

Ασαφή Συστήματα

Εργασία #4 ~ Classification

Δημήτρης Παππάς

ΑΕΜ: 8391

e-mail: dspappas@ece.auth.gr



Περιεχόμενα

4. Εργασία #4: Classification	3
4.1 Part 1 - Εφαρμογή σε απλό dataset.....	3
4.1.1 TSK model 1.....	5
4.1.2 TSK model 2.....	8
4.1.3 TSK model 3.....	11
4.1.4 TSK model 4.....	14
4.1.5 Αξιολόγηση Μοντέλων και Πίνακες Σφαλμάτων.....	17
4.1.6 Συμπεράσματα και Βέλτιστο Μοντέλο	19
4.2 Part 2 - Εφαρμογή σε dataset με υψηλή διαστασιμότητα.....	20
4.2.1 Αποτελέσματα του Grid Search	21
4.2.2 Διαγράμματα του Grid Search	21
4.3.3 Βέλτιστο Μοντέλο	24

4. Εργασία #4: Classification

4.1 Part 1 - Εφαρμογή σε απλό dataset

Σε αυτή την εργασία μελετάμε την ικανότητα των TSK μοντέλων (Takagi Sugeno Kang) στην ταξινόμηση και ομαδοποίηση δεδομένων σε classes, καθώς και στην αξιολόγηση των μοντέλων με διάφορες μετρικές.

Δημιουργούμε 4 μοντέλα:

Model	Ακτίνα των Clusters	Class
TSK model 1	0.1	Dependent
TSK model 2	0.9	Dependent
TSK model 3	0.1	Independent
TSK model 4	0.9	Independent

Το dataset που θα χρησιμοποιήσουμε είναι το “Haberman’s Survival ” από το UCI repository, το οποίο περιέχει 306 δείγματα και αποτελείται από 3 features και 1 output.

Τα μοντέλα υλοποιούνται στο matlab script “classification_Haberman.m”.

Χωρίζουμε το dataset σε τρία υποσύνολα, 60% training data, 20% validation data, 20% checking data.

Εκπαιδεύουμε τέσσερα TSK μοντέλα.

- Στα 1 και 2 ο διαχωρισμός του χώρου του input dataset γίνεται με subtractive clustering στα δεδομένα του συνόλου εκπαίδευσης, που ανήκουν στην εκάστοτε κλάση ξεχωριστά (**Class Dependent**).
- Στα 3 και 4 ο διαχωρισμός του χώρου του input dataset γίνεται με subtractive clustering σε όλα τα δεδομένα του συνολικού training dataset (**Class Independent**) και η εκπαίδευση υλοποιείται με genfis2 και anfis training.

Για κάθε Dependent και Independent μοντέλο, έχουμε δύο μοντέλα που υλοποιούνται με διαφορετικό αριθμό κανόνων. Η τιμή του αριθμού των κανόνων καθορίζεται από την τιμή της ακτίνας των clusters. Επιλέγουμε δύο ακραίες τιμές [0.1 0.9], ώστε ο αριθμός των κανόνων να αλλάζει σε μεγάλο βαθμό.

Για την εκπαίδευση του μοντέλου χρησιμοποιούμε τη συνάρτηση anfis και επιλέγουμε EpochNumber = 100.

Επίσης, δημιουργούμε διαγράμματα για τις τελικές εκβάσεις των μοντέλων και των Συναστίσεων Συμμετοχής, για την καμπύλη εκμάθησης και για το σφάλμα πρόβλεψης.

Οι εισόδοι των μοντέλων είναι:

- age (numerical)
- year – 1900 (numerical)
- nodes (numerical)

ενώ η έξοδος είναι :

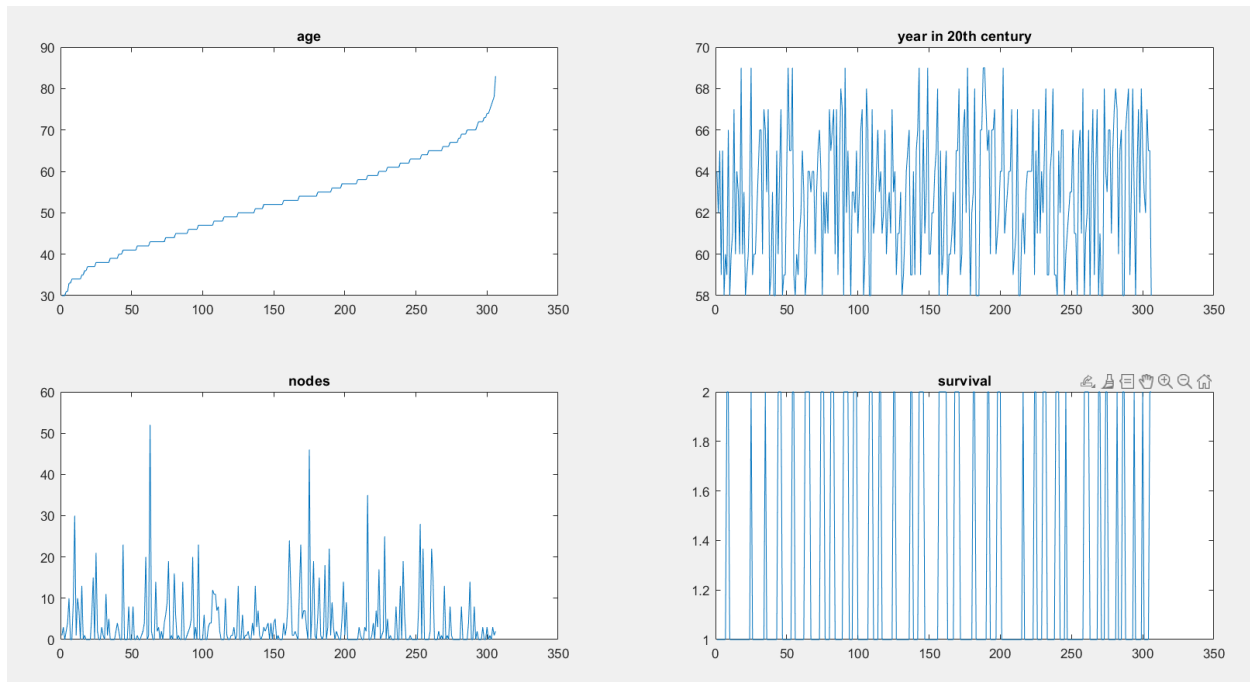
- survival (class attribute)

Η έξοδος αποτελείται από δύο κλάσεις {1,2}.

1 = the patient survived 5 years or longer

2 = the patient died within 5 year

Παρακάτω φαίνονται οι τιμές των εισόδων και της εξόδου.

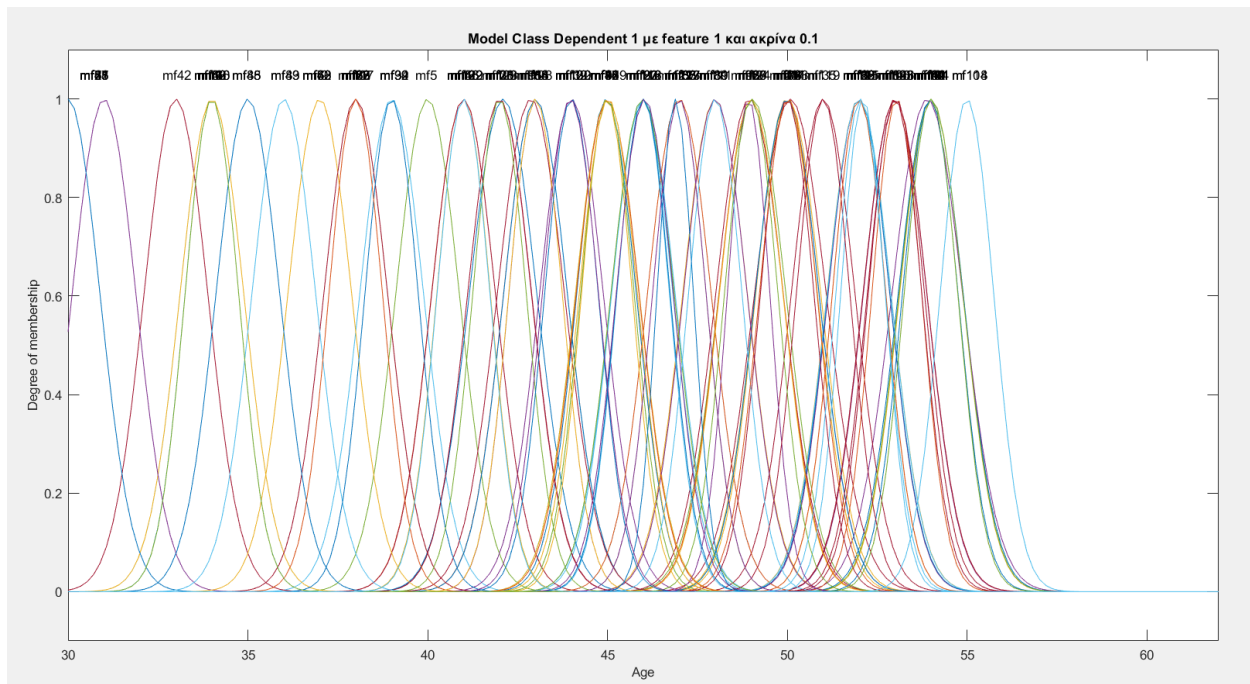


Σχήμα 95: Input & Output Values

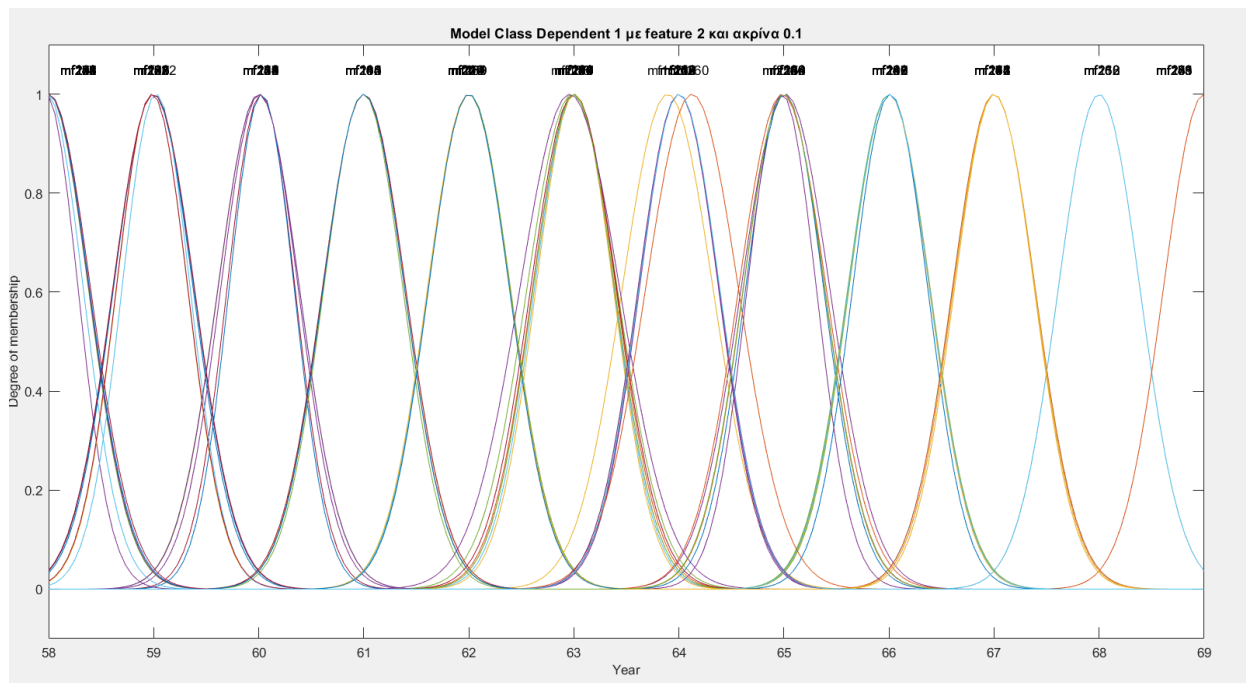
4.1.1 TSK model 1

TSK μοντέλο με ακτίνα 0.1 και class dependent.

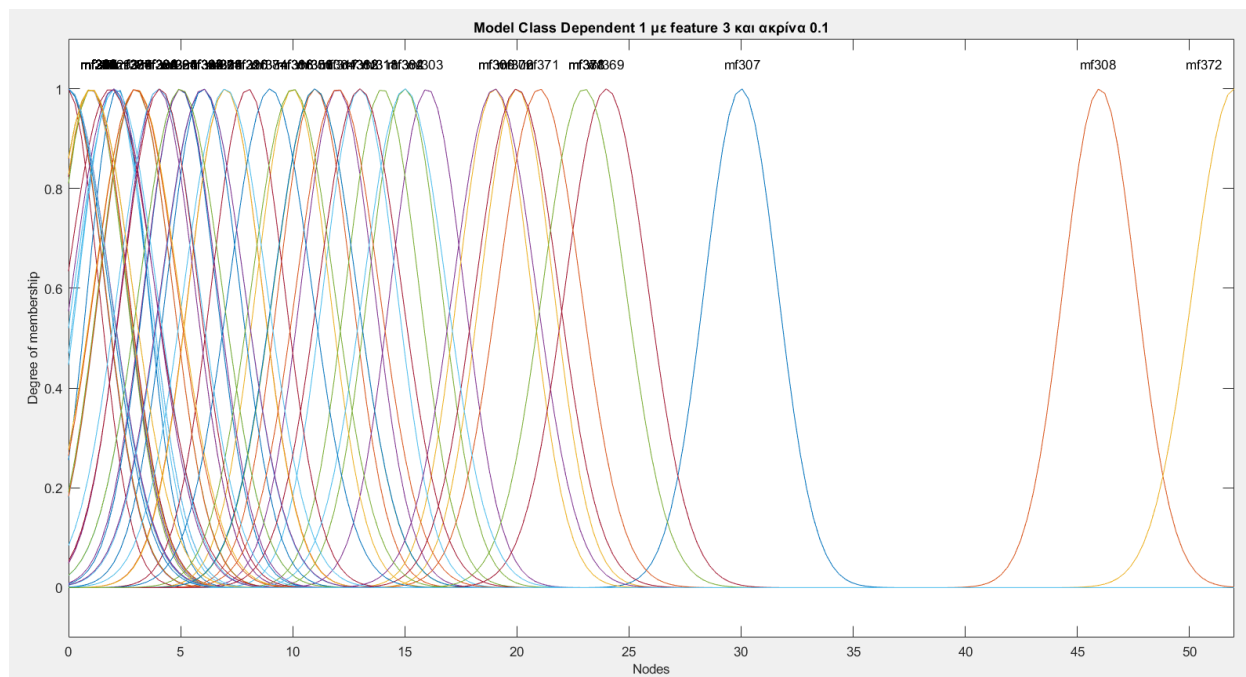
- Συναρτήσεις Συμμετοχής Μετά την Εκπαίδευση



Σχήμα 96: mf After Training – Age – class dependent - radius = 0.1

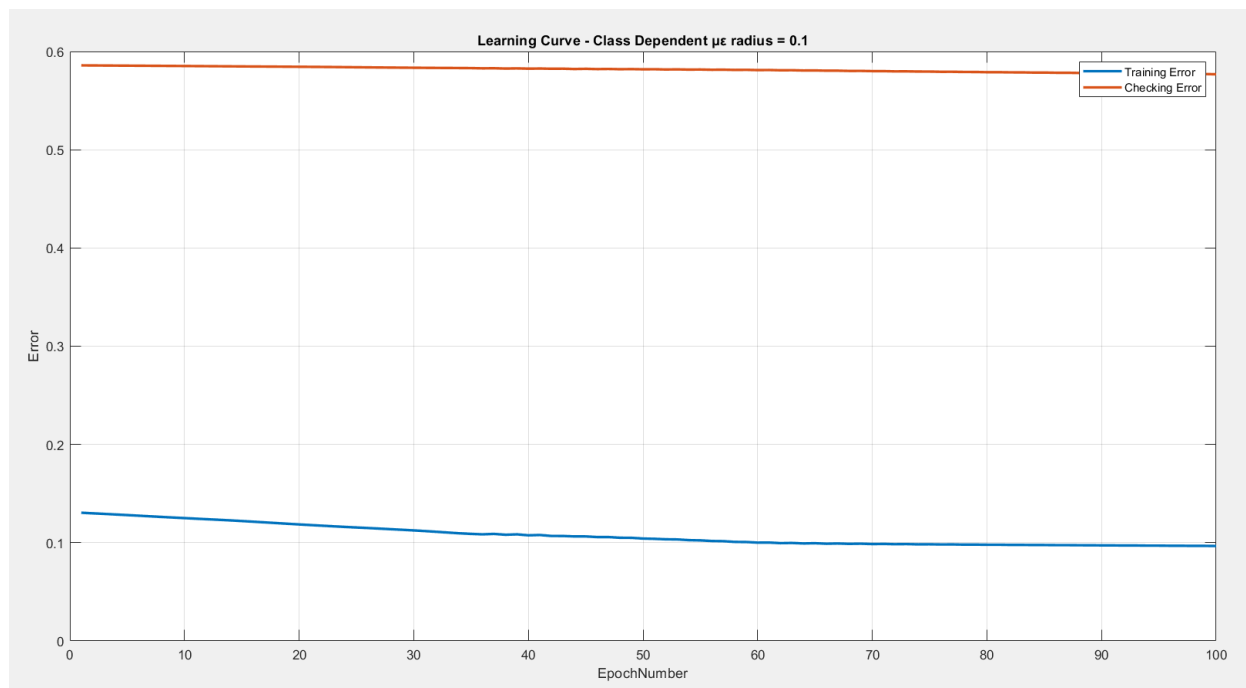


Σχήμα 97: mf After Training – Year – class dependent - radius = 0.1



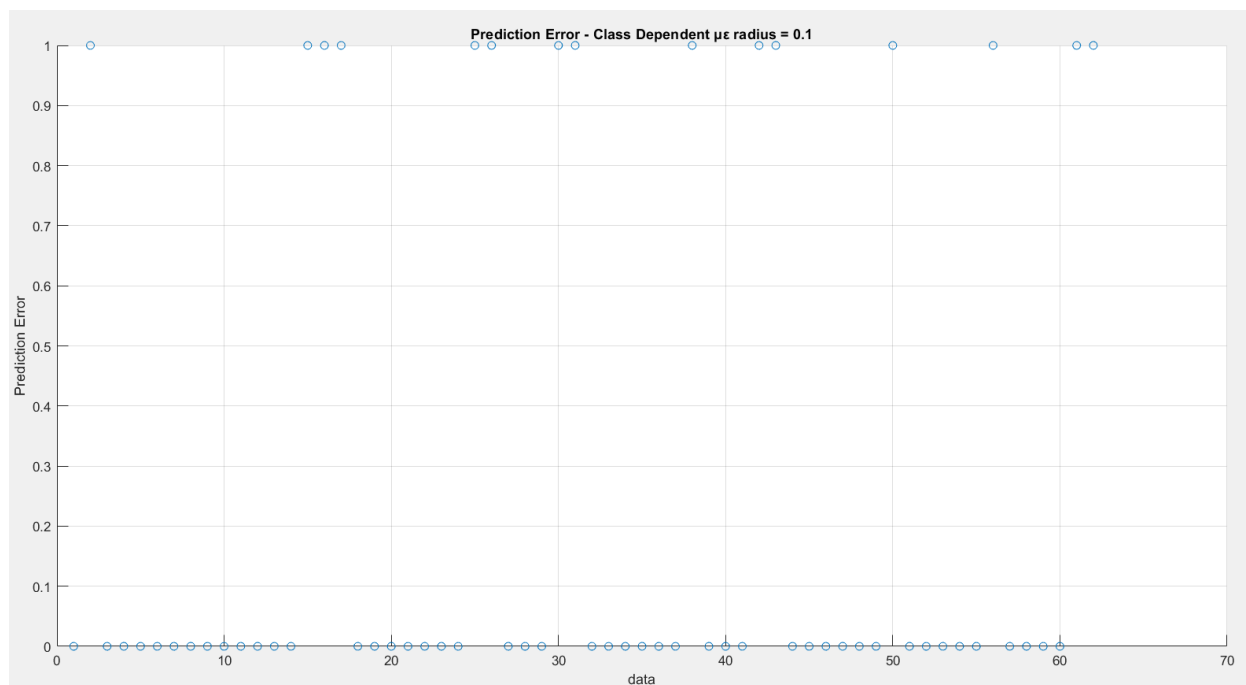
Σχήμα 98: mf After Training – Nodes – class dependent - radius = 0.1

- Learning Curve



Σχήμα 99: Learning Curve - Class Dependent - radius = 0.1

- Prediction Error



Σχήμα 100: Prediction Error - Class Dependent - radius = 0.1

Αξιολόγηση Μοντέλου 1

Μετρικές	Μοντέλο 1
OA	0.7581
PA 1	1.0000
PA 2	0.0625
UA 1	0.7541
UA 2	1.0000
K	0.0900
Rules	132

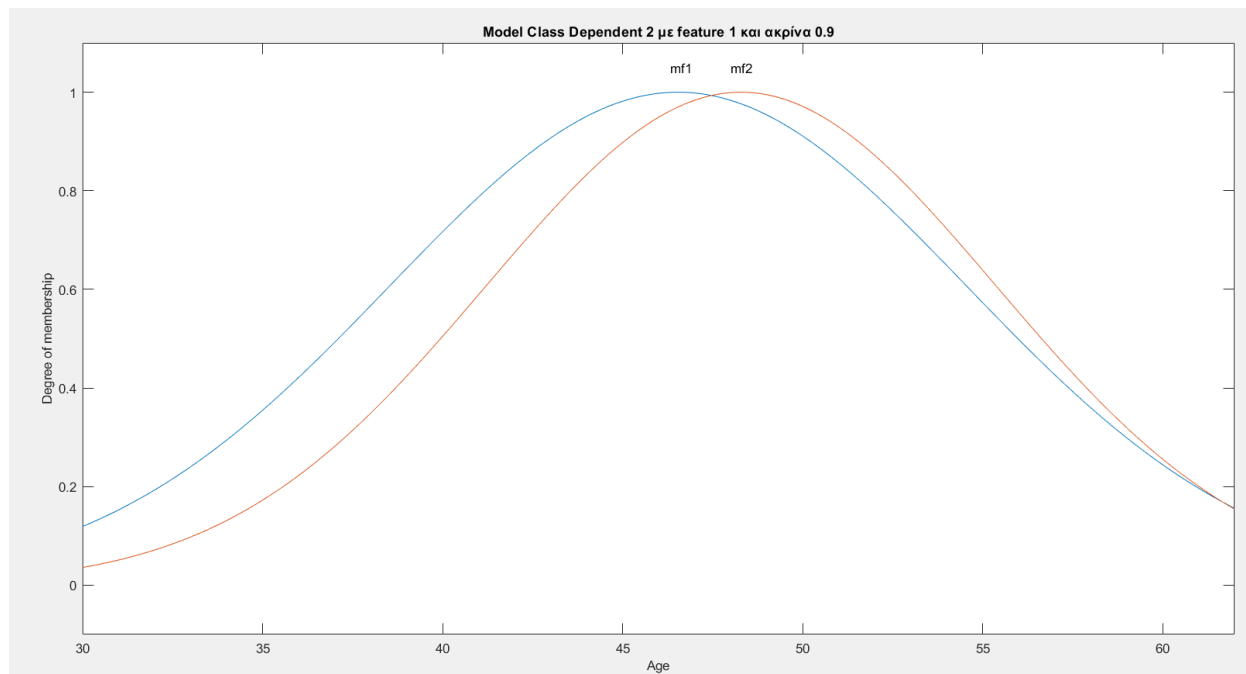
Error Matrix Μοντέλου 1

46	0
15	1

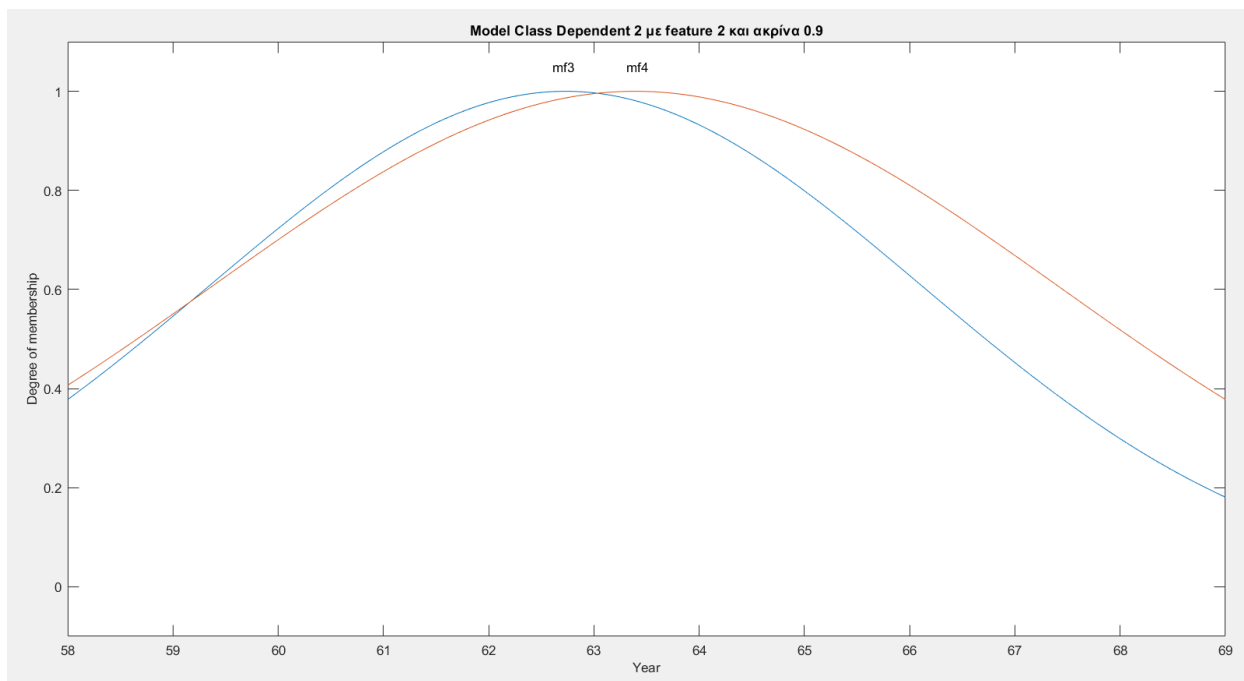
4.1.2 TSK model 2

TSK μοντέλο με ακτίνα 0.9 και class dependent.

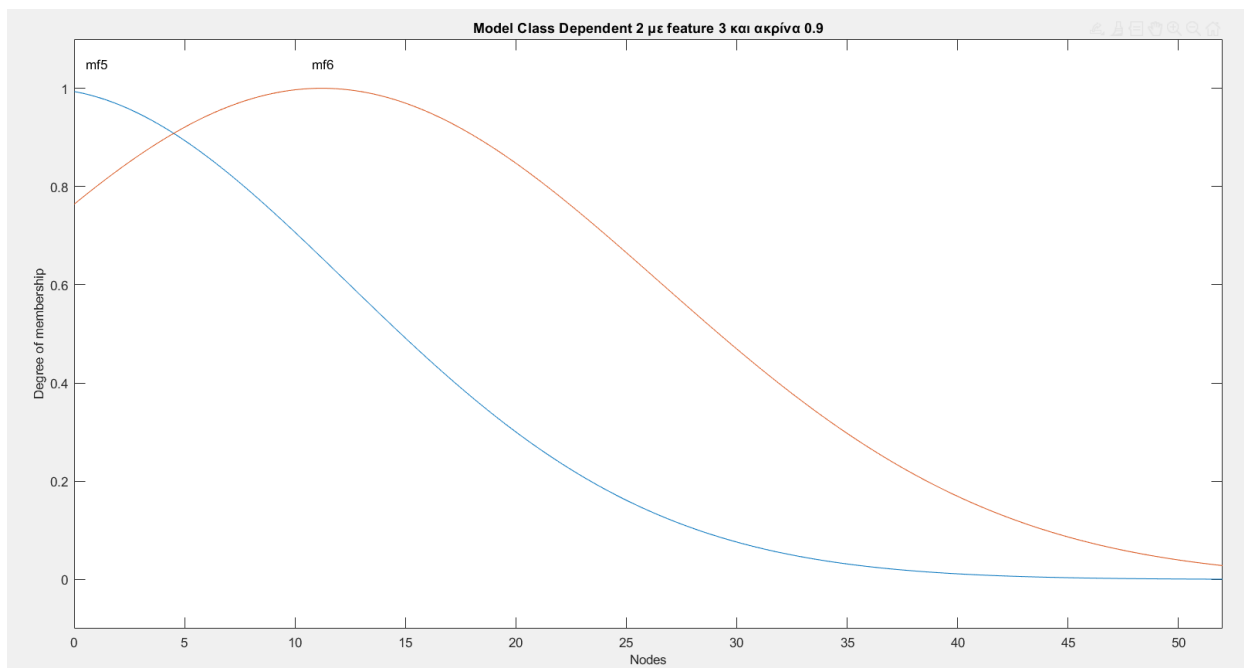
- Συναρτήσεις Συμμετοχής Μετά την Εκπαίδευση



Σχήμα 101: mf After Training – Age – class dependent - radius = 0.9

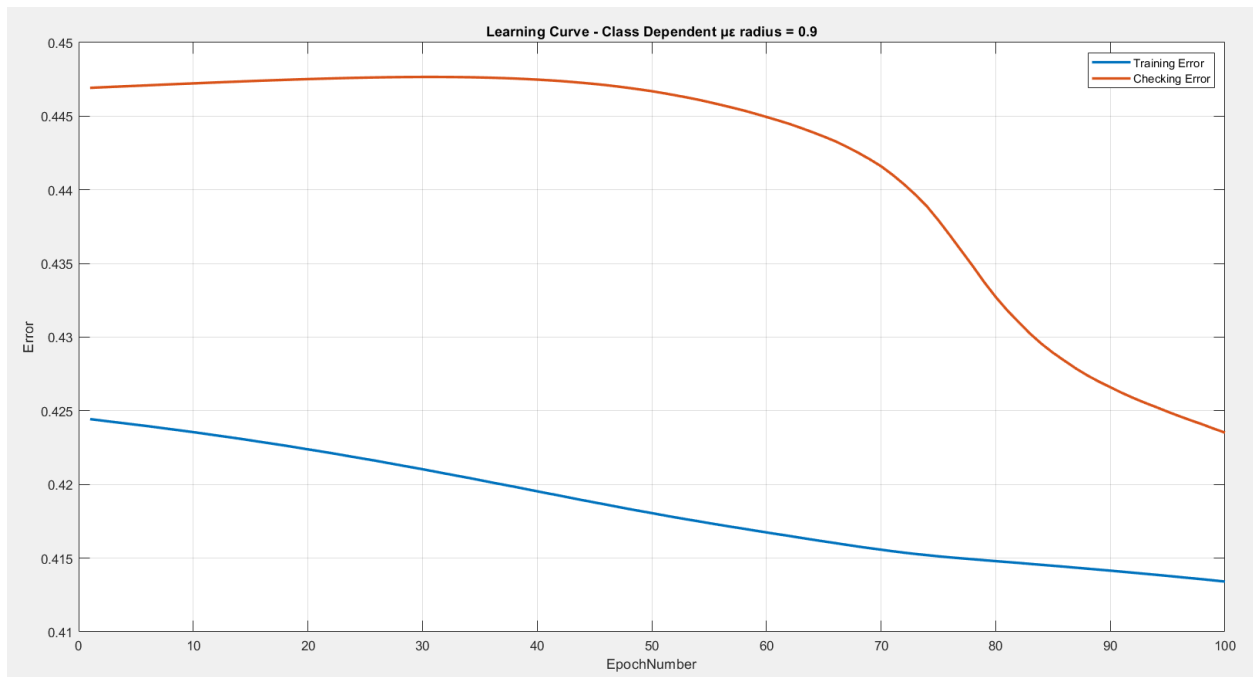


Σχήμα 102: mf After Training – Year – class dependent - radius = 0.9



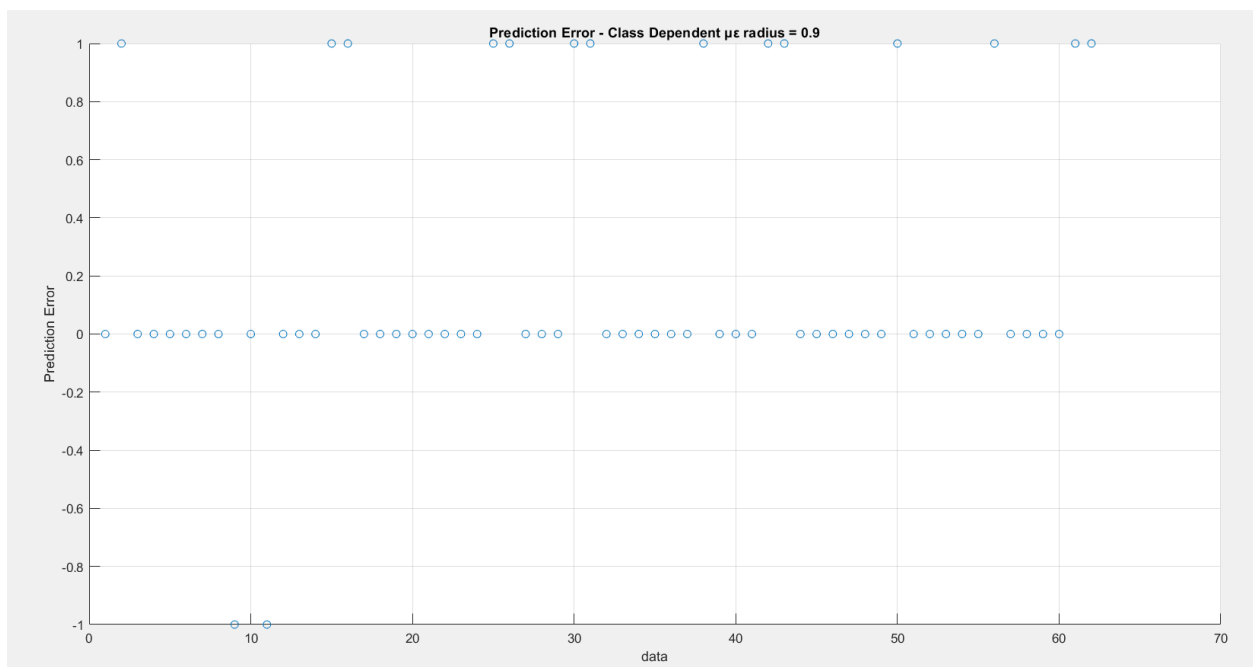
Σχήμα 103: mf After Training –Nodes – class dependent - radius = 0.9

- Learning Curve



Σχήμα 104: Learning Curve - Class Dependent - radius = 0.9

- Prediction Error



Σχήμα 105: Prediction Error - Class Dependent - radius = 0.9

Αξιολόγηση Μοντέλου 2

Μετρικές	Μοντέλο 2
OA	0.7419
PA 1	0.9565
PA 2	0.1250
UA 1	0.7586
UA 2	0.5000
K	0.1079
Rules	2

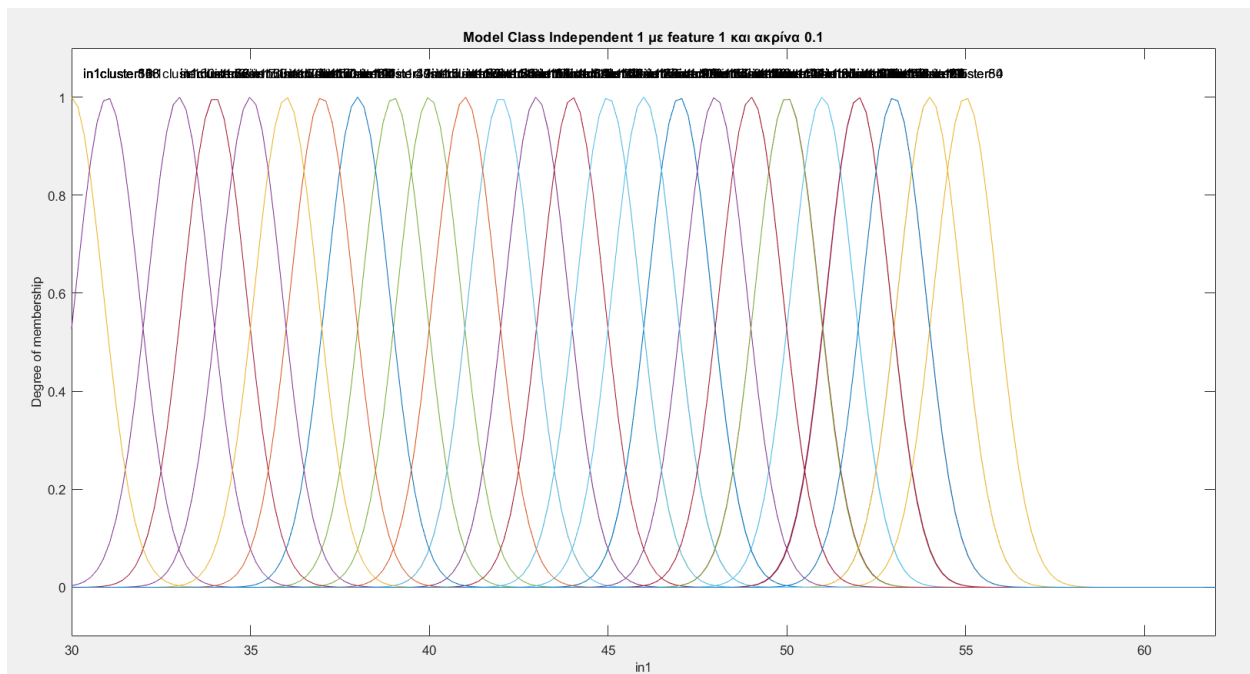
Error Matrix Μοντέλου 2

44	2
14	2

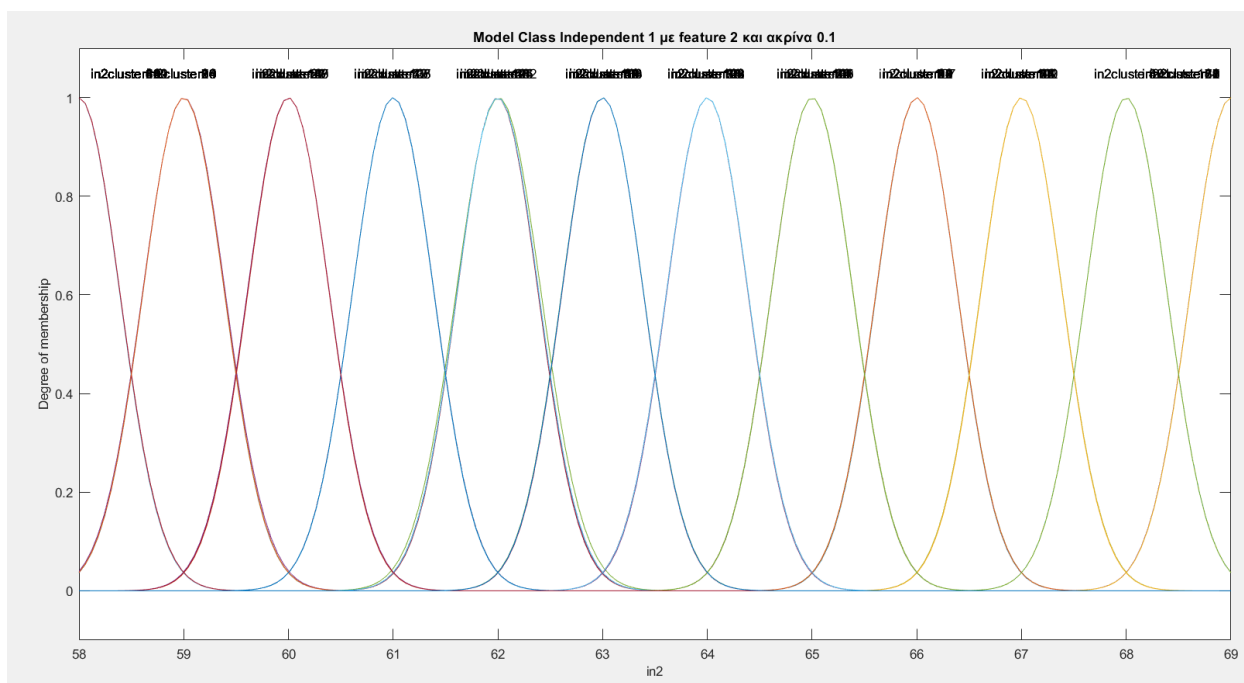
4.1.3 TSK model 3

TSK μοντέλο με ακτίνα 0.1 και class independent.

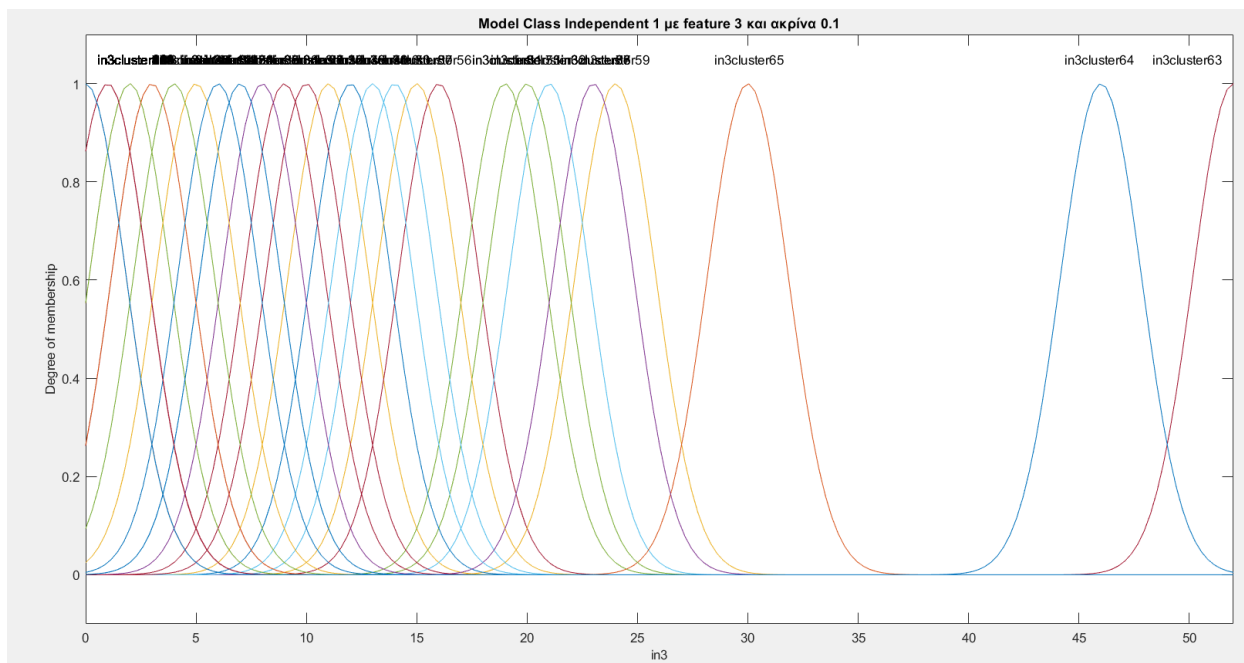
- Συναρτήσεις Συμμετοχής Μετά την Εκπαίδευση



Σχήμα 106: mf After Training – Age – class independent - radius = 0.1

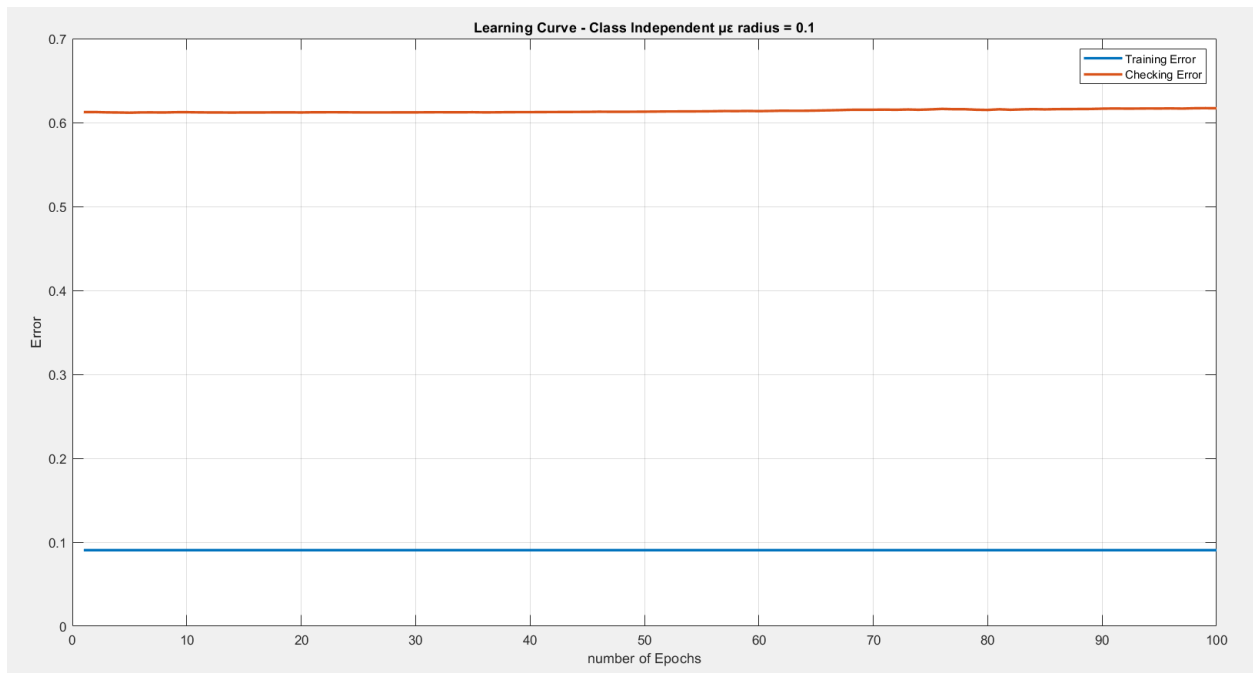


Σχήμα 107: mf After Training – Year – class independent - radius = 0.1



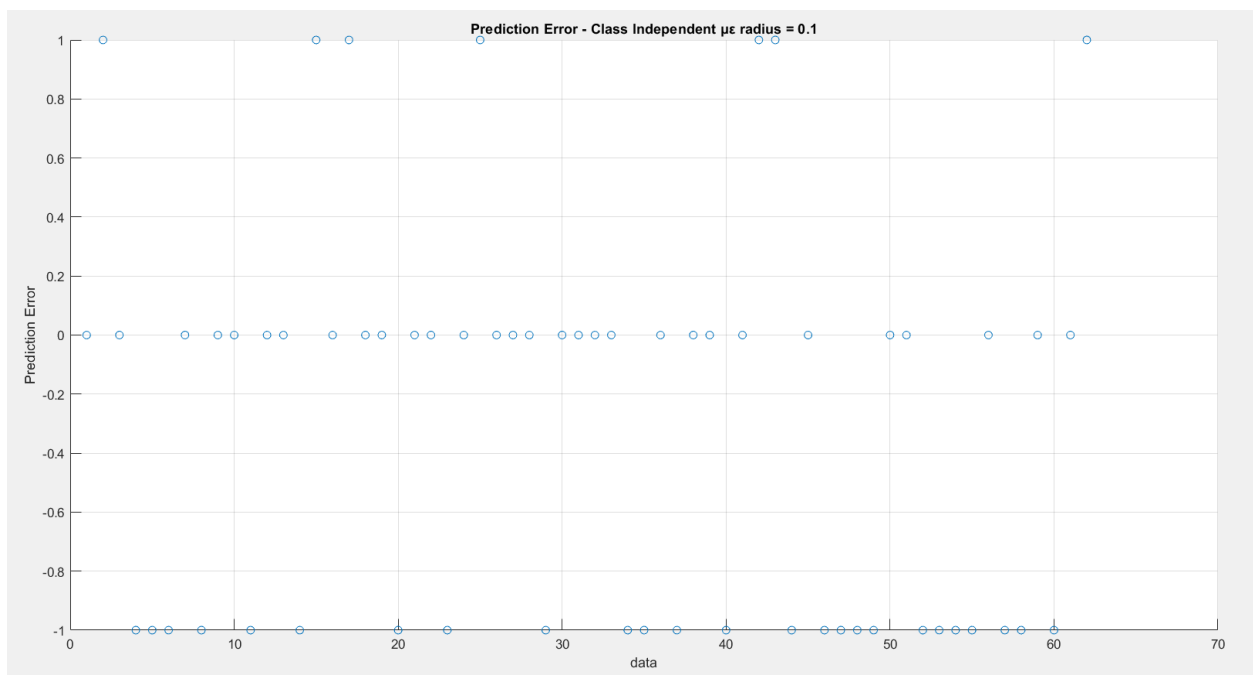
Σχήμα 108: mf After Training – Nodes – class independent - radius = 0.1

- Learning Curve



Σχήμα 109: Learning Curve - Class Independent - radius = 0.1

- Prediction Error



Σχήμα 110: Prediction Error - Class Independent - radius = 0.1

Αξιολόγηση Μοντέλου 3

Μετρικές	Μοντέλο 3
OA	0.4839
PA 1	0.4565
PA 2	0.5625
UA 1	0.7500
UA 2	0.2647
K	0.0139
Rules	134

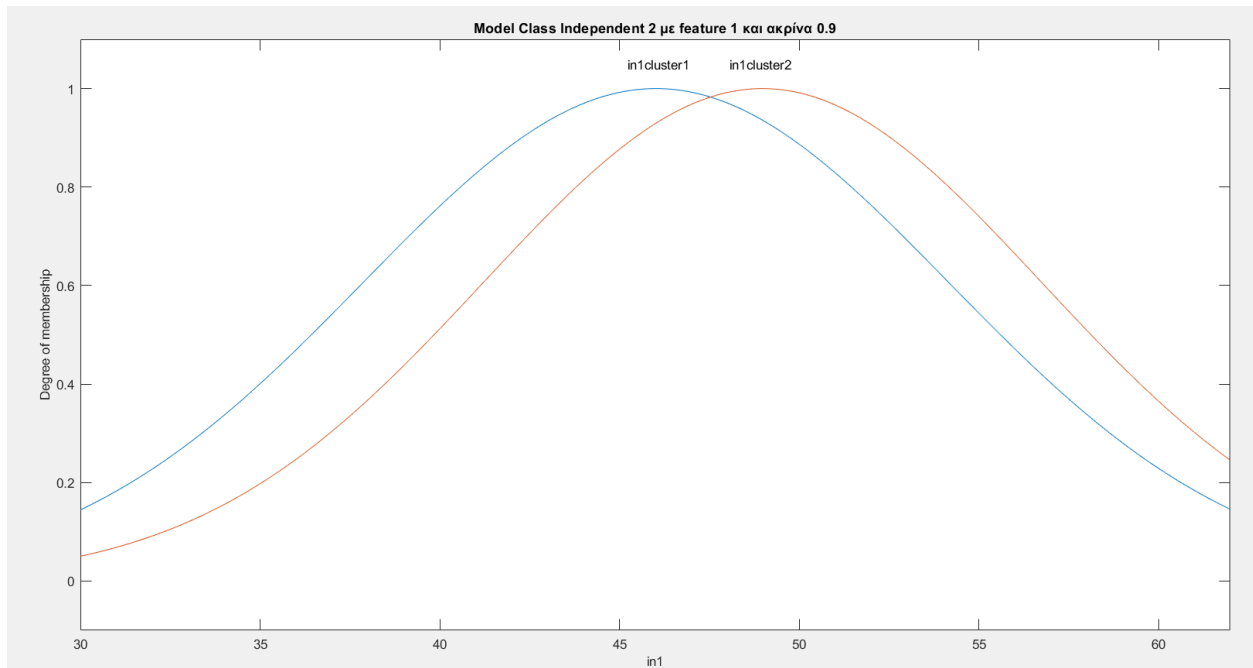
Error Matrix Μοντέλου 3

21	25
7	9

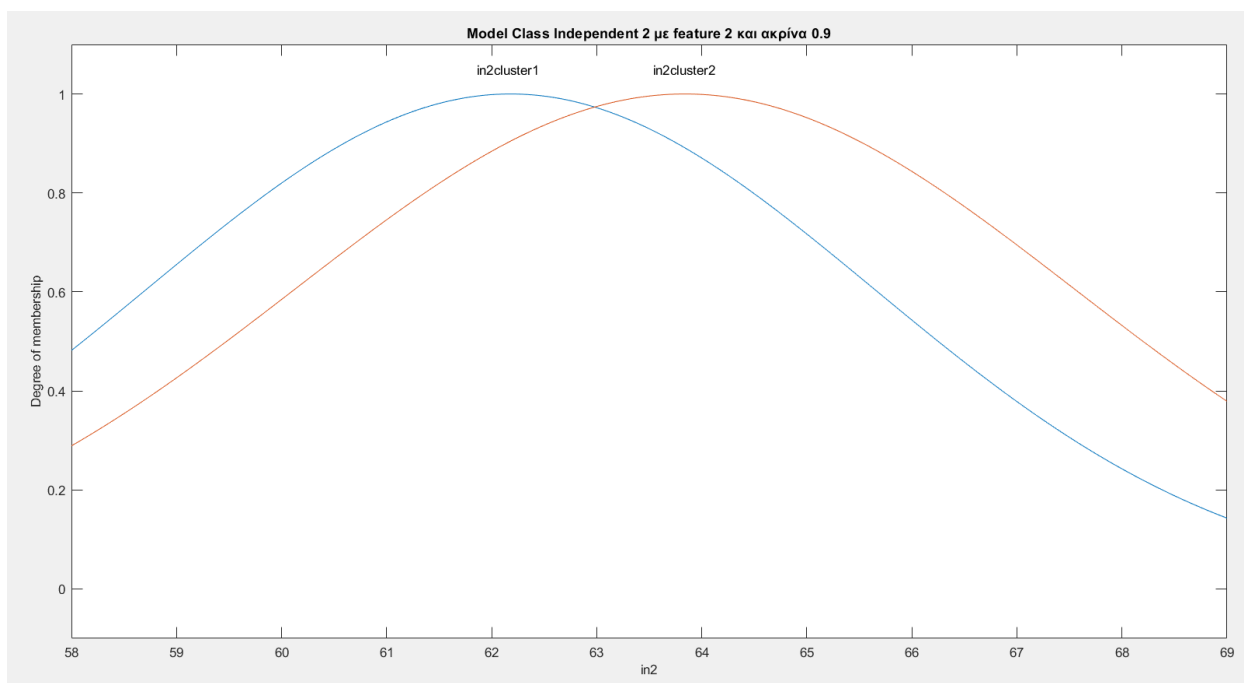
4.1.4 TSK model 4

TSK μοντέλο με ακτίνα 0.9 και class independent.

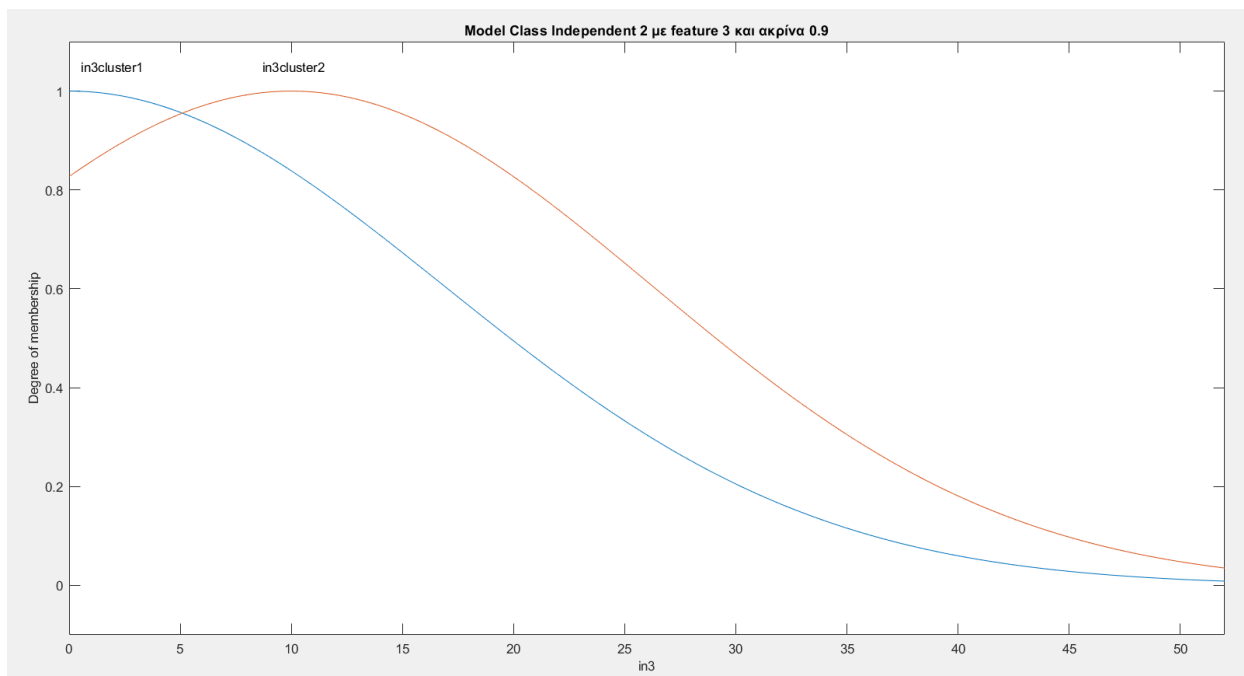
- Συναρτήσεις Συμμετοχής Μετά την Εκπαίδευση



Σχήμα 111: mf After Training – Age – class independent - radius = 0.9

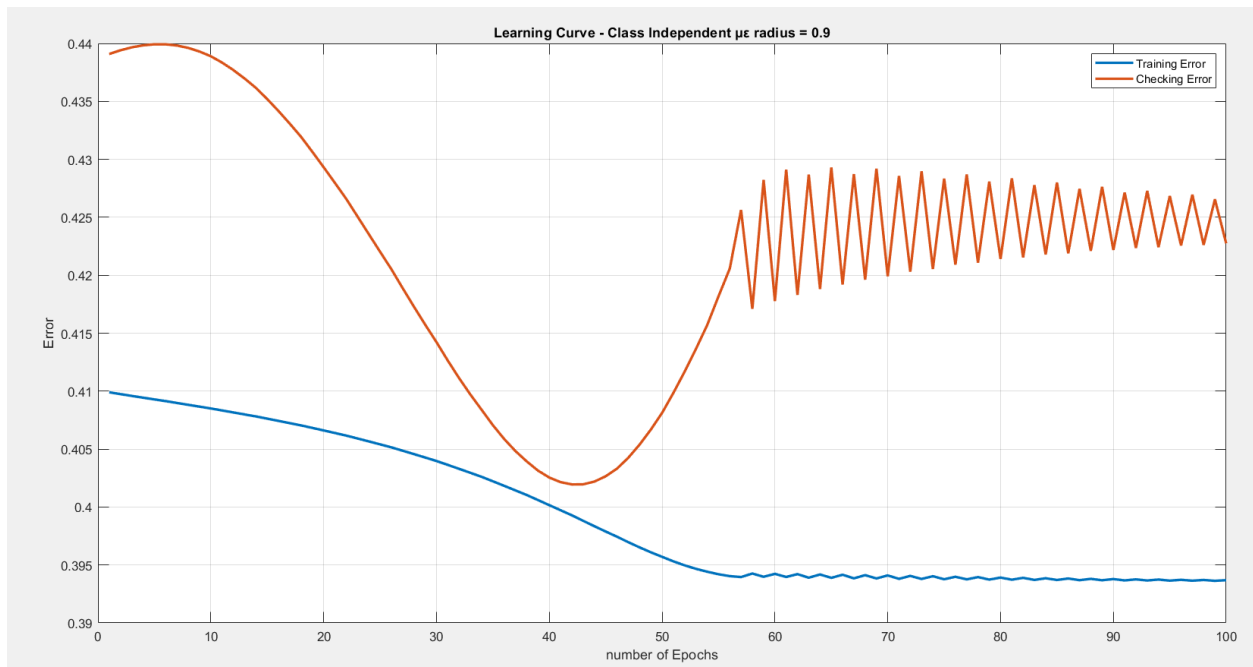


Σχήμα 112: mf After Training – Year – class independent - radius = 0.9



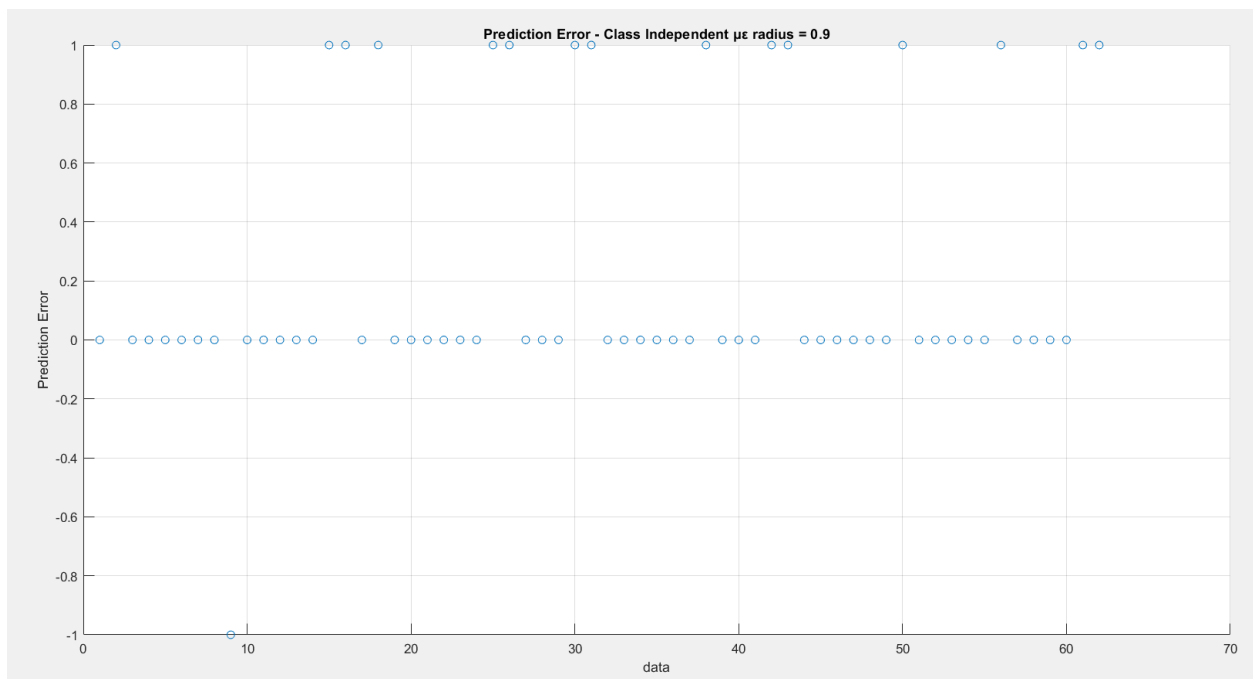
Σχήμα 113: mf After Training – Nodes – class independent - radius = 0.9

- Learning Curve



Σχήμα 114: Learning Curve - Class Independent - radius = 0.9

- Prediction Error



Σχήμα 115: Prediction Error - Class Independent - radius = 0.9

Αξιολόγηση Μοντέλου 4

Μετρικές	Μοντέλο 4
OA	0.7419
PA 1	0.9783
PA 2	0.0625
UA 1	0.7500
UA 2	0.5000
K	0.0570
Rules	2

Error Matrix Μοντέλου 4

45	1
15	1

4.1.5 Αξιολόγηση Μοντέλων και Πίνακες Σφαλμάτων

- Πίνακας Μετρικών Αξιολόγησης

Μοντέλα\Μετρικές	OA	PA 1	PA 2	UA 1	UA 2	K	Rules
Μοντέλο 1	0.7581	1.0000	0.0625	0.7541	1.0000	0.0900	132
Μοντέλο 2	0.7419	0.9565	0.1250	0.7586	0.5000	0.1079	2
Μοντέλο 3	0.4839	0.4565	0.5625	0.7500	0.2647	0.0139	134
Μοντέλο 4	0.7419	0.9783	0.0625	0.7500	0.5000	0.0570	2

- Οι Μετρικές Αξιολόγησης υπολογίζονται από τους παρακάτω τύπους:

$$OA = (TP + TN) / (TP + FP + TN + FN)$$

$$PA1 = TP / (TP + FN)$$

$$PA2 = TN / (TN + FP)$$

$$UA1 = TP / (TP + FP)$$

$$UA2 = TN / (TN + FN)$$

$$K = (N * (TP + TN) - ((TP + FP) * (TP + FN) + (FN + TN) * (FP + TN))) / (N^2 - ((TP + FP) * (TP + FN) + (FN + TN) * (FP + TN)))$$

- Πίνακες Σφαλμάτων

Μοντέλο 1	Predicted Class 1	Predicted Class 2
True Class 1	46	0
True Class 2	15	1

Μοντέλο 2	Predicted Class 1	Predicted Class 2
True Class 1	44	2
True Class 2	14	2

Μοντέλο 3	Predicted Class 1	Predicted Class 2
True Class 1	21	25
True Class 2	7	9

Μοντέλο 4	Predicted Class 1	Predicted Class 2
True Class 1	45	1
True Class 2	15	1

- Οι πίνακες Σφαλμάτων έχουν την εξής δομή:

Μοντέλο	Predicted Class 1	Predicted Class 2
True Class 1	TP	FN
True Class 2	FP	TN

- Ο πίνακας Σφαλμάτων υπολογίζεται με την εντολή “confusionmat(TrueData, PredictedData)”, που επιστρέφει τον confusion matrix.

4.1.6 Συμπεράσματα και Βέλτιστο Μοντέλο

Από τους παραπάνω πίνακες παρατηρούμε ότι τα καλύτερα μοντέλα είναι το 1 και το 2, δηλαδή τα Class Dependent, γιατί έχουν το μικρότερο Overall Accuracy, με τιμές 0.7581 και 0.7419 αντίστοιχα και πετυχαίνουν καλύτερες μετρικές.

Επίσης, έχουν πολύ καλό Producer's Accuracy με τιμές 1 και 0.9565 αντίστοιχα, δηλαδή ομαδοποιούν σωστά τα δείγματα της κλάσης 1.

Και τα δύο μοντέλα έχουν ικανοποιητικές τιμές User's Accuracy, με το μοντέλο 1 να έχει καλύτερο UA 2.

Όμως, το μοντέλο 2 έχει καλύτερη Εκτίμηση Πραγματικής Στατιστικής Παραμέτρου K με τιμή 0.1079, με μικρή διαφορά από το μοντέλο 1 που έχει K με τιμή 0.0900.

Βλέπουμε ότι τα μοντέλα με Class Dependent Subtractive Clustering, που έχουν ίδια ακτίνα για το cluster, πετυχαίνουν καλύτερα αποτελέσματα ομαδοποίησης σε σχέση με τα Class Independent μοντέλα. Δηλαδή, το μοντέλο 1 είναι καλύτερο από το μοντέλο 3 και το μοντέλο 2 είναι καλύτερο από το μοντέλο 4.

Παρατηρούμε ότι τα μοντέλα 1 και 3 με μικρή ακτίνα cluster 0.1 έχουν πολύ μεγάλο αριθμό κανόνων. Αντίθετα, τα μοντέλα 2 και 4 με μεγάλη ακτίνα 0.9 έχουν μόνο δύο κανόνες για κάθε Συνάρτηση Συμμετοχής.

Όμως, είναι φανερό ότι ο αριθμός των κανόνων (άρα και το μέγεθος της ακτίνας των clusters) δεν επηρεάζει την απόδοση του μοντέλου. Αυτό που επηρεάζει σημαντικά την απόδοση είναι η μέθοδος με την οποία έγινε ο διαμερισμός του χώρου των δεδομένων.

Γνωρίζουμε όμως ότι όσο πιο μικρή είναι η ακτίνα των clusters, τόσο περισσότερους κανόνες παράγει το μοντέλο. Αυτό συνεπάγεται ότι το μοντέλο γίνεται όλο και πιο περίπλοκο, με αποτέλεσμα να κινδυνεύει να πάθει overtraining, επομένως είναι πιο δύσκολο να γενικεύσει όσα έμαθε και σε διαφορετικά datasets. Ένας τρόπος αντιμετώπισης αυτού του προβλήματος είναι να αφαιρέσουμε τους κανόνες που κάνουν overlap, ώστε το μοντέλο να έχει λίγους και σημαντικούς κανόνες.

Με βάση όλα τα παραπάνω επιλέγω ως βέλτιστο μοντέλο, το **Μοντέλο 2** με Class Dependent και ακτίνα 0.9 (2 κανόνες).

4.2 Part 2 - Εφαρμογή σε dataset με υψηλή διαστασιμότητα

Σε αυτή την εργασία ερευνούμε τις βέλτιστες παραμέτρους (αριθμό σημαντικότερων features, αριθμό κανόνων, ακτίνα των clusters) σε dataset υψηλής διαστασιμότητας. Το dataset που μελετάμε είναι το “Epileptic Seizure Recognition” από το UCI repository. Περιλαμβάνει 11500 δείγματα, τα οποία περιγράφονται από 178 features και 1 output.

Η έξοδος αποτελείται από πέντε κλάσεις {1,2,3,4,5}.

1 = Recording of seizure activity

2 = They recorder the EEG from the area where the tumor was located

3 = Yes, they identify where the region of the tumor was in the brain and recording the EEG activity from the healthy brain area

4 = Eyes closed, means when they were recording the EEG signal the patient had their eyes closed

5 = Eyes open, means when they were recording the EEG signal of the brain the patient had their eyes open

Είναι απαραίτητο να επιλέξουμε ένα μικρότερο σύνολο features με τα οποία θα εργαστούμε, διότι τα 178 είναι απαγορευτικά μεγάλο νούμερο, λόγω της έκρηξης των κανόνων If-Then.

Αυτό επιτυγχάνεται με την εντολή Relief.

Το μοντέλο υλοποιείται στο script “classification_Epileptic.m”.

Όπως και στην εργασία #3, δοκιμάζουμε διαφορετικές τιμές για τον αριθμό των σημαντικότερων features [5, 7, 9, 11] που θα κρατήσουμε για να τροφοδοτήσουμε το μοντέλο, καθώς και διάφορες τιμές για την ακτίνα radius [0.2, 0.4, 0.6, 0.8, 1] των clusters.

Όσο αυξάνουμε τον αριθμό των features τόσο πιο αργό γίνεται το μοντέλο και σε πολύ υψηλές τιμές μη λειτουργικό.

Χρησιμοποιούμε την τεχνική Grid Search και διαχωρίζουμε τα δεδομένα με την μέθοδο CV Partition k-fold με k=5.

Ο διαχωρισμός του χώρου του input dataset γίνεται με subtractive clustering στα δεδομένα του συνόλου εκπαίδευσης, που ανήκουν στην εκάστοτε κλάση ξεχωριστά (**Class Dependent**).

Τα μοντέλα δημιουργούνται με new fis κατασκευασμένο εξ αρχής, με Membership Functions που ορίζουμε εμείς στο script.

Κάθε μοντέλο εκπαιδεύεται με την μέθοδο anfis και χρησιμοποιούμε 100 EpochNumber.

Για την αξιολόγηση των μοντέλων χρησιμοποιούμε την μετρική Overall Accuracy.

4.2.1 Αποτελέσματα του Grid Search

- Πίνακας Overall Accuracy **OA**

Features\radius	0.2	0.4	0.6	0.8	1
5	0.3629	0.3392	0.3235	0.3142	0.3250
7	0.3810	0.3604	0.3535	0.3435	0.3563
9	0.3939	0.3684	0.3857	0.3786	0.3544
11	0.4017	0.3730	0.3800	0.3901	0.3673

- Πίνακας Αριθμού Κανόνων **Rules**

Features\radius	0.2	0.4	0.6	0.8	1
5	31	10	5.6	5	5
7	42	10.8	6	5.2	5
9	53.2	12.2	7	5.2	5
11	122.2	13.8	7.2	6	5

Παρατηρούμε ότι ο αριθμός των κανόνων αυξάνεται όταν ο αριθμός των features αυξάνεται καιόταν η ακτίνα των clusters μειώνεται.

Σύμφωνα με τα παραπάνω αποτελέσματα πετυχαίνουμε το βέλτιστο Overall Accuracy

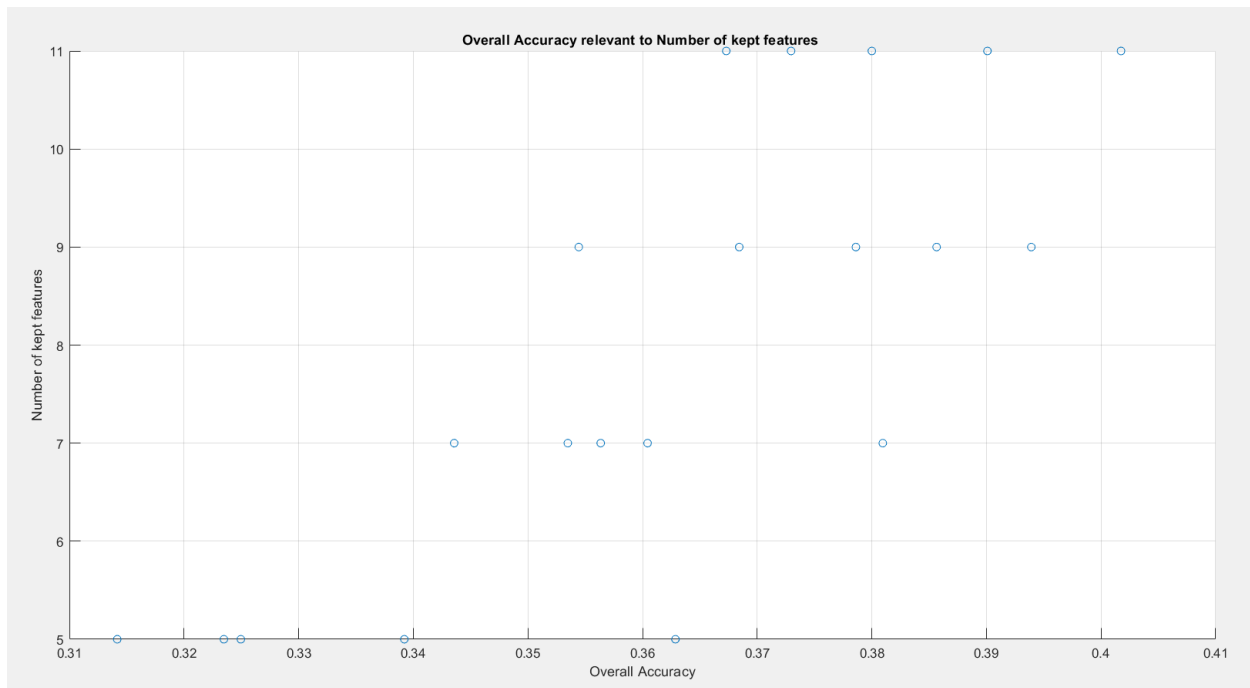
OA = 0.4017 με παραμέτρους:

- features = 11
- rules = 122.2
- radius = 0.2

4.2.2 Διαγράμματα του Grid Search

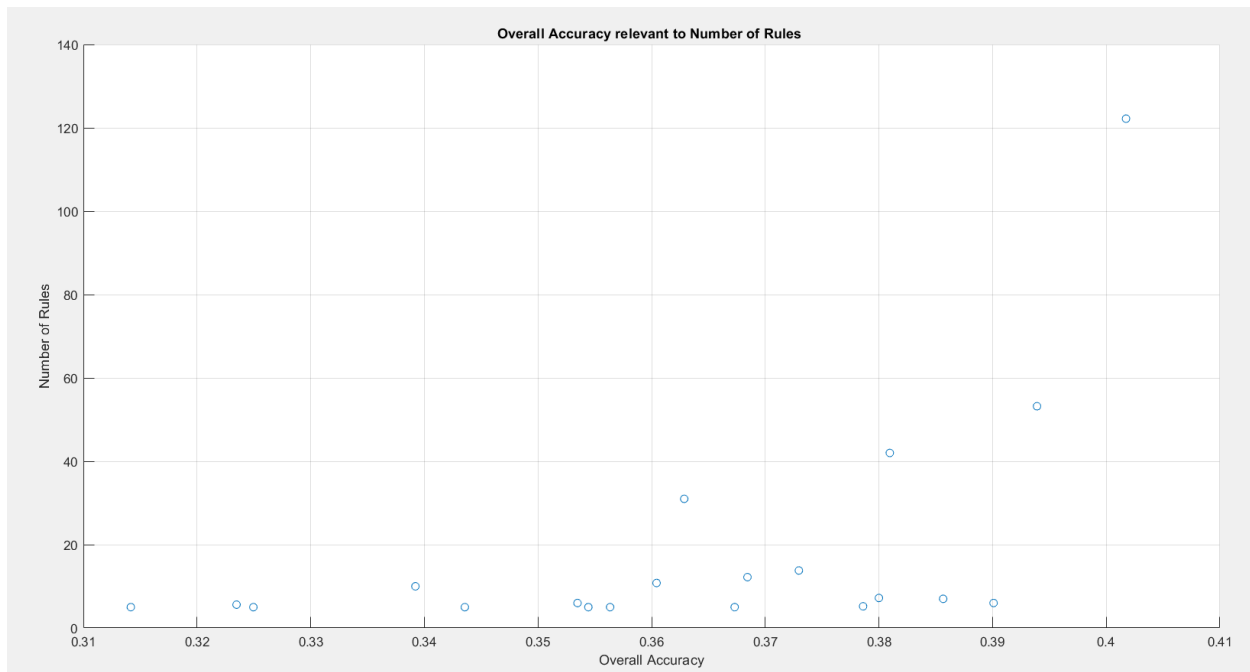
Παρακάτω φαίνονται τα διαγράμματα scatters για το Overall Accuracy συναρτήσει του αριθμού των σημαντικότερων features, του αριθμού των κανόνων και της ακτίνας των clusters.

- Overall Accuracy συναρτήσει του αριθμού των features



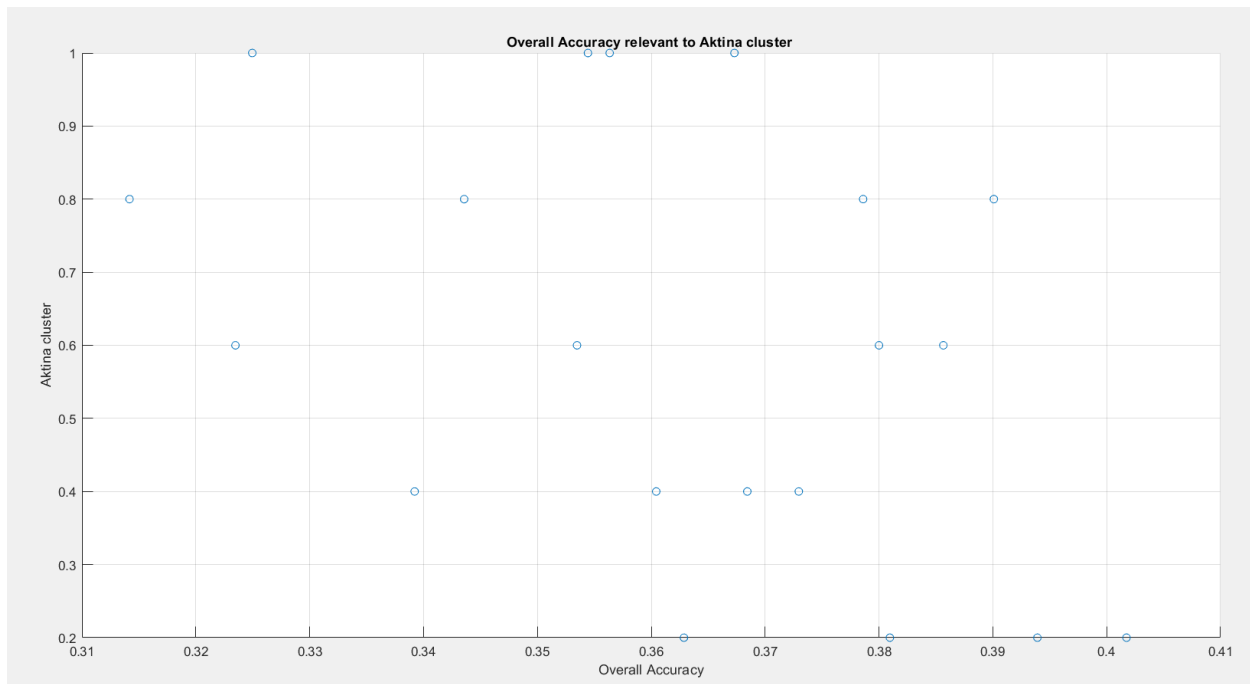
Σχήμα 116: OA - features

- Overall Accuracy συναρτήσει του αριθμού των κανόνων



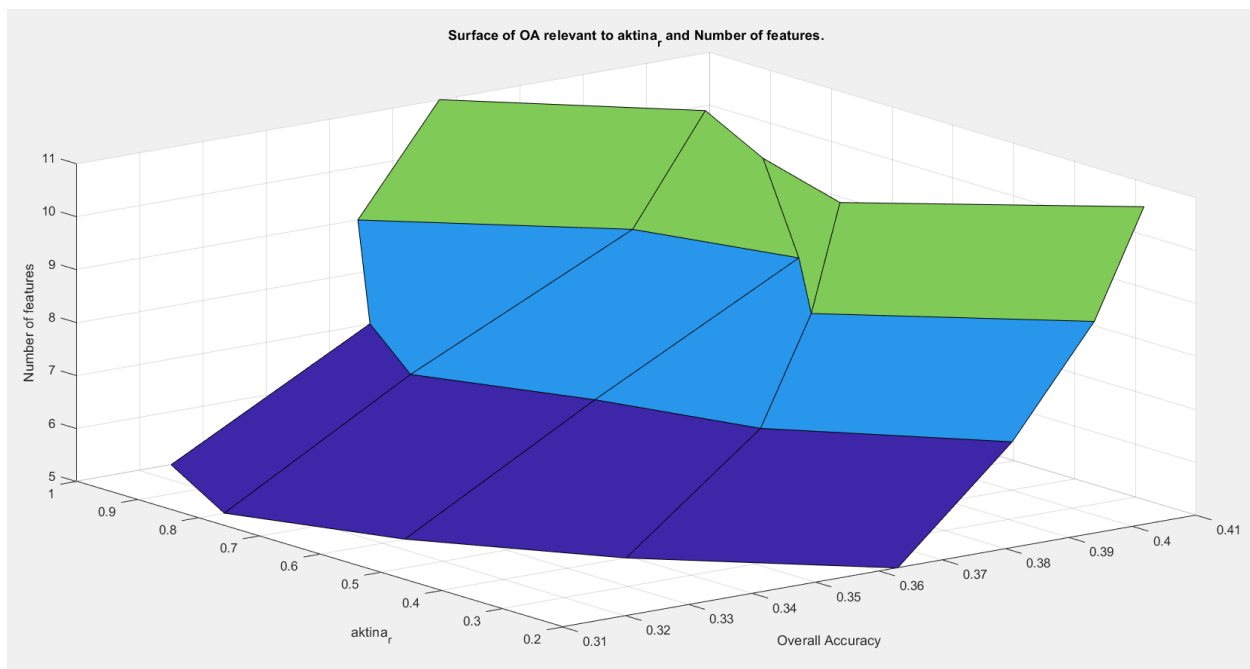
Σχήμα 117: OA - rules

- Overall Accuracy συναρτήσει της ακτίνας των clusters



Σχήμα 118: OA - radius

- Overall Accuracy συναρτήσει των features & radius



Σχήμα 118: OA – features & radius

Από τα παραπάνω διαγράμματα συμπεραίνουμε ότι το OverallAccuracy αυξάνεται όταν ο αριθμός των σημαντικότερων features αυξάνεται και όταν η ακτίνα των clusters μειώνεται.

Όμως, αυτό δεν είναι κανόνας, γιατί μεγάλος αριθμός feature και μικρή ακτίνα μπορεί να οδηγήσει σε overtraining, δηλαδή το μοντέλο να μάθει τέλεια το συγκεκριμένο dataset, αλλά να μην μπορεί να γενικεύσει όσα έμαθε.

Παρατηρούμε ότι παρόλο που έχουμε το καλύτερο OA για μεγάλο αριθμό features = 11, οι τιμές είναι αρκετά κοντά και για μικρό αριθμό features = 5, δεν υπάρχουν δηλαδή ακραίες τιμές.

4.3.3 Βέλτιστο Μοντέλο

Το βέλτιστο μοντέλο επιλέχθηκε από το script “classification_Epileptic.m” και υλοποιείται στο script “classification_Optimal_TSK_model.m”.

Αποτελείται από 11 features και 0.2 ακτίνα.

- Πίνακας Μέσου Όρου Σφαλμάτων Error Matrix Βέλτιστου Μοντέλου (Confusion Matrix)

333.8	72.4	34.8	17.2	0.8
23.2	47.2	256.2	129.2	4.2
3	29	283.8	138.8	5.4
2.8	43.4	173	222	18.8
0.2	7.6	170	255.2	27

- Μέσος Όρος Μετρικών Βέλτιστου Μοντέλου

Mean Μετρικών	Βέλτιστο Μοντέλο
OA	0.3977
PA1	0.7278
PA2	0.1026
PA3	0.6170
PA4	0.4826
PA5	0.0587
UA1	0.9199
UA2	0.2351
UA3	0.3092
UA4	0.2911
UA5	0.4711
K	0.2472
Rules	92

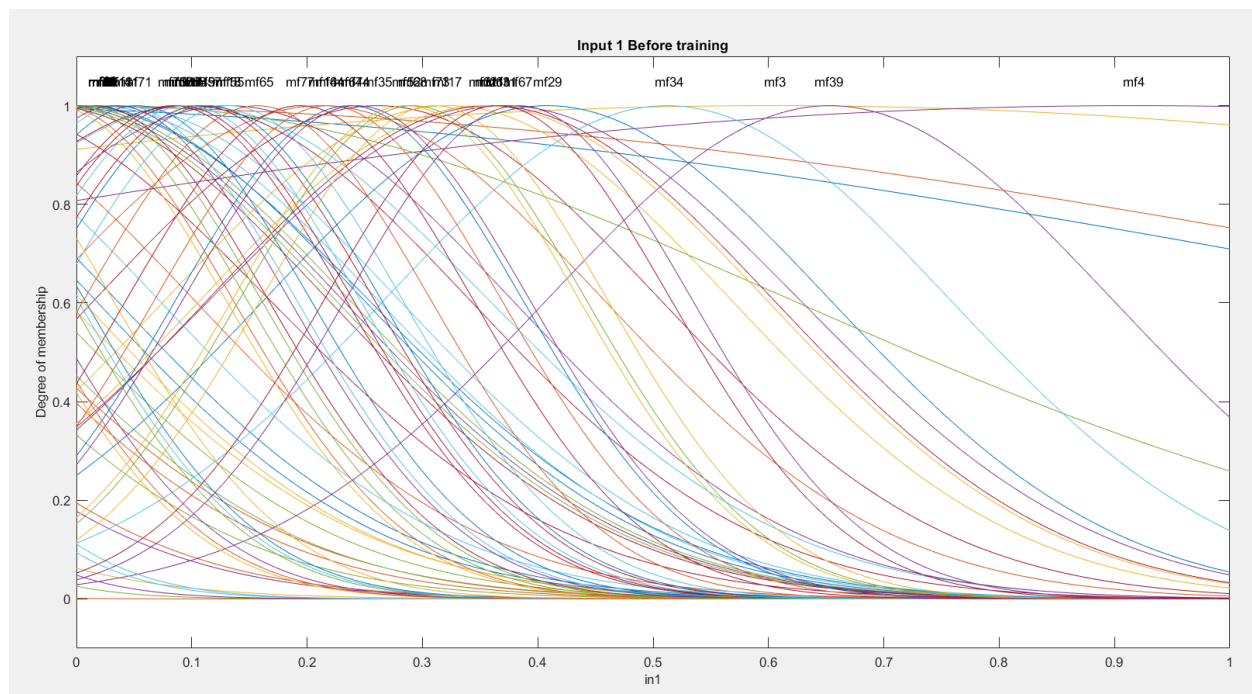
Παρατηρούμε ότι πετυχαίνουμε πολύ ικανοποιητικά αποτελέσματα.

- Ενδεικτικά κάποιες προβλέψεις του μοντέλου και οι πραγματικές του τιμές

Real Vale	Prediction
4	3
1	1
5	4
2	4
3	3
2	1
1	1
2	2
1	4
4	4
5	3
1	3
3	3

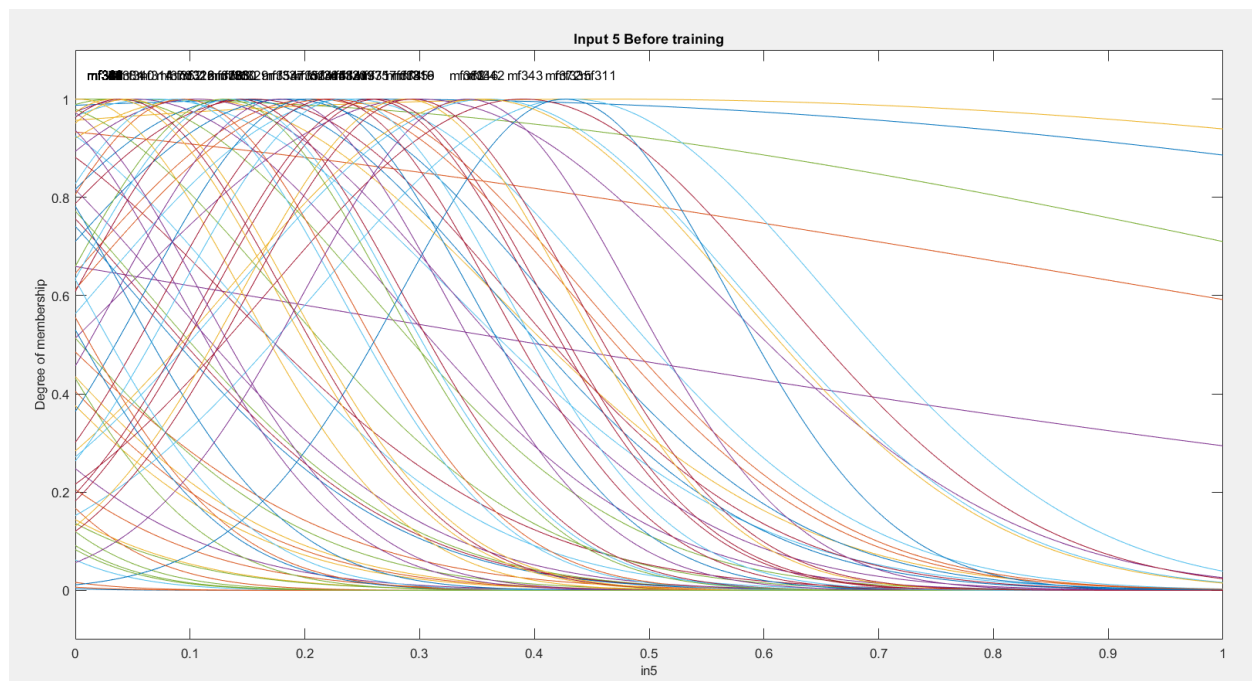
- Διαγράμματα Βέλτιστου Μοντέλου:

Συναρτήσεις Συμμετοχής Πριν την εκπαίδευση – Input 1



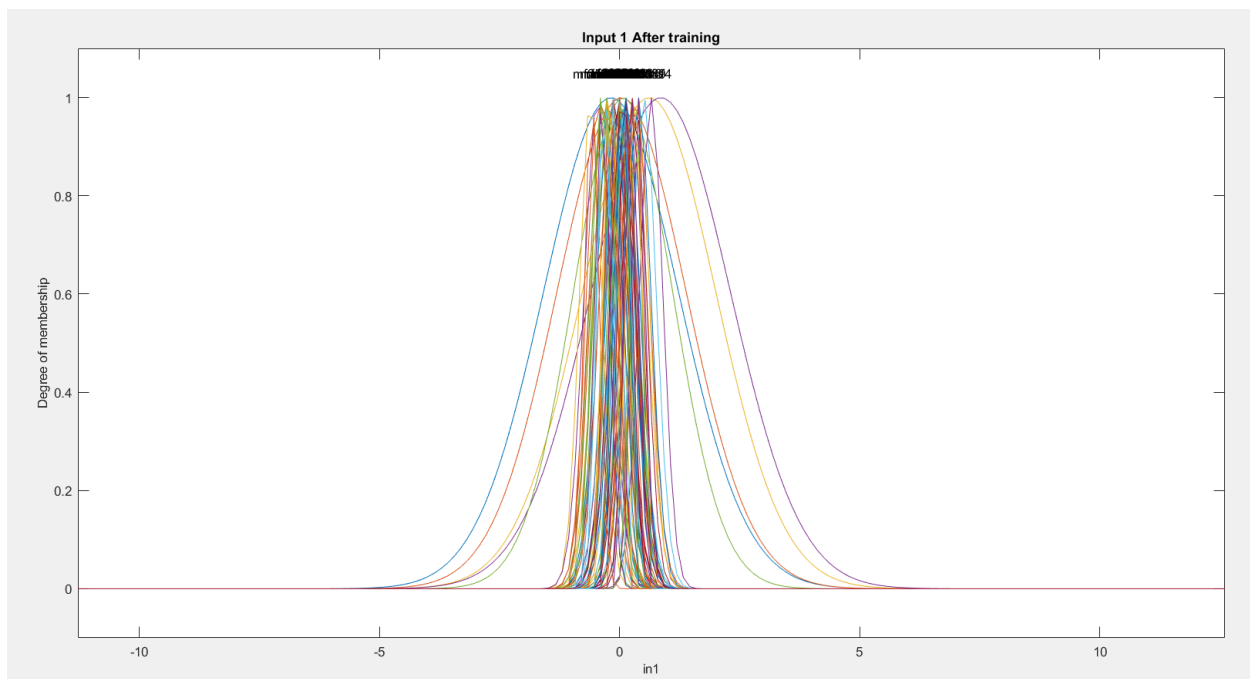
Σχήμα 119: mf Before Training – input 1

Συναρτήσεις Συμμετοχής Πριν την εκπαίδευση – Input 5



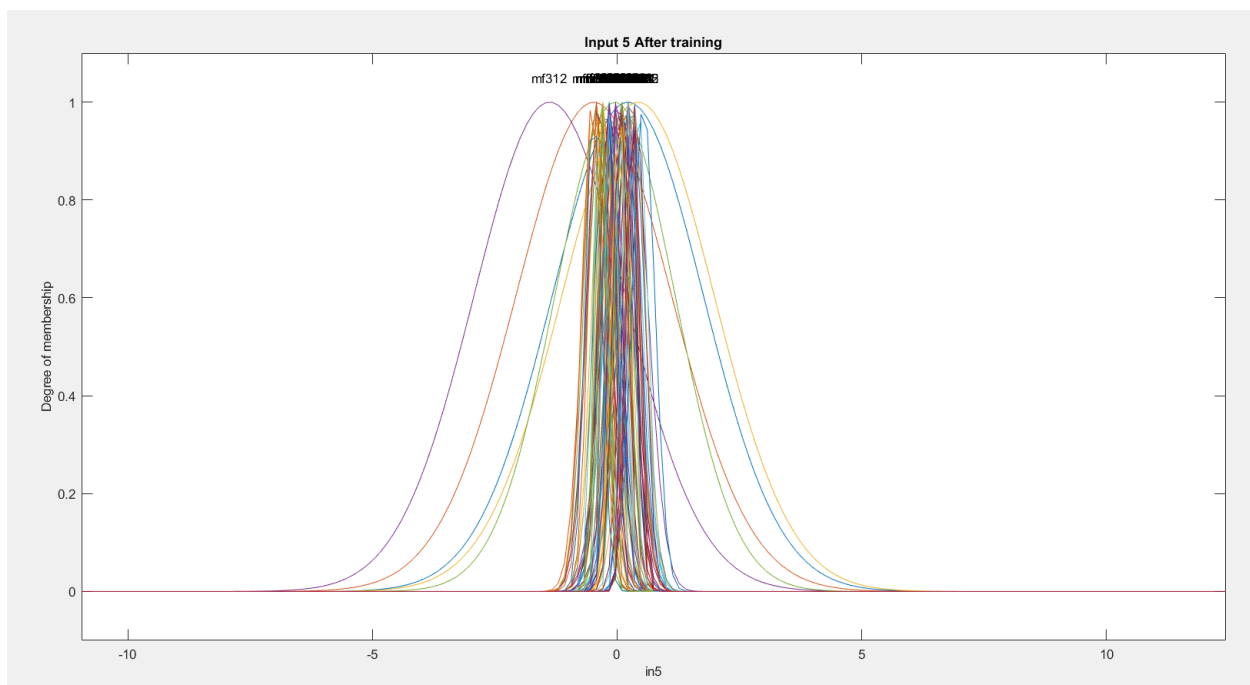
Σχήμα 120: mf Before Training – Input 5

Συναρτήσεις Συμμετοχής Μετά την εκπαίδευση – Input 1



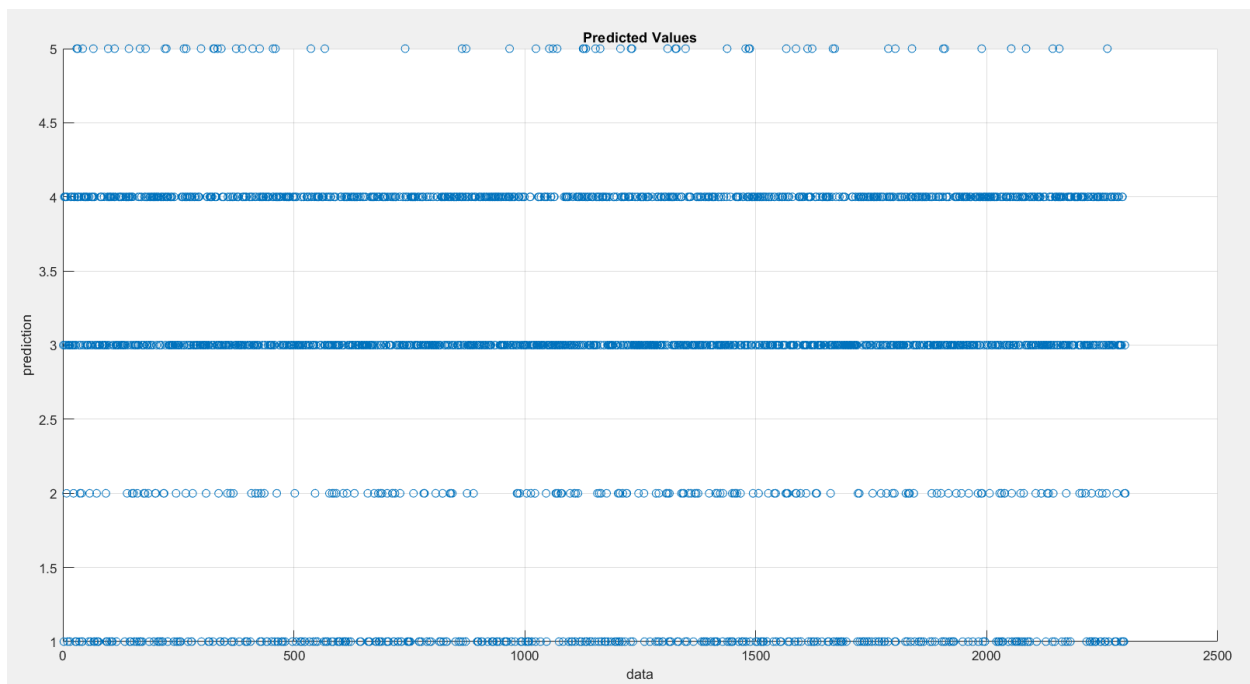
Σχήμα 121: mf After Training – input 1

Συναρτήσεις Συμμετοχής Μετά την εκπαίδευση – Input 5



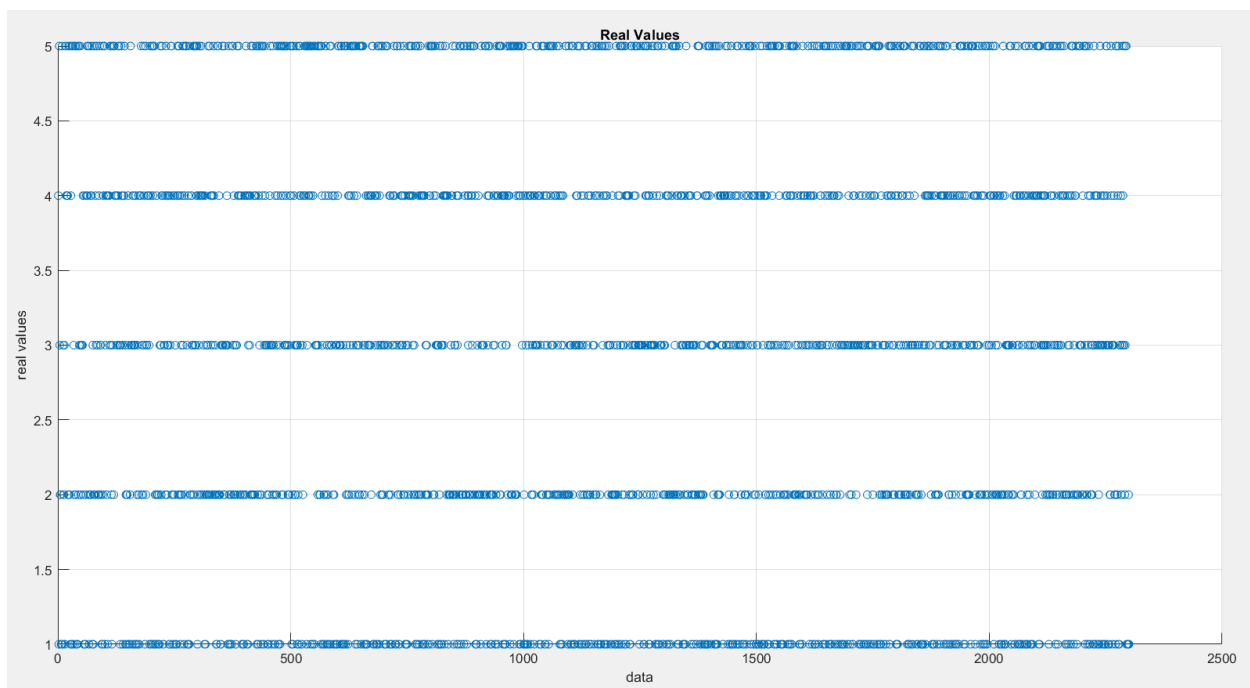
Σχήμα 122: mf After Training – input 5

Τιμές Πρόβλεψης



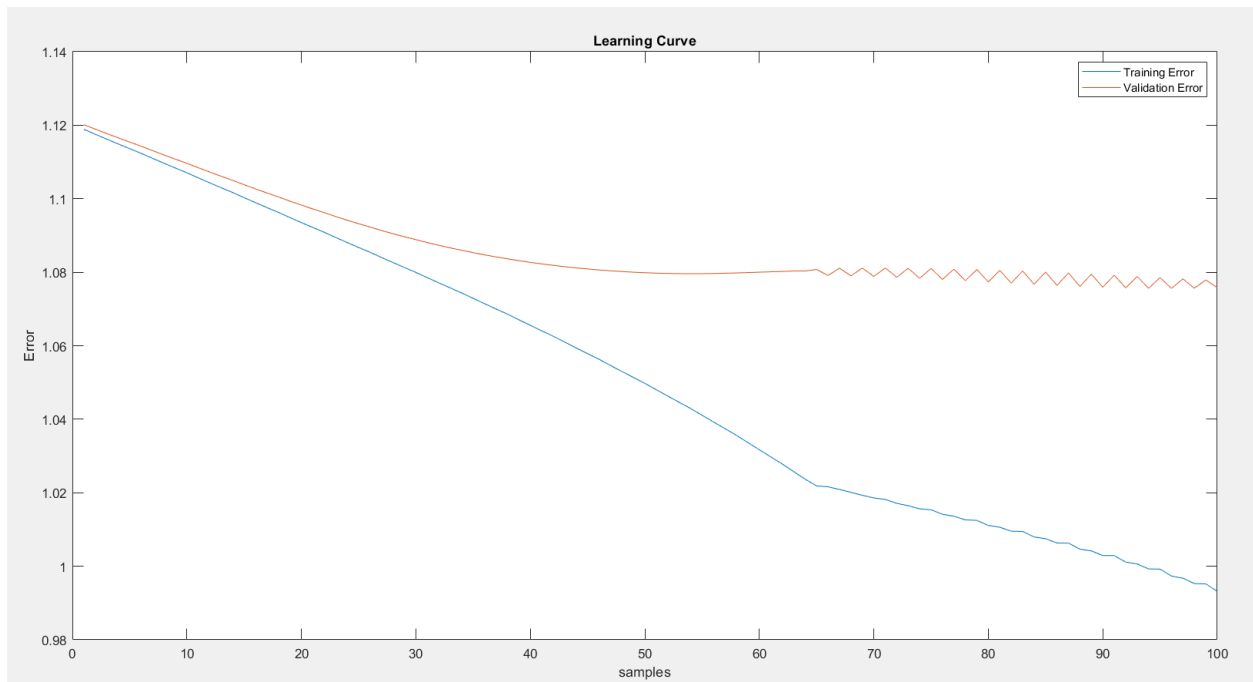
Σχήμα 123: Prediction Values

Πραγματικές Τιμές



Σχήμα 124: Real Values

Καμπύλη Εκμάθησης



Σχήμα 125: Learning Curve

- Συμπεράσματα

Αφού τρέξαμε το script “classification_Optimal_TSK_model.m” με 11 features και 0.2 radius ως το βέλτιστο μοντέλο, εξάγαμε τους παραπάνω πίνακες και διαγράμματα.

Το μοντέλο χρησιμοποιεί κατά μέσο όρο 92 κανόνες με την μέθοδο Subtractive Clustering, σε αντίθεση με την Grid Partition, η οποία θα χρειαζόταν 3^{11} κανόνες, αν υποθέσουμε ότι θα είχαμε 3 ασαφή σύνολα για κάθε είσοδο. Αυτό θα καθιστούσε το μοντέλο μη λειτουργικό.

Παρατηρούμε ότι πολλές Συναρτήσεις Συμμετοχής επικαλύπτονται, πράγμα που σημαίνει ότι αρκετά ασαφή σύνολα είναι αχρείαστα και μπορούν να αφαιρεθούν.

Από τους Πίνακες του Accuracy και του Error Matrix συμπεραίνουμε τα εξής:

- Τα δείγματα που ανήκουν στην κλάση 1, έχουν πιθανότητα να ταξινομηθούν σε αυτή την κλάση με πιθανότητα $PA1 = 72.78\%$. Αντίστοιχα, αυτά που ανήκουν στην κλάση 2 ταξινομούνται στην σωστή κλάση με πιθανότητα $PA2 = 10.26\%$, στην κλάση 3 με $PA3 = 61.70\%$, στην κλάση 4 με $PA4 = 48.26\%$ και στην κλάση 5 με $PA5 = 5.87\%$.
- Αν επιλέξουμε με τυχαίο τρόπο ένα δείγμα από το σύνολο του dataset, το οποίο ταξινομήθηκε στην κλάση 1 από το βέλτιστο μοντέλο, η πιθανότητα να ανήκει όντως στην κλάση 1 είναι $UA1 = 91.99\%$. Αντίστοιχα, ένα δείγμα που ταξινομήθηκε στην κλάση 2, ανήκει όντως σε αυτή την κλάση με πιθανότητα $UA2 = 23.51\%$, στην κλάση 3 με $UA3 = 30.92\%$, στην κλάση 4 με $UA4 = 29.11\%$ και στην κλάση 5 με $UA5 = 47.11\%$.

Με βάση τα παραπάνω συμπεράσματα και το Producer’s Accuracy, το μοντέλο προβλέπει με ικανοποιητική ακρίβεια τις κλάσεις 1 και 3, με μέτρια ακρίβεια την κλάση 4, και με κακή ακρίβεια τις κλάσεις 2 και 5.

Με βάση το User’s Accuracy το μοντέλο προβλέπει σε ικανοποιητικό βαθμό την κλάση 1 και σε μέτριο την κλάση 5.

Η σημαντικότερη κλάση είναι η 1, γιατί είναι αυτή που περιλαμβάνει τα θετικά κρούσματα σε επιληπτική κρίση, ενώ στις υπόλοιπες τέσσερις κλάσεις τα δείγματα είναι αρνητικά. Η ακρίβεια της κλάσης 1 είναι πολύ υψηλή με πιθανότητες $PA1 = 72.78\%$ και $UA1 = 91.99\%$.

Το μοντέλο προβλέπει με μεγάλη ακρίβεια τα θετικά δείγματα (κλάση 1), επομένως λειτουργεί ικανοποιητικά.