



ΑΣΑΦΗ ΣΥΣΤΗΜΑΤΑ

REGRESSION

Κιλάρογλου Ελευθέριος (8501)

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Θεσσαλονίκη 2021

\

Εισαγωγή

Στόχος της εργασίας αυτής είναι να διερευνηθεί η ικανότητα των μοντέλων TSK (Tagaki-Sugeno-Kang) στη μοντελοποίηση πολυμεταβλητών, μη γραμμικών συναρτήσεων. Στο πρώτο κομμάτι της εργασίας χρησιμοποιείται ένα σχετικά μικρό dataset με λίγες διαστάσεις έτσι ώστε να επιφέρει μια απλή διερεύνηση της διαδικασίας εκπαίδευσης και αξιολόγησης μοντέλων αυτού του είδους αλλά και να επιδείξει τον τρόπο ανάλυσης και ερμηνείας των αποτελεσμάτων αυτών. Αντίθετα στο δεύτερο κομμάτι της εργασίας χρησιμοποιείται ένα πολυπλοκότερο dataset με μεγάλο αριθμό διαστάσεων αλλά ταυτόχρονα και μεγάλο σε μέγεθος. Έτσι η διαδικασία εκπαίδευσης και αξιολόγησης των μοντέλων δεν είναι απλή με αποτέλεσμα την ανάγκη να γίνει μια πιο συστηματική προσέγγιση στο πρόβλημα της μοντελοποίησης. Συγκεκριμένα θα χρησιμοποιούνται προεπεξεργαστικά βήματα όπως η επιλογή χαρακτηριστικών (feature selection) για την μείωση των διαστάσεων, καθώς και μεθόδους βελτιστοποίησης των μοντέλων μέσω της διασταυρωμένης επικύρωσης (cross validation) έτσι ώστε να πετύχουμε ένα μοντέλο που να μπορεί να περιγράψει αρκετά καλά το σύστημα δεδομένου των μειωμένων χαρακτηριστικών.

Μέρος 1^ο

Περιγραφή

Το dataset το οποίο επιλέγεται είναι το Airfoil Self-Noise dataset από το UCI repository το οποίο περιλαμβάνει 1503 δείγματα και 6 χαρακτηριστικά. Αρχικά το σύνολο δεδομένων διαχωρίζεται σε σύνολα εκπαίδευσης (training), επικύρωσης (validation) και ελέγχου (checking). Πιο συγκεκριμένα χρησιμοποιείται το 60% των δειγμάτων για το σύνολο εκπαίδευσης ενώ από 20% των δειγμάτων σε κάθε ένα από τα εναπομείναντα υποσύνολα. Στη συνέχεια εκπαιδεύονται 4 TSK μοντέλα με διαφορετικές παραμέτρους. Πιο συγκεκριμένα μεταβάλλεται η μορφή εξόδου καθώς και το πλήθος των συναρτήσεων συμμετοχής για κάθε μεταβλητή εισόδου. Τα μοντέλα αυτά περιλαμβάνουν:

- 1) 2 συναρτήσεις συμμετοχής για κάθε μεταβλητή εισόδου και μια έξοδο τύπου Singleton
- 2) 3 συναρτήσεις συμμετοχής για κάθε μεταβλητή εισόδου και μια έξοδο τύπου Singleton
- 3) 2 συναρτήσεις συμμετοχής για κάθε μεταβλητή εισόδου και μια έξοδο τύπου Polynomial
- 4) 3 συναρτήσεις συμμετοχής για κάθε μεταβλητή εισόδου και μια έξοδο τύπου Polynomial

Οι συναρτήσεις συμμετοχής είναι bell-shaped και η αρχικοποίησή τους γίνεται με τέτοιον τρόπο ώστε τα διαδοχικά ασαφή σύνολα να παρουσιάζουν σε κάθε είσοδο βαθμό επικάλυψης περίπου 0.5. Και τα 4 μοντέλα εκπαιδεύονται με την υβριδική μέθοδο και τέλος γίνεται η αξιολόγηση των μοντέλων με βάση τους δείκτες απόδοσης MSE , $RMSE$, $NMSE$, $NDEI$, R^2 .

Αρχεία Matlab

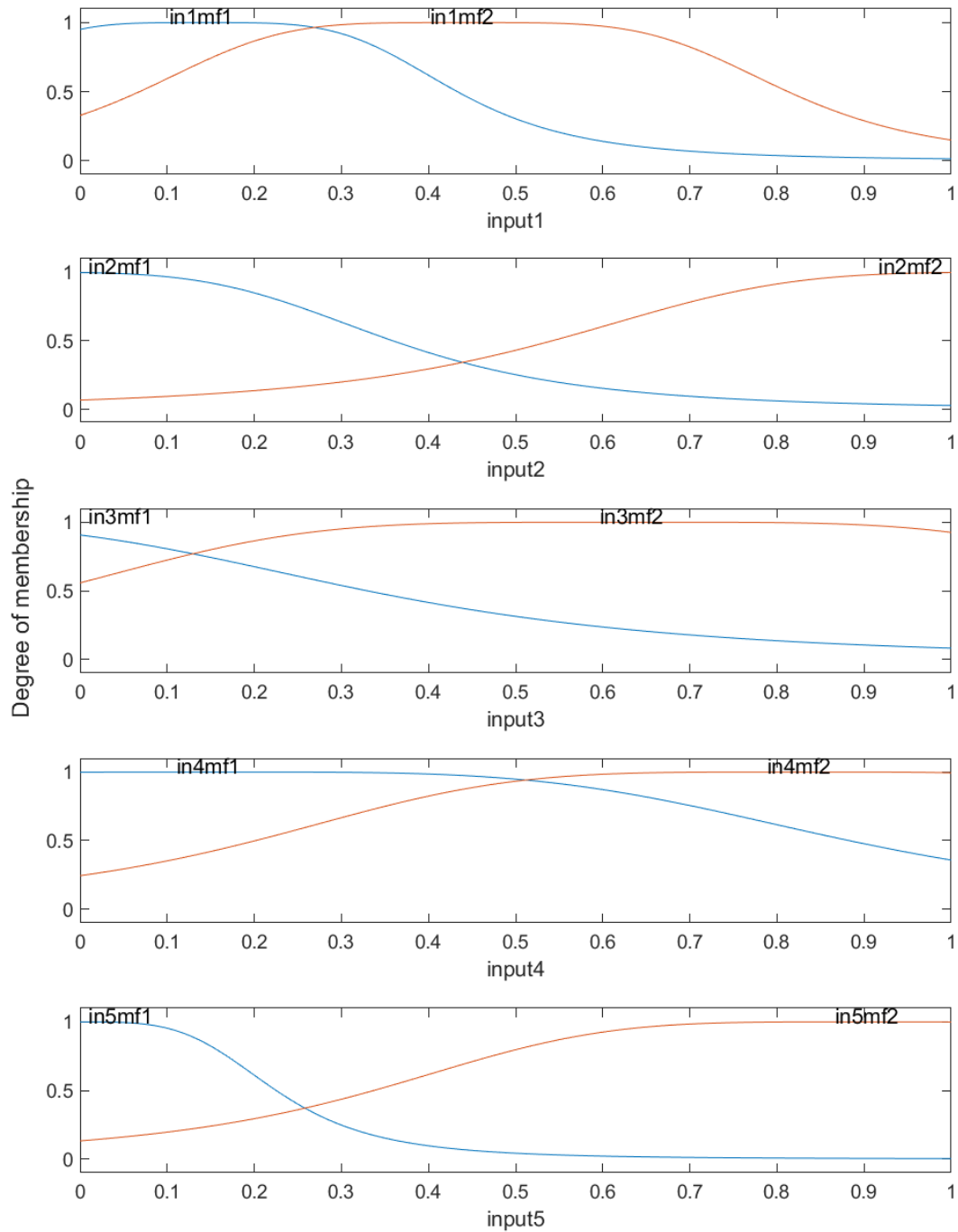
- **main_part_1.m**: Περιέχει το πρόγραμμα το οποίο με την χρήση των βοηθητικών συναρτήσεων που περιγράφονται παρακάτω, εκπαιδεύει και αξιολογεί τα 4 TSK μοντέλα.
- **TSK.m**: Παίρνοντας κατάλληλες παραμέτρους ως εισόδους εκπαιδεύει ένα TSK μοντέλο και επιστρέφει τις τιμές των δεικτών αξιολόγησης ενώ ταυτόχρονα παράγει τα ζητούμενα διαγράμματα.
- **plotErrors.m**: Παράγει τις καμπύλες εκμάθησης στις οποίες φαίνεται η μείωση του σφάλματος σε σχέση με τις εποχές.
- **plotMFs.m**: Παράγει τα διαγράμματα των συναρτήσεων συμμετοχής των εισόδων ενός TSK μοντέλου.
- **plotPrediction.m**: Παράγει τα διαγράμματα στα οποία απεικονίζονται τα σφάλματα πρόβλεψης.
- **split_scale.m**: Χωρίζει το σύνολο δεδομένων σε σύνολα εκπαίδευσης-επικύρωσης-ελέγχου ενώ ταυτόχρονα κανονικοποιεί τις μεταβλητές.

Διαγράμματα:

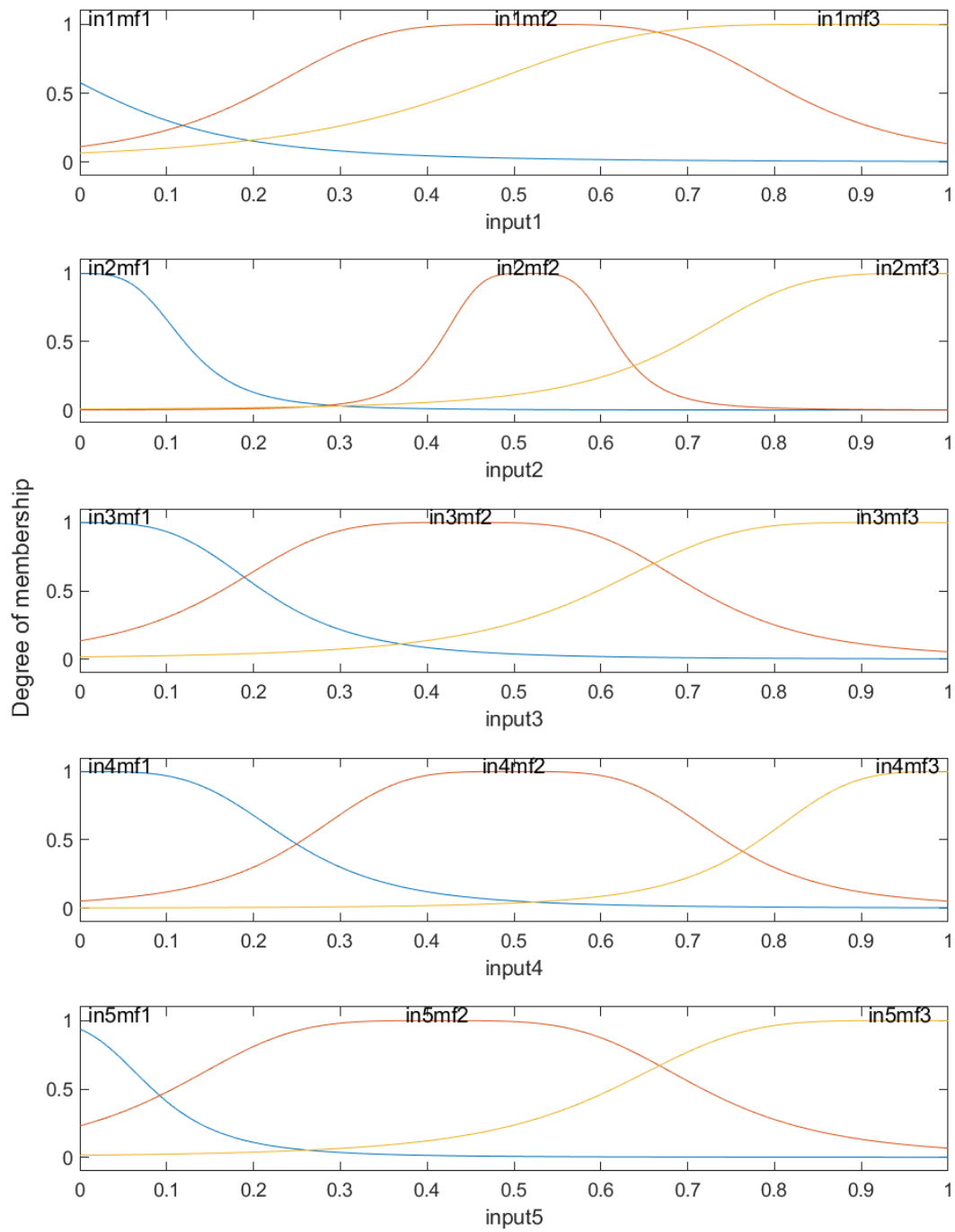
Παρακάτω εμφανίζονται 3 ειδών διαγράμματα για κάθε μια από τις 4 περιπτώσεις.

- i) Ασαφή σύνολα - Διαγράμματα στα οποία απεικονίζονται οι τελικές μορφές των ασαφών συνόλων που προέκυψαν μέσω της διαδικασίας εκπαίδευσης:
- ii) Διαγράμματα εκμάθησης – Φαίνονται οι καμπύλες εκμάθησης στις οποίες φαίνεται η μείωση του σφάλματος RMSE σε σχέση με τις εποχές εκπαίδευσης.
- iii) Διαγράμματα πρόβλεψης - Εδώ με μπλέ χρώμα βλέπουμε τις τιμές πρόβλεψης ενώ με κόκκινο χρώμα τις πραγματικές τιμές. Στον άξονα y έχουμε την τιμή της εξόδου ενώ στον άξονα x είναι το index του δείγματος που εξετάζουμε. Όπως παρατηρούμε ο συνολικός αριθμός των δειγμάτων είναι περίπου 300 δεδομένου ότι τα checking data που εξετάζουμε αποτελούν το 20% του συνολικού dataset όπου περιέχει 1503 δείγμα

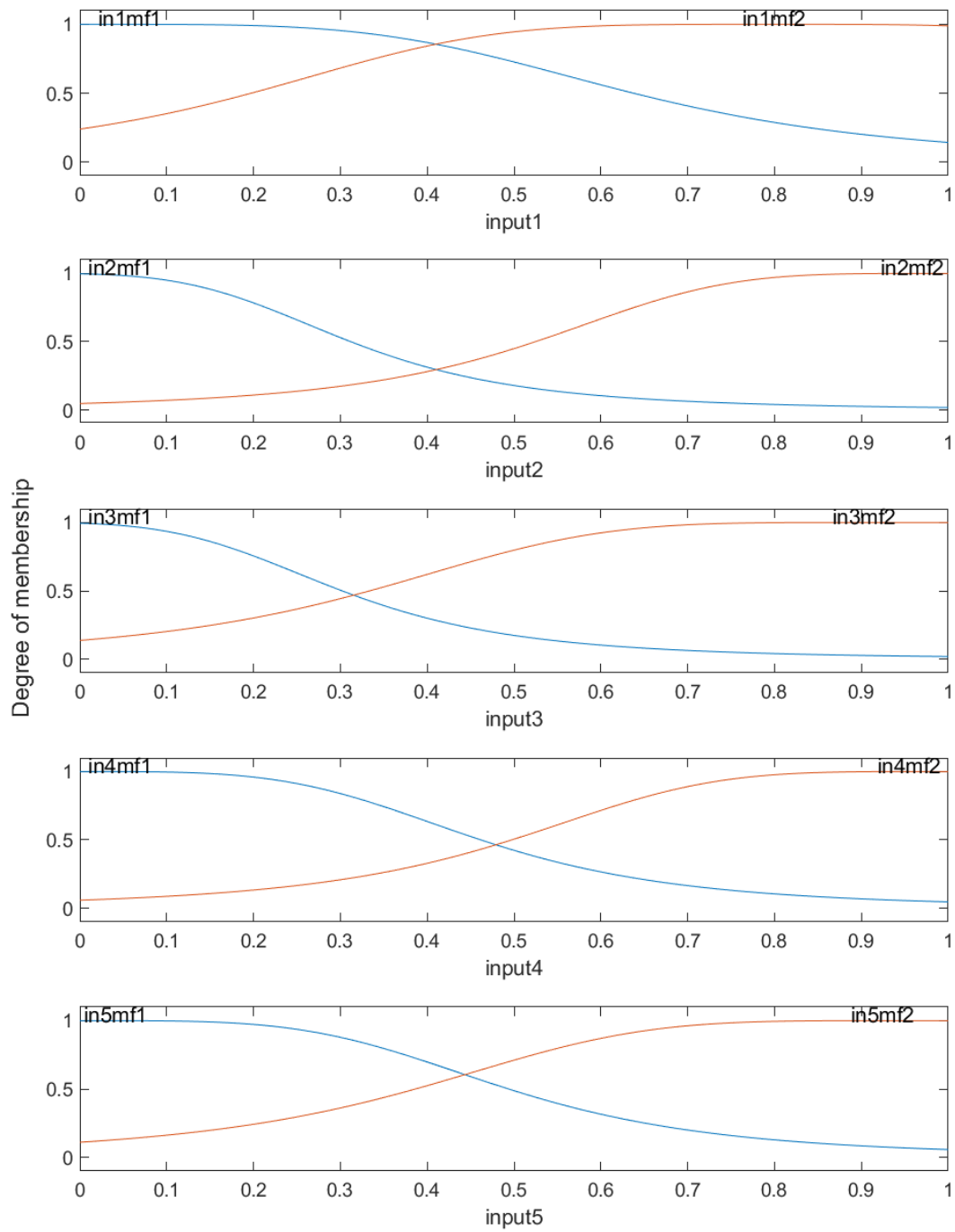
Ασαφή σύνολα:



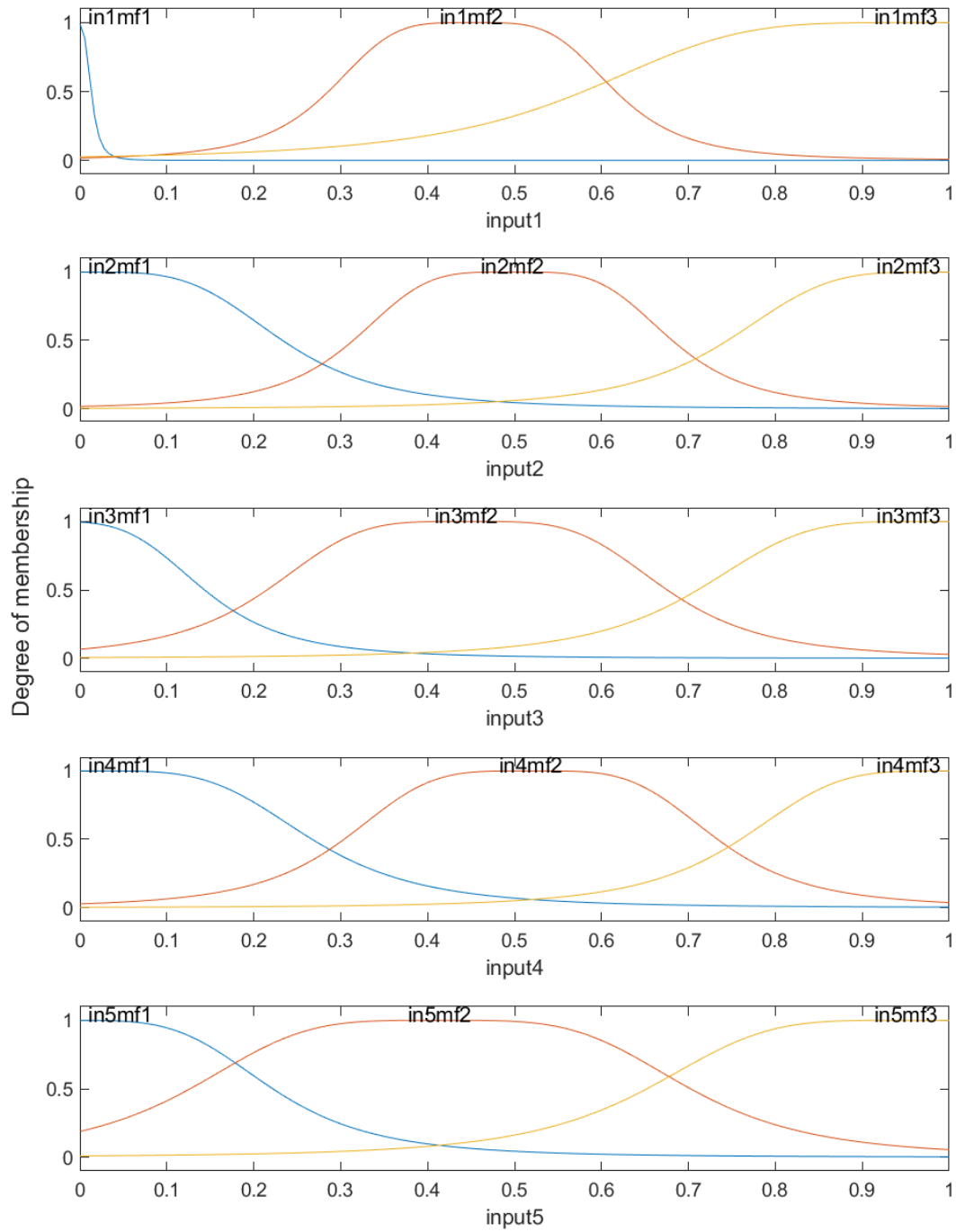
Διάγραμμα 1 - Singleton έξοδος με 2 συναρτήσεις συμμετοχής ανά είσοδο



Διάγραμμα 2 - Singleton έξοδος με 3 συναρτήσεις συμμετοχής ανά είσοδο

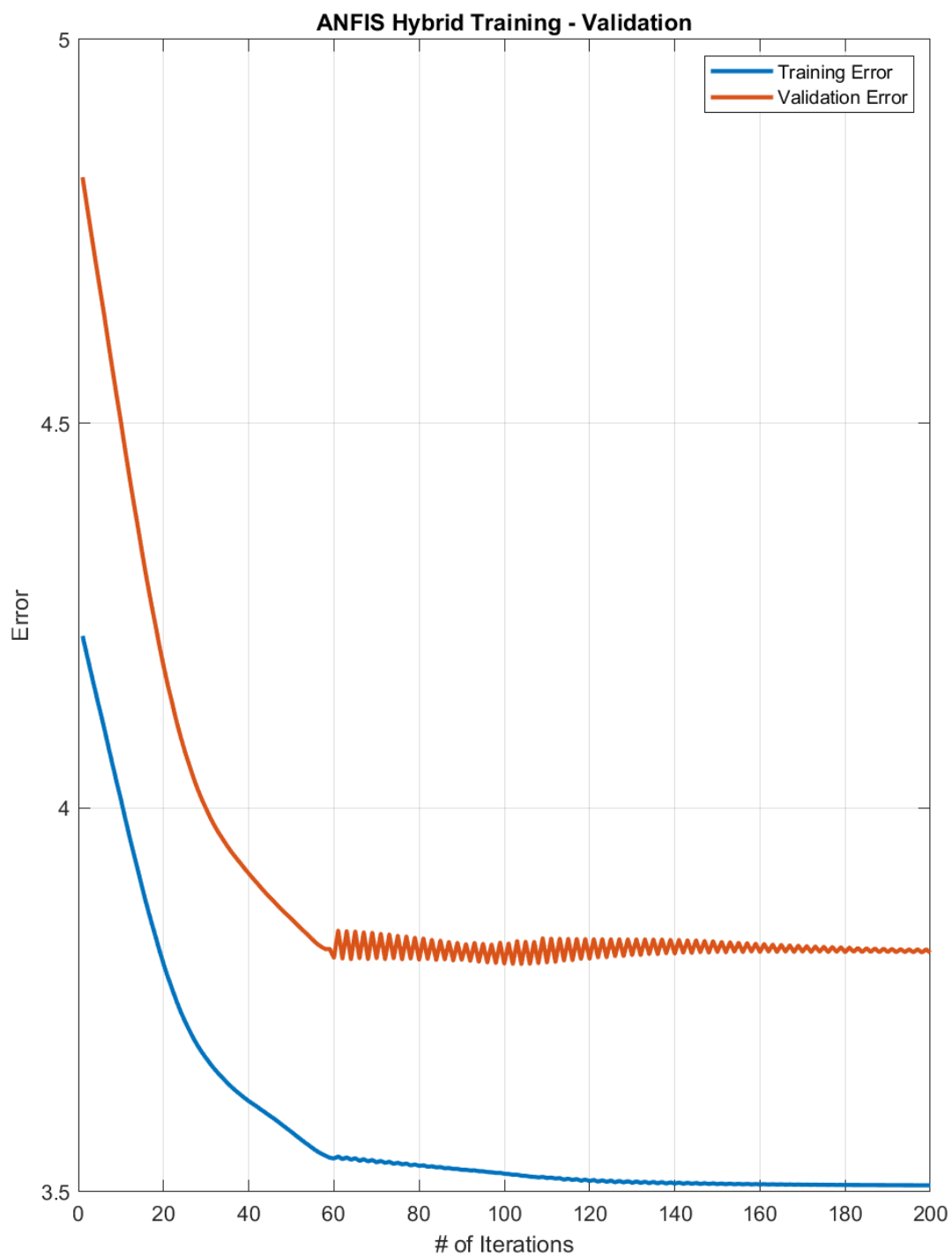


Διάγραμμα 3 - Polynomial έξοδος με 2 συναρτήσεις συμμετοχής ανά είσοδο

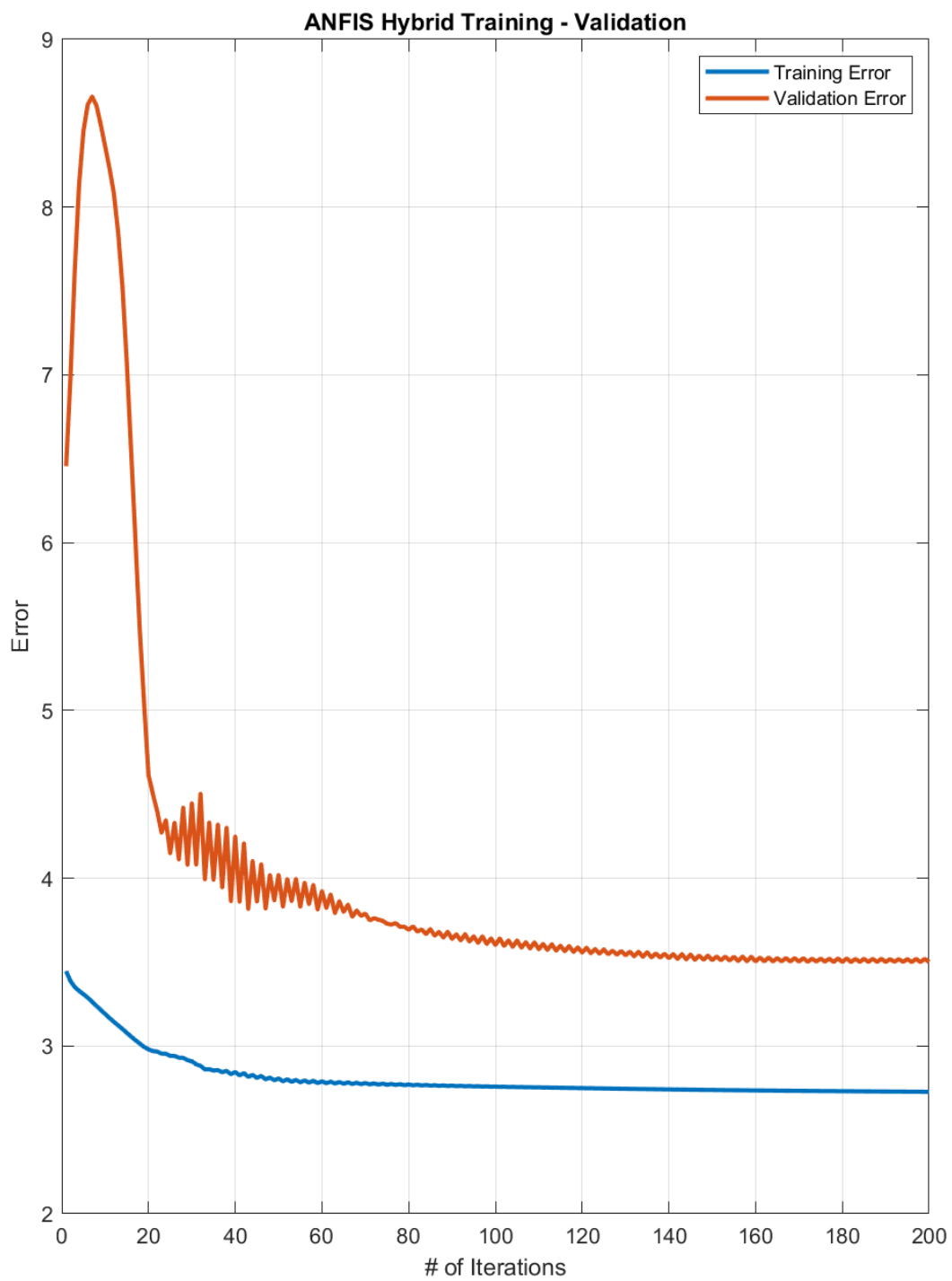


Διάγραμμα 4 - Polynomial έξοδος με 3 συναρτήσεις συμμετοχής ανά είσοδο

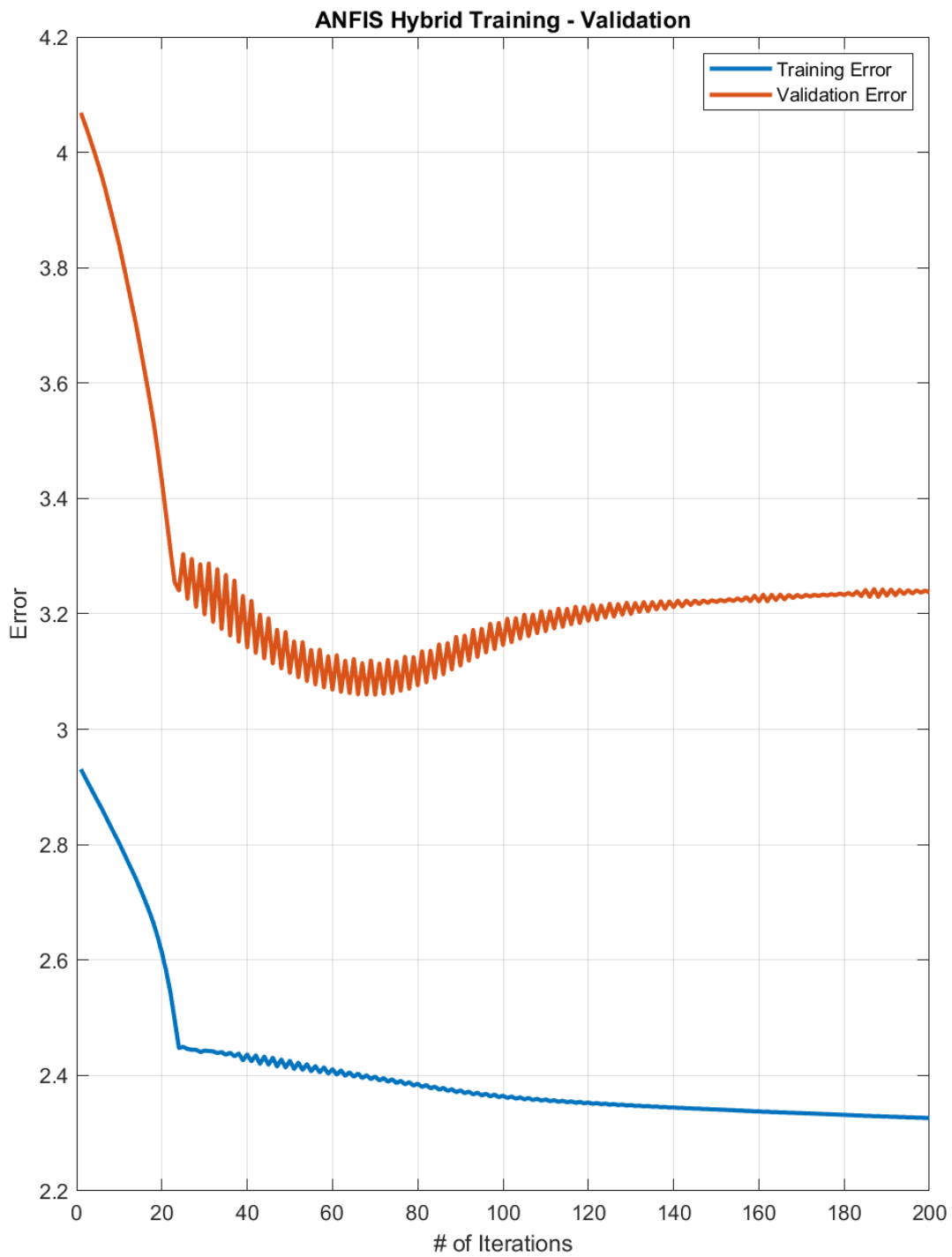
Διαγράμματα εκμάθησης:



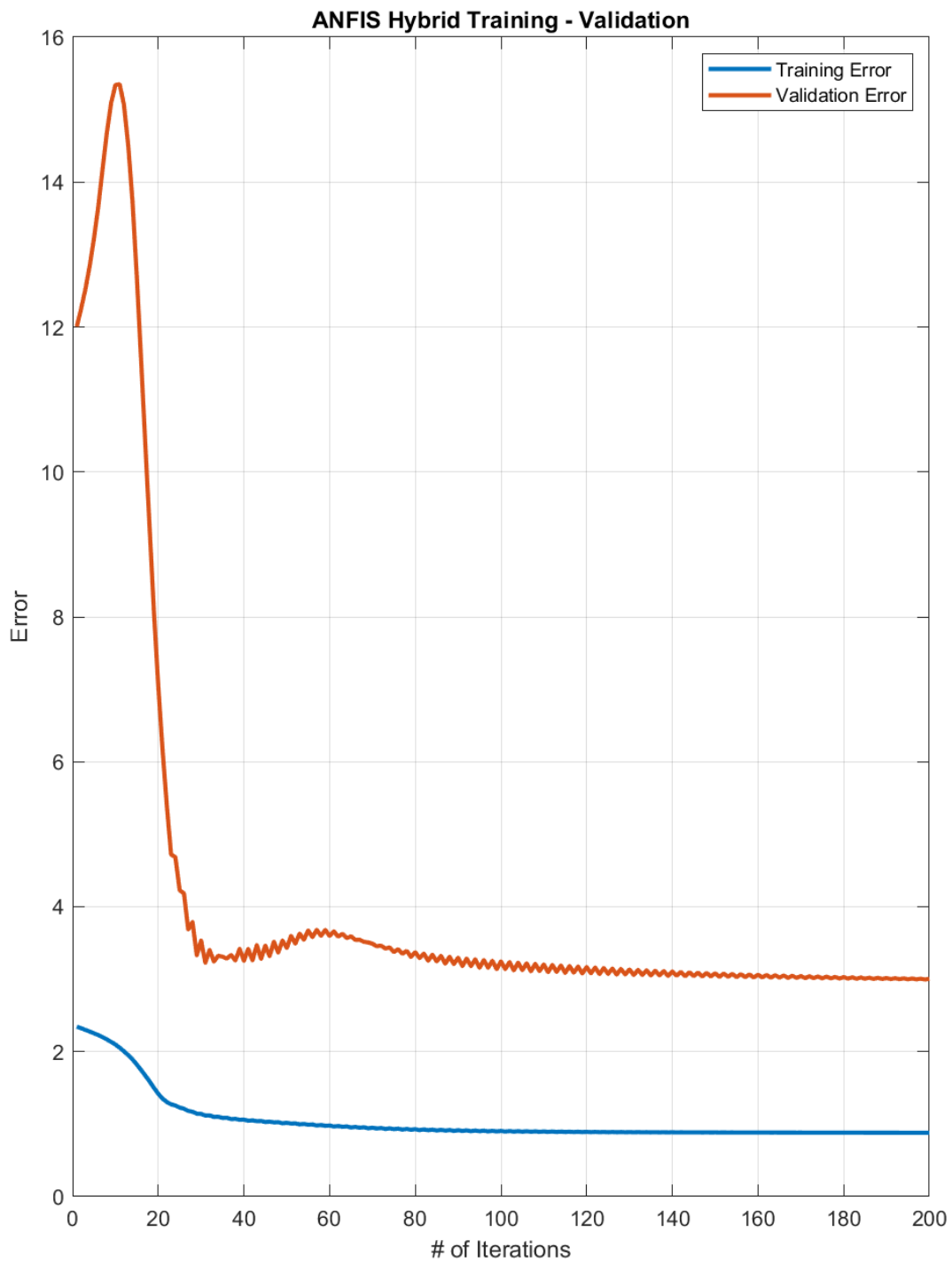
Διάγραμμα 5 - Εκμάθηση με Singleton έξοδο και 2 συναρτήσεις συμμετοχής ανά είσοδο



Διάγραμμα 6 - Εκμάθηση με Singleton έξοδο και 3 συναρτήσεις συμμετοχής ανά είσοδο

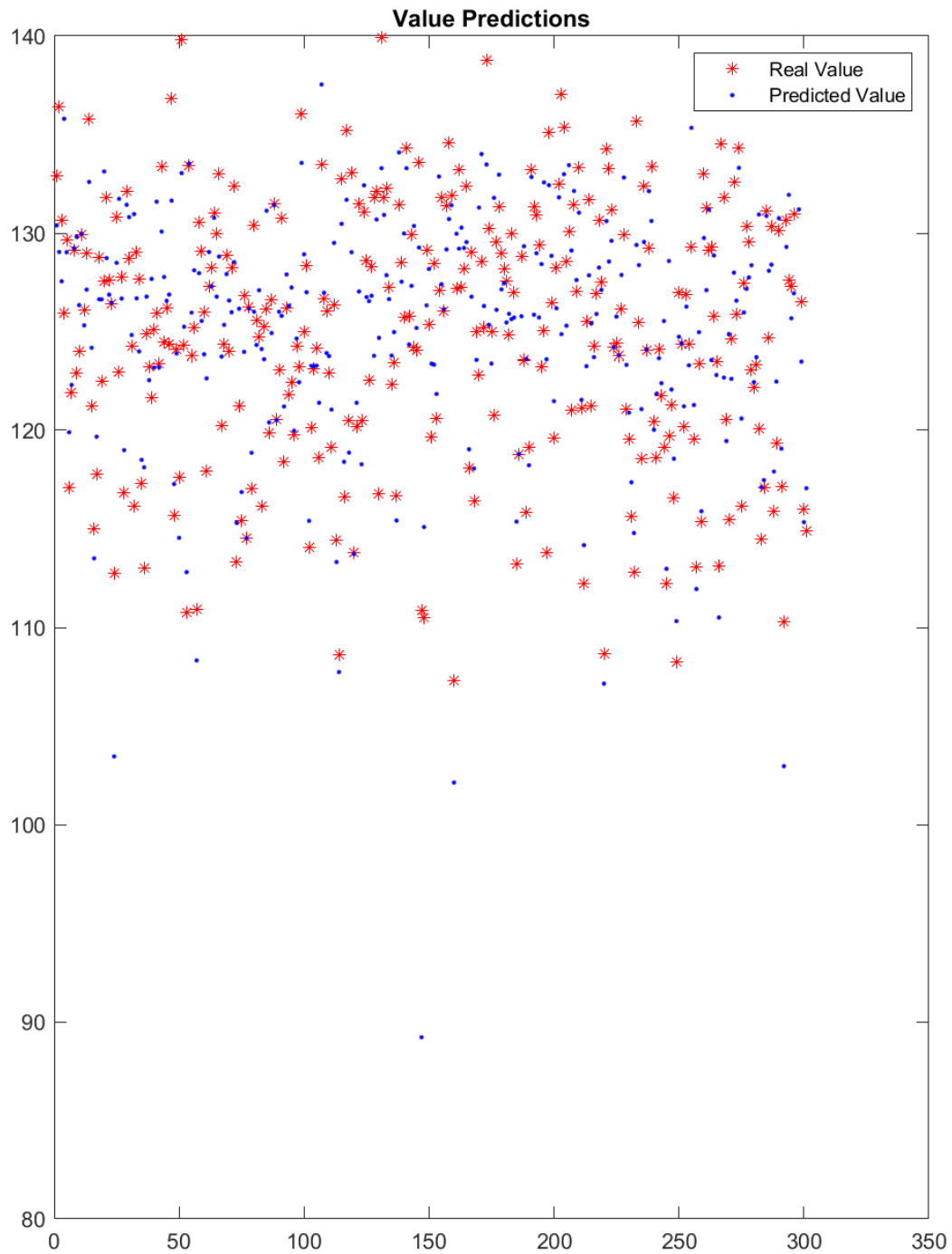


Διάγραμμα 7 - Εκμάθηση με *Polynomial* έξοδο και 2 συναρτήσεις συμμετοχής ανά είσοδο

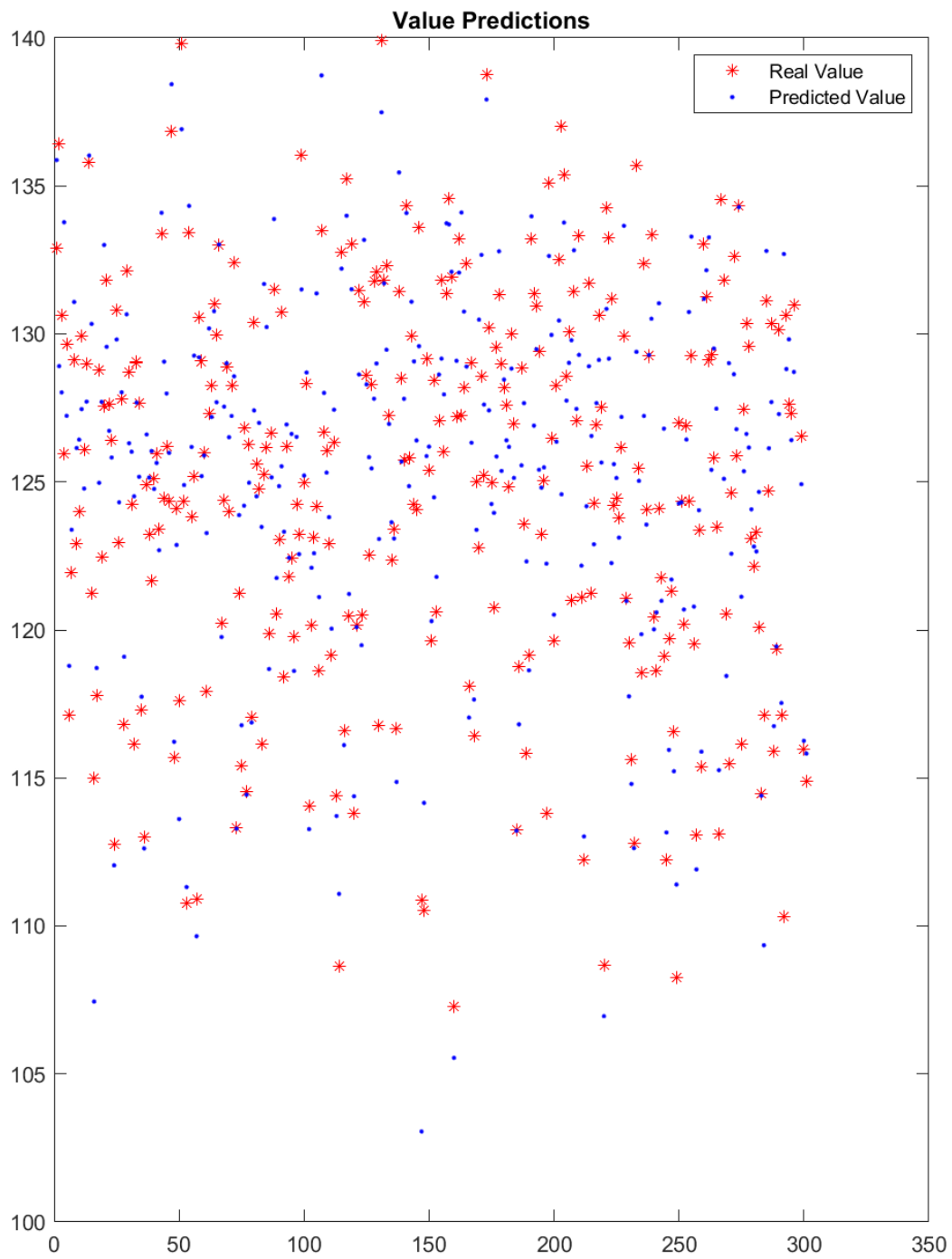


Διάγραμμα 8 - Εκμάθηση με Polynomial έξοδο και 3 συναρτήσεις συμμετοχής ανά είσοδο

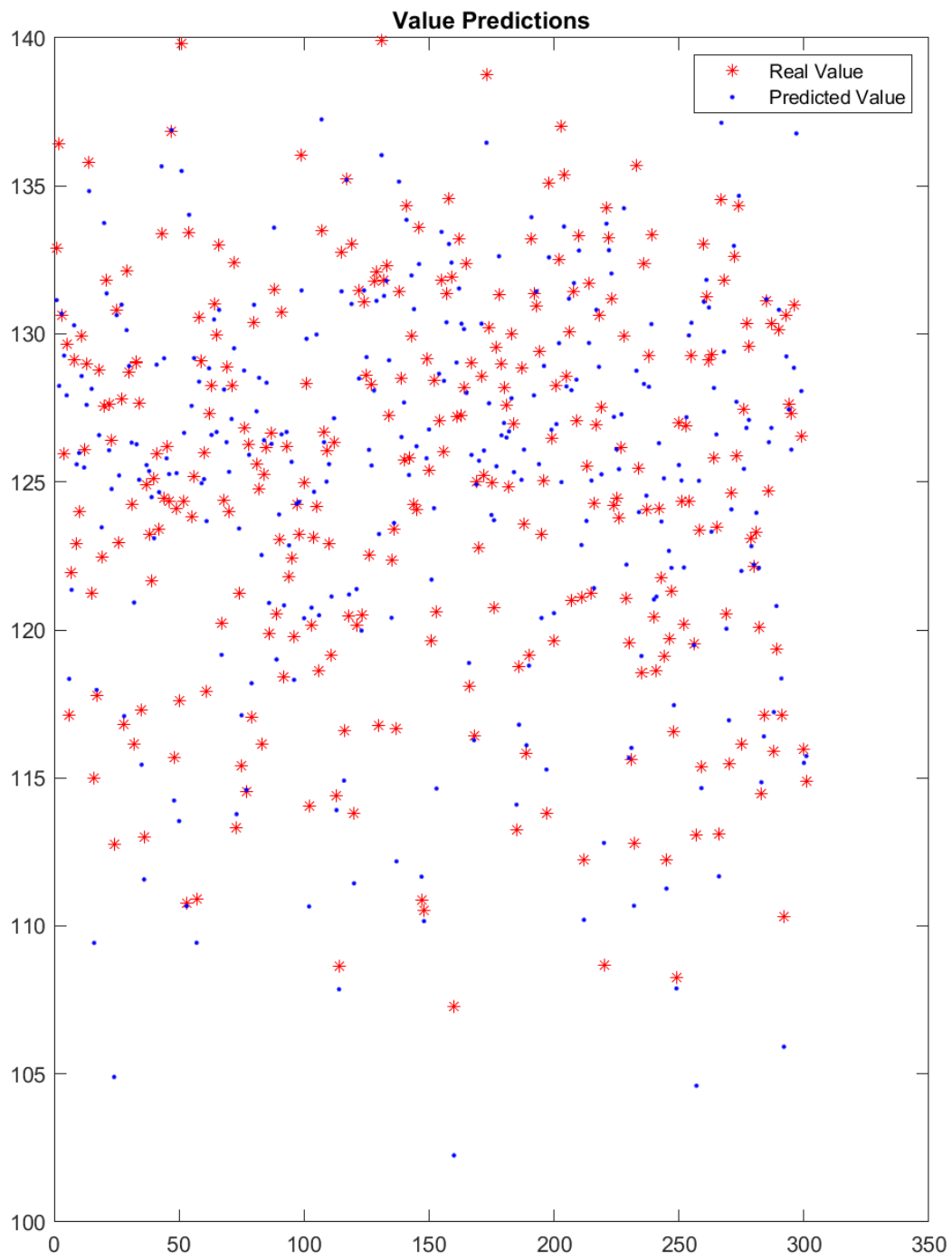
Διαγράμματα πρόβλεψης:



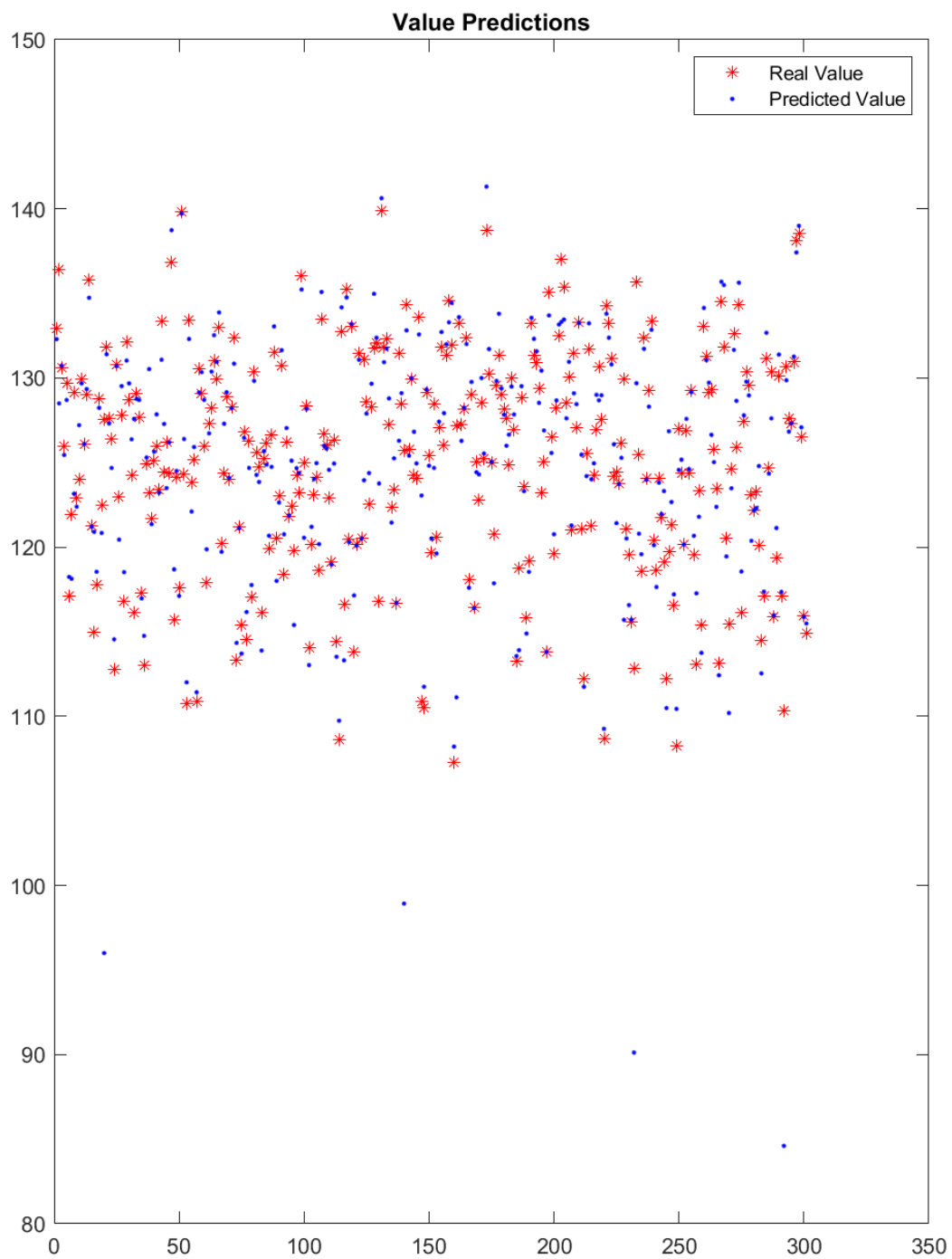
Διάγραμμα 9 - Πρόβλεψη με Singleton έξοδο και 2 συναρτήσεις συμμετοχής ανά είσοδο



Διάγραμμα 10 - Πρόβλεψη με Singleton έξοδο και 3 συναρτήσεις συμμετοχής ανά είσοδο



Διάγραμμα 11 - Πρόβλεψη με Polynomial έξοδο και 2 συναρτήσεις συμμετοχής ανά είσοδο



Διάγραμμα 12 - Πρόβλεψη με Polynomial έξοδο και 3 συναρτήσεις συμμετοχής ανά είσοδο

Αριθμητικές τιμές των δεικτών απόδοσης:

Έξοδος	Αριθμός συναρτήσεων συμμετοχής	MSE	RMSE	R2	NMSE	NDEI
Singleton	2	14.952	3.8667	0.65734	0.34266	0.58537
	3	11.763	3.4297	0.73042	0.26958	0.51924
Polynomial	2	7.2531	2.6932	0.83377	0.16623	0.40771
	3	15.951	3.9939	0.63443	0.36557	0.60462

Παρατηρήσεις

Όπως παρατηρούμε το μοντέλο με Singleton έξοδο και 3 συναρτήσεις συμμετοχής ανά είσοδο παρουσιάζει καλύτερη επίδοση από το μοντέλο με Singleton έξοδο και 2 συναρτήσεις συμμετοχής ανά είσοδο. Αυτό συμβαίνει επειδή 3 συναρτήσεις συμμετοχής έναντι 2 συναρτήσεων συμμετοχής στον χώρο εισόδου, σημαίνει καλύτερος διαχωρισμός των μεταβλητών στον χώρο και κατά συνέπεια μεγαλύτερη ακρίβεια που σημαίνει μικρότερο σφάλμα.

Επίσης το μοντέλο Polynomial έξοδο και 2 συναρτήσεις συμμετοχής ανά είσοδο παρουσιάζει καλύτερη επίδοση από το μοντέλο με Singleton έξοδο και 2 συναρτήσεις συμμετοχής ανά είσοδο. Αυτό οφείλεται στην χρήση πολυωνυμικής συνάρτησης στην έξοδο αντί Singleton, πράγμα που σημαίνει ότι έχουμε μεγαλύτερο αριθμό παραμέτρων τις οποίες ρυθμίζοντας κατάλληλα μας επιτρέπει μεγαλύτερη ακρίβεια έναντι της μιας καλά ρυθμισμένης παραμέτρου.

Το μοντέλο με Polynomial έξοδο και 3 συναρτήσεις συμμετοχής ανά είσοδο όμως παρόλο που αναμένεται για τους παραπάνω λόγους να παρουσιάζει την καλύτερη επίδοση τελικά παρουσιάζει χειρότερη επίδοση από τα υπόλοιπα. Αυτό συμβαίνει επειδή βλέποντας το διάγραμμα πρόβλεψης έχει προβλέψει κάποιες τιμές ως πολύ μικρότερες απ'ότι είναι στην πραγματικότητα και αυτό επιφέρει μεγάλη αλλαγή στο άθροισμα του $MSE = \frac{1}{N} \sum_{i=1}^N e_i^2$ αλλά και των υπολοίπων συντελεστών που φέρουν τον παραπάνω όρο. Αυτό είναι άμεση συνέπεια του overfitting του μοντέλου στα δεδομένα εκπαίδευσης και αυτό φαίνεται από την καμπύλη εκμάθησης όπου το training error παίρνει πολύ μικρές τιμές ακόμα και από τις πρώτες εποχές ενώ το validation error αρχικά παίρνει τεράστιες τιμές ενώ στη συνέχεια μειώνεται έως μια σχετικά μεγάλη τιμή (σε σχέση με το training error) και παραμένει σταθερό χωρίς περαιτέρω μείωση σε αυτή τη μεγάλη τιμή.

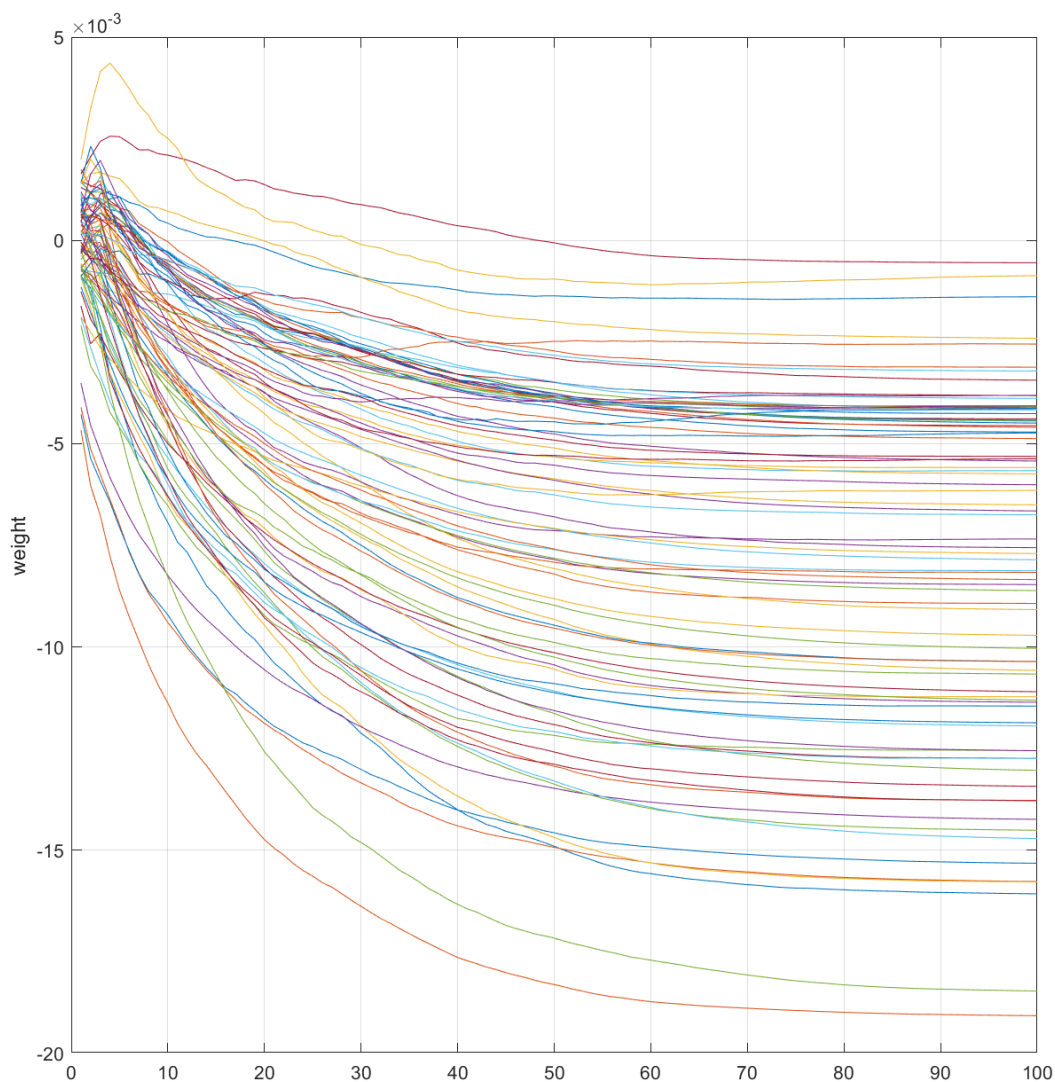
Μέρος 2°

Περιγραφή

Το dataset το οποίο επιλέγεται σε αυτήν την περίπτωση είναι το Superconductivity dataset από το UCI repository το οποίο περιλαμβάνει 21263 δείγματα και 81 χαρακτηριστικά. Αρχικά το σύνολο δεδομένων διαχωρίζεται όπως και στην προηγούμενη περίπτωση σε σύνολα εκπαίδευσης (training), επικύρωσης (validation) και ελέγχου (checking). Πιο συγκεκριμένα χρησιμοποιείται το 60% των δειγμάτων για το σύνολο εκπαίδευσης ενώ από 20% των δειγμάτων σε κάθε ένα από τα εναπομείναντα υποσύνολα. Το μέγεθος όμως του dataset καθιστά απαγορευτική μια απλή εφαρμογή ενός TSK μοντέλου όπως στο πρώτο μέρος. Για τον σκοπό αυτό χρησιμοποιούνται μέθοδοι για την μείωση των χαρακτηριστικών (feature selection) αλλά και ομαδοποίησης δεδομένων (subtractive clustering) ενώ ταυτόχρονα χρησιμοποιούνται και τεχνικές βελτιστοποίησης για την βελτιστοποίηση των παραμέτρων που μας επιφέρουν οι 2 παραπάνω μέθοδοι. Όλες αυτές οι μέθοδοι εφαρμόζονται στα δεδομένα εκπαίδευσης καθώς αυτά είναι τα μόνα τα οποία θέλουμε να δώσουν πληροφορίες στην δημιουργία του μοντέλου μας.

Αρχικά για την μείωση των χαρακτηριστικών χρειαζόμαστε μια μέθοδο αξιολόγησης αυτών με σκοπό μια καλύτερη επιλογή χαρακτηριστικών. Η μέθοδος η οποία χρησιμοποιείται είναι ο αλγόριθμος relief. Ο αλγόριθμος αυτός αρχικά θέτει όλα τα βάρη των χαρακτηριστικών στο 0. Στη συνέχεια επαναληπτικά επιλέγει μια τυχαία παρατήρηση (τυχαίο δείγμα) των δεδομένων και βρίσκει τις k κοντινότερες παρατηρήσεις. Έτσι σε κάθε επανάληψη «τιμωρεί» τα χαρακτηριστικά που έχουν διαφορετικές τιμές σε ίδιες τιμές απόκρισης ενώ «επιβραβεύει» τα χαρακτηριστικά που έχουν διαφορετικές τιμές σε διαφορετικές τιμές της απόκρισης. Τελικά μετά από έναν αριθμό επαναλήψεων και έχοντας τα τελικά βάρη των χαρακτηριστικών τα κατατάσσει με βάση αυτά. Η επιλογή του k όμως δεν είναι πάντοτε απλή υπόθεση. Μια μικρή τιμή του k μπορεί να δώσει αναξιόπιστα αποτελέσματα σε θορυβώδες δεδομένα (noisy data) ενώ με μια τιμή του k ανάλογη με τον αριθμό των δειγμάτων μπορεί η relief να αποτύχει να βρεί τα σημαντικά χαρακτηριστικά.

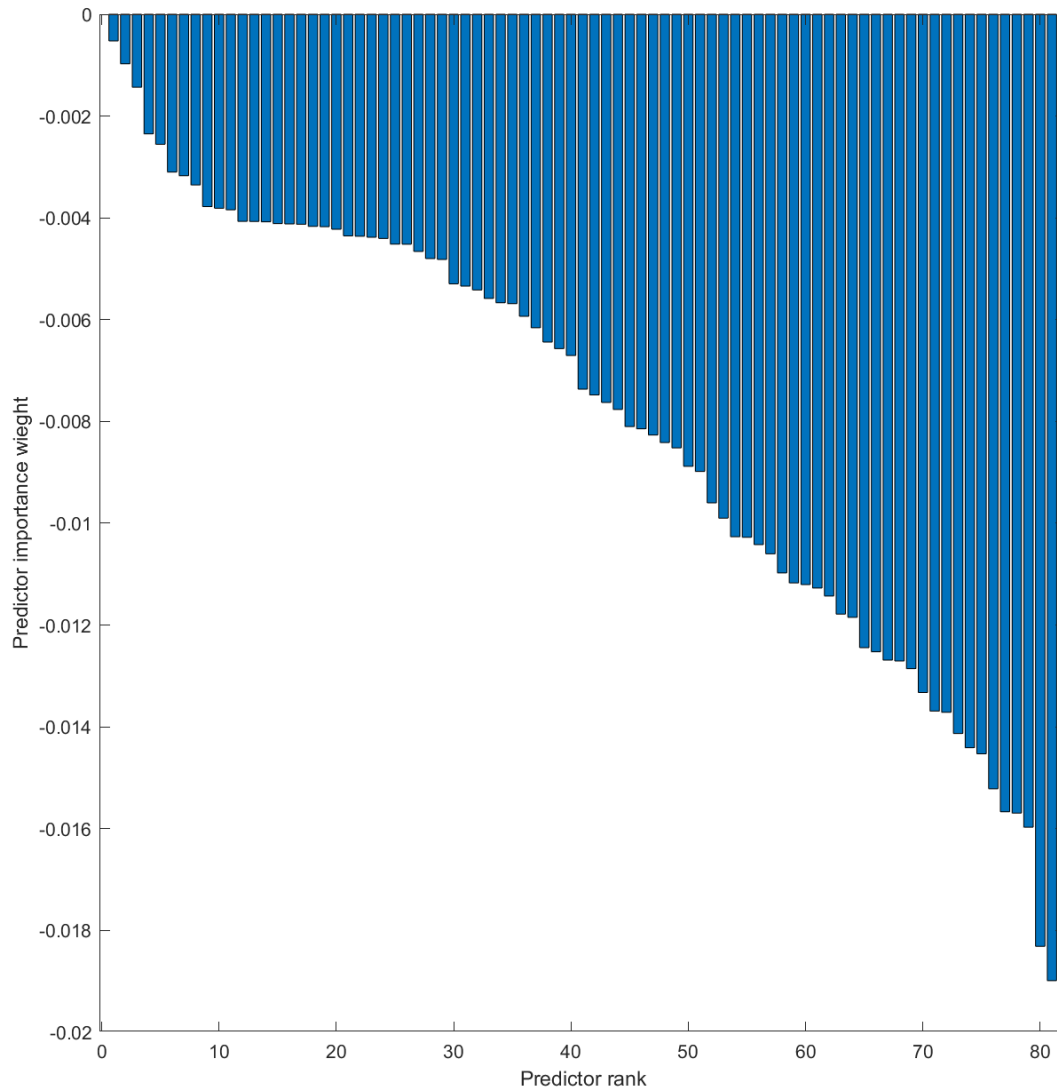
Έτσι για να βρούμε μια καλή τιμή του k φτιάξαμε τις γραφικές παραστάσεις των βαρών των χαρακτηριστικών συναρτήσει του k .



Διάγραμμα 13 - Μεταβολή βαρών των χαρακτηριστικών για διάφορες τιμές του k

Όπως παρατηρούμε μετά περίπου την τιμή $k = 80$ τα βάρη δεν αλλάζουν άλλο. Αυτό σημαίνει πως σε κάθε επανάληψη του αλγόριθμου relief, παίρνοντας υπόψη περισσότερες από 80 κοντινότερες παρατηρήσεις δεν επιφέρει αλλαγή στα βάρη των χαρακτηριστικών. Επιπλέον το 80 δεν είναι τιμή συγκρίσιμη με το μέγεθος των δεδομένων να ανησυχούμε για το φαινόμενο που αναφέρθηκε παραπάνω. Κατα συνέπεια το 80 είναι μια καλή τιμή του k και αυτή η τιμή χρησιμοποιήθηκε για την εύρεση των βέλτιστων χαρακτηριστικών με τον αλγόριθμο relief.

Έχοντας κατατάξει λοιπόν τα χαρακτηριστικά με βάση την σημαντικότητά τους αυτό που μας μένει είναι να βρούμε πόσα χαρακτηριστικά θα επιλέξουμε. Το διάγραμμα των βαρών των χαρακτηριστικών για τιμή $k = 80$ φαίνεται παρακάτω.

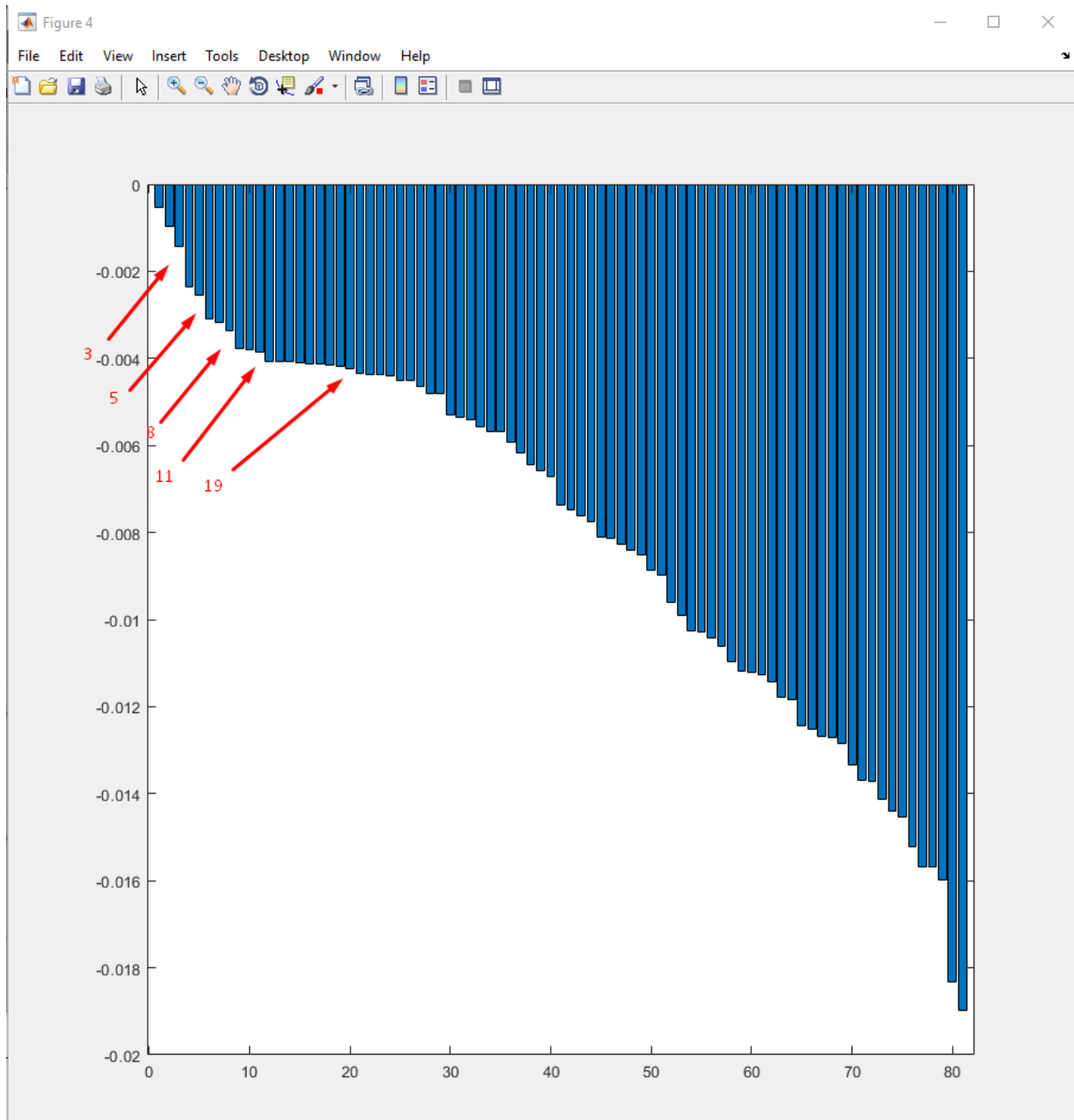


Διάγραμμα 14 - Βάρη των χαρακτηριστικών για $k = 80$

Όπως παρατηρούμε σε κάποια σημεία οι διαφορές των βαρών των χαρακτηριστικών είναι πολύ μικρή. Πολύ μικρή διαφορά βάρους μεταξύ δύο χαρακτηριστικών σημαίνει ίδια σημαντικότητα των 2 χαρακτηριστικών. Για τον λόγο αυτό εάν επιλέξουμε ένα από τα δυο χαρακτηριστικά των οποίων τα βάρη διαφέρουν λίγο μεταξύ τους αλλά δεν επιλέξουμε και το άλλο τότε θα υπήρχε σύγχυση προς την επιλογή αυτή καθώς θα ήταν τυχαία (τυχαία επιλογή μεταξύ των δυο αυτών χαρακτηριστικών στο τελικό σύνολο των χαρακτηριστικών). Έτσι για μια πιο συστηματική

προσέγγιση επιλέγουμε έναν αριθμό χαρακτηριστικών για τα οποία τα βάρη είναι πάνω από μια συγκεκριμένη τιμή. Αυτό σημαίνει πως στο παραπάνω σχήμα επιλέγουμε χαρακτηριστικά ώσπου να επιλέξουμε μια «μεγάλη» αλλαγή μεταξύ ενός χαρακτηριστικού και του επόμενου.

Ενδεικτικά η διαδικασία αυτή φαίνεται στο παρακάτω σχήμα:



Διάγραμμα 15 - Ενδεικτικές τιμές για επιλογή αριθμού χαρακτηριστικών

Έτσι ο αριθμός των χαρακτηριστικών που θα επιλέξουμε θα είναι μια από αυτές τις τιμές. Για την ακρίβεια στο πρόβλημα αυτό χρησιμοποιήθηκαν οι τιμές 3, 5, 8, 11, 19, 28.

Η άλλη μέθοδος η οποία χρησιμοποιήθηκε για την απλούστευση του προβλήματος είναι όπως αναφέρθηκε παραπάνω η ομαδοποίηση των δεδομένων η οποία έγινε μέσω του αλγόριθμου subtractive clustering. Η μέθοδος αυτή έχει σαν κύρια ελεύθερη παράμετρο την ακτίνα των ομάδων randii. Ο αλγόριθμος αυτός περιγράφεται πολύ συνοπτικά παρακάτω. Αρχικά όλα τα δεδομένα κατατάσσονται στον πολυδιάστατο χώρο. Στη συνέχεια βρίσκει το βέλτιστο κέντρο μιας ομάδας (ενός cluster) το οποίο είναι ένα σημείο από το σύνολο των δεδομένων. Για τον σκοπό αυτό χρησιμοποιεί έναν δείκτη (potential) για να αξιολογήσει το κάθε σημείο του συνόλου δεδομένων. Σημεία που έχουν περισσότερα σημεία κοντά τους έχουν μεγαλύτερο potential. Τα κοντινά αυτά σημεία που βρίσκονται εντός της ακτίνας randii επηρεάζουν πολύ το potential ενώ τα σημεία εκτός της ακτίνας αυτής το επηρεάζουν λίγο. Επίσης σημεία που βρίσκονται μακριά από άλλα κέντρα άλλων ομάδων έχουν μεγαλύτερο potential έτσι ώστε να αποφυγεί η αλληλοεπικάλυψη των cluster. Τελικά το σημείο με το μεγαλύτερο potential επιλέγεται ως κέντρο του cluster αυτού και το cluster αυτό έχει ακτίνα randii. Η διαδικασία αυτή συνεχίζεται ώσπου όλα τα δεδομένα να καταταχθούν σε μια ομάδα. Έτσι τελικά τα δεδομένα ομαδοποιούνται και το πρόβλημα απλουστεύεται καθώς καταλήγουμε σε λιγότερους κανόνες στο νέο ασαφή σύστημα που προκύπτει (1 κανόνας για κάθε cluster).

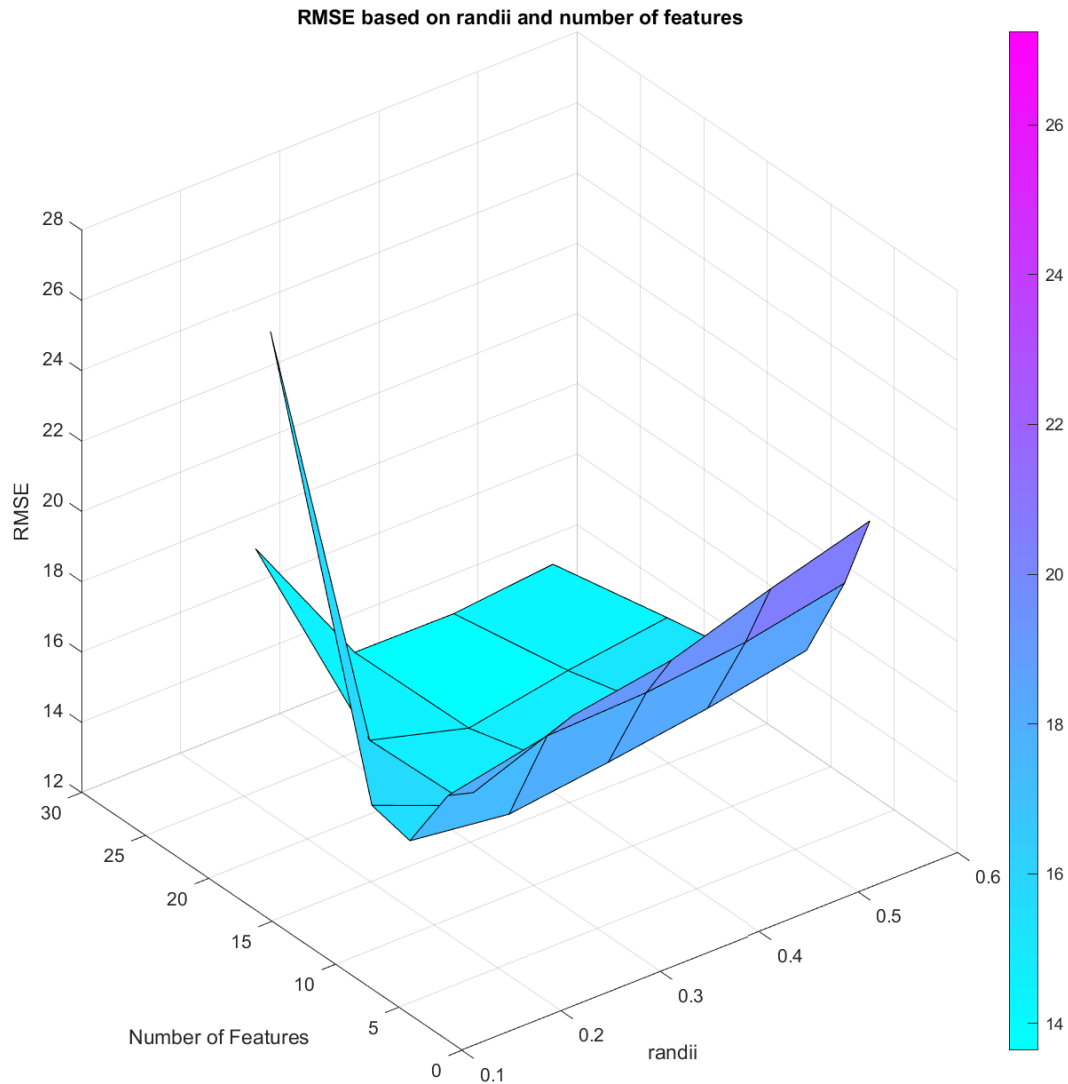
Στο πρόβλημα αυτό η επιλογή του randii έγινε από το σύνολο τιμών 0.15, 0.30, 0.45, 0.60.

Αφού λοιπόν έχοντας ορίσει τα σύνολα τιμών των δύο ελεύθερων παραμέτρων του προβλήματος, χρησιμοποιείται η αναζήτηση πλέγματος (grid search) για την αξιολόγηση των τιμών αυτών. Σύμφωνα με την τεχνική αυτή δημιουργούμε ένα διδιάστατο πλέγμα με τις τιμές των δύο ελεύθερων παραμέτρων μας. Κάθε σημείο του πλέγματος αυτού αντιστοιχεί σε μια επιλογή των δυο αυτών παραμέτρων. Για την αξιολόγηση του σημείου αυτού χρησιμοποιήθηκε η μέθοδος της διασταυρωμένης επικύρωσης (cross validation). Για την ακρίβεια χρησιμοποιήθηκε 5-fold cross validation. Σύμφωνα με τη μέθοδο αυτή το σύνολο εκπαίδευσης χωρίζεται στα 5. Το ένα κομμάτι αυτών των 5 συνόλων το οποίο αποτελείται από το 20% των δεδομένων εκπαίδευσης θα χρησιμοποιηθεί για επικύρωση ενώ τα υπόλοιπα κομμάτια (80%) για το σύνολο εκπαίδευσης του καινούριου συστήματος. Επειδή όμως θέλουμε όσο πιο γενικά αποτελέσματα γίνεται, αντί να χρησιμοποιήσουμε τυχαία ένα κομμάτι για το σύνολο επικύρωσης, χρησιμοποιούμε μια φορά όλα τα κομμάτια (ενώ το υπόλοιπο 80% πάντα λαμβάνεται ως το σύνολο εκπαίδευσης) και στο τέλος παίρνουμε τον μέσο όρο των 5 αυτών περιπτώσεων.

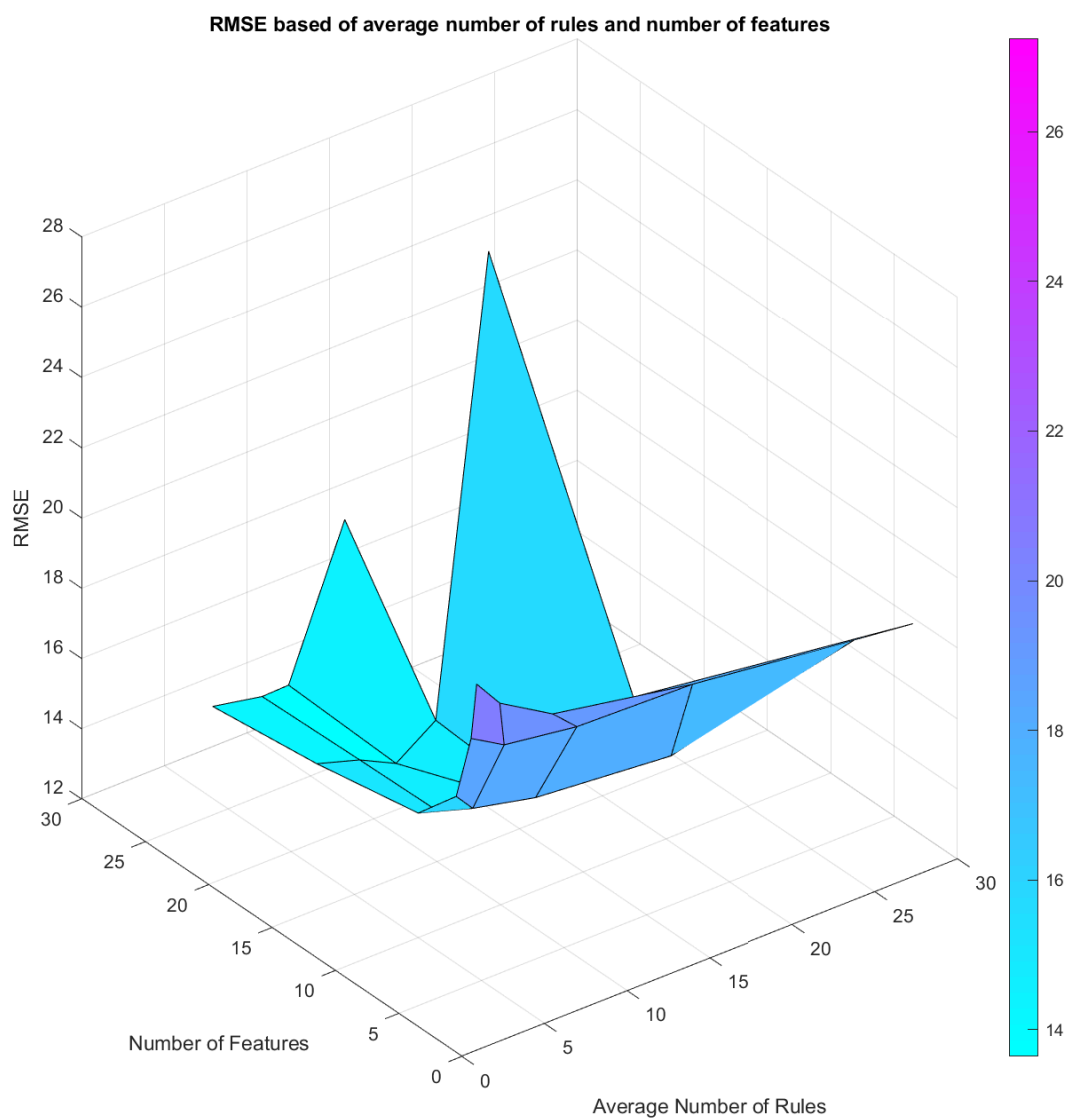
Έτσι σε κάθε μια περίπτωση αφού εκπαιδεύσουμε το μοντέλο και βρούμε το σφάλμα του, το κρατάμε και βρίσκουμε το μέσο όρο του. Αυτό γίνεται για κάθε σημείο του πλέγματος που έχουμε δημιουργήσει και τέλος επιλέγουμε την δυάδα παραμέτρων η οποία επιφέρει το μικρότερο μέσο σφάλμα.

Τα αποτελέσματα της διαδικασίας αυτής φαίνονται παρακάτω:

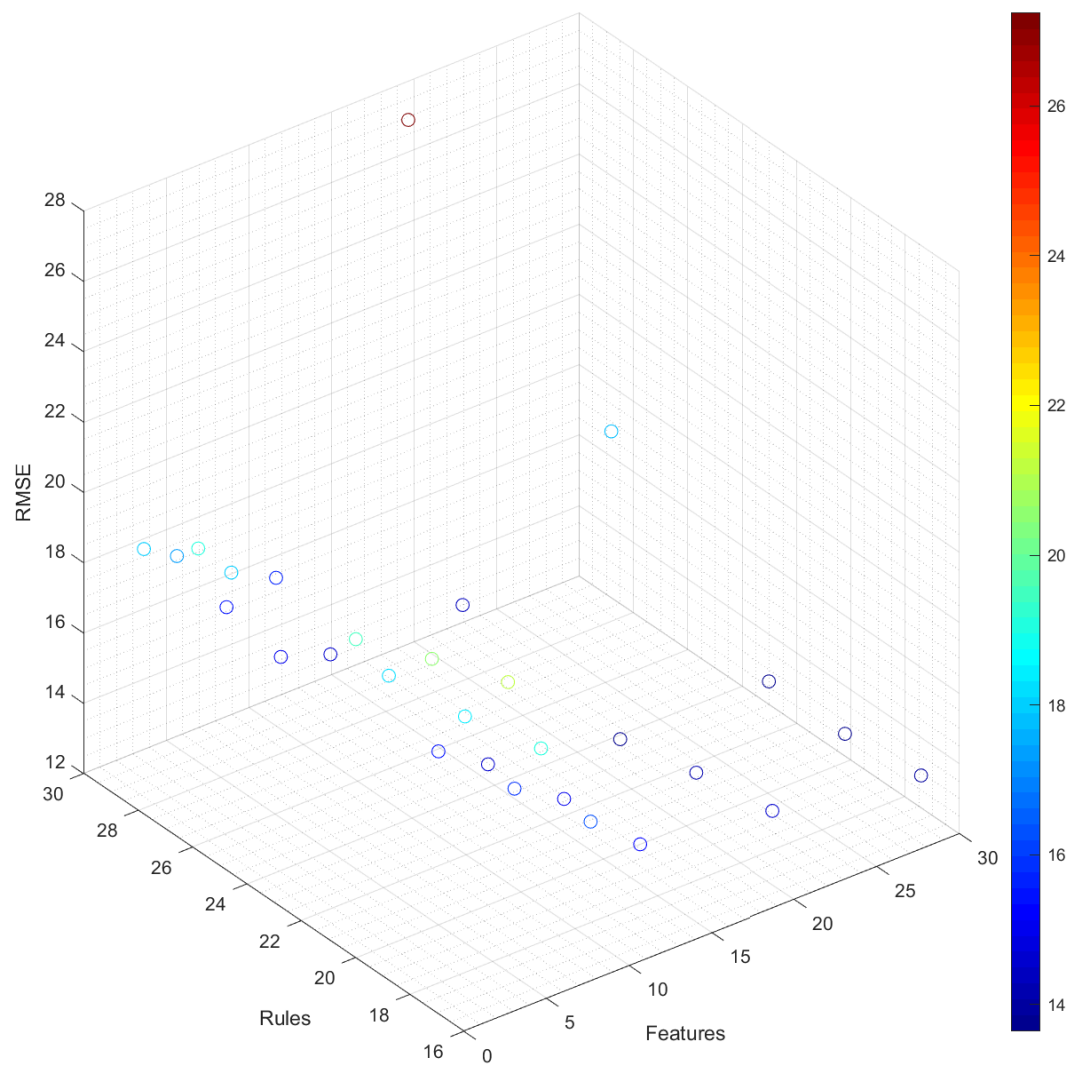
Στο πρώτο διάγραμμα παρουσιάζεται το μέσο σφάλμα RMSE συναρτήσει των 2 παραμέτρων ενώ στο δεύτερο διάγραμμα παρουσιάζεται το μέσο σφάλμα RMSE συναρτήσει του μέσου αριθμού των κανόνων και του αριθμού των επιλεγθέντων χαρακτηριστικών. Τέλος στο τρίτο διάγραμμα φαίνεται μια άλλη οπτική του δεύτερου διαγράμματος:



Διάγραμμα 16 - RMSE συναρτήσει του αριθμού των χαρακτηριστικών και της ακτίνας των cluster



Διάγραμμα 17 - RMSE συναρτήσει του αριθμού των χαρακτηριστικών και του μέσου αριθμού των κανόνων (1)



Διάγραμμα 18 - *

Παρατηρούμε όπως ήταν αναμενόμενο πως κατά κανόνα όσο αυξάνεται ο αριθμός των χαρακτηριστικών αλλά και όσο μειώνεται η ακτίνα των clusters randii (και παράλληλα ο αριθμός των κανόνων) τόσο μειώνεται και το σφάλμα. Η βέλτιστη λύση βρέθηκε για αριθμό χαρακτηριστικών ίσο με 28 και ακτίνα randii ίση με 0.45. Επειδή όμως ο αριθμός των χαρακτηριστικών αυτών είναι ακόμη μεγάλος και το ζητούμενο είναι μια πιο απλουστευμένη λύση, το σύνολο τιμών του αριθμού των χαρακτηριστικών τελικά θεωρήθηκε ίσο με 3, 5, 8, 11. Σε αυτήν την περίπτωση η βέλτιστη λύση βρέθηκε για αριθμό χαρακτηριστικών ίσο με 11 και ακτίνα randii ίση με 0.30.

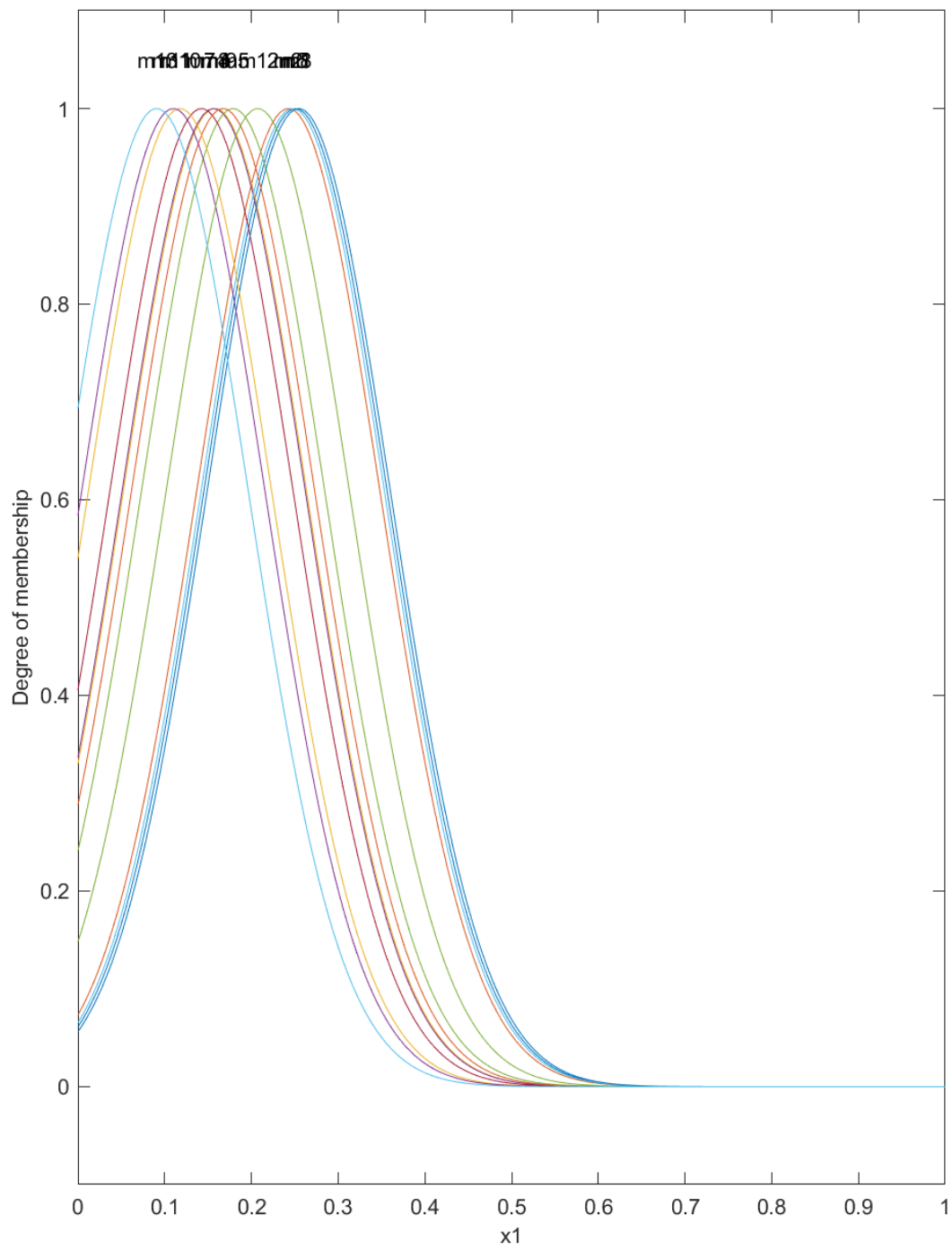
Αρχεία Matlab

- **main_part_2.m**: Περιέχει το πρόγραμμα το οποίο με την χρήση των βοηθητικών συναρτήσεων εφαρμόζει τη διαδικασία που περιγράφηκε παραπάνω και τελικά εκπαιδεύει και αξιολογεί το τελικό TSK μοντέλο.
- **plotErrors.m**: Παράγει τις καμπύλες εκμάθησης στις οποίες φαίνεται η μείωση του σφάλματος σε σχέση με τις εποχές.
- **plotMFs.m**: Παράγει τα διαγράμματα των συναρτήσεων συμμετοχής των εισόδων ενός TSK μοντέλου.
- **plotPrediction.m**: Παράγει τα διαγράμματα στα οποία απεικονίζονται τα σφάλματα πρόβλεψης.
- **split_scale.m**: Χωρίζει το σύνολο δεδομένων σε σύνολα εκπαίδευσης-επικύρωσης-ελέγχου ενώ ταυτόχρονα κανονικοποιεί τις μεταβλητές.
- **Plot_RMSE_surf.m**: Παράγει τα διαγράμματα τα οποία απεικονίζουν τα αποτελέσματα της αναζήτησης πλέγματος με τη χρήση της συνάρτησης surf.
- **Plot_RMSE_scatter3.m**: Παράγει τα διαγράμματα τα οποία απεικονίζουν τα αποτελέσματα της αναζήτησης πλέγματος με τη χρήση της συνάρτησης scatter3.
- **Best_K.m**: Υπολογίζει τα βάρη των χαρακτηριστικών με τη συνάρτηση relief συναρτήσει του k και παράγει τα ανάλογα διαγράμματα χρησιμοποιώντας την **plots_K.m**.

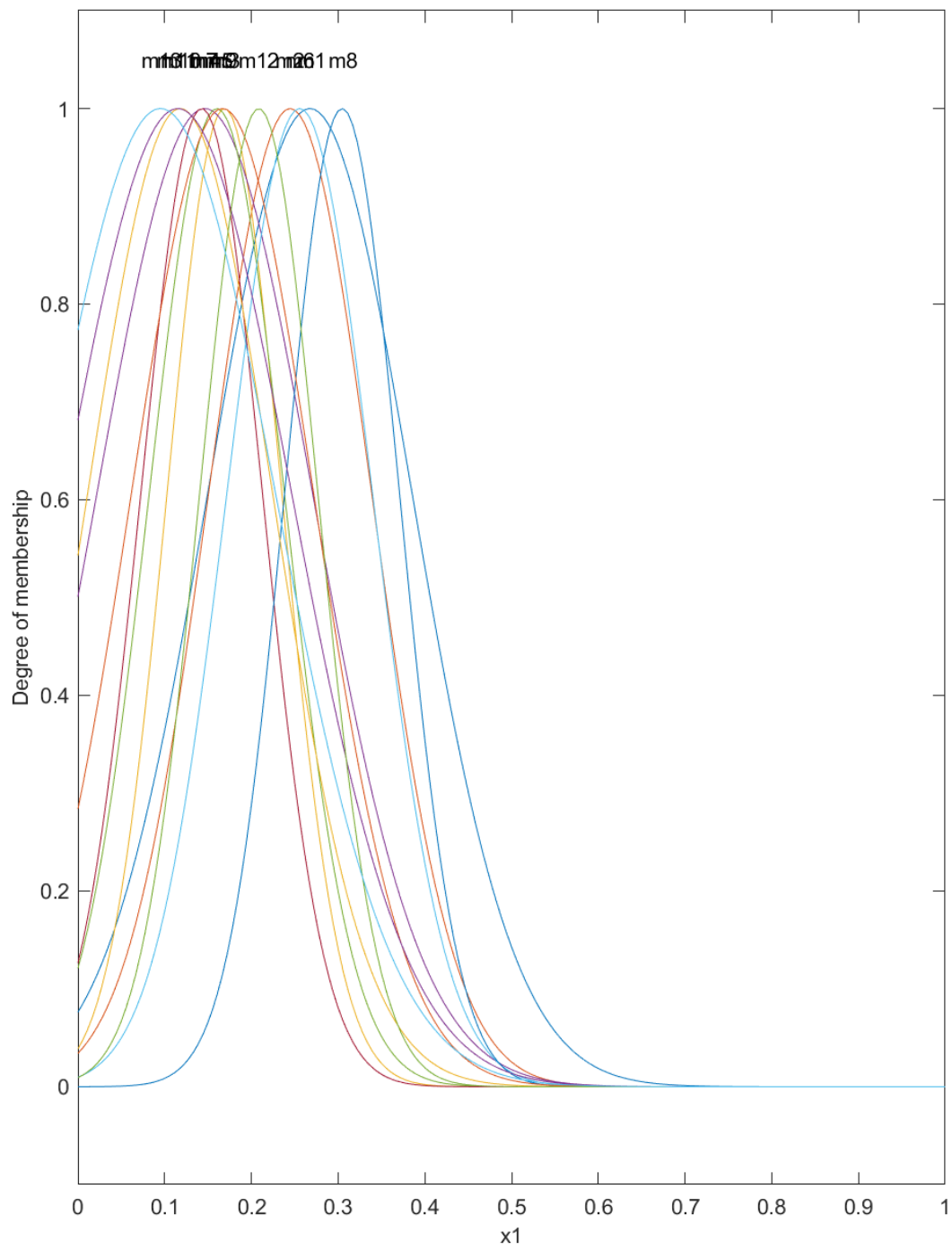
Διαγράμματα

Όπως και στο πρώτο μέρος παίρνουμε τα παρακάτω διαγράμματα τα οποία για ευκολία περιγράφονται ξανά:

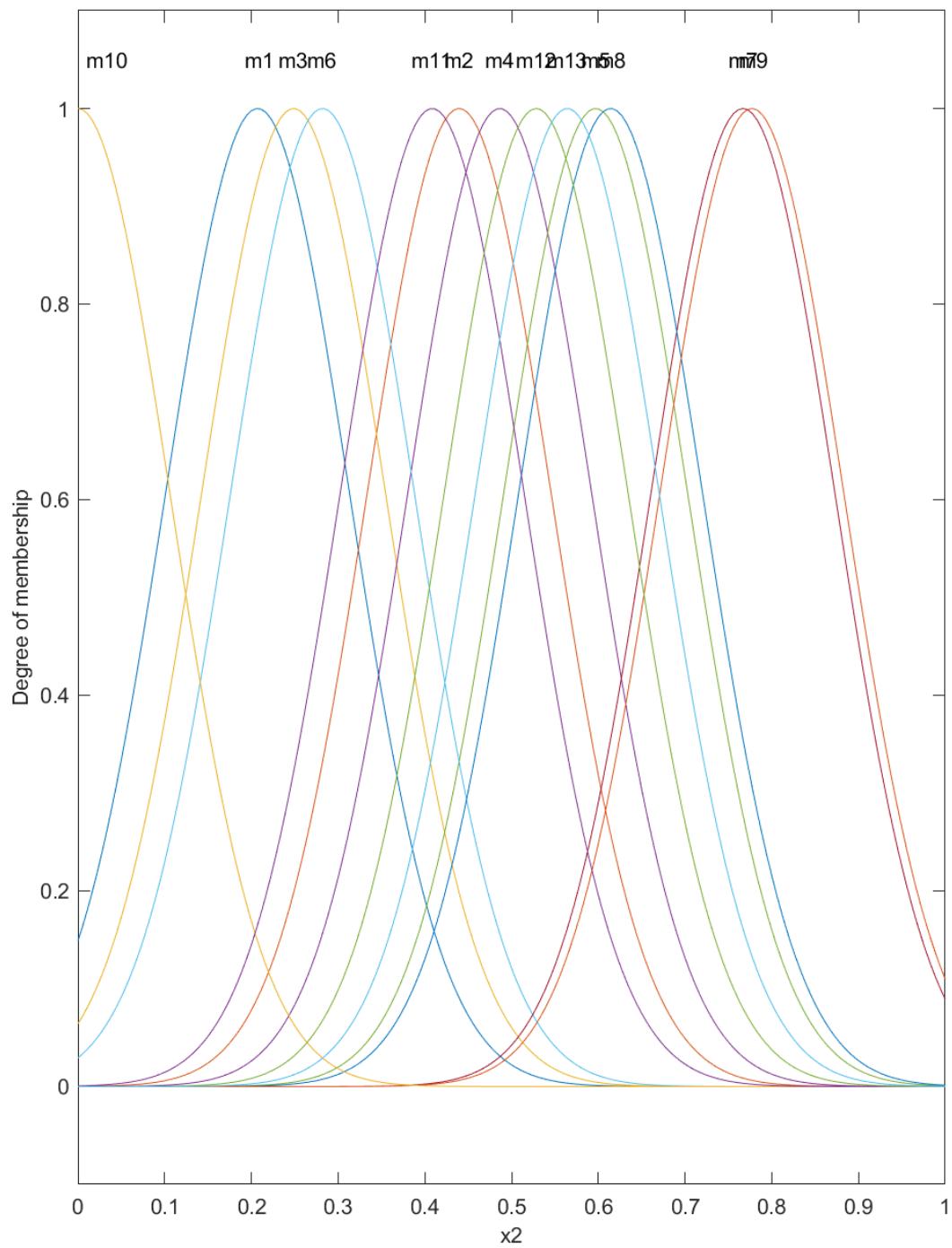
- i) Ασαφή σύνολα – Ενδεικτικά διαγράμματα στα οποία απεικονίζονται οι αρχικές αλλά και οι τελικές μορφές των ασαφών συνόλων που προέκυψαν μέσω της διαδικασίας εκπαίδευσης:
- ii) Διάγραμμα εκμάθησης – Φαίνονται οι καμπύλες εκμάθησης στις οποίες φαίνεται η μείωση του σφάλματος RMSE σε σχέση με τις εποχές εκπαίδευσης.
- iii) Διάγραμμα πρόβλεψης - Εδώ με μπλέ χρώμα βλέπουμε τις τιμές πρόβλεψης ενώ με κόκκινο χρώμα τις πραγματικές τιμές. Στον άξονα y έχουμε την τιμή της εξόδου ενώ στον άξονα x είναι το index του δείγματος που εξετάζουμε.



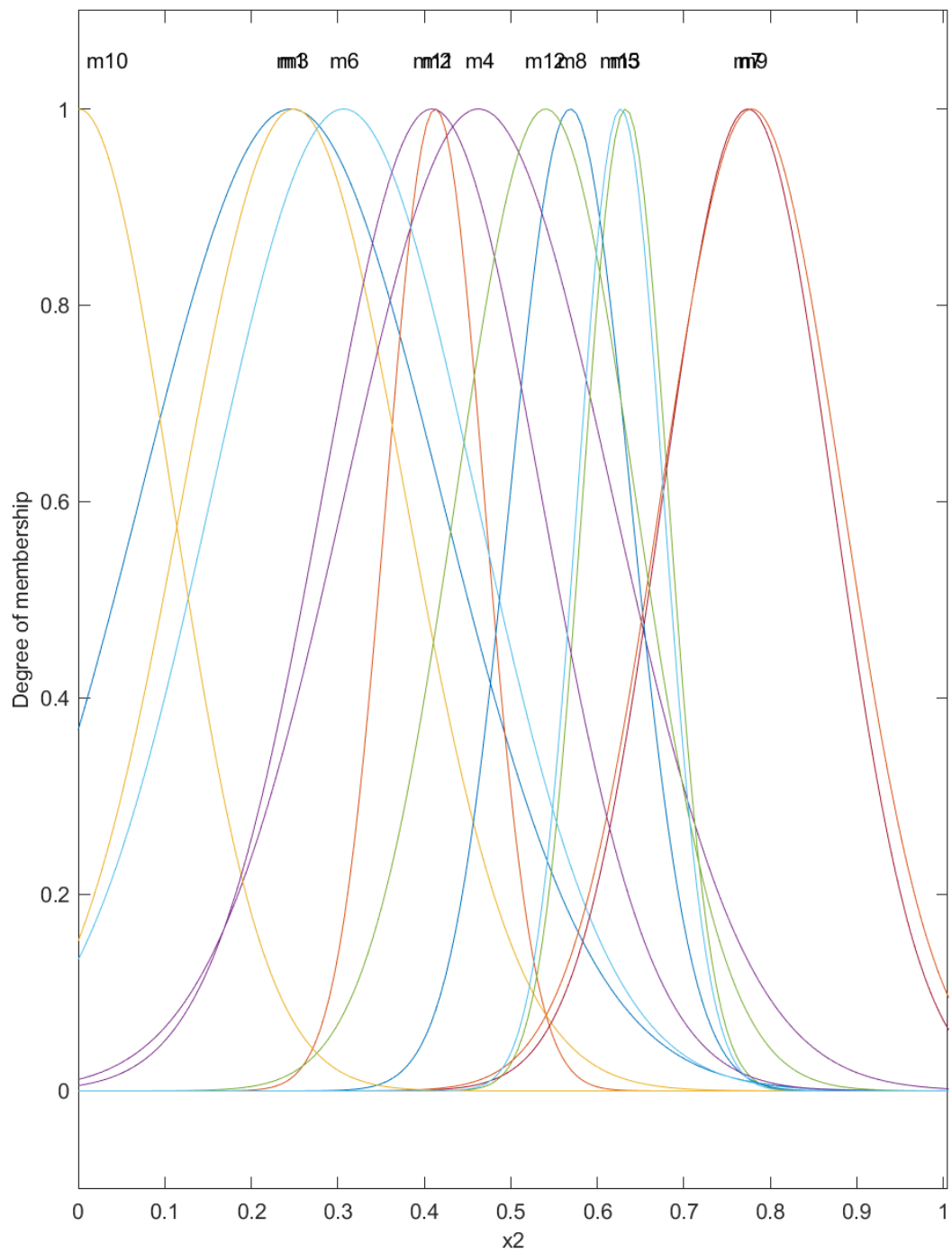
Διάγραμμα 19 - Ασαφή σύνολα σημαντικότερου χαρακτηριστικού πριν την εκπαίδευση



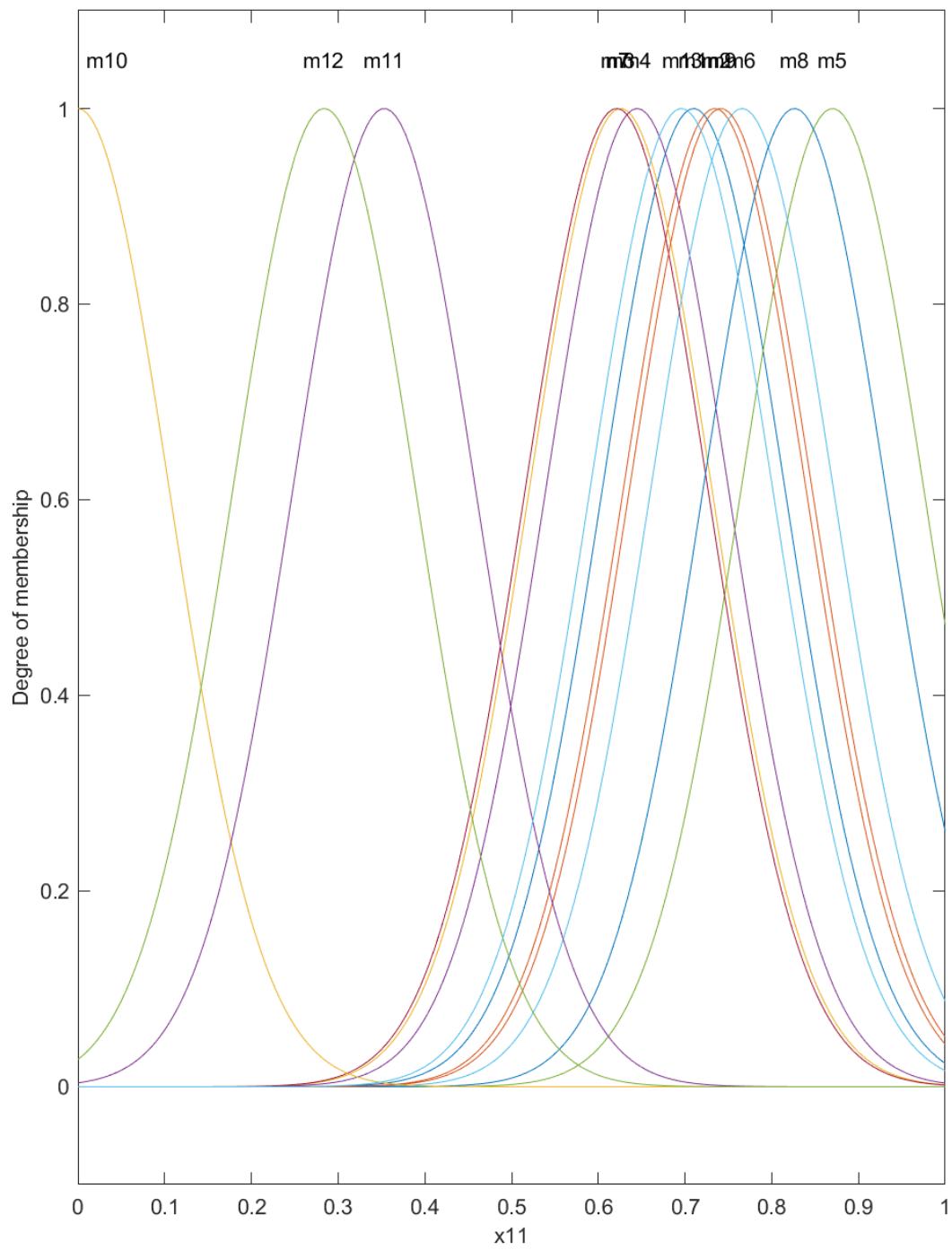
Διάγραμμα 20 - Ασαφή σύνολα σημαντικότερου χαρακτηριστικού μετά την εκπαίδευση



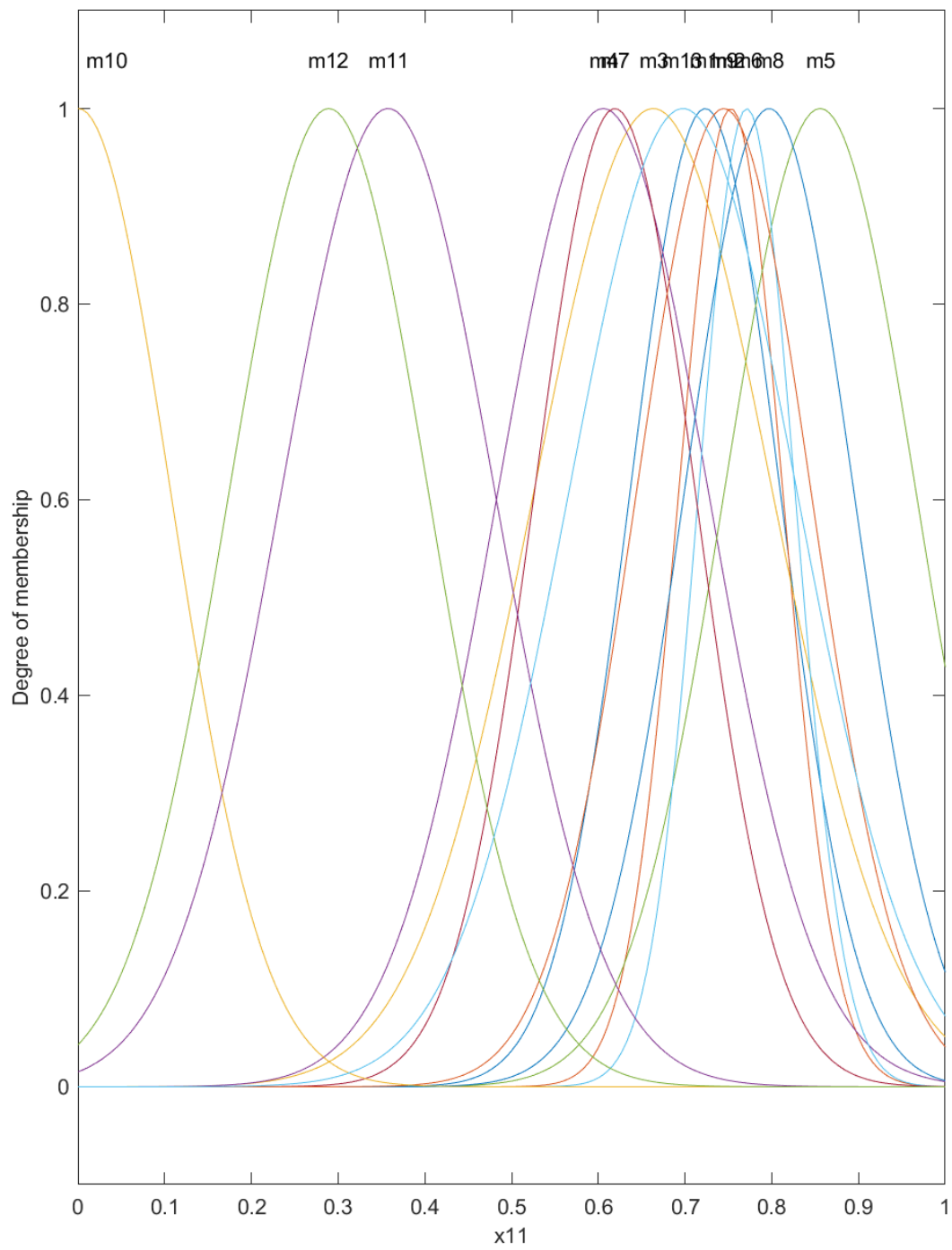
Διάγραμμα 21 - Ασαφή σύνολα δεύτερου σημαντικότερου χαρακτηριστικού πριν την εκπαίδευση



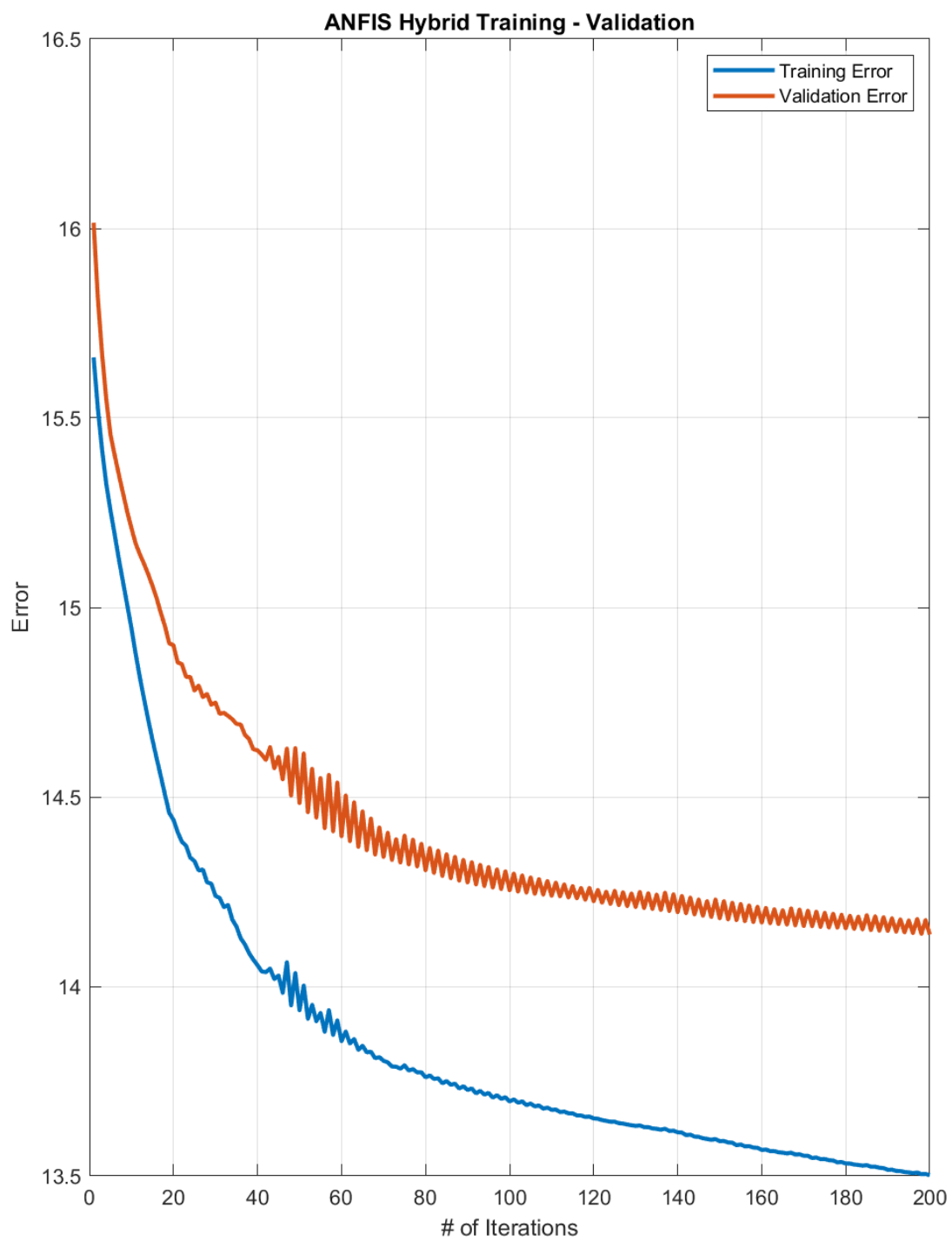
Διάγραμμα 22 - Ασαφή σύνολα δεύτερου σημαντικότερου χαρακτηριστικού μετά την εκπαίδευση



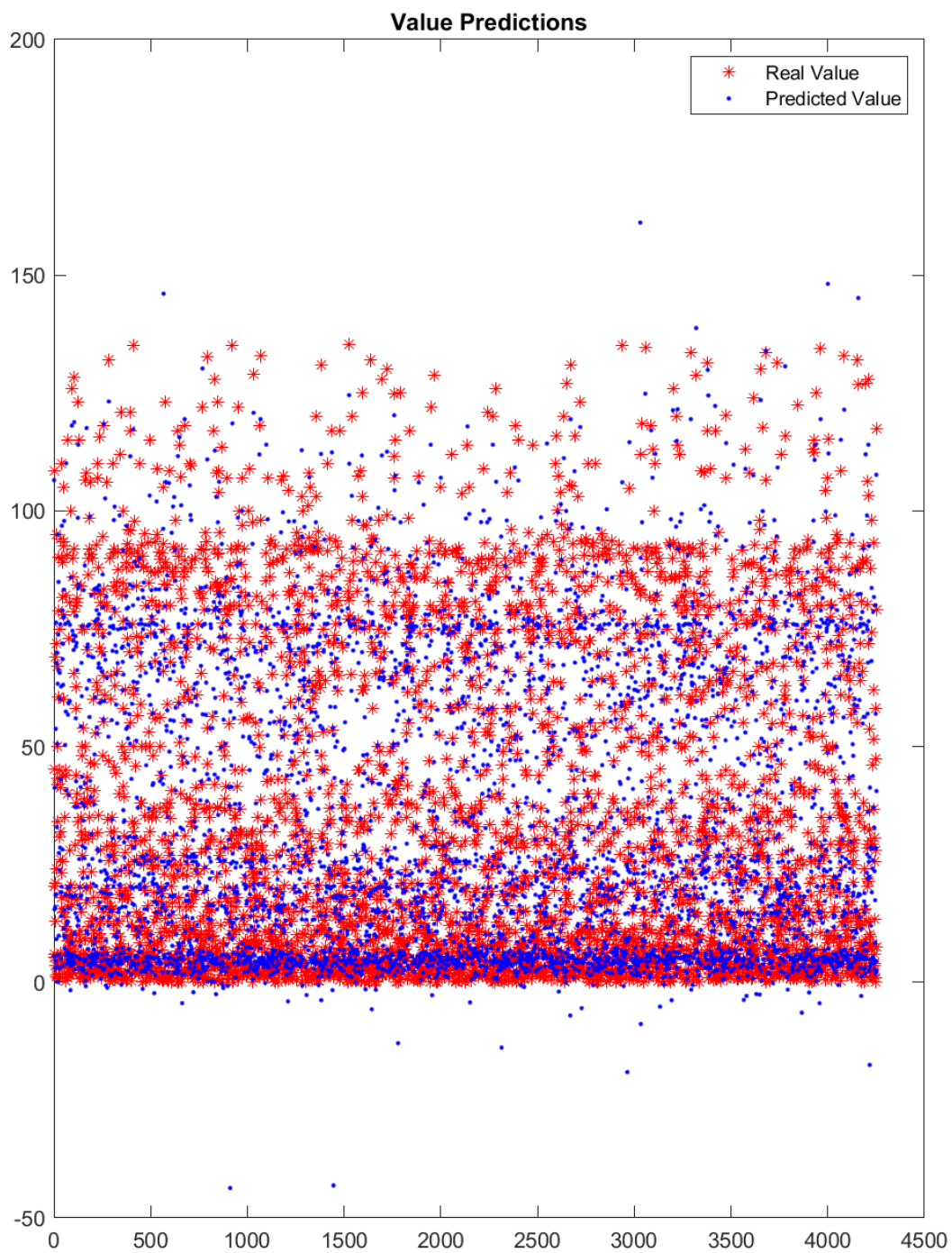
Διάγραμμα 23 - Ασαφή σύνολα λιγότερο σημαντικότερου χαρακτηριστικού πριν την εκπαίδευση



Διάγραμμα 24 - Ασαφή σύνολα λιγότερο σημαντικότερου χαρακτηριστικού μετά την εκπαίδευση



Διάγραμμα 25 - Διάγραμμα εκμάθησης μοντέλου μεγάλης διαστασιμότητας



Διάγραμμα 26 - Διάγραμμα πρόβλεψης μοντέλου μεγάλης διαστασιμότητας

Δείκτες Απόδοσης	
<i>RMSE</i>	14.5715
<i>R2</i>	0.8183
<i>NMSE</i>	0.1817
<i>NDEI</i>	0.4262

Παρατηρήσεις

Όπως αναφέρθηκε και προηγουμένως το αποτέλεσμα που προέκυψε από το grid search ως την περίπτωση που επιφέρει τα καλύτερα αποτελέσματα είναι ο συνδιασμός

- Αριθμός χαρακτηριστικών : 11
- Randii : 0.3

Αυτό έχει ως συνέπεια τη δημιουργία 16 cluster και κατά συνέπεια την ύπαρξη 16 συνολικών κανόνων. Εδώ να αναφερθεί πως εάν για τον ίδιο αριθμό χαρακτηριστικών είχαμε επιλέξει grid partitioning έναντι του subtractive clustering, τότε ακόμη και με την πιο απλή περίπτωση της διαμέρισης του χώρου εισόδου σε 2 μόνο ασαφή σύνολα για κάθε μεταβλητή, το ασαφές σύστημα συμπερασμού θα είχε σύνολο $2^{11} = 2048$ κανόνες. Δηλαδή 128 φορές περισσότερους από τώρα!

Τα σφάλματα που παρουσιάζονται στον παραπάνω πίνακα φαίνονται αρκετά μεγάλα αλλά είναι αποδεκτά δεδομένου την επιλογή μόνο των 11 από τα 81 συνολικά χαρακτηριστικά αλλά και τη δραστική μείωση των κανόνων που μας επέφερε η ομαδοποίηση των δεδομένων με τη χρήση του subtractive clustering.