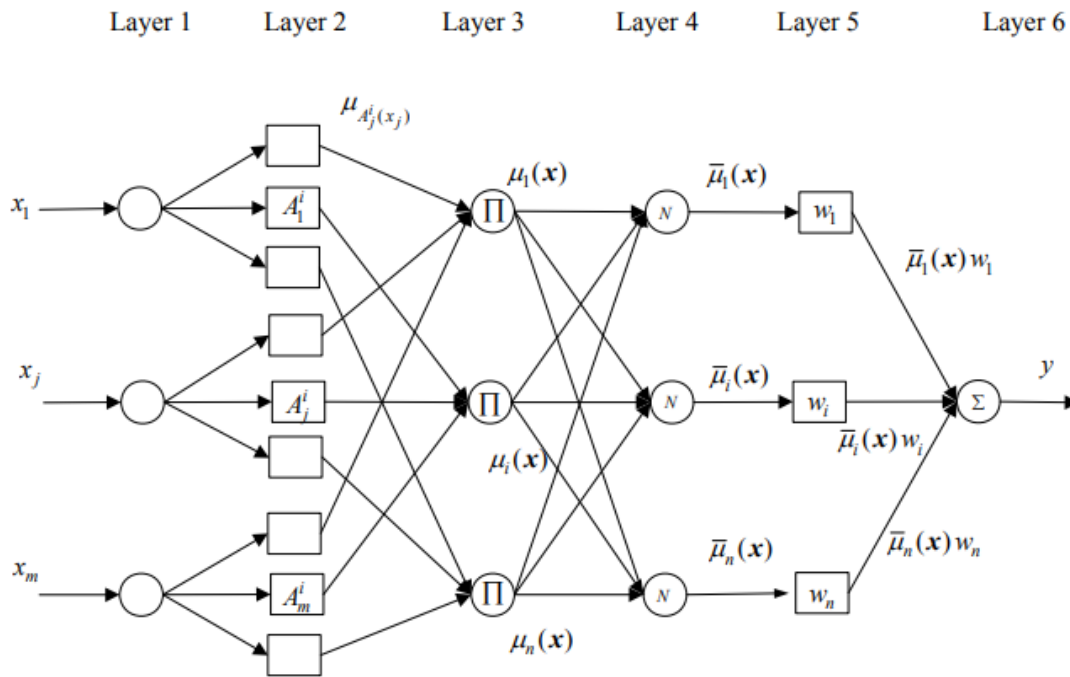


ΑΣΑΦΗ ΣΥΣΤΗΜΑΤΑ



Εργασία 4 σειρά 3

Επίλυση προβλήματος ταξινόμησης με χρήση μοντέλων TSK

Παρασκευαΐδης Κωνσταντίνος

AEM: 7754

konstapf@auth.gr

ΕΙΣΑΓΩΓΗ

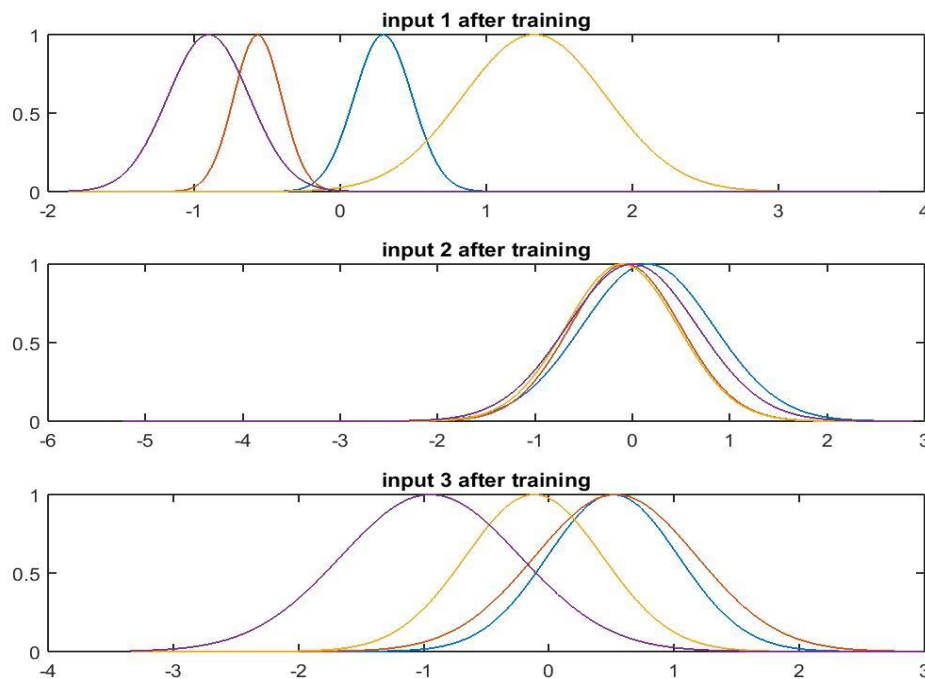
Στόχος της εργασίας αυτής είναι να διερευνηθεί η ικανότητα των TSK ασαφών μοντέλων στην επίλυση προβλημάτων ταξινόμησης. Επιλέγονται δύο σύνολα δεδομένων με σκοπό την ταξινόμηση, από τα διαθέσιμα δεδομένα, δειγμάτων στις εκάστοτε κλάσεις τους, με χρήση ασαφών νευρωνικών μοντέλων. Η εργασία αποτελείται 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.

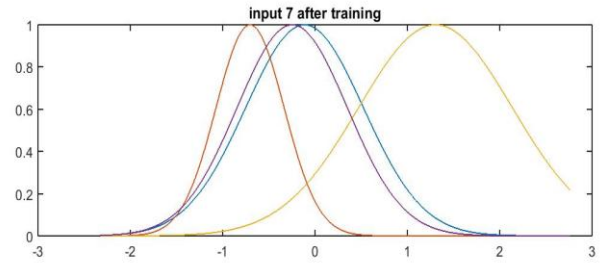
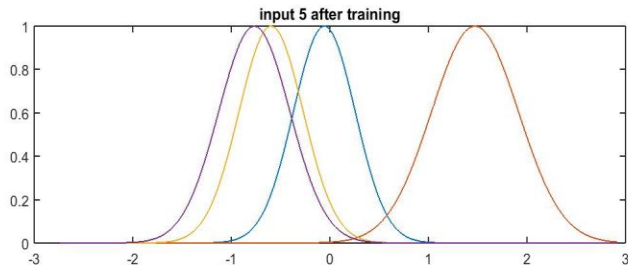
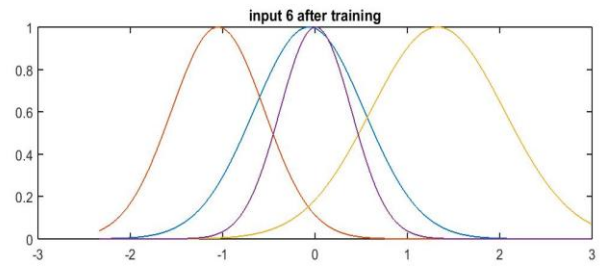
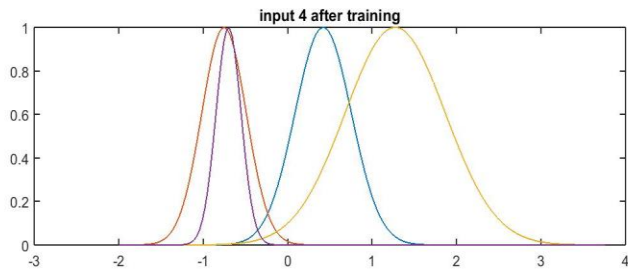
1 UCI dataset

Αρχικά θα εξεταστούν διάφορα μοντέλα TSK όσον αφορά την απόδοσή τους στο σύνολο ελέγχου. Συγκεκριμένα, θα εκπαιδευτούν τέσσερα TSK μοντέλα, στα οποία θα μεταβάλλεται το πλήθος των ασαφών IF-THEN κανόνων. Η διαμέριση του χώρου εισόδου θα γίνει με τη μέθοδο του Fuzzy C-Means (FCM) και τα TSK μοντέλα που θα προκύψουν θα διαφέρουν ως προς την παράμετρο που καθορίζει τον αριθμό των κανόνων.

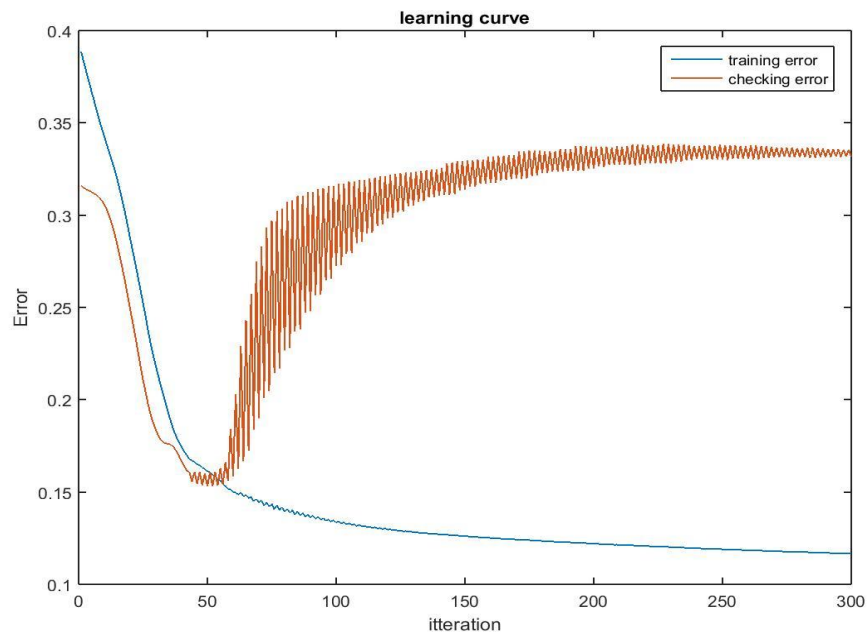
1.1 model_1 NR=4

Τελικές μορφές ασαφών συνόλων





Learning curve



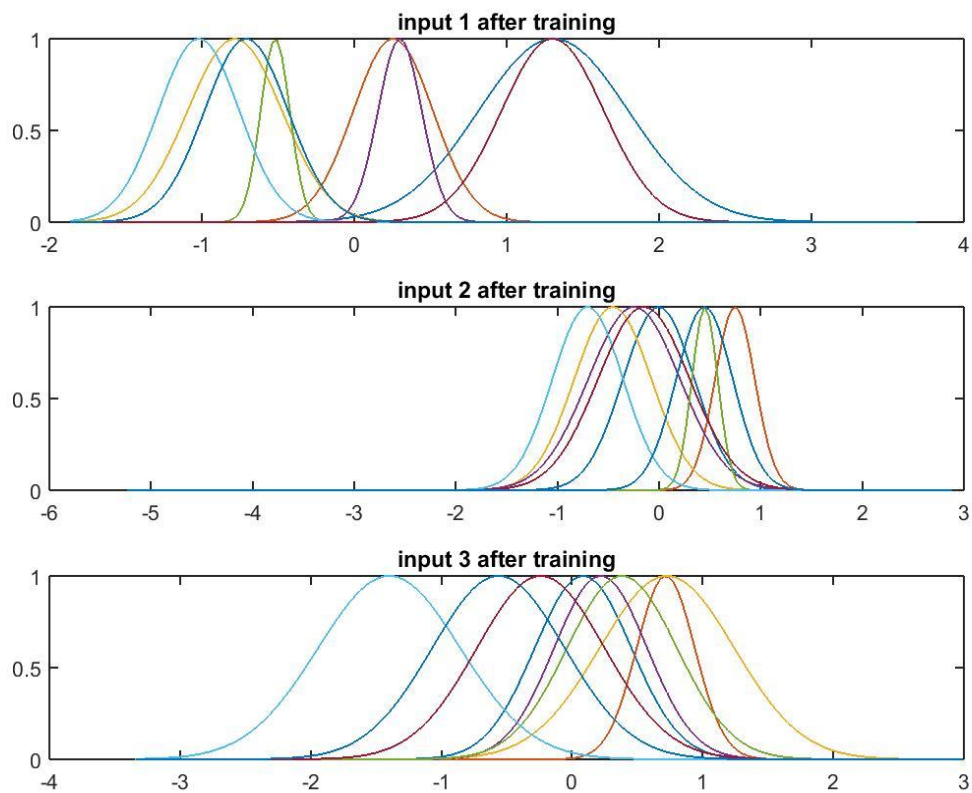
Error matrix		Actual Class			
		C1	C2	C3	C4
Predicted Class	C1	99	0	0	1
	C2	0	96	4	0
	C3	0	2	98	0
	C4	0	0	0	100

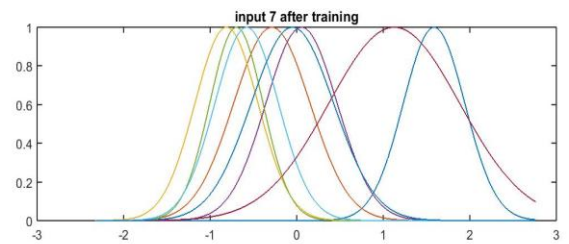
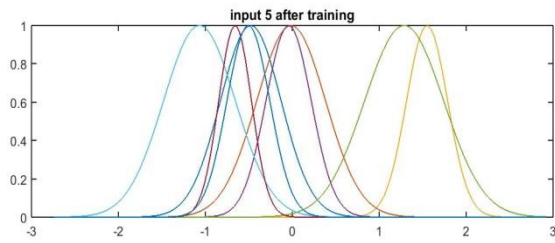
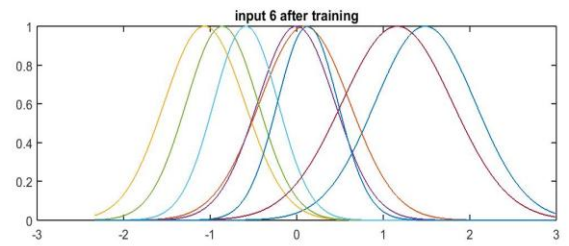
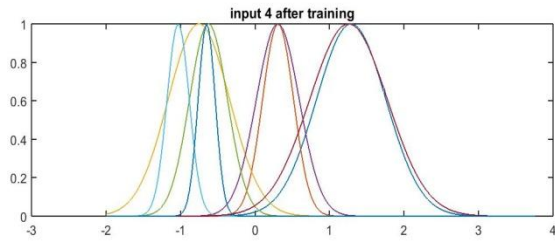
PA		UA	
C1	0.9705	C1	0.99
C2	0.9411	C2	0.6
C3	0.9607	C3	0.98
C4	0.9803	C4	1

OA	0,9825
K	0,9796

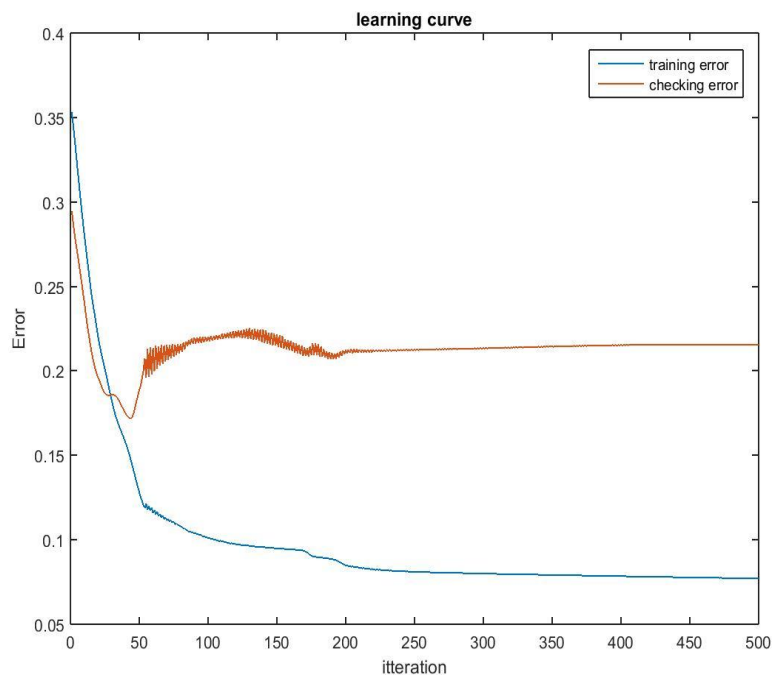
1.2 model_2 NR=8

Τελικές μορφές ασαφών συνόλων





Learning curve



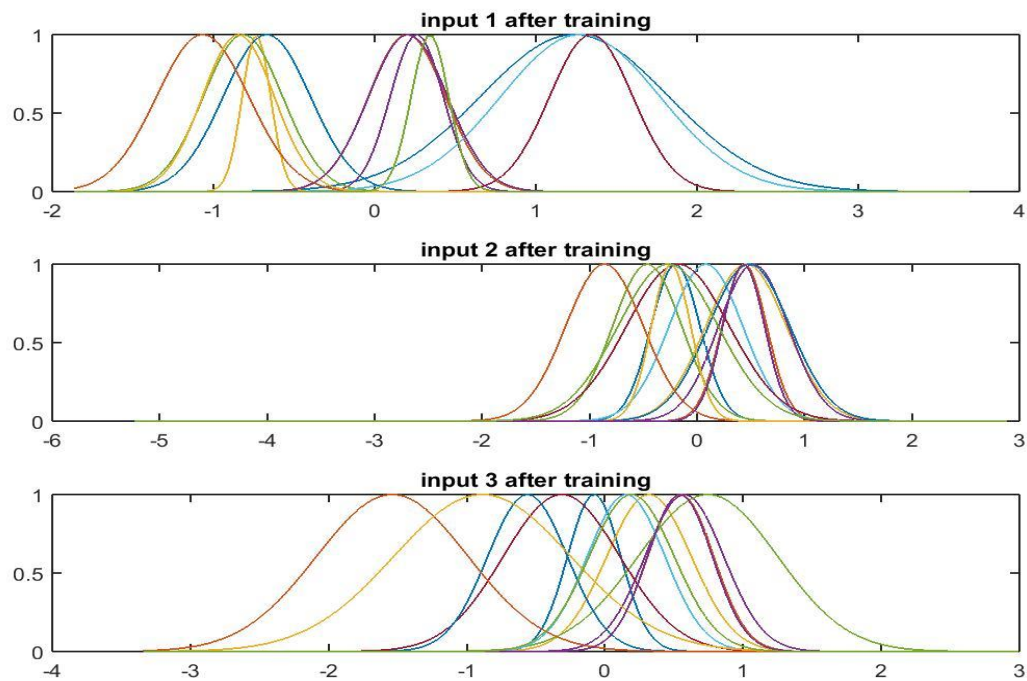
Error matrix		Actual Class			
		C1	C2	C3	C4
Predicted Class	C1	100	0	0	0
	C2	0	97	3	0
	C3	0	2	98	0
	C4	0	0	3	97

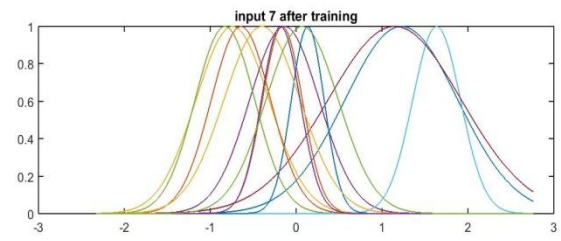
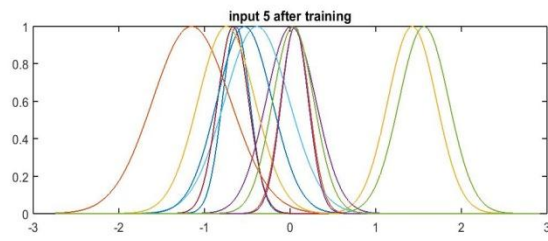
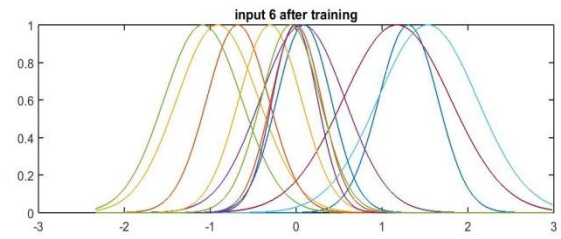
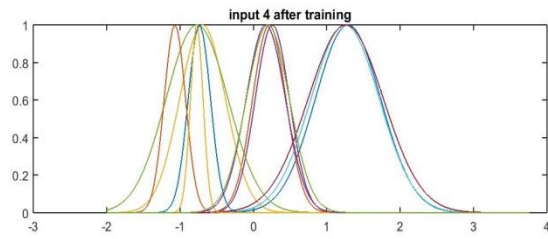
PA		UA	
C1	0.9615	C1	1
C2	0.9326	C2	0.97
C3	0.9423	C3	0.98
C4	0.9326	C4	0.97

OA	0,98
K	0,9733

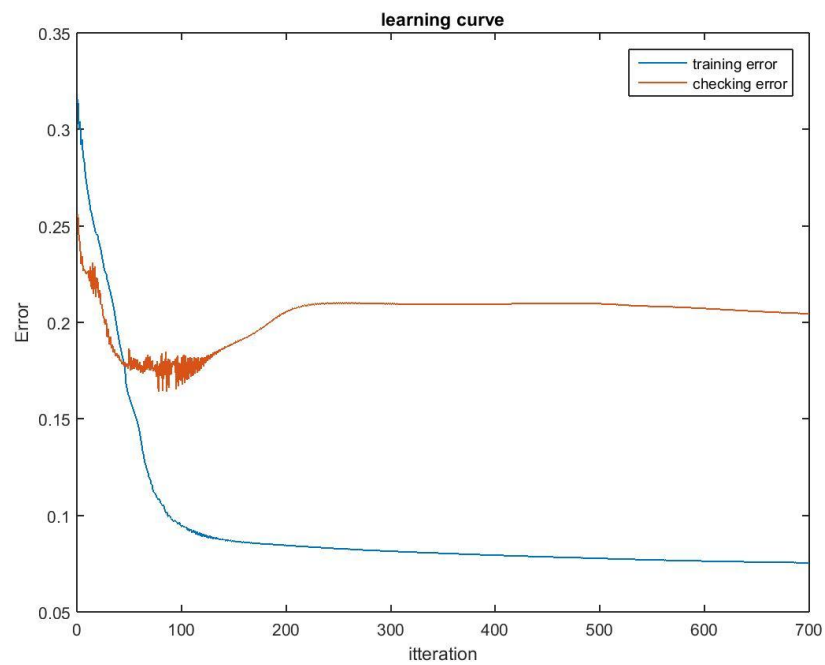
1.3 model_3 NR=12

Τελικές μορφές ασαφών συνόλων





Learninig curve



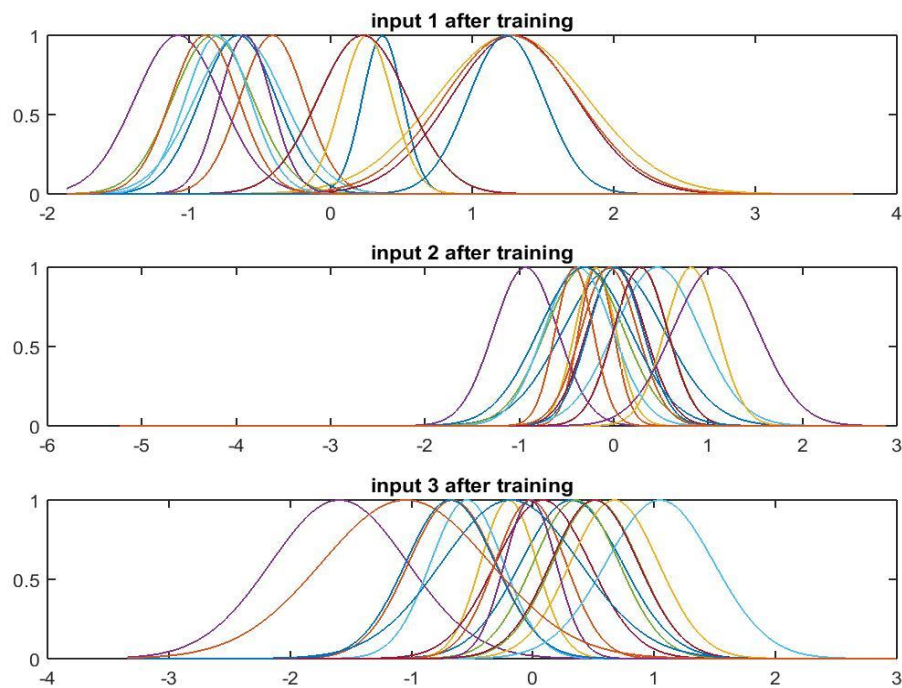
Error matrix		Actual Class			
		C1	C2	C3	C4
Predicted Class	C1	100	0	0	0
	C2	0	96	4	0
	C3	0	3	96	1
	C4	0	0	1	99

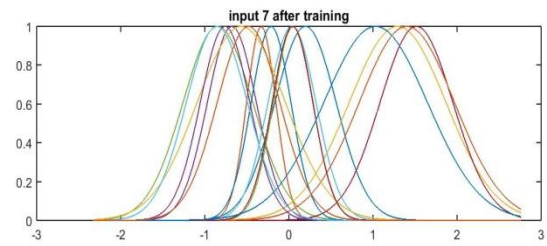
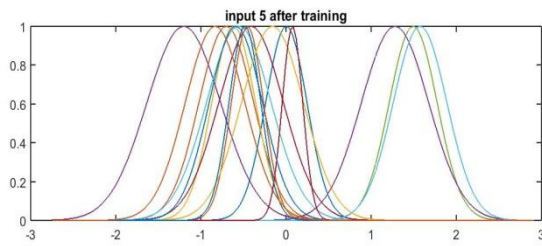
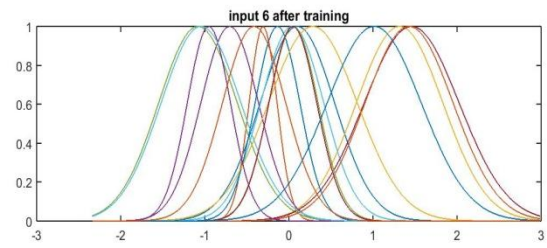
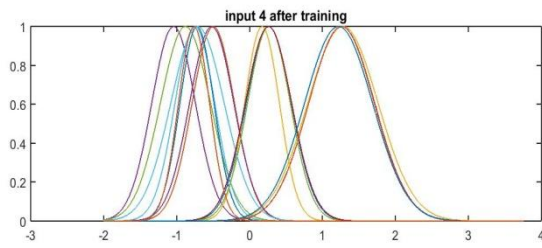
PA		UA	
C1	0.9900	C1	1
C2	0.9504	C2	0.96
C3	0.9504	C3	0.96
C4	0.9801	C4	0.99

OA	0,9775
K	0,970

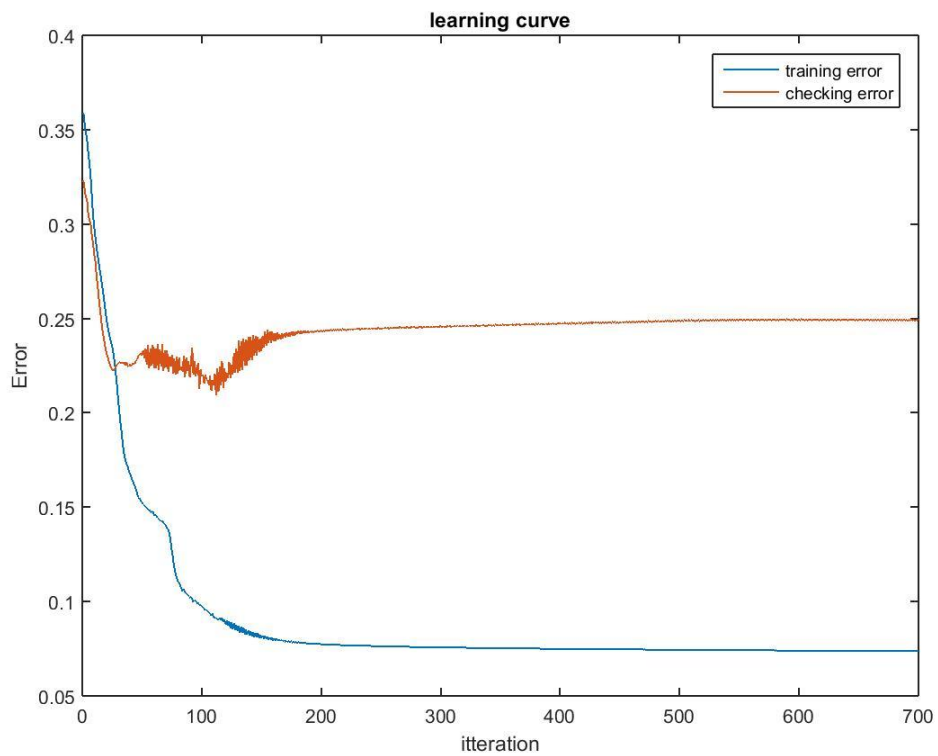
1.4 model_4 NR=16

Τελικές μορφές ασαφών συνόλων





Learning curve



Error matrix		Actual Class			
		C1	C2	C3	C4
Predicted Class	C1	99	0	0	1
	C2	0	96	4	0
	C3	0	2	98	0
	C4	0	0	1	99

PA		UA	
C1	0.9612	C1	0.99
C2	0.9320	C2	0.96
C3	0.9512	C3	0.98
C4	0.9612	C4	0.99

OA	0,98
K	0,9733

Παρατηρήσεις

Από τα διαγράμματα εκμάθησης παρατηρούμε ότι στα μοντέλα εμφανίζεται το φαινόμενο της υπέρ-εκπαίδευσης το οποίο μειώνεται όσο αυξάνουμε τον αριθμό των κανόνων. Με την αύξηση όμως των κανόνων αυξάνονται οι συναρτήσεις συμμετέχεις και η πολυπλοκότητα του συστήματος. Η μεγάλη επικάλυψη των εισόδων δεν βοηθού στην περεταίρω μείωση της υπέρ-εκπαίδευσης. Τα μοντέλα παρόλα αυτά έχουν πολύ υψηλό accuracy. Η προ-επεξεργασία του σετ μπορεί να οδηγούσε σε καλύτερα αποτελέσματα.

2 Εφαρμογή σε dataset με υψηλή διαστασιμότητα

Στη συνέχεια έγινε μια προσέγγιση στο πρόβλημα της μοντελοποίησης μιας άγνωστης συνάρτησης σε data set με υψηλή διαστασιμότητα. Λόγο της έκρηξης του πλήθους των IF-THEN κανόνων σε τέτοια data set η μοντελοποίηση με χρήση TSK καθίστατε δύσκολη. Για αυτό τον λόγο πρέπει να μειώσουμε το πλήθος των κανόνων και το πλήθος των εισόδων.

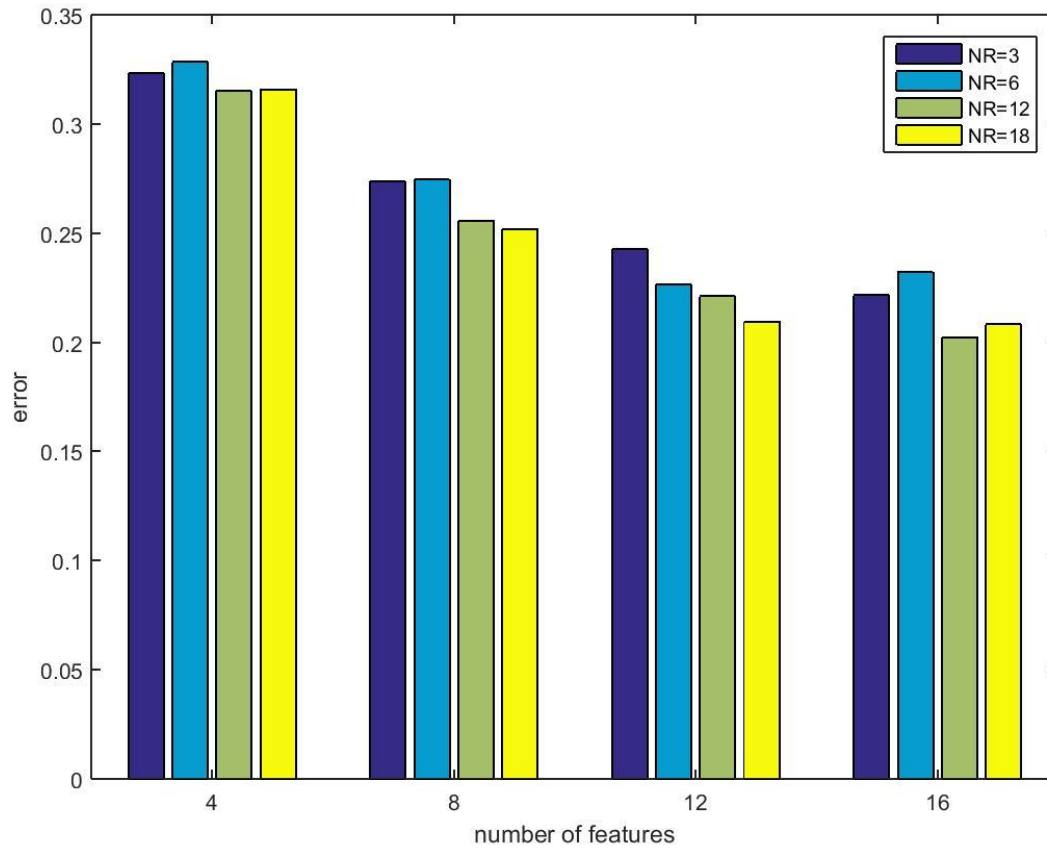
2.1 Επιλογή χαρακτηριστικών

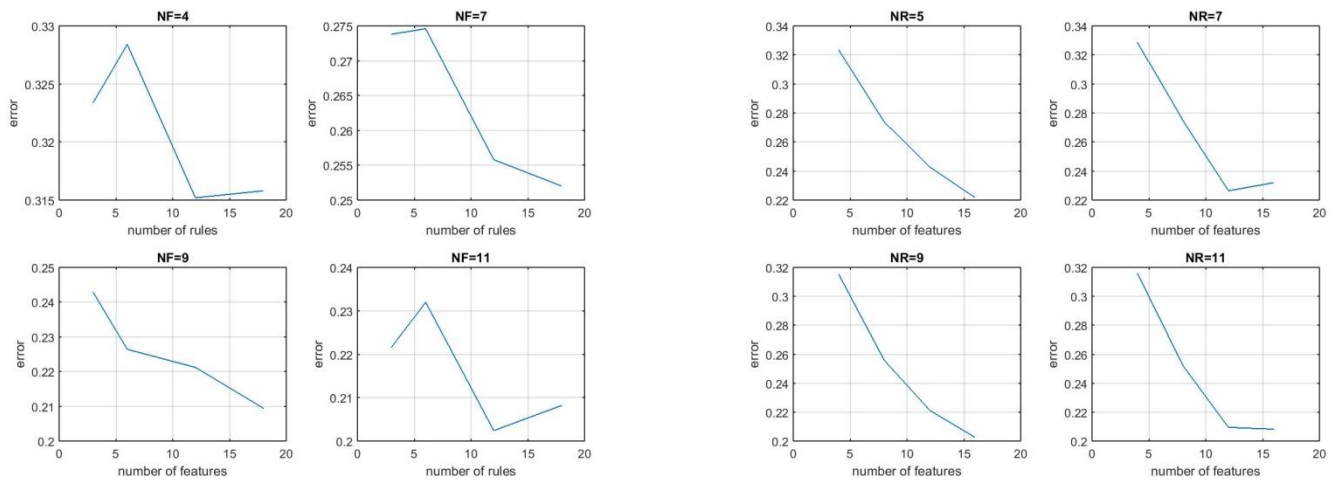
Αρχικά εκτελεστική αναζήτηση πλέγματος (grid search) και αξιολόγηση μέσω 5-πτυχης διασταυρωμένης επικύρωσης (5-fold cross validation) για την επιλογή των βέλτιστων τιμών των παραμέτρων. Ως μέθοδος διαμέρισης διασκορπισμού ο αλγόριθμος Subtractive Clustering και ως μέθοδος επιλογής χαρακτηριστικών επιλέγεται ο αλγόριθμος Relief. η αναζήτηση έγινε για τιμές χαρακτηριστούν από το

σύνολο $NF = \{4, 8, 12, 16\}$ και για αριθμό κανόνων από το σύνολο $NR = \{3, 6, 12, 18\}$. Οι τιμές που επιλέχτηκαν είναι αυτές που είχαν το μικρότερο μέσο σφάλμα, οι οποίες είναι :

- ❖ Αριθμός χαρακτηριστικών $NF=16$
- ❖ Αριθμός κανόνων $NR=12$

Στο επόμενο διάγραμμα φαίνεται το μέσο σφάλμα για όλους τους συνδυασμούς χαρακτηριστικών.

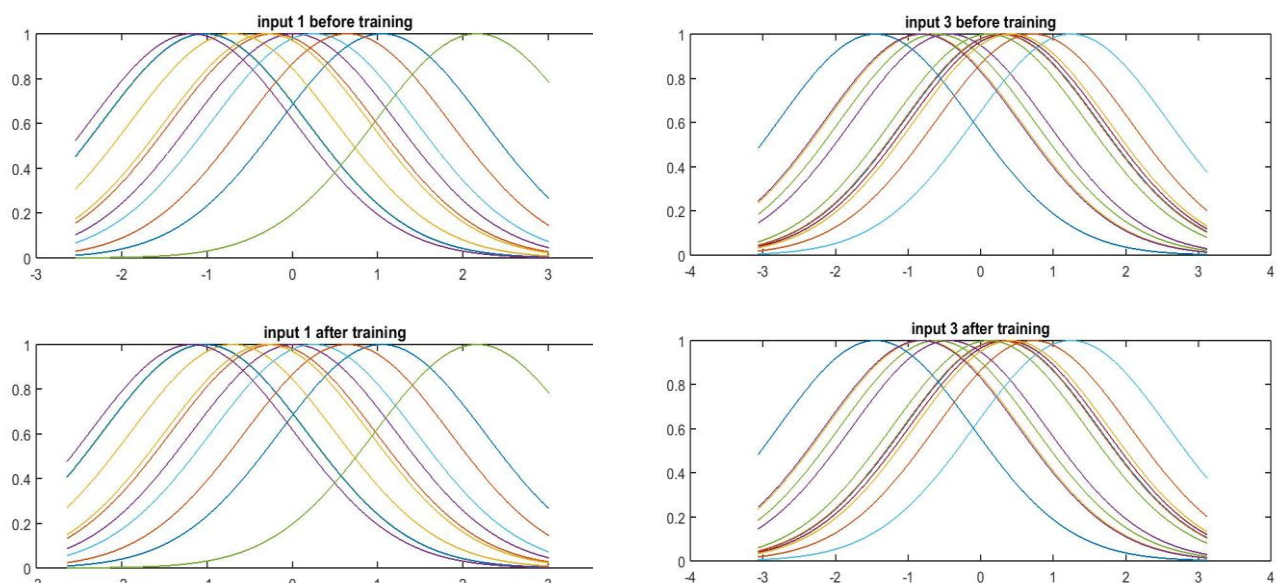


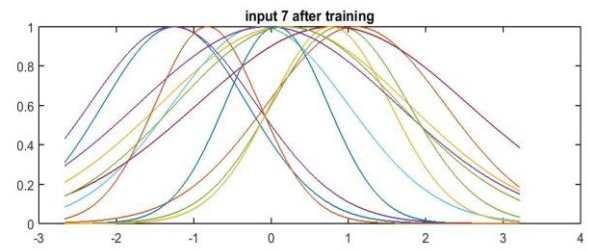
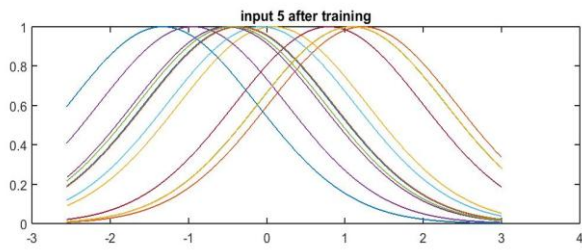
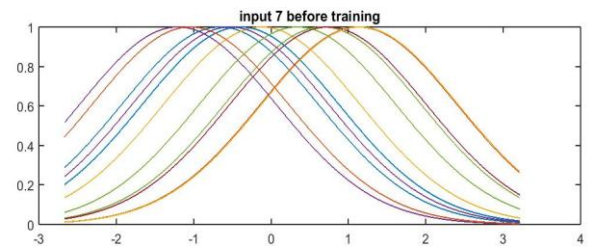


Τα τέσσερα αριστερά διαγράμματα δείχνουν την μεταβολή του σφάλματος ταξινόμησης για σταθερό αριθμό χαρακτηριστικών και τα δεξιά για σταθερό αριθμό κανόνων. Όπως παρατηρούμε για σταθερό αριθμό κανόνων αύξηση των χαρακτηριστικών μειώνει το σφάλμα. Για σταθερό αριθμό χαρακτηριστικών το σφάλμα μειώνεται με την αύξηση των κανόνων όταν έχουμε λίγα χαρακτηριστικά. Όπως παρατηρούμε για τα 11 χαρακτηρίστηκα μετά τους 12 κανόνες το σφάλμα αυξάνεται.

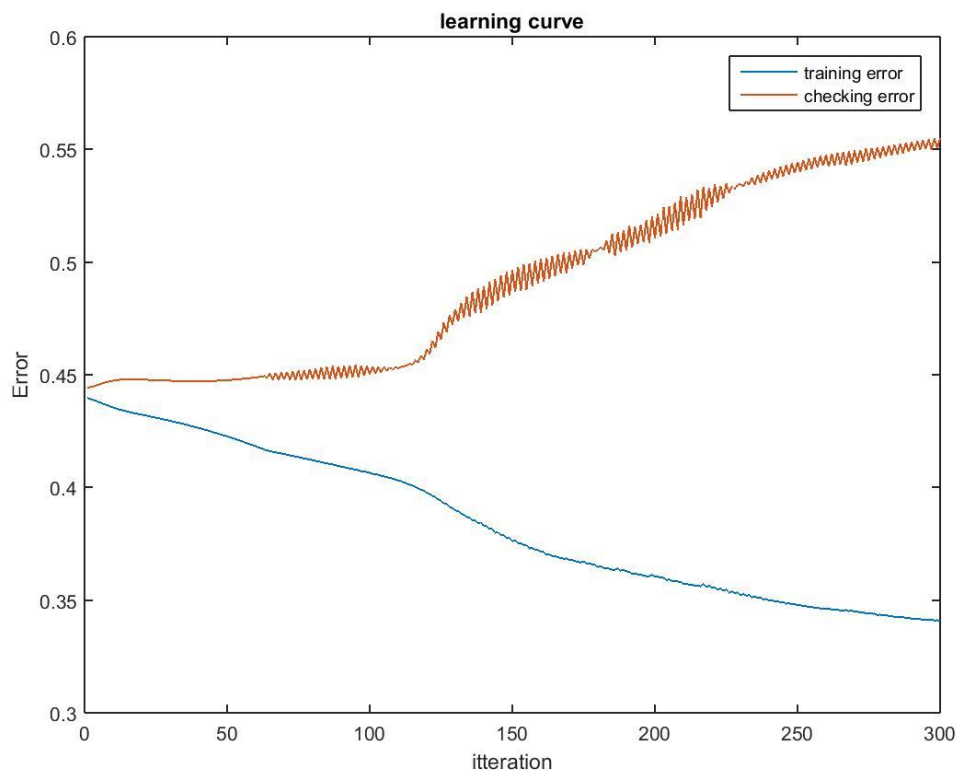
2.2 Εκπαίδευση τελικού μοντέλου

Μορφές ασαφών συνόλων





learning curve



Error matrix		Actual Class		
		C1	C2	C3
Predicted Class	C1	265	56	24
	C2	22	279	12
	C3	24	46	272

PA		UA	
C1	0.6955	C1	0.7681
C2	0.7323	C2	0.8087
C3	0.7139	C3	0.7884

OA	0,8160
K	0,7246

Στο τελικό μοντέλο παρατηρούμε το φαινόμενο της υπέρ-εκπαίδευσης , άρα το μοντέλο χάνει την ικανότητα της γενίκευσης . Μειώνοντας τις επαναλήψεις σημείο που το σφάλμα αξιολόγησης αρχίζει να αποκλίνει δεν έχει καλύτερα αποτελέσματα μιας και το σφάλμα εκπαίδευσης δεν έχει συγκλίνει ακόμα . Παρόλο δηλαδή της μείωσης της πολυπλοκότητας το μοντέλο εμφανίζει υπέρ-εκπαίδευση ίσως λόγο της μεγάλης επικάλυψης των συναρτήσεων εισόδου αλλά και της ύπαρξης θορύβου. Με χρήση grid partitioning θα είχαμε λιγότερες συναρτήσεις εισόδου ο αριθμός όμως των κανόνων θα ήταν πολύ μεγάλος εξαιτίας της υψηλής διασπασιμότητας του σετ δεδομένων δυσκόλευαν την μοντελοποίηση .