



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ & ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

Διπλωματική Εργασία

Ανάπτυξη προσαρμοστικών μεθόδων για online εκπαίδευση νευρωνικών δικτύων τύπου radial basis function



**Φοιτητής: Κολέτσιος Σωτήριος
ΑΜ: 06796**

Επιβλέπων Καθηγητής

**Αλεξανδρίδης Αλέξανδρος
Καθηγητής**

ΑΘΗΝΑ-ΑΙΓΑΛΕΩ, Οκτώβριος 2021



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

Diploma Thesis

Development of adaptive methods for online training of radial basis function neural networks



Student: Koletsios Sotirios
Registration Number: 06796

Supervisor

Alex Alexandridis
Professor

ATHENS-EGALEO, October 2021

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

Αλεξανδρίδης Αλέξανδρος, Καθηγητής	Φαμέλης Ιωάννης, Καθηγητής	Ζώης Ηλίας, Επίκουρος Καθηγητής
(Υπογραφή)	(Υπογραφή)	(Υπογραφή)

Copyright © Με επιφύλαξη παντός δικαιώματος. All rights reserved.

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και ΚΟΛΕΤΣΙΟΣ ΣΩΤΗΡΙΟΣ,
Οκτώβριος, 2021**

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

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

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Κολέτσιος Σωτήριος του Θωμά, με αριθμό μητρώου 06796 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ,

δηλώνω υπεύθυνα ότι:

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

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

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

Ο Δηλών
Κολέτσιος Σωτήριος



Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κ. Αλεξανδρίδη Αλέξανδρο για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου ένα τόσο ενδιαφέρον, και κατά τα δικά μου δεδομένα, δύσκολο και απαιτητικό θέμα καθώς και για την συνεχή βοήθεια που μου παρείχε, για την υπομονή που έδειξε εξηγώντας μου όλες αυτές τις έννοιες. Με την παρακολούθηση και την εκπλήρωση των εργασιών που σου προσφέρουν τα μαθήματα που πρότεινε ο καθηγητής φεύγεις από το θεωρητικό επίπεδο και αρχίζεις πραγματικά να κατανοείς τι σημαίνουν όλες αυτές οι έννοιες. Προχωρώντας την πορεία μου στην σχολή γνώρισα άλλο ένα άτομο το οποίο είχαμε τις ίδιες απόψεις και αγάπη για τέτοιου είδους προβλήματα. Την υποψήφια διδάκτορα Καραμιχαηλίδου Δέσποινα. Δέσποινα θα ήθελα να πω ένα μεγάλο ευχαριστώ και σε σένα για την άψογη συνεργασία και υπομονή που έδειξες με τις «παραξενιές» μου, για τις ατελείωτες ώρες debugging καθώς επίσης για την βοήθειά σου σε θέματα που υστερούσα (αχ αυτά τα Givens Rotations). Ήσουν κατανοητή, με άψογη συμπεριφορά και μεταδοτικότητα που λίγοι άνθρωποι κατέχουν. Κανονικά σε αυτό το σημείο θα έπρεπε να σου ευχηθώ να τελειώσεις το διδακτορικό σου. Όμως νομίζω πως είναι καλύτερα να ευχηθώ να περάσεις την γραμμή του τερματισμού και να συνεχίσεις να έχεις την ίδια κατανόηση, συμπεριφορά και μεταδοτικότητα που έδειξες ότι έχεις σε μένα, για όλους τους μελλοντικούς φοιτητές που θα παρακολουθούν τα μαθήματά σου όπως επίσης και για τα άτομα που θα συνεργαστούν μαζί σου.

Θα ήταν παράλειψη από μεριάς μου να μην ευχαριστήσω την οικογένειά μου, που σε όλη την περίοδο της σχολής με στήριξαν, βάζοντας τις δικές μου ανάγκες πάνω από τις δικές τους. Πλέον οι δικές μου ανάγκες ολοκληρώθηκαν και ήρθε η δικιά σας στιγμή. Μακάρι να μπορέσω να σας βοηθήσω να πραγματοποιήσετε τα όνειρα σας, έτσι όπως με βοηθήσατε εσείς.

Υπάρχει ένα ακόμα άτομο που η βοήθεια και στήριξη ήταν κάτι παραπάνω από τεράστια. Παραμένοντας στην σκιά, με βοήθησε να ανταπεξέλθω σε όλες τις δυσκολίες που προέκυψαν τα τελευταία χρόνια και που συνεχίζουν και προκύπτουν. Ι.Χ. ένα μεγάλο ευχαριστώ και σε σένα. Ήσουν η φωνή της λογικής που με βοήθησε να ξεπεράσω όλες τις δύσκολες στιγμές. Συνέχισε αυτό που κάνεις...

Σωτήριος Κολέτσιος
Τρίκαλα,
Σεπτέμβριος 2021

“It is as if my life were magically run by two electric currents: joyous positive and despairing negative — whichever is running at the moment dominates my life, floods it.”

The Unabridged Journals of Sylvia Plath

Περίληψη

Η εκπαίδευση σε μη στατικά περιβάλλοντα παρουσιάζει σημαντικές δυσκολίες λόγω του drift που παρουσιάζουν οι ροές δεδομένων καθώς και της περιορισμένης διαθεσιμότητας των δεδομένων αυτών. Από την άλλη, η εκπαίδευση σε μη στατικά περιβάλλοντα είναι πολύ σημαντική για την μοντελοποίηση ροών δεδομένων, όπου ένα αποτελεσματικό και ενημερωμένο μοντέλο είναι αναγκαίο καθώς η ροή δεδομένων εξελίσσεται. Αυτήν η διπλωματική παρουσιάζει έναν νέο αλγόριθμο εκπαίδευσης σε τέτοιου είδους περιβάλλοντα. Χρησιμοποιείται η λογική των τεχνητών νευρωνικών δικτύων, και πιο συγκεκριμένα τα δίκτυα radial basis function (RBF), σε συνδυασμό με τον αλγόριθμο Fuzzy Means (FM) για την ενημέρωση των κέντρων και την τεχνική Givens Rotations για την ενημέρωση των συναπτικών βαρών. Ο προτεινόμενος αλγόριθμος δοκιμάστηκε σε 9 πραγματικές και τεχνητές ροές δεδομένων (benchmark) και έγινε σύγκριση με άλλους γνωστούς αλγορίθμους που υπάρχουν στην βιβλιογραφία. Τα αποτελέσματα δείχνουν πως ο προτεινόμενος αλγόριθμος FM-Givens δημιουργεί μοντέλα με καλύτερη απόδοση ακόμα και σε ροές που υπάρχει drift, ενώ ο χρόνος εκπαίδευσης του παραμένει μικρός.

Λέξεις κλειδιά: ροές δεδομένων, φαινόμενο drift, μη στατικά περιβάλλοντα, δυναμικά περιβάλλοντα, online εκπαίδευση, δίκτυα ακτινικής συμμετρίας, ασαφή σύνολα

Abstract

Learning on non-stationary environments is laden with many challenges, as the procedure is usually characterized by drifts and data unavailability; on the other hand, it is of great importance with regard to data stream modeling, where an effective and up-to date model is required as the data stream evolves. This work presents a new method for producing highly accurate models for learning on such environments. The method is based on artificial neural networks (ANNs) and more specifically on the efficient architecture of radial basis function (RBF) networks. A novel RBF online training scheme for real time adaption of both the network structure and parameter values is proposed based on the fuzzy means (FM) algorithm and the Givens rotations technique. Within this integrated framework, it is guaranteed that for each update of the network structure, the optimal values for the synaptic weights are calculated efficiently by maintaining low order matrix updates. The proposed approach is evaluated on 9 real-word and artificial benchmark data streams and is compared to other well-known methodologies from the literature. The results show that the FM-Givens algorithm produces models with highly competitive online accuracy for non-stationary environments in the presence of drifts, while maintaining low model-updating times.

Keywords: concept drift, data stream, fuzzy means, non-stationary environments, online learning, radial basis function network

Περιεχόμενα

Ευχαριστίες	4
Περίληψη	8
Abstract	9
1. Κεφάλαιο 1^ο: Εισαγωγή	12
Δομή Διπλωματικής	14
2. Κεφάλαιο 2^ο: Νευρωνικά Δίκτυα	15
i. Εισαγωγή.....	15
ii. Radial Basis Function Networks (RBFNs)	24
iii. Fuzzy Means	27
3. Κεφάλαιο 3^ο: Online Μάθηση	30
i. Γενικά.....	30
ii. Online Fuzzy Means	34
iii. Givens Fuzzy Means	37
4. Κεφάλαιο 4^ο: Πειραματικό Μέρος	40
i. Περιγραφή συνόλων ροών δεδομένων	40
ii. Αποτελέσματα.....	44
iii. Συζήτηση.....	50
Σχολιασμός πίνακα αποτελεσμάτων	50
Σχολιασμός σχημάτων	51
5. Συμπεράσματα	53
6. Βιβλιογραφία	55

1. Εισαγωγή

Η επιστήμη της τεχνητής νοημοσύνης σημειώνει μεγάλη πρόοδο τα τελευταία χρόνια σε διάφορες εφαρμογές. Στο παρελθόν οι έρευνες είχαν επικεντρωθεί σε ανάπτυξη αλγορίθμων εκπαίδευσης με δεδομένα που προέρχονται από στατικό περιβάλλον. Σε ένα στατικό περιβάλλον τα δεδομένα είναι διαθέσιμα από την αρχή και δεν εισέρχονται σε αυτό με την πάροδο του χρόνου. Ωστόσο, με την πάροδο του χρόνου σημειώνεται μεγάλη πρόοδος στην εκπαίδευση συστημάτων ακόμα και σε δυναμικά περιβάλλοντα.

Δυναμικά θεωρούνται τα περιβάλλοντα που δεν είναι διαθέσιμα τα δεδομένα στο σύστημα από την αρχή αλλά καταφθάνουν σε αυτό σαν μία ροή δεδομένων (data streams). Με την εξέλιξη του Internet of Things όλο και περισσότεροι αισθητήρες και συστήματα έχουν την δυνατότητα για την στιγμιαία αποστολή των μετρήσεων σε βάσεις δεδομένων χρησιμοποιώντας κάποιο δίκτυο. Οι μετρήσεις αυτές είναι τόσο πολλές και φτάνουν ως κάποια ροή δεδομένων. Ένα παράδειγμα ενός τέτοιου συστήματος είναι η παρακολούθηση μίας καλλιέργειας. Χρησιμοποιούνται αισθητήρες που μετράνε διάφορα στοιχεία από αυτήν, όπως για παράδειγμα την ωρίμανση του καρπού, το επίπεδο υγρασίας τόσο της ατμόσφαιρας όσο και του εδάφους, την ηλιοφάνεια, τις ριπές ανέμου κ.α. Μετά από την επεξεργασία αυτών των μετρήσεων προβλέπεται το απαιτούμενο νερό που χρειάζεται ο καρπός, εάν χρειάζεται, τις απώλειες που υπάρχουν από διάφορες καιρικές συνθήκες αλλά και σε πόσες μέρες θα μπορέσει να γίνει η συγκομιδή των καρπών. Σε αυτά τα συστήματα εκτός από δεδομένα που εισέρχονται από τους αισθητήρες που υπάρχουν ήδη στην καλλιέργεια μπορεί να γίνεται και λήψη πληροφοριών από την μετεωρολογική υπηρεσία για την μέγιστη και την ελάχιστη θερμοκρασία της ημέρας, το προβλεπόμενο επίπεδο της υγρασίας, την πιθανότητα βροχόπτωσης στην περιοχή. Λαμβάνοντας όλες αυτές τις πληροφορίες μπορεί να γίνει πιο αποδοτική, παραδείγματος χάριν η ανάγκη ποτίσματος της καλλιέργειας έτσι ώστε να μην ξεραθούν οι καρποί. Αυτό το σύστημα μπορεί με την πάροδο του χρόνου, και με την συνεχή εκπαίδευσή του να βελτιστοποιήσει τα έσοδα ενός αγρότη μειώνοντας τόσο την απώλεια καρπών, όσο και την ανάγκη πρώτων υλών. Επιτρέποντας τέτοια συστήματα να εκπαιδεύονται όχι μόνο από μία καλλιέργεια αλλά ανταλλάσσοντας πληροφορίες με ίδιου τύπου καλλιέργειες, η απόδοση του συστήματος αυξάνεται διότι ο αλγόριθμος εκπαιδεύει καλύτερα το σύστημα από την στιγμή που έχει περισσότερα δεδομένα εισόδου.

Έτσι ένας κλάδος ο οποίος έχει κερδίσει το ενδιαφέρον πολλών ερευνητών είναι η μάθηση σε μη στατικά περιβάλλοντα, που τα δεδομένα εισόδου δεν είναι γνωστά από την αρχή αλλά με την πάροδο του χρόνου εισέρχονται στο σύστημα διαδοχικά, σαν ροές δεδομένων. Συνήθως τα προβλήματα που στοχεύουν αυτοί οι αλγόριθμοι είναι προβλήματα κατηγοριοποίησης, χωρίς να αποκλείονται και άλλα σημαντικά προβλήματα όπως είναι η προσέγγιση μίας συνάρτησης. Τα δεδομένα αυτά έχουν διαφορετικές διακυμάνσεις με την πάροδο του χρόνου. Χρησιμοποιώντας τους υπάρχοντες αλγορίθμους για την εκπαίδευση τέτοιων δικτύων διαπιστώθηκε πως η απόδοσή τους δεν είναι ικανοποιητική [1]. Λόγω της μεγάλης διαφοροποίησης των δεδομένων με την πάροδο του χρόνου, τα δίκτυα που προκύπτουν από τους κλασσικούς αλγορίθμους δεν μπορούν να προσαρμόσουν την δομή τους στα δεδομένα που εισέρχονται στο σύστημα. Για τον σκοπό αυτό ξεκίνησε η ανάπτυξη αλγορίθμων ικανών να προσαρμόζονται στις ροές δεδομένων με την πάροδο του χρόνου και αποτελεσματικά ως

προς την κατηγοριοποίηση. Οι αλγόριθμοι που δημιουργήθηκαν βασίστηκαν στους υπάρχοντες αλγορίθμους [2] με μερικές διαφοροποιήσεις. Για παράδειγμα ο αλγόριθμος ARF (Adaptive random forest) κάνει τον αλγόριθμο random forest ικανό να επεξεργαστεί ροές δεδομένων με μεγάλη διαφοροποίηση στις τιμές εισόδου, με αποτέλεσμα να μπορεί να προβλέπει με επιτυχία την έξοδο ακόμα και σε μεγάλες δομές δεδομένων [3]. Αναφορικά κάποιιοι άλλοι αλγόριθμοι που έχουν εξελιχθεί για να μπορούν να λύνουν προβλήματα κατηγοριοποίησης σε ροές δεδομένων είναι ο αλγόριθμος Naïve Bayes (NB) [4] και ο αλγόριθμος Stochastic gradient descent (SGD) [5]. Ένας από τους σημαντικότερους αλγορίθμους εκπαίδευσης στην κατηγορία των feed forward δικτύων είναι τα δίκτυα ακτινικής συμμετρίας (radial basis function, RBF) [6]. Η δομή τους είναι απλή με ένα επίπεδο εισόδου, ένα κρυφό επίπεδο και ένα επίπεδο εξόδου που συνδέεται γραμμικά με το κρυφό επίπεδο. Λόγω της απλής δομής τους έχουν αποτελέσει το ενδιαφέρον της επιστημονικής κοινότητας. Τα μοντέλα των δικτύων RBF έχουν χρησιμοποιηθεί με επιτυχία για την επίλυση περίπλοκων προβλημάτων παρουσιάζοντας ταχύτητα και αποτελεσματικότητα. Στην βιβλιογραφία αναφέρονται διάφοροι αλγόριθμοι εκπαίδευσης, οι περισσότεροι των οποίων αποτελούνται από δύο στάδια. Στο πρώτο υπολογίζονται το πλήθος και οι τιμές των κέντρων συνήθως με κάποιον αλγόριθμο μη εποπτευόμενης μάθησης (unsupervised learning). Το δεύτερο στάδιο περιλαμβάνει τον υπολογισμό των συναπτικών βαρών του δικτύου εφαρμόζοντας συνήθως γραμμικά ελάχιστα τετράγωνα.

Τα τεχνητά νευρωνικά δίκτυα αποτελούνται από ένα σύνολο διατεταγμένων υπολογιστικών μονάδων επεξεργασίας, που ακολουθούν συγκεκριμένες τεχνικές σχεδίασης, και χρησιμοποιούνται για την επίλυση προβλημάτων κατηγοριοποίησης, πρόβλεψης, χρονοσειρών, ή παλινδρόμησης κ.α. Οι τιμές των παραμέτρων που απαρτίζουν το νευρωνικό δίκτυο προέρχονται μετά από μια διαδικασία που είναι γνωστή ως εκπαίδευση.

Το τελευταίο καιρό οι ανάγκες που έχουν προκύψει για την άμεση εκπαίδευση τεχνητών νευρωνικών δικτύων και λήψη απόφασης έχοντας σαν είσοδο μεγάλες ροές δεδομένων, σε πραγματικό χρόνο, έχουν αυξηθεί. Αυτός ήταν ένας από τους κύριους λόγους επιλογής του συγκεκριμένου θέματος της διπλωματικής, η οποία στοχεύει στην ανάπτυξη μίας νέας προσαρμοστικής (online) μεθόδου εκπαίδευσης νευρωνικών δικτύων χρησιμοποιώντας τα δίκτυα ακτινικής συμμετρίας (radial basis function, RBF). Τα δίκτυα αυτά προσφέρουν μοντέλα τα οποία σημειώνουν μεγάλη πρόοδο το τελευταίο διάστημα, παρουσιάζοντας υψηλότερη ακρίβεια σε πιο σύντομο χρονικό διάστημα σε σχέση με τα απλά δίκτυα πολλαπλών επιπέδων perceptron (multilayer perceptron, MLP) [7].

Υπάρχουν διάφοροι αλγόριθμοι που χρησιμοποιούνται για την εύρεση των κέντρων. Η επιλογή του αλγορίθμου εκπαίδευσης παίζει μεγάλο ρόλο τόσο στην απόδοση όσο και στην ταχύτητα της εκπαίδευσης του δικτύου. Η μέθοδος που χρησιμοποιείται περισσότερο στην βιβλιογραφία είναι ο αλγόριθμος k-means clustering. Στην δημοσίευση [8] γίνεται αναφορά για δίκτυα RBF που χρησιμοποιούν k-means για την εύρεση των κέντρων ενώ για τον υπολογισμό των βαρών κάθε κέντρου χρησιμοποιείται ο αλγόριθμος των ελαχίστων τετραγώνων.

Κάποιες έρευνες επικεντρώθηκαν στην εύρεση ενός αλγορίθμου σε αντικατάσταση του k-means έτσι ώστε να ξεπεραστούν τα προβλήματα που προκύπτουν από αυτόν. Στη δημοσίευση [7] παρουσιάζεται ο αλγόριθμος fuzzy means (FM) ο οποίος έχει καλύτερα απόδοση και αποτελεσματικότητα στην εύρεση του πλήθους και των κέντρων του δικτύου RBF. Κάποια από τα πιο σημαντικά πλεονεκτήματα του αλγορίθμου είναι ότι μπορεί να βρίσκει το πλήθος

και τις θέσεις των κέντρων του δικτύου με αυτοματοποιημένη διαδικασία που βασίζεται στην λογική των ασαφών συνόλων. Επίσης δεν υπάρχει τυχαιότητα όπως στον k-means και ανεξάρτητα από όσες φορές και να τρέξει ο αλγόριθμος θα βγάλει τα ίδια αποτελέσματα.

Σε αυτήν την διπλωματική παρουσιάζεται ένας νέος προσαρμοστικός αλγόριθμος εκπαίδευσης δικτύων RBF, ο οποίος βασίζεται στον αλγόριθμο Fuzzy Means για την εύρεση του πλήθους και των τιμών των κέντρων σε πραγματικό χρόνο ενώ για τον υπολογισμό των βαρών γίνεται χρήση της μεθόδου Givens Rotations έτσι ώστε να γίνει μείωση του χρόνου εκπαίδευσης και αύξηση της αποτελεσματικότητας του δικτύου. Ο προτεινόμενος προσαρμοστικός αλγόριθμος εκπαίδευσης, αξιολογήθηκε σε διάφορες benchmark ροές δεδομένων της βιβλιογραφίας τόσο του τεχνητού όσο του πραγματικού κόσμου που χρησιμοποιούνται για αξιολόγηση προσαρμοστικών μεθόδων.

Δομή Διπλωματικής

Η διπλωματική αναπτύσσεται σε τέσσερα (4) κεφάλαια.

Στο επόμενο κεφάλαιο γίνεται μία εισαγωγή στα νευρωνικά δίκτυα. Παρουσιάζονται οι αναλογίες με τους βιολογικούς νευρώνες, ενώ γίνεται μία αναφορά στις αρχιτεκτονικές των δικτύων καθώς επίσης και στα είδη των μεθόδων μηχανικής μάθησης. Στην συνέχεια του κεφαλαίου παρουσιάζονται τα δίκτυα Radial Basis Function καθώς και ο αλγόριθμος Fuzzy Means.

Στο τρίτο κεφάλαιο γίνεται μία εισαγωγή στην online μάθηση καθώς και τα πλεονεκτήματα που προσφέρει. Ακολουθεί η δομή και ανάλυση του αλγορίθμου Online Fuzzy Means καθώς και του προτεινόμενου αλγορίθμου Online Givens Fuzzy Means.

Στο τέταρτο κεφάλαιο παρουσιάζεται η πειραματική διαδικασία που ακολουθήθηκε καθώς επίσης γίνεται μία περιγραφή των συνόλων ροών δεδομένων που χρησιμοποιήθηκαν για τον έλεγχο της απόδοσης του προτεινόμενου αλγορίθμου. Στην συνέχεια παρουσιάζονται τα αποτελέσματα τις πειραματικής διαδικασίας, και μία ανάλυση των αποτελεσμάτων.

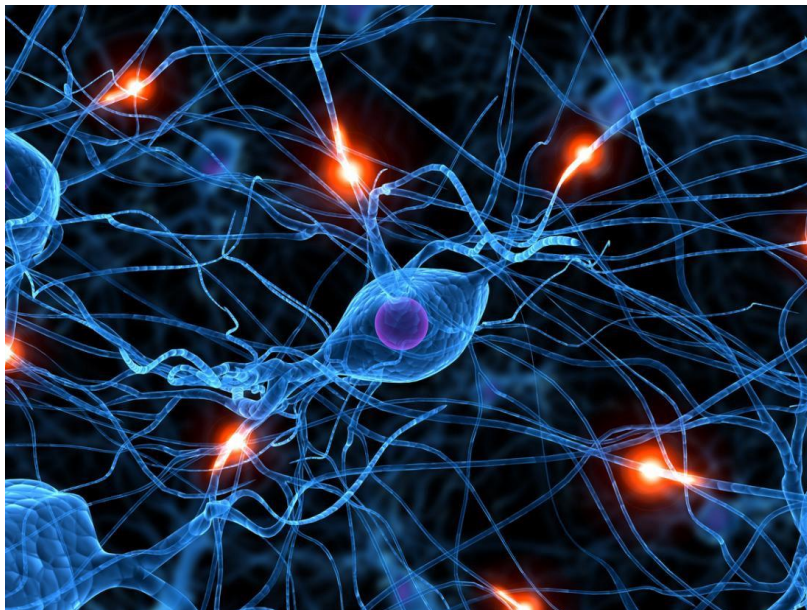
Τέλος στο πέμπτο κεφάλαιο αναλύονται τα συμπεράσματα που προέκυψαν μετά την ανάλυση των αποτελεσμάτων της πειραματικής διαδικασίας ενώ δίνεται έναυσμα για περαιτέρω μελλοντική εργασία.

2. Νευρωνικά Δίκτυα

i. Εισαγωγή

Η ιδέα της κατασκευής και ανάπτυξης των τεχνητών νευρωνικών δικτύων έχει ξεκινήσει από την κατανόηση της συμπεριφοράς του ανθρώπινου εγκεφάλου και την προσομοίωσή της στους ηλεκτρονικούς υπολογιστές. Μπορούμε να χαρακτηρίσουμε τον ανθρώπινο εγκέφαλο σαν έναν πολύπλοκο, καταναμημένο υπολογιστικό σύστημα, το οποίο αποτελείται από εκατοντάδες χιλιάδες νευρώνες που «συνεργάζονται» μεταξύ τους για την επεξεργασία διαφόρων ερεθισμάτων καθώς και την μετατροπή τους σε κάποια κίνηση – απόφαση. Για παράδειγμα, το οπτικό σύστημα αποτελείται από διάφορους νευρώνες που λαμβάνουν πληροφορίες για το περιβάλλον, τις οποίες και επεξεργάζονται ώστε να μας βοηθούν να κινούμαστε στον χώρο παίρνοντας τις κατάλληλες αποφάσεις.

Ο ανθρώπινος εγκέφαλος καταφέρνει να επεξεργάζεται μεγάλο όγκο πληροφοριών σε μικρό χρονικό διάστημα διότι από την γέννηση του ανθρώπου αρχίζει να εξελίσσεται και να δημιουργεί καινούργια «μονοπάτια», ανάλογα με τις εμπειρίες του, έτσι ώστε να καταφέρνει να φτάσει σε ένα επίπεδο που μία απλή λειτουργία αναγνώρισης ενός προσώπου να παίρνει περίπου 100 – 200 ms σε αντίθεση με έναν απλό υπολογιστή που μπορεί να διαρκέσει ακόμα και μέρες.



Σχήμα 1 Mapping human brain - Source: medium.com.

Αυτές οι ιδιαίζουσες ιδιότητες έχουν δώσει το έναυσμα για τη μελέτη και τη δημιουργία μαθηματικών μοντέλων τα οποία θα προσομοιώνουν τη λειτουργία του ανθρώπινου εγκεφάλου για την επίλυση πολύπλοκων προβλημάτων. Τα μοντέλα αυτά ονομάζονται τεχνητά νευρωνικά δίκτυα (artificial neural networks). Η χρήση τέτοιων μοντέλων είναι σημαντική στην εποχή που ζούμε διότι σε συνδυασμό με την υψηλή υπολογιστική ισχύ αλλά

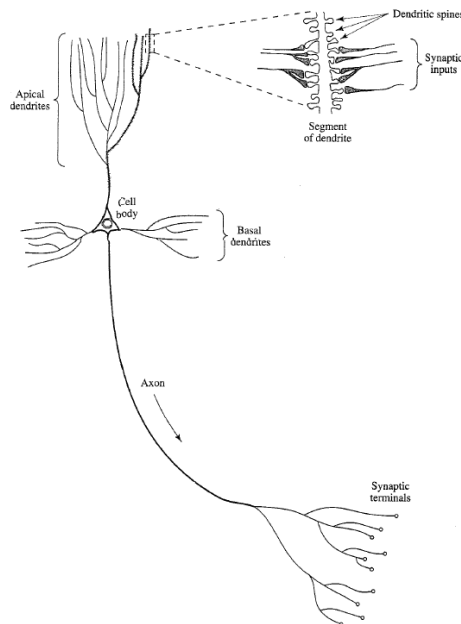
και την ταχύτητα υπολογισμών, μας δίνεται η δυνατότητα επίλυσης σύνθετων προβλημάτων με τη χρήση υπολογιστή στα οποία ακόμη κι ο άνθρωπος δυσκολεύεται.

Ωστόσο, τα πρώτα μοντέλα που είχαν δημιουργηθεί δεν ήταν ικανά για τη μοντελοποίηση σύνθετων προβλημάτων. Όμως με την πάροδο του χρόνου, ολοένα και πιο ακριβείς αρχιτεκτονικές εισάγονται στη βιβλιογραφία. Η χρήση των νευρωνικών δικτύων έχει πολλά πλεονεκτήματα και δυνατότητες. Η πιο βασική θα μπορούσαμε να ισχυριστούμε πως είναι η δυνατότητα ανάπτυξης ενός μοντέλου το οποίο βασίζεται μόνο σε δεδομένα για τη μοντελοποίηση ενός γραμμικού, ή ακόμα κι ενός μη γραμμικού συστήματος. Αυτό μας δίνει το πλεονέκτημα να χρησιμοποιούμε τέτοια μοντέλα για να λύνουμε προβλήματα που δεν είναι γνωστή η σχέση, είτε αυτή είναι γραμμική είτε όχι, μεταξύ της εισόδου και της εξόδου. Έτσι μπορούμε να μοντελοποιήσουμε οποιοδήποτε σύστημα, αρκεί να έχουμε έναν ικανοποιητικό αριθμό δεδομένων εισόδου και εξόδου για να εκπαιδεύσουμε το δίκτυο και να μπορούμε να προβλέψουμε τις εξόδους ακόμα και σε δεδομένα εισόδου που η έξοδος τους δεν είναι γνωστή. Η τελευταία ικανότητα είναι γνωστή και ως γενίκευση της γνώσης και πρόκειται για την ικανότητα του μοντέλου να μπορεί να βγάλει συμπεράσματα για δεδομένα που δεν έχει δει ξανά.

Βιολογικός νευρώνας

Για να μπορέσουμε να κατασκευάσουμε ένα τεχνητό νευρωνικό δίκτυο πρέπει να κατανοήσουμε τον τρόπο με τον οποίο λειτουργεί ο ανθρώπινος εγκέφαλος. Μετά από την δουλειά του Ramon y Cajal το 1911 [10], ο οποίος εισήγαγε την έννοια του νευρώνα, ο τρόπος λειτουργίας του ανθρώπινου εγκεφάλου έγινε πιο κατανοητός. Μία λειτουργία του νευρώνα συμβαίνει περίπου σε μερικά millisecond (10^{-3} s). Όμως λόγω της μεγάλης πολυπλοκότητας του νευρωνικού δικτύου του εγκεφάλου, αυτός ο χρόνος μειώνεται δραματικά. Σύμφωνα με τους Shepherd και Koch, το 1990 [10], ένας ανθρώπινος εγκέφαλος έχει 10 δισεκατομμύρια νευρώνες και 60 τρισεκατομμύρια συνάψεις (ή αλλιώς συνδέσεις μεταξύ τους). Οι συνάψεις είναι βασικές για την λειτουργία των νευρώνων, καθώς βοηθούν στην επικοινωνία μεταξύ τους. Το πιο βασικό είδος σύναψης είναι μία χημική σύναψη, στην οποία ένα ηλεκτρικό σήμα μεταφράζεται σε χημικό και μετά πάλι σε ηλεκτρικό.

Οι νευρώνες υπάρχουν σε μεγάλη ποικιλία από σχήματα και μεγέθη στα διαφορετικά τμήματα του εγκεφάλου. Το Σχήμα 2-1 παρουσιάζει τον νευρώνα σε σχήμα πυραμίδας που είναι μία από τις πιο γνωστές αναπαραστάσεις. Το κύτταρο σε σχήμα πυραμίδας μπορεί να λαμβάνει πάνω από 10000 συναπτικές επαφές και να τις μεταφέρει σε χιλιάδες κύτταρα. Οι περισσότεροι νευρώνες κωδικοποιούν τις εξόδους τους σε μία σειρά από ηλεκτρικές τάσεις και μέσω του άξονα μεταφέρουν το σήμα σε κάποιον άλλο νευρώνα. Ο άξονας είναι αρκετά μακρύς και λεπτός με μεγάλη αντοχή σε ηλεκτρικές τάσεις καθώς και χωρητικότητα. Μπορούμε να θεωρήσουμε τον άξονα του νευρώνα σαν μία γραμμή εκπομπού RC.

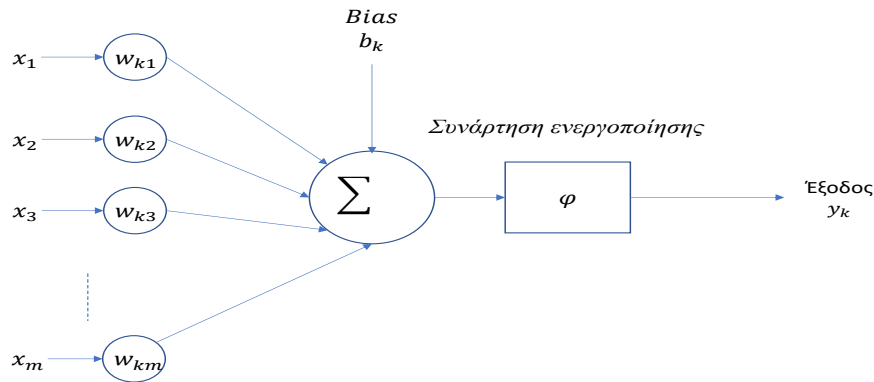


Σχήμα 2-1: Νευρώνας σε σχήμα πυραμίδας.

Μοντέλο τεχνητού νευρώνα

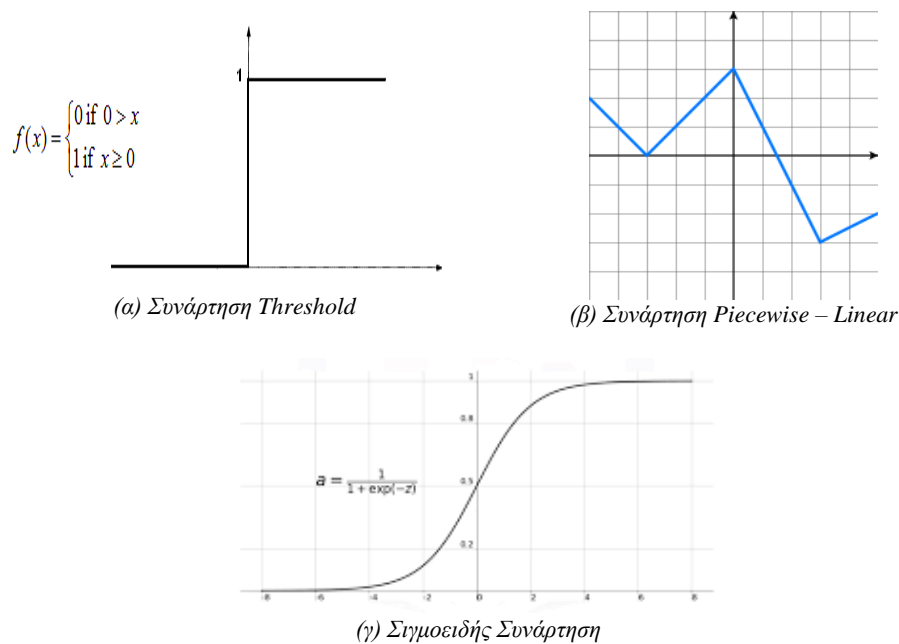
Έχοντας μια εικόνα της λειτουργίας ενός βιολογικού νευρώνα, μπορούμε να κατανοήσουμε καλύτερα τη λειτουργία του αντίστοιχου τεχνητού. Ένας τεχνητός νευρώνας είναι μία περιοχή που γίνεται η επεξεργασία των πληροφοριών και είναι βασικό συστατικό ενός τεχνητού νευρωνικού δικτύου. Το Σχήμα 2-2 απεικονίζει το μοντέλο ενός νευρώνα και μπορούμε να καταλάβουμε πως αποτελείται από τρία (3) βασικά στοιχεία:

1. Μία σειρά από συνάψεις, οι οποίες χαρακτηρίζονται από ένα βάρος. Πιο συγκεκριμένα έστω x_i η είσοδος της σύναψης i , που συνδέει τον νευρώνα k , και πολλαπλασιάζεται με το βάρος της σύναψης w_{ki} . Σε αντίθεση με τις συνάψεις σε έναν ανθρώπινο εγκέφαλο, οι συνάψεις σε ένα τεχνητό νευρωνικό δίκτυο μπορούν να πάρουν ακόμα και αρνητικές τιμές, ενισχύοντας ή καταστέλλοντας το σήμα εισόδου.
2. Έναν αθροιστή που λαμβάνει ως όρισμα το σταθμισμένο άθροισμα των σημάτων με τα βάρη των συνάψεων. Εκτός από τα σήματα εισόδου ο αθροιστής δέχεται σαν είσοδο και έναν συντελεστή προτίμησης, ή αλλιώς Bias b_k .
3. Μία συνάρτηση ενεργοποίησης, η οποία καθορίζει την τιμή της εξόδου του νευρώνα



Σχήμα 2-2: Μοντέλο Ενός Νευρώνα

Υπάρχουν πολλές συναρτήσεις ενεργοποίησης. Στο Σχήμα 2-3 παρουσιάζονται κάποιες από αυτές, όπως η συνάρτηση Threshold, η συνάρτηση Piecewise-Linear, και η σιγμοειδής (Sigmoid) συνάρτηση.



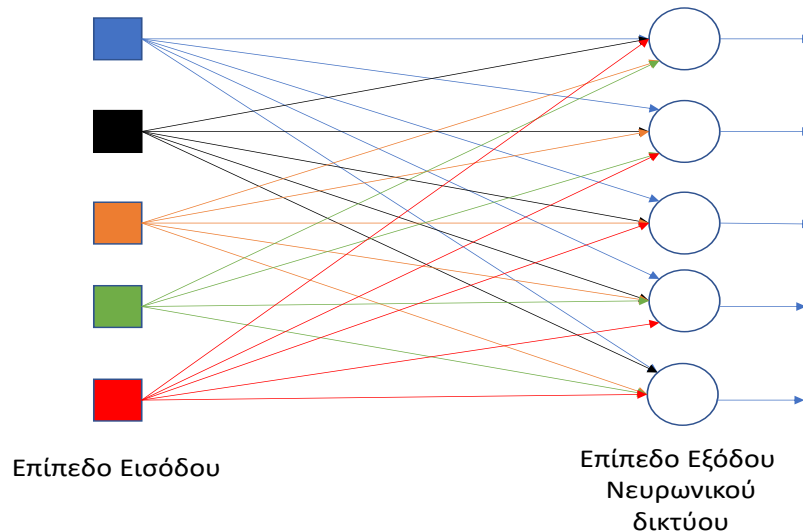
Σχήμα 2-3: Συναρτήσεις ενεργοποίησης

Αρχιτεκτονικές τεχνητού νευρωνικού δικτύου

Γενικά, θα μπορούσαμε να πούμε πως υπάρχουν τρεις (3) βασικοί τύποι αρχιτεκτονικής τεχνητού νευρωνικού δικτύου.

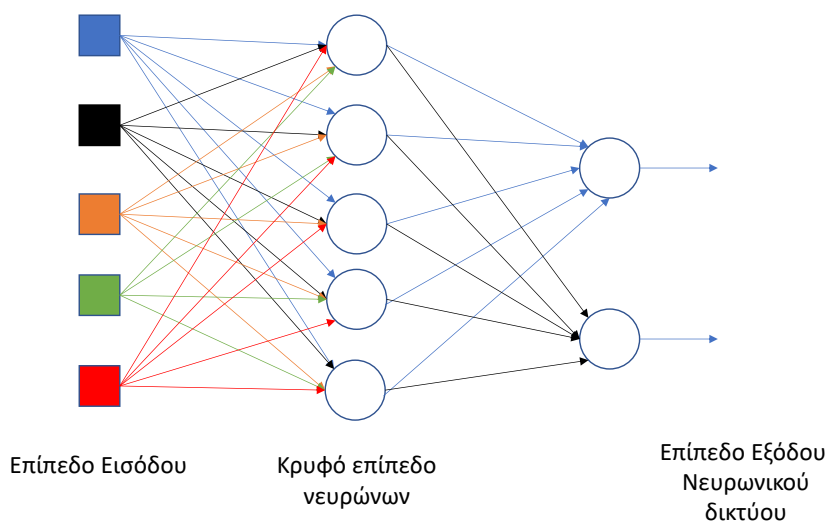
1. Γραμμικά Δίκτυα Feedforward μόνης στοιβάδας (Single - Linear Feedforward Networks): Σε αυτόν τον τύπο αρχιτεκτονικής οι νευρώνες είναι τοποθετημένοι σε

επίπεδα. Στην πιο απλή εκδοχή αυτού του δικτύου έχουμε ένα επίπεδο εισόδου το οποίο συνδέεται με όλες τις εξόδους στο επίπεδο εξόδου. Ονομάστηκε μονής γραμμικής επειδή έχουμε μόνο ένα επίπεδο εξόδου. Ακολουθεί το Σχήμα 2-4 που παρουσιάζει αυτήν την αρχιτεκτονική.



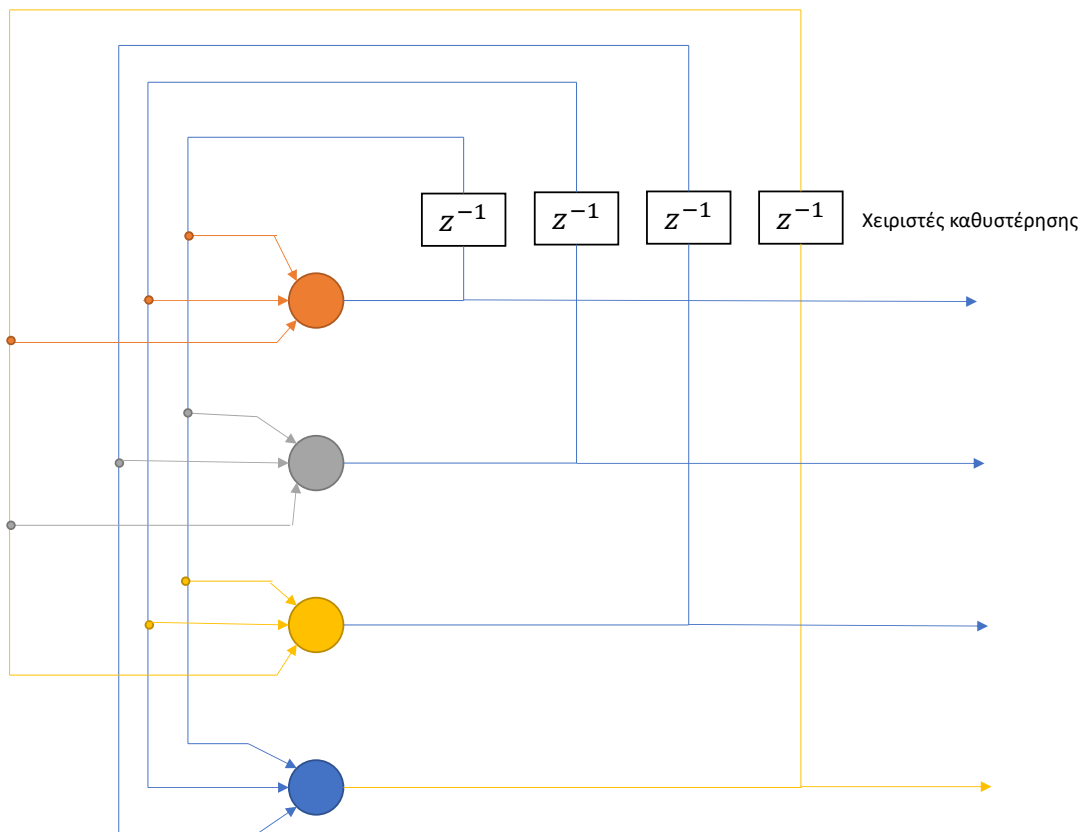
Σχήμα 2-4: Γραμμικό δίκτυο *FeedForward* μονή στοιβάδα

2. Δίκτυα *Feedforward* με πολλά επίπεδα. Το σκεπτικό είναι το ίδιο με το *Single-Linear* απλά σε αυτήν την περίπτωση έχουμε ένα ή και περισσότερα «κρυφά» επίπεδα που ονομάζονται κρυμμένοι νευρώνες (*hidden neurons*). Η λειτουργία των κρυφών νευρώνων είναι να επεμβαίνουν μετά το επίπεδο εισόδου και πριν το επίπεδο εξόδου και να αντλούνε με περισσότερη ακρίβεια στατιστικά μιας τιμής εισόδου η οποία συμμετέχει περισσότερο στο επίπεδο εισόδου. Ακολουθεί το Σχήμα 2-5 που παρουσιάζει αυτήν την αρχιτεκτονική.



Σχήμα 2-5: Νευρωνικό δίκτυο *Feedforward* με ένα κρυφό επίπεδο και ένα επίπεδο εξόδου

3. Αναδρομικά νευρωνικά δίκτυα: Τα δίκτυα αυτά διαφέρουν από τα νευρωνικά δίκτυα feedforward γιατί έχουν έναν τουλάχιστον βρόχο ανατροφοδότησης. Για παράδειγμα, ένα αναδρομικό δίκτυο μπορεί να αποτελείται από ένα επίπεδο με νευρώνες που κάθε ένας από αυτούς τους νευρώνες προσθέτει στην είσοδό του, την έξοδο. Ακολουθεί το Σχήμα 2-6 που απεικονίζει αυτήν την λειτουργία.



Σχήμα 2-6: Αναδρομικά νευρωνικά δίκτυα [10]

Ένα ακόμα χαρακτηριστικό αυτής της αρχιτεκτονικής είναι ότι οι βρόχοι επανάληψης αποτελούνται και από στοιχείο καθυστέρησης (unit-delay elements denoted by z^{-1}) το οποίο μας δίνει μία μη γραμμική συμπεριφορά.

Μέθοδοι μάθησης

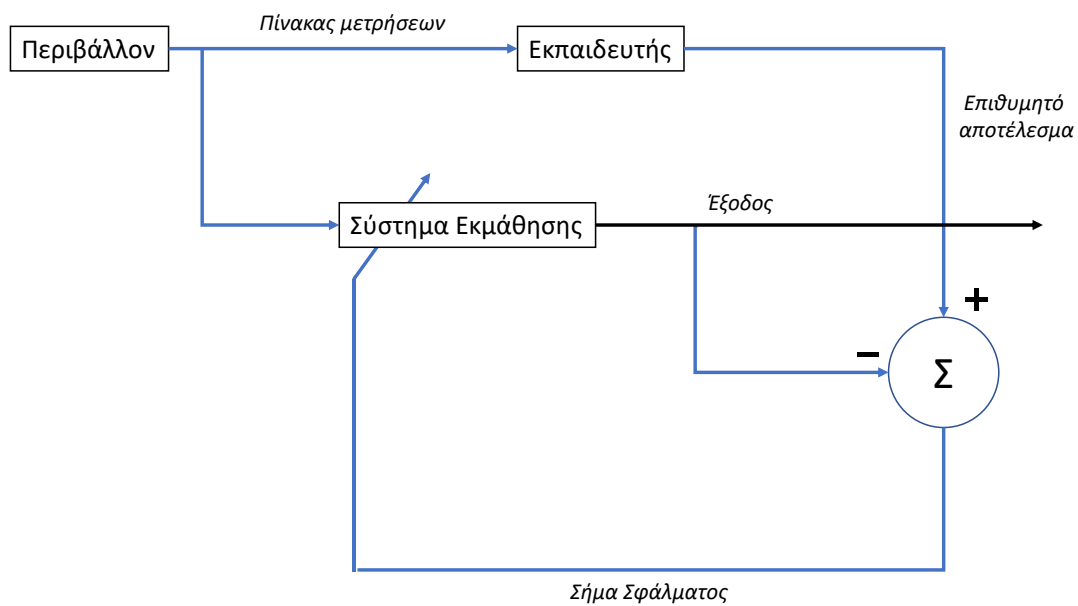
Θα μπορούσαμε να πούμε πως υπάρχουν δύο (2) βασικοί μέθοδοι εκπαίδευσης νευρωνικών δικτύων. Η μέθοδος εκπαίδευσης υπό επιτήρηση και η μέθοδος εκπαίδευσης χωρίς επιτήρηση.

1. Μέθοδος εκπαίδευσης υπό επιτήρηση (Supervised Training Methods)

Αυτοί οι μέθοδοι χρησιμοποιούν ένα είδος «καθηγητή» για να βοηθήσει στην εκπαίδευση του δικτύου λέγοντας στο δίκτυο πιο είναι το επιθυμητό αποτέλεσμα. Η μέθοδος παρουσιάζεται στο Σχήμα 2-7 από όπου καταλαβαίνουμε πως το σύστημα είναι εκτεθειμένο στο περιβάλλον, το οποίο περιγράφεται από ένα πίνακα με

χαρακτηριστικά. Ο πίνακας αυτός είναι διαθέσιμος και στον «καθηγητή» ο οποίος μέσω της εμπειρίας του μπορεί να καταλάβει ποιο είναι το επιθυμητό αποτέλεσμα. Έτσι προκύπτει ένα σφάλμα μεταξύ της επιθυμητής και της υπολογισμένης τιμής, το οποίο χρησιμοποιείται για να γίνει τροποποίηση των βαρών του συστήματος εκπαίδευσης. Για κάθε set από δεδομένα εισόδου υπάρχει ένα set από δεδομένα εξόδου τα οποία χρησιμοποιούνται για την εκπαίδευση του δικτύου.

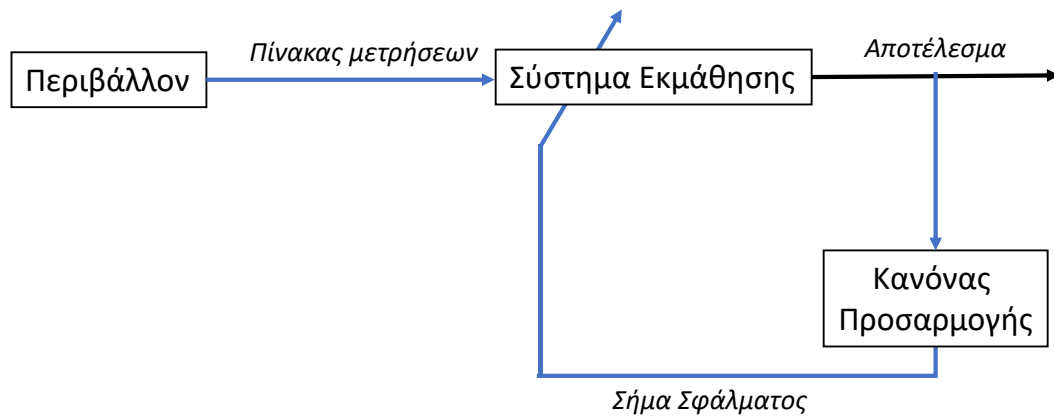
Το βασικό χαρακτηριστικό αυτής της μεθόδου είναι η αναγκαιότητα για ένα set από εισόδους με τις αντίστοιχες εξόδους. Όπως καταλαβαίνουμε υπάρχουν πολλά προβλήματα που δεν είναι εφικτό αυτό, ή ακόμα και πρακτικό.



Σχήμα 2-7: Block διάγραμμα μεθόδου εκπαίδευσης υπό επιτήρηση

2. Μέθοδος μάθησης χωρίς επιτήρηση (Unsupervised Training Methods)

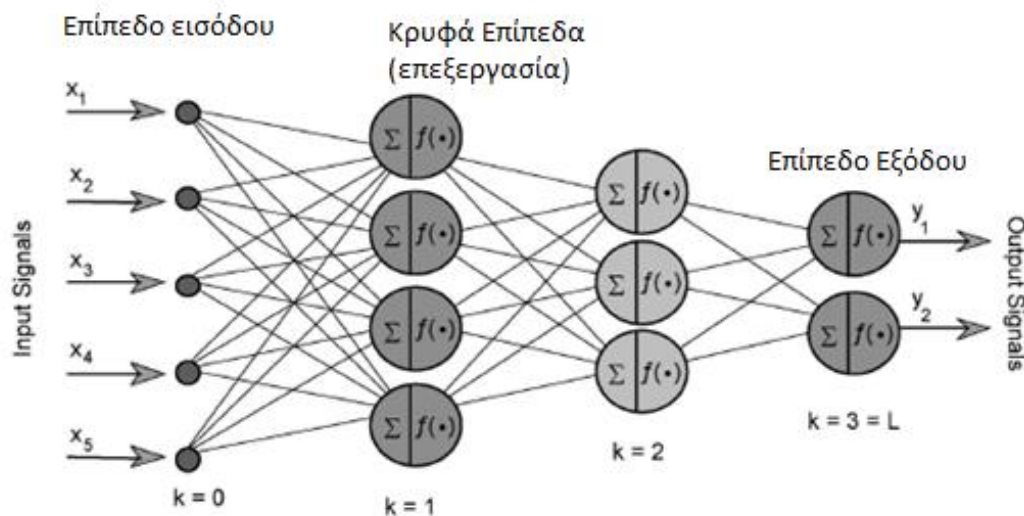
Αυτοί οι μέθοδοι είναι παρόμοιες με τις μεθόδους υπό επιτήρηση, και η διαφορά τους είναι ότι δεν υπάρχει κάποιο είδος «καθηγητή» στην διάρκεια της εκπαίδευσης. Θα μπορούσαμε να το παρομοιάσουμε με τους μαθητές που προσπαθούν να μάθουν μόνοι τους. Η διαδικασία της εκπαίδευσης είναι ένα είδος επανάληψης με ένα σετ από κανόνες προσαρμογής που είναι κύρια για την συμπεριφορά του δικτύου (Σχήμα 2-8). Οι μέθοδοι εκπαίδευσης χωρίς επιτήρηση χρησιμοποιούν και αυτοί έναν πίνακα με τα χαρακτηριστικά του περιβάλλοντος όμως δεν χρειάζεται έναν πίνακα με τις εξόδους για κάθε είσοδο. Για την δημιουργία του σήματος σφάλματος χρησιμοποιούνται οι κανόνες προσαρμογής και έχει τον ρόλο που έχει ο καθηγητής στις μεθόδους με επιτήρηση. Από τα παραπάνω καταλαβαίνουμε πως οι κανόνες προσαρμογής παίζουν πολύ σημαντικό ρόλο για την εκπαίδευση του δικτύου.



Σχήμα 2-8: Block diagram of unsupervised-training model

Feedforward Neural Network

Ένα παράδειγμα ενός απλού νευρωνικού δικτύου feedforward δίνεται στο Σχήμα 2-9. Πρόκειται για τα γνωστά δίκτυα Multilayer Perceptron (MLP).



Σχήμα 2-9: Multilayer feedforward neural network [11]

Στα αριστερά παρουσιάζεται οι εισοδοί του νευρωνικού δικτύου και κατά συνέπεια η είσοδος του πρώτου επιπέδου των νευρώνων, ενώ ακολουθούν διασυνδεδεμένα επίπεδα με νευρώνες τα οποία καταλήγουν στο επίπεδο εξόδου. Παρατηρώντας το παραπάνω σχήμα καταλαβαίνουμε πως κάθε επίπεδο είναι είσοδος για το επόμενο, για αυτόν τον λόγο ο αλγόριθμος πήρε την ονομασία feedforward. Έχει βρεθεί ότι οι συναρτήσεις μεταφοράς των νευρώνων δεν επηρεάζουν την λειτουργία του δικτύου. Τα περισσότερα δίκτυα feedforward χρησιμοποιούν σιγμοειδείς συναρτήσεις μεταφοράς αλλά μπορεί να χρησιμοποιηθεί οποιαδήποτε συνάρτηση. Ένα ακόμα σημαντικό στοιχείο των δικτύων αυτών είναι πως ένας νευρώνας του πρώτου επιπέδου μπορεί να μπει σαν είσοδος στο τρίτο (3) επίπεδο καθώς και

στο δεύτερο. Επιπρόσθετα η έξοδος από κάθε νευρώνα μπορεί να σταλθεί σε ένα ή και σε περισσότερα επίπεδα ή νευρώνες του δικτύου.

Η διαδικασία της τροποποίησης των βαρών σε ένα νευρώνα ή και στο ολόκληρο το δίκτυο είναι βασική για την αποτελεσματικότητα του δικτύου και προκύπτει μέσα από την διαδικασία της εκπαίδευσης. Υπάρχουν πολλοί μέθοδοι εκπαίδευσης ενός τεχνητού νευρωνικού δικτύου, αλλά όλοι τους βασίζονται πάνω στην τροποποίηση των βαρών μεταξύ των νευρώνων. Όταν επιλεχθεί η αρχιτεκτονική ενός δικτύου, ο σχεδιαστής πρέπει να αποφασίσει ποιο βάρος θα μεταβάλλει για να έχει το επιθυμητό αποτέλεσμα. Για να επιτύχουμε σωστή εκπαίδευση θα πρέπει να χωρίσουμε τα δεδομένα σε τρία σύνολα. Το ένα από αυτά θα χρησιμοποιηθεί για την εκπαίδευση (training), το δεύτερο για έλεγχο του επιπέδου εκπαίδευσης και την επιλογή του μοντέλου (validation), ενώ το τρίτο για την αμερόληπτη αξιολόγηση του μοντέλου (testing).

Ο αλγόριθμος Backpropagation

Ο αλγόριθμος της ανάστροφης μετάδοσης λάθους (backpropagation) είναι μία από τις πιο διαδεδομένες τεχνικές για την εκπαίδευση ενός τεχνητού νευρωνικού δικτύου. Ο στόχος του αλγορίθμου είναι να μεταβάλει τα βάρη στο δίκτυο έτσι ώστε να παίρνουμε σωστά αποτελέσματα με όσο το δυνατόν μικρότερο σφάλμα. Ο αλγόριθμος εκπαίδευσης βασίζεται στην γνωστή μέθοδο της καθόδου βασισμένη στην κλίση (gradient descent) και χρησιμοποιεί τα πρώτα παράγωγα ελαχιστοποιώντας τα έτσι ώστε να βρεθεί η βέλτιστη λύση. Λειτουργεί με έναν πίνακα που περιέχει στοιχεία εκπαίδευσης εισόδου x και έναν πίνακα με τα αντίστοιχα στοιχεία εξόδου y . Ο αλγόριθμος προσπαθεί μέσω αυτών των δεδομένων να εκπαιδεύσει το νευρωνικό δίκτυο έτσι ώστε οι έξοδοι του να έχουν περίπου ίδιες εξόδους με τον πίνακα y . Προσπαθεί δηλαδή να μειώσει το σφάλμα όσον τον δυνατόν πιο πολύ, αλλάζοντας τα βάρη των συνάψεων. [11]

$$\Delta w_{ji}^l = \sum_{p=1}^P \Delta w_{ji}^l$$

Ξεκινώντας ο αλγόριθμος θέτει κάποιες τυχαίες αρχικές τιμές για τα βάρη. Εκτελεί το εμπρόσθιο πέρασμα και για το τρέχον δεδομένα, παράγει την έξοδο και στην συνέχεια συγκρίνει την παραγόμενη έξοδο με την πραγματική και υπολογίζει το σφάλμα. Στην συνέχεια αλλάζει τα βάρη των συνάψεων εκτελώντας ανάστροφο πέρασμα σύμφωνα με το σφάλμα που έχει υπολογιστεί. Τα βάρη μπορούν να αλλάξουν μετά την ολοκλήρωση όλων των δεδομένων ή κατά την διάρκεια υπολογισμού του καθενός ξεχωριστά. Η διαδικασία αυτή επαναλαμβάνεται μέχρι το σφάλμα να φτάσει σε μία ελάχιστη τιμή. Εάν το ποσοστό μάθησης είναι πολύ χαμηλό τότε η διαδικασία της εκπαίδευσης, αλλαγής δηλαδή των τιμών από τα συνοπτικά βάρη, μπορεί να διαρκέσει περισσότερο. Εάν το ποσοστό μάθησης είναι πολύ υψηλό, ο αλγόριθμος παρουσιάζει overfitting, και το μοντέλο δεν έχει καλή γενίκευση της γνώσης.

Υπάρχουν πολλοί ακόμα παράγοντες που πρέπει να λαμβάνονται υπόψιν κατά την χρήση του αλγορίθμου αυτού. Ένας από αυτούς είναι η επιλογή της συνάρτησης ενεργοποίησης, η οποία πρέπει να έχει συγκεκριμένες ιδιότητες. Κάποιες από τις ιδιότητες που πρέπει να έχει η συνάρτηση ενεργοποίησης είναι να είναι διαφορίσιμη και μη αυξητική. Συνήθως

χρησιμοποιείται σιγμοειδής συνάρτηση ενεργοποίησης, λόγω σημαντικών ιδιοτήτων που προσφέρει.

Πλεονεκτήματα και μειονεκτήματα του αλγορίθμου Backpropagation

Κάποια από τα βασικά πλεονεκτήματα του αλγορίθμου Backpropagation [11] είναι η ευκολία χρήσης και ρύθμισης λίγων παραμέτρων και έχει εφαρμογή σε πολλά προβλήματα. Όμως η διαδικασία της μάθησης μπορεί να είναι αργή κυρίως σε δίκτυα με μεγάλο αριθμό νευρώνων και κρυφών επιπέδων. Στην βιβλιογραφία έχουν προταθεί διάφορες παραλλαγές των τεχνητών νευρωνικών δικτύων, όπως για παράδειγμα είναι τα νευρωνικά δίκτυα ακτινικής βάσης (radial basis function networks).

ii. Radial Basis Function Networks (RBFNs)

Τα δίκτυα συναρτήσεων πυρήνα χρησιμοποιούνται ευρέως για προβλήματα παρεμβολής (interpolation) και θα μπορούσαμε να πούμε πως έχοντας ένα σύνολο από m διαφορετικούς πίνακες (δεδομένα) έστω x_i , με $i = 1, 2, 3, \dots, m$, στο \mathbb{R}^n και m πραγματικοί αριθμοί, το δίκτυο βρίσκει μία συνάρτηση, έστω s με $\mathbb{R}^n \rightarrow \mathbb{R}$, που να ικανοποιεί τις συνθήκες της παρεμβολής δηλαδή

$$s(x_i) = f_i \quad i = 1, 2, 3, \dots, m \quad (2.ii.1)$$

Η συνάρτηση Φ είναι συμμετρική γύρω από ένα κέντρο μ_c . Δηλαδή ισχύει ότι $\varphi_c = \varphi(\|x - \mu_c\|)$, όπου $\|\cdot\|$ είναι συνήθως η συνάρτηση Gaussian η οποία χαρακτηρίζεται και από ένα πλάτος, σ , οπότε η παραπάνω σχέση γίνεται $\varphi_c = \varphi(\|x - \mu_c\|; \sigma)$. Ένα σετ από δίκτυα RBF μπορεί να αντιπροσωπεύει μία μεγάλη γκάμα συναρτήσεων οι οποίες έχουν κάποια γραμμική συμπεριφορά [12]. Έτσι για την έξοδο του δικτύου ισχύει:

$$\hat{y}_c = \sum_{j=1}^M w_j \varphi(\|x - \mu_j\|) \quad (2.ii.2)$$

Από την παραπάνω σχέση καταλαβαίνουμε πως ένα δίκτυο RBF δεν είναι τίποτα άλλο από ένα δίκτυο feedforward με τρία επίπεδα, το επίπεδο εισόδου, το κρυφό επίπεδο καθώς και το επίπεδο εξόδου. Κάθε κρυφή μονάδα αντιπροσωπεύει μία συνάρτηση πυρήνα που συνδέεται με ένα κέντρο και ένα πλάτος. Κάθε μονάδα εξόδου αθροίζει τις εξόδους από τις κρυφές μονάδες χρησιμοποιώντας τα βάρη. [13]

Εκπαίδευση Δικτύου RBF

Χρησιμοποιώντας μία συνάρτηση κόστους μπορούμε να ενημερώνουμε τις παραμέτρους (κέντρα και βάρη) για κάθε σετ από δεδομένα εισόδου έτσι ώστε να ελαχιστοποιούμε το κόστος.[14]

Το πρώτο βήμα είναι να εκπαιδεύσουμε το κρυφό επίπεδο. Δημιουργείται μία μη γραμμική καμπύλη κατά την διάρκεια της εκπαίδευσης. Για κάθε έναν από τους κόμβους στο κρυφό

επίπεδο, είναι απαραίτητο να βρούμε τις θέσεις των κέντρων, καθώς και τα εύρη των συναρτήσεων ακτινικής συμμετρίας [15].

Στη δεύτερη φάση της εκπαίδευσης πρέπει να ενημερώσουμε τον πίνακα με τα βάρη (w) μεταξύ των κρυφών επιπέδων και των επιπέδων εξόδου. Στα κρυφά επίπεδα κάθε κόμβος αναπαριστά κάθε συνάρτηση βάσης μετασχηματισμού (transformation basis function). Όλες οι συναρτήσεις πρέπει να ικανοποιούν διαχωριστικά την μη – γραμμικότητα ή συνδυασμός των συναρτήσεων πρέπει να ικανοποιούν διαχωριστικά την μη – γραμμικότητα. Έτσι κατά τον μετασχηματισμό του κρυφού επιπέδου εμπεριέχονται όλοι οι μη – γραμμικοί όροι.

Από τα παραπάνω συμπεραίνουμε πως κατά το πρώτο στάδιο της εκπαίδευσης, μπορεί να χρησιμοποιηθεί κάποιος αλγόριθμος μη εποπτευόμενης μάθησης μέσω του οποίου να βρεθούν ο αριθμός και οι θέσεις των κέντρων RBF. Ένας εναλλακτικός και εύκολος τρόπος εύρεσης των κέντρων του δικτύου RBF είναι να ορισθούν όσα κέντρα όσα τα δεδομένα εισόδου, με τις ίδιες συντεταγμένες. Κάτι τέτοιο όμως δεν είναι αποτελεσματικό διότι σε περίπτωση πολλών δεδομένων θα δημιουργηθεί μεγάλος αριθμός κέντρων, ενώ χρησιμοποιώντας τον παραπάνω τρόπο «ομαδοποιούνται» τα δεδομένα και βρίσκουμε ένα μέσο κέντρο μειώνοντας το υπολογιστικό κόστος. Έτσι με τον τρόπο που αναλύεται παραπάνω υπολογίζουμε την τιμή του κέντρου της ομάδας το οποίο στην συνέχεια συσχετίζεται με έναν υποδοχέα για κάθε κρυφό νευρώνα. Εάν υποθέσουμε πως έχουμε N δείγματα, πρέπει να τα διαχωρίσουμε σε M ομάδες, με $N > M$. Έτσι κάθε σμήνος εξόδου είναι υποδοχέας για τον οποίο μπορούμε να υπολογίσουμε την διαφορά ως «το άθροισμα στο τετράγωνο μεταξύ ενός σχετικού υποδοχέα και την κοντινότερη τιμή του κέντρου της πιο κοντινής ομάδας», δηλαδή $1/N \times \|X - t\|^2$.

Ομαδοποίηση με χρήση του αλγορίθμου k-means

Ένας από τους πιο γνωστούς αλγορίθμους που χρησιμοποιούμε για την ομαδοποίηση και την δημιουργία των κέντρων είναι ο αλγόριθμος k-means. Ο αλγόριθμος αυτός διαχωρίζει n παρατηρήσεις σε K ομάδες, έτσι ώστε κάθε παρατήρηση να ανήκει σε μία ομάδα η οποία χαρακτηρίζεται από ένα κέντρο. Η κατηγοριοποίηση σε ομάδες γίνεται βρίσκοντας την μικρότερη απόσταση μεταξύ των δεδομένων και των κέντρων. Η διαδικασία που ακολουθεί ο αλγόριθμος k-means είναι η ακόλουθη:

- Ορίζει έναν αριθμό από K clusters.
- Γίνεται αρχικοποίηση των κέντρων, ανακατεύοντας την βάση δεδομένων και επιλέγοντας στην τύχη K δεδομένα για τα κέντρα χωρίς αντικατάσταση.
- Επαναλαμβάνεται η διαδικασία μέχρι να μην γίνεται καμία αλλαγή στα κέντρα.
- Υπολογίζεται το άθροισμα των τετραγώνων της απόστασης των δεδομένων και όλων των κέντρων
- Ανάθεση κάθε δεδομένου στην κοντινότερη ομάδα (κέντρο)
- Υπολογισμός των κέντρων κάθε ομάδας παίρνοντας την μέση τιμή από όλα τα κέντρα που ανήκουν σε αυτήν την ομάδα.

Στην ουσία ο αλγόριθμος k-means είναι ένας αλγόριθμος ελαχιστοποίησης.

Αλγόριθμος k-means

Επανάλαβε:

Για $i = 1, 2, \dots, m$

Υπολογισμός του εγγύτερου κέντρου μ^i και ανάθεση των δεδομένων εισόδου x στο κέντρο

Τέλος

Για $j = 1, 2, \dots, K$

Ανανέωση κέντρων c_j με τα στοιχεία που ανήκουν στην ομάδα j (τελική θέση κέντρων)

Τέλος

Έως να μην γίνονται αλλαγές στα κέντρα

Εύρεση των συναπτικών βαρών w με χρήση της μεθόδου ελαχίστων τετραγώνων (Linear Least Squares)

Έχοντας υπολογίσει τις παραμέτρους του επιπέδου RBF και λόγω της γραμμικής σχέσης που υπάρχει μεταξύ του επιπέδου εξόδου και του επιπέδου RBF, μπορεί να χρησιμοποιηθεί η μέθοδος των ελαχίστων τετραγώνων για την εύρεση των συναπτικών βαρών με την ακόλουθη σχέση:

$$w^T = Y^T \cdot \Phi \cdot (\Phi^T \cdot \Phi)^{-1}$$

όπου Φ ο πίνακας που περιέχει τις εξόδους από το κρυφό επίπεδο και Y ο πίνακας που περιέχει τις επιθυμητές εξόδους.

Πλεονεκτήματα ενός δικτύου Radial Basis Function σε σχέση με ένα MLP

Υπάρχουν πολλά πλεονεκτήματα χρήσης ενός δικτύου συναρτήσεων πυρήνα (radial basis) σε σχέση με τα δίκτυα πολλαπλών επιπέδων Perceptron (MLP). Τα δίκτυα RBF χαρακτηρίζονται από σημαντικές ιδιότητες, όπως είναι η ικανότητα γενίκευσης (generalization), δηλαδή να ανταποκρίνεται αποτελεσματικά το δίκτυο σε δεδομένα που δεν έχει «δει» κατά το στάδιο της εκπαίδευσης. Μία ακόμα σημαντική ιδιότητα είναι η ικανότητα ολικής προσέγγισης (universal approximation), σύμφωνα με την οποία ένα δίκτυο RBF με ικανό αριθμό μονάδων, μπορεί να προσεγγίσει οποιαδήποτε συνάρτηση με ικανοποιητική ακρίβεια. Επίσης, λόγω της απλής δομής τους, τα δίκτυα RBF χαρακτηρίζονται από σημαντικά πλεονεκτήματα, όπως είναι οι γρήγοροι και αποτελεσματικοί αλγόριθμοι εκπαίδευσης έναντι των αντίστοιχων αλγορίθμων που εφαρμόζονται σε άλλες γνωστές αρχιτεκτονικές, όπως είναι τα δίκτυα MLP.

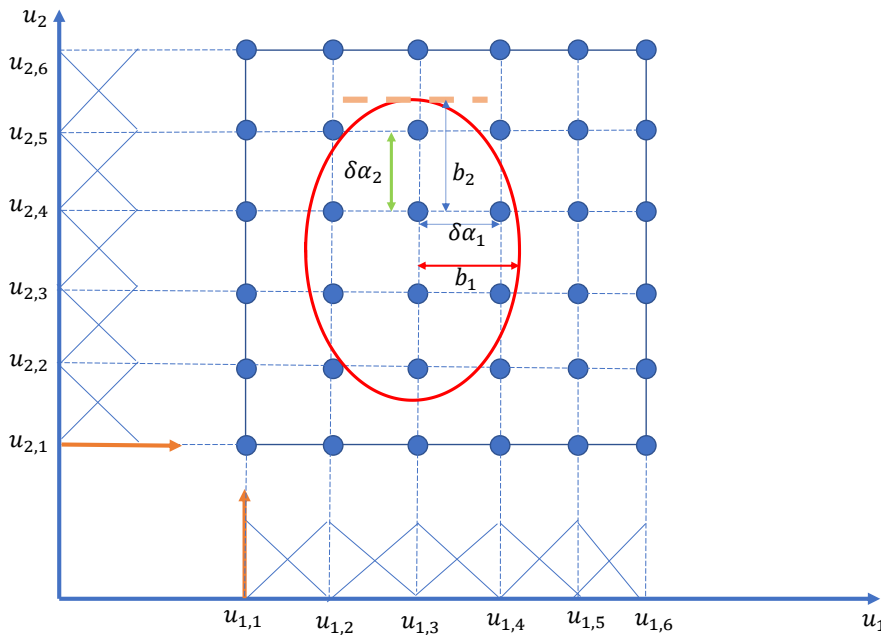
iii. Fuzzy Means

Ο αλγόριθμος Fuzzy Means (FM Algorithm) είναι μία μέθοδος που επιλέγει τα κέντρα του κρυφού επίπεδου (τόσο τον αριθμό όσο και τις τοποθεσίες) ενός δικτύου RBF με αυτόματο τρόπο [7]. Ένα από τα πιο βασικά πλεονεκτήματα είναι ακριβώς αυτή η αυτόματη εύρεση του μεγέθους του δικτύου μέσω της οποίας αποφεύγονται χρονοβόρες επαναληπτικές διαδικασίες μάθησης όπως είναι ο αλγόριθμος k-means. Έτσι ο χρόνος υπολογισμού των παραμέτρων είναι πιο γρήγορος σε σχέση με τον k-means, λόγω της μη επαναληπτικής φύσης του αλγορίθμου FM. Ο αλγόριθμος FM χρησιμοποιείται σε ένα μεγάλο φάσμα εφαρμογών όπως προβλήματα κατηγοριοποίησης, προβλήματα βελτιστοποίησης κλπ.

Έστω ότι έχουμε ένα σύστημα με N κανονικοποιημένες μεταβλητές εισόδου u_i , όπου $i = 1, 2, \dots, N$. Κάθε διάσταση του χώρου εισόδου χωρίζεται σε διαφορετικό αριθμό s_i μονοδιάστατων τριγωνικών ασαφών συνόλων. Κάθε ασαφή σύνολο μπορεί να γραφτεί ως:

$$A_{i,j} = \{a_{i,j}, \delta a_i\} \quad i = 1, 2, \dots, N \quad j = 1, 2, \dots, s_i$$

Όπου a το κέντρο του συνόλου και δa το μισό του αντίστοιχου εύρους.



Σχήμα 2-10: Μη συμμετρική αναπαράσταση διδιάστατου χώρου εισόδου με ελλειπτική συνάρτηση ομαδοποίησης

Η τιμή που έχει ένα αντικείμενο που ανήκει σε ένα ασαφές σύνολο με μία βεβαιότητα ονομάζεται βαθμός αληθείας (degree of truth) και παίρνει τιμές στο διάστημα $[0,1]$. Στην ουσία ο βαθμός αληθείας χαρακτηρίζει τον βαθμό συγγένειας του στοιχείου ενός συνόλου και η τιμή του προκύπτει από την συνάρτηση συμμετοχής (membership function)

Κάθε ασαφής υποχώρος δημιουργείται από τον συνδυασμό N 1-D ασαφών συνόλων, ένα για κάθε διάσταση του χώρου εισόδου.

Αυτήν η τεχνική δημιουργεί ένα σύνολο από S πολυδιάστατους ασαφείς υποχώρους A^l , όπου $l = 1, 2, \dots, S$.

$$S = \prod_{i=1}^N s_i$$

Ο στόχος του αλγορίθμου Fuzzy Means είναι να βρεθούν τα κέντρα του κρυφού επιπέδου σε ένα RBF δίκτυο, επιλέγοντας ένα μικρό μέρος των ασαφών υποχώρων. Το υποσύνολο αντιπροσωπεύει τα υποψήφια κέντρα RBF, και επιλέγεται έτσι ώστε να καλύπτει ομοιόμορφα τον χώρο εισόδου. Ο αλγόριθμος FM επιλέγει τα κέντρα σύμφωνα με την πολυδιάστατη συνάρτηση συγγένειας $\mu_{A^1}(u(k))$ από έναν πίνακα εισόδου $u(k)$ σε ένα fuzzy υποεπίπεδο A^1 . [16] Ισχύει η σχέση:

$$\mu_{A^1}(u(k)) = \begin{cases} 1 - d_r^l(u(k)), & \text{εάν } d_r^l(u(k)) \leq 1 \\ 0, & \text{σε όλες τις άλλες περιπτώσεις} \end{cases}$$

όπου $d_r^l(u(k))$ είναι η απόσταση μεταξύ του fuzzy set A^1 και του δεδομένου εισόδου $u(k)$.

Για κάθε δεδομένο θα προστίθεται νέο κέντρο, όταν για όλους τους υπάρχοντες υποχώρους η συνάρτηση συγγένειας είναι 0.

Σύμφωνα με προηγούμενες δημοσιεύσεις [7] η απόσταση $d_r^l(u(k))$ αποτελεί σημείο της συνάρτησης μίας έλλειψης που δημιουργείται γύρω από το κέντρο και μπορεί να υπολογιστεί από τον τύπο:

$$d_r^l(u(k)) = \frac{\sqrt{\sum_{i=1}^N (a_{i,j}^l - u_i(k))^2}}{\sqrt{N\delta\alpha}}$$

Αμα υποθέσουμε πως έχουμε ένα σύνολο δεδομένων με τις κανονικοποιημένες τιμές N των δεδομένων εισόδου K και μας δίνεται και ο ακέραιος αριθμός των ασαφών συνόλων s , χρησιμοποιώντας τον αλγόριθμο fuzzy means παίρνουμε σαν έξοδο τον αριθμό των κέντρων L καθώς και τις θέσεις αυτών $\hat{U} = [\hat{u}_1, \hat{u}_2, \dots, \hat{u}_L]^T$. Ο τρόπος εκπαίδευσης που χρησιμοποιεί ο αλγόριθμος fuzzy means θα μπορούσαμε να πούμε ότι είναι ο ακόλουθος:

Κατά την είσοδο του πρώτου δεδομένου εκπαίδευσης σχηματίζεται ο πρώτος ασαφής υποχώρος και κατά συνέπεια το πρώτο κέντρο του δικτύου. Υπολογίζεται ο βαθμός συγγένειας από τη συνάρτηση συμμετοχής με το κάθε ασαφές σύνολο σε κάθε διάσταση i (με $i = 1, 2, \dots, N$). Το ασαφές σύνολο j που δίνει τον μεγαλύτερο βαθμό συμμετοχής στη διεύθυνση i θα είναι μέρος του ασαφούς υποχώρου που θα σχηματιστεί, το οποίο στη διεύθυνση i θα έχει ως συντεταγμένη το κέντρο του συγκεκριμένου ασαφούς συνόλου. Οι υπολογισμός αυτός συνεχίζεται για όλες τις διευθύνσεις από $1, \dots, N$, όπου διαμορφώνεται ο πρώτος ασαφής υποχώρος και τελικά το πρώτο κέντρο του δικτύου. Στη συνέχεια εισάγονται τα επόμενα παραδείγματα εκπαίδευσης. Πραγματοποιείται έλεγχος αν οι υπάρχοντες ασαφείς υποχώροι καλύπτουν τα παραδείγματα εκπαίδευσης. Αν κάποιο παράδειγμα εκπαίδευσης δεν καλύπτεται από κανέναν ασαφή υποχώρο, κάτι το οποίο ελέγχεται από την Ευκλείδεια απόσταση και τελικά με την συνάρτηση συμμετοχής, τότε δεν προστίθεται νέο κέντρο L . Σε διαφορετική περίπτωση επαναλαμβάνεται η διαδικασία και προστίθεται νέο κέντρο. Ο αλγόριθμος συνεχίζει με τον ίδιο τρόπο για όλα τα δεδομένα εκπαίδευσης.

Μία διαφοροποίηση του αλγορίθμου FM, σε σχέση με κλασσικό FM που είναι συμμετρικός, είναι ότι χρησιμοποιεί μη-συμμετρικά ασαφή υποσύνολα του χώρου εισόδου (non-symmetric

fuzzy means algorithm). Σε αυτήν την εργασία θα εστιάσουμε στην συμμετρική παραλλαγή του αλγορίθμου FM. Παρακάτω δίνεται ο ψευδοκώδικας του αλγορίθμου.

Ο αλγόριθμος Fuzzy Means:

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

Το 1^ο δεδομένο: $k = 1$

Υπολογισμός του 1^{ου} κέντρου RBF, $L = 1$

Για $i = 1, 2, \dots, N$

Υπολογισμός του ασαφούς συνόλου j το οποίο δίνει τη μεγαλύτερη συγγένεια σε κάθε διεύθυνση i , δηλαδή $A_i^1 = [a_i^1, \delta a] \leftarrow \max_{j=1,2,\dots,s} [\mu_{A_{i,j}}(x_i(1))]$

Τέλος

Πρώτο κέντρο δικτύου, $\widehat{u}_1 = [a_1^1, a_2^1, \dots, a_N^1]$

Για $k = 1, 2, \dots, K$

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

τότε:

Προσθήκη νέου κέντρου, $L \leftarrow L + 1$

Για $i = 1, 2, \dots, N$

Υπολογισμός του ασαφούς συνόλου j το οποίο δίνει τη μεγαλύτερη συγγένεια σε κάθε διεύθυνση i , δηλαδή $A_i^L = [a_i^L, \delta a] \leftarrow \max_{j=1,2,\dots,s} [\mu_{A_{i,j}}(x_i(1))]$

Τέλος

Επόμενο κέντρο δικτύου $\widehat{u}_L = [a_1^L, a_2^L, \dots, a_N^L]$

Τέλος

Τέλος

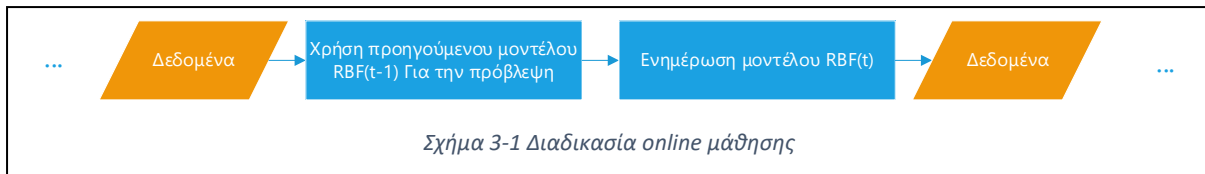
3. Online Μάθηση

i. Γενικά

Τα τελευταία χρόνια υπάρχει μεγάλη αύξηση της ψηφιοποίησης των δεδομένων καθώς και αύξηση του πλήθους των δεδομένων. Με την είσοδο των έξυπνων συσκευών καθώς και του IoT στην καθημερινότητα είναι αναμενόμενο η επιστήμη να προχωρά με ραγδαίους ρυθμούς στο κομμάτι την Online μάθησης. Πλέον οι περισσότερες συσκευές συλλέγουν δεδομένα, όπως για παράδειγμα το κινητό, τα έξυπνα ρολόγια που συλλέγουν συνεχώς δεδομένα σχετικά με την φυσική κατάσταση του χρήστη, είτε αυτά είναι οι καρδιακοί του παλμοί, το επίπεδο του οξυγόνου στο αίμα, τα βήματα που έχει κάνει και φυσικά η τοποθεσία του χρήστη. Τα παραπάνω παραδείγματα αναφέρονται στον όγκο των δεδομένων που μπορεί να προέρχονται από κάποιον χρήστη, όμως έχουν αναπτυχθεί για πολλές εφαρμογές σχετικά με την πρόβλεψη του καιρού μέσω αισθητήρων, κλιματικές αλλαγές καθώς και αλλαγές στο ηλιακό μας σύστημα. Η χρησιμότητα της μάθησης σε μη στατικά περιβάλλοντα δεν σταματάει εκεί. Λόγω της ανάλυσης των δεδομένων σε πραγματικό χρόνο η χρησιμότητα των δικτύων αυτών είναι πολύ σημαντική. Ακόμα θα μπορούσε να χρησιμοποιηθεί για στρατιωτικές εφαρμογές. Αναλύοντας τις εικόνες από drone και δορυφόρους μπορούν να γίνουν προβλέψεις για την κίνηση των αντίπαλων χωρών, την εύρεση ενός τρομοκρατικού πυρήνα μέχρι και την τροχιά που πρέπει να ακολουθήσει ένας πύραυλος για να μειωθούν οι απώλειες άμαχου πληθυσμού. Τέτοια συστήματα ήδη χρησιμοποιούνται για την εύρεση της βέλτιστης τοποθεσίας που πρέπει να συγκρουστεί κάποιος πύραυλος για περιορισμό της έκρηξης μέσα στο οίκημα.

Από τα παραπάνω καταλαβαίνουμε πως τα δεδομένα είναι πάρα πολλά και σε μερικές περιπτώσεις μπορεί να απαιτείται μεγάλη ακρίβεια για την λήψη των αποφάσεων. Επομένως κρίνεται αναγκαίο να χρησιμοποιηθούν μέθοδοι μηχανικής μάθησης σε πραγματικό χρόνο. Οι αλγόριθμοι – τεχνικές που είχαν αναπτυχθεί βασίζονται στη λογική της batch επεξεργασίας δεδομένων. Δεν μπορούν λοιπόν, τουλάχιστον χωρίς κάποιες τροποποιήσεις, να χρησιμοποιηθούν για την online μάθηση διότι χρειάζονται όλα τα δεδομένα για να εκπαιδύσουμε και να αξιολογήσουμε το μοντέλο. Ακόμα μερικοί από αυτούς τους αλγορίθμους, απαιτούν την διάσπαση των δεδομένων εισόδου σε τρία υποσύνολα (Training – Validation – Testing) έτσι ώστε να ολοκληρωθεί η διαδικασία της εκπαίδευσης και τέλος τον υπολογισμό της πρόβλεψης.

Από την άλλη, για την Online μάθηση είναι αναγκαίο ο αλγόριθμος που χρησιμοποιείται να εκπαιδεύει συνεχώς το δίκτυο έτσι ώστε να βελτιώνεται σε συνάρτηση με τον χρόνο, διότι τα δεδομένα λαμβάνονται συνεχώς σε πραγματικό χρόνο όποτε και το μοντέλο οφείλει να προσαρμόζεται καταλλήλως. Στην online μάθηση, ή αλλιώς εκμάθηση σε μη στάσιμα περιβάλλοντα, τα δεδομένα δεν είναι εξαρχής διαθέσιμα, αντιθέτως συλλέγονται συνεχώς ενώ ταυτόχρονα υπάρχει και το φαινόμενο drift [17]. Το φαινόμενο αυτό δημιουργείται από την φυσική τάση των δεδομένων να εξελίσσονται κατά την πάροδο του χρόνου, ενώ υπάρχει ο κίνδυνος μετά από κάποιο χρονικό διάστημα το μοντέλο να είναι απαρχαιωμένο εξαιτίας της συνεχούς αλλαγής των πληροφοριών της ροής. Στο Σχήμα 3-1 περιγράφεται η διαδικασία της μάθησης σε μη στατικά περιβάλλοντα.



Όμως προκύπτουν προβλήματα κατά την λήψη των δεδομένων από τον αισθητήρα. Το πιο βασικά θα μπορούσαμε να πούμε ότι είναι:

- Η μεγάλη ταχύτητα με την οποία μπορεί να φτάνουν τα δεδομένα στη ροή δεδομένων μας.
- Το περιορισμένο μέγεθος της μνήμης και οι απαιτήσεις που υπάρχουν.
- Το φαινόμενο drift.
- Η ακρίβεια των αποτελεσμάτων και η διαχείριση του χρόνου.

Μάθηση σε στάσιμα περιβάλλοντα (Off-line setting)

Όταν έχουμε ένα offline σύνολο δεδομένων χρησιμοποιούμε ένα batch αλγόριθμο για να δημιουργήσουμε το μοντέλο. Το μοντέλο αυτό χρησιμοποιεί ένα υποσύνολο των δεδομένων για να πραγματοποιήσει την εκπαίδευση. Στην συνέχεια χρησιμοποιεί ένα δεύτερο υποσύνολο για να υπολογίσει την αποτελεσματικότητά του. Με αυτόν τον τρόπο αποφεύγεται το φαινόμενο over-fitting.

Μία άλλη τεχνική που χρησιμοποιείται είναι ο αυξητικός αλγόριθμος ο οποίος αντί να χρησιμοποιεί όλα τα δεδομένα, τα χωρίζει σε πλειάδες (tuples) και εκπαιδεύει μία σειρά από μοντέλα αντίστοιχα. Έτσι εάν ονομάσουμε τα μοντέλα m θα ισχύει ότι κάθε μοντέλο m_i εκπαιδεύεται χρησιμοποιώντας το προηγούμενο μοντέλο και για ένα περιορισμένο αριθμό από πλειάδες. Θα μπορούσαμε να πούμε πως ισχύει η σχέση:

$$m_i = \text{train}(m_{i-1}, (x_i, y_i), \dots, (x_{i-p+1}, y_{i-p+1})) \quad (3.1.1)$$

Όπου, \mathbf{x} : δεδομένα εισόδου, \mathbf{y} : δεδομένα εξόδου, \mathbf{i} : αριθμός μοντέλου, \mathbf{p} : αριθμός πλειάδας .

Με αυτήν την λογική καταλαβαίνουμε πως μόνο το τελευταίο μοντέλο χρησιμοποιείται για το testing, δηλαδή να ελεγχθεί η αποτελεσματικότητά του μοντέλου.

Η παραπάνω τεχνική είναι ιδιαίτερα αποτελεσματική σε σενάρια με δεδομένα Big Data διότι τα μοντέλα εκπαιδεύονται ξανά όταν έρχεται η επόμενη πλειάδα χωρίς να υπάρχει η ανάγκη να κρατηθούν όλες οι προηγούμενες τιμές, παρά μόνο το τελευταίο μοντέλο που έχει εκπαιδευτεί. [2]

Μάθηση σε μη στάσιμα περιβάλλοντα (Online Setting)

Όταν έχουμε online setting δεν υπάρχει σύνολο αλλά ροή δεδομένων (data stream), η οποία θεωρητικά είναι άπειρη, με διάφορες πλειάδες $s_i = (x_i, y_i)$ οι οποίες φτάνουν στο μοντέλο μας διαδοχικά και δημιουργούν μία ακολουθία από $S = (s_1, s_2, \dots, s_t, \dots)$ με το t να είναι η χρονική στιγμή που έρχεται το δεδομένο το οποίο θα χρησιμοποιήσει το προηγούμενο μοντέλο για να προβλέψει το y_t με το αντίστοιχο x_t .

Μία βασική διαφορά είναι πως σε αντίθεση με το offline setting όλα τα μοντέλα χρησιμοποιούνται πρώτα για να υπολογιστεί η αποτελεσματικότητα – απόδοση, και στην συνέχεια για την εκπαίδευση. Ακόμα δεν υπάρχει διαχωρισμός σε training και testing δεδομένα διότι όλα τα δεδομένα χρησιμοποιούνται και κατά την εκπαίδευση και στην συνέχεια για την προσαρμογή του επόμενου μοντέλου.

Η απόδοση του μοντέλου θα μπορούσαμε να ισχυριστούμε πως είναι το πιο βασικό σημείο σύγκρισης μεταξύ των δύο περιπτώσεων (offline – online) διότι είναι σημαντικό ακόμα και από τις πρώτες πλειάδες δεδομένων να παράγονται σωστές προβλέψεις.

Ενδεικτικά αναφέρονται τρεις αλγόριθμοι online μάθησης οι οποίοι χρησιμοποιούνται στη βιβλιογραφία για τη μοντελοποίηση ροών δεδομένων:

- Naïve Bayes,
- Stochastic Gradient Descent
- Incremental support vector machine,

οι οποίοι αναλύονται αναλυτικά παρακάτω.

Αλγόριθμος Naive Bayes (NB_{Gauss})

Ο αλγόριθμος NB_{GAUSS} αποτελείται από παράλληλους άξονες με Gaussian κατανομή για κάθε κλάση και κάνει μία πρόβλεψη σχετικά με το πόσο τα δεδομένα ταιριάζουν στην κλάση [18]. Επειδή οι κλάσεις είναι αραιές μεταξύ τους, ο συγκεκριμένος αλγόριθμος είναι πολύ αποτελεσματικός σε σχέση με τον χρόνο εκμάθησης καθώς και απαίτησης μνήμης. Μπορεί να εκπαιδευτεί με λίγα δεδομένα [19]. Ο αλγόριθμος έχει χρησιμοποιηθεί ως φίλτρο SPAM [20] καθώς και για ταξινόμηση εγγράφων [21].

Stochastic Gradient Descent (SGD_{Lin})

Ο αλγόριθμος SGD_{Lin} είναι ένας αποτελεσματικός αλγόριθμος μάθησης με το να διακρίνει τα μοντέλα ελαχιστοποιώντας μία συνάρτηση κόστους, όπως Hinge ή Logistic. Χρησιμοποιείται για να εκπαιδευσουμε ένα γραμμικό μοντέλο ελαχιστοποιώντας την συνάρτηση κόστους. Μπορεί να έχει μεγάλη απόδοση για αραιά με πολλές διαστάσεις δεδομένα, όμως πρέπει να υπάρχει κάποια γραμμική σχέση μεταξύ τους.

Incremental support vector machine (ISVM)

Ο αλγόριθμος ISVM [22] είναι από τους πιο γνωστούς αλγορίθμους SVM (Support Vector Machine – μηχανές διανυσμάτων υποστήριξης). Χρησιμοποιεί ελάχιστα αποτελέσματα από υποψήφιους φορείς, οι οποίοι προκύπτουν ανάλογα με τα μελλοντικά δεδομένα. Όσο λιγότεροι είναι οι υποψήφιοι φορείς τόσο το πιο πιθανό να αποτύχει ο αλγόριθμος στην προαγωγή τους σε διάνυσμα υποστήριξης (support vector).

Χρησιμοποιώντας τα δίκτυα Multilayer Perceptron (MLP) για μάθηση data stream

Σύμφωνα με πρόσφατες έρευνες έχουν γίνει δοκιμές για να εφαρμοστεί κάποιος μηχανισμός forgetting σε νευρωνικά δίκτυα, ο οποίος έχει ως στόχο τη προσαρμογή του μοντέλου με την πάροδο της ροής δεδομένων. Για παράδειγμα, κάτι τέτοιο εφαρμόζεται στην εργασία [23], η οποία αναπτύσσει MLP δίκτυα με ένα κρυφό επίπεδο σταθερού μεγέθους εκπαιδευμένα με τον αλγόριθμο LBFGS (Limited memory Broyden-Fletcher-Goldfarb-Shanno) [24]. Στην αναφερόμενη έρευνα έχει παρατηρηθεί πως όταν υπάρχει ξαφνική αλλαγή στη ροή δεδομένων,

το κρυφό επίπεδο επηρεάζεται με παρόμοιο τρόπο που επηρεάζεται όταν αλλάξει το μέγεθος των δεδομένων αλλά στην προκειμένη περίπτωση λιγότεροι νευρώνες ελέγχουν την δυνατότητα εκπαίδευσης. Όταν η ροή των δεδομένων είναι πιο ομαλή έχουμε λιγότερη μείωση στην απόδοση του δικτύου. Από τα παραπάνω καταλαβαίνουμε πως μία αλλαγή στα δεδομένα εισόδου, ή ακόμα και στο μέγεθος του δικτύου (πόσοι νευρώνες υπάρχουν στο κρυφό επίπεδο) μπορεί να καθυστερήσει την απόδοση του δικτύου. Επιπλέον παρατηρήθηκε πως σε μικρότερο όγκο δεδομένων είναι πιο εύκολο να γίνει κατάταξη των δεδομένων σε κλάσεις ενώ όπως αναφέραμε και στην αρχή της παραγράφου είναι σημαντικό να υπάρχει και κάποιος μηχανισμός forgetting έτσι ώστε το δίκτυο να «ξεχνάει» τα παλιά δεδομένα και να προσαρμόζει τις παραμέτρους του σύμφωνα με την πληροφορία που υπάρχει στα τρέχοντα δεδομένα.

Αλγόριθμος εκπαίδευσης GGAP-RBF

Αυξανόμενο κριτήριο

Ο αλγόριθμος εκπαίδευσης GGAP-RBF δημιουργεί με τρόπο αυξητικό το κρυφό επίπεδο. Το δίκτυο RBF ξεκινάει χωρίς κανένα νευρώνα στο κρυφό επίπεδο. Όσο τα δεδομένα εισέρχονται ένα-ένα στο δίκτυο κατά την διάρκεια της εκπαίδευσης μπορεί να προστεθεί κάποιος καινούργιος νευρώνας σύμφωνα με το αυξανόμενο κριτήριο. Το αυξανόμενο κριτήριο αποτελείται από δύο (2) μέρη:

Το πρώτο κριτήριο βεβαιώνει πως ο καινούργιος νευρώνας θα προστεθεί εάν το δεδομένο εισόδου είναι αρκετά μακριά από τους υπάρχοντες νευρώνες.

Το δεύτερο κριτήριο είναι πως ο καινούργιος νευρώνας αποδίδει μεγαλύτερη ακρίβεια από τους προηγούμενους.

Αξίζει να σημειωθεί πως τα κέντρα του δικτύου RBF θα είναι υποσύνολο των δεδομένων, μιας και ο μηχανισμός προσθήκης κέντρου διαλέγει τα κέντρα από τα διαθέσιμα δεδομένα.

Κριτήριο αφαίρεσης

Εάν η σημαντικότητα του νευρώνα k είναι μικρότερη από την ακρίβεια e_{min} , τότε ο νευρώνας k δεν είναι σημαντικός στην λήψη των αποφάσεων και θα πρέπει να αφαιρεθεί αλλιώς θα παραμείνει στο δίκτυο. Έτσι λοιπόν θα πρέπει να υπολογίζεται η σημαντικότητα κάθε νευρώνα στο τρέχον δεδομένο και εάν είναι μικρότερη από την ακρίβεια τότε θα πρέπει να αφαιρείται από το δίκτυο. Είναι προφανές πως αυτή η διαδικασία δημιουργεί μεγάλο υπολογιστικό κόστος.

Αλλαγή παραμέτρων και κριτήριο αφαίρεσης για τον κοντινότερο νευρώνα

Για να αυξηθεί περισσότερο η ταχύτητα εκπαίδευσης έχει αποδειχθεί πως δεν χρειάζεται να γίνεται παραμετροποίηση όλων των παραμέτρων των νευρώνων παρά μόνο του κοντινότερου νευρώνα (χρησιμοποιώντας την Ευκλείδεια απόσταση) που δημιουργήθηκε από το πιο πρόσφατο δεδομένο εισόδου [25]. Εάν δεν έχει δημιουργηθεί κάποιος νέος νευρώνας τότε ο αλγόριθμος θα πρέπει να ελέγξει τον κοντινότερο σε αυτόν νευρώνα σύμφωνα με το κριτήριο αφαίρεσης.

Με τα παραπάνω επιτυγχάνεται να μειωθεί ο χρόνος εκπαίδευσης που απαιτείται χωρίς να μειώνετε η απόδοση του δικτύου RBF στις περισσότερες περιπτώσεις.

ii. Online Fuzzy Means

Εισαγωγή

Όπως αναφέρθηκε παραπάνω, στα δίκτυα RBF είναι σημαντικό να οριστεί ο αριθμός των νευρώνων στο κρυφό επίπεδο. Για να απαλλαγούμε από αυτό το πρόβλημα έγιναν αρκετές έρευνες για την ανάπτυξη αλγορίθμων εύρεσης των νευρώνων του κρυφού επιπέδου μέσω μιας αυτόματης διαδικασίας. Αυτό γίνεται με χρήση του αλγορίθμου ορθογώνιων ελαχίστων τετραγώνων (orthogonal least squares algorithm) όπως αυτός παρουσιάζεται στο [8], όπου το κρυφό επίπεδο δημιουργείται με τρόπο αυξητικό. Η δομή ενός αλγορίθμου καθώς και η παραμετροποίηση αυτού γίνεται με χρήση των παραπάνω αλγορίθμων ενώ η βελτιστοποίηση με χρήση γενετικών αλγορίθμων [26]. Οι περισσότεροι από αυτούς τους αλγορίθμους έχουν μεγάλο υπολογιστικό κόστος. Στην δημοσίευση [27] του 2002, προτείνεται μία καινούργια μέθοδος βασισμένη σε διαχωρισμό fuzzy του συνόλου εισόδου. Η συγκεκριμένη μέθοδος είναι πολύ πιο γρήγορη από τις προηγούμενες μεθόδους εκπαίδευσης δικτύων και έχει αποδειχθεί ότι είναι πιο αποτελεσματική σε μη γραμμικά συστήματα.

Ένα κοινό πρόβλημα στους παραπάνω αλγορίθμους είναι ότι δημιουργούνε στατικά δίκτυα ενώ η τεχνική εκπαίδευσης για τη μοντελοποίηση ροών δεδομένων πρέπει να μπορεί να μεταβάλλει τους παραμέτρους του δικτύου. Σε πολλές εφαρμογές θα ήταν καλύτερο ένας online αλγόριθμος εκπαίδευσης να μπορεί να μεταβάλλει τις παραμέτρους του μοντέλου, έτσι ώστε το δίκτυο να παράγει σωστές προβλέψεις, ακόμα και όταν εφαρμόζεται σε δυναμικό περιβάλλον. Τα πλεονεκτήματα που προσφέρονται από ένα τέτοιο αλγόριθμο είναι πολύ χρήσιμα για την ανάπτυξη της προσαρμοστικότητας του μοντέλου. Τέτοια μοντέλα έχουν ήδη αναπτυχθεί [28] με ένα βασικό μειονέκτημα, τα μοντέλα αυτά αποδίδουν όταν πρόκειται για συστήματα που είναι αμετάβλητα στον χρόνο.

Fuzzy Means και δίκτυα RBF

Στη δημοσίευση [29] αναπτύχθηκε ένα ολοκληρωμένο πλαίσιο για online προσαρμοστικότητα σε δίκτυα RBF που μπορούν να χρησιμοποιηθούν για δυναμική σχεδίαση χρονικών εξαρτημένων μη γραμμικών συστημάτων. Η μεθοδολογία που προτείνεται χρησιμοποιεί σχεδιασμό και δυναμικό τρόπο μεταβολής των βαρών έτσι ώστε να αλλάζει το δίκτυο με την πάροδο του χρόνου και να προσαρμόζεται σε διαφορετικές περιοχές λειτουργίας. Ο αλγόριθμος βασίζεται σε διαχωρισμό των δεδομένων εισόδου σε fuzzy sets (ασαφή σύνολα) για τον αυτόματο προσδιορισμό του κρυφού επιπέδου του δικτύου, ενώ τα βάρη που συνδέουν το επίπεδο εισόδου με το επίπεδο εξόδου ενημερώνονται με την μέθοδο των αναδρομικών ελαχίστων τετραγώνων (recursive least squares RLS).

Πιο συγκεκριμένα ο αλγόριθμος αποτελείται από δύο (2) βασικά στοιχεία:

- Την ενημέρωση των βαρών σύνδεσης του κρυφού επιπέδου και του επιπέδου εξόδου.
- Αλλαγές στην δομή του κρυφού επιπέδου.

Ακόμα στην τεχνική εκπαίδευσης που προτείνεται σε αυτήν την διπλωματική, δεν απαιτείται μία εκ των προτέρων γνώση του μεγέθους του δικτύου. Το δίκτυο ξεκινάει με μηδέν νευρώνες στο κρυφό επίπεδο και διαδοχικά τροποποιεί το επίπεδο για κάθε δεδομένο εισόδου. Είναι προφανές πως οι προβλέψεις που γίνονται στα αρχικά στάδια του δικτύου δεν είναι αξιόπιστες, αλλά όσο περισσότερα δεδομένα εισέρχονται στο δίκτυο, τόσο αυξάνεται η αξιοπιστία και η

απόδοση του συστήματος. Χρησιμοποιώντας fuzzy sets στο επίπεδο εισόδου απαλασσόμαστε από τον αλγόριθμο k-means, που αναφέρθηκε σε προηγούμενη ενότητα. Έχει αποδειχθεί πως χρησιμοποιώντας fuzzy sets στο επίπεδο εισόδου καθώς και για την εύρεση των κέντρων του δικτύου, μειώνεται κατά πολύ ο χρόνος που απαιτείται για την εκπαίδευση και γίνεται ο αλγόριθμος πιο κατάλληλος σε μη στατικά περιβάλλοντα, δηλαδή για online environments.

Η λογική του αλγορίθμου είναι η ακόλουθη:

- Αρχικοποίηση αλγορίθμου

Κατά την αρχικοποίηση, τα πεδία των μεταβλητών εισόδου χωρίζονται σε έναν αριθμό από τριγωνικά fuzzy sets s και δίνονται οι τιμές για τις παραμέτρους λειτουργίας του αλγορίθμου, οι οποίες είναι: ο αριθμός N_d που ορίζει τον χρόνο από συνεχόμενα δεδομένα που το κέντρο δεν χρησιμοποιείται πριν αφαιρεθεί από το κρυφό επίπεδο του δικτύου, ο αριθμός N_s που αντιπροσωπεύει το μέγεθος του κινούμενου «παραθύρου» το οποίο χρησιμοποιείται για την αποθήκευση προηγούμενων δεδομένων εισόδου – εξόδου, και τέλος ο παράγοντας forgetting λ που χρησιμοποιείται από την μέθοδο RLS. Μία ακόμα βασική παράμετρος του δικτύου είναι το L που δηλώνει τον αριθμό των νευρώνων του κρυφού δικτύου, που κατά την αρχικοποίηση ορίζεται μηδέν.

- Ορισμός του κέντρου του πρώτου νευρώνα του κρυφού επιπέδου καθώς και το βάρος.

Με το που έρχεται το πρώτο δεδομένο η τιμή L παίρνει την τιμή 1 και ορίζεται το fuzzy υποεπίπεδο A^1 που είναι πιο κοντά στο δεδομένο εισόδου. Στην συνέχεια υπολογίζεται το βάρος του πρώτου νευρώνα χρησιμοποιώντας τον αλγόριθμο ελαχίστων τετραγώνων..

- Έλεγχος ανάγκης προσθήκης νέου κέντρου.

Όταν ένα δεδομένο εισόδου εισέρχεται στο δίκτυο γίνεται έλεγχος αν τα υπάρχοντα κέντρα το καλύπτουν, ή υπάρχει ανάγκη δημιουργίας νέου κέντρου. Η διαδικασία ακολουθεί την εξής λογική. Εάν η απόσταση του δεδομένου εισόδου από έστω και ένα κέντρο είναι μικρότερη από ένα όριο τότε τα υπάρχοντα κέντρα «καλύπτουν» το δεδομένο εισόδου. Σε αντίθετη περίπτωση, δημιουργείται ένα νέο κέντρο καθώς και υπολογίζεται το αντίστοιχο βάρος του.

- Ορισμός της τοποθεσίας του νέου νευρώνα του κρυφού επιπέδου.

Όταν προστίθεται ένας νέος νευρώνας στο κρυφό επίπεδο η τιμή L αυξάνεται κατά 1 μονάδα και υπολογίζεται το υποσύνολο fuzzy A^L που είναι πιο κοντά στο νέο δεδομένο εισόδου χρησιμοποιώντας το κριτήριο της ελάχιστης απόστασης.

- Ανανέωση του πίνακα ιστορικού ενεργοποίησης

Αυτό το βήμα γίνεται όταν δεν προστίθεται καινούργιος νευρώνας στο κρυφό επίπεδο και ανανεώνεται ο πίνακας του ιστορικού ενεργοποίησης $h^{lo} = k$ όπου k ο αριθμός του δεδομένου εισόδου.

- Έλεγχος διαγραφής νευρώνα του κρυφού επιπέδου

Σε αυτό το βήμα υπολογίζεται η μικρότερη τιμή του πίνακα του ιστορικού ενεργοποίησης, ο οποίος ορίζει τότε ένα κέντρο χρησιμοποιήθηκε τελευταία φορά. Εάν ο αριθμός που δεν έχει

ενεργοποιηθεί το κέντρο είναι μεγαλύτερος από τον αριθμό του παραθύρου N_d , τότε θα πρέπει να διαγραφεί ο συγκεκριμένος νευρώνας.

- Διαγραφή του νευρώνα που δεν έχει χρησιμοποιηθεί

Εάν ο αλγόριθμος έχει αποφασίσει πως πρέπει να διαγραφεί ένας νευρώνας του κρυφού επιπέδου, η αντίστοιχη σειρά διαγράφεται από τον πίνακα με τα κέντρα και αφαιρείται από τον πίνακα ιστορικού ενεργοποίησης.

- Γίνεται χρήση του αλγορίθμου ελαχίστων τετραγώνων για τον υπολογισμό των νέων βαρών

Όταν διαγράφεται ή προστίθεται ένας νευρώνας στο κρυφό επίπεδο πρέπει να ξαναγίνει υπολογισμός του αντίστοιχου βάρους χρησιμοποιώντας απλά ελάχιστα τετράγωνα και όχι RLS, ενώ RLS χρησιμοποιείται στην περίπτωση που ούτε εισάγεται ούτε διαγράφεται κόμβος.

iii. Givens Fuzzy Means

Παρατηρούμε πως με την πάροδο του χρόνου όλο και περισσότερο χρησιμοποιείται η αρχιτεκτονική των δικτύων RBF σε συνδυασμό με αποτελεσματικούς αλγόριθμους για να γίνει βελτιστοποίηση τόσο των αποτελεσμάτων, όσο και του χρόνου πρόβλεψης και της υπολογιστικής ισχύος. Έχει παρατηρηθεί ακόμα ότι χρησιμοποιώντας την μέθοδο του αλγορίθμου Fuzzy Means επιτυγχάνεται μεγάλη βελτίωση στα παραπάνω στοιχεία. Το μόνο που θα μπορούσαμε να πούμε πως κάνει τον αλγόριθμο όχι και τόσο γρήγορο είναι η ανάγκη του επαναυπολογισμού των βαρών όλων των νευρώνων, κατά την προσθήκη ή την αφαίρεση αυτών, με χρήση του αλγορίθμου των ελαχίστων τετραγώνων. Στην παρούσα εργασία προσπαθήσαμε να κρατήσουμε τα θετικά της μεθόδου που προτείνεται στη δημοσίευση [29] αφαιρώντας το στοιχείο που προσθέτει κάποια περιττή καθυστέρηση, καθώς και αύξηση της υπολογιστικής ισχύς. Για να το καταφέρουμε αυτό έπρεπε να βρεθεί ένας άλλος αλγόριθμος για τον υπολογισμό των βαρών κατά την προσθήκη ή την αφαίρεση κάποιου νευρώνα, να μην χρειάζεται λοιπόν να κάνουμε ελάχιστα τετράγωνα.

Είναι γνωστό και από τα παραπάνω πως ο Online Fuzzy Means είναι ένας αλγόριθμος που μπορεί να χρησιμοποιηθεί για την μοντελοποίηση ροών δεδομένων σε δυναμικά περιβάλλοντα. Ας υποθέσουμε πως έχουμε μία ροή δεδομένων S όπως παρουσιάζεται παρακάτω:

$$S = s_1, s_2, s_3, \dots, s_t, s_{t+1} \quad (3.iii.1)$$

με

$$s_i = (x_i, y_i)$$

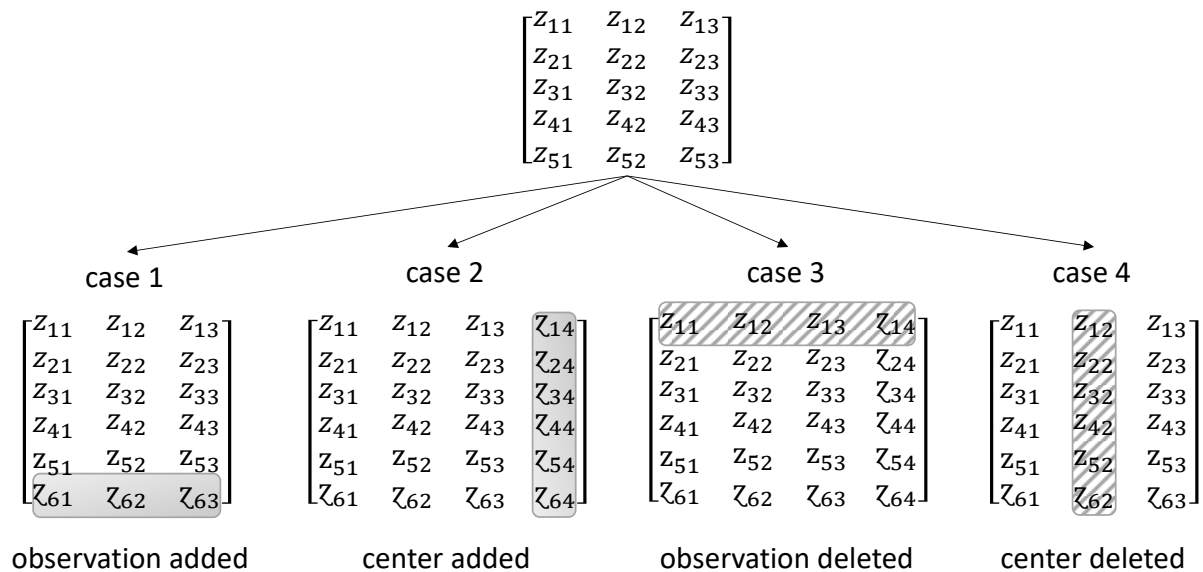
όπου x_i τα δεδομένα εισόδου ενώ y_i η κλάση στην οποία ανήκει αυτό το δεδομένο i . Ακόμα ορίζουμε ως N την διάσταση του x_i και ως M το συνολικό αριθμό των κλάσεων.

Ο αλγόριθμος που προτείνεται σε αυτήν την διπλωματική θα δημιουργεί σε κάθε χρονική στιγμή διαδοχικά μοντέλα RBF. Τα μοντέλα αυτά θα έχουν N νευρώνες εισόδου και M νευρώνες εξόδου, χωρίς να γίνεται ορισμός του μεγέθους του κρυφού επιπέδου αλλά δημιουργώντας το δυναμικά με τον αλγόριθμο Fuzzy Means (FM). Ο FM θα αποφασίζει για την προσθήκη ή την διαγραφή ενός κόμβου RBF με βάση τα εισερχόμενα δεδομένα. Κατά την είσοδο θα γίνεται έλεγχος άμα τα δεδομένα εισόδου καλύπτονται από το δίκτυο. Αν δεν καλύπτονται, θα δημιουργείται καινούργιος νευρώνας στο κρυφό επίπεδο. Στην συνέχεια γίνεται έλεγχος διαγραφής ενός περιττού κόμβου, ελέγχεται δηλαδή αν κάποιος από τους κόμβους RBF δεν συνεισφέρει στο μοντέλο έτσι ώστε να διαγραφεί. Η διαγραφή του κόμβου γίνεται, όπως και στην δημοσίευση [29], μετά από ένα συνεχόμενο χρόνο N_d ο κόμβος RBF δεν χρησιμοποιείται. Κατά αυτόν τον τρόπο υπάρχει πάντα διαθέσιμο κάποιο μοντέλο να κάνει την πρόβλεψη. Μετά την πρόβλεψη γίνεται η παραμετροποίηση των βαρών. Όπως αναφέρθηκε και παραπάνω, πραγματοποιώντας τον υπολογισμό χρησιμοποιώντας την μέθοδο των ελαχίστων τετραγώνων έχουμε μεγάλο υπολογιστικό κόστος, για αυτό προτείνεται να χρησιμοποιηθεί η παραγοντοποίηση QR σε ένα παράθυρο με σταθερό μέγεθος το οποίο ορίζεται ως N_s . Ο πίνακας λοιπόν που σχετίζεται με τα ελάχιστα τετράγωνα κάνει μικρές αλλαγές σε δύο συνεχή δεδομένα εισόδου χρησιμοποιώντας το πλαίσιο Givens rotation για την

εφαρμογή παραγοντοποίησης QR. Έστω λοιπόν Z ο πίνακας που περιέχει τα αποτελέσματα της παραγοντοποίησης QR.

Υπάρχουν τέσσερα πιθανά σενάρια για την ανανέωση του Z :

- Ένα νέο δεδομένο εισέρχεται στο σύστημα αυξάνοντας τον Z κατά μία σειρά,
- Το νέο δεδομένο που εισέρχεται δεν καλύπτεται από τα υπάρχοντα κέντρα RBF και έτσι ο Z αυξάνεται κατά μία στήλη,
- Όταν είναι γεμάτο το παράθυρο που έχει οριστεί ως N_s αφαιρείται το πιο παλιό δεδομένο μειώνοντας τον πίνακα Z κατά μία σειρά,
- Διαγράφοντας έναν κόμβο που δεν χρειάζεται μειώνεται ο πίνακας Z κατά μία στήλη.



Ο υπολογισμός των W γίνεται με χρήση του αλγορίθμου με τα ελάχιστα τετράγωνα χρησιμοποιώντας QR. Ισχύει λοιπόν η σχέση:

$$W = \tilde{R}^{-1} \tilde{Q}^T \tilde{Y}$$

Χρησιμοποιώντας της εξισώσεις που παρουσιάζονται στον αλγόριθμο Givens rotations ισχύει ότι: $Z = QR$, όπου Q είναι ένας ορθογώνιος πίνακας και R ένας άνω τριγωνικός. Σε κάθε μία από τις τέσσερις περιπτώσεις που αναφέρονται παραπάνω υπάρχει μία μικρή μετατροπή του πίνακα Z χρησιμοποιώντας Givens Rotations.

Παρακάτω παρουσιάζεται ο ψευδοκώδικας που χρησιμοποιήθηκε για τον αλγόριθμο Online Fuzzy Means Givens:

Είσοδος: Ροή Δεδομένων $S = s_1, s_2, \dots, s_t, \dots$
 N_s : παράμετρος κινούμενου παραθύρου,
 N_d : παράμετρος διαγραφής νευρώνα,
Έξοδος: RBF_t : RBF δίκτυο κατηγοροποίησης σε κάθε χρονική στιγμή t

- 1: Είσοδος πρώτης παρατήρησης
 $t = 1$
- 2: Ενημέρωση του δικτύου RBF_t
- 3: Εφαρμογή του αλγορίθμου FM για επιλογή του κέντρου RBF
- 4: Υπολογισμός των βαρών
- 5: **Επανάλαβε** για κάθε νέα παρατήρηση
 $t = t + 1$
- 6: Εύρεση εξόδου χρησιμοποιώντας το δίκτυο $RBF_{t-1}(s_t)$
- 7: Ενημέρωση του RBF_t κατηγοροποιητή
- 8: Πρώτη περίπτωση: Εφαρμογή του Givens για ανανέωση της παραγοντοποίησης QR
- 9: **Εάν** $t \geq N_s$
- 10: Αφαίρεση της παλαιότερης παρατήρησης.
- 11: Τρίτη περίπτωση: Εφαρμογή του Givens για ανανέωση της παραγοντοποίησης QR
- 12: **Εάν** η παρούσα παρατήρηση δεν καλύπτεται από τα υπάρχοντα κέντρα
- 13: Εφαρμογή του αλγορίθμου FM για δημιουργία νέων κέντρων RBF
- 14: Δεύτερη περίπτωση: Εφαρμογή του Givens για ανανέωση της παραγοντοποίησης QR
- 15: **if** $inactivity_node_k \geq N_d$
- 16: Αφαίρεση του περιττού κόμβου $node_k$
- 17: Τέταρτη Περίπτωση: Εφαρμογή του Givens για ανανέωση της παραγοντοποίησης QR
- 18: Υπολογισμός των βαρών
- 19: **μέχρι** την τελευταία παρατήρηση εισόδου στη ροή δεδομένων.

4. Πειραματικό Μέρος

i. Περιγραφή συνόλων ροών δεδομένων

Ο προτεινόμενος αλγόριθμος αξιολογήθηκε σε 9 σύνολα ροών δεδομένων της βιβλιογραφίας που αφορούν προβλήματα κατηγοριοποίησης, τόσο του πραγματικού όσο και του τεχνητού κόσμου. Στην συνέχεια, ακολουθεί μία σύντομη περιγραφή των συνόλων των ροών.

Ροές πραγματικών δεδομένων

Δύο από τις ροές που χρησιμοποιούνται βασίζονται σε πραγματικά δεδομένα. Αναλυτικά:

- Poker-Isn: είναι μία βάση δεδομένων που περιέχει τι χαρτιά μπορεί να έχει κάποιος παίκτης του poker στο χέρι του. Αποτελείται από 10 αριθμούς εισόδου και 10 κλάσεις κατηγοριοποίησης. Χρησιμοποιεί drift σε περίπτωση που αλλάζει κάποιο χαρτί από το χέρι του παίκτη.
- Weather: είναι μία βάση δεδομένων που δημιουργήθηκε από την National Oceanic and Atmospheric Administration (NOAA) της Αμερικής.

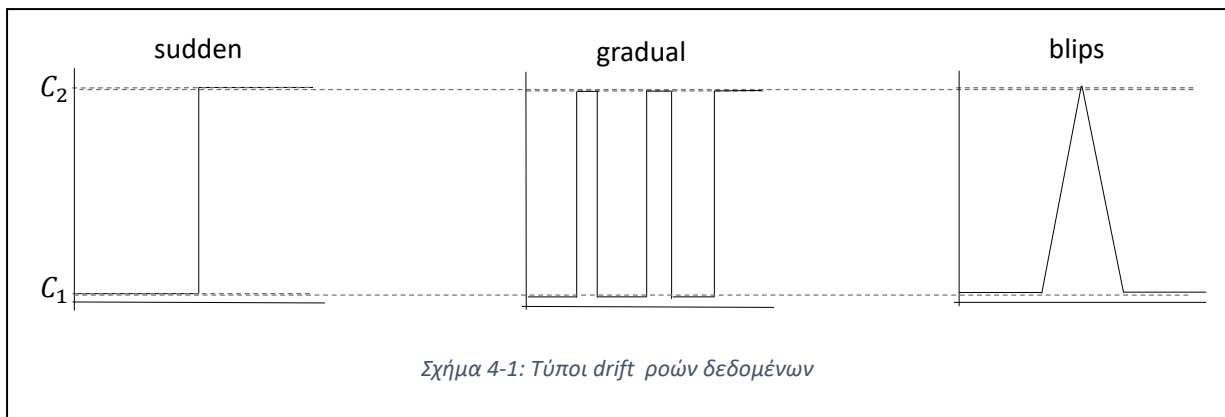
Οι υπόλοιπες ροές που χρησιμοποιήθηκαν βασίζονται σε δεδομένα που έχουν παραχθεί από τον υπολογιστή. Αναλυτικά:

- RFBBlips: είναι μία βάση δεδομένων που αποτελείται από 20 αριθμούς εισόδου με περίπου 40000 αριθμό παραδειγμάτων με drift χρησιμοποιώντας RBF.
- HyperplaneSlow: είναι μία βάση δεδομένων που αποτελείται από 5 αριθμούς εισόδου με περίπου 500000 παραδείγματα με gradual drift και δημιουργήθηκε χρησιμοποιώντας την γεννήτρια Hyperplane.
- SeaSuddenFaster: είναι μία βάση δεδομένων που αποτελείται από 3 αριθμούς εισόδου και περιέχει περίπου 1000000 παραδείγματα με sudden drift και δημιουργήθηκε χρησιμοποιώντας τον αλγόριθμο Streaming Ensemble [30].

Πίνακα 4-1 Ροές Δεδομένων που χρησιμοποιήθηκαν για τον έλεγχο του αλγορίθμου.

Dataset	Αριθμός εισόδων	Αριθμός παραδειγμάτων	Τύπος drift
RFBBlips	20	40000	Blips
HyperplaneSlow	5	500000	Gradual drift
SeaSuddenFaster	3	1000000	Sudden drift
Poker-Isn	11	1025010	Unknown
Weather	8	18159	Unknown

Οι διάφοροι τύποι drift που υπάρχουν στις παραπάνω ροές δεδομένων απεικονίζονται στο Σχήμα 4-1.



Όπως έχουμε αναφέρει και παραπάνω σε μη στατικά συστήματα δεν χρειάζεται να γίνει διαχωρισμός των δεδομένων στις 3 κατηγορίες (training, validation, testing). Τα δεδομένα χρησιμοποιούνται ένα-ένα πρώτα για πρόβλεψη και στην συνέχεια για την εκπαίδευση του δικτύου, σύμφωνα με τη μέθοδο test-then-train, η οποία έχει το πλεονέκτημα της πλήρους εκμετάλλευσης των διαθέσιμων δεδομένων. Στον παρακάτω πίνακα παρουσιάζονται κάποια βασικά στοιχεία των βάσεων αυτών.

Μέτρηση απόδοσης του συστήματος

Για την μέτρηση της απόδοσης του συστήματος ακολουθήθηκε η παρακάτω λογική. Μία θεωρητικά άπειρη σειρά από δυνάδες $S = (s_1, s_2, \dots, s_t \dots)$ με $s_i = (x_i, y_i)$ όπως αναφέραμε και παραπάνω χρησιμοποιώντας τα x_i καθώς και το προηγούμενο εκπαιδευμένο δίκτυο h_{t-1} γίνεται η πρόβλεψη $\hat{y}_t = h_{t-1}(x_t)$. Αμέσως μετά αποκαλύπτεται η πραγματική τιμή του y_t και υπολογίζεται η διαφορά $L(\hat{y}_t, y_t)$. Η απόδοση ενός μη στατικού συστήματος μέχρι μία χρονική τιμή t δίνεται από τον τύπο [2]:

$$E(S) = \frac{1}{t} \sum_{i=1}^t L(y_i, \hat{y}_i) \quad (4.i.1)$$

Όπου L μία συνάρτηση κόστους. Στην συγκεκριμένη περίπτωση ως συνάρτηση κόστους επιλέχθηκε η ακρίβεια (accuracy):

$$Accuracy = \frac{\text{σωστές προβλέψεις}}{\text{συνολο προβλέψεων}} \quad (4.i.2)$$

Είναι σημαντικό να τονιστεί σε αυτό το σημείο πως όλα τα μοντέλα χρησιμοποιούνται για την εύρεση της απόδοσης του συστήματος αλλά κάθε ένα από αυτά κάνει την πρόβλεψη για το επόμενο παράδειγμα. Και όπως αναφέρθηκε και παραπάνω δεν γίνεται διαχωρισμός των δεδομένων σε training και testing αντιθέτως όλα τα δεδομένα χρησιμοποιούνται για την εκπαίδευση και στην συνέχεια για την προσαρμογή του συστήματος.

Για την σύγκριση τις απόδοσης του αλγορίθμου χρησιμοποιήθηκαν διάφοροι αλγόριθμοι κάποιους από τους οποίους έχουμε αναφέρει και παραπάνω. Έγινε σύγκριση όλων των ροών δεδομένων με τον MRAN, Stochastic gradient descent (SGD), Adaptive Random Forest (ARF) καθώς και τον Naïve Bayes classifier (NB).

Αλγόριθμος εκπαίδευσης MRAN

Ο αλγόριθμος εκπαίδευσης MRAN [31] είναι ένας ακολουθιακός αλγόριθμος (sequential algorithm) εκπαίδευσης για δίκτυα Gaussian RBF και ορίζει τον αριθμό των κρυφών νευρώνων αυτόματα ανάλογα με κριτήρια αύξησης, ενώ την ίδια στιγμή επιτυγχάνει μικρά δίκτυα με διαγραφή κρυφών νευρώνων που δεν χρησιμοποιούνται πλέον.

Αλγόριθμος εκπαίδευσης SGD

Ο αλγόριθμος εκπαίδευσης SGD μπορεί να χαρακτηριστεί ως στοχαστική προσέγγιση μίας καθοδικής βαθμίδας βελτιστοποίησης διότι αλλάζει την πραγματική βαθμίδα με μία βαθμίδα προσέγγισης. Όταν πρόκειται για πολυδιάστατα προβλήματα βελτιστοποίησης μειώνεται το κόστος υπολογισμού, κάνοντας πιο γρήγορες επαναλήψεις αλλά με λιγότερο ποσοστό σύγκλισης.

Αλγόριθμος εκπαίδευσης ARF

Ο αλγόριθμος εκπαίδευσης ARF [3] είναι μία παραλλαγή του αλγορίθμου Random Forest [32] ο οποίος έχει χρησιμοποιηθεί στην βιβλιογραφία για να λύσει διάφορα προβλήματα μηχανικής μάθησης. Σε αντίθεση με τον RF έχει αναπτυχθεί για να μπορεί να αντιμετωπίσει το φαινόμενο drift χωρίς να υπάρχει καμία σχέση μεταξύ των δέντρων που δημιουργεί.

Αλγόριθμος εκπαίδευσης NB

Ο αλγόριθμος εκπαίδευσης Naïve Bayes αποτελείται από μία οικογένεια από πιθανολογικούς ταξινομητές χρησιμοποιώντας το θεώρημα του Bayes με μεγάλη ελευθερία μεταξύ στα χαρακτηριστικά. Οι NB classifiers είναι εξαιρετικά επεκτάσιμοι, ενώ απαιτούν έναν αριθμό παραμέτρων σε γραμμική σχέση με τον αριθμό από τις.

Παράμετροι του προτεινόμενου αλγορίθμου

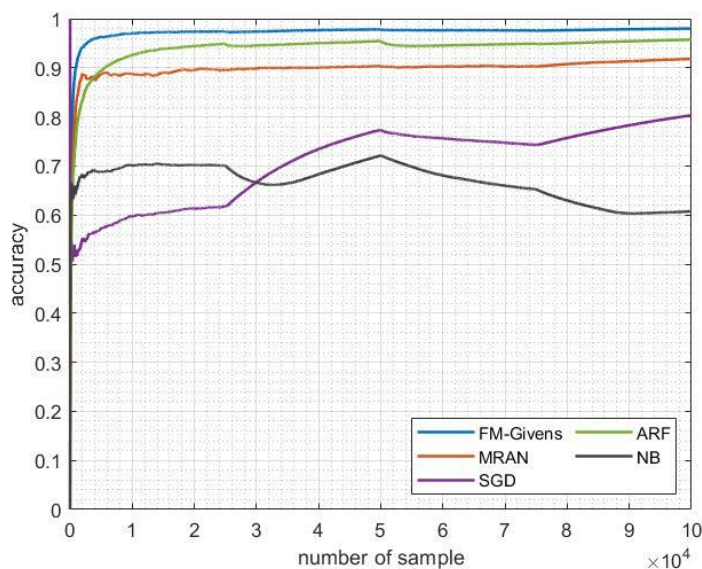
Παράμετρος	Σύμβολο	Τιμή
Αριθμός fuzzy sets	s	4:150
Αριθμός παραθύρου προηγούμενων παρατηρήσεων	N_s	1000-2000
Αριθμός επαναλήψεων απενεργοποιημένου νευρώνα	N_d	2000

Ο προτεινόμενος αλγόριθμος αποτελείται από τρεις παραμέτρους. Η πρώτη είναι ο αριθμός των fuzzy sets (s). Η δεύτερη παράμετρος (N_s) περιέχει τις τελευταίες παρατηρήσεις μέσω των οποίων γίνεται η εκπαίδευση του δικτύου. Η τρίτη και η τελευταία παράμετρος είναι ο αριθμός των διαδοχικών επαναλήψεων στις οποίες ένας κόμβος παραμένει ανενεργός. Όσο μεγαλύτερη είναι η δεύτερη παράμετρος, τόσο πιο πιθανό είναι να μην υπάρχει καλή διαχείριση του drift της ροής δεδομένων. Ακόμα αλλάζοντας τιμές στην τρίτη παράμετρο υπάρχει μεγάλη αλλαγή στην απόδοση του συστήματος. Όσο πιο μεγάλη είναι αυτή η παράμετρος τόσο περισσότερο παραμένει στο σύστημα ένας ανενεργός κόμβος αυξάνοντας το υπολογιστικό κόστος. Αξίζει να σημειωθεί πως ο αριθμός N_s βοηθάει στον επαναυπολογισμό

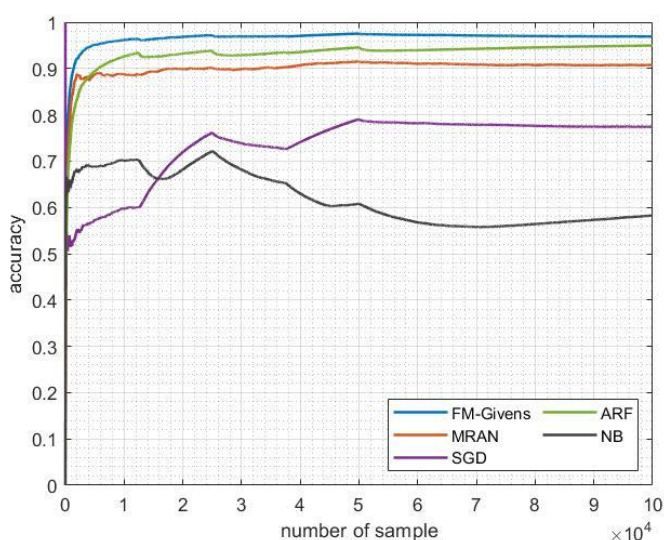
των βαρών μεταξύ του κρυφού επιπέδου και του επιπέδου εξόδου σε περίπτωση διαγραφής ή προσθήκης ενός νέου κέντρου. Έτσι χρησιμοποιώντας τις προηγούμενες τιμές εισόδου και εξόδου καθώς και την πραγματική έξοδο του συστήματος υπολογίζεται η τιμή για κάθε βάρος μεταξύ του κρυφού επιπέδου και του επιπέδου εξόδου κάθε νευρώνα. Οι τιμές των παραμέτρων που χρησιμοποιούνται σε αυτήν την μέθοδο παρουσιάζονται στον Πίνακα 4-2.

ii. Αποτελέσματα

Παρακάτω παρουσιάζονται τα διαγράμματα που προκύπτουν μετά από την εκπαίδευση με χρήση του αλγορίθμου που δημιουργήθηκε καθώς και από τους άλλους αλγορίθμους. Τα διαγράμματα απεικονίζουν την απόδοση του συστήματος που έχει δημιουργηθεί από τους διάφορους αλγορίθμους σε σχέση με τον αριθμό των δεδομένων που χρειάστηκαν για την εκπαίδευση (τα δεδομένα εισόδου δηλαδή).

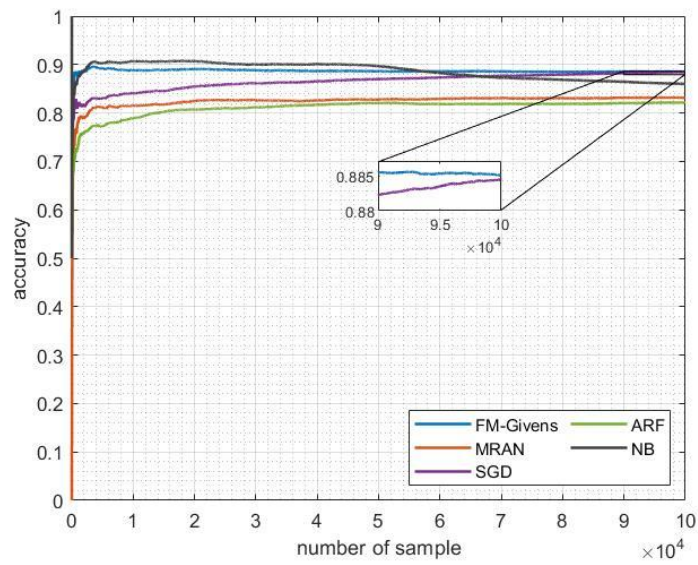


Σχήμα 4-2. Απόδοση των συστημάτων που προκύπτουν από τους διάφορους αλγορίθμους για την ροή δεδομένων (a) RBF blips.

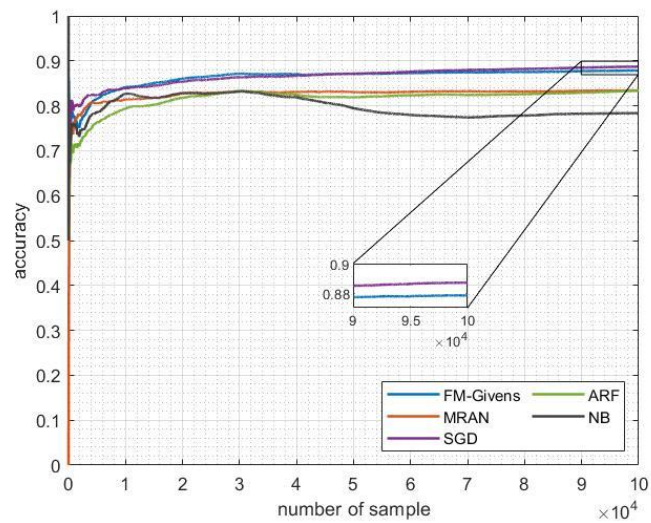


Σχήμα 4-3 Απόδοση των συστημάτων που προκύπτουν από τους διάφορους αλγορίθμους για την ροή δεδομένων (b) RBF GradualRecurring.

Ανάπτυξη προσαρμοστικών μεθόδων για online εκπαίδευση νευρωνικών δικτύων τύπου radial basis function

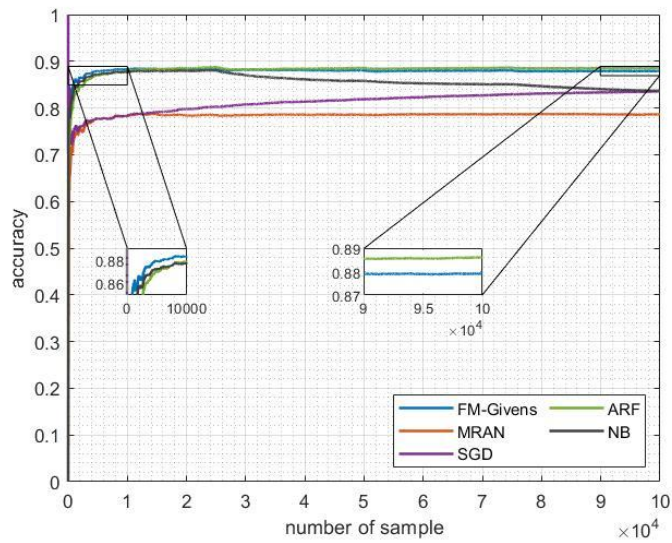


Σχήμα 4-4 (cont'd). Απόδοση των συστημάτων που προκύπτουν από τους διάφορους αλγορίθμους για την ροή δεδομένων (c) Hyperplane slow.

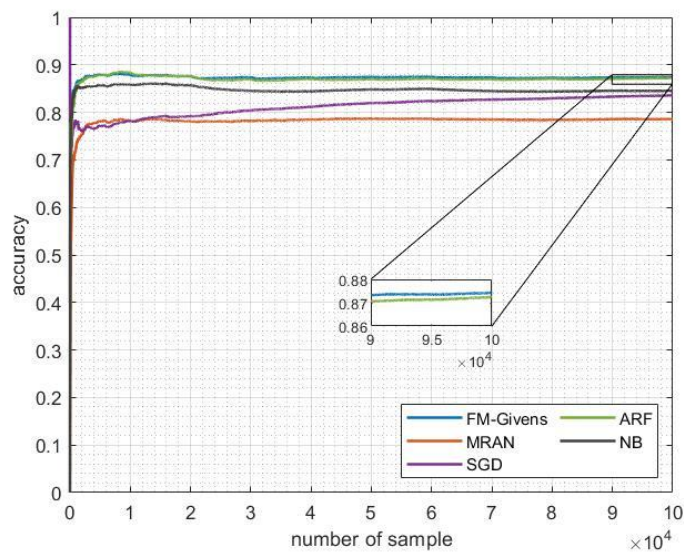


Σχήμα 4-5 (cont'd). Απόδοση των συστημάτων που προκύπτουν από τους διάφορους αλγορίθμους για την ροή δεδομένων (d) Hyperplane Faster.

Ανάπτυξη προσαρμοστικών μεθόδων για online εκπαίδευση νευρωνικών δικτύων τύπου radial basis function

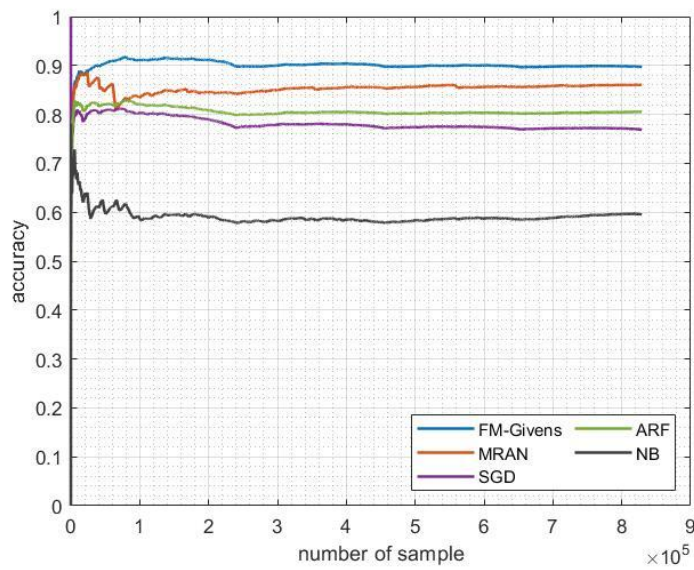


Σχήμα 4-6 (cont'd). Απόδοση των συστημάτων που προκύπτουν από τους διάφορους αλγορίθμους για την ροή δεδομένων (e) SeaSudden.

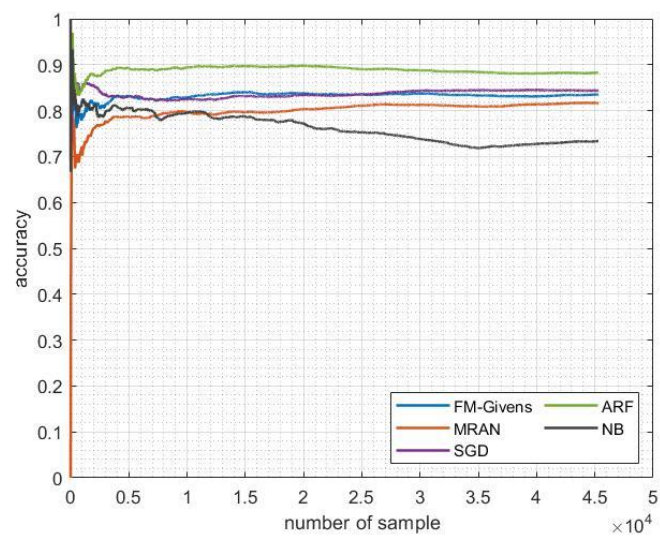


Σχήμα 4-7 (cont'd). Απόδοση των συστημάτων που προκύπτουν από τους διάφορους αλγορίθμους για την ροή δεδομένων (f) SeaSudden Faster.

Ανάπτυξη προσαρμοστικών μεθόδων για online εκπαίδευση νευρωνικών δικτύων τύπου radial basis function

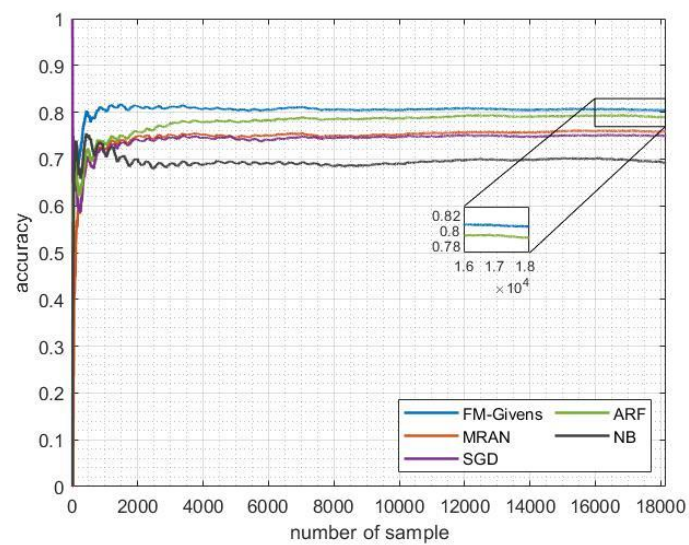


Σχήμα 4-8 Απόδοση των συστημάτων που προκύπτουν από τους διάφορους αλγορίθμους για την ροή δεδομένων (g) poker-1sn.



Σχήμα 4-9 Απόδοση των συστημάτων που προκύπτουν από τους διάφορους αλγορίθμους για την ροή δεδομένων (h) elecNormNew.

Ανάπτυξη προσαρμοστικών μεθόδων για online εκπαίδευση νευρωνικών δικτύων τύπου radial basis function



Σχήμα 4-10 Απόδοση των συστημάτων που προκύπτουν από τους διάφορους αλγορίθμους για την ροή δεδομένων (i) weather.

Πίνακα 3 - 1: Πίνακας αποτελεσμάτων

Data stream		FM-Givens	MRAN	SGD	ARF	NB
RBF_Blips	accuracy	0.9798	0.9177	0.8029	0.9570	0.6077
	nodes	163	58	-	-	-
	time (s)	7939	20673	539	5645	105
RBF_GradualRecurring	accuracy	0.9691	0.9078	0.7740	0.9496	0.5824
	nodes	176	230	-	-	-
	time (s)	7266	29126	572	9046	83
Hyperplane_Slow	accuracy	0.8851	0.8316	0.8845	0.8214	0.8600
	nodes	95	1240	-	-	-
	time (s)	16078	71670	328	24240	94
Hyperplane_Faster	accuracy	0.8783	0.8338	0.8871	0.8333	0.7837
	nodes	95	1125	-	-	-
	time (s)	16010	65553	318	23106	99
SEASudden	accuracy	0.8791	0.7862	0.8356	0.8861	0.8365
	nodes	96	466	-	-	-
	time (s)	12007	9224	319	7575	81
SEASudden_Faster	accuracy	0.8742	0.7861	0.8356	0.8724	0.8454
	nodes	131	490	-	-	-
	time (s)	13259	10124	302	10442	70
elecNormNew	accuracy	0.8353	0.8165	0.8435	0.8825	0.7336
	nodes	295	15	-	-	-
	time (s)	3411	47	77	2689	13
poker-lsn	accuracy	0.8970	0.8605	0.7686	0.8053	0.5955
	nodes	89	177	-	-	-
	time (s)	55057	206530	19793	102413	4301
weather	accuracy	0.8043	0.7590	0.7502	0.7898	0.6922
	nodes	79	142	-	-	-
	time (s)	716	845	26	1823	5

The highest performance per dataset achieved is shown in boldface type.

iii. Συζήτηση

Σχολιασμός πίνακα αποτελεσμάτων

- RBF_blips Dataset

Από τον Πίνακα 3 - 1 παρατηρούμε πως η ακρίβεια είναι μεγαλύτερη με χρήση του προτεινόμενου αλγορίθμου σε σχέση με τους άλλους αλγόριθμους. Πιο συγκεκριμένα μπορεί με τον αλγόριθμο ARF να γίνεται η εκπαίδευση σε μικρότερο χρόνο αλλά η ακρίβεια των αποτελεσμάτων αυξάνεται κατά 3%. Οι αλγόριθμοι που είναι πιο γρήγοροι όπως για παράδειγμα ο NB έχει πολύ μικρή ακρίβεια. Αξίζει να σημειωθεί πως ο αλγόριθμος MRAN χρησιμοποιεί το ένα τρίτο του αριθμού των νευρώνων του FM – Givens αλλά ο χρόνος είναι κατά πολύ μεγαλύτερος. Σύμφωνα με τα παραπάνω είναι προφανές πως στη ροή δεδομένων RBF_blips με $s = 15$, $N_s = 1000$ και $N_d = 2000$ ο αλγόριθμος παρουσιάζει καλύτερη απόδοση από τους υπολοίπους σε έναν μεσαίο χρόνο.

- Hyperplane_Slow Dataset

Χρησιμοποιώντας πάλι τα αποτελέσματα από τον Πίνακα 3 - 1 παρατηρούμε μικρό αριθμό νευρώνων ενώ η απόδοση είναι σχετικά ίδια με τον αλγόριθμο SGD ο οποίος πραγματοποιεί την εκπαίδευση σε πολύ μικρότερο χρόνο. Έτσι ο προτεινόμενος αλγόριθμος με $s = 4$, $N_s = 2000$ και $N_d = 2000$ για την ροή δεδομένων Hyperplane_slow έχει καλύτερη απόδοση από τους άλλους αλγορίθμους, εκτός από τον αλγόριθμο SGD που η απόδοση μπορεί να είναι σχετικά ίδια απαιτεί όμως περισσότερο χρόνο.

- SeaSudden_Faster Dataset

Παρατηρούμε πως ο αλγόριθμος MRAN απαιτεί περισσότερο χρόνο από τον προτεινόμενο αλγόριθμο, περισσότερους νευρώνες και έχει μικρότερη απόδοση. Όπως και στην ροή δεδομένων RBF_blips ο ARF παρουσιάζει σχεδόν ίδια απόδοση με τον προτεινόμενο σε λιγότερο χρόνο. Οι υπόλοιποι μέθοδοι (SGD, NB) παρουσιάζουν μικρότερη ακρίβεια με λιγότερο χρόνο. Συνεπώς ο προτεινόμενος αλγόριθμος με $s = 10$, $N_s = 2000$ και $N_d = 2000$ για την ροή δεδομένων SeaSudden_faster έχει καλύτερη απόδοση από τους υπόλοιπους αλλά υστερεί σε χρόνο σε σχέση με τον ARF.

- Poker-Isn Dataset

Σε αυτήν την ροή δεδομένων, που τα δεδομένα είναι πραγματικά και δεν έχουν παραχθεί στον υπολογιστή παρατηρούμε πως η απόδοση του προτεινόμενου αλγορίθμου FM – Givens είναι καλύτερη, ενώ η αμέσως μικρότερη απόδοση προκύπτει από τον αλγόριθμο MRAN ο οποίος απαιτεί περίπου 5 φορές τον χρόνο που απαιτεί ο FM – Givens και 2 φορές το πλήθος των νευρώνων που απαιτεί ο προτεινόμενος αλγόριθμος. Οι άλλοι μέθοδοι (SGD, ARF, NB) θα μπορούσαμε να πούμε πως υστερούν σημαντικά ως προς την απόδοση που δίνουν, και πιο συγκεκριμένα ο επόμενος καλύτερος σε απόδοση είναι ο ARF που ο χρόνος εκπαίδευσης είναι περίπου 3 φορές μεγαλύτερος. Καταλήγουμε στο συμπέρασμα πως για την ροή δεδομένων poker-Isn, που αποτελείται από πραγματικά δεδομένα, με $s = 4$, $N_s = 2000$ και $N_d = 2000$ ο προτεινόμενος αλγόριθμος FM – Givens είναι καλύτερος ως προς τον χρόνο, την απόδοση,

αλλά και τον αριθμό των απαιτούμενων νευρώνων σε σχέση με τους αλγόριθμους που συγκρίνονται.

- Weather Dataset

Όπως και στη ροή *roker-lsn* έτσι και εδώ πρόκειται για ροή πραγματικών δεδομένων, και όχι δεδομένων που έχουν παραχθεί από τον υπολογιστή, ο προτεινόμενος αλγόριθμος FM – Givens μας δίνει την καλύτερη απόδοση με μικρότερο αριθμό νευρώνων καθώς και χρόνο από τον αμέσως καλύτερο σε απόδοση MRAN. Έτσι για την ροή δεδομένων *weather*, που αποτελείται από πραγματικά δεδομένα, με $s = 12$, $N_s = 1000$, $N_d = 2000$, ο προτεινόμενος αλγόριθμος FM – Givens είναι καλύτερος ως προς την απόδοση, τον αριθμό των απαιτούμενων νευρώνων και τον χρόνο από τους άλλους αλγόριθμους.

Σχολιασμός σχημάτων

Εκτός από τα αποτελέσματα του Πίνακα 3 - 1 αξίζει να σχολιασθούν και τα διαγράμματα που απεικονίζονται στα Σχήμα 4-2, Σχήμα 4-3, Σχήμα 4-4, Σχήμα 4-5, Σχήμα 4-6, Σχήμα 4-7, Σχήμα 4-8, Σχήμα 4-9, Σχήμα 4-10. Καταλαβαίνουμε πως οι αλγόριθμοι παρουσιάζουν, κατά την έναρξη λήψης πληροφοριών, μία αυξητική απόδοση η οποία αρχίζει και μειώνεται κατά την άφιξη περισσότερων παρατηρήσεων.

Αναλυτικά θα μπορούσαμε να πούμε πως στο Σχήμα 4-2 οι αλγόριθμοι NB και SGD παρουσιάζουν μεγάλη αυξομείωση στην τιμή της απόδοσης, κάτι το οποίο οφείλεται στο φαινόμενο *drift*. Έτσι βγάζουμε το συμπέρασμα πως οι συγκεκριμένοι αλγόριθμοι δεν μπορούν να διαχειριστούν το *drift* που υπάρχει. Το ίδιο παρατηρείται και στο Σχήμα 4-5 που έχει τον ίδιο τύπο *drift*. Οι υπόλοιποι αλγόριθμοι που δοκιμάστηκαν φαίνεται πως διαχειρίζονται καλύτερα τον τύπο *drift* των ροών δεδομένων *roker-lsn* και RBF blips.

Πηγαίνοντας στον *gradual* τύπο *drift* που παρουσιάζεται στη ροή δεδομένων *Hyperplane Slow* παρατηρούμε στο Σχήμα 4-4 πως οι αλγόριθμοι MRAN, ARF παρουσιάζουν σχεδόν την ίδια απόδοση με μία μικρή καθοδική πορεία. Ο αλγόριθμος NB φαίνεται πως στην αρχή παρουσιάζει καλή απόδοση, αλλά με την αύξηση των παρατηρήσεων έχει μεγαλύτερη καθοδική πορεία, κάτι το οποίο μας δείχνει πως ο συγκεκριμένος αλγόριθμος δεν μπορεί να παράγει αξιοποιήσιμα αποτελέσματα λόγω της κακής απόδοσης κατά την αύξηση του αριθμού παρατηρήσεων. Κάτι αντίθετο θα μπορούσαμε να ισχυριστούμε πως συμβαίνει με τον αλγόριθμο SGD ο οποίος μπορεί στις πρώτες παρατηρήσεις να έχει χαμηλή απόδοση, κατά την αύξηση αυτών όμως βελτιώνει την απόδοσή του και παρουσιάζει μία ανοδική πορεία. Ο προτεινόμενος αλγόριθμος παρατηρούμε πως από τα πρώτα κίολας δεδομένα έχει καλή απόδοση η οποία θα μπορούσαμε να πούμε πως διατηρείται κατά την αύξηση των δεδομένων κάτι το οποίο μας φανερώνει πως ο FM – Givens διαχειρίζεται καλύτερα τον *gradual* τύπο *drift*.

Στην συνέχεια στον *sudden* τύπο *drift* που παρουσιάζεται στην ροή δεδομένων *SeaSudden Faster* παρατηρούμε στο Σχήμα 4-7 πως οι αλγόριθμοι που μπορούν να διαχειριστούν καλύτερα αυτόν τον τύπο *drift* είναι ο FM – Givens και ο ARF. Η απόδοση των υπολοίπων είναι μικρότερη με τον SGD να αυξάνει την απόδοσή του κατά την μεγάλη αύξηση των δεδομένων. Ο FM – Givens φαίνεται πως παρουσιάζει την πιο καλή απόδοση, όπως παρατηρούμε και στον Πίνακα 3 - 1. Έτσι ο FM – Givens είναι αυτός που διαχειρίζεται καλύτερα ροές δεδομένων που παρουσιάζουν *drift* τύπου *sudden*.

Τελικά συγκρίνοντας όλα τα σχήματα αυτής της ενότητας που παρουσιάζουν διαφορετικό τύπο drift, όπως φαίνεται και στον Πίνακα 4-1, ακόμα και για ροή που δεν έχει καθόλου drift ο FM – Givens είναι αυτός που παρουσιάζει την καλύτερη απόδοση ακόμα και με λίγες παρατηρήσεις.

5. Συμπεράσματα

Η μάθηση σε δυναμικά περιβάλλοντα εμφανίζεται ως μία σημαντική πρόκληση. Στην βιβλιογραφία έχουν αναπτυχθεί διάφορες προσαρμοστικές τεχνικές, ανάμεσά τους και μοντέλα νευρωνικών δικτύων. Σε αντίθεση με τα στατικά περιβάλλοντα τα δεδομένα εισάγονται στο μοντέλο σαν ροές με αποτέλεσμα το σύστημα να πρέπει να μεταβάλλεται συνεχώς αλλάζοντας όχι μόνο τις παραμέτρους του δικτύου αλλά και των αριθμό των νευρώνων και των συναπτικών τους βαρών. Οι κλαστικοί αλγόριθμοι καταφέρνουν να προσαρμόζονται σε τέτοια περιβάλλοντα χωρίς όμως να μπορούν να αντιμετωπίσουν το φαινόμενο drift που υπάρχει στις ροές δεδομένων, καταναλώνοντας μεγάλη υπολογιστική ισχύ ενώ απαιτούν περισσότερο αριθμό παρατηρήσεων για την εκπαίδευση. Στόχος αυτής της διπλωματικής είναι η ανάπτυξη ενός νέου προσαρμοστικού αλγορίθμου εκπαίδευσης νευρωνικών δικτύων τύπου RBF που να μπορεί να ανταπεξέλθει στις δυσκολίες που παρουσιάζει η μάθηση σε δυναμικά περιβάλλοντα αντιμετωπίζοντας και το φαινόμενο drift.

Ο προτεινόμενος προσαρμοστικός αλγόριθμος εκπαίδευσης Fuzzy – Givens αξιολογήθηκε σε 9 σύνολα ροών δεδομένων της βιβλιογραφίας και μέσω των πειραμάτων που πραγματοποιήθηκαν καταλαβαίνουμε πως είναι πολύ καλύτερος από τους άλλους αλγορίθμους που συγκρίνεται (MRAN, SGD, ARF, NB) ως προς την απόδοση, τον αριθμό των απαιτούμενων νευρώνων αλλά και στον χρόνο που απαιτείται για την παραμετροποίηση του μοντέλου.

Σε προσομοιωμένες ροές δεδομένων που δημιουργήθηκαν από τον υπολογιστή παρατηρούμε πως πάλι ο προτεινόμενος αλγόριθμος FM – Givens είναι πιο αποδοτικός από τους υπόλοιπους ενώ η δυσκολία που υπάρχει στους άλλους αλγορίθμους, λόγω του drift, αντιμετωπίζεται.

Ο αλγόριθμος που προτείνεται μπορεί να προσαρμόζει τόσο την δομή του όσο και τις παραμέτρους του δικτύου κατά την διάρκεια εισαγωγής νέων δεδομένων από την ροή. Βασίζεται στην συνάρτηση radial basis function που είναι γνωστή για την απλότητα της δομής των δικτύων που δημιουργεί, καθώς και για την απόδοσή τους. Όταν χρησιμοποιείται για μάθηση σε μη στατικά περιβάλλοντα το μοντέλο του δικτύου RBF προσαρμόζεται συνεχώς στα καινούργια δεδομένα έτσι ώστε να αντιμετωπίζει τα drifts. Ακόμα ελέγχει αν υπάρχει ανάγκη προσθήκης ή αφαίρεσης κάποιου κόμβου κατά την διάρκεια της εισαγωγής νέων δεδομένων στο δίκτυο. Τα συναπτικά βάρη προσαρμόζονται με την βοήθεια των Givens rotations με αποτέλεσμα ο αλγόριθμος να είναι πιο αποδοτικός από τους προηγούμενους. Όπως αναφέρουμε και παραπάνω η απόδοση του αλγορίθμου σε μη στατικά περιβάλλοντα αξιολογήθηκε σε διαφορετικές benchmark ροές δεδομένων της βιβλιογραφίας διαφορετικού τύπου drift, τεχνητών προβλημάτων και μη, καθώς και διαφορετικού μεγέθους. Σε σχέση με άλλους αλγορίθμους παρουσιάζει μεγάλη βελτίωση του χρόνου εκπαίδευσης, των απαιτούμενων δεδομένων που χρειάζεται για αυτήν καθώς και της απόδοσης. Σύμφωνα με το παραπάνω κεφάλαιο καταλαβαίνουμε πως μπορεί να ανταποκρίνεται σε διαφορετικούς τύπους drift, ένα μεγάλο πλεονέκτημα σε σχέση με τους άλλους αλγορίθμους. Κάνοντας βελτιστοποίηση των λειτουργικών παραμέτρων, η απόδοση του δικτύου κυμαίνεται από 80% - 98% χρησιμοποιώντας λιγότερους κόμβους από τους άλλους αλγορίθμους.

Αξίζει να σημειωθεί πως κάνοντας περισσότερες δοκιμές βαθμονόμησης με συνδυασμούς των λειτουργικών παραμέτρων, υπάρχει μεγάλη πιθανότητα ο προτεινόμενος αλγόριθμος Fuzzy Means – Givens (FM-Givens) να γίνει ακόμα πιο αποδοτικός.

6. Βιβλιογραφία

- [1] J. Stefanowski and D. Brzezinski, “Stream Classification,” in *Encyclopedia of Machine Learning and Data Mining*, Boston, MA: Springer US, 2017, pp. 1191–1199.
- [2] V. Losing, B. Hammer, and H. Wersing, “Incremental on-line learning: A review and comparison of state of the art algorithms,” *Neurocomputing*, vol. 275, pp. 1261–1274, 2018, doi: 10.1016/j.neucom.2017.06.084.
- [3] H. M. Gomes *et al.*, “Adaptive random forests for evolving data stream classification,” *Mach. Learn.*, vol. 106, no. 9–10, pp. 1469–1495, Oct. 2017, doi: 10.1007/s10994-017-5642-8.
- [4] C. Salperwyck, V. Lemaire, and C. Hue, “Incremental Weighted Naive Bays Classifiers for Data Stream,” 2015, pp. 179–190.
- [5] E. Jothimurugesan, A. Tahmasbi, P. Gibbons, and S. Tirthapura, “Variance-Reduced Stochastic Gradient Descent on Streaming Data,” in *Advances in Neural Information Processing Systems*, 2018, vol. 31, [Online]. Available: <https://proceedings.neurips.cc/paper/2018/file/cebd648f9146a6345d604ab093b02c73-Paper.pdf>.
- [6] S. Haykin, *Neural Networks: A Comprehensive Foundation (3rd Edition)*, vol. 13, no. 4. 1999.
- [7] A. Alexandridis, E. Chondrodima, and H. Sarimveis, “Radial basis function network training using a nonsymmetric partition of the input space and particle swarm optimization,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 24, no. 2, pp. 219–230, 2013, doi: 10.1109/TNNLS.2012.2227794.
- [8] S. CHEN, S. A. BILLINGS, C. F. N. COWAN, and P. M. GRANT, “Practical identification of NARMAX models using radial basis functions,” *Int. J. Control*, vol. 52, no. 6, pp. 1327–1350, Dec. 1990, doi: 10.1080/00207179008953599.
- [9] D. F. Mayers, G. H. Golub, and C. F. van Loan, “Matrix Computations.,” *Mathematics of Computation*, vol. 47, no. 175. p. 376, 1986, doi: 10.2307/2008107.
- [10] S. Haykin, “Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7.,” *The Knowledge Engineering Review*. 1994, doi: 10.1017/s0269888998214044.
- [11] K. L. Priddy and P. E. Keller, *Artificial Neural Networks: An Introduction*. 2009.
- [12] Y. Lu, N. Sundararajan, and P. Saratchandran, “Performance evaluation of a sequential minimal Radial Basis Function (RBF) neural network learning algorithm,” *IEEE Transactions on Neural Networks*, vol. 9, no. 2. pp. 308–318, 1998, doi: 10.1109/72.661125.
- [13] A. Alexandridis, “Evolving RBF neural networks for adaptive soft-sensor design,” *Int. J. Neural Syst.*, vol. 23, no. 6, 2013, doi: 10.1142/S0129065713500299.
- [14] J. Liu, *Radial basis function (RBF) neural network control for mechanical systems: Design, analysis and matlab simulation*, vol. 9783642348. 2013.
- [15] R. Chandradevan, “Radial basis functions Neural Networks - All we need to know,”

- Towards Data Science*, 2017. <https://towardsdatascience.com/radial-basis-functions-neural-networks-all-we-need-to-know-9a88cc053448>.
- [16] J. Nie, “Fuzzy control of multivariable nonlinear servomechanisms with explicit decoupling scheme,” *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 2, pp. 304–311, May 1997, doi: 10.1109/91.580803.
- [17] J. Gama, R. Sebastião, and P. P. Rodrigues, “On evaluating stream learning algorithms,” *Mach. Learn.*, vol. 90, no. 3, pp. 317–346, Mar. 2013, doi: 10.1007/s10994-012-5320-9.
- [18] H. ZHANG, “EXPLORING CONDITIONS FOR THE OPTIMALITY OF NAÏVE BAYES,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 19, no. 02, pp. 183–198, Mar. 2005, doi: 10.1142/S0218001405003983.
- [19] C. Salperwyck and V. Lemaire, “Learning with few examples: An empirical study on leading classifiers,” in *The 2011 International Joint Conference on Neural Networks*, Jul. 2011, pp. 1010–1019, doi: 10.1109/IJCNN.2011.6033333.
- [20] V. Metsis, I. Androutsopoulos, and G. Paliouras, “Spam Filtering with Naive Bayes - Which Naive Bayes?,” 2006, [Online]. Available: <http://www.ceas.cc/2006/listabs.html%5C#15.pdf>.
- [21] S. L. Ting, W. H. Ip, and A. H. C. Tsang, “Is Naïve bayes a good classifier for document classification?,” *Int. J. Softw. Eng. its Appl.*, vol. 5, no. 3, pp. 37–46, 2011.
- [22] T. Poggio and G. Cauwenberghs, “Incremental and decremental support vector machine learning,” *Adv. Neural Inf. Process. Syst.*, vol. 13, pp. 409–412, 2001.
- [23] P. Ksieniewicz, M. Woźniak, B. Cyganek, A. Kasprzak, and K. Walkowiak, “Data stream classification using active learned neural networks,” *Neurocomputing*, vol. 353, pp. 74–82, 2019, doi: 10.1016/j.neucom.2018.05.130.
- [24] A. Mokhtari and A. Ribeiro, “Global convergence of online limited memory BFGS,” *J. Mach. Learn. Res.*, vol. 16, pp. 3151–3181, 2015.
- [25] G. Bin Huang, P. Saratchandran, and N. Sundararajan, “A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation,” *IEEE Trans. Neural Networks*, vol. 16, no. 1, pp. 57–67, 2005, doi: 10.1109/TNN.2004.836241.
- [26] S. A. Billings and G. L. Zheng, “Radial basis function network configuration using genetic algorithms,” *Neural Networks*, vol. 8, no. 6, pp. 877–890, Jan. 1995, doi: 10.1016/0893-6080(95)00029-Y.
- [27] H. Sarimveis, A. Alexandridis, G. Tsekouras, and G. Bafas, “A Fast and Efficient Algorithm for Training Radial Basis Function Neural Networks Based on a Fuzzy Partition of the Input Space,” *Ind. Eng. Chem. Res.*, vol. 41, no. 4, pp. 751–759, Feb. 2002, doi: 10.1021/ie010263h.
- [28] M. Pottmann and D. E. Seborg, “A nonlinear predictive control strategy based on radial basis function models,” *Comput. Chem. Eng.*, vol. 21, no. 9, pp. 965–980, Jun. 1997, doi: 10.1016/S0098-1354(96)00340-7.
- [29] H. Sarimveis, A. Alexandridis, S. Mazarakis, and G. Bafas, “A new algorithm for developing dynamic radial basis function neural network models based on genetic algorithms,” in *Computer Aided Chemical Engineering*, vol. 10, no. C, 2002, pp. 949–

954.

- [30] W. N. Street and Y. Kim, “A streaming ensemble algorithm (SEA) for large-scale classification,” in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01*, 2001, pp. 377–382, doi: 10.1145/502512.502568.
- [31] G. G. Amenu, M. Markus, P. Kumar, and M. Demissie, “Hydrologic Applications of MRAN Algorithm,” *J. Hydrol. Eng.*, vol. 12, no. 1, pp. 124–129, 2007, doi: 10.1061/(asce)1084-0699(2007)12:1(124).
- [32] L. Breiman, “Random Forest,” *Mach. Learn.*, no. 45, pp. 5–32, 2001, doi: <https://doi.org/10.1023/A:1010933404324>.