

ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ

ΕΡΓΑΣΙΑ 4^η – TSK CLASSIFICATION

ΘΩΜΑΣ ΚΥΡΙΑΚΟΣ ΠΡΑΒΙΝΟΣ, 9937

1. ΕΙΣΑΓΩΓΗ

Η εργασία στοχεύει στη διερεύνηση της ικανότητας των μοντέλων TSK στην επίλυση προβλημάτων ταξινόμησης. Είναι διαθέσιμα δύο datasets, το πρώτο εκ των οποίων θα χρησιμοποιηθεί για μια απλή διερεύνηση της διαδικασίας εκπαίδευσης και αξιολόγησης μοντέλων αυτού του είδους, καθώς και για μια επίδειξη τρόπων ανάλυσης και ερμηνείας των αποτελεσμάτων. Το δεύτερο είναι ένα σύνολο δεδομένων υψηλότερης πολυπλοκότητας που θα χρησιμοποιηθεί για μια πληρέστερη διαδικασία μοντελοποίησης, η οποία θα περιλαμβάνει μεταξύ άλλων προ-επεξεργαστικά βήματα όπως επιλογή χαρακτηριστικών (feature selection), καθώς και μεθόδους βελτιστοποίησης των μοντέλων μέσω της διασταυρωμένης επικύρωσης (cross validation). Η συγγραφή της αναφοράς βασίστηκε στα δεδομένα της τελευταίας εκτέλεσης των προγραμμάτων πριν την παράδοσή της.

2. ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ Α

Το πρώτο μέρος της εργασίας απαιτεί την εκπαίδευση τεσσάρων μοντέλων με διαφορετικά χαρακτηριστικά, όπως φαίνεται στον παρακάτω πίνακα. Το σύνολο δεδομένων Haberman's Survival από το αποθετήριο UCI, το οποίο περιέχει 306 δείγματα και 3 χαρακτηριστικά, το τελευταίο εκ των οποίων είναι το εξαρτημένο χαρακτηριστικό ή χαρακτηριστικό-στόχος. Αρχικά, τα δεδομένα πρέπει να χωριστούν σε τρεις μη επικαλυπτόμενες ομάδες: εκπαίδευση (D_{trn}), επικύρωση (D_{val}) και δοκιμή (D_{test}). Το πρώτο σύνολο, όπως υποδηλώνουν τα ονόματα, χρησιμοποιείται για την εκπαίδευση, το δεύτερο για την αποφυγή της υπερεκπαίδευσης και το τρίτο για τη δοκιμή της απόδοσης του τελικού μοντέλου. Σύμφωνα με τη θεωρία ο διαχωρισμός γίνεται σε 60%, 20%, και 20%, του αρχικού συνόλου δεδομένων, στο κάθε υποσύνολο αντίστοιχα. Για την εκπαίδευση των μοντέλων διαλέγουμε αυθαίρετα δύο ακραίες τιμές ακτίνων clusters, ώστε ο αριθμός των κανόνων ανάμεσα στα μοντέλα να παρουσιάζει σημαντική διαφορά. Τα τέσσερα μοντέλα εκπαιδεύονται χρησιμοποιώντας ένα υβριδικό μοντέλο, σύμφωνα με την οποία οι

παράμετροι των συναρτήσεων συμμετοχής βελτιστοποιούνται μέσω της μεθόδου της οπισθοδιάδοσης (backpropagation algorithm), ενώ οι παράμετροι της πολυωνυμικής συνάρτησης εξόδου βελτιστοποιούνται μέσω της μεθόδου των ελαχίστων τετραγώνων (Least Squares).

Περιγραφή Μοντέλων		
Μοντέλο	Τύπος Μοντέλου	Ακτίνα Clusters
TSK_model_1	Class Dependent	0.3
TSK_model_2	Class Independent	0.3
TSK_model_3	Class Dependent	0.9
TSK_model_4	Class Independent	0.9

Για την ακρίβεια της εκτίμησης της πραγματικής συνάρτησης από καθένα από τα παραπάνω μοντέλα, θα χρησιμοποιηθούν οι εξής δείκτες απόδοσης:

- Error Matrix: Ο πίνακας σφαλμάτων ταξινόμησης (αλλιώς και Πίνακας Σύγχυσης (Confusion Matrix) είναι ένας $k \times k$ πίνακας, με k τον αριθμό των κλάσεων.
- Overall Accuracy (OA): Η συνολική ακρίβεια ενός ταξινομητή ορίζεται ως το ποσοστό των ορθώς ταξινομημένων δειγμάτων ως προς το συνολικό πλήθος των δειγμάτων.
- Producer's accuracy – User's accuracy (PA-UA): Δύο δείκτες που παρουσιάζουν ενδιαφέρον και αναφέρονται στην απόδοση του ταξινομητή όσον αφορά κάθε κλάση ξεχωριστά, είναι η ακρίβεια παραγωγού και η ακρίβεια χρήστη.
- K^{hat} : Αποτελεί εκτίμηση της πραγματικής στατιστικής παραμέτρου.

(Οι μαθηματικοί τύποι δίνονται στην εκφώνηση της εργασίας)

3. ΕΚΠΑΙΔΕΥΣΗ

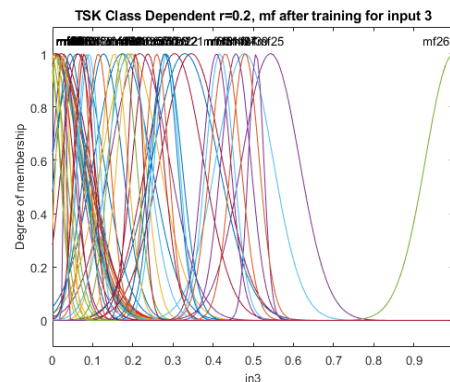
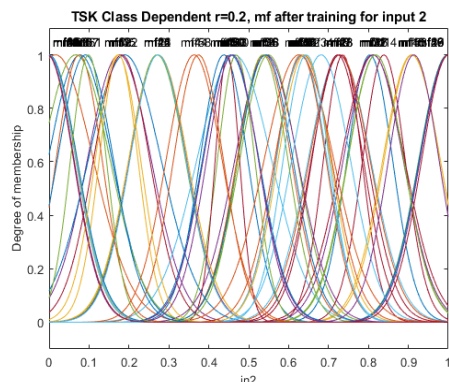
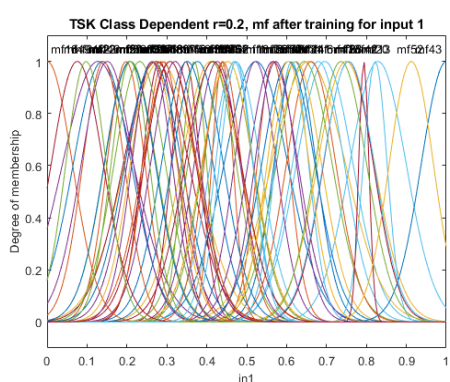
Γίνεται αρχικοποίηση των μοντέλων, σύμφωνα με τις δοθείσες προδιαγραφές και μετά από την ολοκλήρωση της εκπαίδευσης λαμβάνουμε τα εξής :

- Τελικές μορφές των ασαφών συνόλων
- Διαγράμματα μάθησης
- Διάγραμμα σφαλμάτων πρόβλεψης
- Τιμές των δεικτών απόδοσης OA, PA, UA, K^{hat}

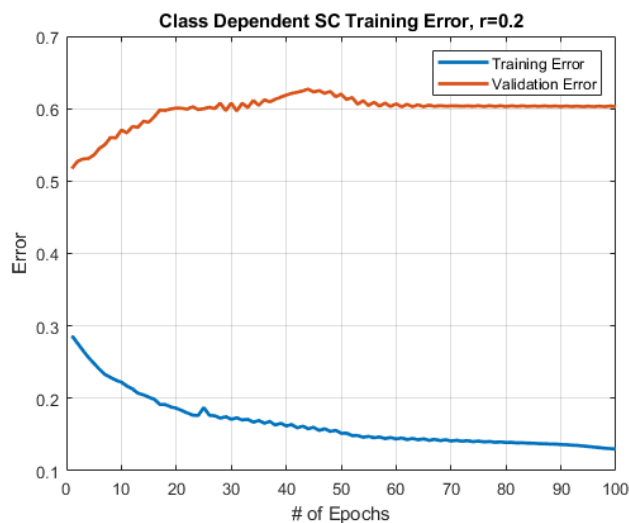
Κάθε μοντέλο εκπαιδεύεται για 100 Epochs.

3.1. ΜΟΝΤΕΛΟ 1

Τελικές μορφές των ασαφών συνόλων



Διάγραμμα Μάθησης

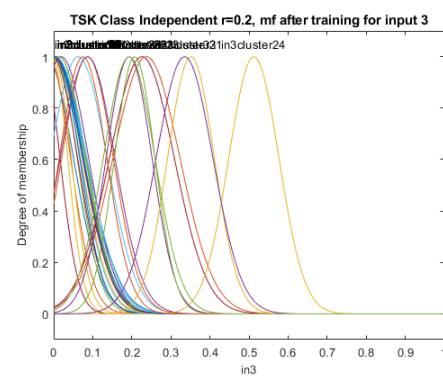
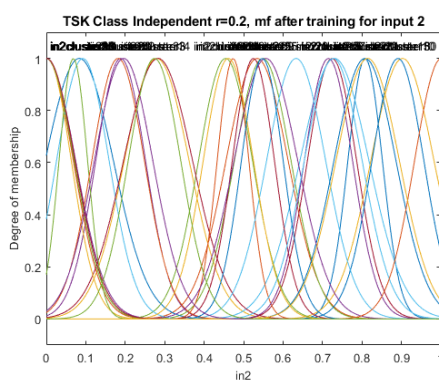
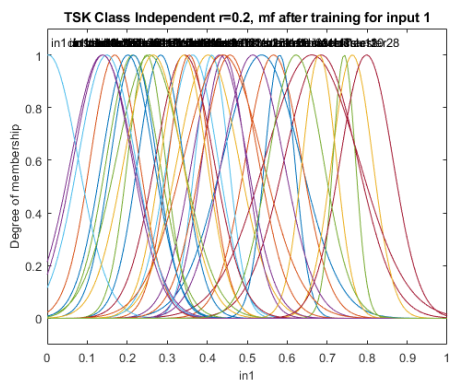


Error Matrix Μοντέλου 1

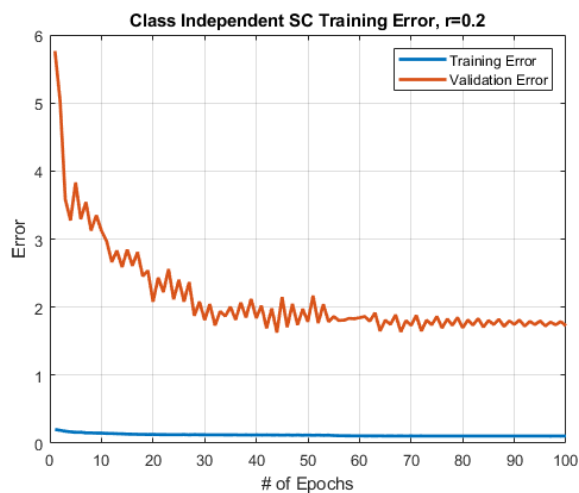
	Actual 1	Actual 2
Predicted 1	38	10
Predicted 2	7	8

3.2. ΜΟΝΤΕΛΟ 2

Τελικές μορφές των ασαφών συνόλων



Διάγραμμα Μάθησης

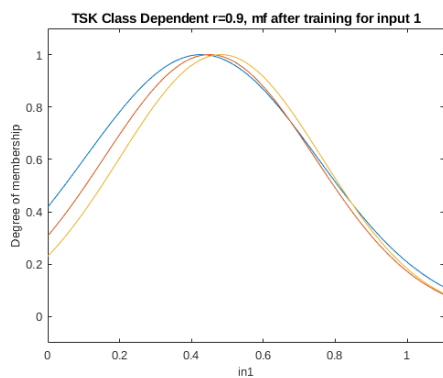


Error Matrix Μοντέλου 2

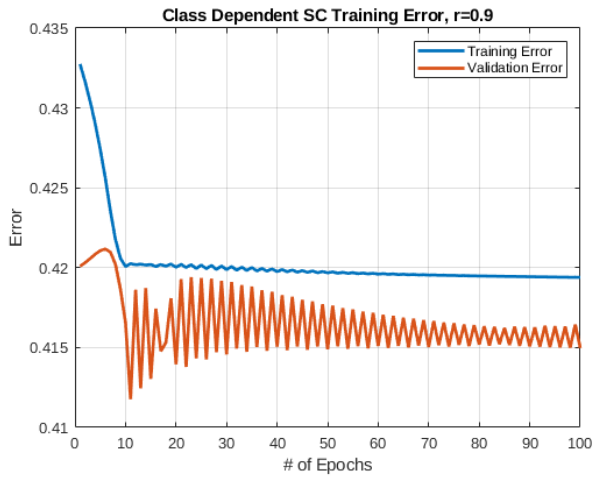
	Actual 1	Actual 2
Predicted 1	36	10
Predicted 2	6	9

3.3. ΜΟΝΤΕΛΟ 3

Τελικές μορφές των ασαφών συνόλων



Διάγραμμα Μάθησης

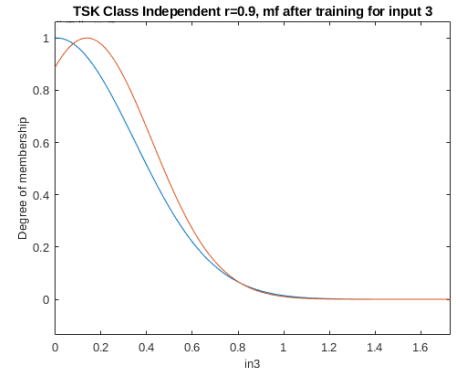


Error Matrix Μοντέλου 3

	Actual 1	Actual 2
Predicted 1	42	4
Predicted 2	8	6

3.4. ΜΟΝΤΕΛΟ 4

Τελικές μορφές των ασαφών συνόλων



Διάγραμμα Μάθησης



Error Matrix Μοντέλου 4

	Actual 1	Actual 2
Predicted 1	42	4
Predicted 2	10	6

3.5. ΑΠΟΤΕΛΕΣΜΑΤΑ

	OA	PA	UA	K^{hat}	Rules
Μοντέλο 1	0.7377	0.7547 0.625	0.9302 0.278	0.2481	45
Μοντέλο 2	0.7152	0.7551 0.607	0.8605 0.333	0.2146	22
Μοντέλο 3	0.8033	0.8400 0.636	0.9130 0.466	0.4172	3
Μοντέλο 4	0.7705	0.8077 0.555	0.9130 0.333	0.2848	2

Σε **bold** σημειώνονται οι βέλτιστες τιμές της κάθε μετρικής για όλα τα μοντέλα. Παρατηρείται πως οι περισσότερες βέλτιστες τιμές δεικτών ανήκουν για το Μοντέλο 3. Δεν μπορεί να βγει κάποιο συμπέρασμα για το πως επηρεάζεται η απόδοση του μοντέλου σύμφωνα με το πλήθος των κανόνων. Βεβαίως, αν συγκριθούν τα μοντέλα βάσει τύπου (Class Dependent / Class Independent), φαίνεται πως η μέθοδος Dependent Clustering παράγει περισσότερους κανόνες και έχει εξίσου ή και καλύτερη απόδοση από τα Class Independent μοντέλα. Επίσης, παρατηρείται πως τα μοντέλα 1 και 2 έχουν μεγαλύτερο πλήθος συναρτήσεων συμμετοχής και ταυτόχρονα μεγαλύτερο ποσοστό επικάλυψης μεταξύ διαδοχικών συναρτήσεων συμμετοχής και έχουν μικρότερο OA από τα άλλα δύο μοντέλα. Σύμφωνα με τις παραπάνω παρατηρήσεις, μια μέθοδος που θα μπορούσε να βελτιώσει την σχεδίαση του τμήματος υπόθεσης είναι να αφαιρούνται οι συναρτήσεις μεταφοράς που παρουσιάζουν μεγάλη επικάλυψη. Έτσι το μοντέλο θα απλοποιείται και θα αυξάνεται η δυνατότητα γενίκευσης του.

4. ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ B

Για την εργασία, επιλέγεται το Seizure Recognition dataset από το UCI Repository, το οποίο περιλαμβάνει 11.500 δείγματα, με κάθε ένα να περιγράφεται από 179 χαρακτηριστικά. Για την αντιμετώπιση του φαινομένου rule explosion, εφαρμόζονται δύο σημαντικές μέθοδοι:

Επιλογή Χαρακτηριστικών: Εδώ, αποφασίζεται πόσα από τα 81 αρχικά χαρακτηριστικά θα χρησιμοποιηθούν για την εκπαίδευση των μοντέλων.

Διαμέριση Διασκορπισμού (Clustering): Αυτή η μέθοδος επηρεάζει την ακτίνα των clusters και, κατά συνέπεια, το πλήθος των κανόνων που θα προκύψουν.

Για την επιλογή των καλύτερων παραμέτρων, χρησιμοποιείται η μέθοδος grid search, κατά την οποία δημιουργείται ένα δισδιάστατο πλέγμα, όπου κάθε σημείο του αντιστοιχεί σε μια συνδυασμένη τιμή για τις

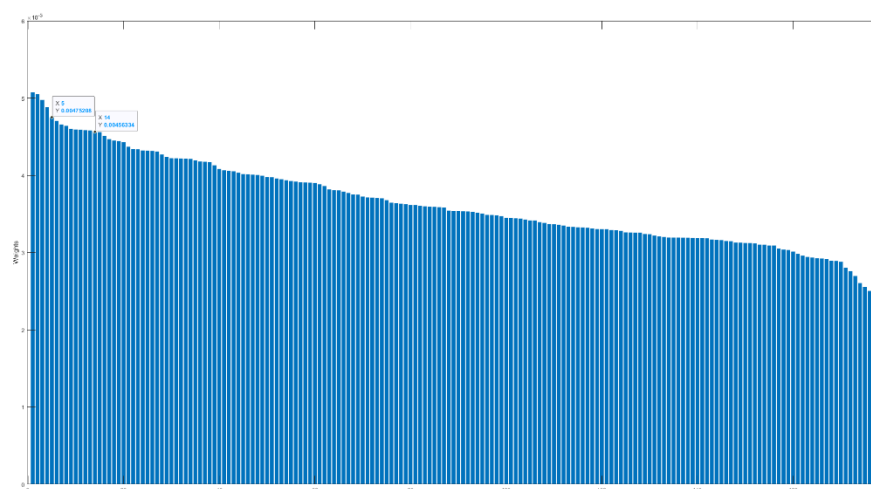
παραμέτρους. Σε αυτά τα σημεία, εφαρμόζεται μια μέθοδο αξιολόγησης για να αξιολογηθεί η απόδοση του μοντέλου. Πιο συγκεκριμένα, η μέθοδος αξιολόγησης που χρησιμοποιείται είναι η διασταυρωμένη επικύρωση (Cross Validation). Αυτό σημαίνει ότι για κάθε συνδυασμό παραμέτρων, διαχωρίζεται το σύνολο εκπαίδευσης σε δύο υποσύνολα, ένα για την εκπαίδευση του μοντέλου και ένα για την αξιολόγησή του. Αυτή η διαδικασία επαναλαμβάνεται πολλές φορές (συνήθως 5 ή 10), και τελικά υπολογίζεται ο μέσος όρος του σφάλματος του μοντέλου.

Για κάθε διαδικασία διασταυρωμένης επικύρωσης, τα μοντέλα εκπαιδεύονται για 50 Epochs.

5.GRID SEARCH

Για την επιλογή χαρακτηριστικών, εφαρμόζεται στα δεδομένα η συνάρτηση του MATLAB, `relieff` για `k-nearest neighbors`. Η συνάρτηση αυτή επιστρέφει τους δείκτες των χαρακτηριστικών των δειγμάτων με σειρά σημαντικότητας βάσει ενός βάρους, όπως φαίνεται στο παρακάτω σχήμα. Παρατηρείται πως ήδη μέχρι το 140 σημαντικότερο χαρακτηριστικό, το βάρος είναι πολύ μικρό. Επομένως δεν θα χρειαστεί να χρησιμοποιηθούν περισσότερα χαρακτηριστικά. Επίσης για λόγους πρακτικότητας δεν ξεπερνάμε αυτήν την τιμή διότι ο χρόνος εκπαίδευσης αυξάνεται εκθετικά στο περιβάλλον MATLAB, σε σχέση με τον αριθμό χαρακτηριστικών.

Βάρη σημαντικότητας χαρακτηριστικών:

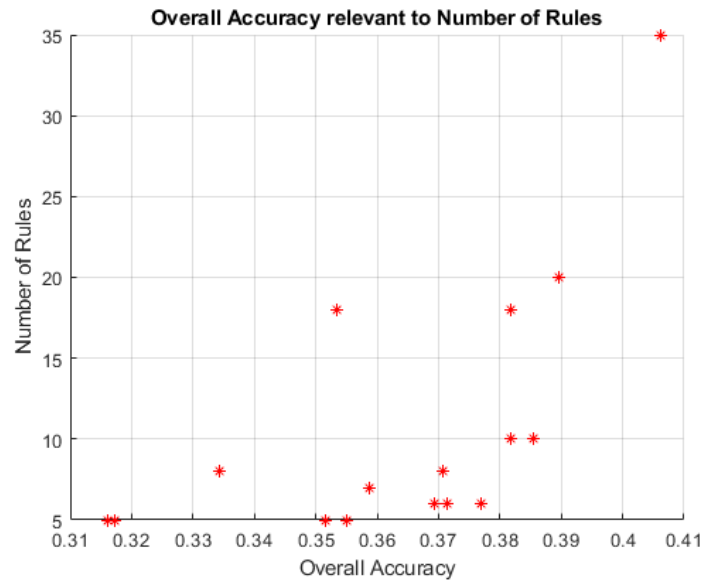


(Τα figures υπάρχουν σε φυσιολογική ανάλυση στον φάκελο του παραδοτέου)

Επομένως για τις παραμέτρους του πλέγματος αυθαίρετα διαλέγουμε:

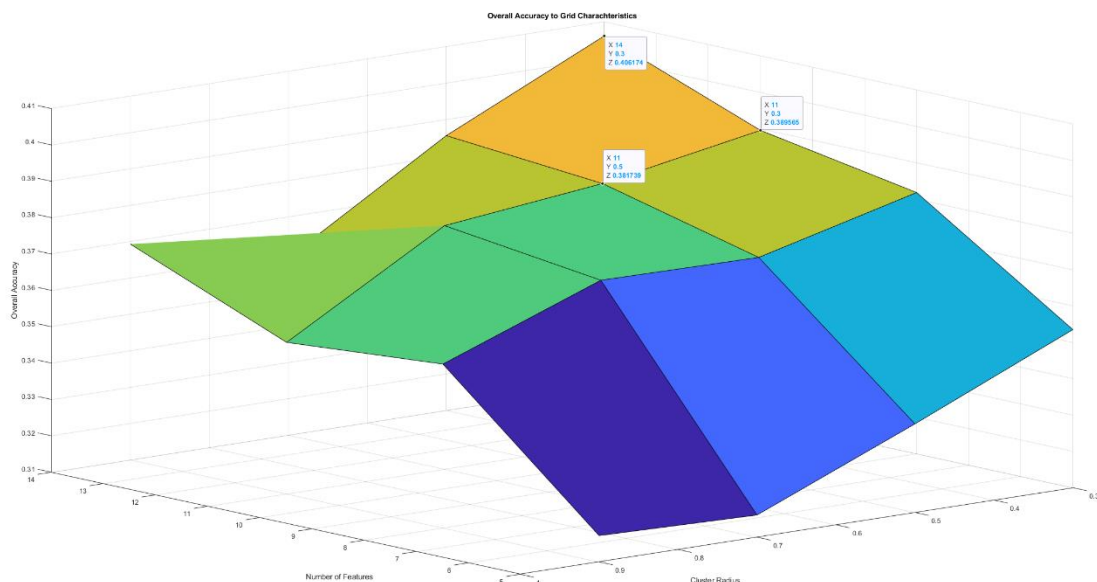
- Αριθμός χαρακτηριστικών = {5, 8, 11, 14}
- Ακτίνα cluster $r_{\text{cluster}} = \{0.3, 0.5, 0.7, 0.9\}$

Στη συνέχεια, παίρνουμε το γράφημα RMSE συναρτήσεως του πλήθους κανόνων:



Παρατηρώντας το γράφημα, φαίνονται 16 μετρήσεις καθώς υπολογίζεται ο μέσος όρος της ακρίβειας από όλα τα k -fold. Βγαίνει το συμπέρασμα πως καθώς αυξάνεται το πλήθος των κανόνων, αυξάνεται και η ακρίβεια του μοντέλου εκθετικά.

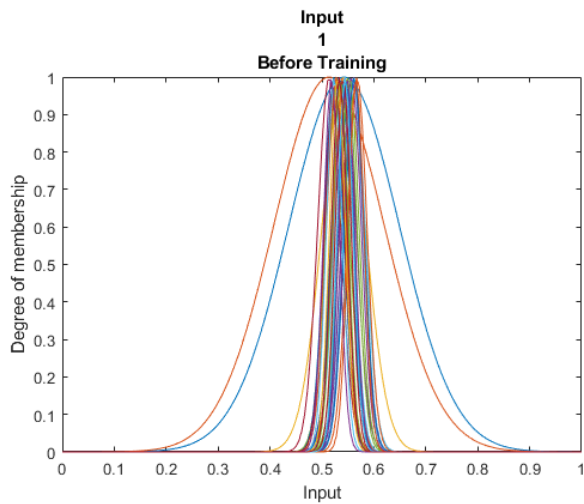
Τα υποψηφία σημεία εύρεσης του βέλτιστου μοντέλου μπορεί να διαπιστωθούν από την επιφάνεια του παρακάτω σχήματος:



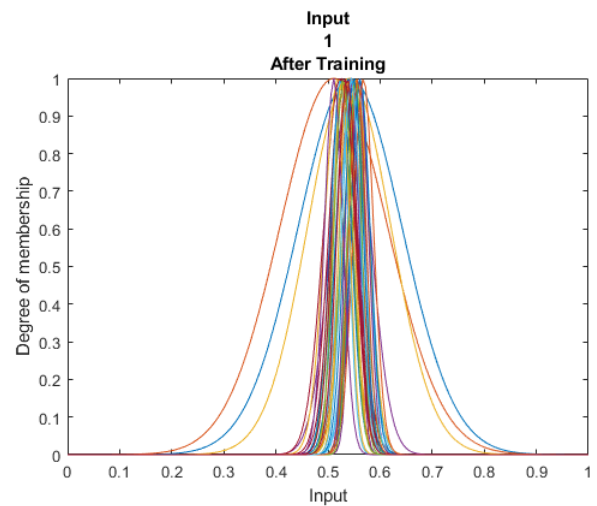
Επιλέγεται μοντέλο με αριθμό χαρακτηριστικών $\text{feat}_{\text{num}} = 14$ και μέγεθος ακτίνας $\text{cluster}, r_{\text{cluster}} = 0.3$, το εκπαιδεύουμε για τα αρχικά δείγματα εκπαίδευσης αλλά κρατάμε μόνο τα 11 πιο σημαντικά χαρακτηριστικά. Από την εκπαίδευση του μοντέλου έχουμε:

Αρχικές και τελικές μορφές Ασαφών Συνόλων:

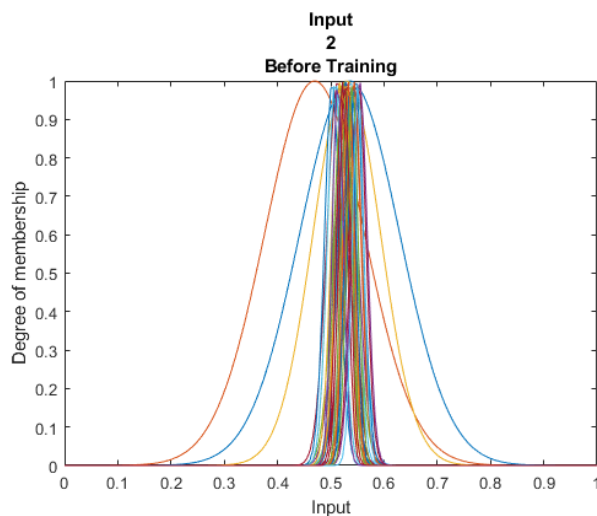
Είσοδος 1 Αρχικό



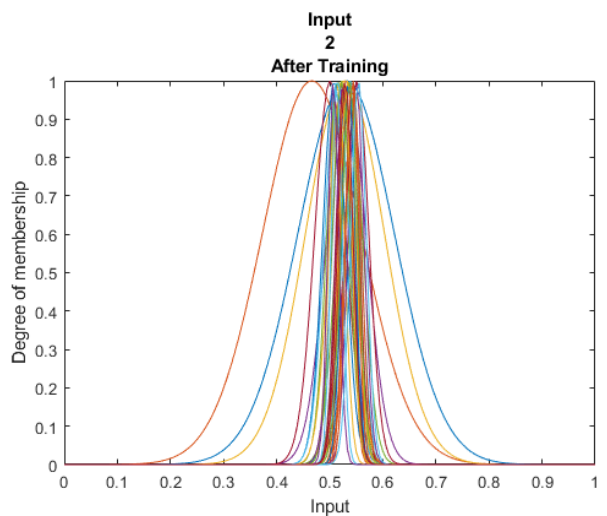
Είσοδος 1 Τελικό



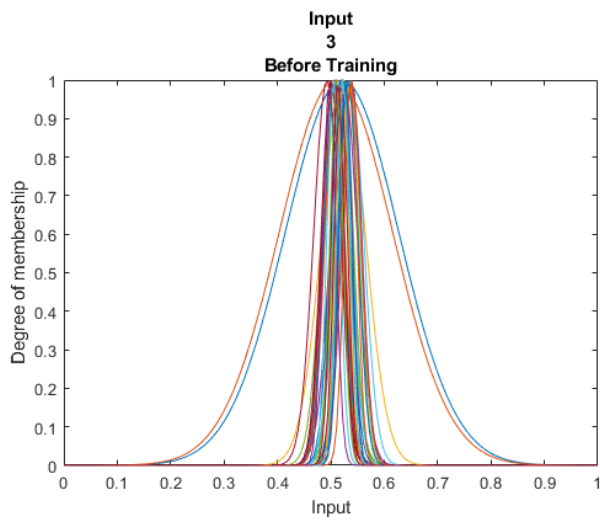
Είσοδος 2 Αρχικό



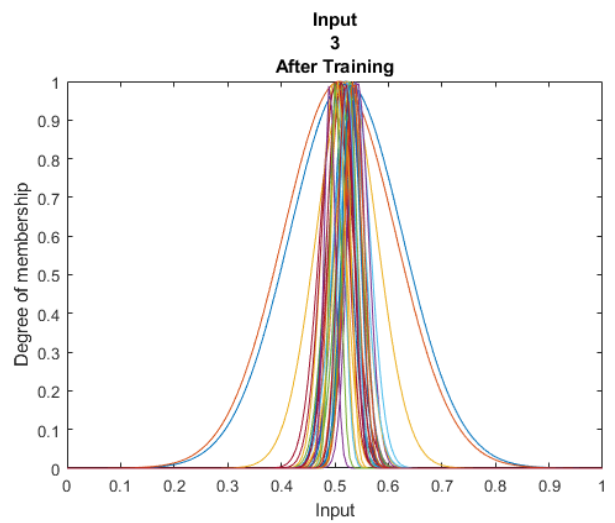
Είσοδος 2 Τελικό



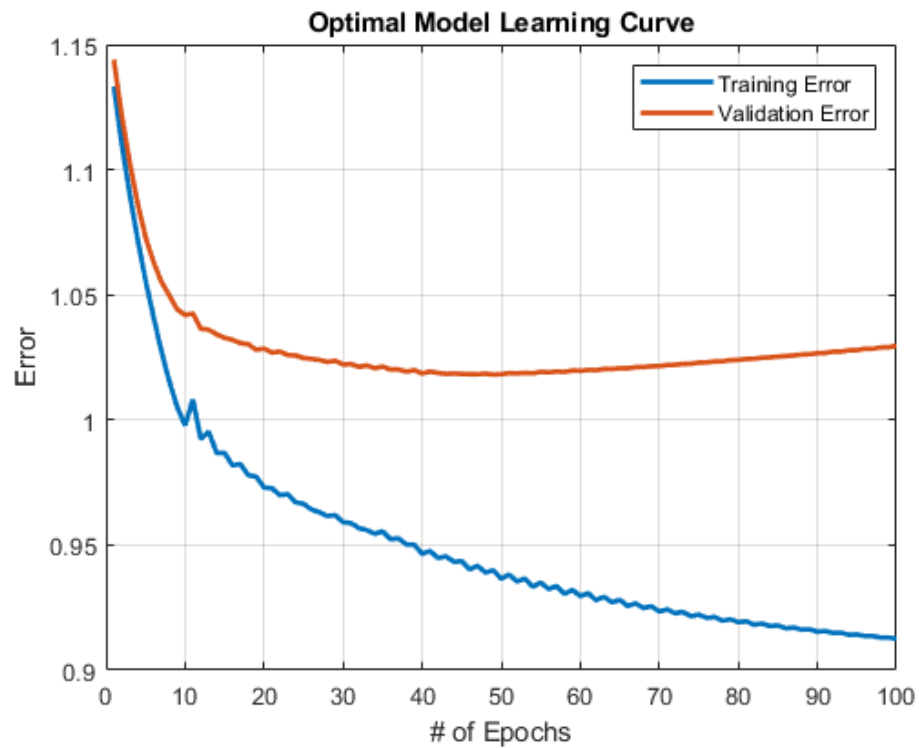
Είσοδος 3 Αρχικό



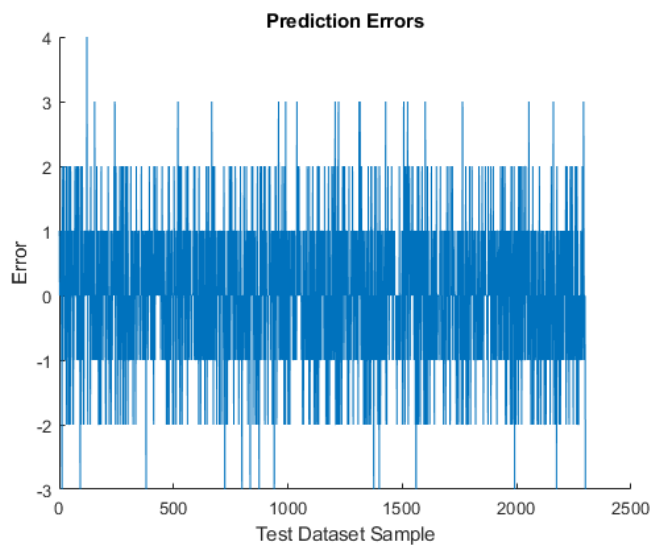
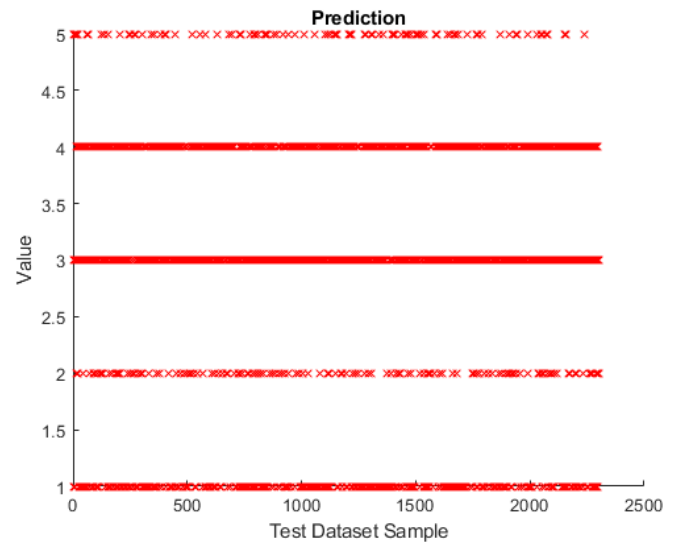
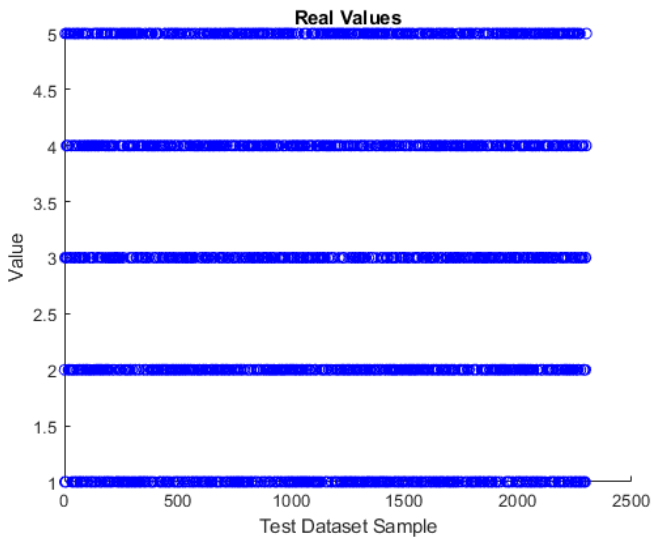
Είσοδος 3 Τελικό



Διάγραμμα Μάθησης:



Διαγράμματα Πραγματικών – Προβλεπόμενων τιμών



Δείκτες Απόδοσης Βέλτιστου Μοντέλου:

OA	κ^{hat}	PA	UA
0.3878	0.2334	0.7309	0.9130
		0.0717	0.1868
		0.6152	0.3096
		0.4720	0.2855
		0.0721	0.5000

Με βάση το διάγραμμα εκμάθησης, δεν παρατηρείται κάποιο φαινόμενο υπερεκπαίδευσης παρόλο που παρουσιάζεται μεγάλη επικάλυψη μεταξύ της πληθώρας των συναρτήσεων συμμετοχής. Από τον παραπάνω πίνακα, φαίνεται πως όταν ένα δείγμα ανήκει στην κλάση 1, τότε έχει 73, 09% πιθανότητα να ταξινομηθεί σωστά σε αυτήν την κλάση (PA_1). Στην περίπτωση που επιλεχθεί τυχαία ένα δείγμα από το σύνολο των δειγμάτων που το μοντέλο έχει ταξινομήσει στην κλάση 1, τότε η πιθανότητα αυτό το δείγμα να ανήκει ορθώς σε αυτήν την κλάση είναι 91,30%. Αντίστοιχα συμπεράσματα προκύπτουν και για τα δείγματα των υπόλοιπων κλάσεων. Στην περίπτωση που επιλέγονταν η μέθοδος `grid partitioning` θα υπήρχαν λιγότερες συναρτήσεις συμμετοχής για τις μεταβλητές εισόδου του μοντέλου, αλλά ο αριθμός των κανόνων θα ήταν εκθετικά μεγαλύτερος εξαιτίας της υψηλής διαστασιμότητας του dataset, οπότε θα δυσκόλευε την διαδικασία εκπαίδευσης του μοντέλου.