

ΑΣΑΦΗ ΣΥΣΤΗΜΑΤΑ

ΕΡΓΑΣΙΑ 4
Classification

ΚΑΜΠΟΥΡΗΣ ΜΙΧΟΣ ΔΙΟΝΥΣΟΣ
7691

ΦΕΒΡΟΥΑΡΙΟΣ 2024

1. Εφαρμογή σε απλό Dataset

Ξεκινάμε χωρίζοντας τα δεδομένα σε τρία υποσύνολα εκπαίδευσης-επικύρωσης-ελέγχου με τη διαφορά ότι στα classification προβλήματα προσαρμόζουμε το κώδικα ώστε σε κάθε σύνολο να βρίσκονται ισοπίθानα δεδομένα από κάθε κλάση.

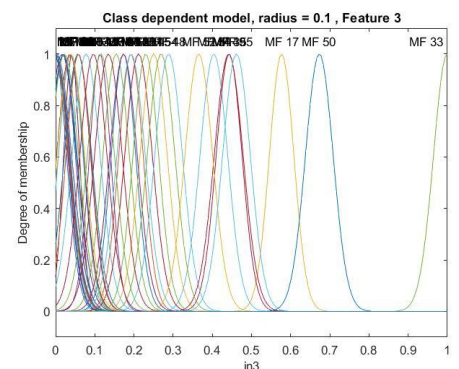
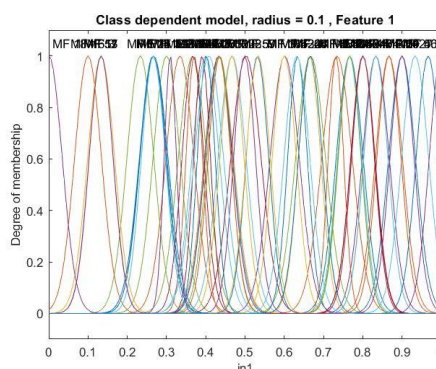
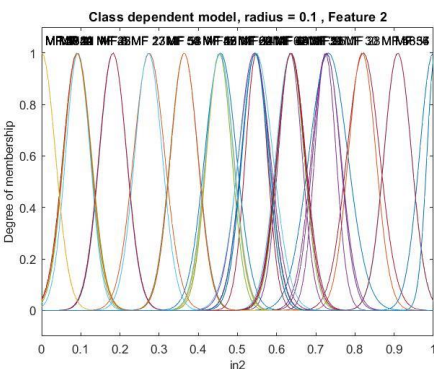
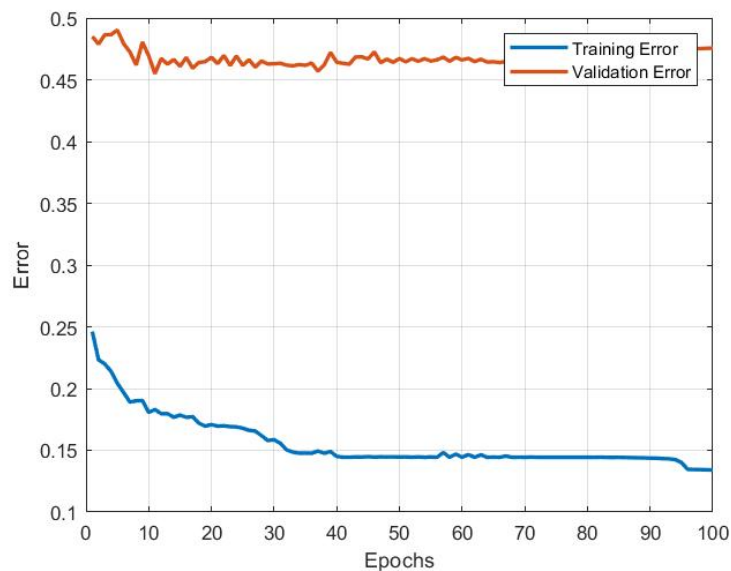
Ο διαχωρισμός και η κανονικοποίηση γίνεται με το αρχείο `split.m`

Τέλος εξετάζουμε τις δύο ελεύθερες μεταβλητές, την ακτίνα επιρροής και το σε ποια δεδομένα εκτελείτε το subtractive clustering (class dependent και class independent). Για την ακτίνα επιρροής επιλέγουμε δύο ακραίες τιμές (0.1 και 0.9) ώστε να έχουμε μεγάλη διαφοροποίηση στον αριθμό κανόνων. Έτσι έχουμε τα τέσσερα μοντέλα τα οποία εκπαιδεύουμε για 100 epochs το καθένα.

Ο κώδικας της εκπαίδευσης βρίσκεται στο αρχείο `main_simple.m`

Μοντέλο class dependent με $r_a=0.1$

Στα παρακάτω διαγράμματα βλέπουμε την καμπύλη μάθησης και τις τελικές μορφές των ασαφών συνόλων:



Πίνακας σφαλμάτων ταξινόμησης:

Predicted/true	Class 1	Class 2
Class 1	36	12
Class 2	6	7

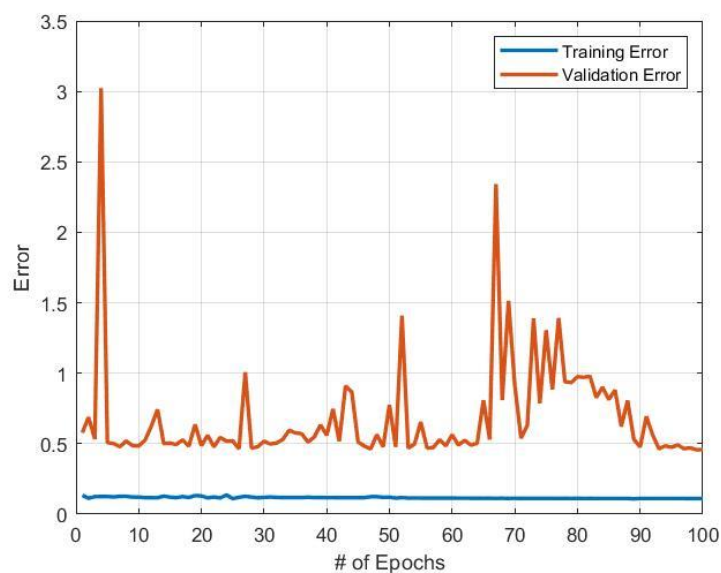
Δείκτες απόδοσης:

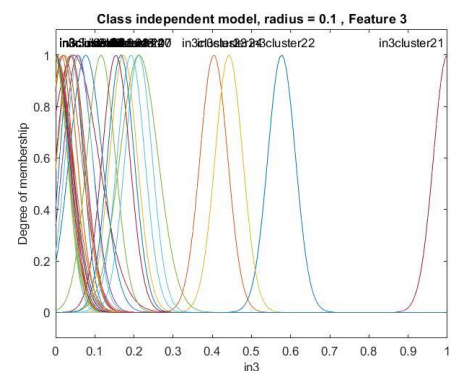
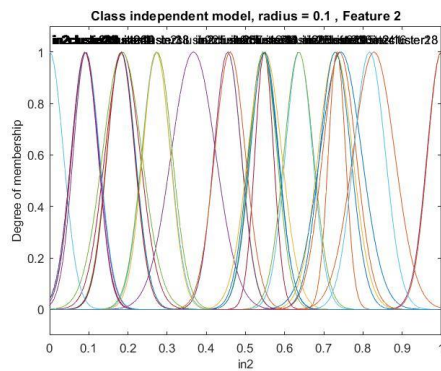
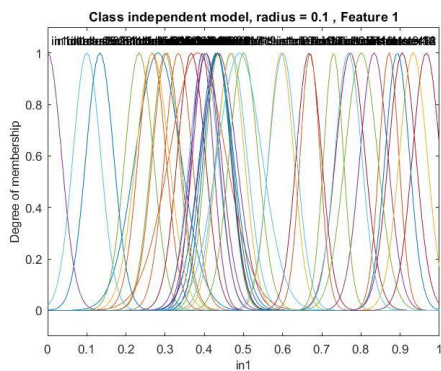
OA	0.7049
PA class 1	0.7500
PA class 2	0.5385
UA class 1	0.8571
UA class 2	0.3684
k	0.2469

Αριθμός κανόνων: 64

Μοντέλο class dependent με $r_a=0.9$

Στα παρακάτω διαγράμματα βλέπουμε την καμπύλη μάθησης και τις τελικές μορφές των ασαφών συνόλων:





Πίνακας σφαλμάτων ταξινόμησης:

Predicted/true	Class 1	Class 2
Class 1	48	0
Class 2	10	3

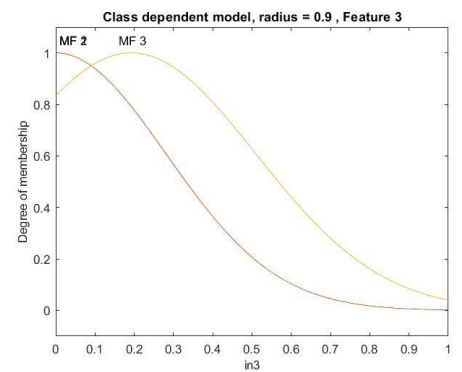
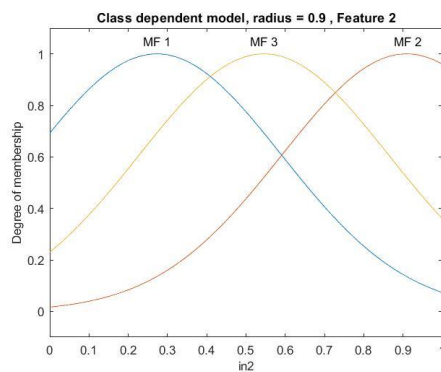
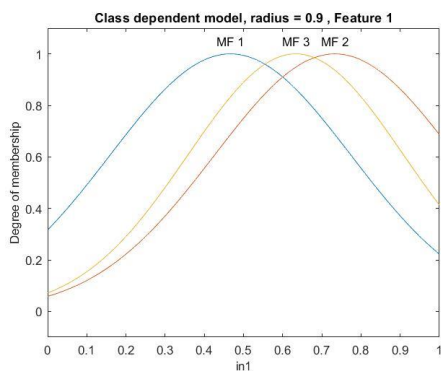
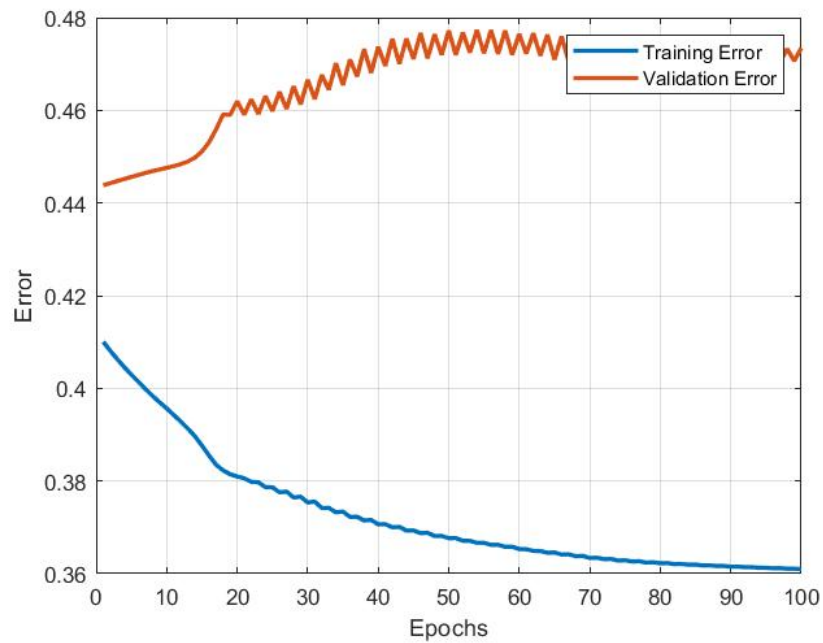
Δείκτες απόδοσης:

OA	0.8361
PA class 1	1
PA class 2	0.2308
UA class 1	0.8276
UA class 2	1
k	0.3207

Αριθμός κανόνων: 43

Μοντέλο class independent με $r_a=0.1$

Στα παρακάτω διαγράμματα βλέπουμε την καμπύλη μάθησης και τις τελικές μορφές των ασαφών συνόλων:



Πίνακας σφαλμάτων ταξινόμησης:

Predicted/true	Class 1	Class 2
Class 1	37	11
Class 2	7	6

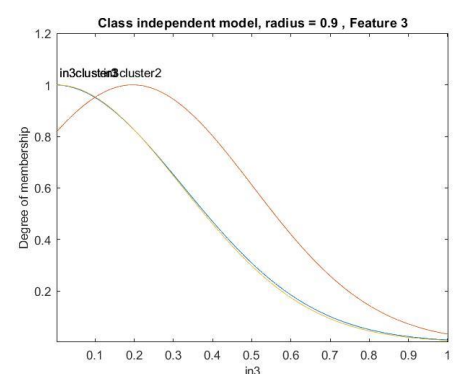
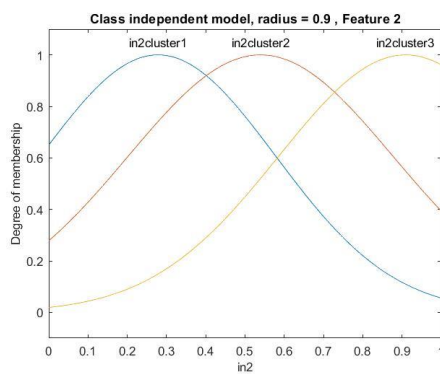
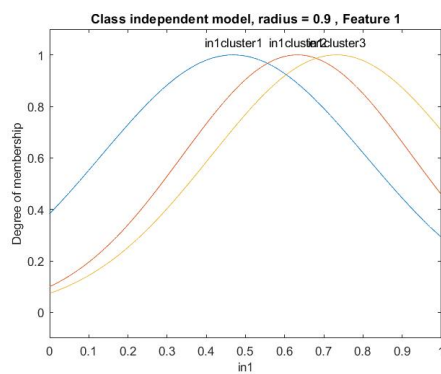
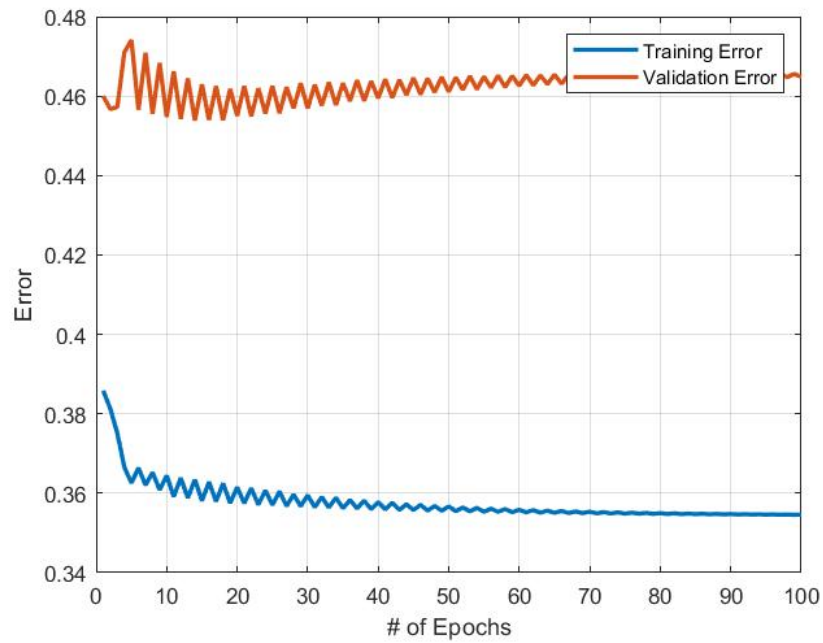
Δείκτες απόδοσης:

OA	0.7049
PA class 1	0.7708
PA class 2	0.4615
UA class 1	0.8409
UA class 2	0.3529
k	0.2089

Αριθμός κανόνων: 3

Μοντέλο class independent με $r_a=0.9$

Στα παρακάτω διαγράμματα βλέπουμε την καμπύλη μάθησης και τις τελικές μορφές των ασαφών συνόλων:



Πίνακας σφαλμάτων ταξινόμησης:

Predicted/true	Class 1	Class 2
Class 1	44	4
Class 2	9	4

Δείκτες απόδοσης:

OA	0.7869
PA class 1	0.9167
PA class 2	0.3077
UA class 1	0.8302
UA class 2	0.5000
k	0.2610

Αριθμός κανόνων: 3

Συμπεράσματα

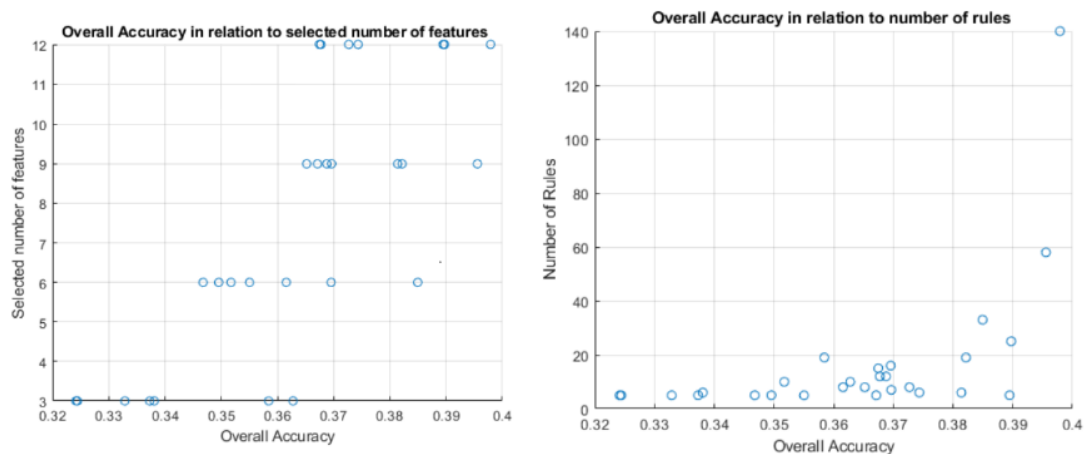
Μπορούμε με βεβαιότητα να πούμε ότι το class dependent μοντέλο με $r_a=0.9$ είναι το καλύτερο με overall accuracy 83.61% και ακολουθεί το class independent μοντέλο με $r_a=0.9$ με overall accuracy 78.69% από όπου καταλαβαίνουμε την σημαντικότητα της ακτίνας επιρροής. Επίσης παρατηρούμε ότι όλα τα μοντέλα εμφανίζουν τις χειρότερες μετρικές στην δεύτερη κλάση γεγονός που μας οδηγεί στο συμπέρασμα ότι τα μοντέλα μας δεν έχουν εκπαιδευτεί καλά για την δεύτερη κλάση το οποίο είναι λογικό καθώς τα δεδομένα της δεύτερης κλάσης είναι πολύ λιγότερα από αυτά της πρώτης. Τέλος όσον αφορά την επίδραση του αριθμού κανόνων στην απόδοση του ταξινομητή μπορούμε να πούμε ότι το μεγάλο πλήθος κανόνων έχει αρνητική επίδραση στην απόδοση όπως βλέπουμε στα μοντέλα 1 και 3 όπου παρατηρείτε και επικάλυψη των προβολών των ασαφών συνόλων κάθε cluster αυτό θα μπορούσε να βελτιωθεί αυξάνοντας την ακτίνα επιρροής και άρα μειώνοντας τον αριθμό των κανόνων.

2. Εφαρμογή σε dataset με υψηλή διαστασιμότητα

Στη δεύτερη φάση της εργασίας μας δίνεται ένα dataset πολύ μεγαλύτερων διαστάσεων από το πρώτο, δηλαδή με πολύ περισσότερα δείγματα τα οποία έχουν το καθένα πολύ περισσότερα χαρακτηριστικά. Όπως και στο πρώτο μέρος γίνεται διαχωρισμός του αρχικού συνόλου δεδομένων σε τρία υποσύνολα εκπαίδευσης-επικύρωσης-ελέγχου. Στη συνέχεια και εδώ έγκειται η διαφορά με το πρώτο μέρος της εργασίας θα πρέπει να διαλέξουμε τιμές για δύο παραμέτρους α) τον αριθμό των χαρακτηριστικών που θα χρησιμοποιήσουμε στην εκπαίδευση του μοντέλου και β) την ακτίνα επιρροής των cluster.

Για την επιλογή τους θα ακολουθήσουμε τη μέθοδο της αναζήτησης πλέγματος και για η αξιολόγησή τους θα γίνει με τη διαδικασία στη διασταυρωμένης επικύρωσης την οποία θα επαναλάβουμε πέντε φορές για κάθε σημείο του πλέγματος και στο τέλος θα διαλέξουμε το σημείο, δηλαδή τις τιμές των παραμέτρων, που μας έδωσε το μικρότερο μέσο σφάλμα.

Ο κώδικας για αυτή τη διαδικασία βρίσκεται στο αρχείο paramset.m

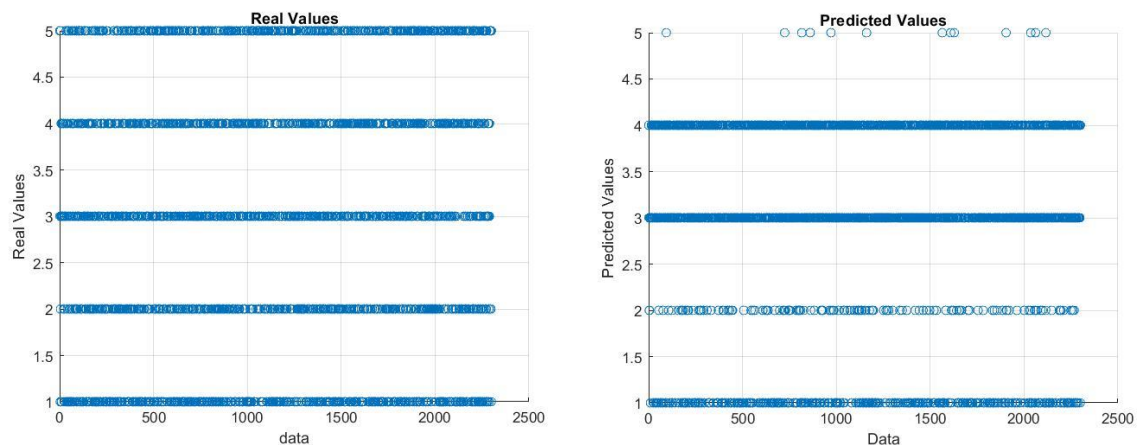


Στα διαγράμματα παρατηρούμε ότι πετυχαίνουμε καλύτερο OA για τον μεγαλύτερο αριθμό χαρακτηριστικών, ενώ όσον αφορά τον αριθμό των κανόνων βλέπουμε ότι το OA αυξάνεται όσο μεγαλώνει και ο αριθμός τους δηλαδή μικραίνει η ακτίνα επιρροής.

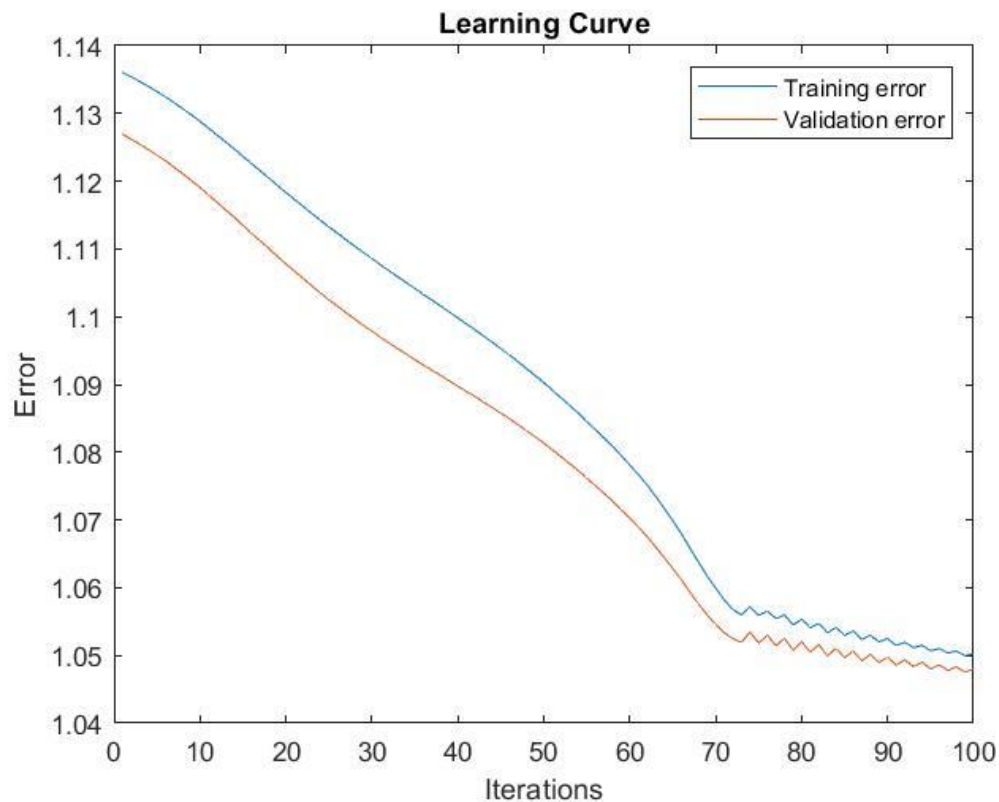
Εκπαίδευση τελικού μοντέλου (num.Features=12 , $r_a=0.8$)

Ο κώδικας εκπαίδευσης του τελικού μοντέλου βρίσκεται στο αρχείο `main_high.m`

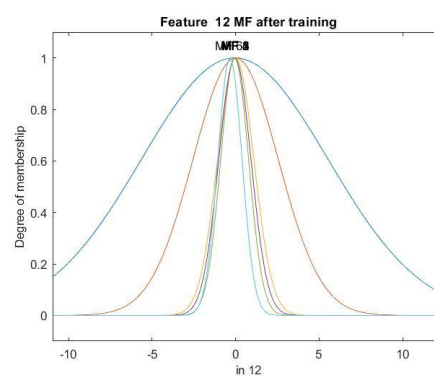
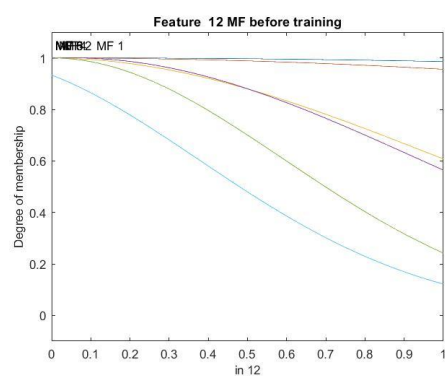
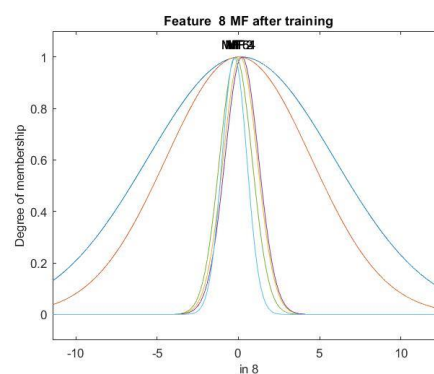
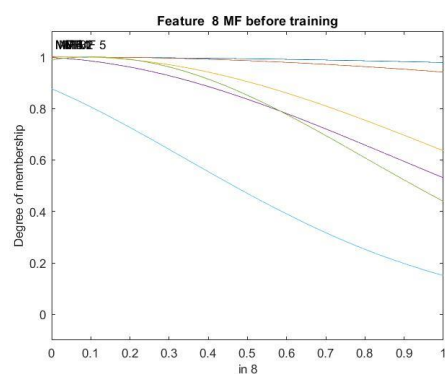
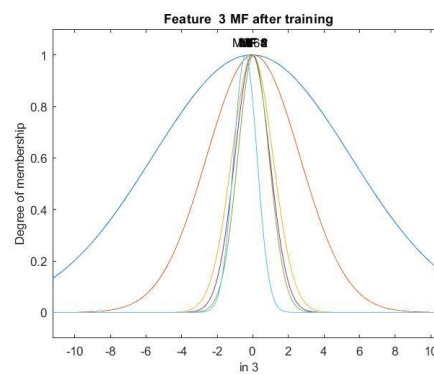
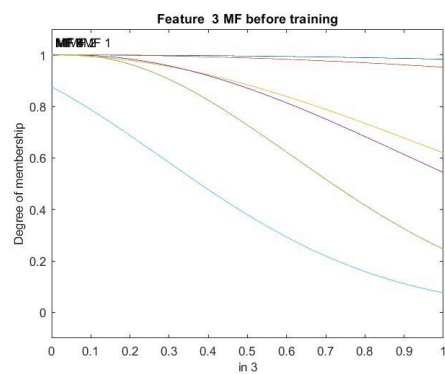
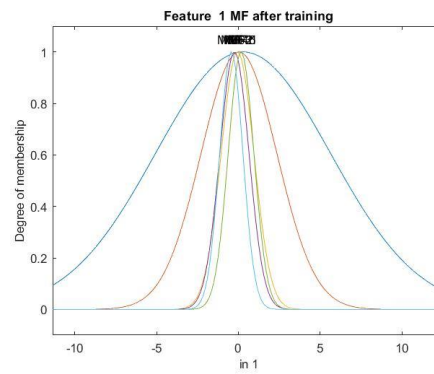
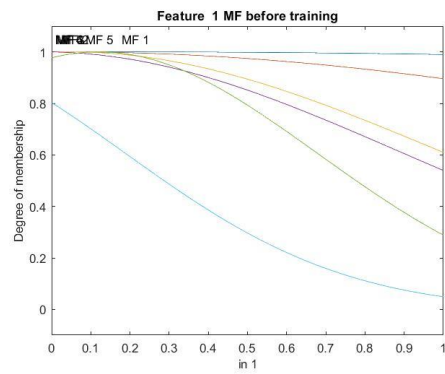
Στα παρακάτω διαγράμματα έχουμε μια πρώτη εικόνα για τις τιμές που προβλέψαμε με το μοντέλο μας σε σχέση με τις πραγματικές τιμές. Από όπου μπορούμε να συμπεράνουμε ότι σίγουρα το μεγαλύτερο πρόβλημα εντοπίζεται στις κλάσεις 2 και 5.



Καμπύλη μάθησης:



Επίσης παραθέτουμε μερικές από τις καμπύλες των ασαφών συνόλων πριν (αριστερά) και μετά (δεξιά) την εκπαίδευση.



Πίνακας σφαλμάτων ταξινόμησης:

	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	355	50	35	18	2
Class 2	44	22	252	139	3
Class 3	8	22	290	139	11
Class 4	6	28	177	237	12
Class 5	0	0	182	278	0

Ο πίνακας σφαλμάτων ταξινόμησης επιβεβαιώνει τις αρχικές μας παρατηρήσεις για τα προβλήματα του μοντέλου μας και βλέπουμε όντως ότι το μεγαλύτερο πρόβλημα είναι στις κλάσεις 2 και 5 και ειδικότερα στην κλάση 5 όπου δεν έχουμε ούτε μια σωστή αντιστοίχιση δείγματος. Αντίθετα βλέπουμε ότι ο ταξινομητής λειτουργεί πολύ καλά για την πρώτη κλάση όπου τα περισσότερα δείγματα που ανήκουν σε αυτήν ταξινομούνται σωστά ενώ από τις υπόλοιπες κλάσεις ελάχιστα δείγματα ταξινομούνται εσφαλμένα στη πρώτη κλάση.