

## **Εργασία 4 – Classification**

---

### **Αναφορά**

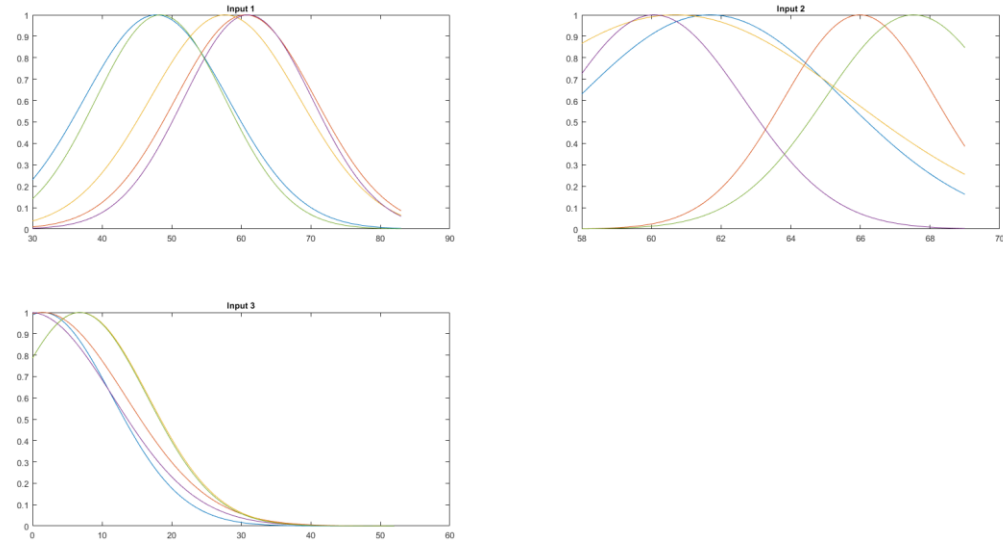
Ασαφή Συστήματα / Υπολογιστική Νοημοσύνη

Άρης Ελευθέριος Παπαγγέλης, [ariselefp@ece.auth.gr](mailto:ariselefp@ece.auth.gr)  
ΑΕΜ: 8883

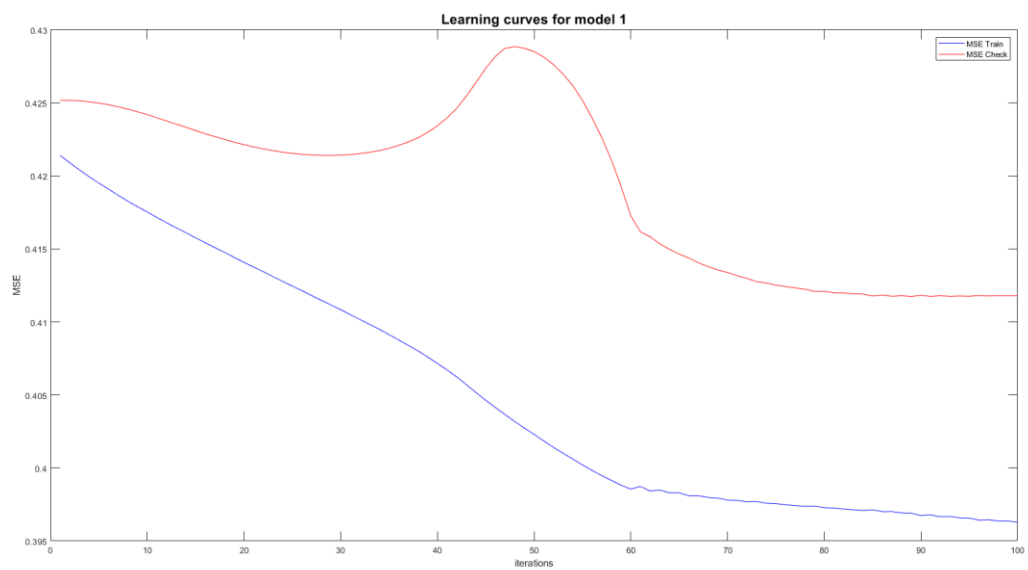
## Μέρος Α

Μοντέλο 1 (radius 0.6, 5 κανόνες, class independent clustering)

Συναρτήσεις συμμετοχής



Διάγραμμα μάθησης



## Μετρικές αξιολόγησης

Πίνακας σφαλμάτων

40	12
2	7

$$OA = 0.7705$$

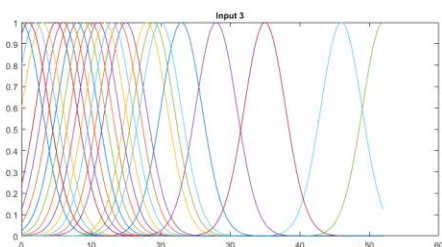
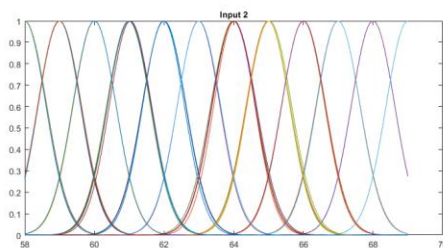
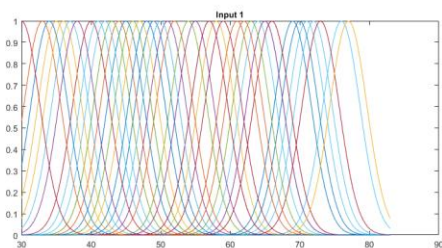
$$PA = [0.9524, \quad 0.3684]$$

$$UA = [0.7692, \quad 0.7778]$$

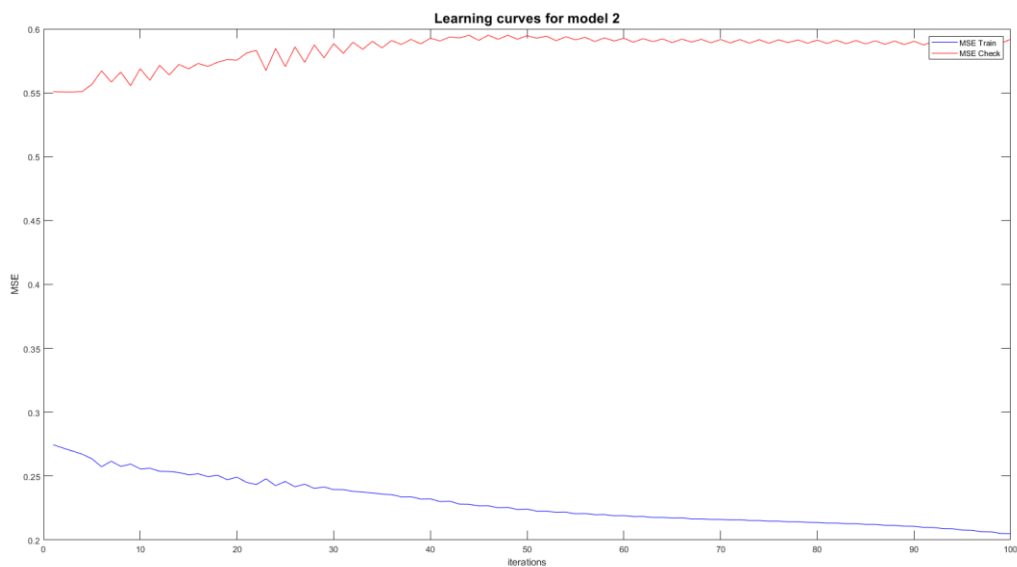
$$k = 0.3748$$

Μοντέλο 2 (radius 0.16, 78 κανόνες, class independent clustering)

## Συναρτήσεις συμμετοχής



## Διάγραμμα μάθησης



## Μετρικές αξιολόγησης

Πίνακας σφαλμάτων

35	14
7	5

$$OA = 0.6557$$

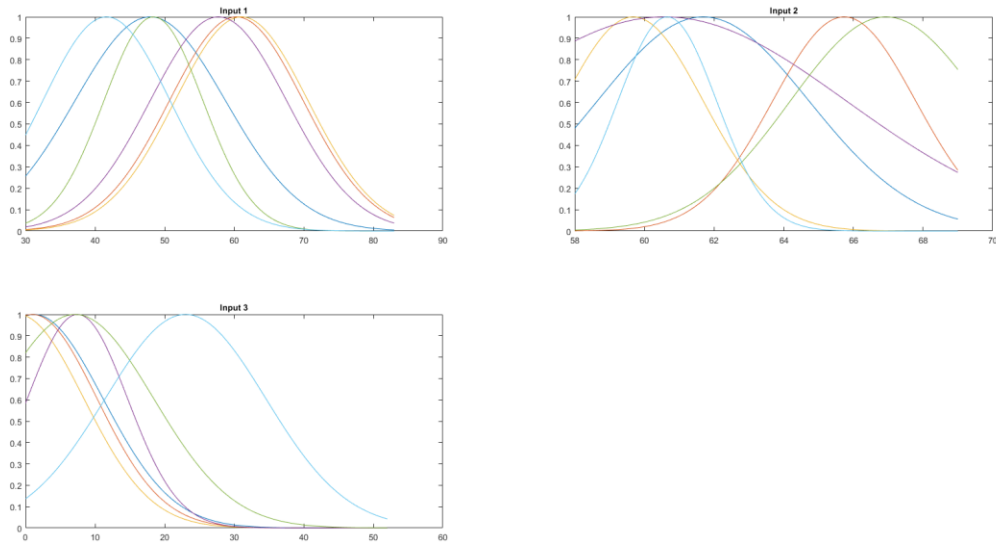
$$PA = [0.8333, \quad 0.2632]$$

$$UA = [0.7143, \quad 0.4167]$$

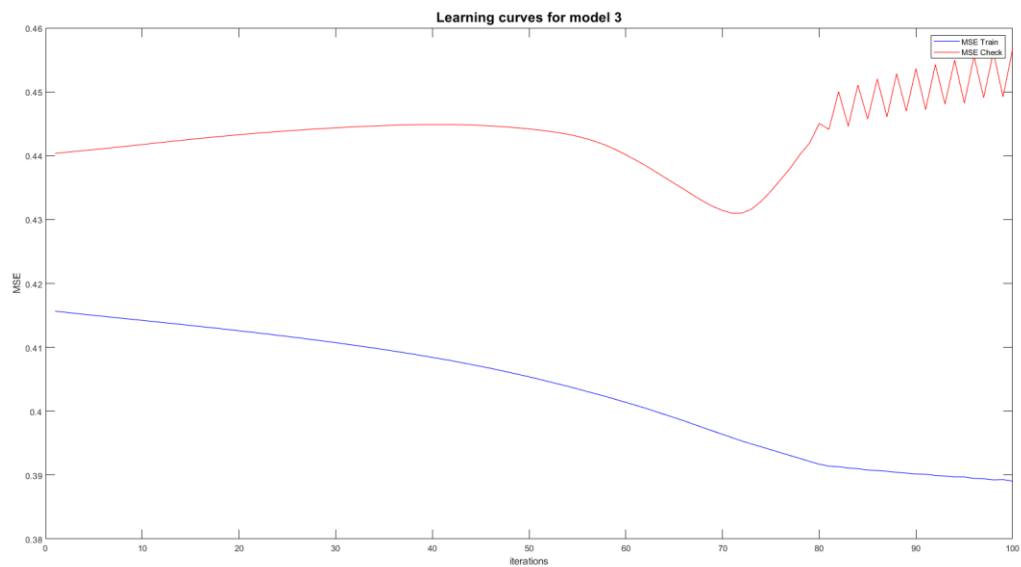
$$k = 0.1073$$

## Μοντέλο 3 (radius 0.6, 6 κανόνες, class dependent clustering)

### Συναρτήσεις συμμετοχής



### Διάγραμμα μάθησης



## Μετρικές αξιολόγησης

Πίνακας σφαλμάτων

41	13
1	6

$$OA = 0.7705$$

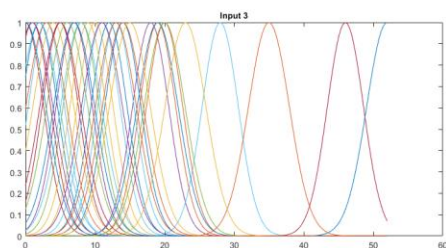
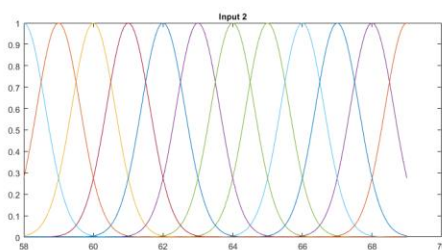
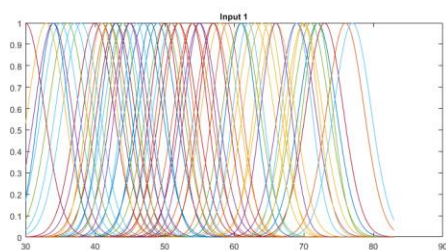
$$PA = [0.9762, \quad 0.3158]$$

$$UA = [0.7593, \quad 0.8571]$$

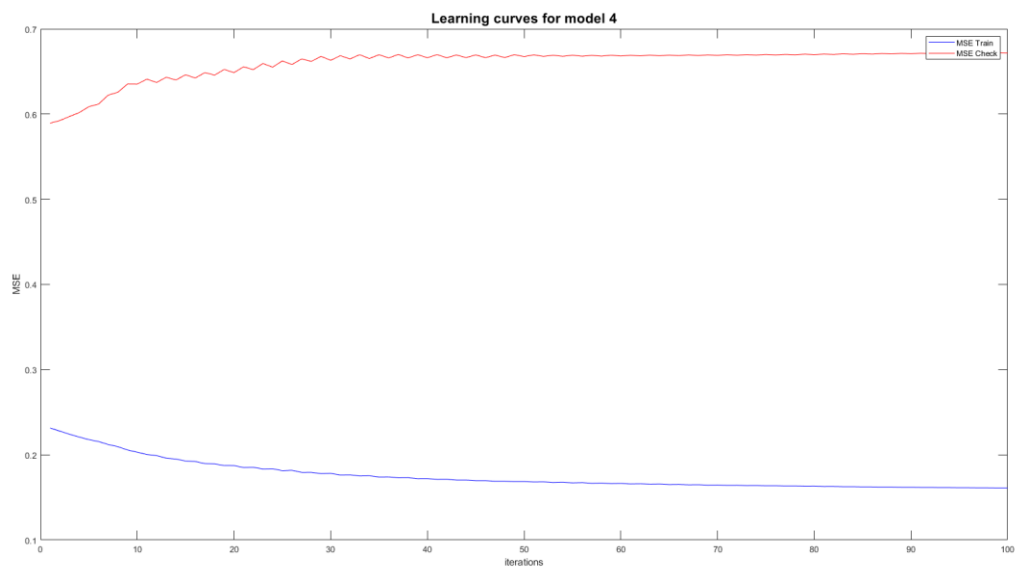
$$k = 0.3530$$

Μοντέλο 4 (radius 0.16, 86 κανόνες, class dependent clustering)

## Συναρτήσεις συμμετοχής



## Διάγραμμα μάθησης



## Μετρικές αξιολόγησης

Πίνακας σφαλμάτων

38	14
4	5

$$OA = 0.7049$$

$$PA = [0.9048, \quad 0.2632]$$

$$UA = [0.7308, \quad 0.5556]$$

$$k = 0.1962$$

## Παρατηρήσεις

Γενικά, η εκπαίδευση των μοντέλων έγινε πολλές φορές για να μπορέσουν να εξαχθούν ασφαλή συμπεράσματα για το πως επηρεάζουν τα αποτελέσματα οι διάφορες παράμετροι της εκπαίδευσης, με τα αποτελέσματα αναμενόμενα να διαφέρουν κάθε φορά ανάλογα με τη διαμέριση του dataset στα υποσύνολα εκπαίδευσης, επικύρωσης και ελέγχου. Επιλέχθηκαν τα παραπάνω αποτελέσματα ως αντιπροσωπευτικά, από όλες τις φορές που έγινε εκτέλεση του κώδικα.

Συνολικά παρατηρείται μέτρια επίδοση για όλα τα μοντέλα που δοκιμάστηκαν, με το ποσοστό ακρίβειας (OA) να κυμαίνεται από 65% για το χειρότερο έως 77% για το καλύτερο μοντέλο. Σίγουρα για αυτό ευθύνεται τόσο η μέθοδος του subtractive clustering που χρησιμοποιήθηκε, αλλά κυρίως το μικρό μέγεθος του dataset, ειδικά στην κλάση 2 που είχε πολύ λιγότερα δείγματα από την κλάση 1. Και πράγματι, από τα παραπάνω αποτελέσματα, παρατηρείται πως για όλα τα μοντέλα είναι πολύ πιο δύσκολη η σωστή ταξινόμηση δειγμάτων στην κλάση 2 από ότι στην κλάση 1, ακόμα και εξασφαλίζοντας πως η συχνότητα εμφάνισης της κλάσης 2 ήταν η ίδια και για τα τρία υποσύνολα.

Για μικρότερο μέγεθος ακτίνας, υπάρχει μεγαλύτερος αριθμός κανόνων. Ως προς την επίδραση του αριθμού των κανόνων, παρατηρείται πως, στο συγκεκριμένο dataset, έχουμε καλύτερα αποτελέσματα για μικρότερο αριθμό κανόνων, δηλαδή για μεγάλες τιμές της ακτίνας. Αντίθετα, τα πολυπλοκότερα μοντέλα των πολλών κανόνων παρουσιάζουν χειρότερα αποτελέσματα. Βέβαια, η απόδοση μεταξύ των διαφορετικών μοντέλων είναι σχετικά κοντά, αλλά αυτό που αναφέρεται παραπάνω ίσχυε γενικά για κάθε εκτέλεση του κώδικα, οπότε μπορεί να χαρακτηριστεί ως ασφαλές συμπέρασμα.

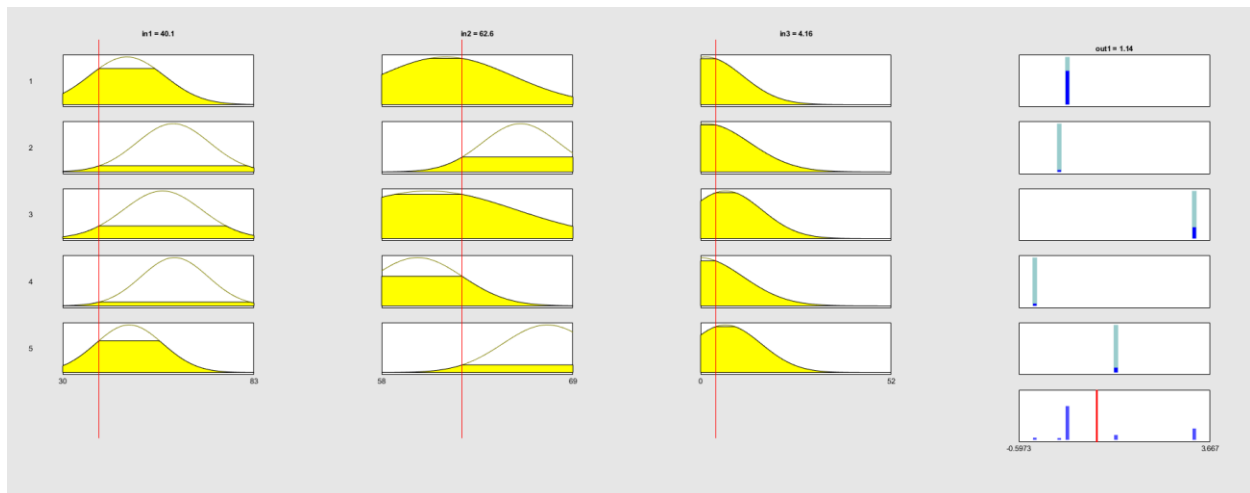
Επιπλέον, σε γενικές γραμμές η εφαρμογή class dependent clustering συνήθως επέφερε μια ελαφρά βελτίωση στα αποτελέσματα για την ίδια τιμή ακτίνας, σε σχέση με την αντίστοιχη εφαρμογή class independent clustering (χωρίς βέβαια αυτό να ισχύει αυστηρά για κάθε διαφορετική εκτέλεση του κώδικα, όπως ίσχυε για το παραπάνω συμπέρασμα για τον αριθμό κανόνων). Η μικρή έως καθόλου βελτίωση οφείλεται πιθανόν στο γεγονός πως υπάρχουν μόνο 2 κλάσεις στο συγκεκριμένο dataset. Για περισσότερες κλάσεις θα εμφανιζόταν μεγαλύτερη βελτίωση.

Ενδεχομένως με μεγαλύτερο σύνολο δεδομένων να μπορούσε να γίνει μια πιο ολοκληρωμένη αξιολόγηση. Με βάση τα διαθέσιμα δεδομένα ελέγχου, τα μοντέλα κατατάσσονται σε σειρά από το καλύτερο στο χειρότερο ως εξής:

Μοντέλο 1 = Μοντέλο 3 -> Μοντέλο 4 -> Μοντέλο 2

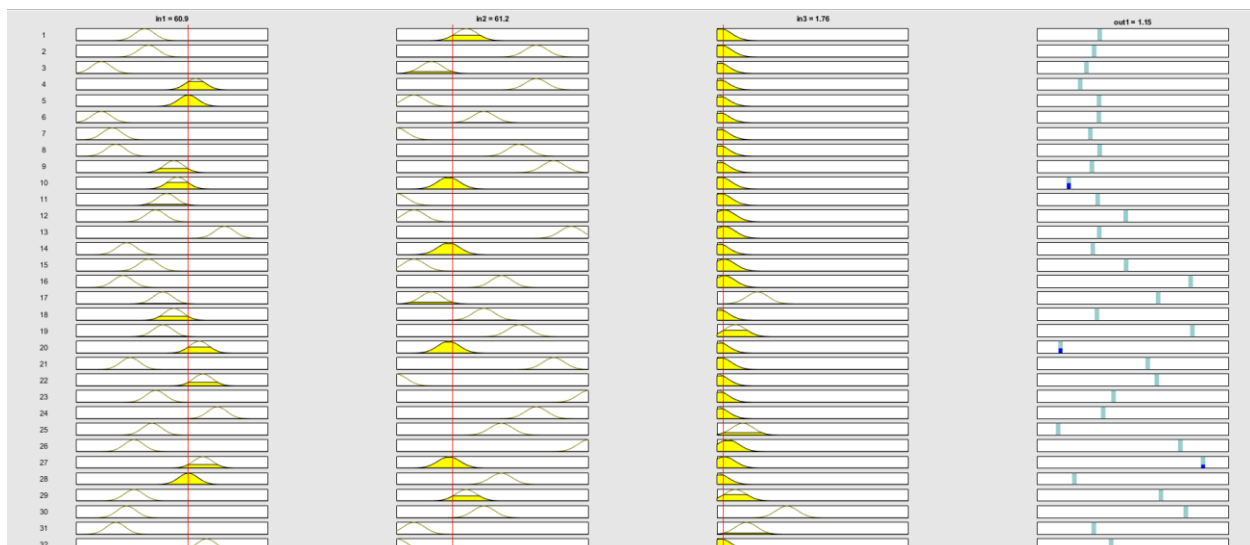


Ως προς τους κανόνες και την μεταξύ τους επικάλυψη παίρνουμε ως παράδειγμα τους κανόνες που διαμορφώνονται από το μοντέλο 1 των 5 κανόνων:



Παρατηρείται επικάλυψη μεταξύ των προβολών των συνόλων εισόδου για τους διαφορετικούς κανόνες, πράγμα που διαπιστώνεται σε μεγαλύτερο βαθμό για την είσοδο 3, στην οποία οι προβολές είναι αρκετά όμοιες για όλους τους κανόνες. Αυτό σημαίνει ότι μικρή μεταβολή στην είσοδο 3 δεν θα αλλάξει σημαντικά την έξοδο, πράγμα το οποίο δεν ισχύει για άλλες εισόδους, όπως για παράδειγμα την 2.

Το φαινόμενο αυτό παρατηρείται και για τα υπόλοιπα σύνολα κανόνων. Παρατίθεται ένα ακόμη παράδειγμα, για το μοντέλο 2 των 78 κανόνων, όπου και πάλι εμφανίζεται επικάλυψη, ιδιαίτερα για την είσοδο 3 (φαίνονται οι πρώτοι 32 κανόνες):



## Μέρος Β

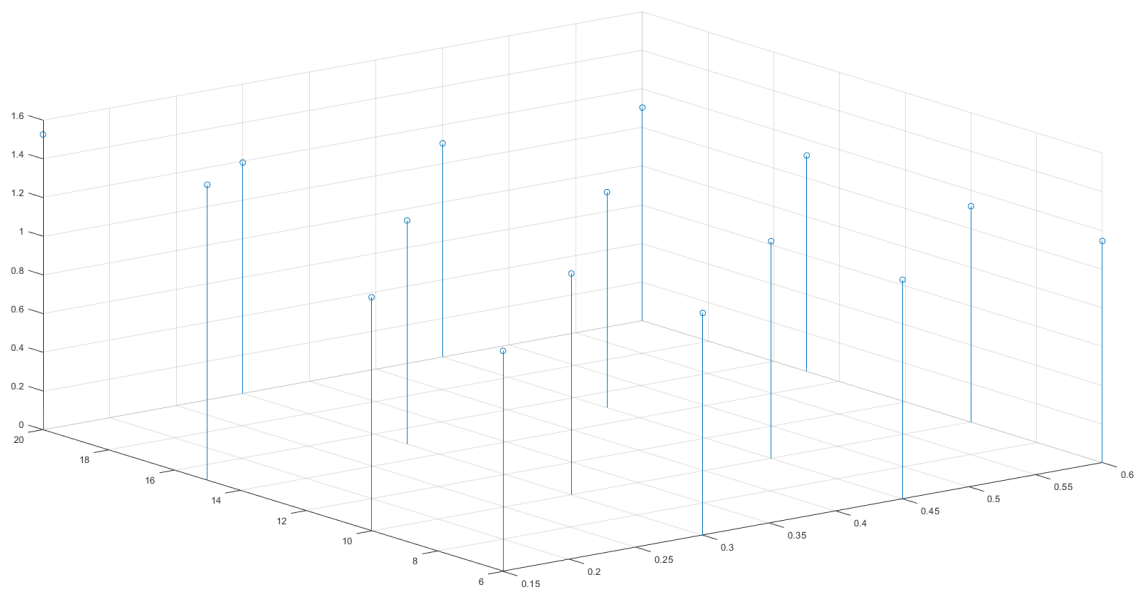
Τα σύνολα τα οποία επιλέχθηκαν, είναι τα εξής:

$$N_F = [6, 10, 15, 20]$$

$$C_R = [0.15, 0.3, 0.45, 0.6]$$

Να σημειωθεί εδώ ότι εξαιτίας του γεγονότος ότι ως μέθοδος διαμέρισης χρησιμοποιήθηκε η *genfis2*, σε κάθε επανάληψη του αλγορίθμου ο τελικός αριθμός κανόνων που προκύπτει είναι διαφορετικός.

### Διάγραμμα σφαλμάτων



NF \ CR	0.15	0.3	0.45	0.6
6	<b>1.1048</b>	1.1484	1.1327	1.1453
10	1.2078	1.1429	1.1226	1.1160
15	1.5278	1.1557	1.1148	1.1165
20	1.5257	1.1936	1.1403	<b>1.1027</b>

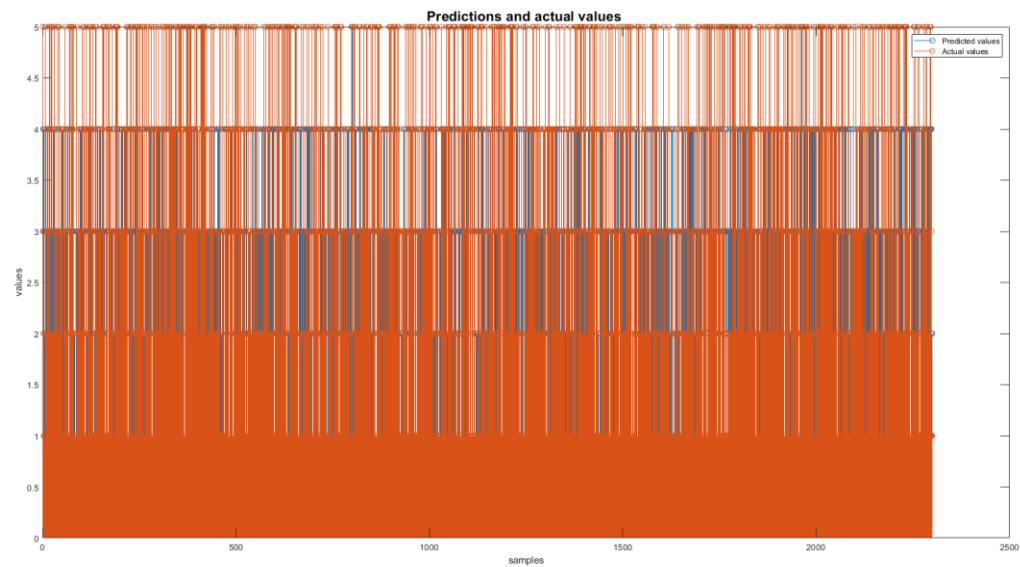
Να σημειωθεί πως, λόγω απαγορευτικά μεγάλου χρόνου εκτέλεσης, τα παραπάνω αποτελέσματα προέκυψαν από μοντέλα που εκπαιδεύτηκαν για 1 epoch, ώστε να βρεθεί το βέλτιστο που παρακάτω θα εκπαιδευθεί για περισσότερες εποχές.

Η καλύτερη επίδοση επιτυγχάνεται για ακτίνα 0.6 και 20 χαρακτηριστικά εισόδου, ενώ πολύ κοντά βρίσκεται και η επίδοση για ακτίνα 0.15 και 6 χαρακτηριστικά εισόδου. Επομένως, κρίνεται σκόπιμο να δοκιμάσουμε και τους δύο συνδυασμούς για το τελικό μοντέλο.

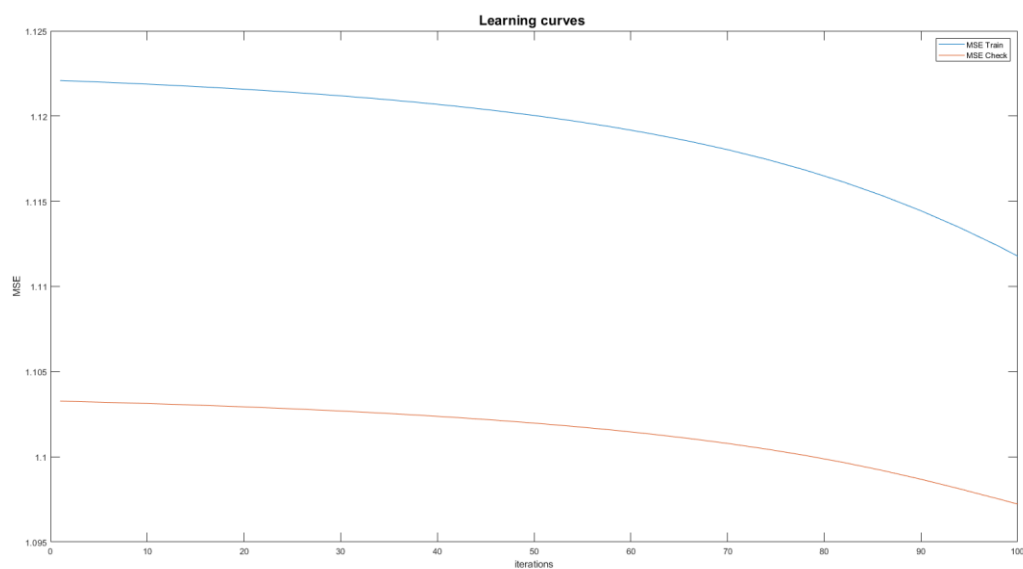
Δοκιμάζοντας και τους δύο συνδυασμούς, βρέθηκε πως για 20 χαρακτηριστικά και ακτίνα 0.6 επιτυγχάνεται  $OA = 0.3270$ , ενώ για 6 χαρακτηριστικά και ακτίνα 0.15 επιτυγχάνεται  $OA = 0.3678$ . Επομένως, αποφασίζεται να κρατηθεί ο δεύτερος συνδυασμός για τα παρακάτω αποτελέσματα.

Το τελικό μοντέλο δηλαδή περιλαμβάνει 6 εισόδους και 62 κανόνες. (διαφέρει ο αριθμός κανόνων απο εκτέλεση σε εκτέλεση, όπως αναφέρθηκε παραπάνω)

### Προβλέψεις και Πραγματικές τιμές



### Διάγραμμα μάθησης



## Μετρικές αξιολόγησης

Πίνακας σφαλμάτων

	Actual: Class 1	Actual: Class 2	Actual: Class 3	Actual: Class 4	Actual: Class 5
Predicted: Class 1	261	14	2	8	0
Predicted: Class 2	92	31	29	41	6
Predicted: Class 3	54	310	325	186	218
Predicted: Class 4	24	119	117	226	231
Predicted: Class 5	0	0	0	3	3

$$OA = 0.3678$$

$$PA = [0.6056, 0.0654, 0.6871, 0.4871, 0.0066]$$

$$UA = [0.9158, 0.1558, 0.2973, 0.3152, 0.5000]$$

$$k = 0.2076$$

Η καλύτερη επίδοση που μπόρεσε να επιτευχθεί ήταν  $OA = 0.3678$ .

Παρατηρούμε πως τα αποτελέσματα του βέλτιστου μοντέλου εξακολουθούν να είναι πολύ κακά, με το μοντέλο σχεδόν να αδυνατεί να ταξινομήσει σωστά στην κλάση 5, ενώ και για τις άλλες κλάσεις υπάρχουν πολλές λανθασμένες ταξινομήσεις. Επιπλέον, παρουσιάζει μια μεγάλη τάση ταξινόμησης δειγμάτων ως κλάσης 3 και σε μικρότερο βαθμό κλάσης 4, ακόμα και αν δεν είναι.

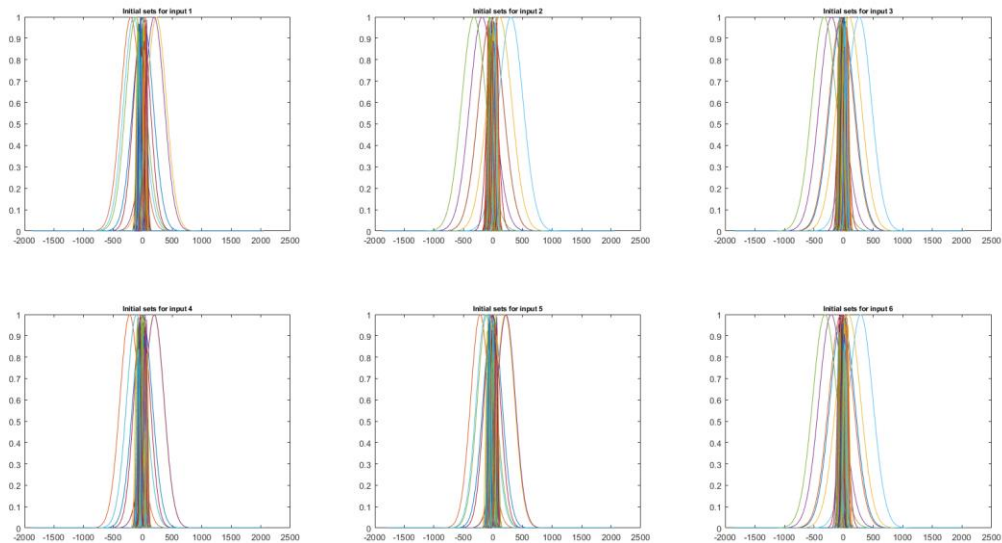
Αυτό οφείλεται στο πολύπλοκο dataset, για το οποίο, παρά την επιλογή των σημαντικότερων χαρακτηριστικών, δεν μπορεί να εκπαιδευτεί ικανό μοντέλο που να λειτουργεί με υψηλή ακρίβεια.

Πάντως, για την κλάση 1 παρουσιάζονται πολύ καλύτερα αποτελέσματα, και πράγματι αυτή είναι η σημαντικότερη κλάση που πρέπει να ταξινομείται σωστά, εφόσον η κλάση 1 είναι αυτή που δείχνει την καταγραφή επιληπτικής κρίσης στο EEG. Άρα, το μοντέλο εκπληρώνει το σκοπό του έως ένα βαθμό.

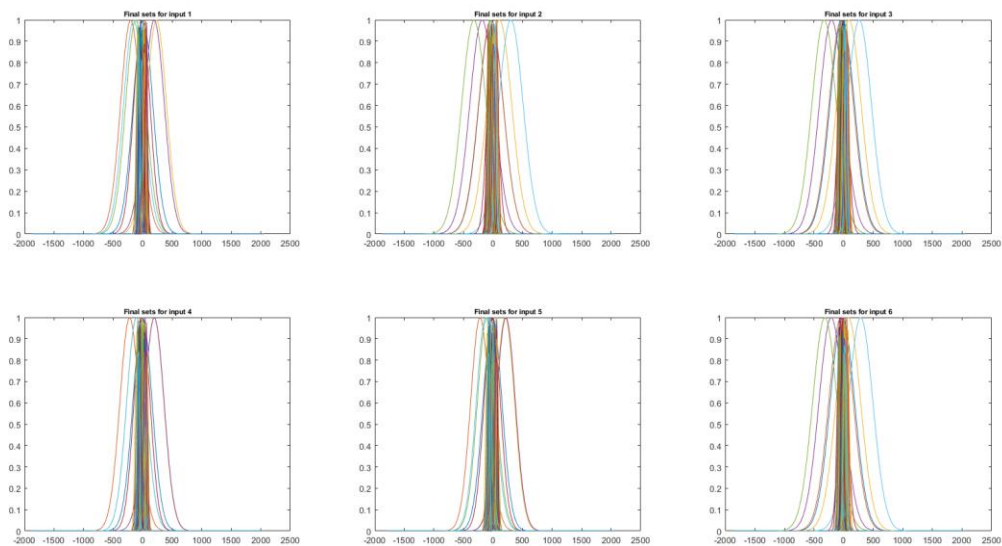
Εξάλλου, όπως αναφέρεται και στην περιγραφή του dataset στο UCI, αρκετοί ερευνητές κάνουν την απλοποίηση να θεωρούν δύο κλάσεις, την κλάση 1 για την καταγραφή επιληπτικής κρίσης, και την υπερκλάση  $\{2,3,4,5\}$  για γεγονότα άσχετα με την καταγραφή επιληπτικής κρίσης. Ενδεχομένως αν γινόταν και εδώ αυτή η απλοποίηση να επιτυγχάνονταν πολύ καλύτερα αποτελέσματα. \*\*\*

## Συναρτήσεις συμμετοχής

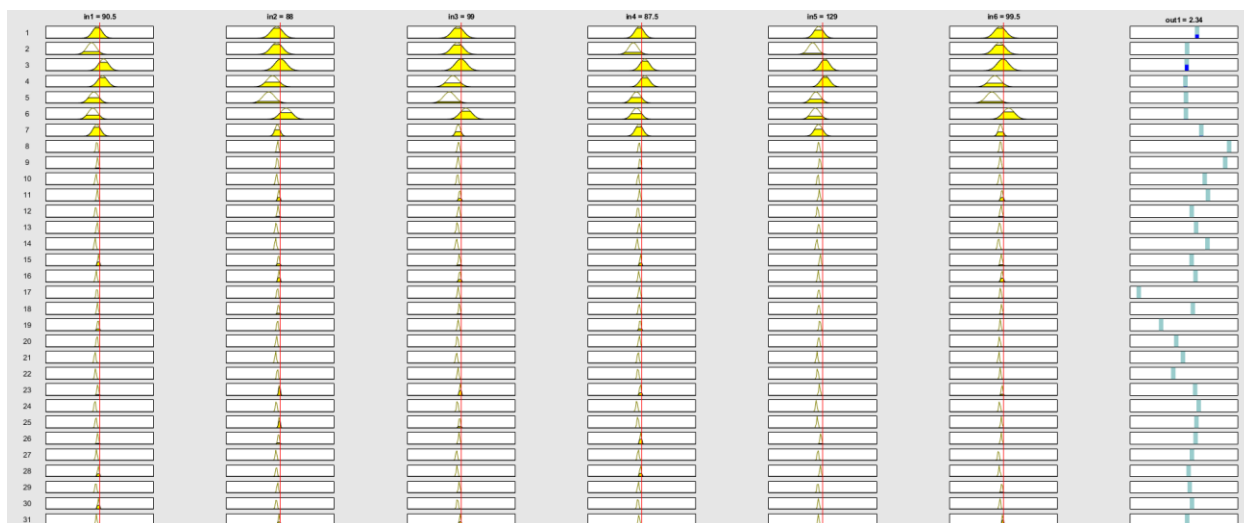
Μερικές από τις αρχικές συναρτήσεις συμμετοχής:



Μερικές από τις τελικές συναρτήσεις συμμετοχής:



## Επικάλυψη συνόλων



Αν και δεν υπάρχει το φαινόμενο που παρατηρήθηκε στο Μέρος Α, όπου μία είσοδος να έχει ουσιαστικά την ίδια προβολή για κάθε κανόνα που διαμορφώνει το μοντέλο, παρατηρείται επικάλυψη μεταξύ διαφορετικών κανόνων (φαίνονται οι πρώτοι 31 κανόνες, και οι υπόλοιποι συνεχίζουν στο ίδιο μοτίβο μετά τον κανόνα 31)

Για παράδειγμα, οι κανόνες 1 και 3, έχουν σχεδόν τις ίδιες προβολές ανά είσοδο, πράγμα που σημαίνει ότι ο ένας από τους 2 θα μπορούσε να παραληφθεί, μειώνοντας έτσι την πολυπλοκότητα του συστήματος και διατηρώντας την ίδια πληροφορία. Το ίδιο ισχύει και για όλους του κανόνες απο τον κανόνα 8 και κάτω, που όπως φαίνεται πολλοί από αυτούς έχουν σχεδόν ίδιες προβολές ανά είσοδο. Αυτοί οι κανόνες, μάλιστα, ενεργοποιούνται σε μια πολύ μικρή περιοχή στο κέντρο του πεδίου ορισμού των εισόδων, και φαίνεται από το σχήμα πως ή θα ικανοποιούνται σχεδόν όλοι, ή κανένας. Επομένως γενικά η συμπεριφορά του συστήματος διαμορφώνεται σε μεγαλύτερο βαθμό από τους πρώτους 7 κανόνες.

Κατά τα άλλα μπορεί επίσης να παρατηρηθεί το φαινόμενο της απενεργοποίησης της πλειοψηφίας των κανόνων, ανάλογα με την είσοδο, καθώς είναι σπάνιο να μην μηδενίζεται έστω μία εκ των εισόδων ενός κανόνα για την πλειοψηφία των κανόνων. Ειδικά για τιμές των εισόδων που βρίσκονται στα άκρα του πεδίου ορισμού τους, όλοι οι κανόνες μηδενίζονται.

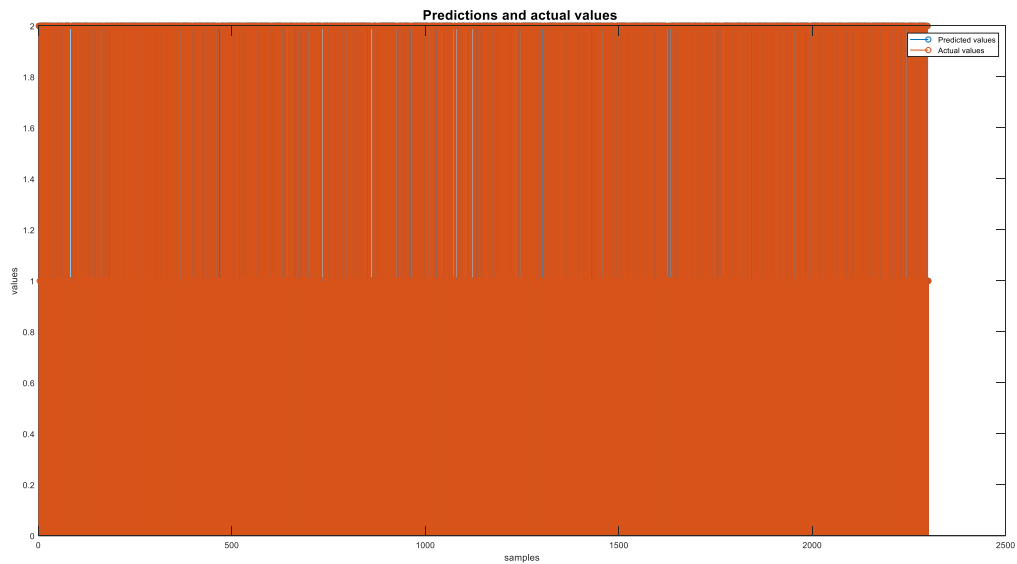
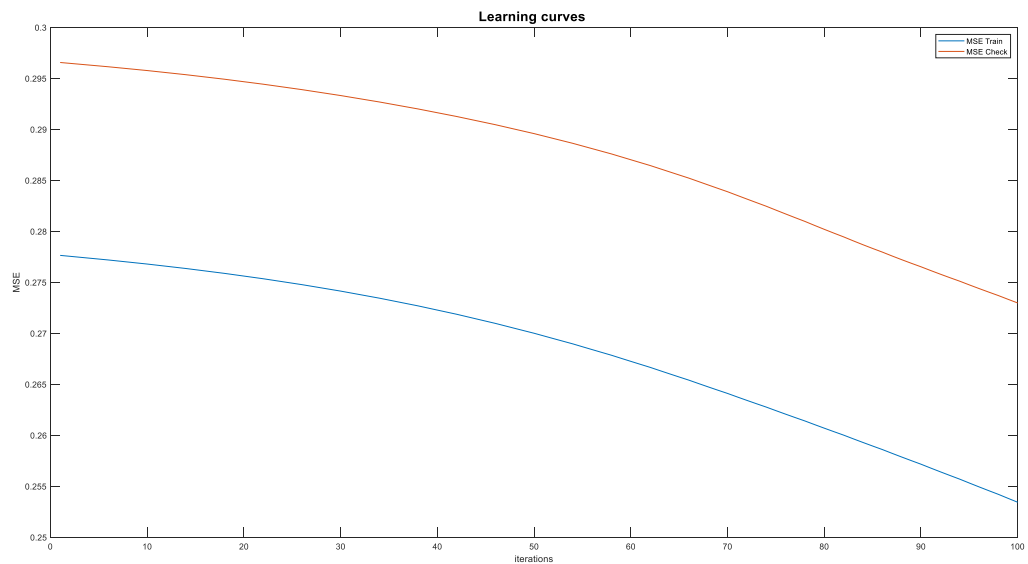
## Σχολιασμός

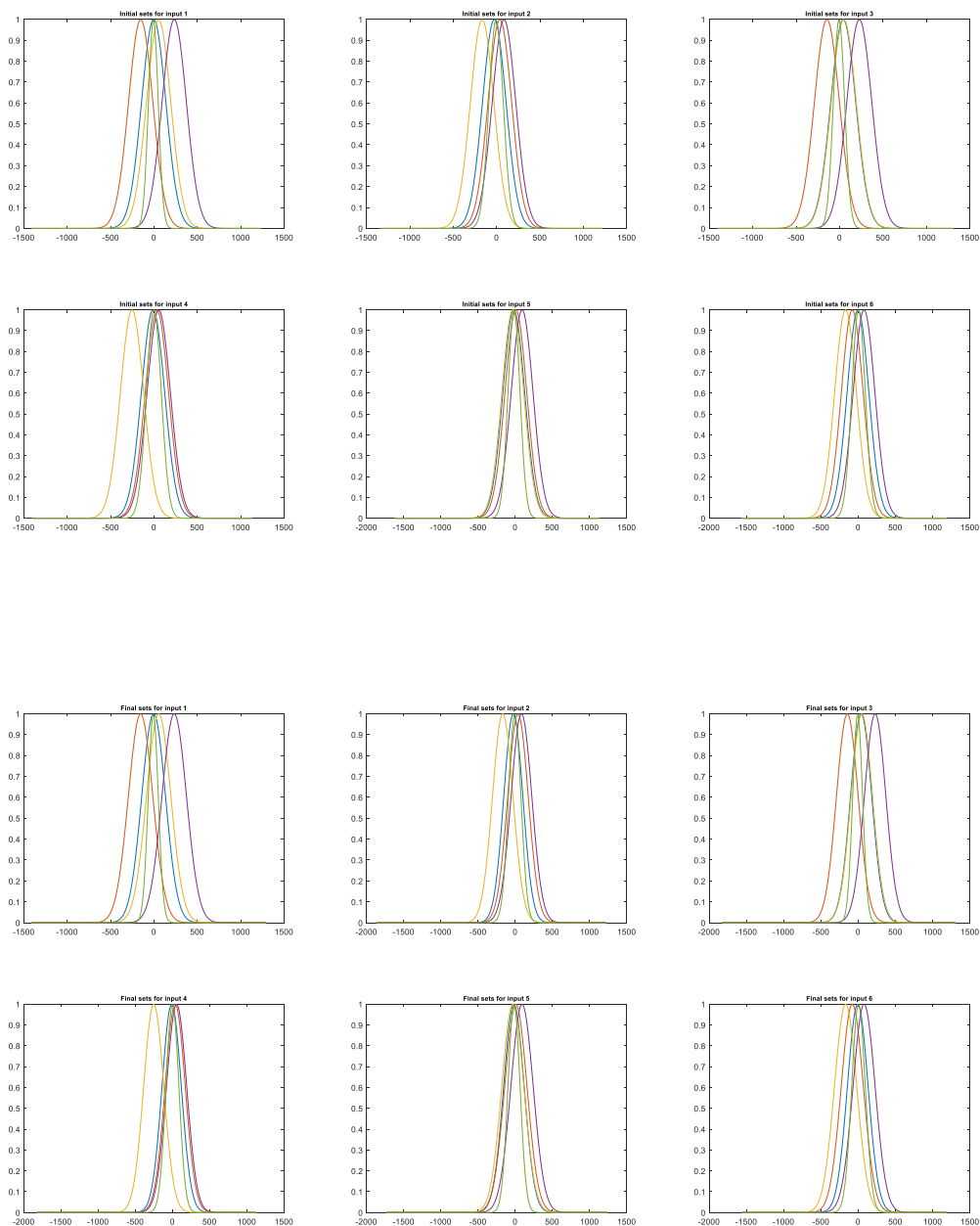
Συγκρίνοντας με την κλασική μέθοδο του grid partitioning, θα απαιτούνταν  $2^{178}$  ή  $3^{178}$  κανόνες, τη στιγμή που τώρα χρησιμοποιούνται μόλις 62, οπότε η χρήση του εν λόγω συστήματος εμφανίζεται ως ένας καλός συνδυασμός ακρίβειας/πολυπλοκότητας.

Γενικότερα, όσο περισσότερο διαμερίζεται ο χώρος σε νέα ασαφή σύνολα, τόσο μεγαλύτερη επικάλυψη θα υπάρχει στα νέα αυτά ασαφή σύνολα, πράγμα που θα έχει ως συνέπεια να ενεργοποιείται σημαντικά περισσότερος αριθμός κανόνων που όμως συνολικά δε θα προσφέρει σημαντική επιπλέον πληροφορία.

\*\*\*

Για προσωπική μου περιέργεια, παρότι δεν ζητείται στα πλαίσια της εργασίας, δοκίμασα να κάνω το binary classification που αναφέρεται παραπάνω, δηλαδή θεωρώντας την κλάση 1 και την υπερκλάση {2,3,4,5}. Κάνοντας αυτή την παραδοχή, πάλι για 6 χαρακτηριστικά και ακτίνα 0.15 για να είναι συγκρίσιμα τα αποτελέσματα, έχουμε τα παρακάτω:





	Actual: Class 1	Actual: Class {2,3,4,5}
Predicted: Class 1	342	74
Predicted: Class {2,3,4,5}	89	1795



$$OA = 0.9291$$

$$PA = [0.7935, 0.9604]$$

$$UA = [0.8221, 0.9528]$$

$$k = 0.7641$$

Αποδεικνύεται άρα αυτό που εικάστηκε στην αντίστοιχη ενότητα της εργασίας, ότι δηλαδή για 2 κλάσεις τα αποτελέσματα του μοντέλου θα ήταν πολύ καλύτερα, επιτυγχάνοντας μεγάλα ποσοστά ακριβείας.

Ενδεχομένως κάνοντας μια νέα αναζήτηση πλέγματος, να επιτυγχάνονταν ακόμα μεγαλύτερη ακρίβεια για διαφορετικές τιμές των παραμέτρων (αριθμός χαρακτηριστικών, ακτίνα). Οι βέλτιστες παράμετροι πολύ πιθανόν να διαφέρουν, εφόσον σε αυτή την περίπτωση υπάρχουν 2 κλάσεις αντί για 5. Για παράδειγμα, για τυχαίο συνδυασμό 20 χαρακτηριστικών και ακτίνας 0.6 που δοκιμάστηκε, το OA έφτασε στο 0.95 περίπου και το k στο 0.84.