

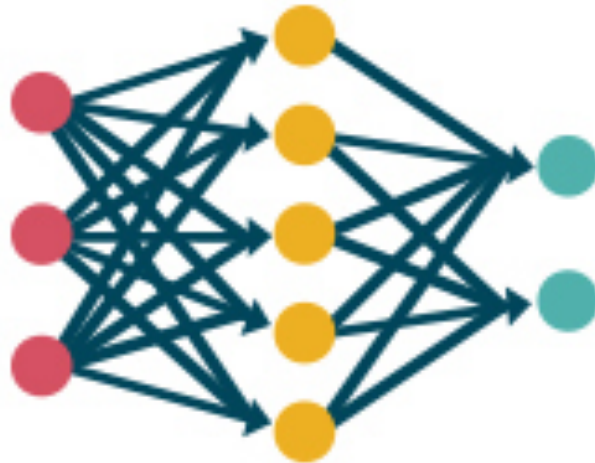
Υπολογιστική Νοημοσύνη

Εργασία 4 - Classification

Καράμπελας Σάββας

AEM:9005

savvaska@ece.auth.gr



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Περιγραφή Εργασίας

Στα πλαίσια της εργασίας αυτής στόχος μας είναι να διερευνήσουμε την ικανότητα των μοντέλων TSK στην επίλυση προβλημάτων ταξινόμησης(classification). Συγκεκριμένα, επιλέγονται δύο σύνολα δεδομένων από το UCI repository με σκοπό τη ταξινόμηση, από τα διαθέσιμα δεδομένα, δειγμάτων στις εκάστοτε κλάσεις τους, με χρήση ασαφών νευρωνικών μοντέλων. Η εργασία αποτελείται από δύο μέρη, το πρώτο από τα οποία προορίζεται για μια απλή διερεύνηση της διαδικασίας εκπαίδευσης και αξιολόγησης των TSK μοντέλων, ενώ το δεύτερο περιλαμβάνει μια πιο συστηματική προσέγγιση στο πρόβλημα της εκμάθησης από δεδομένα, σε συνδυασμό με προεπεξεργαστικά βήματα όπως επιλογή χαρακτηριστικών(feature selection) και μεθόδους βελτιστοποίησης των μοντέλων μέσω της διασταυρωμένης επικύρωσης(cross validation).

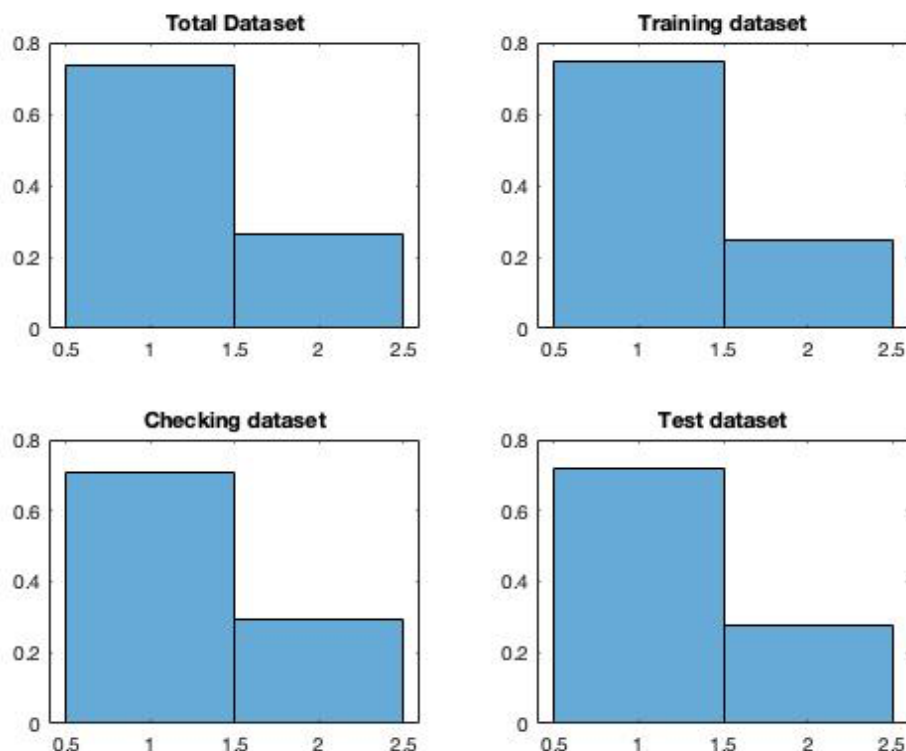
Εφαρμογή σε Απλό dataset

Στη πρώτη φάση της εργασίας, επιλέγουμε από το UCI repository, το Haberman's Survival dataset, το οποίο περιλαμβάνει 306 δείγματα(instances) με 3 χαρακτηριστικά(attributes) το καθένα. Σαν έξοδο έχουμε μόνο μία τιμή που περιέχει τον αριθμό της κλάσης που ανήκει το δείγμα. Παρατηρώντας το σύνολο δεδομένων πριν ακόμα ξεκινήσουμε, βλέπουμε ότι έχουμε μόνο δύο κλάσεις στις οποίες μπορεί να ανήκει ένα δείγμα. Τα βήματα τα οποία θα ακολουθήσουμε είναι τα παρακάτω:

- 1) Θα διαχωρίσουμε το dataset σε σύνολα εκπαίδευσης, επικύρωσης και ελέγχου. Είναι απαραίτητο τα τρία αυτά σύνολα να είναι μη επικαλυπτόμενα. Το σύνολο εκπαίδευσης θα είναι τα δεδομένα με βάση τα οποία θα λάβει χώρα η εκπαίδευση, το σύνολο επικύρωσης θα χρησιμοποιηθεί για επικύρωση και αποφυγή του φαινομένου υπερεκπαίδευσης και το τελευταίο χρησιμοποιείται για τον έλεγχο της απόδοσης του τελικού μοντέλου που προκύπτει. Αναλυτικότερα, θα έχουμε το 60% των δειγμάτων ως υποσύνολο εκπαίδευσης και από 20 % του συνόλου των δειγμάτων για κάθε ένα από τα δύο εναπομείναντα υποσύνολα. Τα δεδομένα μας τα κανονικοποιούμε με κατάλληλη συνάρτηση του MATLAB ώστε να λαμβάνουν τιμές εντός του διαστήματος $[0,1]$.

Σε αυτό το σημείο είναι μεγάλης σημασίας να επιβεβαιώσουμε με έναν έλεγχο ότι η συχνότητα εμφάνισης δειγμάτων που ανήκουν σε μία

συγκεκριμένη κλάση, σε καθένα από τα τρία σύνολα διαμέρισης, να είναι όσο το δυνατόν πιο «όμοια» με την αντίστοιχη συχνότητα εμφάνισης τους στο αρχικό σύνολο δεδομένων. Αυτό επιβεβαιώνεται με τη σχεδίαση του ιστογράμματος εμφάνισης των κλάσεων τόσο για το αρχικό σύνολο δεδομένων όσο και για τα τρία υποσύνολα.



Σχήμα 1: Ιστογράμμοι συχνότητας εμφάνισης των κλάσεων στα σύνολα δεδομένων

2) Θα διεξάγουμε διαδικασία εκπαίδευσης των TSK μοντέλων με διαφορετικές παραμέτρους και θα εξεταστούν όσον αφορά την απόδοση τους στο σύνολο ελέγχου. Συγκεκριμένα, θα εκπαιδευτούν τέσσερα TSK μοντέλα, στα οποία θα μεταβάλλεται το πλήθος των ασαφών IF-THEN κανόνων. Σκοπός είναι να μελετηθεί η επίδραση της διαμέρισης του χώρου εισόδου-σε συνάρτηση με την πολυπλοκότητα που αυτή επιφέρει, στην απόδοση του ταξινομητή. Η διαμέριση του χώρου εισόδου θα γίνει με τη μέθοδο του Subtractive Clustering και τα TSK μοντέλα θα διαφέρουν ως προς την παράμετρο που καθορίζει τον αριθμό των κανόνων, δηλαδή την ακτίνα των cluster. Θα σχεδιάσουμε δύο περιπτώσεις και κάθε μία θα αποτελείται από την εκπαίδευση δύο μοντέλων. Στην πρώτη περίπτωση το subtractive clustering θα εκτελεστεί για όλα τα δεδομένα του συνόλου εκπαίδευσης(class independent), ενώ στη δεύτερη θα εξεταστεί ο διαμερισμός του χώρου εισόδου εφαρμόζοντας clustering

στα δεδομένα του συνόλου εκπαίδευσης που ανήκουν στην εκάστοτε κλάση ξεχωριστά(class dependent). Αυτό γίνεται διότι θέλουμε να αυξηθεί η ερμηνευσιμότητα των μοντέλων και να παράξουμε «καθαρότερα» clusters (άρα και κατά συνέπεια κανόνες).

Σημειώνεται ότι και τα 4 μοντέλα θα εκπαιδευτούν με την υβριδική μέθοδο, σύμφωνα με την οποία οι παράμετροι των συναρτήσεων συμμετοχής βελτιστοποιούνται μέσω της μεθόδου οπισθοδιάδοσης (backpropagation algorithm) , ενώ οι παράμετροι της συνάρτησης εξόδου βελτιστοποιούνται μέσω της μεθόδου των ελάχιστων τετραγώνων (Least Squares).

Καθώς θέλουμε δύο μοντέλα που να διαφέρουν αισθητά ως προς το πλήθος των κανόνων, η ακτίνα που καθορίζει το μέγεθος των clusters πρέπει να λάβει δύο ακραίες τιμές. Για το λόγο αυτό έγινε μία διερεύνηση τρέχοντας έναν αλγόριθμο επαναληπτικά και για ακτίνες από 0.6 μέχρι και 0.1 με βήμα μείωσης 0.0025. Στο τέλος, χρησιμοποιώντας ένα πίνακα που κρατούσε τις τιμές των δεικτών αξιολόγησης συγκρίνουμε τα διάφορα αποτελέσματα και επιλέγουμε δύο μοντέλα που διαφέρουν αισθητά ως προς το πλήθος κανόνων αλλά ταυτόχρονα έχουν ικανοποιητική απόδοση με βάση τους δείκτες που περιγράφονται παρακάτω. Οι επιλογές της ακτίνας(άρα και κανόνων) έχουν παρουσιαστεί πιο κάτω αναλυτικά για τις περιπτώσεις class dependent και class independent.

3) Για την αξιολόγηση της ταξινόμησης των δειγμάτων από τα διάφορα μοντέλα, θα χρησιμοποιηθούν οι εξής δείκτες απόδοσης:

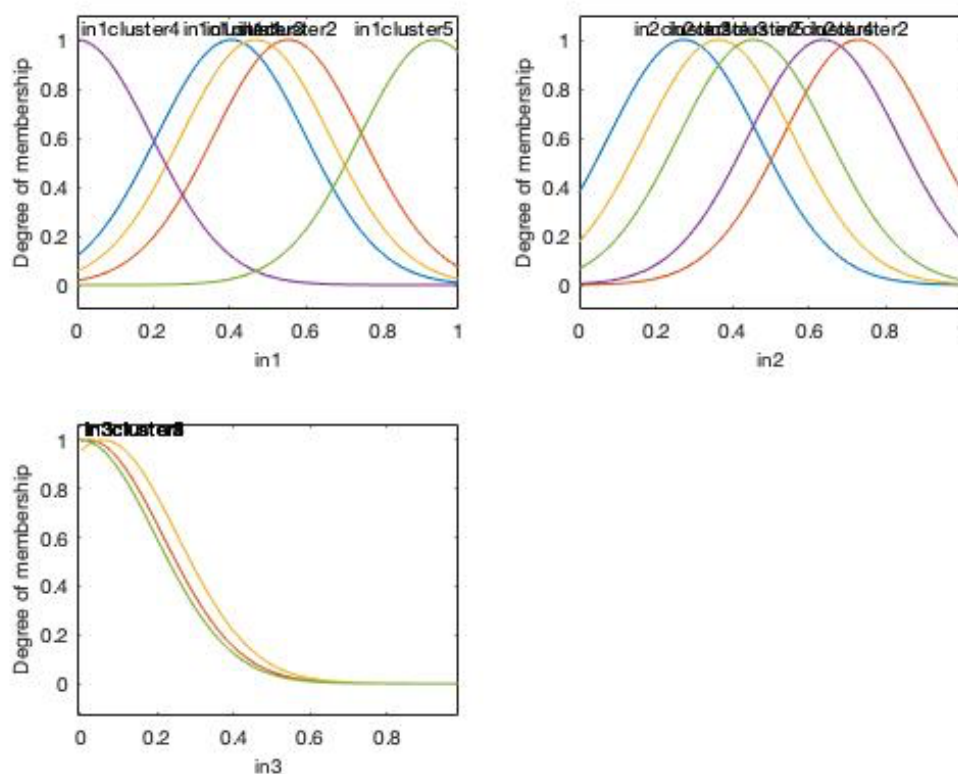
- Error Matrix: Είναι ο πίνακας σφαλμάτων ταξινόμησης και είναι ένας $k \times k$ πίνακας, με k τον αριθμό των κλάσεων ο οποίος βοηθά στην οπτικοποίηση της απόδοσης ενός ταξινομητή και μέσω του οποίου αποκτούμε πρόσβαση σε μια σειρά από δείκτες απόδοσης.
- Overall Accuracy: Είναι η συνολική ακρίβεια του ταξινομητή και ορίζεται ως το ποσοστό των ορθώς ταξινομημένων δειγμάτων ως προς το συνολικό πλήθος των δειγμάτων.
- Producer's Accuracy-User's Accuracy: Οι δείκτες αυτοί αναφέρονται στην απόδοση του ταξινομητή όσον αφορά κάθε κλάση ξεχωριστά.
- \hat{k} : Αποτελεί εκτίμηση της πραγματικής στατιστικής παραμέτρου.

Τα αποτελέσματα λοιπόν που έχουμε από την διαδικασία εκπαίδευσης και αξιολόγησης φαίνονται παρακάτω:

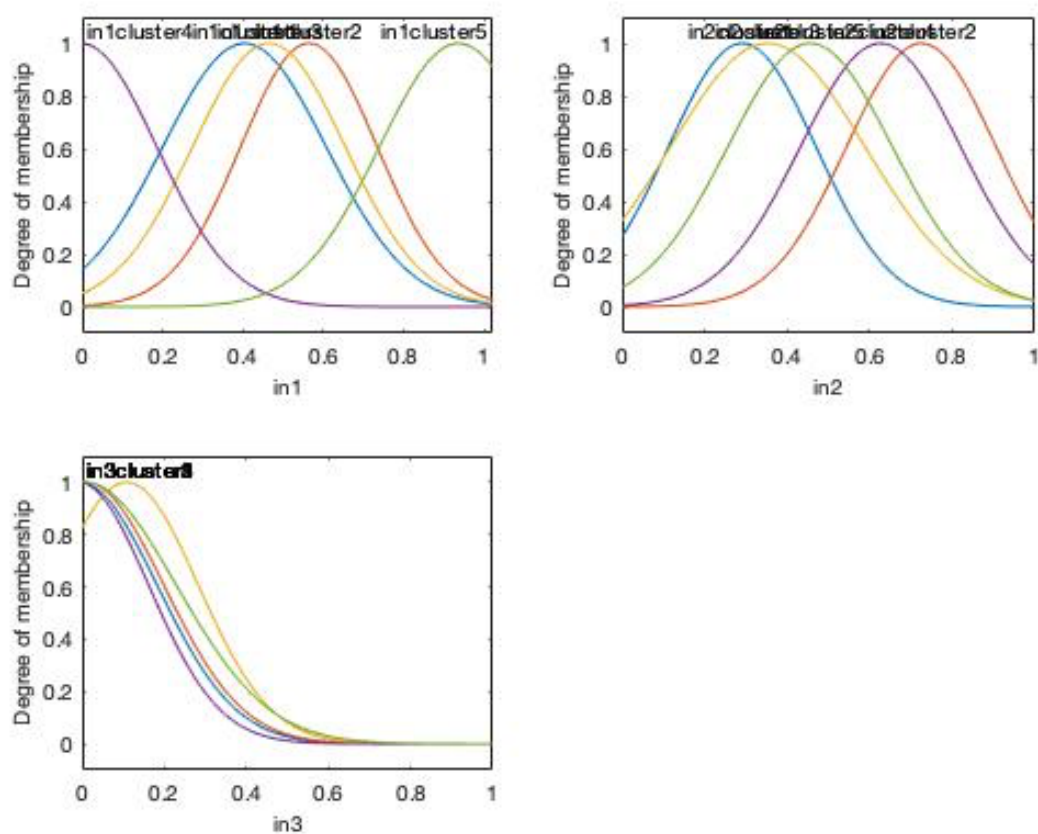
Class Independent:

Για αυτή τη περίπτωση με βάση την διερεύνηση που πραγματοποιήθηκε στη διαδικασία εκπαίδευσης επιλέχθηκε ένα μοντέλο με ακτίνα $r_a=0.5525$ που δημιουργεί 5 κανόνες και ένα μοντέλο με $r_a=0.2425$ που δημιουργεί 22 κανόνες.

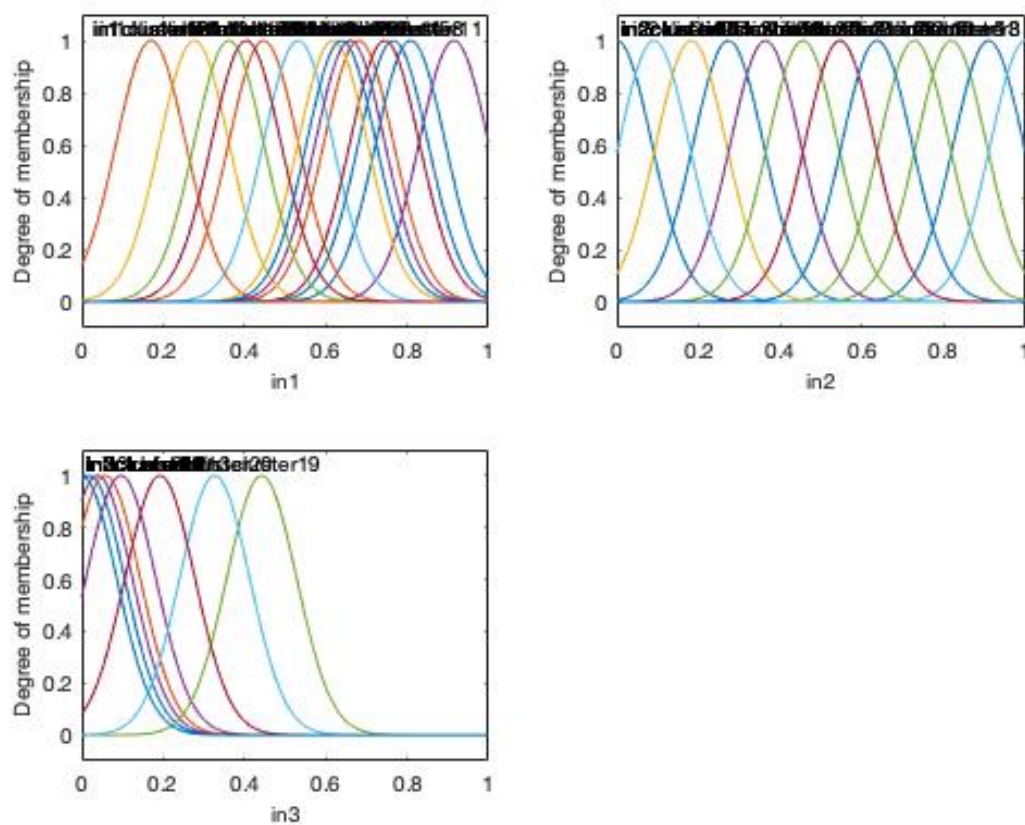
- i. Παρουσιάζονται διαγράμματα στα οποία απεικονίζονται οι τελικές μορφές των συναρτήσεων συμμετοχής για τις εισόδους 1, 2 και 3 των μοντέλων.



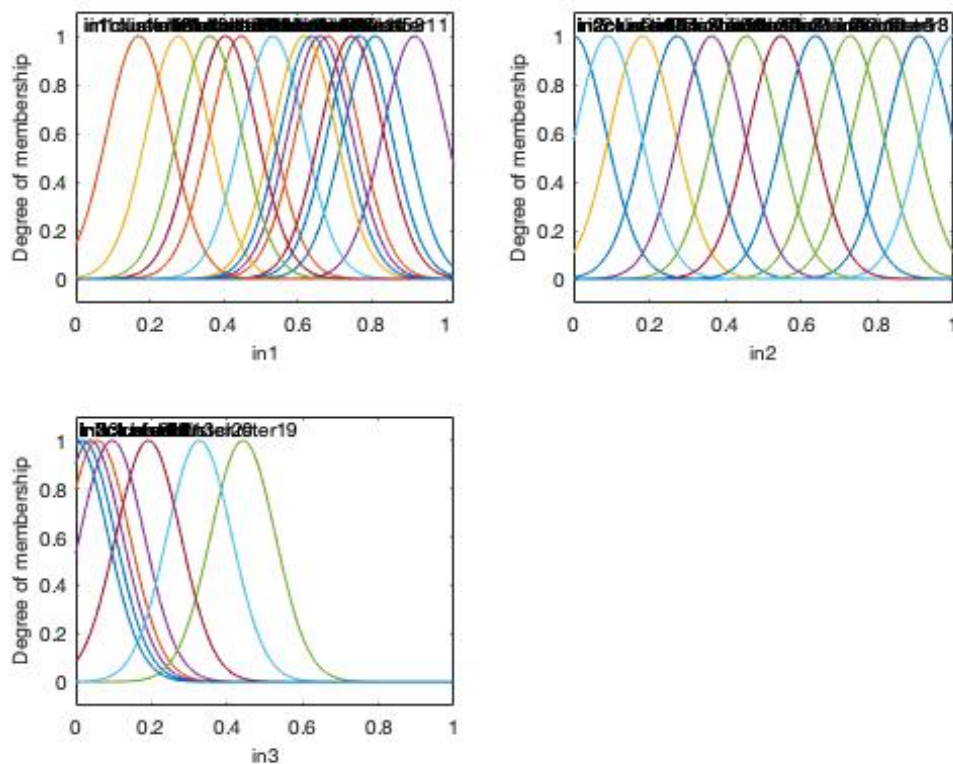
Σχήμα 2: Αρχική Μορφή Συναρτήσεων Συμμετοχής μοντέλου με 5 κανόνες



Σχήμα 3: Τελική Μορφή Συναρτήσεων Συμμετοχής μοντέλου με 5 κανόνες

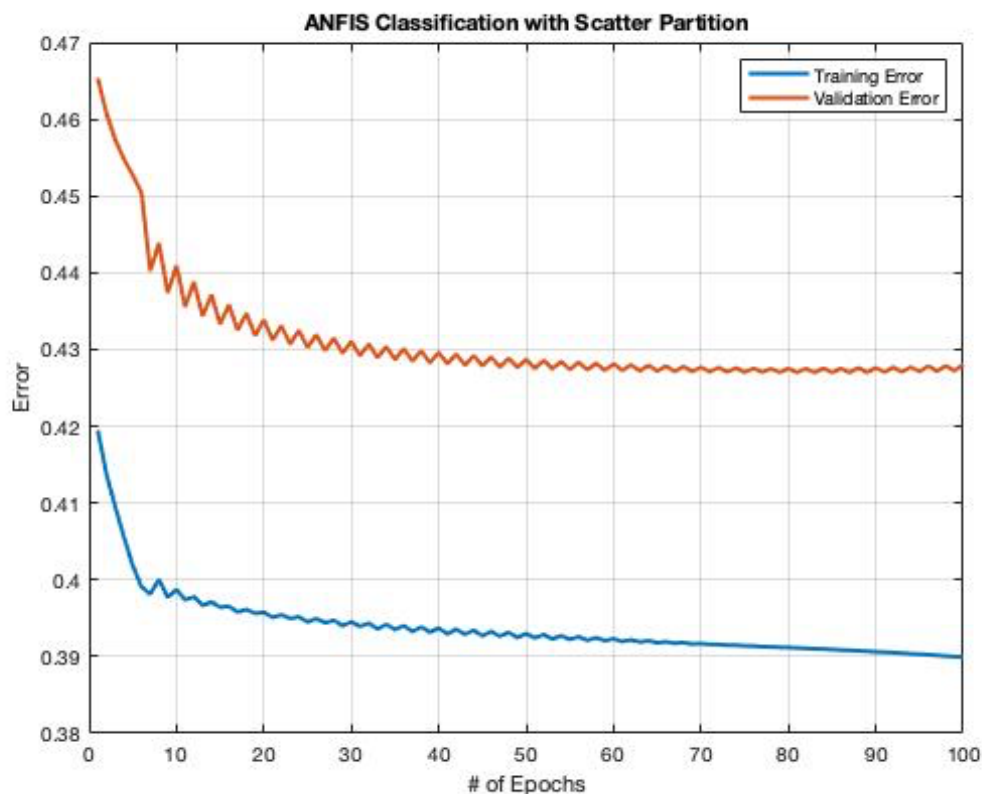


Σχήμα 4: Αρχική Μορφή Συναρτήσεων Συμμετοχής μοντέλου με 22 κανόνες

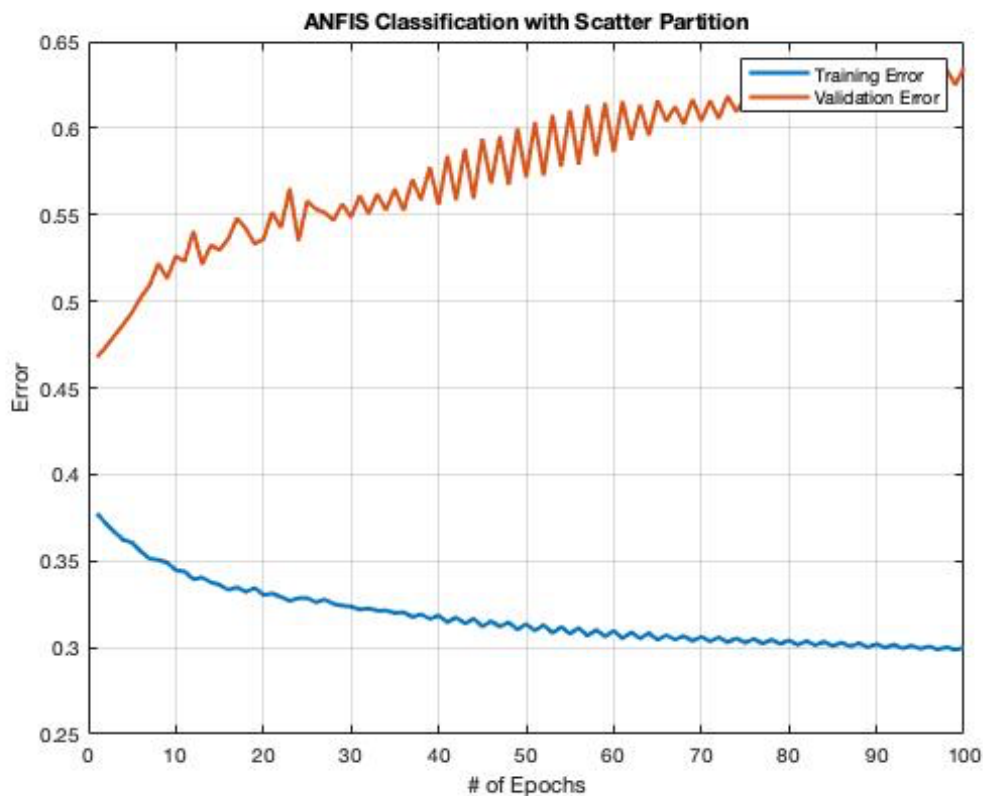


Σχήμα 5: Τελική Μορφή Συναρτήσεων Συμμετοχής μοντέλου με 22 κανόνες

- ii. Συνεχίζοντας φαίνονται τα διαγράμματα μάθησης (learning curves) όπου απεικονίζεται το σφάλμα του μοντέλου συναρτήσει του αριθμού των επαναλήψεων (iterations).



Σχήμα 6: Καμπύλη εκμάθησης μοντέλου με 5 κανόνες



Σχήμα 7: Καμπύλη εκμάθησης μοντέλου με 22 κανόνες

- iii. Παρακάτω δίνεται ο πίνακας σφαλμάτων ταξινόμησης για τα μοντέλα class independent καθώς επίσης και οι διάφορες μετρικές αξιολόγησης:

	AC1	AC2
PC1	43	12
PC2	4	2

Σχήμα 8: Πίνακας Σφάλματος μοντέλου με 5 κανόνες

Δείκτης	Τιμή
OA	0.7377
PA	[0.9148, 0.8571]
UA	[0.7818, 0.6666]
k	0.0722

Σχήμα 9: Δείκτες απόδοσης μοντέλου με 5 κανόνες

	AC1	AC2
PC1	42	13
PC2	5	1

Σχήμα 10: Πίνακας Σφάλματος μοντέλου με 22 κανόνες

Δείκτης	Τιμή
OA	0.7049
PA	[0.8936, 0.9285]
UA	[0.7636, 0.8333]
k	-0.0437

Σχήμα 11: Δείκτες απόδοσης μοντέλου με 22 κανόνες

Παρατηρούμε ότι το μοντέλο με τους 5 κανόνες και ακτίνα $r_a = 0.5525$ έχει καλύτερη απόδοση κρίνοντας από τους πίνακες που περιέχουν τις τιμές των δεικτών αξιολόγησης. Το μοντέλο έχει 0.7377 συνολική απόδοση, και η οποία όταν έχω περισσότερους κανόνες μειώνεται. Δηλαδή, παρατηρούμε ότι όσο μεγαλώνει το πλήθος των κανόνων τόσο μειώνεται η απόδοση του ταξινομητή. Ακόμα, αν και η διαφορά για τα δύο μοντέλα είναι πολύ μικρή (επιβεβαιώθηκε και με ένα μοντέλο με πολύ μεγάλο πλήθος κανόνων), όσο αυξάνεται το πλήθος των κανόνων τόσο περισσότερα δείγματα ταξινομούνται στη κλάση 2. Παρουσιάζεται ένα μοντέλο με 66 κανόνες και $r_a = 0.1650$ για να αποδειχθεί ο παραπάνω ισχυρισμός:

	AC1	AC2
PC1	33	12
PC2	14	2

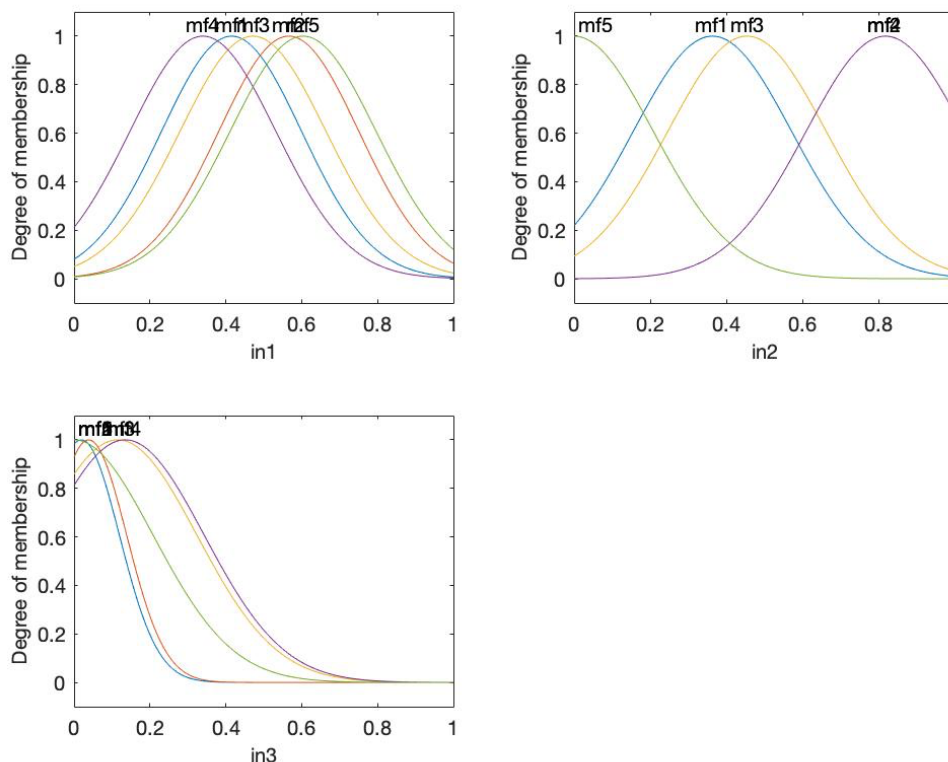
Σχήμα 12: Πίνακας Σφάλματος μοντέλου με 66 κανόνες

Εδώ φαίνεται πιο καθαρά ότι πράγματι με την αύξηση των κανόνων ο ταξινομητής τοποθετεί περισσότερα δείγματα στη κλάση 2. Αυτό το φαινόμενο προκαλεί υψηλή επικάλυψη των συναρτήσεων συμμετοχής και είναι πιθανό να μη καλύπτεται καλά ο χώρος των εισόδων από κάποια τιμή και μετά. Μια μέθοδος που θα βελτίωνε πιθανά τα αποτελέσματα, θα ήταν να ορίσουμε τις δύο κλάσεις με τέτοιο τρόπο ώστε να καλύπτουν καλύτερα το χώρο των εισόδων και των εξόδων, και άρα να πετύχουμε μια πιο ομοιόμορφη κατανομή στη ταξινόμηση.

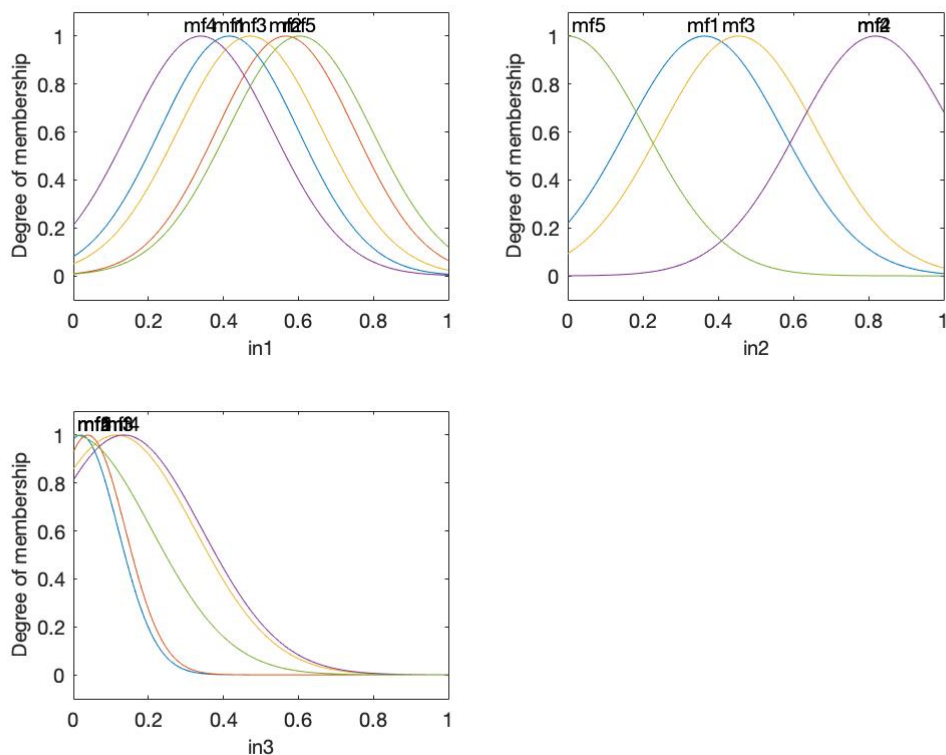
Class Dependent

Για αυτή τη περίπτωση με βάση την διερεύνηση που πραγματοποιήθηκε στη διαδικασία εκπαίδευσης επιλέχθηκε ένα μοντέλο με ακτίνα $r_a=0.5900$ που δημιουργεί 5 κανόνες και ένα μοντέλο με $r_a=0.3025$ που δημιουργεί 25 κανόνες.

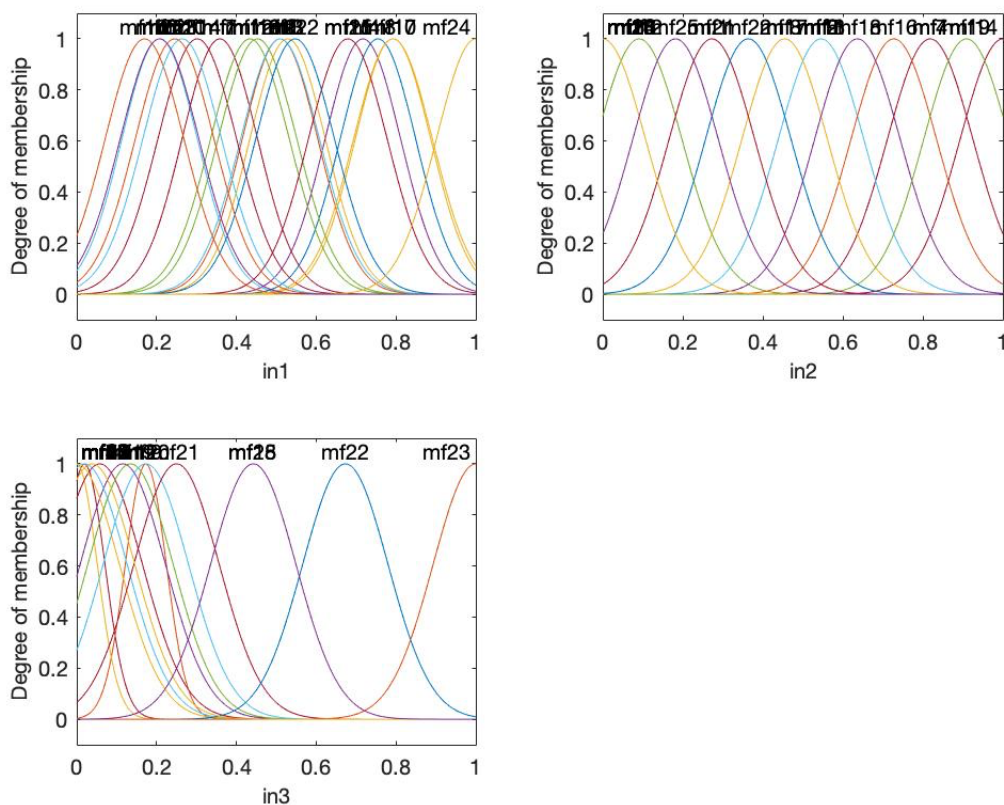
- i. Παρουσιάζονται διαγράμματα στα οποία απεικονίζονται οι τελικές μορφές των συναρτήσεων συμμετοχής για τις εισόδους 1, 2 και 3 των μοντέλων.



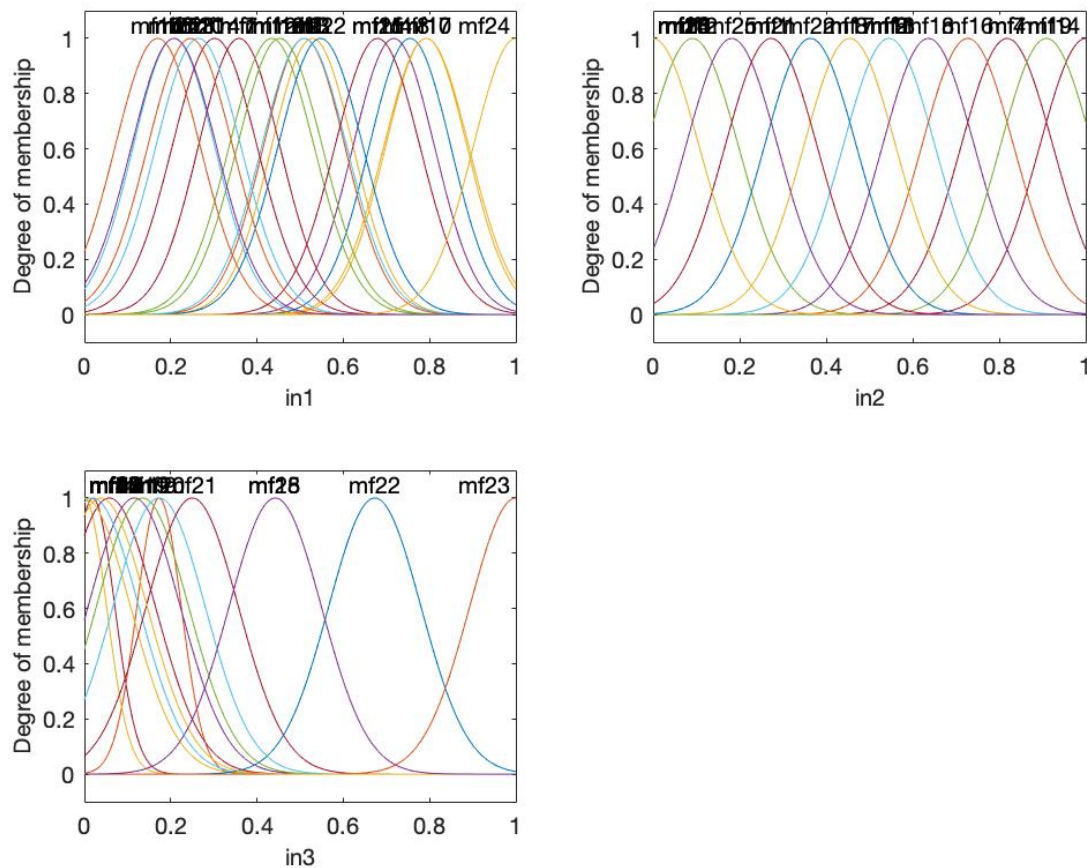
Σχήμα 13: Αρχική Μορφή Συναρτήσεων Συμμετοχής μοντέλου με 5 κανόνες *Class Dependent*



Σχήμα 14: Τελική Μορφή Συναρτήσεων Συμμετοχής μοντέλου με 5 κανόνες *Class Dependent*

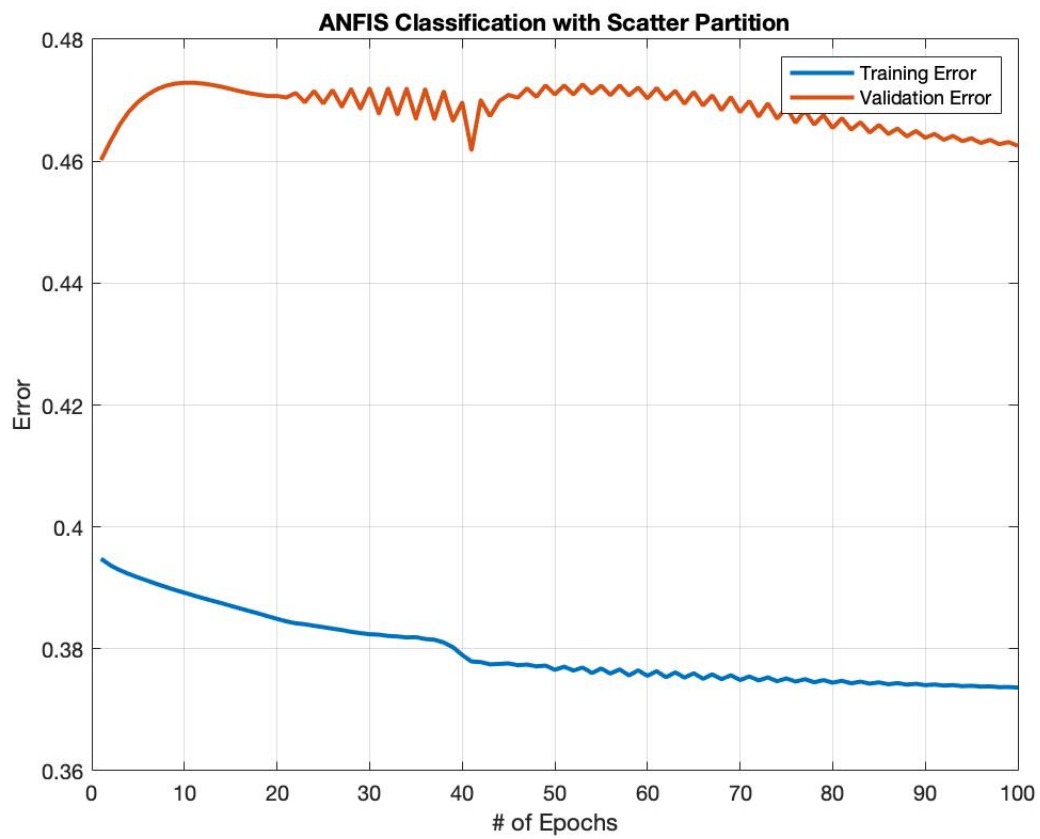


Σχήμα 15: Αρχική Μορφή Συναρτήσεων Συμμετοχής μοντέλου με 25 κανόνες *Class Dependent*

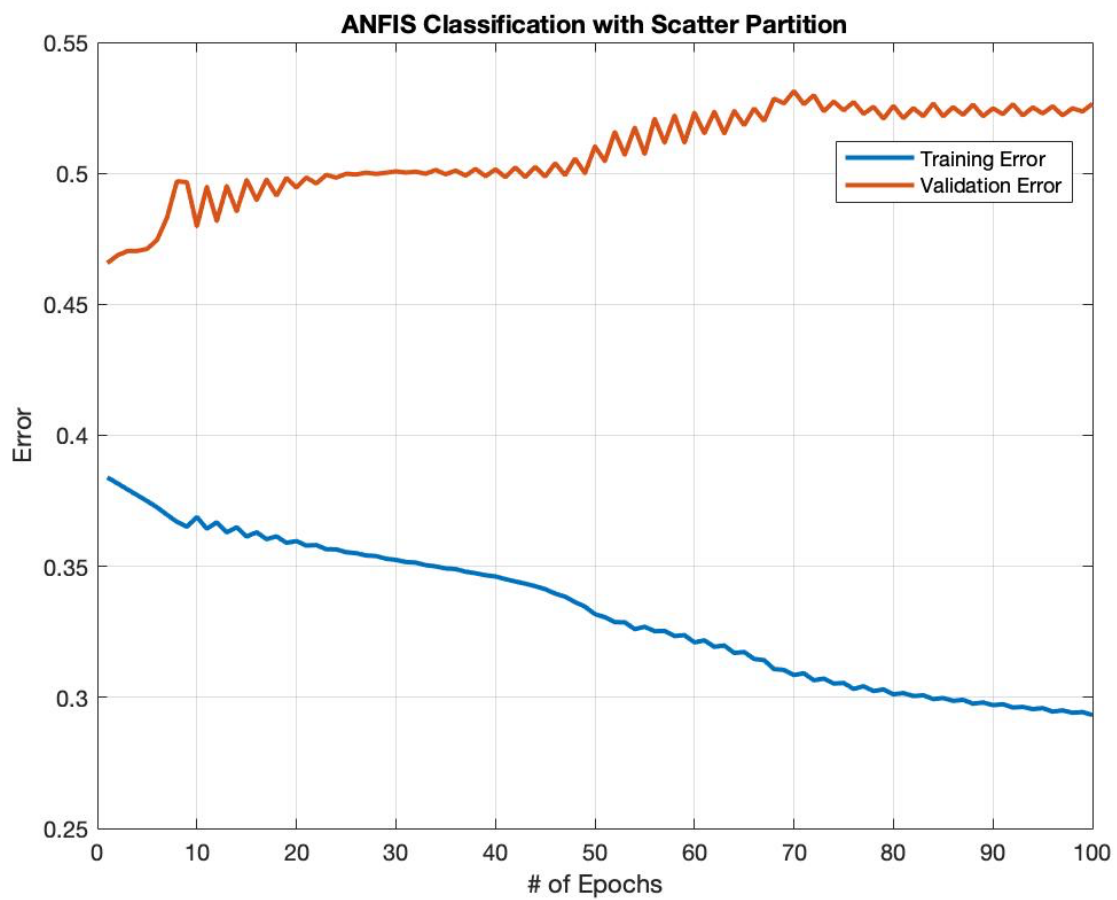


Σχήμα 16: Τελική Μορφή Συναρτήσεων Συμμετοχής μοντέλου με 25 κανόνες *Class Dependent*

- ii. Συνεχίζοντας φαίνονται τα διαγράμματα μάθησης (learning curves) όπου απεικονίζεται το σφάλμα του μοντέλου συναρτήσει του αριθμού των επαναλήψεων (iterations).



Σχήμα 17: Καμπύλη εκμάθησης μοντέλου με 5 κανόνες



Σχήμα 18: Καμπύλη εκμάθησης μοντέλου με 25 κανόνες

- iii. Παρακάτω δίνεται ο πίνακας σφαλμάτων ταξινόμησης για τα μοντέλα class dependent καθώς επίσης και οι διάφορες μετρικές αξιολόγησης:

	AC1	AC2
PC1	42	12
PC2	4	3

Σχήμα 19: Πίνακας Σφάλματος μοντέλου με 5 κανόνες

Δείκτης	Τιμή
OA	0.7377
PA	[0.9130, 0,8]
UA	[0.7777, 0.5714]
k	0.1378

Σχήμα 20: Δείκτες απόδοσης μοντέλου με 5 κανόνες

	AC1	AC2
PC1	43	12
PC2	3	3

Σχήμα 21: Πίνακας Σφάλματος μοντέλου με 25 κανόνες

Δείκτης	Τιμή
OA	0.7541
PA	[0.9347, 0.8]
UA	[0.7818, 0.5]
k	0.1689

Σχήμα 22: Δείκτες απόδοσης μοντέλου με 25 κανόνες

Παρατηρούμε ότι το μοντέλο με τους 25 κανόνες και ακτίνα $r_a = 0.3025$ έχει καλύτερη απόδοση κρίνοντας από τους πίνακες που περιέχουν τις τιμές των δεικτών αξιολόγησης. Το μοντέλο έχει 0.7541 συνολική απόδοση. Παρατηρήθηκε ότι αυξάνοντας το πλήθος των κανόνων και ανάλογα την ακτίνα που τους δημιουργεί η συνολική ακρίβεια σε κάποιες περιπτώσεις είναι λίγο μεγαλύτερη ενώ σε άλλες μπορεί να είναι και μικρότερη. Αυτό που παρατηρούμε ωστόσο είναι ότι οι αρχικές και οι τελικές μορφές των συναρτήσεων συμμετοχής για τις εισόδους δεν διαφέρουν ιδιαίτερα. Δηλαδή, με την class dependent μέθοδο οι συναρτήσεις τοποθετούνται αρκετά καλύτερα και καλύπτουν πιο ομοιόμορφα το χώρο εισόδων. Υπάρχει πάλι το φαινόμενο επικάλυψης ωστόσο σε ένα βαθμό.

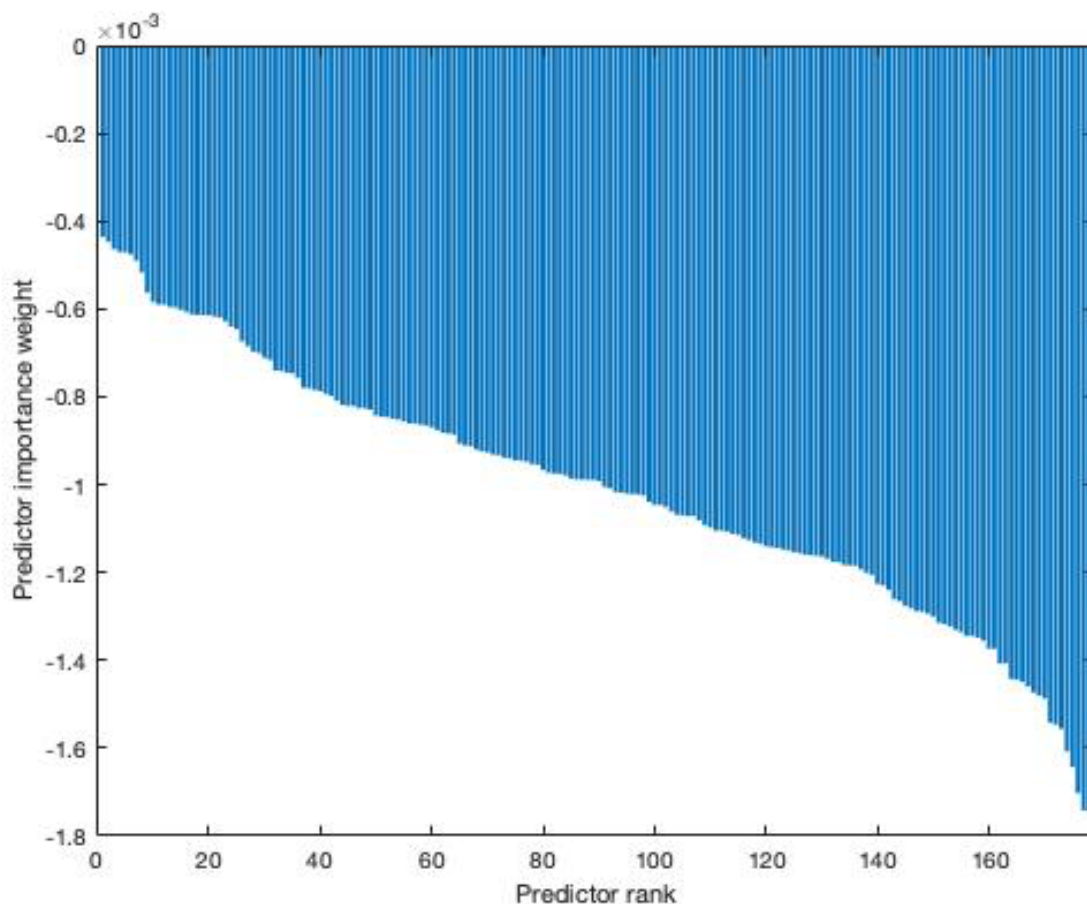
Εφαρμογή σε dataset με υψηλή διαστασιμότητα

Στη δεύτερη φάση της εργασίας θα ακολουθηθεί μια πιο συστηματική προσέγγιση στο πρόβλημα της χρήσης ασαφών νευρωνικών σε προβλήματα ταξινόμησης. Για το σκοπό αυτό χρησιμοποιήθηκε ένα dataset με υψηλότερο βαθμό διαστασιμότητας. Πιο συγκεκριμένα χρησιμοποιήθηκε το Epileptic Seizure Recognition dataset από το UCI repository, το οποίο περιλαμβάνει 11500 δείγματα καθένα από τα οποία περιγράφεται από 179 χαρακτηριστικά. Το χαρακτηριστικό της τελευταίας στήλης βέβαια αποτελεί τον αριθμός της κλάσης στην οποία ανήκει. Είναι φανερό ότι το μέγεθος του συνόλου δεδομένων καθιστά δυσκολότερη μία απλή εφαρμογή ενός TSK μοντέλου, όπως αυτή που πραγματοποιήθηκε στο προηγούμενο μέρος της εργασίας. Είναι αναγκαίο λοιπόν να χρησιμοποιηθούν μέθοδοι μείωσης της διαστασιμότητας και του αριθμού των IF-THEN κανόνων. Αυτό θα το πετύχουμε μέσω της επιλογής χαρακτηριστικών και της χρήσης ασαφούς ομαδοποίησης. Οι δύο αυτές μέθοδοι όμως, παρά την ελάττωση της πολυπλοκότητας που επιφέρουν, εισάγουν στο πρόβλημα δύο ελεύθερες παραμέτρους:

- Αριθμός χαρακτηριστικών: Το πλήθος των χαρακτηριστικών που θα χρησιμοποιηθούν στην εκπαίδευση των μοντέλων.
- Ακτίνα των clusters r_a : Η παράμετρος που καθορίζει την ακτίνα επιρροής των clusters και κατ' επέκταση το πλήθος των κανόνων που θα προκύψουν.

Η αναλυτική, λοιπόν, διαδικασία μοντελοποίησης του προβλήματος ακολουθεί τα παρακάτω βήματα που περιγράφονται :

1. Το σύνολο δεδομένων διαχωρίζεται σε τρία υποσύνολα. Ένα υποσύνολο εκπαίδευσης που περιέχει το 60% των δεδομένων, ένα υποσύνολο επικύρωσης που περιέχει το 20% των δεδομένων και ένα σύνολο ελέγχου που περιέχει το 20% των δεδομένων επίσης. Εφαρμόζουμε κανονικοποίηση στα δεδομένα εισόδου ώστε να βρίσκονται στο διάστημα $[0,1]$, χρησιμοποιώντας κατάλληλη συνάρτηση στο MATLAB.
2. Προκειμένου να διευκολύνουμε την επιλογή των χαρακτηριστικών εφαρμόζουμε τον αλγόριθμο Relieff. Πιο συγκεκριμένα με τον αλγόριθμο αυτό ταξινομούμε τα χαρακτηριστικά σε φθίνουσα σειρά σημαντικότητας. Δηλαδή, το πρώτο είναι το πιο σημαντικό και το τελευταίο το λιγότερο σημαντικό. Αυτό το βήμα είναι μεγάλης σημασίας διότι καθώς θέλουμε να μειώσουμε τη διαστασιμότητα του συνόλου δεδομένων και να το προσεγγίσουμε με ένα ισοδύναμο μικρότερο, πρέπει να επιλέξουμε τα πιο σημαντικά χαρακτηριστικά που συνεισφέρουν το μεγαλύτερο βαθμό στο καθορισμό της εξόδου. Στο σχήμα 23 βλέπουμε ένα ιστόγραμμα που παρουσιάζει το αποτέλεσμα του αλγορίθμου Relieff.



Σχήμα 23: Αποτέλεσμα αλγορίθμου Relieff για ταξινόμηση των χαρακτηριστικών ως προς τη σημαντικότητά τους.

3. Στη συνέχεια προκειμένου να βρούμε τις βέλτιστες τιμές για τις δύο ελεύθερες παραμέτρους του προβλήματος μας εκτελούμε αναζήτηση πλέγματος(grid search) και αξιολόγηση μέσω 5-πτυχης διασταυρωμένης επικύρωσης (5-fold cross validation). Σημειώνεται ότι ως μέθοδος ομαδοποίησης για τη δημιουργία των IF-THEN κανόνων χρησιμοποιήθηκε ο αλγόριθμος Subtractive Clustering, και στην αναζήτηση πλέγματος με διασταυρωμένη επικύρωση τα δεδομένα εκπαίδευσης διαχωρίζονταν σε 80% για εκπαίδευση και 20% για επικύρωση. Τέλος, μετά από διεξαγωγή του grid search για διάφορες τιμές χαρακτηριστικών και τιμών ακτίνας παρουσιάζονται τα αποτελέσματα της αναζήτησης πλέγματος για $NF=[3, 8, 12, 16, 20]$ και $ra=[0.6, 0.55, 0.5, 0.45, 0.4]$. Το NF είναι τα δυνατά πλήθη χαρακτηριστικών και ra είναι οι δυνατές τιμές της ακτίνας των clusters. Οι τιμές αυτές επιλέχθηκαν διότι για μεγάλα πλήθη χαρακτηριστικών και πολύ μικρές τιμές ακτίνας, παρατηρήθηκαν τα εξής:

A) Απαγορευτικός χρόνος υπολογισμού του αποτελέσματος, που καθιστά μη πρακτική τη διαδικασία.

B) Για πολύ μικρές τιμές της ακτίνας παρατηρήθηκε ότι η συνάρτηση anfis επέστρεφε πολλές τιμές σφαλμάτων NaN. Αυτό οδηγούσε όπως γίνεται αντιληπτό σε αδυναμία υπολογισμού της εξόδου μετέπειτα.

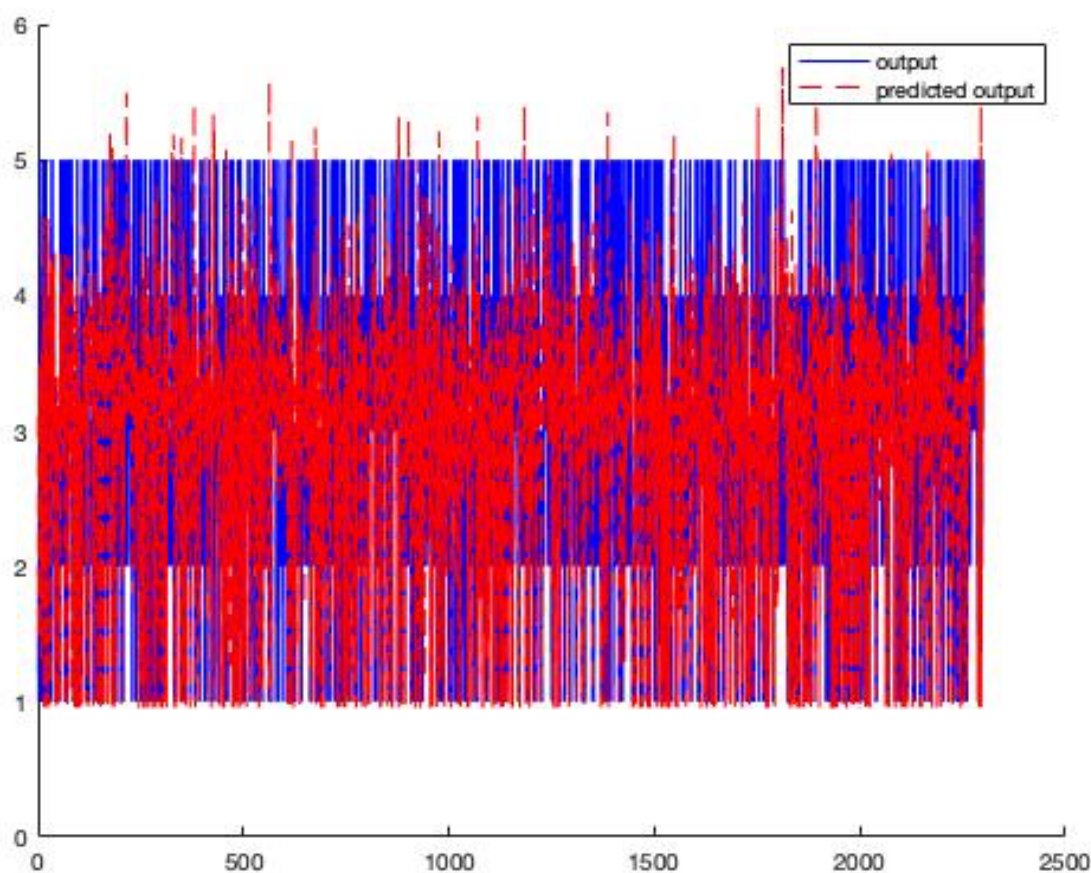
Στο σχήμα 24 παρουσιάζεται ο πίνακας που περιέχει τις τιμές του μέσου σφάλματος επικύρωσης για όλους τους συνδυασμούς χαρακτηριστικών και ακτινών. Έχει σημειωθεί ο συνδυασμός που δίνει το χαμηλότερο μέσο σφάλμα επικύρωσης και είναι για 20 χαρακτηριστικά και ακτίνα cluster $ra=0.5$. Η τιμή του χαμηλότερου αυτού σφάλματος είναι 1.0049. Οι κανόνες που δημιουργούνται είναι 12.

NF\ra	0.6	0.55	0.5	0.45	0.4
4	1.1108	1.1103	1.1091	1.1094	1.1091
8	1.0578	1.0490	1.0471	1.0484	1.0499
12	1.0505	1.0516	1.0498	1.0427	1.0408
16	1.0301	1.0276	1.0253	1.0232	1.0198
20	1.0203	1.0182	1.0049	1.0079	1.0131

Σχήμα 24: Πίνακας με τα αποτελέσματα της Αναζήτησης Πλέγματος

4. Χρησιμοποιώντας τις βέλτιστες τιμές των ελεύθερων παραμέτρων, που βρήκαμε στο προηγούμενο βήμα θα εκπαιδεύσουμε με τις ίδιες προδιαγραφές όπως και προηγουμένως(SC) το τελικό TSK μοντέλο για το πρόβλημα μας. Σε αυτό το σημείο τονίζουμε ότι σαν σύνολο εκπαίδευσης χρησιμοποιούμε όλο το σετ εκπαίδευσης και αντίστοιχα σαν σύνολο επικύρωσης, όλο το σετ δειγμάτων επικύρωσης. Παρακάτω έχουν προστεθεί τα διάφορα διαγράμματα τα οποία ζητούνται σαν αποτελέσματα:

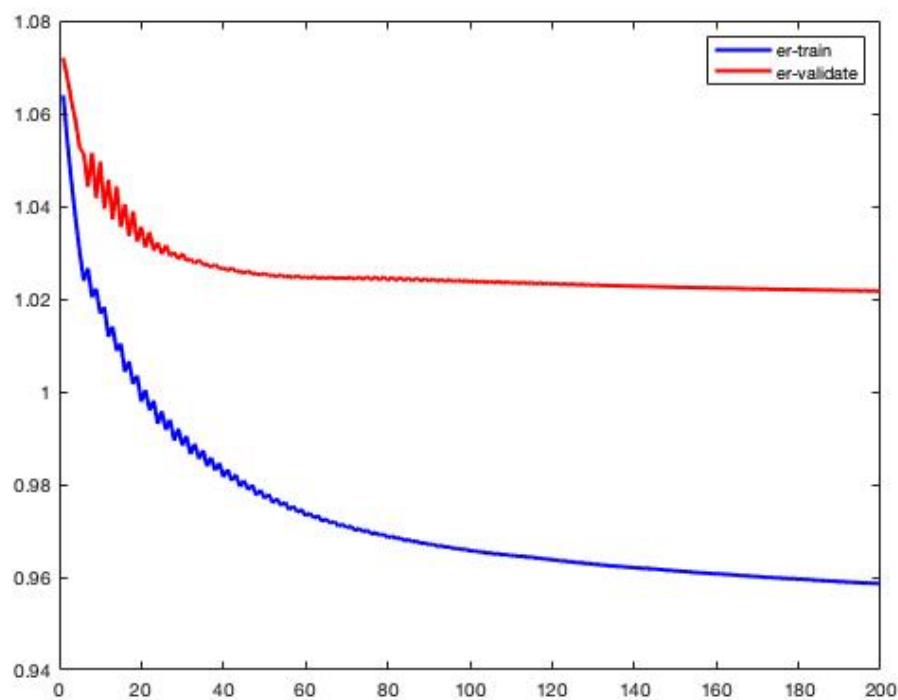
- I. Διαγράμματα που αποτυπώνονται οι προβλέψεις του τελικού μοντέλου καθώς και οι πραγματικές τιμές της εξόδου. Για να σχεδιαστεί αυτό το διάγραμμα αντιπαραβάλλουμε το τελικό μας μοντέλο με το υποσύνολο ελέγχου.



Σχήμα 25: Πρόβλεψη μοντέλου σε σχέση με τη πραγματική έξοδο

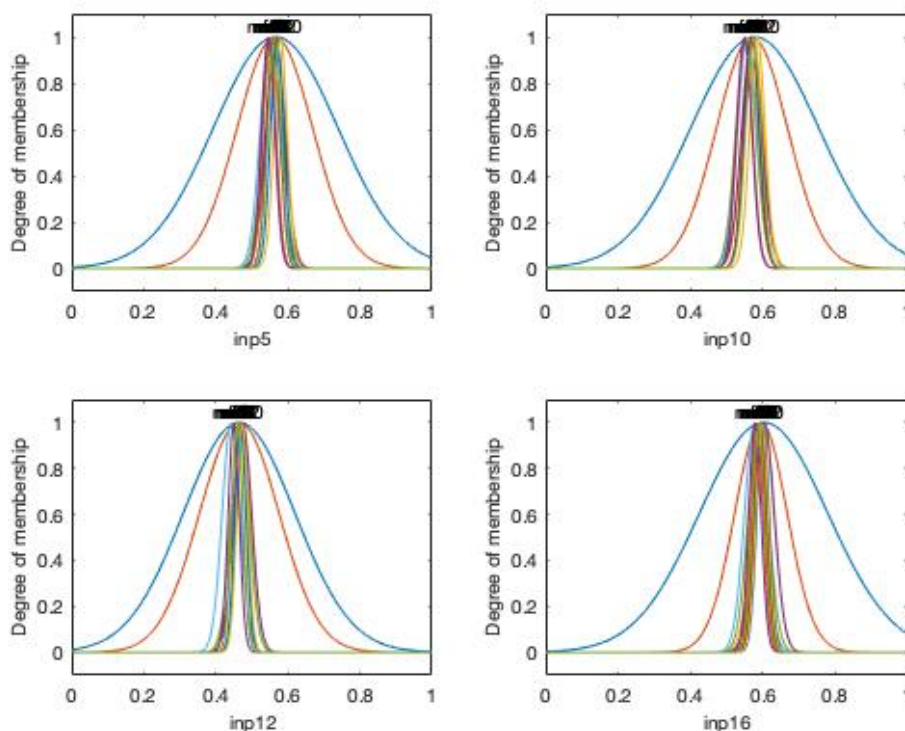
Παρατηρούμε ότι υπάρχει αρκετή απόκλιση ανάμεσα στις προβλέψεις του μοντέλου για την έξοδο και τη πραγματική τιμή της.

- II. Διάγραμμα εκμάθησης όπου απεικονίζεται το σφάλμα συναρτήσεως του αριθμού επαναλήψεων:

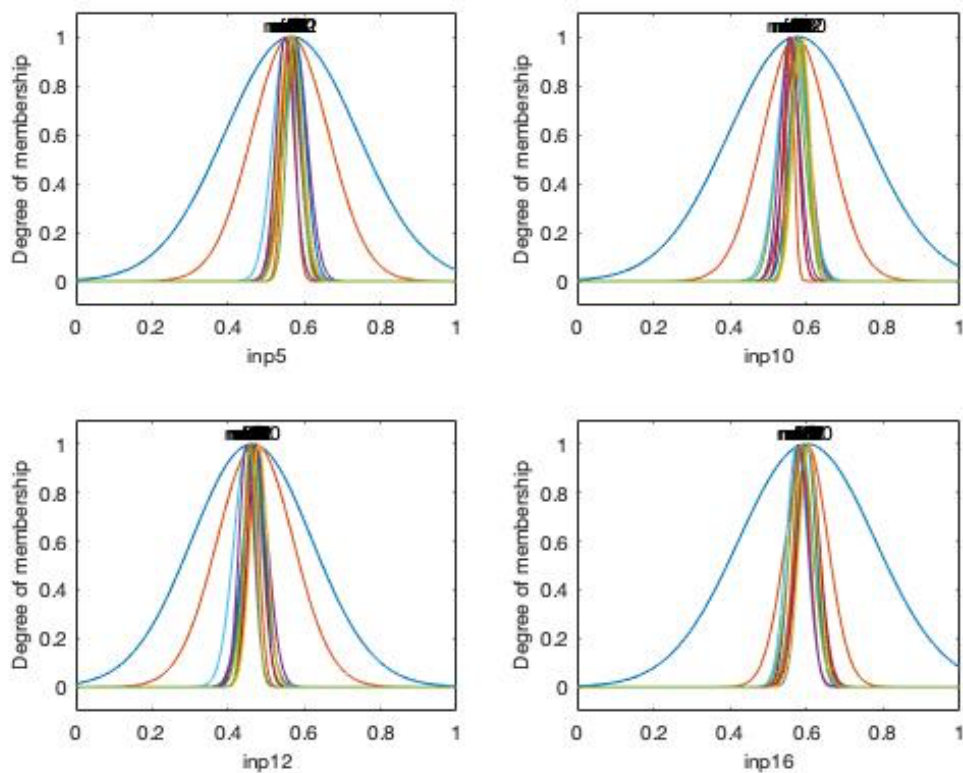


Σχήμα 26: Καμπύλη εκμάθησης τελικού TSK μοντέλου

- III. Εδώ έχω επισυνάψει τις αρχικές και τελικές μορφές(μετά την εκπαίδευση) των συναρτήσεων συμμετοχής για τις μεταβλητές εισόδου 5,10,12 και 16.



Σχήμα 27: Συναρτήσεις συμμετοχής πριν το training



Σχήμα 28: Συναρτήσεις συμμετοχής μετά το training

Παρατηρούμε ότι οι συναρτήσεις συμμετοχής μετά την διαδικασία εκπαίδευσης έχουν μεταβληθεί ελάχιστα σε σχέση με την αρχική τους κατάσταση. Ακόμα, παρατηρούμε ότι είναι συγκεντρωμένες όλες προς το κέντρο και επικαλύπτονται έντονα.

IV. Παρακάτω δίνεται ο πίνακας σφαλμάτων ταξινόμησης για το τελικό μοντέλο καθώς επίσης και οι διάφορες μετρικές αξιολόγησης:

	AC1	AC2	AC3	AC4	AC5
PC1	341	15	2	1	0
PC2	102	23	24	30	4
PC3	45	311	330	213	141
PC4	4	86	127	166	236
PC5	1	4	5	42	47

Σχήμα 29: Πίνακας Σφαλμάτων Ταξινόμησης Βέλτιστου Μοντέλου

Μετρική	Τιμή
OA	0.3943
PA	[0.6917, 0.0524, 0.6762, 0.3673, 0.1098]
UA	[0.9499, 0.1257, 0.3173, 0.2682, 0.4747]
k	0.2377

Σχήμα 30: Δείκτες απόδοσης μοντέλου

Από τους παραπάνω πίνακες μπορούμε να αποφανθούμε ότι το μοντέλο μας έχει πολύ μικρή ακρίβεια, ποιο συγκεκριμένα μόλις 39.43%! Εάν δούμε βέβαια τον αντίστοιχο πίνακα για τους δείκτες απόδοσης των μοντέλων του πρώτου μέρους των εργασιών, δημιουργείται η αίσθηση ότι ίσως δεν είναι τόσο ακριβές αυτό το μοντέλο σε σχέση με τα προηγούμενα. Ωστόσο, δεν πρέπει να ξεχνάμε ότι το τελικό μας μοντέλο σε αυτό το μέρος της εργασίας εκπαιδεύτηκε με μόλις 20 χαρακτηριστικά(11.23%) από τα 178 που περιγράφουν κάθε δείγμα. Αυτό σε συνδυασμό με το ότι η ακτίνα των cluster $ra = 0.5$ μας δημιουργεί μόνο 12 ασαφείς κανόνες, μειώνει την ακρίβεια του μοντέλου μας σε ένα βαθμό καθώς η έξοδος εξαρτάται από πολύ μεγαλύτερο πλήθος χαρακτηριστικών.

Εάν συγκρίνουμε τον αντίστοιχο αριθμό κανόνων που θα είχαμε αν για το ίδιο πλήθος χαρακτηριστικών είχαμε επιλέξει grid partitioning με δύο ή τρία ασαφή σύνολα ανά είσοδο έχουμε:

Με δύο συναρτήσεις συμμετοχής ανά είσοδο θα είχαμε 1048576 κανόνες για τα ίδια χαρακτηριστικά ενώ για τρεις συναρτήσεις συμμετοχής ανά είσοδο το MATLAB δεν έχει τη απαραίτητη μνήμη για το μέγεθος του μοντέλου! Είναι κατανοητό ότι για τέτοιο πλήθος κανόνων, ο χρόνος της διαδικασίας εκπαίδευσης είναι απαγορευτικός. Εκτός αυτού είναι πολύ πιθανό ότι για τέτοιο πλήθος κανόνων θα είχαμε φαινόμενα υπερεκπαίδευσης με αποτέλεσμα το μοντέλο να χάσει τη δυνατότητα γενίκευσης. Δηλαδή, το μοντέλο χάνει την ικανότητα της επιτυχούς πρόβλεψης της εξόδου για άγνωστες εισόδους.

Τέλος, όπως φαίνεται από το πίνακα με τα σφάλματα ταξινόμησης αλλά και τα διαγράμματα για τις συναρτήσεις συμμετοχής, οι κανόνες είναι συγκεντρωμένοι γύρω από τις κλάσεις 3 και 4 και όσο αυξάνεται το πλήθος των κανόνων θα ταξινομούνται όλο και περισσότερα δείγματα στις κλάσεις αυτές με αποτέλεσμα να έχουν υψηλή επικάλυψη οι κλάσεις αυτές και οι υπόλοιπες κλάσεις δεν καλύπτονται αρκετά. Αυτό σημαίνει ότι δεν υπάρχουν αρκετοί κανόνες να τις ενεργοποιούν .