλεψη των τιμών μετοχών του χρηματιστηρίου. Το πρώτο στρώμα του δικτύου αποτελείται από το στρώμα των δεδομένων εισόδου $\mathbf{x} = [x_1, x_2, \dots, x_p]^T$ τα οποία καταλήγουν μέσω των συναπτικών βαρών $w_{ij}, i = 1, \dots, p, j = 1, \dots, q$ στο στρώμα με τους κρυφούς κόμβους του δικτύου όπου περνάν μέσα από q στο πλήθος πολώνυμα Legendre. Η έξοδος j -στου

κρυφού κόμβου είναι τότε η $P_j\left(\sum_{i=1}^p w_{ij} x_i + \theta_j\right)$ όπου $P_j(x)$ το πολώνυμο Legendre j -στου

βαθμού. Για να υπολογιστεί η έξοδος του δικτύου τα δεδομένα του κρυφού κόμβου αθροίζονται μέσω των συναπτικών βαρών $c_j, j = 1, 2, \dots, q$ και τότε προκύπτει η έξοδος του

δικτύου: $\hat{y} = \sum_{j=0}^q c_j P_j\left(\sum_{i=1}^p w_{ij} x_i + \theta_j\right)$. Ως μέθοδος εκπαίδευσης του δικτύου τους

χρησιμοποιείται η μέθοδος Gradient Descent για να ελαχιστοποιηθεί το σφάλμα

$E(t) = \frac{1}{2} \varphi(t) (y(t) - \hat{y}(t))^2$ όπου $\varphi(t)$ μια συνάρτηση που παράγει τυχαίους αριθμούς για αντοχή στο χρόνο (της κάθε επανάληψης) (random time strength function). Η συνάρτηση $\varphi(t)$ συνηθίζεται να χρησιμοποιείται σε οικονομικά μοντέλα προγνώσεων [171]. Πρέπει να σημειώσουμε ότι στα δεδομένα εισόδου του ΤΝΔ εφάρμοσαν την αλλαγή διαστήματος :

$$s'(t) = \frac{s(t) - \min(s(t))}{\max(s(t)) - \min(s(t))}$$
 που μετατρέπει τα δεδομένα εισόδου σε δεδομένα με τιμές εντός του διαστήματος $[0, 1]$.

Το LeNN το χρησιμοποίησαν οι Nanda et al. [28] για να προβλέψουν τη ποιότητα του αέρα ως συνάρτηση μετεωρολογικών δεδομένων και της παρουσίας ρύπων στην ατμόσφαιρα.

Ο Lin [172] χρησιμοποίησε ένα ΤΝΔ με πολυώνυμα Legendre για να προσομοιώσει την κίνηση του σερβοκινητήρα (servo*) από ένα ηλεκτρικό μοτοποδήλατο (scooter).

Σε όλες τις παραπάνω αναφορές τα δεδομένα εισόδου εισέρχονται απευθείας μέσα από τους κόμβους του δικτύου που περιέχουν τις συναρτήσεις των πολυωνύμων Legendre εκτός από την [29] όπου οι συγγραφείς έχουν παρατηρήσει τα δεδομένα εισόδου του ΤΝΔ έχουν τιμές στο διάστημα $[0, 1]$ και δεν ισχύει όμως το ίδιο για τα δεδομένα $\sum_{i=1}^p w_{ij} x_i + \theta_j$ που χρησιμοποιούνται ως μεταβλητή εισόδου στα πολυώνυμα Legendre. Τα παραπάνω έχουν το εξής πρόβλημα: τα πολυώνυμα Legendre είναι ορθογώνια μόνο στο διάστημα $[-1, 1]$ (βλ. εξίσωση (4.4)) και ως εκ τούτου πρέπει να γίνει μετασχηματισμός στα δεδομένα που εισέρχονται στα πολυώνυμα αυτά ή θα πρέπει τα δεδομένα που εισέρχονται μέσα στα πολυώνυμα να μην είναι άμεση έκφραση (direct expansion) των δεδομένων εισόδου του ΤΝΔ. Το πρόβλημα αυτό έχει διορθωθεί με το προτεινόμενο ΤΝΔ της παρούσας ενότητας.

* Οι κινητήρες servo είναι μικρές συσκευές που έχουν έναν εξωτερικό άξονα. Αυτός ο άξονας μπορεί να μετακινηθεί σε διάφορες θέσεις αν αποσταλεί στον servo ένα κωδικοποιημένο σήμα. Όσο υπάρχει αυτό το σήμα στην γραμμή εισόδου του Servo τόσο αυτός θα διατηρεί τον άξονα του σε μια συγκεκριμένη θέση. Όταν αλλάζει το σήμα προκαλεί στον servo να μεταβάλλει την γωνία του άξονα. Οι servos χρησιμοποιούνται σε τηλεχειριζόμενα αεροπλάνα, αυτοκίνητα και στην ρομποτική.

4.2 Πολυώνυμα Legendre

Τα πολυώνυμα Legendre έχουν προκύψει από την προσπάθεια εύρεσης πολυωνύμων που να ικανοποιούν την διαφορική εξίσωση του Legendre [149, 167]:

$$\frac{d}{dx} \left[(1-x^2) \frac{d}{dx} P_n(x) \right] + n(n+1)P_n(x) = 0. \quad (4.1)$$

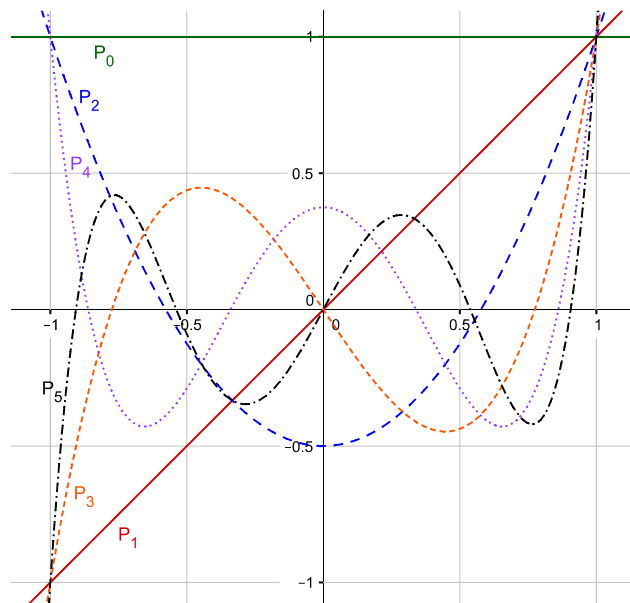
όπου $n = 0, 1, 2, \dots$ και $P_n(x)$ ένα n -στού βαθμού πολυώνυμο. Οι λύσεις της παραπάνω διαφορικής εξίσωσης είναι τα πολυώνυμα που ορίζονται ως :

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} [(x^2 - 1)^n], \quad n = 0, 1, \dots \quad (4.2)$$

από την οποία για $n = 0, 1, 2, 3, 4, 5$ προκύπτουν τα 6 πρώτα πολυώνυμα:

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) \\ P_3(x) &= \frac{1}{2}(5x^3 - 3x) \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3) \\ P_5(x) &= \frac{1}{8}(63x^5 - 70x^3 + 15x) \end{aligned} \quad (4.3)$$

με γραφικές παραστάσεις για $x \in [-1, 1]$ αυτές του Σχήματος 4.1.



Σχήμα 4.1 Τα πολυώνυμα Legendre $P_n(x)$ βαθμών $n = 0$ ως και $n = 5$ στο διάστημα $[-1, 1]$.

Τα πολυώνυμα αυτά είναι ορθογώνια στο διάστημα $[-1,1]$ ικανοποιώντας τη σχέση εσωτερικού γινομένου:

$$\langle P_m(x), P_n(x) \rangle = \int_{-1}^1 P_m(x)P_n(x)dx = \frac{2}{2n+1} \delta_{mn} = \begin{cases} \frac{2}{2n+1}, & m = n \\ 0, & m \neq n \end{cases} \quad (4.4)$$

όπου το δ_{mn} είναι το δέλτα του Kronecker το οποίο ορίζεται ως:

$$\delta_{mn} = \begin{cases} 1, & m \neq n \\ 0, & m = n \end{cases} \quad (4.5)$$

Επιπρόσθετα, για προγραμματιστικούς σκοπούς μπορούν να υπολογιστούν από την αναδρομική σχέση που ονομάζεται τύπος του Bonnet:

$$\left\{ \begin{array}{l} P_0(x) = 1, \quad P_1(x) = x \\ P_n(x) = \frac{1}{n} [(2n-1)xP_{n-1}(x) - (n-1)P_{n-2}(x)], \quad n \geq 2 \end{array} \right\} \quad (4.6)$$

4.3 Το Πολυωνυμικό δίκτυο με χρήση πολυωνύμων Legendre

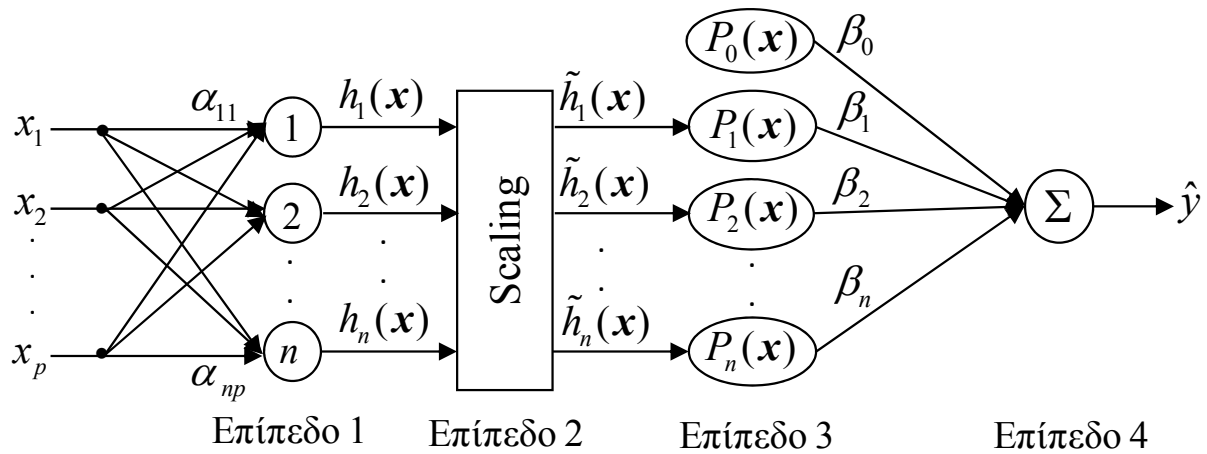
Υποθέτουμε ότι τα δεδομένα που έχουμε είναι της μορφής:

$$S = \{(\mathbf{x}_k, y_k) : \mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kp}]^T, y_k \in \mathbb{R}, k = 1, 2, \dots, N\} \quad (4.7)$$

όπως και στις προηγούμενες παραγράφους, με p τη διάσταση του χώρου εισόδου και N το πλήθος των δεδομένων εκπαίδευσης του ΤΝΔ. Το ΤΝΔ που δημιουργήθηκε και χρησιμοποιεί πολυώνυμα Legendre είναι αυτό που παρουσιάζεται στο Σχήμα 4.2 και αποτελείται από 4 επίπεδα (layers).

Πρέπει να σημειωθεί ότι τα προτεινόμενο ΤΝΔ δεν χρησιμοποιεί τα πολυώνυμα Legendre απευθείας στα δεδομένα εισόδου (direct expansion) όπως γίνεται σε αυτά που παρουσιάστηκαν στη βιβλιογραφία (βλ. παράγραφο 4.1). Λόγω της λειτουργίας των

Legendre πολυωνύμων μόνο στο διάστημα $[-1,1]$ τα πολυώνυμα Legendre εφαρμόζονται σε γραμμικούς συνδυασμούς των δεδομένων εισόδου οι οποίοι έχουν τιμές στο $[-1,1]$. Ακολουθεί η ανάλυση λειτουργίας του προτεινόμενου ΤΝΔ.



Σχήμα 4.2 Η τοπολογία του Νευρωνικού δικτύου Legendre.

Αν συμβολιστεί με n ο μέγιστος βαθμός πολυωνύμων Legendre που θα χρησιμοποιηθεί στο παρόν ΤΝΔ, τότε στο 1^ο επίπεδο αυτού παράγει το γραμμικό συνδυασμό των δεδομένων εισόδου :

$$h_\ell(\mathbf{x}) = \sum_{j=1}^p \alpha_{\ell j} x_j, \quad \ell = 1, 2, \dots, n, \quad j = 1, 2, \dots, p \quad (4.8)$$

όπου $\alpha_{\ell j}$ συντελεστές βαρύτητας που συνδέουν τα δεδομένα εισόδου με το πρώτο στρώμα του δικτύου. Στο 2^ο επίπεδο του ΤΝΔ στα δεδομένα $h_\ell(\mathbf{x})$ που προέκυψαν από το πρώτο επίπεδο εφαρμόζεται αλλαγή διαστήματος (scaling) ώστε να προκύψουν δεδομένα $\tilde{h}_\ell(\mathbf{x})$ τα οποία να ανήκουν στο διάστημα $[-1,1]$ επειδή -όπως προαναφέρθηκε- τα πολυώνυμα Legendre είναι ορθογώνια μόνο στο διάστημα αυτό. Για να γίνει αυτό ακολουθήθηκε η παρακάτω διαδικασία, η οποία διατηρεί τη γραμμικότητα της συνάρτησης $h_\ell(\mathbf{x})$.

Έστω ότι το j -στό δεδομένο εισόδου $x_j, j = 1, 2, \dots, p$ παίρνει τιμές στο πεδίο ορισμού:

$D_j = [x_j^L, x_j^U]$ δηλαδή $x_j^L \leq x_j \leq x_j^U$, το οποίο διάστημα πρέπει να σημειωθεί ότι είναι σταθερό για ένα σύνολο δεδομένων S της σχέσης (4.7) και υπολογίζεται απλά ως:

$D_j = \left[\min_{1 \leq k \leq N} x_{kj}, \max_{1 \leq k \leq N} x_{kj} \right]$. Ομοίως και οι συντελεστές βαρύτητας $\alpha_{\ell j}$ μπορούν να περιοριστούν στο διάστημα $A = [\alpha_L, \alpha_U]$, δηλαδή $\alpha_L \leq \alpha_{\ell j} \leq \alpha_U$. Το διάστημα A παραμένει σταθερό σε όλη τη διαδικασία εκπαίδευσης του ΤΝΔ και επιλέγεται με δοκιμές ώστε το ΤΝΔ που προκύπτει να έχει ικανοποιητική απόδοση.

Προκειμένου στα δεδομένα $h_\ell(\mathbf{x})$ να εφαρμοστεί αλλαγή διαστήματος ώστε να ανήκουν στο $[-1,1]$ αρκεί να υπολογιστεί ότι οι γραμμικοί συνδυασμοί των $\alpha_{\ell j} \in [\alpha_L, \alpha_U]$ με τα $x_j \in [x_j^L, x_j^U]$ βρίσκονται στο διάστημα: $[L_j, U_j]$, όπου :

$$\begin{aligned} L_j &= \min \{ a_L x_j^L, a_L x_j^U, a_U x_j^L, a_U x_j^U \} \\ U_j &= \max \{ a_L x_j^L, a_L x_j^U, a_U x_j^L, a_U x_j^U \} \end{aligned} \quad (4.9)$$

δηλαδή: $L_j \leq a_{\ell j} x_j \leq U_j, \forall \ell = 1, 2, \dots, n, \forall j = 1, 2, \dots, p$ και άρα

$$\sum_{j=1}^p L_j \leq \sum_{j=1}^p a_{\ell j} x_j \leq \sum_{j=1}^p U_j \Rightarrow Q_L \leq h_\ell(\mathbf{x}) \leq Q_U \quad (4.10)$$

όπου $Q_L = \sum_{j=1}^p L_j$ και $Q_U = \sum_{j=1}^p U_j$. Συνεπώς καταλήγουμε ότι τα $h_\ell(\mathbf{x})$ της εξίσωσης (4.8)

βρίσκονται στο διάστημα $[Q_L, Q_U]$. Αυτή τη στιγμή το μοναδικό που χρειάζεται ώστε να αλλάξουν τα δεδομένα $h_\ell(\mathbf{x})$ τιμές: $[Q_L, Q_U] \rightarrow [-1,1]$ είναι ο ακόλουθος μετασχηματισμός αλλαγής διαστήματος (scaling):

$$\tilde{h}_\ell(\mathbf{x}) = \frac{2}{Q_U - Q_L} h_\ell(\mathbf{x}) - \frac{Q_U + Q_L}{Q_U - Q_L} \quad (4.11)$$

με χρήση του οποίου προκύπτουν δεδομένα $\tilde{h}_\ell(\mathbf{x}) \in [-1,1]$. Για να αποδειχθεί ότι η τελευταία εξίσωση είναι γραμμική ως προς τα $\alpha_{\ell j}$ και τα x_j αρκεί να θέσουμε ως :

$R = \frac{2}{Q_U - Q_L}$ και $\Omega = \frac{Q_U + Q_L}{Q_U - Q_L}$ και τότε η εξίσωση (4.11) μέσω της (4.8) γράφεται ως

εξής:

$$\tilde{h}_\ell(\mathbf{x}) = R \sum_{j=1}^p a_{\ell j} x_j - \Omega = \sum_{j=1}^p R a_{\ell j} x_j - \Omega \quad (4.12)$$

Στο 3^ο στρώμα του ΤΝΔ υπάρχουν n το πλήθος κόμβοι οι οποίοι χρησιμοποιούν τα πολυώνυμα Legendre ως συναρτήσεις ενεργοποίησης του ΤΝΔ ως:

$$P_\ell(\mathbf{x}) = P_\ell(\tilde{h}_\ell(\mathbf{x})) = P_\ell\left(\sum_{j=1}^p R a_{\ell j} x_j - \Omega\right), \ell = 1, 2, \dots, n \quad (4.13)$$

καθώς και το πολυώνυμο $P_0(x) = 1$ το οποίο επειδή είναι σταθερό λειτουργεί ως bias του ΤΝΔ.

Τέλος, στο 4^ο επίπεδο υπολογίζεται η εκτιμώμενη έξοδος του ΤΝΔ ως :

$$\hat{y} = \beta_0 + \sum_{\ell=1}^n \beta_\ell P_\ell(\tilde{h}_\ell(\mathbf{x})) = \beta_0 + \sum_{\ell=1}^n \beta_\ell P_\ell\left(\sum_{j=1}^p R a_{\ell j} x_j - \Omega\right) \quad (4.14)$$

όπου $\beta_\ell, 0 \leq \ell \leq n$ συντελεστές βαρύτητας του ΤΝΔ.

Η εξίσωση (4.14) γράφεται και ως:

$$\hat{y} = \beta_0 + \beta_1 P_1(\tilde{h}_1(\mathbf{x})) + \beta_2 P_2(\tilde{h}_2(\mathbf{x})) + \dots + \beta_n P_n(\tilde{h}_n(\mathbf{x})) \quad (4.15)$$

η οποία πρέπει να σημειωθεί ότι δεν αποτελεί έκφραση ως σειρά πολυωνύμων Legendre αλλά αποτελεί προσέγγιση. Θα αποτελούσε σειρά πολυωνύμων Legendre μόνο στην περίπτωση που τα δεδομένα $\tilde{h}_\ell(\mathbf{x}), \ell = 1, 2, \dots, n$ ήταν όλα ίσα μεταξύ τους, το οποίο δεν ισχύει εδώ εφόσον συναπτικά βάρη $a_{\ell j}$ της εξίσωσης (4.12) δεν είναι μεταξύ τους ίσα .

4.4 Εκπαίδευση του δικτύου

Η εκπαίδευση του ΤΝΔ που παρουσιάστηκε στηρίζεται στην εύρεση των συναπτικών βαρών του δικτύου $a_{\ell j}$ και β_ℓ έτσι ώστε το ολικό τετραγωνικό σφάλμα του δικτύου:

$$J_{SE} = \sum_{k=1}^N |y_k - \hat{y}_k|^2 \text{ να ελαχιστοποιηθεί. Για να γίνει η ελαχιστοποίηση αυτή όμως ώστε οι}$$

ποσότητες $\tilde{h}_\ell(\mathbf{x})$ της εξίσωσης (4.12) να παραμείνουν στο διάστημα $[-1, 1]$, η εύρεση των $a_{\ell j}$ πρέπει να γίνει υπό τον περιορισμό (constrained optimization):

$$a_{\ell j} \in A \Leftrightarrow a_L \leq a_{\ell j} \leq a_U \quad \forall \ell, j \quad (4.16)$$

ενώ για τα βάρη β_ℓ δεν υπάρχει κάποιος περιορισμός (unconstrained optimization).

Στη δημοσίευση [168] η εκπαίδευση του ΤΝΔ έγινε με τη χρήση της μεθόδου Barrier [173], στο παρόν κείμενο όμως ακολουθείται μια διαφορετική διαδικασία. Προκειμένου να εκπαιδευτεί το παρόν ΤΝΔ χρησιμοποιήθηκε ο αλγόριθμος GABC [30] που είναι η G-best παραμετροποίηση του αλγόριθμου Artificial Bee Colony (ABC) ο οποίος αναλύθηκε στην παράγραφο 1.1.3.2. Στο συγκεκριμένο ΤΝΔ οι λύσεις \mathbf{v}_i των εξισώσεων (1.16) και (1.19) που πρέπει να προσδιοριστούν είναι της μορφής:

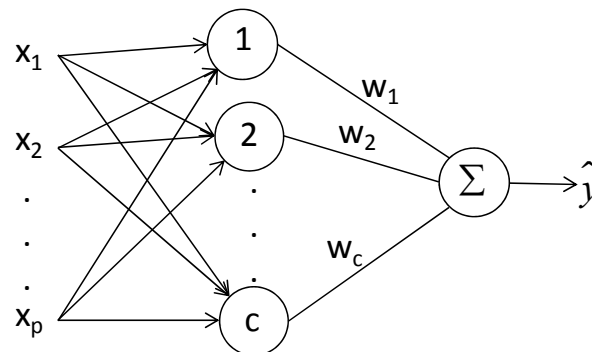
$$\mathbf{v} = [\alpha_{11}, \alpha_{12}, \dots, \alpha_{np}, \beta_0, \beta_1, \dots, \beta_n]^T \quad (4.17)$$

Ενώ στον αλγόριθμο GABC όπως περιγράφηκε στην παράγραφο 1.1.3.2, ελέγχεται οι λύσεις \mathbf{v} να ανήκουν σε ένα συγκεκριμένο χώρο λύσεων (αυτό γίνεται στη φάση των απασχολούμενων-μελισσών του αλγορίθμου GABC), στο συγκεκριμένο ΤΝΔ ο αλγόριθμος GABC μετατράπηκε ώστε να ελέγχει μόνο για τις πρώτες $n \cdot p$ το πλήθος συνιστώσες (τα συναπτικά βάρη a_{ij}) να ανήκουν εντός του χώρου $A = [a_L, a_U]$ ενώ για τις τελευταίες $n+1$ συνιστώσες δεν εφαρμόστηκε κανένας έλεγχος κατά τη διάρκεια των επαναλήψεων του GABC.

4.5 Αριθμητικές προσομοιώσεις σε σύνολα δεδομένων του UCI Repository

Προκειμένου να φανεί η απόδοση του προτεινόμενου Νευρωνικό δίκτυο σε σχέση με άλλα της βιβλιογραφίας, δύο ακόμη νευρωνικά δίκτυα δημιουργήθηκαν και χρησιμοποιήθηκαν για σκοπούς σύγκρισης.

(α) το ΤΝΔ δίκτυο Conditional Fuzzy C Means RBF που προτείνεται από τον Pedrycz στο [104], και περιγράφηκε στην παράγραφο 1.3.1.1 (Σχήμα 4.3) . Το ΤΝΔ αυτό είναι ακτινικής βάσης και στηρίζεται στη συσταδοποίηση του χώρου εισόδου-εξόδου. Για την εύρεση των παραμέτρων του ΤΝΔ (w_j στο σχήμα της τοπολογίας του ΤΝΔ) χρησιμοποιήθηκε η μέθοδος των ελαχίστων τετραγώνων. Το παρόν ΤΝΔ θα αναφέρεται απλά ως RBF.



Σχήμα 4.3 Το ΤΝΔ ακτινικής βάσης Conditional Fuzzy C Means με το οποίο έγινε σύγκριση με το προτεινόμενο ΤΝΔ Legendre

(β) Ένα εμπροσθοτροφοδοτούμενο νευρωνικό δίκτυο το οποίο χρησιμοποιεί ως συνάρτηση ενεργοποίησης την $f(x) = \tanh \frac{x}{2}$ και έχει $(p+1) \cdot c$ στο πλήθος παραμέτρους, με τοπολογία αυτή που παρουσιάζεται στο Σχήμα 1.13. Αυτό το ΤΝΔ αυτό θα αναφέρεται ως FFNN. Για την εκπαίδευσή του FFNN, δηλαδή για την εύρεση των συναπτικών βαρών $a_{ij}, w_j, 1 \leq i \leq p, 1 \leq j \leq c$ (βλ. Σχήμα 1.13) χρησιμοποιήθηκε η μέθοδος PSO (βλ. [38] και παράγραφο 1.1.3.1) για να προσδιορισθούν λύσεις της μορφής:

$$\mathbf{x} = [a_{11}, a_{12}, \dots, a_{pc}, \beta_1, \dots, \beta_c]^T \quad (4.18)$$

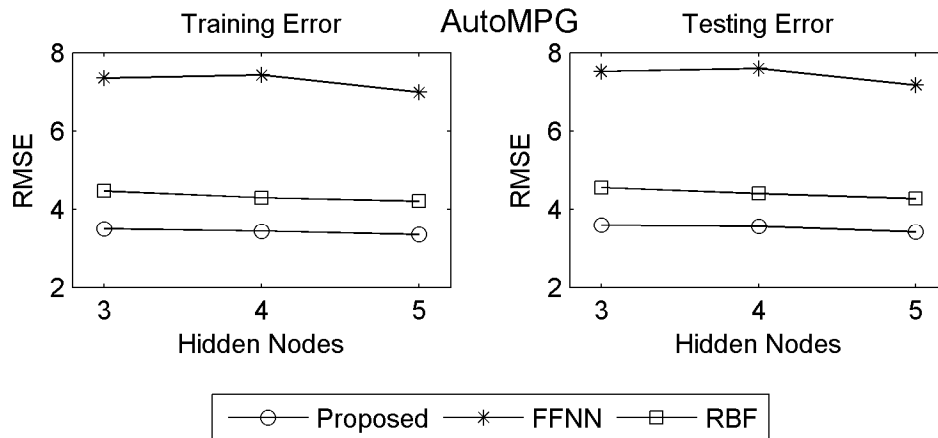
έτσι ώστε να ελαχιστοποιείται το σφάλμα RMSE του δικτύου.

Σε όλες τις προσομοιώσεις έγιναν έλεγχοι για 3, 4 και 5 κρυφούς κόμβους (πλήθος πολυωνύμων n στο προτεινόμενο και κρυφών κόμβων c στα δύο ΤΝΔ με τα οποία γίνεται σύγκριση). Για το προτεινόμενο ΤΝΔ τα διαστήματα που επιλέχθηκαν τα αρχικά συναπτικά βάρη α_{ij} και $\beta_\ell, 0 \leq \ell \leq n$ των εξισώσεων (4.8) και (4.14) αντίστοιχα του δικτύου ήταν το $[-10,10]$ για την εκπαίδευσή του χρησιμοποιήθηκε στον αλγόριθμο GABC ως μέγεθος αποικίας μελισσών $CS = 50$, ως μέγιστος αριθμός επαναλήψεων ο αριθμός 350 και το ψ_{ij} της εξίσωσης (1.19) ως τυχαίος αριθμός στο διάστημα $[0,1]$. Για το FFNN τα αρχικά συναπτικά βάρη α_{ij} και $w_j, 1 \leq i \leq p, 1 \leq j \leq c$ επιλέχθηκαν μέσα στο διάστημα $[-10,10]$, το πλήθος σμήνους του αλγόριθμου PSO ήταν 100, οι παράμετροι $\omega, \varphi_p, \varphi_g$ της εξίσωσης (1.10) είχαν τις τιμές: $\omega = 2, c_1 = 2, c_2 = 2$ και το πλήθος επαναλήψεων του αλγορίθμου επιλέχθηκε να είναι το 100. Για το RBF η μόνη παράμετρος που χρειάστηκε να οριστεί είναι η ασαφής παράμετρος $m = 2$ (βλ. εξισώσεις (1.55) - (1.57)).

Σε όλες τις προσομοιώσεις χρησιμοποιήθηκε η μέθοδος επαλήθευσης 10-fold cross validation με 10 αρχικοποιήσεις (initializations). Τα αποτελέσματα των ΤΝΔ παρουσιάζονται στους πίνακες 4.1-4.6 όπου παρουσιάζεται το σφάλμα $RMSE$ στη μορφή: (μέση τιμή) \pm (τυπική απόκλιση). Μετά από κάθε πίνακα ακολουθεί σχήμα με τις μέσες τιμές των σφαλμάτων.

Πίνακας 4.1 Τα αριθμητικά αποτελέσματα του ΤΝΔ της ενότητας 4.2 καθώς και δύο ακόμη ΤΝΔ με τα οποία γίνεται σύγκριση για το σύνολο δεδομένων **AutoMPG** του UCI Repository. Τα αποτελέσματα δίνονται στη μορφή: (Μέση τιμή) \pm (τυπική απόκλιση)

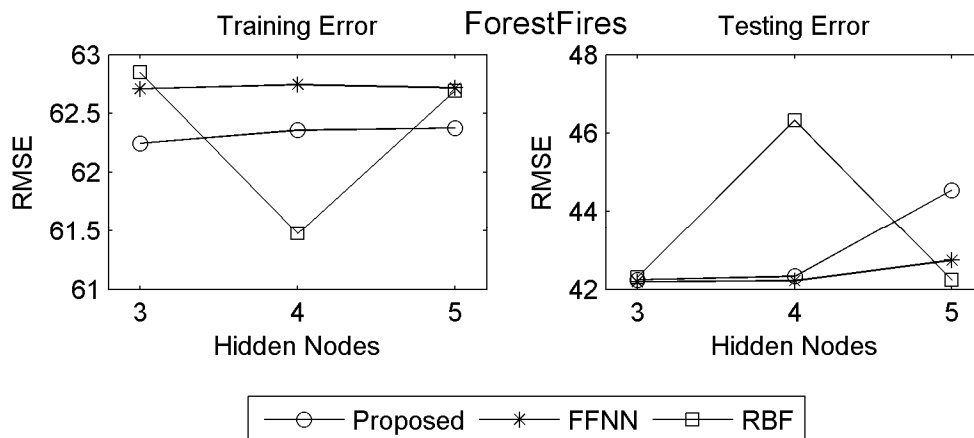
	Κρυφοί κόμβοι	Σφάλμα RMSE στα δεδομένα εκπαίδευσης (training)	Σφάλμα RMSE στα δεδομένα έλεγχου (testing)
Προτεινόμενο	3	3.4933 \pm 0.2780	3.5932 \pm 1.1948
	4	3.4322 \pm 0.2089	3.5550 \pm 1.1692
	5	3.3463 \pm 0.1816	3.4095 \pm 1.1993
RBF	3	4.4711 \pm 0.1799	4.5663 \pm 1.5186
	4	4.2977 \pm 0.1696	4.4041 \pm 1.4294
	5	4.2115 \pm 0.1773	4.2756 \pm 1.4990
FFNN	3	7.3538 \pm 0.9360	7.5211 \pm 2.1169
	4	7.4307 \pm 0.8863	7.5942 \pm 1.9997
	5	6.9904 \pm 1.1605	7.1742 \pm 2.0410



Σχήμα 4.4 Οι μέσες τιμές των αποτελεσμάτων του προηγούμενου πίνακα.

Πίνακας 4.2 Τα αριθμητικά αποτελέσματα του ΤΝΔ της ενότητας 4.2 καθώς και δύο ακόμη ΤΝΔ με τα οποία γίνεται σύγκριση για το σύνολο δεδομένων ForestFires του UCI Repository. Τα αποτελέσματα δίνονται στη μορφή: (Μέση τιμή) ± (τυπική απόκλιση)

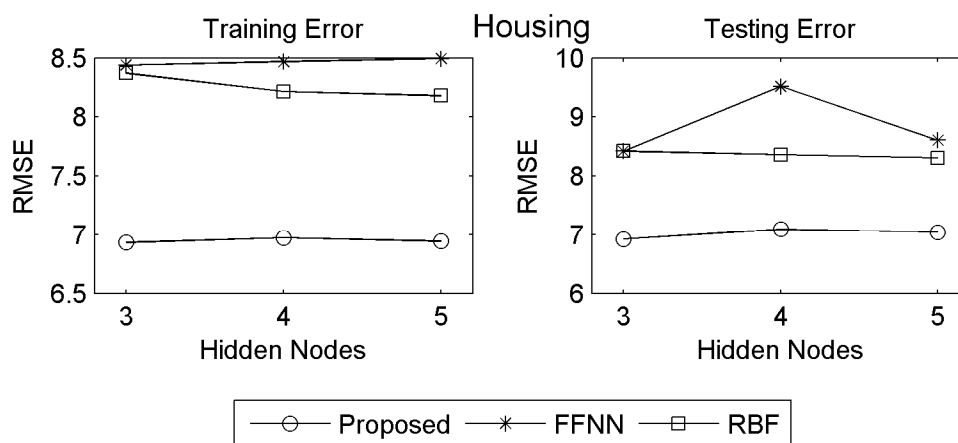
	Κρυφοί κόμβοι	Σφάλμα RMSE στα δεδομένα εκπαίδευσης (training)	Σφάλμα RMSE στα δεδομένα έλεγχου (testing)
Προτεινόμενο	3	62.2453 ± 8.3239	42.2381 ± 49.4908
	4	62.3574 ± 8.4046	42.3328 ± 49.3760
	5	62.3782 ± 8.3975	44.5432 ± 54.3305
RBF	3	62.8476 ± 8.2535	42.2969 ± 49.1077
	4	61.4744 ± 8.2087	46.3312 ± 47.6212
	5	62.6925 ± 8.3636	42.2315 ± 49.4742
FFNN	3	62.7100 ± 8.2286	42.1949 ± 49.2555
	4	62.7464 ± 8.2762	42.2148 ± 49.4116
	5	62.7170 ± 8.2876	42.7371 ± 49.4145



Σχήμα 4.5 Οι μέσες τιμές των αποτελεσμάτων του προηγούμενου πίνακα.

Πίνακας 4.3 Τα αριθμητικά αποτελέσματα του ΤΝΔ της ενότητας 4.2 καθώς και δύο ακόμη ΤΝΔ με τα οποία γίνεται σύγκριση για το σύνολο δεδομένων **Housing** του UCI Repository. Τα αποτελέσματα δίνονται στη μορφή: (Μέση τιμή) ± (τυπική απόκλιση)

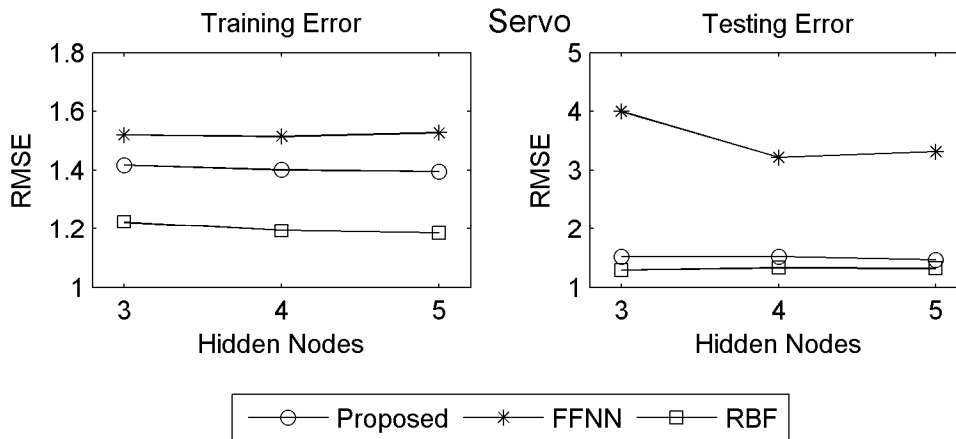
	Κρυφοί κόμβοι	Σφάλμα RMSE στα δεδομένα εκπαίδευσης (training)	Σφάλμα RMSE στα δεδομένα έλεγχου (testing)
Προτεινόμενο	3	6.9327 ± 0.5281	6.9256 ± 2.6806
	4	6.9722 ± 0.5118	7.0905 ± 2.6239
	5	6.9434 ± 0.4531	7.0447 ± 2.5887
RBF	3	8.3706 ± 0.3049	8.4206 ± 2.7753
	4	8.2159 ± 0.3065	8.3607 ± 2.7942
	5	8.1810 ± 0.3137	8.3061 ± 2.7712
FFNN	3	8.4396 ± 0.4913	8.4227 ± 2.9695
	4	8.4688 ± 0.4350	9.5179 ± 7.4965
	5	8.4937 ± 0.4742	8.6064 ± 3.6374



Σχήμα 4.6 Οι μέσες τιμές των αποτελεσμάτων του προηγούμενου πίνακα.

Πίνακας 4.4 Τα αριθμητικά αποτελέσματα του ΤΝΔ της ενότητας 4.2 καθώς και δύο ακόμη ΤΝΔ με τα οποία γίνεται σύγκριση για το σύνολο δεδομένων **Servo** του UCI Repository. Τα αποτελέσματα δίνονται στη μορφή: (Μέση τιμή) ± (τυπική απόκλιση)

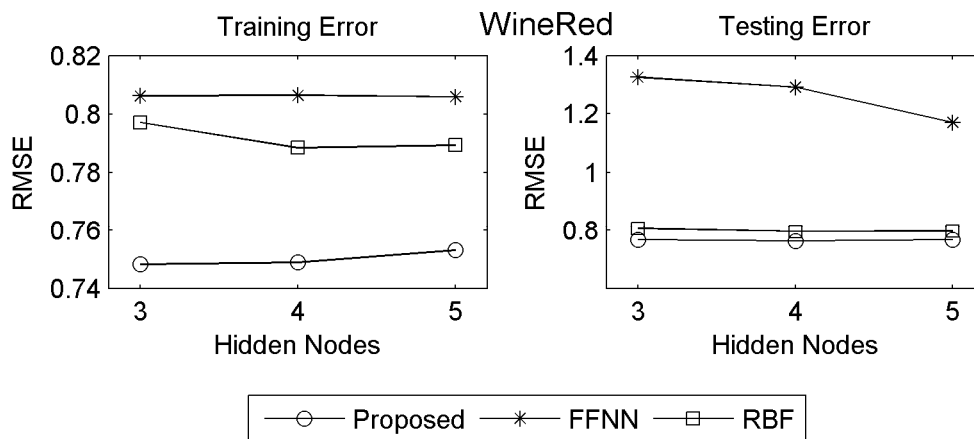
	Κρυφοί κόμβοι	Σφάλμα RMSE στα δεδομένα εκπαίδευσης (training)	Σφάλμα RMSE στα δεδομένα έλεγχου (testing)
Προτεινόμενο	3	1.4183 ± 0.0544	1.5190 ± 0.3738
	4	1.4013 ± 0.0590	1.5178 ± 0.3700
	5	1.3963 ± 0.0521	1.4669 ± 0.3653
RBF	3	1.2225 ± 0.0336	1.2890 ± 0.2998
	4	1.1932 ± 0.0347	1.3270 ± 0.3020
	5	1.1846 ± 0.0350	1.3193 ± 0.3037
FFNN	3	1.5199 ± 0.0659	4.0018 ± 9.5241
	4	1.5153 ± 0.0580	3.2199 ± 7.7853
	5	1.5277 ± 0.0521	3.3190 ± 7.6749



Σχήμα 4.7 Οι μέσες τιμές των αποτελεσμάτων του προηγούμενου πίνακα.

Πίνακας 4.5 Τα αριθμητικά αποτελέσματα του ΤΝΔ της ενότητας 4.2 καθώς και δύο ακόμη ΤΝΔ με τα οποία γίνεται σύγκριση για το σύνολο δεδομένων **WhineRed** του UCI Repository. Τα αποτελέσματα δίνονται στη μορφή: (Μέση τιμή) ± (τυπική απόκλιση)

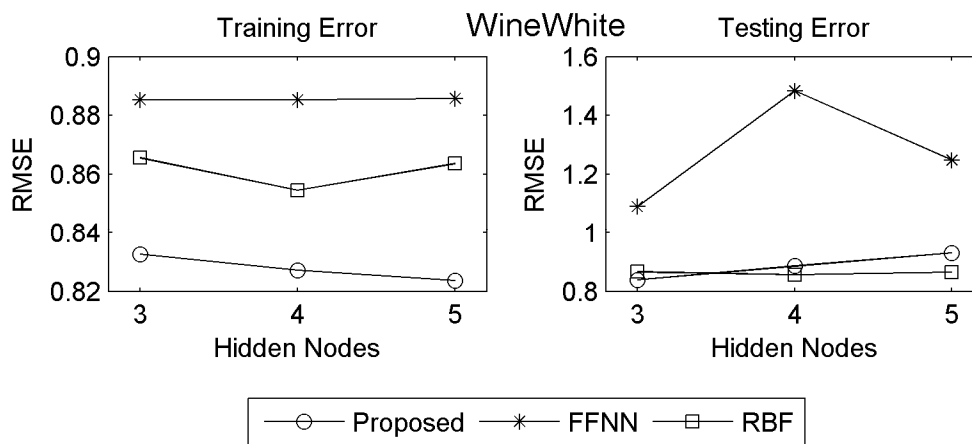
	Κρυφοί κόμβοι	Σφάλμα RMSE στα δεδομένα εκπαίδευσης (training)	Σφάλμα RMSE στα δεδομένα έλεγχου (testing)
Προτεινόμενο	3	0.7482 ± 0.0215	0.7664 ± 0.1037
	4	0.7489 ± 0.0195	0.7624 ± 0.0871
	5	0.7530 ± 0.0154	0.7664 ± 0.0870
RBF	3	0.7971 ± 0.0086	0.8050 ± 0.0727
	4	0.7884 ± 0.0087	0.7949 ± 0.0772
	5	0.7893 ± 0.0085	0.7967 ± 0.0756
FFNN	3	0.8062 ± 0.0091	1.3258 ± 3.5251
	4	0.8064 ± 0.0086	1.2919 ± 2.6936
	5	0.8059 ± 0.0091	1.1714 ± 3.5744



Σχήμα 4.8 Οι μέσες τιμές των αποτελεσμάτων του προηγούμενου πίνακα.

Πίνακας 4.6 Τα αριθμητικά αποτελέσματα του ΤΝΔ της ενότητας 4.2 καθώς και δύο ακόμη ΤΝΔ με τα οποία γίνεται σύγκριση για το σύνολο δεδομένων **WineWhite** του UCI Repository. Τα αποτελέσματα δίνονται στη μορφή: (Μέση τιμή) ± (τυπική απόκλιση)

	Κρυφοί κόμβοι	Σφάλμα RMSE στα δεδομένα εκπαίδευσης (training)	Σφάλμα RMSE στα δεδομένα έλεγχου (testing)
Προτεινόμενο	3	0.8326 ± 0.0126	0.8382 ± 0.0669
	4	0.8271 ± 0.0128	0.8858 ± 0.4215
	5	0.8236 ± 0.0134	0.9302 ± 0.8825
RBF	3	0.8656 ± 0.0075	0.8658 ± 0.0702
	4	0.8546 ± 0.0067	0.8564 ± 0.0681
	5	0.8636 ± 0.0073	0.8653 ± 0.0701
FFNN	3	0.8853 ± 0.0070	1.0903 ± 1.4500
	4	0.8853 ± 0.0071	1.4844 ± 2.3362
	5	0.8858 ± 0.0078	1.2485 ± 2.0239



Σχήμα 4.9 Οι μέσες τιμές των αποτελεσμάτων του προηγούμενου πίνακα.

Όπως φαίνεται και από τους παραπάνω πίνακες και σχήματα, το προτεινόμενο ΤΝΔ δίνει καλύτερα αποτελέσματα σε σχέση με τα άλλα δύο και έχει και την τάση να δίνει καλύτερα όσο το πλήθος των κρυφών κόμβων του δικτύου αυξάνει. Στο σύνολο AutoMPG και στα δεδομένα εκπαίδευσης και στα δεδομένα έλεγχου, καλύτερα αποτελέσματα δίνει το προτεινόμενο για $c = 5$ κρυφούς κόμβους και ακολουθεί το RBF. Στο σύνολο ForestFires στα δεδομένα εκπαίδευσης δίνει τα καλύτερα αποτελέσματα το RBF για $c = 4$ ενώ στα δεδομένα έλεγχου το FFNN για $c = 3$. Γενικά το τελευταίο σύνολο δεδομένων θα μπορούσαμε να ισχυριστούμε ότι χρειάζεται όσο το δυνατόν πιο απλά ΤΝΔ για να χειριστεί. Στο σύνολο Housing το προτεινόμενο υπερτερεί των άλλων δύο δικτύων για $c = 3$ και ακολουθεί το RBF. Στο σύνολο Servo υπερτερεί το RBF και ακολουθεί το προτεινόμενο. Στο σύνολο WineRed το προτεινόμενο δίνει τα καλύτερα αποτελέσματα και στα δεδομένα εκπαίδευσης και στα δεδομένα έλεγχου και ακολουθεί το RBF. Και τέλος στο WineWhite το

προτεινόμενο δίνει τα καλύτερα αποτελέσματα στα δεδομένα εκπαίδευσης για $c = 5$ ενώ στα δεδομένα ελέγχου για $c = 3$. Γενικά στα σύνολα του UCI Repository στα οποία έγιναν οι προσομοιώσεις της παραγράφου αυτής το προτεινόμενο δίκτυο έδωσε πολύ ικανοποιητικά αποτελέσματα στις περισσότερες περιπτώσεις· εξαίρεση αποτελούν τα σύνολα δεδομένων ForestFires και Servo.

4.6 Αριθμητικές προσομοιώσεις στο πρόβλημα πρόβλεψης της περιστροφής μιας παραλίας (beach rotation)

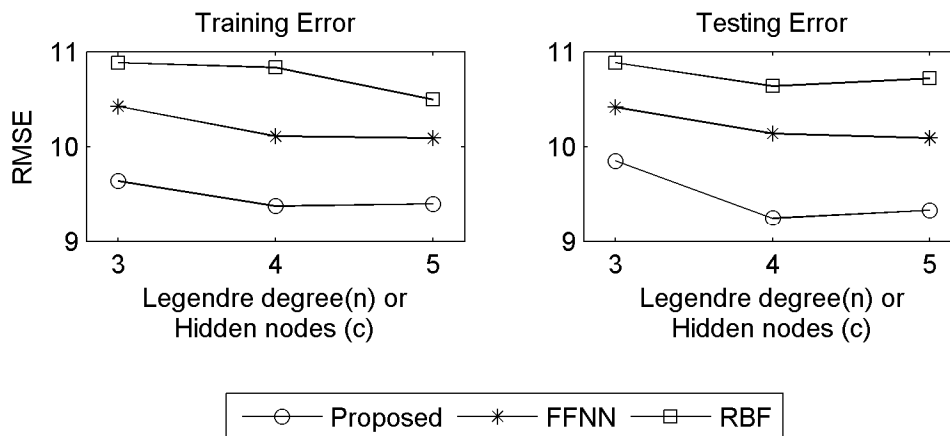
Στην παρούσα παράγραφο θα δοθούν οι αριθμητικές προσομοιώσεις που έγιναν στο σύνολο δεδομένων 'Beach Rotation'. Το σύνολο δεδομένων Beach Rotation αποτελείται από $N = 4148$ δεδομένα εισόδου-εξόδου της μορφής : $\{\mathbf{x}_k, y_k\}$ όπου $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{k6}]^T$ και $y_k \in \mathbb{R}$ και προσπαθεί να προβλέψει την περιστροφή του μετώπου μιας παραλίας (που ονομάζεται φαινόμενο Beach Rotation) βασισμένο στα στοιχεία ενός υφάλου που υπάρχει στην περιοχή συλλογής πρωταρχικών δεδομένων καθώς και σε δύο κυματικές συνθήκες. Ο ορισμός του φαινομένου Beach Rotation, η περιοχή συλλογής πρωταρχικών δεδομένων και η διαδικασία δημιουργίας του συνόλου δεδομένων Beach Rotation βρίσκονται αναλυτικά στο Παράρτημα Β.3.

Προκειμένου να γίνουν αριθμητικές προσομοιώσεις στα δεδομένα Beach Rotation δημιουργήθηκε και εκπαιδεύτηκε το προτεινόμενο ΤΝΔ όπως αυτό αναλύθηκε στις παραγράφους 4.3 και 4.4. Οι τιμές των παραμέτρων που χρησιμοποιήθηκαν στην εξίσωση (4.16) είναι $\alpha_L = -10$ και $\alpha_U = 10$. και ο μέγιστος αριθμός επαναλήψεων του αλγορίθμου GABC ορίστηκε ως 350.

Για σκοπούς σύγκρισης, δημιουργήθηκαν τα ΤΝΔ που ονομάστηκαν FFNN και RBF και αναλύθηκαν στην παράγραφο 4.5. Σε όλες τις προσομοιώσεις χρησιμοποιήθηκε η μέθοδος επαλήθευσης 10-fold cross validation με 10 αρχικοποιήσεις (initializations), οι δοκιμές εκτελέστηκαν με πλήθος πολυωνύμων Legendre $n = 3, 4, 5$ για το προτεινόμενο ΤΝΔ και αντίστοιχα $c = 3, 4, 5$ ως πλήθος κρυφών κόμβων στα FFNN και RBF. Όλα τα αποτελέσματά παρουσιάζονται στον πίνακα 4.7 στη μορφή (μέση τιμή) \pm (τυπική απόκλιση) του σφάλματος *RMSE* και οι μέσες τιμές των *RMSE* παρουσιάζονται στο Σχήμα 4.11. Πρέπει να σημειωθεί ότι τα αποτελέσματα δεν είναι τα ίδια με αυτά των Rigos et al.[168] μιας και στην [168] έχει γίνει χρήση της μεθόδου Barrier [173] στην διαδικασία της εκπαίδευσης του δικτύου.

Πίνακας 4.7 Τα αριθμητικά αποτελέσματα του ΤΝΔ της ενότητας 4.2 καθώς και δύο ακόμη ΤΝΔ με τα οποία γίνεται σύγκριση για το σύνολο δεδομένων **Beach Rotation**. Τα αποτελέσματα είναι σφάλματα *RMSE* και δίδονται στη μορφή: (Μέση τιμή) ± (τυπική απόκλιση)

	Κρυφοί κόμβοι (n ή c)	Σφάλμα <i>RMSE</i> στα δεδομένα εκπαίδευσης (training)	Σφάλμα <i>RMSE</i> στα δεδομένα έλεγχου (testing)
Προτεινόμενο	3	9.6412 ± 0.1921	9.8519 ± 0.1793
	4	9.3776 ± 0.2038	9.2492 ± 0.1823
	5	9.4021 ± 0.1490	9.3345 ± 0.2440
RBF	3	10.4243 ± 0.1107	10.4145 ± 0.1646
	4	10.1126 ± 0.0794	10.1391 ± 0.1392
	5	10.0920 ± 0.0902	10.0930 ± 0.1323
FFNN	3	10.8839 ± 0.8212	10.8849 ± 0.9075
	4	10.8349 ± 0.7830	10.6389 ± 0.7330
	5	10.4968 ± 0.7252	10.7192 ± 0.8291

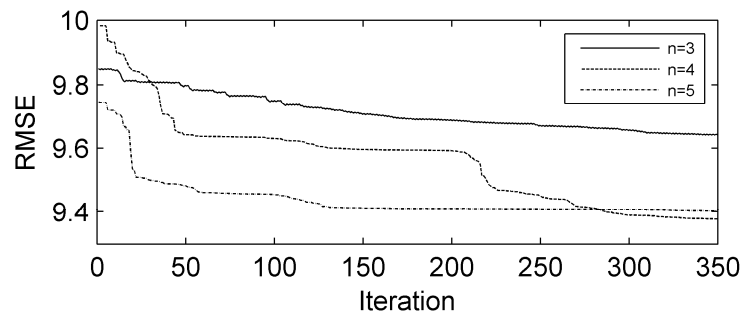


Σχήμα 4.10 Οι μέσες τιμές των αποτελεσμάτων των τριών ΤΝΔ στο σύνολο δεδομένων Beach Rotation.

Από τα αποτελέσματα φαίνεται ότι το προτεινόμενο πολυωνυμικό νευρωνικό δίκτυο έχει καλύτερη απόδοση σε σύγκριση με τα άλλα δύο δίκτυα. Τα καλύτερα αποτελέσματα τόσο και για τα δεδομένα εκπαίδευσης αλλά και για τα δεδομένα έλεγχου δίδονται από το προτεινόμενο δίκτυο για $n = 4$. Αξίζει επίσης να σημειωθούν τα εξής συμπεράσματα :

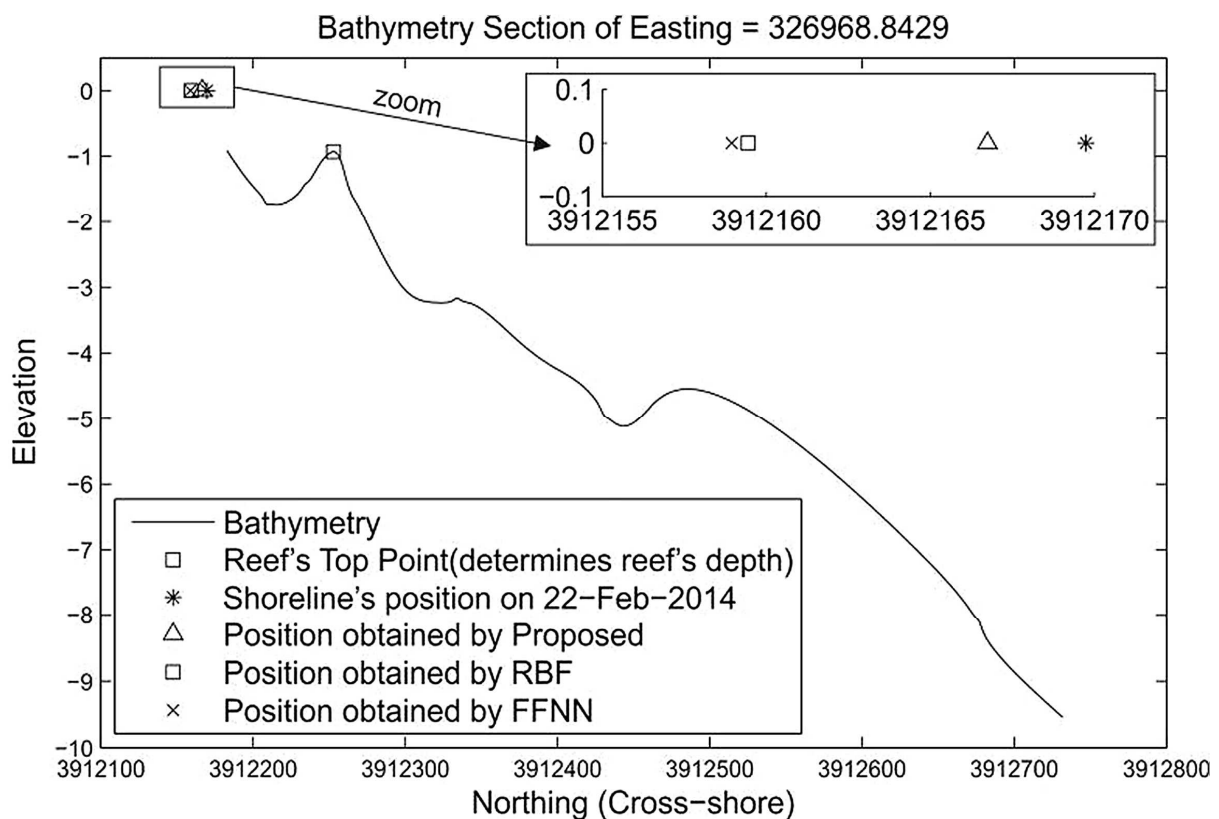
- (α) η διαφορά μεταξύ του πολυωνύμου Legendre και των άλλων δύο δικτύων φαίνεται να είναι σημαντική ιδιαίτερα στα δεδομένα έλεγχου
- (β) το RBF υπερτερεί του FFNN και στις δύο περιπτώσεις
- (γ) παρόλο που το καλύτερο αποτέλεσμα προκύπτει από το προτεινόμενο δίκτυο για $n = 4$, η γενική τάση είναι να προκύπτουν μικρότερα σφάλματα *RMSE* καθώς το πλήθος n των κρυφών κόμβων του πολυωνύμου αυξάνει.

Στο Σχήμα 4.12 παρουσιάζεται η σύγκλιση της μεθόδου GABC, όπου φαίνεται σε κάθε επανάληψη της μεθόδου GABC να μικραίνει το RMSE σφάλμα του ΤΝΔ.



Σχήμα 4.11 Το σφάλμα RMSE του ΤΝΔ κατά τις επαναλήψεις της μεθόδου GABC.

Είναι πολύ ενδιαφέρον να δούμε πώς τα παραπάνω αποτελέσματα μεταφράζονται στη φυσική ερμηνεία του φαινομένου Beach Rotation. Το Σχήμα 4.12 προέρχεται από τα δεδομένα του Beach Rotation (βλ. Παράρτημα Β.3) και παρουσιάζει τις προβλέψεις που προκύπτουν στην πρόβλεψη της θέσης μίας ακτογραμμής από τις τρεις μεθόδους. Με βάση αυτό το σχήμα, το προτεινόμενο δίκτυο αποδίδει την καλύτερη πρόβλεψη στο φαινόμενο Beach Rotation.



Σχήμα 4.12 Ένα παράδειγμα της θέσης της ακτογραμμής όπως έχει προβλεφθεί από τα τρία ΤΝΔ.

Παρόλο που συνολικά η πρόβλεψη με τη χρήση του προτεινόμενου Legendre πολυωνυμικού ΤΝΔ μπορεί να μην φαίνεται να είναι πολύ ικανοποιητική (βάση του RMSE που προκύπτει ≈ 9.4 μέτρα), θα πρέπει να σημειωθούν τα ακόλουθα. Πρώτον, η θέση της ακτογραμμής της παραλίας της Αμμουδάρας (βλ. Παράρτημα Β.3) χαρακτηρίζεται από υψηλή χωροχρονική μεταβλητότητα, με τη θέση της να αλλάζει 3-8 μέτρα μέσα σε μια δεκάμηνη περίοδο παρατηρήσεων [174]. Οι προβλέψεις του προτεινόμενου δικτύου μπορούν να θεωρηθούν ικανοποιητικές σε τέτοιο υψηλά μη γραμμικό παράκτιο σύστημα, ιδιαίτερα καθώς σε πολλές τομές οι προβλέψεις του δικτύου είναι πολύ πιο κοντά στη θέση της παρατηρούμενης ακτογραμμής (βλ. Σχήμα 4.12). Δεύτερον, συγκεκριμένα τμήματα της ακτογραμμής έχουν δείξει μεγάλη μεταβλητότητα διάβρωσης/εναπόθεσης ιζήματος γεγονός που υποδηλώνει σημαντική επιρροή από μικρές διαφορές της μορφολογίας του υφάλου και της κατεύθυνσης της προσέγγισης της ακτής από τα κύματα. Η υδροδυναμική μοντελοποίηση της περιοχής αυτής έχει δείξει ότι οι μικρές διαφορές στη γωνία προσέγγισης των κυμάτων δίνουν αρκετά διαφορετικά παράκτια υδροδυναμικά αποτελέσματα (κυμάτων και των ρευμάτων που προκαλούνται από τα κύματα), ακόμη και για την περίπτωση των κυμάτων που προέρχονται από το ανοιχτό πέλαγος (offshore) με ίδια χαρακτηριστικά (σημαντικά ύψη κύματος H_s και περιόδους T_p) [174]. Τα χαρακτηριστικά των κυμάτων που προέρχονται από το ανοιχτό πέλαγος (δεδομένα H_s και T_p) δεν περιείχαν πληροφορίες σχετικά με την κατεύθυνση των κυμάτων και όποτε αυτά που χρησιμοποιήθηκαν, υπολογίστηκαν ως βόρεια κύματα (αυτή είναι η διεύθυνση των κυμάτων που επηρεάζουν την συγκεκριμένη παραλία). Έτσι, οι παρατηρούμενες διαφορές θα μπορούσαν να οφείλονται κυρίως στην απουσία επιπλέον παραμέτρων του κύματος (γωνία πρόσπτωσης του κύματος στην παραλία) στα δεδομένα Beach Rotation που χρησιμοποιήθηκαν ως μεταβλητές εισόδου του δικτύου. Ως μελλοντική έρευνα, θα ήταν ενδιαφέρον να δοκιμασθεί πώς θα ήταν τα παραπάνω αποτελέσματα αν και η κατεύθυνση του κύματος συμπεριλαμβανόταν ως μεταβλητή εισόδου.

5. Συγκριτικά Αποτελέσματα Των Πολυωνυμικών Δικτύων Του Παρόντος Κειμένου Και Μελλοντική Εργασία

5.1 Σύγκριση των προτεινόμενων μεθόδων

Στην παρούσα διδακτορική διατριβή διεξήχθη έρευνα και αποδείχθηκε η χρησιμότητα τριών προτύπων δικτύων (δύο ΤΝΔ και ενός νευροασαφούς συστήματος) με χρήση ορθογωνίων πολυωνύμων. Στο ένα τεχνητό νευρωνικό δίκτυο (κεφάλαιο 2) έγινε επίσης και χρήση μιας πρωτότυπης συσταδοποίησης δεδομένων εισόδου-εξόδου.

Όλες οι προτεινόμενοι μέθοδοι συγκρίθηκαν πάνω σε σύνολα δεδομένα (α) του UCI repository και (β) σε κάθε ενότητα δημιουργήθηκε ένα σύνολο δεδομένων έπειτα από επεξεργασία δεδομένων που είχαν συλλεχθεί από το τμήμα επιστημών της θάλασσας του παν/μίου Αιγαίου. Πιο συγκεκριμένα, οι επεξεργασίες αυτές περιείχαν :

(i) επεξεργασία εικόνων (κεφ.2)

(ii) επεξεργασία τοπογραφικών και βαθυμετρικών δεδομένων (κεφ.3)

(ii) επεξεργασία τοπογραφικών, βαθυμετρικών, μετεωρολογικών δεδομένων καθώς και δεδομένων που προέκυψαν από επεξεργασία εικόνων (κεφ.4, βλ. και παράρτημα Β.3).

Ο πίνακας 5.1 παρουσιάζει μια σύνοψη των μεθόδων που ακολουθήθηκαν στο παρόν κείμενο καθώς και των εφαρμογών τους σε φυσικά προβλήματα. Όπως παρουσιάστηκε στις παραγράφους 2.6, 3.6 και 4.6 τα τρία πρωτότυπα δίκτυα πέτυχαν να μοντελοποιήσουν το φυσικό πρόβλημα που έπρεπε κάθε φορά να προσομοιωθεί.

Πίνακας 5.1 Σύνοψη των μεθόδων που παρουσιάστηκαν στο παρόν κείμενο

Δίκτυο με χρήση πολυωνύμων	Τύπος Δικτύου	Διαδικασία εκπαίδευσης	Εφαρμογή σε σύνολο δεδομένων εκτός του UCI Repository	Στόχος της εφαρμογή σε σύνολο δεδομένων εκτός του UCI Repository
Chebyshev (κεφάλαιο 2)	TNΔ	Gradient Descent με χρήση του Armijo Rule	Shoreline Extraction	Εξαγωγή της θέσης της ακτογραμμής από ψηφιακές εικόνες
Hermite (κεφάλαιο 3)	Νευροασαφές σύστημα	Gradient Descent με χρήση του Armijo Rule	Coastal Erosion	Πρόβλεψη μετατόπισης/οπισθοχώρησης ακτογραμμής από χωρικά και μετεωρολογικά δεδομένα
Legendre (κεφάλαιο 4)	TNΔ	Αλγόριθμος GABC	Beach Rotation	Πρόβλεψη της περιστροφής του μετάπου μιας παραλίας από χωρικά και κυματικά δεδομένα

Στα κεφάλαια 2 και 4 τα πολυώνυμα Chebyshev και Legendre αντίστοιχα έπρεπε να εφαρμόζονται σε δεδομένα στο διάστημα $[-1,1]$ και αυτό οδήγησε σε μετασχηματισμούς δεδομένων, αντίθετα στο κεφάλαιο 3 δεν υπήρξε τέτοιος μετασχηματισμός για τα πολυώνυμα Hermite. Στο κεφάλαιο 2 εκτός του πολυωνυμικού δικτύου έγινε συνδυασμός με ένα TNΔ συναρτήσεων ακτινικής βάσης, στο 3^ο κεφ. συνδυάστηκε με ένα νευροασαφές σύστημα ενώ στο 4^ο κεφ. το δίκτυο δεν είχε κάποιο συνδυαστικό στάδιο/φάση. Η έξοδος \hat{y} ονομάζεται άμεση έκφραση σε πολυωνυμικές σειρές (των πολυωνύμων $P_i, i=1,2,\dots,n$) αν είναι της μορφής :

$$\hat{y} = a_0 + a_1 P_1(s_1) + a_2 P_2(s_2) + \dots + a_n P_n(s_n) \quad (5.1)$$

και όλα τα $s_i, i=1,2,\dots,n$ είναι ίσα μεταξύ τους, αυτό συνέβη στο κεφάλαιο 2, δεν συμβαίνει όμως στα δίκτυα των κεφαλαίων 3 και 4.

Προκειμένου να γίνει η σύγκριση στα αποτελέσματα των τριών προτεινόμενων μεθόδων επιλέγησαν και σύνολα δεδομένων για τα οποία έχουν πραγματοποιηθεί αριθμητικές προσομοιώσεις στις παραγράφους 2.5, 3.5 και 4.5. Επιλέγησαν τα σύνολα AutoMPG και Servo και στους πίνακες 5.2 ως 5.4 φαίνονται οι μέσες τιμές σφαλμάτων RMSE (σε δεδομένα εκπαίδευσης) και για τις τρεις προτεινόμενες μεθόδους.

Πίνακας 5.2 Μέσα σφάλματα (RMSEs) από τα 3 δίκτυα το παρόντος κειμένου για το σύνολο δεδομένων **AutoMPG**

Μέσα RMSE σφάλματα στα δεδομένα εκπαίδευσης Δίκτυο: Το Chebyshev TNΔ της ενότητας 2				
c	$r = 1$	$r = 2$	$r = 3$	$r = 4$
4	6.2959	5.4238	5.3518	5.2401
6	6.2958	5.4208	5.3262	5.2392
8	6.2958	5.4194	5.3089	5.2392
10	6.2957	5.4186	5.2974	5.2390
12	6.2957	5.4180	5.2899	5.2381
15	6.2957	5.4174	5.2814	5.2376

Μέσα RMSE σφάλματα στα δεδομένα εκπαίδευσης Δίκτυο: Το Hermite νευροασαφές σύστημα TNΔ της ενότητας 3				
c	$n = 2$	$n = 3$	$n = 4$	$n = 5$
3	14.3932	13.7056	10.2818	6.0472
5	14.1115	13.6235	8.8370	4.7187
6	14.3391	6.2165	5.6607	4.2533
8	13.0714	6.3500	4.8458	3.9313
9	8.8316	6.2573	4.8265	3.9542
10	8.2367	5.9019	4.7526	3.9554

Μέσα RMSE σφάλματα στα δεδομένα εκπαίδευσης Δίκτυο: Το Legendre TNΔ της ενότητας 4	
c	
3	3.4933
4	3.4322
5	3.3463

Πίνακας 5.3 Μέσα σφάλματα (RMSEs) από τα δίκτυα των κεφαλαίων 2 και 3 του παρόντος κειμένου για το σύνολο δεδομένων **Servo**

Μέσα RMSE σφάλματα στα δεδομένα εκπαίδευσης Δίκτυο: Το Chebyshev TNΔ της ενότητας 2				
c	$r = 1$	$r = 2$	$r = 3$	$r = 4$
4	1.5620	1.3320	1.2505	1.2640
6	1.5595	1.3214	1.2364	1.2484
8	1.5587	1.3156	1.2326	1.2384
10	1.5567	1.3132	1.2275	1.2333
12	1.5576	1.3115	1.2254	1.2269
15	1.5577	1.3090	1.2229	1.2246

Μέσα RMSE σφάλματα στα δεδομένα εκπαίδευσης Δίκτυο: Το Hermite νευροασαφές σύστημα TNΔ της ενότητας 3				
c	$n = 2$	$n = 3$	$n = 4$	$n = 5$
3	1.4418	1.4166	1.2293	1.3161
5	1.4005	1.2913	1.1856	1.2347
6	1.3857	1.2537	1.1630	1.2104
8	1.3747	1.1832	1.1070	1.1414
9	1.3625	1.1808	1.0985	1.1170
10	1.3592	1.1668	1.0828	1.0546

Πίνακας 5.4 Μέσα σφάλματα (RMSEs) από το δίκτυο του κεφαλαίου 4 του παρόντος κειμένου για το σύνολο δεδομένων **Servo**

c	Μέσα RMSE σφάλματα στα δεδομένα εκπαίδευσης Δίκτυο: Το Legendre TNA της ενότητας 4
3	1.4183
4	1.4013
5	1.3963

Όπως φαίνεται από τους πίνακες αυτούς, στο σύνολο AutoMPG το TNA με χρήση πολυωνύμων Legendre της ενότητας 3 επιτυγχάνει ικανοποιητική απόδοση με μικρό πλήθος κρυφών κόμβων με σφάλμα $RMSE \approx 3.3$ ενώ οι άλλες δύο μέθοδοι δίνουν σφάλματα $RMSE \approx 5.2$ και $RMSE \approx 3.9$. Αντιθέτως, στο σύνολο Servo χρειάζεται πιο πολύπλοκο δίκτυο για να φτάσει το σφάλμα σε μικρά επίπεδα, και αυτό το επιτυγχάνει το δίκτυο της ενότητας 3. Οι παραπάνω παρατηρήσεις δεν δείχνουν ότι το δίκτυο της ενότητας 2 δεν έχει καλή απόδοση, αρκεί κάποιος να παρατηρήσει ότι στο σύνολο δεδομένων ForestFires το δίκτυο της ενότητας 2 καταλήγει σε πολύ καλύτερα σφάλματα από τα άλλα δίκτυα ($RMSE \approx 52.66$ για το δίκτυο της ενότητας 2 και $RMSE \approx 62.15$ και 62.38 για αυτά των ενότητων 3 και 4 αντίστοιχα).

Προκειμένου να αξιολογηθούν οι προτεινόμενοι αλγόριθμοι στα αποτελέσματά τους (στα δεδομένα εκπαίδευσης) στο σύνολο δεδομένων AutoMPG με μεθόδους στατιστικής πραγματοποιήθηκε ο στατιστικός έλεγχος/τεστ Friedman Aligned Ranks [175]. Ο έλεγχος αυτός πραγματοποιήθηκε με το λογισμικό STAC [176], ένα διαδικτυακό λογισμικό το οποίο έχει δημιουργηθεί για την αξιολόγηση αποτελεσμάτων αλγορίθμων. Στο STAC δόθηκαν οι μέσες τιμές (λόγω δέκα αρχικοποιήσεων) από καθένα από τα δέκα folds στα οποία έχουν τρέξει οι παραπάνω αλγόριθμοι. Το στατιστικό τεστ αυτό ελέγχει τις υποθέσεις:

Μηδενική υπόθεση (H_0): Οι μέσες τιμές των αποτελεσμάτων των δύο ή περισσότερων αλγορίθμων είναι οι ίδιες .

Εναλλακτική υπόθεση (H_A): Οι μέσες τιμές των αποτελεσμάτων των δύο ή περισσότερων αλγορίθμων είναι διαφορετικές

Το αποτέλεσμα του τεστ ήταν να απορριφθεί η μηδενική υπόθεση ($p < 0.00001$), δηλαδή μπορούμε να ισχυριστούμε ότι οι αλγόριθμοι δίνουν διαφορετικά αποτελέσματα.

5.2 Μελλοντική εργασία

Ως μελλοντική εργασία θα μπορούσαν να αναπτυχθούν τα παρακάτω:

(i) Χρήση άλλων ορθογωνίων πολυωνύμων πέρα των τριών που χρησιμοποιήθηκαν, όπως τα: Jacobi, Laguerre, Gegenbauer, κ.ά. (βλ. πίνακα 1.1)

(ii) Χρήση δικτύων με άλλες -πρωτότυπες και μη- τοπολογίες

(iii) Χρήση άλλων μεθόδων βελτιστοποίησης στη διαδικασία εκπαίδευσης των δικτύων, όπως για παράδειγμα οι αλγόριθμοι νοημοσύνης σμήνους : Ant Colony Optimisation (ACO), Bat Algorithm (BA), Artificial Fish Swarm, Wolf Search Algorithm κ.ά. (βλ. παράγραφο 1.1.3)

(iv) Χρήση των δικτύων σε πρωτότυπα σύνολα δεδομένων. Στο παρόν έγγραφο έγιναν αριθμητικές προσομοιώσεις σε σύνολα δεδομένων που αφορούν την κίνηση των κυμάτων και των ακτογραμμών . Θα είχε ενδιαφέρον όμως να γίνει εφαρμογή και σε άλλα πρωτότυπα -περιβαλλοντολογικά και μη- δεδομένα.

Παράρτημα Α

Δεδομένα του UCI Repository

Το UCI Machine Learning repository [100] είναι μια συλλογή με βάσεις δεδομένων η οποία χρησιμοποιείται από τους ερευνητές που ασχολούνται με το Machine Learning και είναι διαθέσιμη στο διαδίκτυο στην ιστοσελίδα: <https://archive.ics.uci.edu>

Τα δεδομένα τα οποία συλλέχθηκαν από την ιστοσελίδα αυτή με σκοπό να γίνουν οι προσομοιώσεις στην παρούσα διδακτορική διατριβή, είναι δεδομένα παλινδρόμησης (regression) και οι περιγραφές τους παρουσιάζονται στον παρακάτω πίνακα.

Πίνακας Π.1 Τα δεδομένα παλινδρόμησης (regression) από το UCI Machine Learning repository που χρησιμοποιούνται στην παρούσα διδακτορική διατριβή

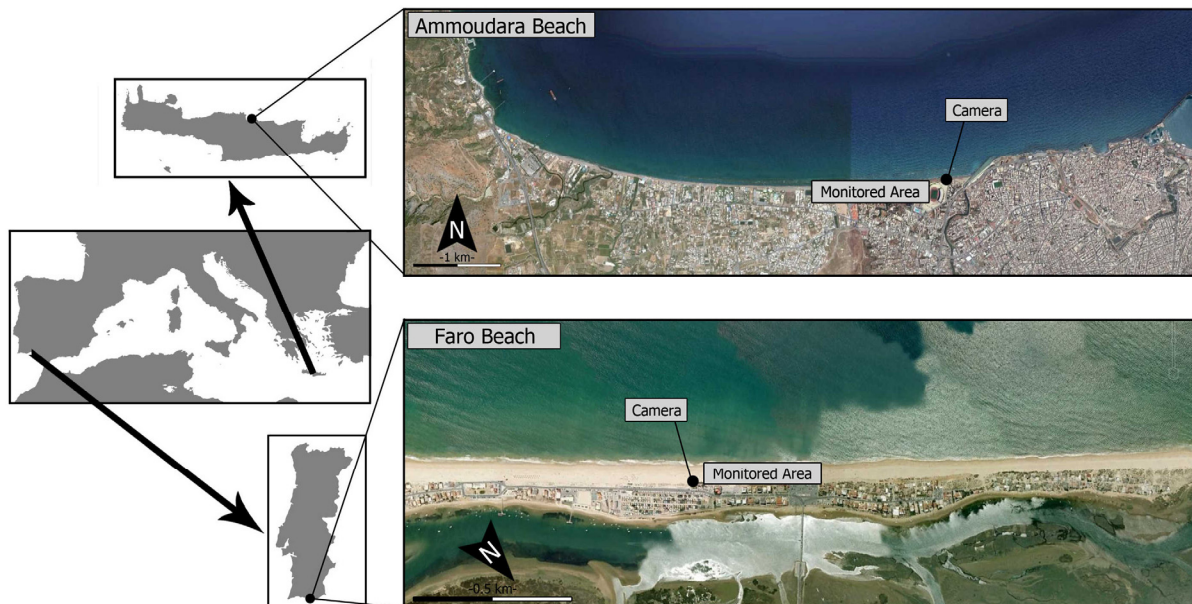
Σύνολο δεδομένων	Πλήθος δεδομένων (N)	Διάσταση χώρου εισόδου (p)	Διάσταση χώρου εξόδου (s)	Σύντομη περιγραφή της βάσης δεδομένων
AutoMPG	392	7	1	Χαρακτηριστικά αυτοκινήτων και ως έξοδος η κατανάλωση καυσίμου σε μίλια ανά γαλόνι. Περιέχει γεωγραφικές συντεταγμένες, ημερομηνίες και μετεωρολογικές συνθήκες και το εμβαδό καμένης έκτασης (ή μηδέν αν δεν εκδηλώθηκε πυρκαγιά) στο πάρκο Montesinho της Πορτογαλίας.
ForestFires	517	12	1	Δεδομένα για τις τιμές και άλλα χαρακτηριστικά των σπιτιών της Βοστώνης των ΗΠΑ. Ως έξοδος είναι το κατά κεφαλήν ποσοστό εγκληματικότητας στην πόλη. Πρόβλεψη της ώρας που χρειάζεται ένα servo να ανταποκριθεί ανάλογα με κάποιες μηχανικές ρυθμίσεις της συνδεσμολογίας του.
Housing	452	13	1	Δεδομένα του Πορτογαλικού κρασιού "Vinho Verde" και ως έξοδος η ποιότητα του ως βαθμός μεταξύ 0 και 10.
Servo	167	4	1	
Wine Quality Red και White	1599 και 4898	11	1	

Παράρτημα Β

Τα Σύνολα Δεδομένων Shoreline Extraction, Coastal Erosion και Beach Rotation.

Β.1 Τα δεδομένα για την Εξαγωγή Ακτογραμμής από Ψηφιακές Εικόνες (Shoreline Extraction)

Τα δεδομένα που παρουσιάζονται στην παράγραφο αυτή έχουν συλλεχθεί από δύο περιοχές μελέτης οι οποίες είναι η παραλία της Αμμουδάρας της Κρήτης και η παραλία Faro της Πορτογαλίας (Σχήμα Β.1). Στην παράγραφο αυτή περιγράφεται η διαδικασία η οποία ακολουθήθηκε και δημιούργησε εικόνες τύπου SIGMA, οι οποίες κατόπιν επεξεργάστηκαν με τη μέθοδο που αναφέρεται στην παράγραφο 2.6 και δημιουργούν το σύνολο δεδομένων Shoreline Extraction, το οποίο χρησιμοποιείται στην παράγραφο 2.6.



Σχήμα Β.1 Οι περιοχές μελέτης από όπου προέκυψαν τα δεδομένα Shoreline Extraction που χρησιμοποιήθηκαν στην παράγραφο 2.6.

Ο σκοπός της έρευνας των Rigos et al. [1] ήταν η αυτοματοποιημένη εξαγωγή της θέσης της ακτογραμμής από ψηφιακές εικόνες. Για το λόγο αυτό στις περιοχές μελέτης τοποθετήθηκαν

IP-κάμερες οι οποίες παρήγαγαν εικόνες ανάλυσης 2048×1536 εικονοστοιχείων και ήταν συνδεδεμένες με Η/Υ οι οποίοι συλλέγαν τις εικόνες για 10 λεπτά κάθε ώρα (τις φωτεινές ώρες της ημέρας) με συχνότητα 1 Hz. Αποτέλεσμα της διαδικασίας αυτής ήταν για κάθε ώρα να συλλέγονται $Q = 600$ διαδοχικές έγχρωμες εικόνες τύπου RGB (εικόνες τύπου snapshot). Κατόπιν, τα 50 υπόλοιπα λεπτά της ώρας ο Η/Υ δεν σύλλεγε εικόνες αλλά μετέτρεπε τις εικόνες σε εικόνες τύπου TIMEX (time averages) και SIGMA (το όνομα των οποίων προέρχεται από το σύμβολο σ της τυπικής απόκλισης) ακολουθώντας τα επόμενα βήματα επεξεργασίας:

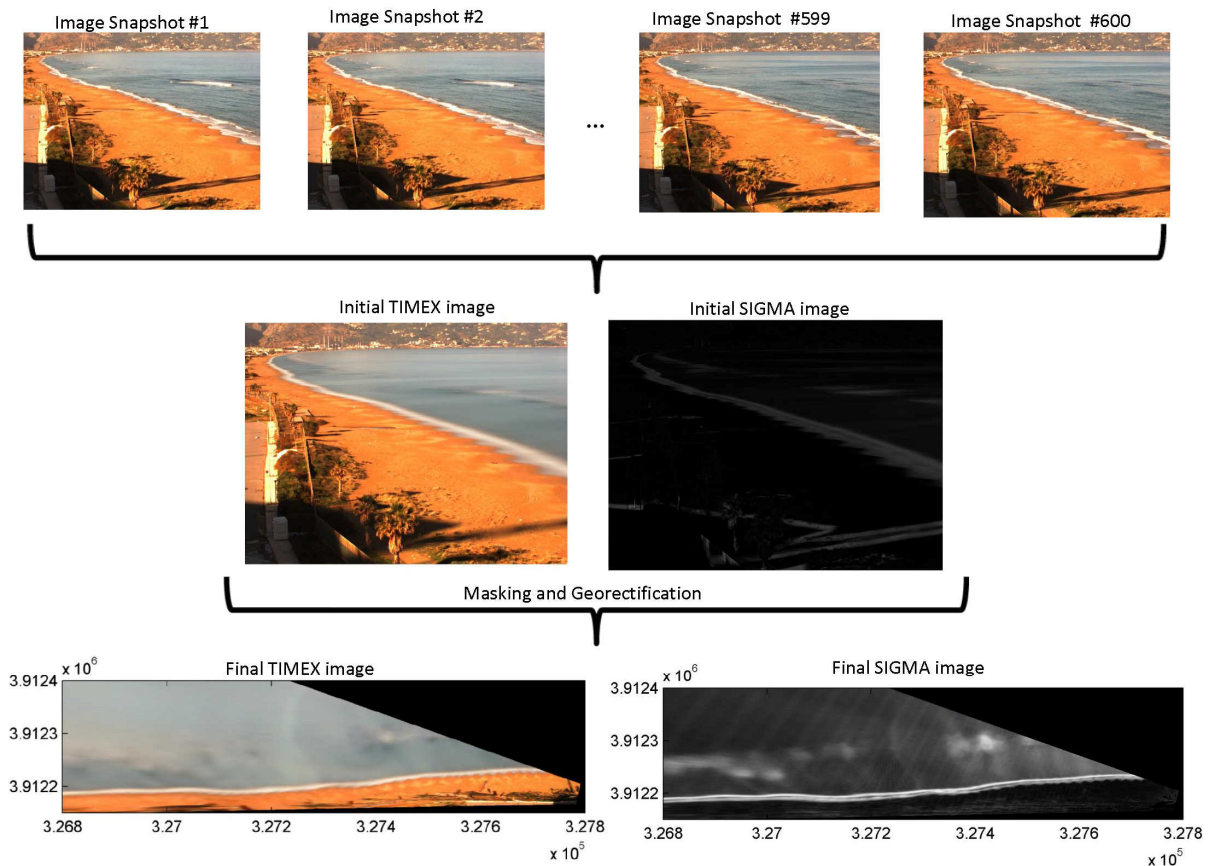
(α) Πρώτα από τις διαδοχικές εικόνες διατηρούνταν μόνο η περιοχή ενδιαφέροντος (περιοχή που να περιέχει την ακτογραμμή, range of interest ή ROI) ενώ άλλες περιοχές όπως π.χ. χώροι στάθμευσης διαγραφόντουσαν από τις εικόνες. Η διαδικασία αυτή ονομάζεται masking των εικόνων.

(β) Μετέπειτα στις εικόνες αυτές εφαρμόζονταν μέθοδος ορθοαναγωγής (georectification) [158, 160, 177-179] ώστε να προκύψουν εικόνες οι οποίες έχουν τη μορφή σαν να έχουν συλλεχθεί από κάποιον που να βρίσκεται κάθετα στο επίπεδο της γης (σαν να έχουν συλλεχθεί από δορυφόρο).

(γ) Στο τελευταίο βήμα υπολογίζονταν οι εικόνες TIMEX και SIGMA. Οι εικόνες TIMEX είναι τύπου RGB και προκύπτουν υπολογίζοντας για κάθε κανάλι τους (το Red, το Green και το Blue) το μέσο όρο των τιμών καθενός καναλιού για τις $Q = 600$ διαδοχικές εικόνες. Οι εικόνες SIGMA προκύπτουν μετατρέποντας πρώτα τις διαδοχικές εικόνες τύπου snapshot σε εικόνες αποχρώσεων του γκρι* (grayscale) και κατόπιν υπολογίζοντας για κάθε εικονοστοιχείο της εικόνας την τυπική του απόκλιση για τις $Q = 600$ διαδοχικές τιμές που έχει πάρει.

Η διαδικασία αυτή περιγράφεται στο Σχήμα Β.2

* Ο συνήθης τρόπος που μετατρέπει μια 3-κάναλη εικόνα RGB σε εικόνα αποχρώσεων του γκρι είναι μέσω της εξίσωσης: $Grayscale = 0.2989 \cdot R + 0.5870 \cdot G + 0.1140 \cdot B$, βλέπε π.χ. τις συναρτήσεις 'rgb2gray' στο Matlab, 'convert' στο πακέτο PIL (Python Imaging Library) στην Python κ.ά.



Σχήμα B.2 Η διαδικασία δημιουργίας εικόνων τύπου TIMEX και SIGMA από διαδοχικές εικόνες.

Πρέπει να σημειωθεί ότι οι εικόνες TIMEX και SIGMA που προκύπτουν είναι τύπου intensity-images (βλ. [157]) το οποίο σημαίνει πως οι τιμές τους είναι μια συνεχής μεταβλητή (επειδή έχει προκύψει από αριθμητικές πράξεις) και όχι ακέραιοι αριθμοί στο διάστημα $[1, 256]$ όπως συμβαίνει συνήθως με τις ψηφιακές εικόνες. Οι εικόνες TIMEX είναι τύπου RGB με τιμές (όχι αναγκαστικά ακέραιους) στο διάστημα $[1, 256]$ ενώ για τις SIGMA δεν υπάρχει άνω φράγμα για τις τιμές που θα προκύψουν.

Για να αποφευχθούν οι μεγάλες τιμές στους άξονες συντεταγμένων, όπως αυτές του Σχήματος B.2, έγινε μετατροπή των συντεταγμένων σε τοπικό σύστημα αναφοράς. Χρησιμοποιώντας ως αρχή των αξόνων συντεταγμένων το σημείο με συντεταγμένες (12255, -295575) στο γεωγραφικό σύστημα συντεταγμένων Datum 73, EPSG 27493, για την περιοχή του Faro και ως αρχή των αξόνων το σημείο (327897, 3912170) στο σύστημα συντεταγμένων UTM της ζώνης 35 N (προβολικό σύστημα: WGS84), οι συντεταγμένες (σε

m) στις εικόνες που προκύπτουν είναι σαν αυτές που παρουσιάστηκαν στα Σχήματα 2.9, 2.12 και 2.13.

Η μετέπειτα επεξεργασία των εικόνων που προέκυψαν αναλύεται στην παράγραφο 2.6 και τα δεδομένα Shoreline Extraction που προέκυψαν από την επεξεργασία αυτή δημοσιεύτηκαν στο [1] και είναι διαθέσιμα στο σύνδεσμο <http://isla.aegean.gr/news.php?lg=2> .

B.2 Τα δεδομένα για την Μοντελοποίηση/Πρόβλεψη μετατόπισης/οπισθοχώρησης ακτογραμμής (Coastal Erosion)

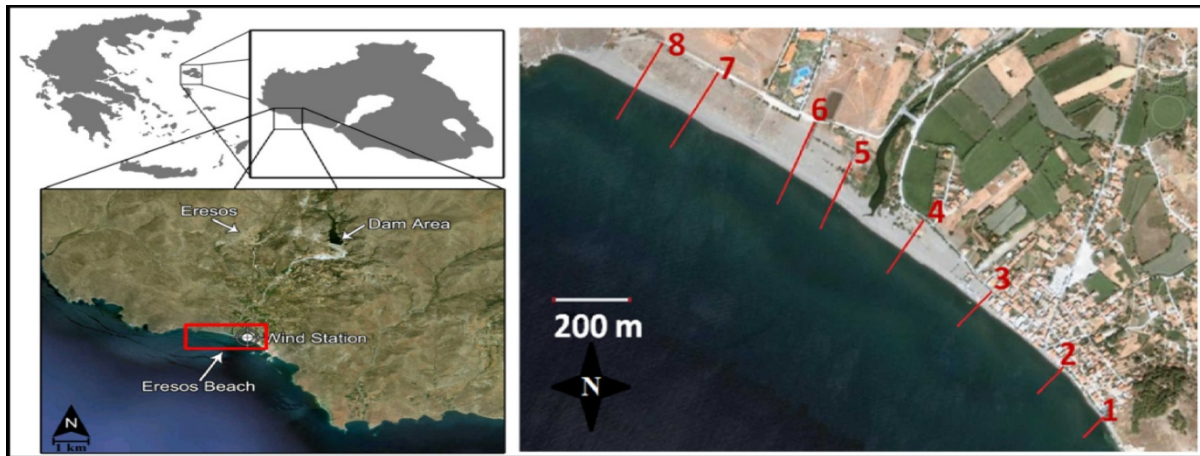
Τι είναι η διάβρωση των παραλιών (Coastal Erosion)

Η διάβρωση των ακτών/παραλιών (Coastal Erosion) είναι ένα από τα σημαντικότερα περιβαλλοντικά προβλήματα στις παράκτιες ζώνες σε όλο τον κόσμο και έχουν καταστροφικές συνέπειες για την οικονομία, καθώς το μεγαλύτερο πλήθος του παγκόσμιου πληθυσμού συγκεντρώνεται στις παράκτιες περιοχές [180]. Η διάβρωση των ακτών είναι το αποτέλεσμα της διαδικασίας της μεταφοράς του ιζήματος της ακτής. Λαμβάνει χώρα κατά τη διάρκεια που πραγματοποιείται κάποιο από τα επόμενα φαινόμενα: ισχυροί άνεμοι, μεγάλα κύματα, υψηλή παλίρροια, συνθήκες θύελλας, και τα αποτελέσματα στις ακτές οφείλονται στην επιτάχυνση των διεργασιών μεταφοράς ιζημάτων σε περιοχές κοντά στην ακτή [181]. Το φαινόμενο αυτό αναμένεται να επιδεινωθεί στο μέλλον λόγω της υπερθέρμανσης του πλανήτη και ανόδου της στάθμης της θάλασσας. Λόγω των παραπάνω, η πρόβλεψη της διάβρωσης των ακτών (από μορφολογικής απόψεως) σε περίοδο καταιγίδων είναι ένα επείγον ζήτημα. Λόγω αυτού, έχουν αναπτυχθεί διάφορα αριθμητικά μοντέλα που είναι σε θέση να προσομοιώνουν τις υδροδυναμικές και μορφοδυναμικές διεργασίες που συμβαίνουν στις ακτές, σε περίοδο καταιγίδων [182-185]. Ωστόσο, η εφαρμογή αυτών των αριθμητικών μοντέλων υπόκειται σε ορισμένους περιορισμούς, όπως: (α) οι μη γραμμικότητες, λόγω υδροδυναμικών, ιζηματολογικών και μορφολογικών δεδομένων της κάθε περιοχής, και (β) το υψηλό υπολογιστικό κόστος.

Περιοχή και μέθοδος συλλογής πρωταρχικών δεδομένων

Η περιοχή συλλογής πρωταρχικών δεδομένων είναι η περιοχή της Ερεσού, που βρίσκεται στη ΝΔ πλευρά της νήσου Λέσβου (Σχήμα Β.3). Το Σχήμα Β.4 δείχνει ότι κατά τη διάρκεια των τελευταίων ετών, η παραλία είναι υπό αυξανόμενη διάβρωση. Η συλλογή πρωταρχικών δεδομένων αποτελείται από τοπογραφικές και βαθυμετρικές αποτυπώσεις της περιοχής που έχουν πραγματοποιηθεί σε ετήσια βάση και δεδομένα από ένα μετεωρολογικό σταθμό που λειτουργεί από το Μάρτιο του 2009 και μετράει ωριαία ταχύτητα και κατεύθυνση του ανέμου. Επιπλέον, ένα εξελιγμένο σύστημα καταγραφής μέσω καμερών έχει εγκατασταθεί και συνέλλεξε δεδομένα κατά τη διάρκεια των ετών 2010-2015, καταγράφοντας καθημερινές αλλαγές της ακτογραμμής. Οι τοπογραφικές και βαθυμετρικές μετρήσεις (συνολικά 7 καταγραφές, 1 κάθε 1.5 μήνες) έχουν συλλεχθεί από τους 8 σταθμούς (S1-S8), που απεικονίζονται στο Σχ. Β.3, μεταξύ Σεπτεμβρίου 2010 και Σεπτεμβρίου 2014. Αυτές οι

μετρήσεις έχουν εξαχθεί από ένα RTK-DGPS (τύπος GPS μεγάλης ακριβείας) που είχε συνδεθεί με ένα μικρό σκάφος εξοπλισμένο με ένα βαθύμετρο.

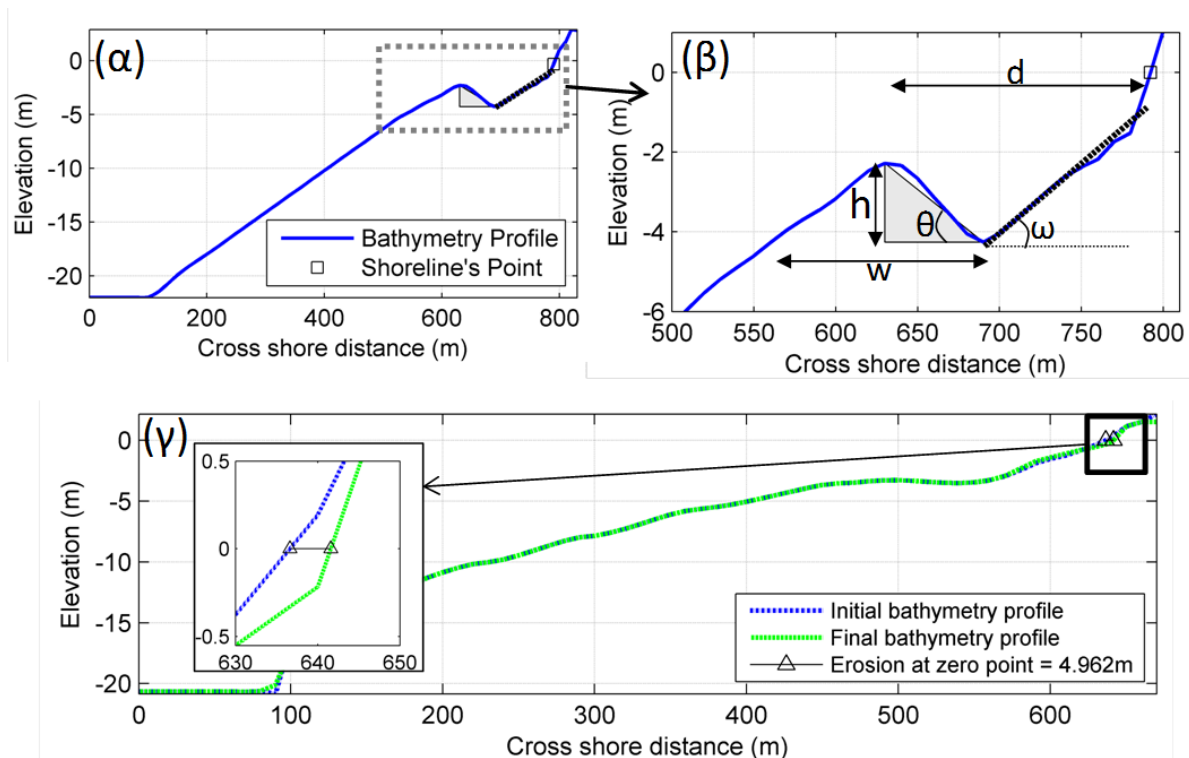


Σχήμα Β.3 Η τοποθεσία της παραλίας Ερεσού Λέσβου (αριστερά) και οκτώ τοπογραφικές τομές της παραλίας αυτής(δεξιά).



Σχήμα Β.4 Η παραλία Ερεσός της Λέσβου το Μάρτιο του 2010 (αριστερά) και το Μάρτιο του 2013 (δεξιά).

Οι βασικές παράμετροι που εμπλέκονται στην παράκτια διάβρωση είναι το προφίλ της ακτής κάθετα στην ακτογραμμή (crossshore) και η παράλληλη στην ακτή υποθαλάσσια ιζηματική μπάρα. Η ιζηματική μπάρα είναι μία υποθαλάσσια κορυφογραμμή του ιζήματος που έχει δημιουργηθεί από τα κύματα και τα ρεύματα σε μια περιοχή κοντά στη ακτή. Τέτοιου είδους μπάρες είναι γενικά παράλληλες προς την ακτή και βρίσκονται σε βάθη μικρότερα των 5-10 m. Στο Σχήμα Β.5(α) απεικονίζεται ένα προφίλ βαθυμετρίας, μιας τοπογραφικής τομής από αυτές του Σχήματος Β.3 καθώς και η θέση της ακτογραμμής (σημείο με βάθος μηδέν) και η ιζηματική μπάρα που αναφέρθηκε. Κατά τη διάρκεια του χειμώνα, οι υδροδυναμικές συνθήκες είναι πιο έντονες και η ιζηματική μπάρα κινείται από την ξηρά προς το πέλαγος. Το αντίθετο συμβαίνει κατά τη διάρκεια της θερινής περιόδου.



Σχήμα Β.5 (α) Ένα βαθυμετρικό προφίλ όπου φαίνεται μια ιζηματική μπάρα στο σχήμα (β) Η ιζηματική μπάρα σε μεγέθυνση και πέντε χαρακτηριστικά της (γ) η διάβρωση της παραλίας είναι η διαφορά των σημείων της ακτογραμμής στα δύο προφίλ.

Ο τρόπος που το βαθυμετρικό προφίλ επηρεάζει την παράκτια διάβρωση εξαρτάται από τα χαρακτηριστικά της ιζηματικής μπάρας, όπως: w = 'πλάτος ιζηματικής μπάρας', h = 'ύψος μπάρας', d = 'η απόσταση από την κορυφή της μπάρας μέχρι το σημείο με βάθος μηδέν', θ = 'η γωνία της κλίσης της μπάρας από τη μεριά της ακτής', και ω = 'η κλίση της βαθυμετρίας από το σημείο μηδέν έως την βάση της μπάρας'. Τα χαρακτηριστικά αυτά απεικονίζονται στο Σχ. Β.5 (β). Τα κύματα 'σπάνε' κοντά στην ακτή όταν το βάθος γίνεται αρκετά μικρό, δηλαδή στη περίπτωση που μελετάμε στο σημείο όπου βρίσκεται η ιζηματική μπάρα. Τα υψηλότερα κύματα που παράγονται κατά τη διάρκεια μιας έντονης καταιγίδας σπάνε σε αυτό το σημείο και όχι στην ξηρά, όπως θα συνέβαινε αν η μπάρα δεν υπήρχε, και το ποσό της ενέργειας των κυμάτων φθάνει μειωμένο στην ακτογραμμή. Ως εκ τούτου, μια ιζηματική μπάρα μπορεί να θεωρηθεί ως η «φυσική άμυνα» των ακτών από τη διάβρωση. Στο παρόν κείμενο, χρησιμοποιούνται οι έννοιες 'αρχικό προφίλ βαθυμετρίας' και 'τελικό προφίλ βαθυμετρίας' για να υποδηλώσουν τα βαθυμετρικά προφίλ πριν και μετά από μία ή περισσότερες καταιγίδες. Η απόσταση μεταξύ των σημείων της ακτογραμμής στις δύο αυτές περιπτώσεις

ορίζει την παράκτια διάβρωση. Το Σχήμα Β.5 (γ) δείχνει ένα τέτοιο παράδειγμα, όπου η αντίστοιχη διάβρωση είναι ίση με 4.9 μέτρα.

Οι παράμετροι ως τώρα οι οποίες καθορίζουν και ποσοτικοποιούν τα χαρακτηριστικά ενός βαθυμετρικού προφίλ είναι οι :

$$x_1 = h = \text{το ύψος αυτής της μπάρας,}$$

$$x_2 = \frac{w}{2} = \text{το μισό πλάτος της μπάρας}$$

$$x_3 = d = \text{η απόσταση της κορυφής της μπάρας από το σημείο που δείχνει το όριο θάλασσας-στεριάς}$$

$$x_4 = \tan \theta \text{ όπου } \theta \text{ η γωνία που δείχνει την κλίση αυτής της μπάρας}$$

$$x_5 = \tan \omega \text{ όπου } \omega \text{ η γωνία που δείχνει την κλίση του βυθού μεταξύ του ορίου της ακτογραμμής και της βάσης της ιζηματικής μπάρας}$$

Αυτές οι παράμετροι εξήχθησαν από τις τοπογραφικές και βαθυμετρικές μετρήσεις που έχουν γίνει στους σταθμούς S1-S8 (Σχήμα Β.3). Επιπρόσθετα, χρησιμοποιήθηκαν τρεις ακόμη παράμετροι οι οποίες χαρακτηρίζουν τα χαρακτηριστικά μιας καταγίδας και έχουν εξαχθεί από τα δεδομένα που παρείχε ένας μετεωρολογικός σταθμός που είχε εγκατασταθεί στην παραλία της Ερεσού Λέσβου. Οι τρεις παράμετροι αυτές είναι οι:

$$x_6 = \text{η διάρκεια (σε ώρες) μιας καταγίδας}$$

$$x_7 = \text{το σημαντικό ύψος κύματος } H_s$$

$$x_8 = \text{η κυματική περίοδος με τη μέγιστη ενέργεια (peak wave period) } T_p$$

Ως τρόπος αναγνώρισης μιας καταγίδας μέσα από μετεωρολογικές καταγραφές χρησιμοποιήθηκε η μέθοδος των Sánchez-Arcilla et al. [186] ενώ ο υπολογισμός των H_s και T_p από μετεωρολογικά δεδομένα έγινε με τη μέθοδο που περιγράφεται από τους Chatzipavlis et al.[187]. Επιπρόσθετα χρησιμοποιήθηκε και η παράμετρος :

$$x_9 = \text{η διάμεσος του μέγεθος των κόκκων του ιζήματος (D}_{50})$$

η οποία χαρακτηρίζει το ίζημα μιας παραλίας.

Οι παράμετροι $x_1 - x_9$ είναι τα δεδομένα εισόδου και η παράμετρος $y = \text{'διάβρωση παραλίας (σε μέτρα)'} (βλ. Σχήμα Β.5(γ))$ είναι η μεταβλητή εξόδου του συνόλου δεδομένων Coastal Erosion. Το σύνολο δεδομένων αποτελείται από $N = 588$ δεδομένα εισόδου $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{k9}]^T$ όπου $k = 1, 2, \dots, N$. Τα δεδομένα εξόδου y παρήχθησαν από το

λογισμικό XBeach όταν σε αυτό δόθηκαν τα δεδομένα εισόδου x_k που προαναφέρθηκαν. Η εφαρμογή του αλγορίθμου XBeach επιστρέφει τη βαθυμετρία μιας παραλίας μετά από το πέρασμα μιας καταιγίδας, οπότε τα δεδομένα εξόδου y υπολογίζονται από τη διαφορά των δύο βαθυμετρικών προφίλ στο ύψος της επιφάνειας της θάλασσας (Σχήμα Β.5 (γ)).

Το σύνολο δεδομένων Coastal Erosion χρησιμοποιείται στην παράγραφο 3.6 και στην αναφορά [161] και είναι διαθέσιμα και στην ιστοσελίδα <http://isla.aegean.gr/news.php?>.

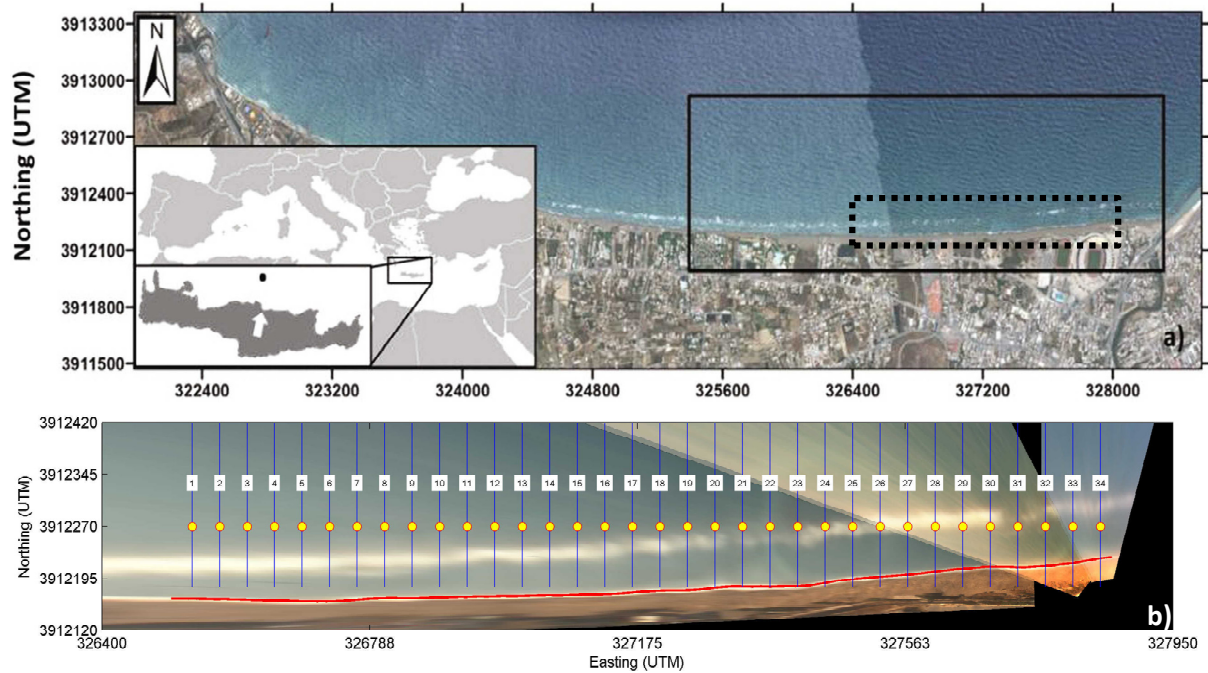
B.3 Τα δεδομένα για την Περιστροφή μιας Παραλίας (Beach Rotation)

Περιγραφή του φαινομένου της περιστροφής μιας Παραλίας

Το φαινόμενο Beach Rotation είναι η περιστροφή του μετώπου μιας παραλίας η οποία προκαλείται λόγω της αλλαγής της ενέργειας των κυμάτων [188]. Το φαινόμενο ελέγχεται από την αλληλεπίδραση της παράκτιας μορφολογίας και του κύματος και μπορεί να οδηγήσει σε μεγάλες αλλαγές στη θέση της ακτογραμμής και, ως εκ τούτου, σε αλλαγές της κάτοψης της παραλίας, η οποία ωστόσο δεν οδηγεί σε μακροπρόθεσμη απώλεια ή εναπόθεση παραλιακού ιζήματος. Οι παραλίες συχνά επιστρέφουν στην αρχική τους κάτοψη με τις αλλαγές αυτές να είναι συνήθως εποχιακές [189-191]. Παρά το γεγονός ότι το φαινόμενο Beach Rotation έχει μοντελοποιηθεί ως η διαδικασία μεταφοράς ιζημάτων μιας παραλίας, οι πρόσφατες έρευνες δείχνουν μια πιο πολύπλοκη ανταπόκριση των παραλιών στην κυματική ενέργεια, σύμφωνα με την οποία η παραλιακή μεταβλητότητα στις ροές ιζημάτων που κινούνται κάθετα στην ακτή μπορεί επίσης να είναι σημαντικές [191, 192]. Οι διαδικασίες του φαινομένου Beach Rotation είναι πιο περίπλοκες στην περίπτωση παραλιών που έχουν φυσικούς ή τεχνητούς υφάλους [193], καθώς η δυναμική των ιζημάτων και οι μορφοδυναμικές αυτών των παραλιών ελέγχονται επίσης από το βάθος και τη μορφολογία του υφάλου. Οι μετασχηματισμοί και το σπάσιμο των κυμάτων πάνω στον ύφαλο μπορεί να προκαλέσουν υψηλά μη γραμμικές επιδράσεις [174, 194]. Ως αποτέλεσμα, οι συνηθισμένες μεθοδολογίες μοντελοποίησης απαιτούν πολύπλοκες μαθηματικές δομές με εξαιρετικά υψηλό υπολογιστικό κόστος [189, 190, 192, 195].

Περιοχή και μέθοδος συλλογής πρωταρχικών δεδομένων

Η περιοχή συλλογής των πρωταρχικών δεδομένων είναι το ανατολικό τμήμα της παραλίας της Αμμουδάρας, μιας παραλίας 6.1 χιλιομέτρων που βρίσκεται στα δυτικά του λιμανιού του Ηρακλείου της Κρήτης (Σχήμα B.6). Η παραλία αυτή βρίσκεται πίσω από έναν ύφαλο τύπου beachrock (παραλιακός ψαμμίτης), ο οποίος είναι σχεδόν παράλληλος με την ακτή, και έχει πλάτος 15-50 m και η απόστασή του από την ακτογραμμή είναι 40-70 m.



Σχήμα Β.6 . (α) Η παραλία της Αμμουδάρας στο Ηράκλειο Κρήτης. Η θέση της σημαδούρας τύπου buoy παρουσιάζεται ως μαύρη κουκίδα. Το ορθογώνιο πλαίσιο δείχνει την περιοχή καταγραφής του συστήματος καμερών και το διακεκομμένο ορθογώνιο πλαίσιο την περιοχή για την οποία υπάρχει και θέση της ακτογραμμής (β) Οι 34 τοπογραφικές τομές και η θέση μιας ακτογραμμής στην περιοχή συλλογής δεδομένων, ως φόντο είναι εικόνα που έχουν συλλεχθεί από το σύστημα των καμερών.

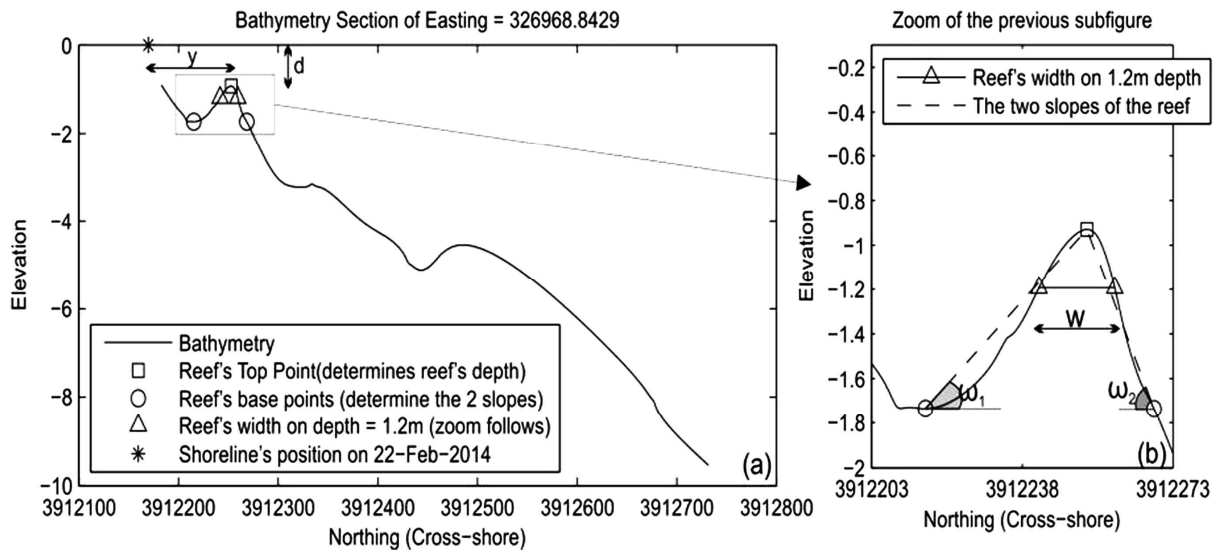
Στο Σχήμα Β.6 (β) παρουσιάζεται μια εικόνα που έχει συλλεχθεί από την περιοχή μελέτης και πάνω της φαίνονται οι 34 κάθετες στην ακτογραμμή τοπογραφικές τομές που χρησιμοποιήθηκαν στα παρόν πειράματα. Κάθε τέτοια τομή έχει δεδομένα από ένα προφίλ βαθυμετρίας, όπως αυτό που παρουσιάζεται στο Σχήμα Β.7 (α). Είναι ευρέως αποδεκτό ότι τα ειδικά μορφολογικά χαρακτηριστικά ενός υφάλου μπορούν να επηρεάσουν το φαινόμενο Beach Rotation. Ο ύφαλος ενεργεί με τρόπο παρόμοιο με ένα βυθισμένο κυματοθραύστη, απορροφώντας την ενέργεια του κύματος [174, 195, 196]. Τα αριθμητικά δεδομένα που προκύπτουν από τη μορφολογία του υφάλου και χρησιμοποιούνται στο σύνολο δεδομένων Beach Rotation είναι (Σχήμα Β.7) :

d : το βάθος του υφάλου (σε μέτρα) από την επιφάνεια της θάλασσας

ω_1 και ω_2 οι κλίσεις του υφάλου από τη μεριά της ακτής και τη μεριά του πελάγους αντίστοιχα

w : το πλάτους του υφάλου σε βάθος 1.2 μέτρα από την επιφάνεια της θάλασσας.

Η επιλογή του βάθος των 1.2 μέτρων αποφασίστηκε μετά από μια εξειδικευμένη επεξεργασία των δεδομένων που έδειξε ότι το πλάτος του υφάλου σε αυτό το βάθος του νερού είχε τις πιο σημαντικές επιπτώσεις.

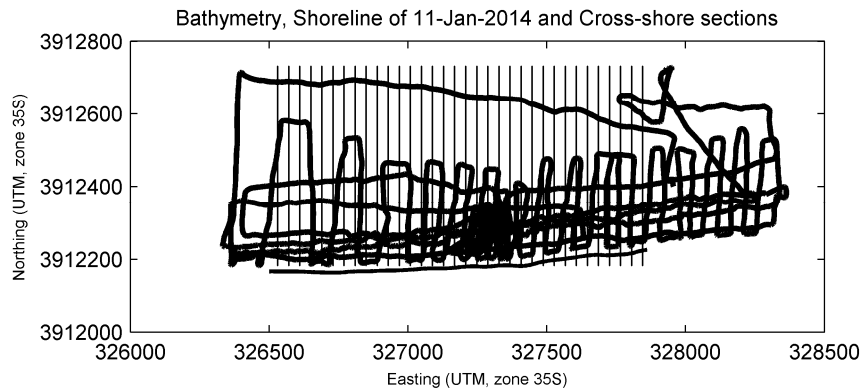


Σχήμα Β.7 (α) Μια κάθετη στην ακτογραμμή τοπογραφική τομή Cross-shore bathymetric profile (βλ. σχήμα Β.6 (b)), η οποία δείχνει τον υφάλου (β) Μεγέθυνση του υφάλου και χαρακτηριστικά της μορφολογίας του.

Εκτός από τις παραπάνω παραμέτρους που ποσοτικοποίησαν τα χαρακτηριστικά της βαθυμετρίας του υφάλου, χρησιμοποιήθηκαν δύο ακόμη παράμετροι που περιγράφουν τις κυματικές συνθήκες, το σημαντικό ύψος κύματος* που συμβολίζεται ως H_s (σε μέτρα), και η κυματική περίοδος με τη μέγιστη ενέργεια (peak wave period[†]) που συμβολίζεται ως T_p (σε δευτερόλεπτα). Να σημειωθεί ότι τα δύο τελευταία είναι δύο σημαντικά μεγέθη τα οποία ελέγχουν άμεσα τη μορφοδυναμική της παραλίας [188, 190, 194]. Όλες οι παραπάνω παράμετροι αποτελούν τα δεδομένα εισόδου του συνόλου δεδομένων Beach Rotation. Ως μεταβλητή εξόδου χρησιμοποιήθηκε η απόσταση (σε μέτρα) από την κορυφή του υφάλου μέχρι την ακτογραμμή και συμβολίζεται με y όπως και στο Σχήμα Β.7 (α). Η μεταβλητότητα αυτής της παραμέτρου y (απόσταση cross-shore) ορίζει το φαινόμενο Beach Rotation [188-191, 194]. Εν ολίγοις, οι μεταβλητές εισόδου είναι: $x_1 = d$, $x_2 = \tan \omega_1$, $x_3 = \tan \omega_2$, $x_4 = w$, $x_5 = H_s$, $x_6 = T_p$ και η μεταβλητή εξόδου είναι η y .

* Υπολογίζεται από χρονοσειρές κυμάτων και είναι ο μέσος όρος του 1/3 των ψηλότερων κυμάτων.

† Υπολογίζεται μέσω φασματικής ανάλυσης μιας χρονοσειράς της κατανομής μεμονωμένων κυμάτων.



Σχήμα Β.8 Η πορεία του σκάφους που σύλλεξε τα βαθυμετρικά δεδομένα, η θέση των 34 τοπογραφικών τομών και η θέση μιας ακτογραμμής (πάνω).

Η πειραματική μεθοδολογία συλλογής των δεδομένων αποτελείται από υψηλής ακρίβειας βαθυμετρικά δεδομένα κοντά στην ακτή και από μια μεγάλης διάρκειας χρονοσειρά (δεδομένα 10 μηνών από Ιαν.2014-Νοέ.2014) της θέσης της ακτογραμμής και των κυματικών συνθηκών. Πιο συγκεκριμένα, τα βαθυμετρικά δεδομένα ελήφθησαν μέσω ενός βυθόμετρου Hi-Target HD 370 και ενός διαφορικού GPS (Topcon Hipper RTK-DGPS) που είχαν τοποθετηθεί σε ένα φουσκωτό σκάφος. Χρησιμοποιώντας μεθόδους παρεμβολής και συγκεκριμένα με τη μέθοδο Biharmonic spline interpolation [197], από αυτά τα βαθυμετρικά δεδομένα προήλθαν 34 τοπογραφικές τομές σαν αυτή που παρουσιάστηκε στο Σχήμα Β.6(β) και Β.7. Στο Σχήμα Β.8 φαίνεται η θέση των δεδομένων που έχει συλλέξει το βαθύμετρο, η θέση των 34 τομών και η θέση μίας ακτογραμμής. Η θέση της ακτογραμμής για την περίοδο των 10 μηνών λήφθηκε από ένα σύστημα που αποτελείται από 3 PointGrey FLEA-2 κάμερες που έχουν εγκατασταθεί στην περιοχή μελέτης, και καταγράφουν την παραλία σε μήκος 1400 m (βλ. Σχήμα Β.6(b)). Μια λεπτομερής περιγραφή του συστήματος των καμερών και της διαδικασίας που αναπτύχθηκε για την εξαγωγή της ακτογραμμής από τις εικόνες παρέχεται στην [174]. Τα παραπάνω πειράματα παρείχαν τα δεδομένα για τις μεταβλητές εισόδου $x_1 - x_4$ και για τη μεταβλητή εξόδου y . Τα δεδομένα των μεταβλητών x_5, x_6 λήφθηκαν από μια κυματική σηματοδότη τύπου buoy (POSEIDON E1-M3A), που βρίσκεται περίπου 35 χιλιόμετρα βόρεια της παραλίας (στις συντεταγμένες 35.66° N και 24.99° E) σε βάθος 1440 m (Σχήμα Β.6(a)), που έχει εγκατασταθεί και διαχειρίζεται από το Εθνικό Κέντρο Θαλασσιών Ερευνών (ΕΛΚΕΘΕ).

Συνολικά τα πειράματα παρήγαγαν $N = 4148$ δεδομένα εισόδου-εξόδου της μορφής $\{\mathbf{x}_k; y_k\}$ όπου $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{k6}]^T$ τα οποία χρησιμοποιήθηκαν στην [168] και στην παράγραφο 4.6

του παρόντος κειμένου. Το πλήθος 4148 προκύπτει από το γινόμενο των 34 τοπογραφικών τομών επί 122 ημερήσιες θέσεις ακτογραμμής (από την περίοδο των 10 μηνών).

Παράρτημα Γ

Δημοσιεύσεις

Ακολουθεί λίστα με τις δημοσιεύσεις σε διεθνή περιοδικά και συνέδρια (με κριτές) που έγιναν κατά τη διάρκεια της παρούσας διδακτορικής διατριβής.

Άρθρα σε διεθνή περιοδικά:

- Monioudi I., Velegrakis A.F., Chatzipavlis A., Rigos A., Karambas Th., Vousdoukas M.I., Hasiotis Th., Koukouroufli N., Manoutsoglou E., Peduzzi P., Poulos S. and Collins M., " Assessment of beach erosion due to sea level rise: The case of the Aegean Archipelago (Eastern Mediterranean) " , Natural Hazards and Earth System Sciences (NHES), DOI: 10.1007/s11069-016-2415-9
- A. Tamvakis, G.E. Tsekouras, A. Rigos, C. Kalloniatis, C.-N. Anagnostopoulos, and G. Anastassopoulos , A Methodology to Carry Out Voting Classification Tasks Using a Particle Swarm Optimization-Based Neuro-Fuzzy Competitive Learning Network , Evolving Systems , doi: 10.1007/s12530-016-9160-x
- Anastasios Rigos, George E. Tsekouras , Michalis I. Vousdoukas, Antonios Chatzipavlis, Adonis Velegrakis, "A Chebyshev polynomial RBF neural network for automated shoreline extraction from coastal imagery", Integrated Computer-Aided Engineering, vol. 23, no. 2, pp. 141-160, 2016 [DOI: 10.3233/ICA-150507]
- Dimitris Papaioannou, Antonis Gakis, Nikolaos Athanasoulis, Anastasios Rigos, Alexandra A. Mamali, "A review of urban sustainability criteria under Global Warming stress", Interdisciplinary Environmental Review., DOI: <http://dx.doi.org/10.1504/IER.2015.069405>
- Anastasios Rigos, Michalis I. Vousdoukas, Olympos P. Andreadis, George E. Tsekouras, and Adonis F. Velegrakis, On the systematic implementation of artificial neural networks in

the classification of variance images and shoreline extraction, Fresenius Environmental Bulletin, Volume 23 – No 11. 2014

- Dimitrios Tsolakis, George E. Tsekouras, Antonios D. Niros, and Anastasios Rigos. 2012. On the systematic development of fast fuzzy vector quantization for grayscale image compression, Neural Networks 36 (December 2012), 83-96. DOI : 10.1016/j.neunet.2012.09.009

Άρθρα σε συνέδρια :

- Anastasios Rigos, George E. Tsekouras, Antonios Chatzipavlis, Adonis F.Velegrakis, "Modeling Beach Rotation Using a Novel Legendre Polynomial Feedforward Neural Network Trained by Nonlinear Constrained Optimization", 12th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2016), September 2016, Thessaloniki, Greece
- A. Karditsa, S. Poulos, A. Velergakis, O. Andreadis, A. Rigos, G. Alexandrakis, S. Peirakis and G. Ghionis, "Development of an inventory of Hellenic beaches ", 41st CIESM Congress, Kiel, Germany, 12 September - 16 September 2016
- Monioudi, I.N , Chatzipavlis, A., Psarros, F., Rigos, A., Andreadis, O., Hasiotis, T., Velegrakis, A.F. " Evaluation of beach erosion risk of the Aegean Archipelago islands ", ISISA Islands of the World XIV Conference 2016 : Niss(i)ology and Utopia: back to the roots of Island Studies, 23 - 27 May 2016, Mytilene, Lesvos Island, Greece
- E. Oikonomou, P. Drakopoulos, O. Andreadis, T. Hasiotis, F. Psarros, A. Rigos, V. Trigonis, E. Vasillakis , S. Poulos and D. Pafitis, " Comparison of bathymetry extracting models from WorldView 2 imagery in Zakynthos island, Greece, and the effect of deglinting methods in the depth estimation accuracy", Fourth International Conference on Remote Sensing and Geoinformation of Environment' 4-8 April, 2016 - Cyprus, (RSCy2016)
- George E. Tsekouras, Anastasios Rigos, Antonios Chatzipavlis, and Adonis Velegrakis, A Neural-Fuzzy Network Based on Hermite Polynomials to Predict the Coast Erosion, 16th

International Conference on Engineering Applications of Neural Networks (EANN 2015),
Island of Rhodes, Greece, 25-28 September 2015

- Anastasios Rigos , Aristidis Vaiopoulos, George Skianis, George Tsekouras, and Panos Drakopoulos, A novel approach for automated shoreline extraction from remote sensing images using low level programming, Geophysical Research Abstracts Vol. 17, EGU2015, 2015
- Rigos A., Kokkoris G.D., Velegrakis A., Hatzopoulos J. N. , Andreadis O., Trygonis, V., “ Beach database of the Aegean Archipelago” , 11ο Πανελλήνιο Συμπόσιο Ωκεανογραφίας & Αλιείας ,13-17 Μαΐου 2015 , Μυτιλήνη ,pp. 1085-1088
- Βελεγράκης Α.Φ., Μονιούδη Ι.Ν., Χατζηπαυλής Α., Χατζόπουλος Ι.Ν., Ρήγος Α. και Ψαρρός Φ., 2015. Ανάπτυξη Μεθοδολογίας για την Εκτίμηση της Οπισθοχώρησης των Νησιωτικών Παραλιών του Αιγαίου Πελάγους κάτω από Αλλαγές της Θαλάσσιας Στάθμης, 11ο Πανελλήνιο Συμπόσιο Ωκεανογραφίας & Αλιείας, Μυτιλήνη, Λέσβος, Ελλάδα, 13-17 Μαΐου 2015, pp. 1089-1092
- Trygonis, V., Ghionis, G. Andreadis, O., Vousdoukas, M., Ntemogiannis, I., Rigos, A, Psarros, F., Velegrakis A., Hasiotis, T. and Poulos, S.E., “Monitoring beach attendance of Paralía Katerini, Greece, using a coastal imaging system”, 11ο Πανελλήνιο Συμπόσιο Ωκεανογραφίας & Αλιείας ,13-17 Μαΐου 2015 , Μυτιλήνη, pp. 737-740
- Χατζηπαυλής, Α., Καραμπάς, Θ., Ανδρεάδης, Ο., Βελεγράκης Α.Φ. και Ρήγος Α., 2014. Προσομοίωση των Υδρο-Μορφοδυναμικών Κυματογενών Διεργασιών στην Παραλία Σκάλας Ερεσού. Πρακτικά Πανελλήνιου Συνεδρίου «Η Χρήση των Νέων Τεχνολογιών στην Πρόληψη και τη Διαχείριση Φυσικών Καταστροφών – Ο Ρόλος της Πολιτικής Προστασίας» (SafeRhodes 2014), Ρόδος, Ελλάδα, 24-26 Οκτώβρη 2014, 12 σελ.
- Anastasios Rigos, Olympos P. Andreadis, Manousakis Andreas, Michalis I. Vousdoukas, George E. Tsekouras, and Adonis Velegrakis , Shoreline Extraction from Coastal Images Using Chebyshev Polynomials and RBF Neural Networks, 10th International Conference on

Artificial Intelligence Applications and Innovations (AIAI 2014) Island of Rhodes, Greece, September 19-21, 2014, pp.593-603

- Velegrakis, A.F., Monioudi, I.N., Chatzipavlis, A., Hatzopoulos, J.N., Hasiotis, Th., Psarros, F., Rigos, A., Vousdoukas, M.I., Lekkas, D., Ghionis, G., Peduzzi, P., Chatenoux, B. . Development of a methodology for the estimation of beach vulnerability to sea level rise - an application to the Aegean Sea . 19th International Sedimentological Congress: at the crossroad of new frontiers, 18.08 - 22.08.2014, Geneva, Switzerland.ICS 2014 Sedimentology. (Poster Presentation).
- Anastasios Rigos, Michalis Ioannis Vousdoukas, George E. Tsekouras, Adonis Velegrakis , On the systematic implementation of artificial neural networks in the classification of variance images and shoreline extraction , Fourth International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE 2013) and SECOTOX Conference, Mykonos island, Greece, June 24 to 28, 2013., pp.536-541
- A. F. Velegrakis, I. N. Monioudi, S.E. Poulos, D. Lekkas, J. N. Hatzopoulos, Th. Hasiotis, Th. Karambas, A. Kontogianni, M. I. Vousdoukas, O. Andreadis, A. Rigos, G. Tsekouras, F. Psarros, G. Ghionis, A. Karditsa, A. Chatzipavlis, S. Anastasiadis, G. Kokkoris, S. Petrakis, M. N. Tsimplis, P. Peduzzi, M. B. Collins, Physical and economic impacts of sea level changes on the island beaches of the Aegean Archipelago (ISLA): Concepts and preliminary results. 13th International Conference on Environmental Science and Technology (CEST2013), 5-7 September 2013, Athens, paper ID: cest13_0819.

Βιβλιογραφία

- [1] Rigos, A., G.E. Tsekouras, M.I. Vousdoukas, A. Chatzipavlis, and A.F. Velegrakis, *A Chebyshev polynomial radial basis function neural network for automated shoreline extraction from coastal imagery* Integrated Computer-Aided Engineering, 2016. **23**(2): p. 141-160.
- [2] Ivakhnenko, A. and G. Ivakhnenko, *The review of problems solvable by algorithms of the group method of data handling (GMDH)*. Pattern Recognition And Image Analysis C/C Of Raspoznavaniye Obrazov I Analiz Izobrazhenii, 1995. **5**: p. 527-535.
- [3] Oh, S.-K. and W. Pedrycz, *The design of self-organizing neural networks based on PNs and FPNs with the aid of genetic optimization and extended GMDH method*. International Journal of Approximate Reasoning, 2006. **43**(1): p. 26-58.
- [4] Lee, T.-T. and J.-T. Jeng, *The Chebyshev-polynomials-based unified model neural networks for function approximation*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1998. **28**(6): p. 925-935.
- [5] Oh, S.-K., W. Pedrycz, and B.-J. Park, *Polynomial neural networks architecture: analysis and design*. Computers & Electrical Engineering, 2003. **29**(6): p. 703-725.
- [6] Oh, S.-K., W.-D. Kim, W. Pedrycz, and B.-J. Park, *Polynomial-based radial basis function neural networks (P-RBF NNs) realized with the aid of particle swarm optimization*. Fuzzy Sets and Systems, 2011. **163**(1): p. 54-77.
- [7] Oh, S.-K. and W. Pedrycz, *Self-organizing polynomial neural networks based on polynomial and fuzzy polynomial neurons: analysis and design*. Fuzzy Sets and Systems, 2004. **142**(2): p. 163-198.
- [8] Huang, W., S.-K. Oh, and W. Pedrycz, *Design of hybrid radial basis function neural networks (HRBFNNs) realized with the aid of hybridization of fuzzy clustering method (FCM) and polynomial neural networks (PNNs)*. Neural Networks, 2014. **60**: p. 166-181.
- [9] Oh, S.-K., W.-D. Kim, W. Pedrycz, and S.-C. Joo, *Design of K-means clustering-based polynomial radial basis function neural networks (pRBF NNs) realized with the aid of particle swarm optimization and differential evolution*. Neurocomputing, 2012. **78**(1): p. 121-132.
- [10] Er, M.J., W. Chen, and S. Wu, *High-speed face recognition based on discrete cosine transform and RBF neural networks*. IEEE Transactions on Neural Networks, 2005. **16**(3): p. 679-691.
- [11] Machado Filho, O.M. and N.F. Ebecken, *Visual RBF network design based on Star Coordinates*. Advances in Engineering Software, 2009. **40**(9): p. 913-919.
- [12] Meghabghab, G. and A. Kandel, *Stochastic simulations of web search engines: RBF versus second-order regression models*. Information Sciences, 2004. **159**(1): p. 1-28.
- [13] Hu, T. and S.Y. Sung, *Predicting Spatial Data with RBF Networks*. International journal of neural systems, 2004. **14**(02): p. 117-123.
- [14] Qiao, J.-F. and H.-G. Han, *A repair algorithm for radial basis function neural network and its application to chemical oxygen demand modeling*. International Journal of Neural Systems, 2010. **20**(01): p. 63-74.
- [15] Wu, D., K. Warwick, Z. Ma, M.N. Gasson, J.G. Burgess, S. Pan, and T.Z. Aziz, *Prediction of Parkinson's disease tremor onset using a radial basis function neural network based on particle swarm optimization*. International journal of neural systems, 2010. **20**(02): p. 109-116.
- [16] Zhi-Ming, Z., L. Rong-Yan, and W. Lei. *Adaptive watermark scheme with RBF neural networks*. in *Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference on*. 2003. IEEE.
- [17] Andréasson, N., A. Evgrafov, and M. Patriksson, *An Introduction to Continuous Optimization*. 2005: Professional Publishing Svc.

- [18] Armijo, L., *Minimization of functions having Lipschitz continuous first partial derivatives*. Pacific Journal of Mathematics, A Non-profit Corporation, 1966. **16**(1): p. 1-3.
- [19] Nocedal, J. and S. Wright, *Numerical Optimization*. 2006: Springer New York.
- [20] Ma, L. and K. Khorasani, *Constructive feedforward neural networks using Hermite polynomial activation functions*. IEEE Transactions on Neural Networks, 2005. **16**(4): p. 821-833.
- [21] Ma, L. and K. Khorasani. *Adaptive structure feed-forward neural networks using polynomial activation functions*. in *Applications and Science of Computational Intelligence III*. 2000.
- [22] Beliczynski, B., *A method of multivariable Hermite basis function approximation*. Neurocomputing, 2012. **96**: p. 12-18.
- [23] Shi, X., S. Duan, L. Wang, T. Huang, and C. Li, *A novel memristive electronic synapse-based Hermite chaotic neural network with application in cryptography*. Neurocomputing, 2015. **166**: p. 487-495.
- [24] Linh, T.H., S. Osowski, and M. Stodolski. *On-line heart beat recognition using Hermite polynomials and neuro-fuzzy network*. in *Instrumentation and Measurement Technology Conference, 2002. IMTC/2002. Proceedings of the 19th IEEE*. 2002. IEEE.
- [25] Rasiah, A., R. Togneri, and Y. Attikiouzel, *Modelling 1-D signals using Hermite basis functions*. IEE Proceedings-Vision, Image and Signal Processing, 1997. **144**(6): p. 345-354.
- [26] Patra, J.C., P.K. Meher, and G. Chakraborty, *Nonlinear channel equalization for wireless communication systems using Legendre neural networks*. Signal Processing, 2009. **89**(11): p. 2251-2262.
- [27] Patra, J.C. and C. Bornand. *Nonlinear dynamic system identification using Legendre neural network*. in *The 2010 international joint conference on neural networks (IJCNN)*. 2010. IEEE.
- [28] Nanda, S.K., D.P. Tripathy, and S. Mahapatra, *Application of Legendre neural network for air quality prediction*. 2011.
- [29] Liu, F. and J. Wang, *Fluctuation prediction of stock market index by Legendre neural network with random time strength function*. Neurocomputing, 2012. **83**: p. 12-21.
- [30] Zhu, G. and S. Kwong, *Gbest-guided artificial bee colony algorithm for numerical function optimization*. Applied Mathematics and Computation, 2010. **217**(7): p. 3166-3173.
- [31] Sundaram, R.K., *A First Course in Optimization Theory*. 1996: Cambridge University Press.
- [32] Anton, H. and C. Rorres, *Elementary Linear Algebra with Applications*. 2005: Wiley.
- [33] Chong, E.K.P. and S.H. Zak, *An Introduction to Optimization*. 2013: Wiley.
- [34] Kröse, B. and P. van der Smagt, *An introduction to neural networks*. 1996.
- [35] Hassanien, A.E. and E. Emary, *Swarm Intelligence: Principles, Advances, and Applications*. 2016: CRC Press.
- [36] Panigrahi, B.K., Y. Shi, and M.H. Lim, *Handbook of Swarm Intelligence: Concepts, Principles and Applications*. 2011: Springer Berlin Heidelberg.
- [37] Bonabeau, E., M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*. 1999: OUP USA.
- [38] Kennedy, J. and R. Eberhart. *Particle swarm optimization*. in *Neural Networks, 1995. Proceedings., IEEE International Conference on*. 1995.
- [39] Karaboga, D., *An idea based on honey bee swarm for numerical optimization*. 2005, Technical report-tr06, Erciyes university, engineering faculty, computer engineering department.
- [40] Zadeh, L.A., *Fuzzy sets*. Information and control, 1965. **8**(3): p. 338-353.
- [41] Yang, X.-S., *A new metaheuristic bat-inspired algorithm*, in *Nature inspired cooperative strategies for optimization (NICSO 2010)*. 2010, Springer. p. 65-74.
- [42] Li, X.-l., Z.-j. Shao, and J.-x. Qian, *An optimizing method based on autonomous animats: fish-swarm algorithm*. System Engineering Theory and Practice, 2002. **22**(11): p. 32-38.
- [43] Tang, R., S. Fong, X.-S. Yang, and S. Deb. *Wolf search algorithm with ephemeral memory*. in *Digital Information Management (ICDIM), 2012 Seventh International Conference on*. 2012. IEEE.

- [44] Eberhart and S. Yuhui. *Particle swarm optimization: developments, applications and resources*. in *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*. 2001.
- [45] Kennedy, J.F., J. Kennedy, R.C. Eberhart, and Y. Shi, *Swarm Intelligence*. 2001: Morgan Kaufmann Publishers.
- [46] Blum, C. and X. Li, *Swarm intelligence in optimization*, in *Swarm Intelligence*. 2008, Springer. p. 43-85.
- [47] Socha, K. and C. Blum, *An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training*. *Neural Computing and Applications*, 2007. **16**(3): p. 235-247.
- [48] Hasan, S., T.S. Quo, S.M. Shamsuddin, and R. Sallehuddin. *Artificial Neural Network Learning Enhancement Using Artificial Fish Swarm Algorithm*. in *Proceedings of the 3rd International Conference on computing and Informatics, ICOCI*. 2011.
- [49] Tsai, P.W., J.S. Pan, B.Y. Liao, M.J. Tsai, and V. Istanda. *Bat algorithm inspired algorithm for solving numerical optimization problems*. in *Applied Mechanics and Materials*. 2012. Trans Tech Publ.
- [50] Mirjalili, S., S.M. Mirjalili, and A. Lewis, *Grey wolf optimizer*. *Advances in Engineering Software*, 2014. **69**: p. 46-61.
- [51] Yamany, W., E. Emary, and A.E. Hassanien. *Wolf search algorithm for attribute reduction in classification*. in *Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on*. 2014. IEEE.
- [52] Qasem, S.N. and S.M.H. Shamsuddin. *Improving performance of radial basis function network based with particle swarm optimization*. in *2009 IEEE Congress on Evolutionary Computation*. 2009. IEEE.
- [53] Shao, Y., Q. Chen, and H. Jiang. *RBF neural network based on particle swarm optimization*. in *International Symposium on Neural Networks*. 2010. Springer.
- [54] Tsekouras, G.E. and J. Tsimikas, *On training RBF neural networks using input-output fuzzy clustering and particle swarm optimization*. *Fuzzy Sets and Systems*, 2013. **221**: p. 65-89.
- [55] Karaboga, D. and B. Akay, *A comparative study of Artificial Bee Colony algorithm*. *Appl. Math. Comput.*, 2009. **214**(1): p. 108-132.
- [56] Song, X., H. Gu, L. Tang, S. Zhao, X. Zhang, L. Li, and J. Huang, *Application of artificial bee colony algorithm on surface wave data*. *Computers & Geosciences*, 2015. **83**: p. 219-230.
- [57] Cruz, D.P.F., R.D. Maia, A. Szabo, and L.N.d. Castro. *A bee-inspired algorithm for optimal data clustering*. in *2013 IEEE Congress on Evolutionary Computation*. 2013.
- [58] Ferreira Cruz, D.P., R. Dourado Maia, L.A. da Silva, and L.N. de Castro, *BeeRBF: A bee-inspired data clustering approach to design RBF neural network classifiers*. *Neurocomputing*, 2016. **172**: p. 427-437.
- [59] Karaboga, D., B. Akay, and C. Ozturk, *Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Networks*, in *Modeling Decisions for Artificial Intelligence: 4th International Conference, MDAI 2007, Kitakyushu, Japan, August 16-18, 2007. Proceedings*, V. Torra, Y. Narukawa, and Y. Yoshida, Editors. 2007, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 318-329.
- [60] Karaboga, D. and B. Akay, *Artificial Bee Colony (ABC) Algorithm on Training Artificial Neural Networks*. 2007 IEEE 15th Signal Processing and Communications Applications, 2007.
- [61] Karaboga, D., B. Gorkemli, C. Ozturk, and N. Karaboga, *A comprehensive survey: artificial bee colony (ABC) algorithm and applications*. *Artificial Intelligence Review*, 2014. **42**(1): p. 21-57.
- [62] Legendre, A.M., *Recherches sur l'attraction des spheroides homogenes*, 1876.
- [63] Chihara, T.S., *An Introduction to Orthogonal Polynomials*. 2014: Dover Publications.
- [64] Szegő, G., *Orthogonal Polynomials*. 1939: American Mathematical Society.
- [65] Gautschi, W., G.H. Golub, and G. Opfer, *Applications and Computation of Orthogonal Polynomials: Conference at the Mathematical Research Institute Oberwolfach, Germany March 22-28, 1998*. 2012: Birkhäuser Basel.

- [66] Haran, S.M.J., *Arithmetical Investigations: Representation Theory, Orthogonal Polynomials, and Quantum Interpolations*. 2008: Springer Berlin Heidelberg.
- [67] Ismail, M., *Classical and Quantum Orthogonal Polynomials in One Variable*. 2005: Cambridge University Press.
- [68] Nikiforov, A.F., S.K. Suslov, and V.B. Uvarov, *Classical Orthogonal Polynomials of a Discrete Variable*. 2012: Springer Berlin Heidelberg.
- [69] Jackson, D., *Fourier Series and Orthogonal Polynomials*. 2004: Dover Publications.
- [70] Dunkl, C.F. and Y. Xu, *Orthogonal Polynomials of Several Variables*. 2014: Cambridge University Press.
- [71] Nevai, P., *Orthogonal Polynomials: Theory and Practice*. 2012: Springer Netherlands.
- [72] Freud, G., *Orthogonal Polynomials*. 2014: Elsevier Science.
- [73] Dominici, D. and R.S. Maier, *Special Functions and Orthogonal Polynomials: AMS Special Session on Special Functions and Orthogonal Polynomials, April 21-22, 2007, Tucson, Arizona*. 2008: American Mathematical Society.
- [74] Fox, L. and I.B. Parker, *Chebyshev Polynomials in Numerical Analysis*. 1972: Oxford University Press.
- [75] Snyder, M.A., *Chebyshev methods in numerical approximation*. 1966: Prentice-Hall.
- [76] Rivlin, T.J., *The Chebyshev Polynomials*. 1974: Wiley.
- [77] Mason, J.C. and D.C. Handscomb, *Chebyshev Polynomials*. 2002: CRC Press.
- [78] Haykin, S., *Neural Networks: A Comprehensive Foundation*. 1998: Prentice Hall PTR. 842.
- [79] Kohavi, R. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. in *Ijcai*. 1995.
- [80] Witten, I.H., E. Frank, and M.A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. 2011: Elsevier Science.
- [81] Hecht-Nielsen, R., *Neurocomputing: picking the human brain*. IEEE Spectrum, 1988. **25**(3): p. 36-41.
- [82] Hafner, M., M. Schüler, O. Nelles, and R. Isermann, *Fast neural networks for diesel engine control design*. Control Engineering Practice, 2000. **8**(11): p. 1211-1221.
- [83] Chapin, J.K., K.A. Moxon, R.S. Markowitz, and M.A. Nicolelis, *Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex*. Nature neuroscience, 1999. **2**(7): p. 664-670.
- [84] Cochocki, A. and R. Unbehauen, *Neural networks for optimization and signal processing*. 1993: John Wiley & Sons, Inc.
- [85] Li, Y. and W. Ma. *Applications of Artificial Neural Networks in Financial Economics: A Survey*. in *Computational Intelligence and Design (ISCID), 2010 International Symposium on*. 2010.
- [86] Papik, K., B. Molnar, R. Schaefer, Z. Dombovari, Z. Tulassay, and J. Feher, *Application of neural networks in medicine-a review*. Medical Science Monitor, 1998. **4**(3): p. MT538-MT546.
- [87] Rowley, H.A., S. Baluja, and T. Kanade, *Neural network-based face detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998. **20**(1): p. 23-38.
- [88] Schalkoff, R.J., *Digital image processing and computer vision*. Vol. 286. 1989: Wiley New York.
- [89] Dony, R.D. and S. Haykin, *Neural network approaches to image compression*. Proceedings of the IEEE, 1995. **83**(2): p. 288-303.
- [90] Dahl, G.E., D. Yu, L. Deng, and A. Acero, *Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition*. IEEE Transactions on Audio, Speech, and Language Processing, 2012. **20**(1): p. 30-42.
- [91] Gong, Y., *Speech recognition in noisy environments: A survey*. Speech communication, 1995. **16**(3): p. 261-291.
- [92] Sutskever, I., J. Martens, and G.E. Hinton. *Generating text with recurrent neural networks*. in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011.

- [93] Gerson, I., O. Karaali, G. Corrigan, and N. Massey. *Neural network speech synthesis*. in *Sixth Australian International Conference on Speech Science and Technology*. 1996.
- [94] Yuanhui, Z., L. Yuchang, and S. Chunyi, *Using neural network to extract knowledge from database*, in *Principles of Data Mining and Knowledge Discovery: First European Symposium, PKDD '97 Trondheim, Norway, June 24–27, 1997 Proceedings*, J. Komorowski and J. Zytkow, Editors. 1997, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 376-383.
- [95] Agrawal, R., J. Gehrke, D. Gunopulos, and P. Raghavan, *Automatic subspace clustering of high dimensional data for data mining applications*. Vol. 27. 1998: ACM.
- [96] Berkhin, P., *A survey of clustering data mining techniques*, in *Grouping multidimensional data*. 2006, Springer. p. 25-71.
- [97] Celenk, M., *A color clustering technique for image segmentation*. Computer Vision, Graphics, and Image Processing, 1990. **52**(2): p. 145-170.
- [98] Bezdek, J.C., J. Keller, R. Krisnapuram, and N. Pal, *Fuzzy models and algorithms for pattern recognition and image processing*. Vol. 4. 2006: Springer Science & Business Media.
- [99] Wu, J., *Advances in K-means Clustering: A Data Mining Thinking*. 2012: Springer Berlin Heidelberg.
- [100] Lichman, M., *{UCI} Machine Learning Repository [http://archive.ics.uci.edu/ml]*. University of California, Irvine, School of Information and Computer Sciences, 2013.
- [101] Bezdek, J.C., *Pattern recognition with fuzzy objective function algorithms*. 1981, New York: Plenum Press.
- [102] Bezdek, J.C., R.J. Hathaway, M.J. Sabin, and W.T. Tucker, *Convergence theory for fuzzy c-means: counterexamples and repairs*. IEEE Transactions on Systems, Man, and Cybernetics, 1987. **17**(5): p. 873-877.
- [103] Pedrycz, W., *Conditional Fuzzy C-Means*. Pattern Recognition Letters, 1996. **17**(6): p. 625-631.
- [104] Pedrycz, W., *Conditional fuzzy clustering in the design of radial basis function neural networks*. IEEE Trans Neural Netw, 1998. **9**(4): p. 601-12.
- [105] Moody, J. and C.J. Darken, *Fast learning in networks of locally-tuned processing units*. Neural Comput., 1989. **1**(2): p. 281-294.
- [106] Saha, A. and J.D. Keeler, *Algorithms for better representation and faster learning in radial basis function networks*, in *Advances in neural information processing systems 2*, S.T. David, Editor. 1990, Morgan Kaufmann Publishers Inc. p. 482-489.
- [107] Niros, A.D. and G.E. Tsekouras, *A novel training algorithm for RBF neural network using a hybrid fuzzy clustering approach*. Fuzzy Sets and Systems, 2012. **193**: p. 62-84.
- [108] Branke, J. *Evolutionary algorithms for neural network design and training*. in *In Proceedings of the First Nordic Workshop on Genetic Algorithms and its Applications*. 1995. Citeseer.
- [109] Cantú-Paz, E. and C. Kamath, *An empirical comparison of combinations of evolutionary algorithms and neural networks for classification problems*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2005. **35**(5): p. 915-927.
- [110] Ilonen, J., J.-K. Kamarainen, and J. Lampinen, *Differential evolution training algorithm for feed-forward neural networks*. Neural Processing Letters, 2003. **17**(1): p. 93-105.
- [111] Montavon, G., G. Orr, and K.R. Müller, *Neural Networks: Tricks of the Trade*. 2012: Springer Berlin Heidelberg.
- [112] Pao, Y., *Adaptive pattern recognition and neural networks*. 1989.
- [113] Singh, S. and K. Srivastava, *Degree of insecurity estimation in a power system using functional link neural network*. European transactions on electrical power, 2002. **12**(5): p. 353-359.
- [114] Misra, B. and S. Dehuri, *Functional Link Artificial Neural Network for Classification Task in Data Mining 1*. 2007.
- [115] Patra, J.C. and R.N. Pal, *A functional link artificial neural network for adaptive channel equalization*. Signal Processing, 1995. **43**(2): p. 181-195.

- [116] Patra, J.C., R.N. Pal, B. Chatterji, and G. Panda, *Identification of nonlinear dynamic systems using functional link artificial neural networks*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1999. **29**(2): p. 254-262.
- [117] Pao, Y.-H., S.M. Phillips, and D.J. Sobajic, *Neural-net computing and the intelligent control of systems*. International Journal of Control, 1992. **56**(2): p. 263-289.
- [118] Dehuri, S. and S.-B. Cho, *A comprehensive survey on functional link neural networks and an adaptive PSO–BP learning for CFLNN*. Neural Computing and Applications, 2010. **19**(2): p. 187-205.
- [119] Ivakhnenko, A., *Polynomial theory of complex systems*. IEEE Transactions on Systems, Man, and Cybernetics, 1971(4): p. 364-378.
- [120] Chapman, R. and M.A.A. Rahman, *Chebyshev-polynomial-based Schur algorithm*. IEE Proceedings F - Radar and Signal Processing, 1990. **137**(1): p. 11-14.
- [121] Klir, G.J. and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. 1995: Prentice Hall PTR.
- [122] Vieira, J., F.M. Dias, and A. Mota. *Neuro-fuzzy systems: a survey*. in *5th WSEAS NNA International Conference on Neural Networks and Applications, Udine, Italia*. 2004.
- [123] Chen, C.-H., C.-J. Lin, and C.-T. Lin, *A functional-link-based neurofuzzy network for nonlinear system control*. IEEE Transactions on Fuzzy Systems, 2008. **16**(5): p. 1362-1378.
- [124] Wang, X., A. Abraham, and K.A. Smith. *Web Traffic Mining Using a Concurrent Neuro-Fuzzy Approach*. in *HIS*. 2002.
- [125] Ocampo-Duque, W., R. Juraske, V. Kumar, M. Nadal, J.L. Domingo, and M. Schuhmacher, *A concurrent neuro-fuzzy inference system for screening the ecological risk in rivers*. Environmental Science and Pollution Research, 2012. **19**(4): p. 983-999.
- [126] Jang, J.-S., *ANFIS: adaptive-network-based fuzzy inference system*. IEEE transactions on systems, man, and cybernetics, 1993. **23**(3): p. 665-685.
- [127] Lin, C.-J. and C.-T. Lin, *An ART-based fuzzy adaptive learning control network*. IEEE Transactions on Fuzzy Systems, 1997. **5**(4): p. 477-496.
- [128] Berenji, H.R. and P. Khedkar, *Learning and tuning fuzzy logic controllers through reinforcements*. IEEE Transactions on neural networks, 1992. **3**(5): p. 724-740.
- [129] Nauck, D. and R. Kruse, *Neuro-fuzzy systems for function approximation*. Fuzzy sets and systems, 1999. **101**(2): p. 261-271.
- [130] Jung, C. and C.T. Lin, *An on-line self-constructing neural fuzzy inference network and its application*. IEEE Trans. Fuzzy Syst, 1998. **6**: p. 12-32.
- [131] Kasabov, N.K. and Q. Song, *Dynamic evolving fuzzy neural networks with 'm-out-of-n' activation nodes for on-line adaptive systems*. 1999: Technical Report TR99/0, Departement of Information Science, University of Otago.
- [132] Viharos, Z.J. and K.B. Kis, *Survey on Neuro-Fuzzy systems and their applications in technical diagnostics and measurement*. Measurement, 2015. **67**: p. 126-136.
- [133] Akritas, P., I. Antoniou, and V. Ivanov, *Identification and prediction of discrete chaotic maps applying a Chebyshev neural network*. Chaos, Solitons & Fractals, 2000. **11**(1): p. 337-344.
- [134] Zou, A.-M. and K.D. Kumar, *Adaptive attitude control of spacecraft without velocity measurements using Chebyshev neural network*. Acta Astronautica, 2010. **66**(5): p. 769-779.
- [135] Shaik, F.A., S. Purwar, and B. Pratap, *Real-time implementation of Chebyshev neural network observer for twin rotor control system*. Expert Systems with Applications, 2011. **38**(10): p. 13043-13049.
- [136] Patra, J.C., *Chebyshev neural network-based model for dual-junction solar cells*. IEEE Transactions on Energy Conversion, 2011. **26**(1): p. 132-139.
- [137] Patra, J.C. and A.C. Kot, *Nonlinear dynamic system identification using Chebyshev functional link artificial neural networks*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2002. **32**(4): p. 505-511.

- [138] Purwar, S., I.N. Kar, and A.N. Jha, *On-line system identification of complex systems using Chebyshev neural networks*. Applied Soft Computing, 2007. **7**(1): p. 364-372.
- [139] Basios, V., A.Y. Bonushkina, and V. Ivanov, *A method for approximating one-dimensional functions*. Computers & Mathematics with Applications, 1997. **34**(7): p. 687-693.
- [140] Rumelhart, D.E., G.E. Hinton, and R.J. Williams, *Learning internal representations by error propagation*. 1985, DTIC Document.
- [141] Antoniou, I. and B. Qiao, *Spectral decomposition of the chaotic logistic map*. Nonlinear World, 1997. **4**: p. 135-144.
- [142] Jackson, E.A., *Perspectives of Nonlinear Dynamics*. 1992: Cambridge University Press.
- [143] Oh, S.-K. and W. Pedrycz, *Fuzzy Polynomial Neuron-Based Self-Organizing Neural Networks*. International Journal of General Systems, 2003. **32**(3): p. 237-250.
- [144] Jackson, J.E., *A User's Guide to Principal Components*. 2003: Wiley.
- [145] Peason, K., *On lines and planes of closest fit to systems of point in space*. Philosophical Magazine, 1901. **2**: p. 559-572.
- [146] Sivanandam, S.N. and S.N. Deepa, *Introduction to Genetic Algorithms*. 2007: Springer Berlin Heidelberg.
- [147] Haupt, R.L. and S.E. Haupt, *Practical Genetic Algorithms*. 2004: Wiley.
- [148] Uykan, Z., C. Guzelis, M.E. Çelebi, and H.N. Koivo, *Analysis of input-output clustering for determining centers of RBFN*. IEEE transactions on neural networks, 2000. **11**(4): p. 851-858.
- [149] Bell, W.W., *Special Functions for Scientists and Engineers*. 2004: Dover Publications.
- [150] Gil, A., J. Segura, and N.M. Temme, *Numerical Methods for Special Functions*. 2007: Society for Industrial and Applied Mathematics.
- [151] Han, J., J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. 2011: Elsevier Science.
- [152] Kolen, M.J. and R.L. Brennan, *Test Equating, Scaling, and Linking: Methods and Practices*. 2014: Springer New York.
- [153] Forsythe, G.E., *Generation and use of orthogonal polynomials for data-fitting with a digital computer*. Journal of the Society for Industrial and Applied Mathematics, 1957. **5**(2): p. 74-88.
- [154] Clenshaw, C., *Curve fitting with a digital computer*. The Computer Journal, 1960. **2**(4): p. 170-173.
- [155] Rigos, A., M.I. Vousdoukas, O. Andreadis, G.E. Tsekouras, and A.F. Velegrakis, *On the systematic implementation of artificial neural networks in the classification of variance images and shoreline extraction*. Fresenius Environmental Bulletin, 2013. **23**(11): p. 2677-2686.
- [156] Niros, A.D. and G.E. Tsekouras, *A novel training algorithm for RBF neural network using a hybrid fuzzy clustering approach*. Fuzzy Sets Syst., 2012. **193**: p. 62-84.
- [157] Gonzalez, R.C., R.E. Woods, and S.L. Eddins, *Digital Image Processing Using MATLAB*. 2013: McGraw Hill Education.
- [158] Vousdoukas, M.I., P.M. Ferreira, L.P. Almeida, G. Dodet, F. Psaros, U. Andriolo, R. Taborda, A.N. Silva, A. Ruano, and Ó.M. Ferreira, *Performance of intertidal topography video monitoring of a meso-tidal reflective beach in South Portugal*. Ocean Dynamics, 2011. **61**(10): p. 1521-1540.
- [159] Rigos, A., O.P. Andreadis, A. Manousakis, M.I. Vousdoukas, G.E. Tsekouras, and A. Velegrakis, *Shoreline Extraction from Coastal Images Using Chebyshev Polynomials and RBF Neural Networks*. in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. 2014. Springer.
- [160] Plant, N.G. and R.A. Holman, *Intertidal beach profile estimation using video images*. Marine Geology, 1997. **140**(1): p. 1-24.
- [161] Tsekouras, G.E., A. Rigos, A. Chatzipavlis, and A. Velegrakis, *A Neural-Fuzzy Network Based on Hermite Polynomials to Predict the Coastal Erosion*, in *Engineering Applications of Neural*

- Networks: 16th International Conference, EANN 2015, Rhodes, Greece, September 25-28 2015. Proceedings*, L. Iliadis and C. Jayne, Editors. 2015, Springer International Publishing: Cham. p. 195-205.
- [162] Andrews, L.C., *Special Functions of Mathematics for Engineers*. 1992: SPIE Optical Engineering Press.
- [163] Lebedev, N.N. and R.A. Silverman, *Special Functions and Their Applications*. 1965: Prentice-Hall.
- [164] Tsekouras, G., H. Sarimveis, and G. Bafas, *A simple algorithm for training fuzzy systems using input-output data*. *Advances in Engineering Software*, 2003. **34**(5): p. 247-259.
- [165] Tsekouras, G., H. Sarimveis, E. Kavakli, and G. Bafas, *A hierarchical fuzzy-clustering approach to fuzzy modeling*. *Fuzzy Sets and Systems*, 2005. **150**(2): p. 245-266.
- [166] Roelvink, D., A. Reniers, A. van Dongeren, J. van Thiel de Vries, R. McCall, and J. Lescinski, *Modelling storm impacts on beaches, dunes and barrier islands*. *Coastal Engineering*, 2009. **56**(11-12): p. 1133-1152.
- [167] El Attar, R., *Special functions and orthogonal polynomials*. 2006: Lulu Press, Morrisville, NC.
- [168] Rigos, A., G.E. Tsekouras, A. Chatzipavlis, and A.F. Velegrakis, *Modeling Beach Rotation Using a Novel Legendre Polynomial Feedforward Neural Network Trained by Nonlinear Constrained Optimization*. in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. 2016. Springer.
- [169] Luenberger, D.G., *Introduction to Dynamic Systems: Theory, Models, and Applications*. 1979: Wiley.
- [170] Alligood, K., T. Sauer, and J.A. Yorke, *Chaos: An Introduction to Dynamical Systems*. 2012: Springer Berlin Heidelberg.
- [171] Wang, J., *Stochastic process and its application in finance*. 2007, Tsinghua University Press and Beijing Jiaotong University Press, Beijing.
- [172] Lin, C.-H., *Novel adaptive recurrent Legendre neural network control for PMSM servo-drive electric scooter*. *Journal of Dynamic Systems, Measurement, and Control*, 2015. **137**(1): p. 011010.
- [173] Luenberger, D.G. and Y. Ye, *Linear and Nonlinear Programming*. 2015: Springer International Publishing.
- [174] Velegrakis, A., V. Trygonis, A. Chatzipavlis, T. Karambas, M. Vousdoukas, G. Ghionis, I. Monioudi, T. Hasiotis, O. Andreadis, and F. Psarros, *Shoreline variability of an urban beach fronted by a beachrock reef from video imagery*. *Natural Hazards*, 2016. **83**(1): p. 201-222.
- [175] Hodges, J. and E.L. Lehmann, *Rank methods for combination of independent experiments in analysis of variance*. *The Annals of Mathematical Statistics*, 1962. **33**(2): p. 482-497.
- [176] Rodríguez-Fdez, I., A. Canosa, M. Mucientes, and A. Bugarín. *STAC: a web platform for the comparison of algorithms using statistical tests*. in *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*. 2015. IEEE.
- [177] Hartley, R. and A. Zisserman, *Multiple view geometry in computer vision*. 2003: Cambridge university press.
- [178] Forsyth, D.A. and J. Ponce, *Computer vision: a modern approach*. 2002: Prentice Hall Professional Technical Reference.
- [179] Hartley, R.I., *Theory and practice of projective rectification*. *International Journal of Computer Vision*, 1999. **35**(2): p. 115-127.
- [180] Prigent, C., F. Papa, F. Aires, C. Jimenez, W. Rossow, and E. Matthews, *Changes in land surface water dynamics since the 1990s and relation to population pressure*. *Geophysical Research Letters*, 2012. **39**(8).
- [181] Marcos, M., G. Jordà, D. Gomis, and B. Pérez, *Changes in storm surges in southern Europe from a regional model under climate change scenarios*. *Global and Planetary Change*, 2011. **77**(3): p. 116-128.

- [182] Harley, M., C. Armaroli, and P. Ciavola, *Evaluation of XBeach predictions for a real-time warning system in Emilia-Romagna, Northern Italy*. Journal of Coastal Research, 2011(64): p. 1861.
- [183] Leont'yev, I., *Numerical modelling of beach erosion during storm event*. Coastal Engineering, 1996. **29**(1): p. 187-200.
- [184] Karambas, T.V. and E. Karathanassi, *Boussinesq modeling of longshore currents and sediment transport*. Journal of Waterway, Port, Coastal and Ocean Engineering, American Society of Civil Engineers (ASCE), 2004. **130**(6): p. 277-286.
- [185] Roelvink, D., A. Reniers, A. Van Dongeren, J.v.T. de Vries, R. McCall, and J. Lescinski, *Modelling storm impacts on beaches, dunes and barrier islands*. Coastal engineering, 2009. **56**(11): p. 1133-1152.
- [186] Sánchez-Arcilla, A., J. Gomez Aguar, J.J. Egozcue, M. Ortego, P. Galiatsatou, and P. Prinos, *Extremes from scarce data: the role of Bayesian and scaling techniques in reducing uncertainty*. Journal of Hydraulic Research, 2008. **46**(S2): p. 224-234.
- [187] Chatzipavlis, A., I. Monioudi, A.F. Velegrakis, and A. Baeten. *Case Scenario Investigation of Beach Nourishment in Eresos Beach, Lesvos, Greece*. 2012. International Society of Offshore and Polar Engineers.
- [188] Thomas, T., M. Phillips, and A. Williams, *A centurial record of beach rotation*. Journal of Coastal Research, 2013. **65**(sp1): p. 594-599.
- [189] da Fontoura Klein, A.H., L. Benedet Filho, and D.H. Schumacher, *Short-term beach rotation processes in distinct headland bay beach systems*. Journal of Coastal Research, 2002: p. 442-458.
- [190] Ranasinghe, R., R. McLoughlin, A. Short, and G. Symonds, *The Southern Oscillation Index, wave climate, and beach rotation*. Marine Geology, 2004. **204**(3): p. 273-287.
- [191] Thomas, T., N. Rangel-Buitrago, M.R. Phillips, G. Anfuso, and A.T. Williams, *Mesoscale Morphological Change, Beach Rotation and Storm Climate Influences along a Macrotidal Embayed Beach*. Journal of Marine Science and Engineering, 2015. **3**(3): p. 1006-1026.
- [192] Harley, M., I. Turner, and A. Short, *New insights into embayed beach rotation: The importance of wave exposure and cross-shore processes*. Journal of Geophysical Research: Earth Surface, 2015. **120**(8): p. 1470-1484.
- [193] Gallop, S.L., C. Bosserelle, I. Eliot, and C.B. Pattiaratchi, *The influence of limestone reefs on storm erosion and recovery of a perched beach*. Continental Shelf Research, 2012. **47**: p. 16-27.
- [194] Gallop, S.L., C. Bosserelle, I. Eliot, and C.B. Pattiaratchi, *The influence of coastal reefs on spatial variability in seasonal sand fluxes*. Marine Geology, 2013. **344**: p. 132-143.
- [195] Lowe, R.J., C. Hart, and C.B. Pattiaratchi, *Morphological constraints to wave-driven circulation in coastal reef-lagoon systems: A numerical study*. Journal of Geophysical Research: Oceans, 2010. **115**(C9).
- [196] Alexandrakis, G., G. Ghionis, and S. Poulos, *The effect of beach rock formation on the morphological evolution of a beach. The case study of an eastern Mediterranean beach: Ammoudara, Greece*. Journal of Coastal Research, 2013. **69**(sp1): p. 47-59.
- [197] Sandwell, D.T., *Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data*. Geophysical research letters, 1987. **14**(2): p. 139-142.