

Εργασία
Στα Ασαφή Συστήματα
(Group3_Ser3)

ΟΝΟΜΑ: Αντώνης
ΕΠΩΝΥΜΟ: Μυρσινιάς
ΑΕΜ:8873
ΕΤΟΣ:2020

Περιεχόμενα

Πρόβλημα.....	3
Πρώτη εφαρμογή σε UCI dataset.....	3
Διαχωρισμός του dataset	4
Εκπαίδευση των TSK μοντέλων με διαφορετικές παραμέτρους	4
Σχήμα 1	4
TSK Μοντέλο 1.....	5
Σχήμα 2: Αρχικές Συναρτήσεις Συμμετοχής Μοντέλο 1	5
Σχήμα 3: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση Μοντέλο 1	6
Σχήμα 4: Καμπύλες εκμάθησης Μοντέλο 1.....	6
Σχήμα 5: Σφάλματα Πρόβλεψης Μοντέλο 1	7
TSK Μοντέλο 2.....	8
Σχήμα 6: Αρχικές Συναρτήσεις Συμμετοχής Μοντέλο 2	8
Σχήμα 7: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση Μοντέλο 2	9
Σχήμα 8: Καμπύλες εκμάθησης Μοντέλο 2.....	9
Σχήμα 9: Σφάλματα Πρόβλεψης Μοντέλο 2	10
TSK Μοντέλο 3.....	11
Σχήμα 10: Αρχικές Συναρτήσεις Συμμετοχής Μοντέλο 3	11
Σχήμα 11: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση Μοντέλο 3	12
Σχήμα 12: Καμπύλες εκμάθησης Μοντέλο 3.....	12
Σχήμα 13: Σφάλματα Πρόβλεψης Μοντέλο 3	13
TSK Μοντέλο 4.....	14
Σχήμα 14: Αρχικές Συναρτήσεις Συμμετοχής Μοντέλο 4	14
Σχήμα 15: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση Μοντέλο 4	15
Σχήμα 16: Καμπύλες εκμάθησης Μοντέλο 4.....	15
Σχήμα 17: Σφάλματα Πρόβλεψης Μοντέλο 4	16
Αξιολόγηση μοντέλων	16
Πίνακας 1: Μετρικές Αξιολόγησης Μοντέλων	16
Εφαρμογή σε dataset με υψηλή διαστασιμότητα.....	17
Διαχωρισμός του dataset και μείωση διαστάσεων και κανόνων.....	18
Πίνακας 2: Μέση τιμή MSE μοντέλων	19
Σχήμα 18: Ιστόγραμμα των MSE των μοντέλων	20
Βέλτιστο Μοντέλων (16 features, 15 rules)	20
Σχήμα 19: Ενδεικτικές συναρτήσεις συμμετοχής πριν την εκπαίδευση	21

<i>Σχήμα 20: Ενδεικτικές συναρτήσεις συμμετοχής μετά την εκπαίδευση</i>	21
<i>Σχήμα 21: Καμπύλες εκμάθησης.....</i>	22
<i>Σχήμα 22: Σφάλμα Πρόβλεψης</i>	22
<i>Αξιολόγηση μοντέλων</i>	23
<i>Πίνακας 3: Μετρικές Αξιολόγησης Μοντέλων</i>	23
<i>Αρχεία.....</i>	24

Πρόβλημα

Στόχος της εργασίας αυτής είναι να διευρυνθεί η ικανότητα των μοντέλων TSK στη μοντελοποίηση πολυμεταβλητών, μη γραμμικών συναρτήσεων. Συγκεκριμένα, επιλέγονται δύο σύνολα δεδομένων από το UCI repository με σκοπό την εκτίμηση της μεταβλητής στόχου από τα διαθέσιμα δεδομένα, με χρήση ασαφών μοντέλων. Το πρώτο σύνολο δεδομένων θα χρησιμοποιηθεί για μια απλή διεύρυνση της διαδικασίας εκπαίδευσης και αξιολόγησης μοντέλων αυτού του είδους, καθώς και για μια επίδειξη τρόπων ανάλυσης και ερμηνείας των αποτελεσμάτων. Το δεύτερο, πολυπλοκότερο σύνολο δεδομένων θα χρησιμοποιηθεί για μια πληρέστερη διαδικασία μοντελοποίησης, η οποία θα περιλαμβάνει μεταξύ άλλων προεπεξεργαστικά βήματα όπως επιλογή χαρακτηριστικών (feature selection), καθώς και μεθόδους βελτιστοποίησης των μοντέλων μέσω της διασταυρωμένης επικύρωσης (cross validation).

Πρώτη εφαρμογή σε UCI dataset

Στην πρώτη φάση της εργασίας, επιλέγεται από το UCI repository το Combined Power Plant (CCPP) dataset, το οποίο περιλαμβάνει 9568 δείγματα (instances) και 4 χαρακτηριστικά (features). Τα δεδομένα συλλέχθηκαν από ένα εργοστάσιο παραγωγή ηλεκτρικής ενέργειας κατά τη διάρκεια 6 ετών και χαρακτηρίζονται από τις εξής τέσσερις μεταβλητές:

- Μέση ωριαία θερμοκρασία (T)
- Μέση ωριαία πίεση (AP)
- Μέση ωριαία σχετική υγρασία (RH)
- Μέση ωριαία exhaust vacuum

Η μεταβλητή της οποίας την τιμή θέλουμε να προβλέψουμε χρησιμοποιώντας τις παραπάνω τέσσερις μεταβλητές χαρακτηριστικών, είναι η ενεργειακή απόδοση του σταθμού (energy output).

Διαχωρισμός του dataset

Αρχικά, πραγματοποιούμε τον διαχωρισμό του dataset σε τρία μη επικαλυπτόμενα υποσύνολα $\{D_{trn}, D_{val}, D_{chk}\}$ ως εξής:

- $D_{trn} = 60\%$ του αρχικού σετ (Σετ το οποίο θα χρησιμοποιηθεί κατά την εκπαίδευση του μοντέλου).
- $D_{val} = 20\%$ του αρχικού σετ (Σετ το οποίο θα χρησιμοποιηθεί για την επικύρωση και αποφυγή του φαινομένου της υπερεκπαίδευσης).
- $D_{chk} = 20\%$ του αρχικού σετ (Σετ που θα χρησιμοποιηθεί για τον έλεγχο της απόδοσης του τελικού μοντέλου).

Εκπαίδευση των TSK μοντέλων με διαφορετικές παραμέτρους

Σε αυτό το στάδιο θα γίνει η σύγκριση τεσσάρων διαφορετικών μοντέλων TSK. Τα χαρακτηριστικά των μοντέλων παρουσιάζονται στον παρακάτω σχήμα.

	Πλήθος συναρτήσεων συμμετοχής	Μορφή εξόδου
<i>TSK_model_1</i>	2	Singleton
<i>TSK_model_2</i>	3	Singleton
<i>TSK_model_3</i>	2	Polynomial
<i>TSK_model_4</i>	3	Polynomial

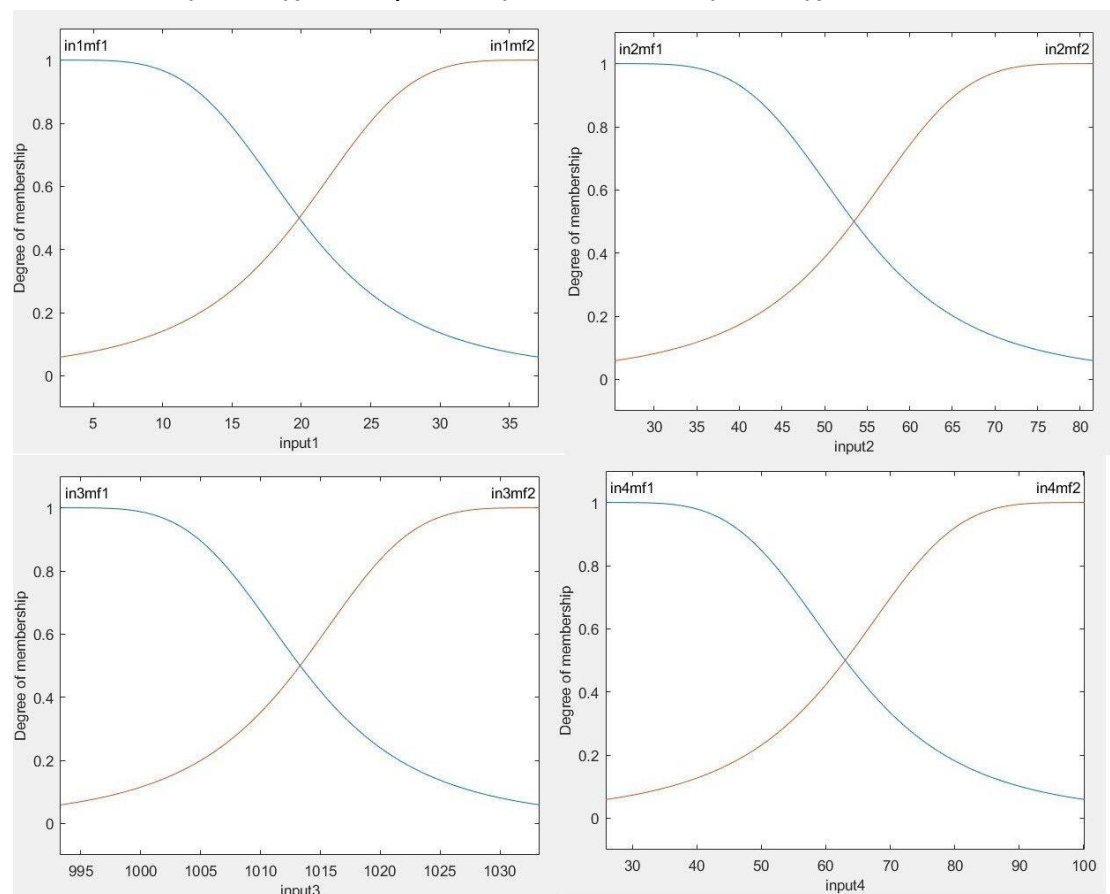
Σχήμα 1

Τα παραπάνω 4 μοντέλα θα εκπαιδευτούν με την υβριδική μέθοδο, σύμφωνα με την οποία οι παράμετροι των συναρτήσεων συμμετοχής βελτιστοποιούνται μέσω της μεθόδου της οπισθοδιάδοσης (backpropagation), ενώ οι παράμετροι της πολυωνυμικής συνάρτησης εξόδου βελτιστοποιούνται μέσω της μεθόδου ελαχίστων τετραγώνων (Least Squares). Οι συναρτήσεις συμμετοχής είναι bell-shaped και η αρχικοποίηση τους έχει γίνει με τέτοιο τρόπο, ώστε τα διαδοχικά ασαφή σύνολα να παρουσιάζουν σε κάθε είσοδο βαθμό επικάλυψης περίπου 0.5.

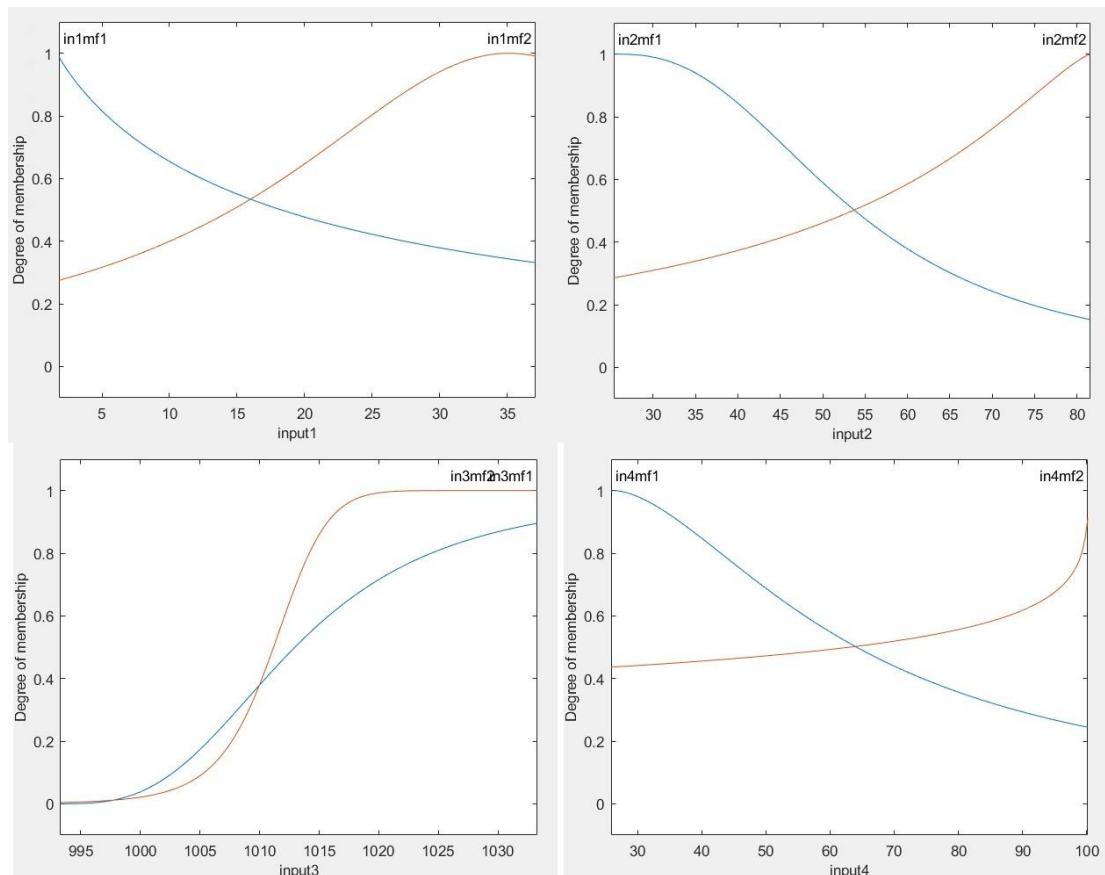
Δημιουργούμε τα με τη συνάρτηση `genfis1()`, του Matlab, τα παραπάνω μοντέλα με την μέθοδο Grid Partition, δίνοντας ως είσοδο τα δεδομένα εκπαίδευσης. Η εκπαίδευση τους γίνεται με την χρήση της εντολής `anfis()`. Για την εκπαίδευσή τους έχουν χρησιμοποιηθεί 350 εποχές. Οι μετρικές που θα χρησιμοποιηθούν για την αξιολόγηση των μοντέλων είναι: MSE, RMSE, R^2 , NMSE, NDEI.

TSK Μοντέλο 1

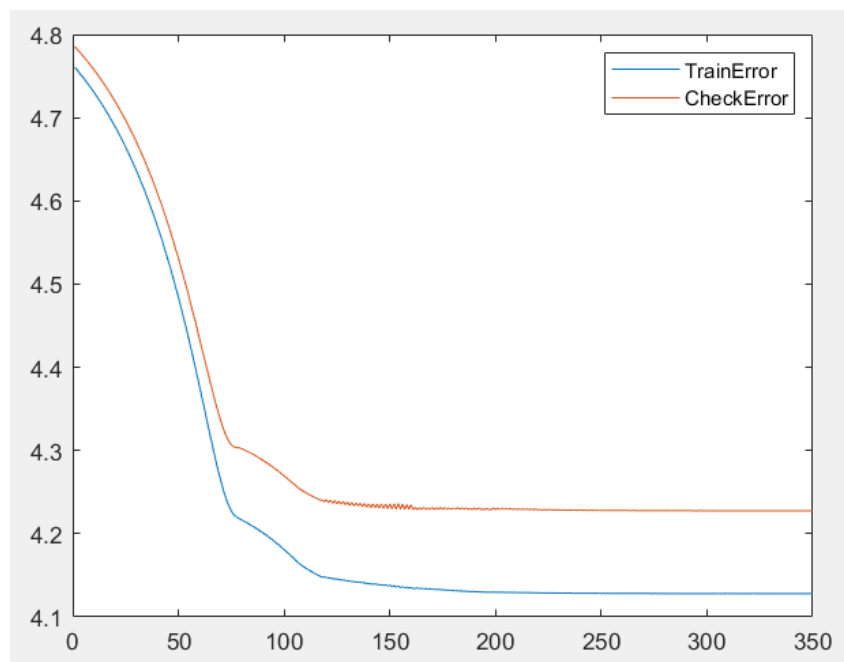
Το μοντέλο 1 έχει δύο συναρτήσεις συμμετοχής τύπου Bell-Shaped με επικάλυψη 0.5 για κάθε μεταβλητή εισόδου ενώ η έξοδός του είναι μορφής Singleton. Οι συναρτήσεις συμμετοχής φαίνονται πριν την εκπαίδευση στο Σχ.2 και μετά την εκπαίδευση στο Σχ.3



Σχήμα 2: Αρχικές Συναρτήσεις Συμμετοχής Μοντέλο 1

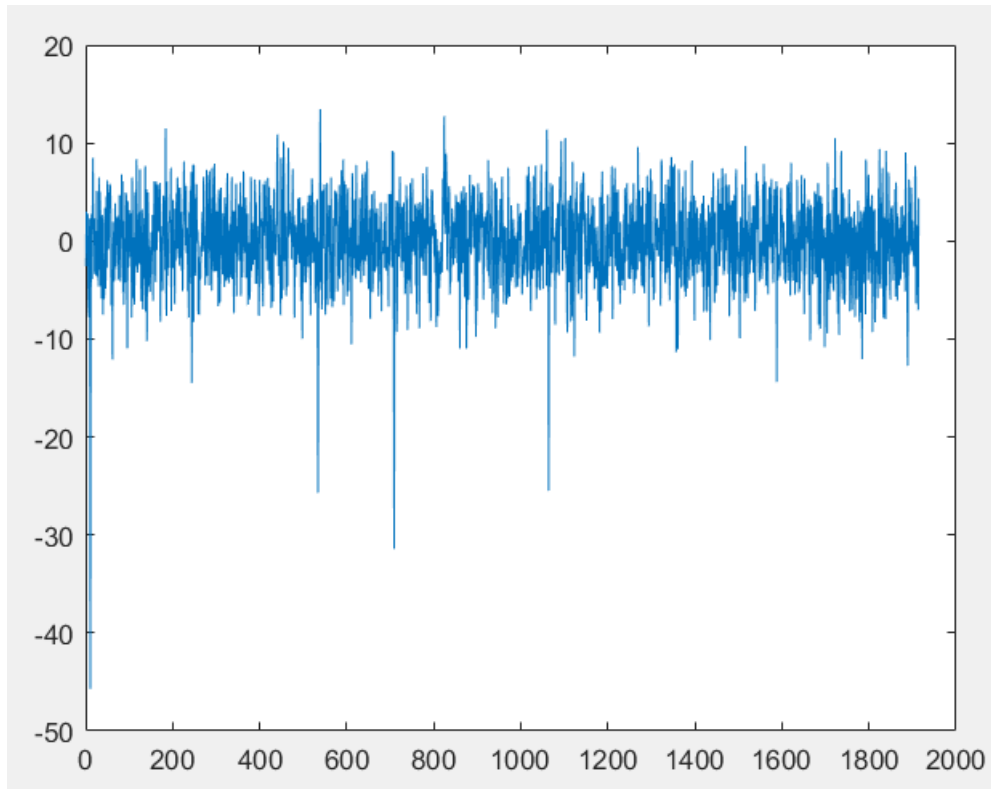


Σχήμα 3: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση Μοντέλο 1
 Παρακάτω παρουσιάζονται οι καμπύλες εκμάθησης (TrainError, CheckError).



Σχήμα 4: Καμπύλες εκμάθησης Μοντέλο 1

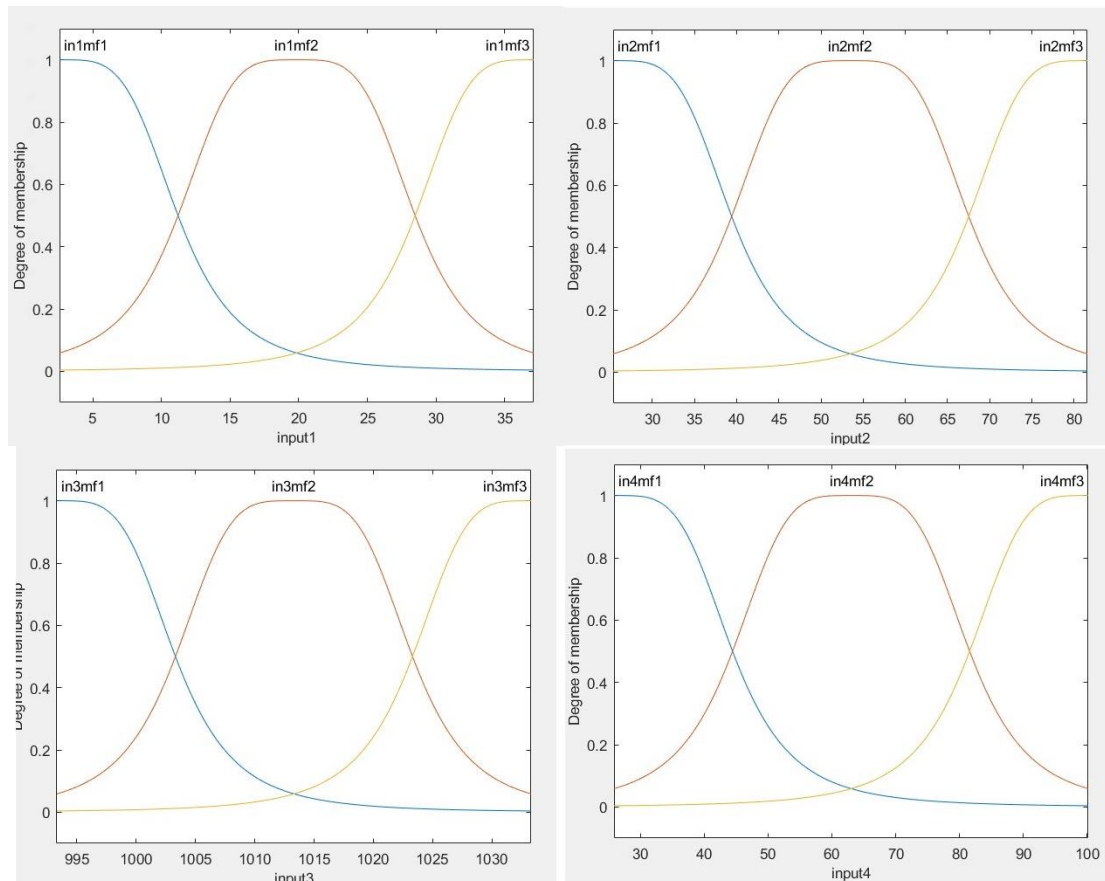
Στη συνέχεια, παρουσιάζεται το διάγραμμα, στο οποίο αποτυπώνεται το σφάλμα πρόβλεψης.



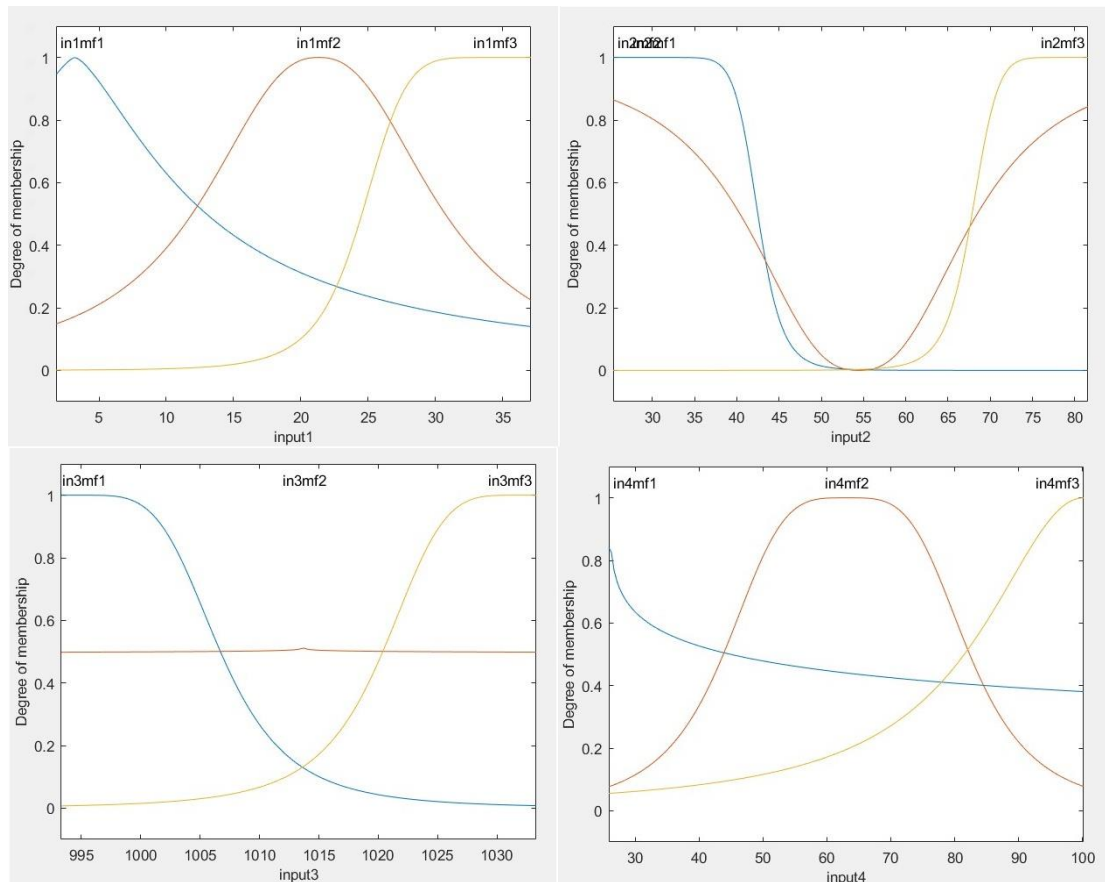
Σχήμα 5: Σφάλματα Πρόβλεψης Μοντέλο 1

TSK Μοντέλο 2

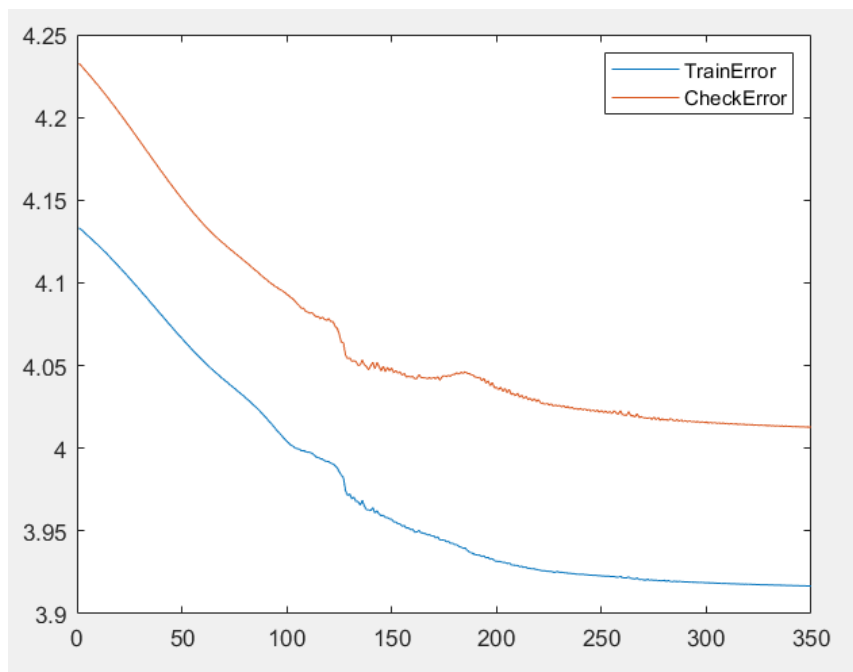
Το μοντέλο 2 έχει τρεις συναρτήσεις συμμετοχής τύπου Bell-Shaped με επικάλυψη 0.5 για κάθε μεταβλητή εισόδου ενώ η έξοδός του είναι μορφής Singleton. Οι συναρτήσεις συμμετοχής φαίνονται πριν την εκπαίδευση στο Σχ.6 και μετά την εκπαίδευση στο Σχ.7.



Σχήμα 6: Αρχικές Συναρτήσεις Συμμετοχής Μοντέλο 2

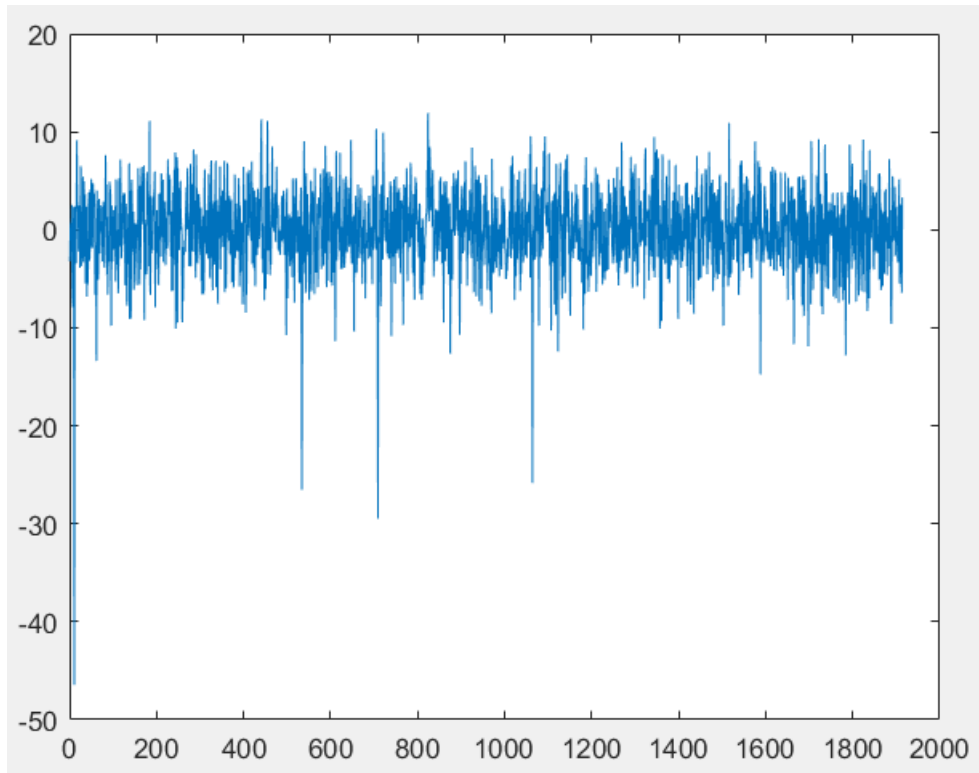


Σχήμα 7: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση Μοντέλο 2
 Παρακάτω παρουσιάζονται οι καμπύλες εκμάθησης (TrainError, CheckError).



Σχήμα 8: Καμπύλες εκμάθησης Μοντέλο 2

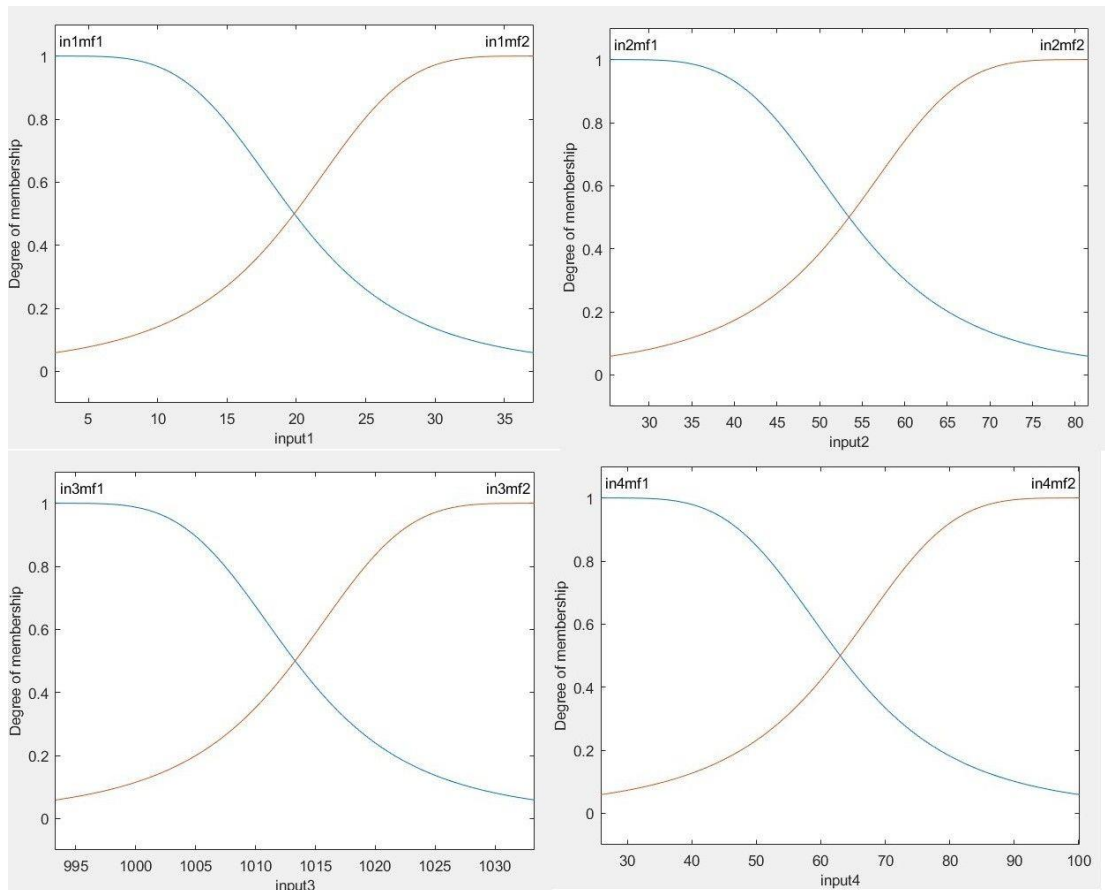
Στη συνέχεια, παρουσιάζεται το διάγραμμα, στο οποίο αποτυπώνεται το σφάλμα πρόβλεψης.



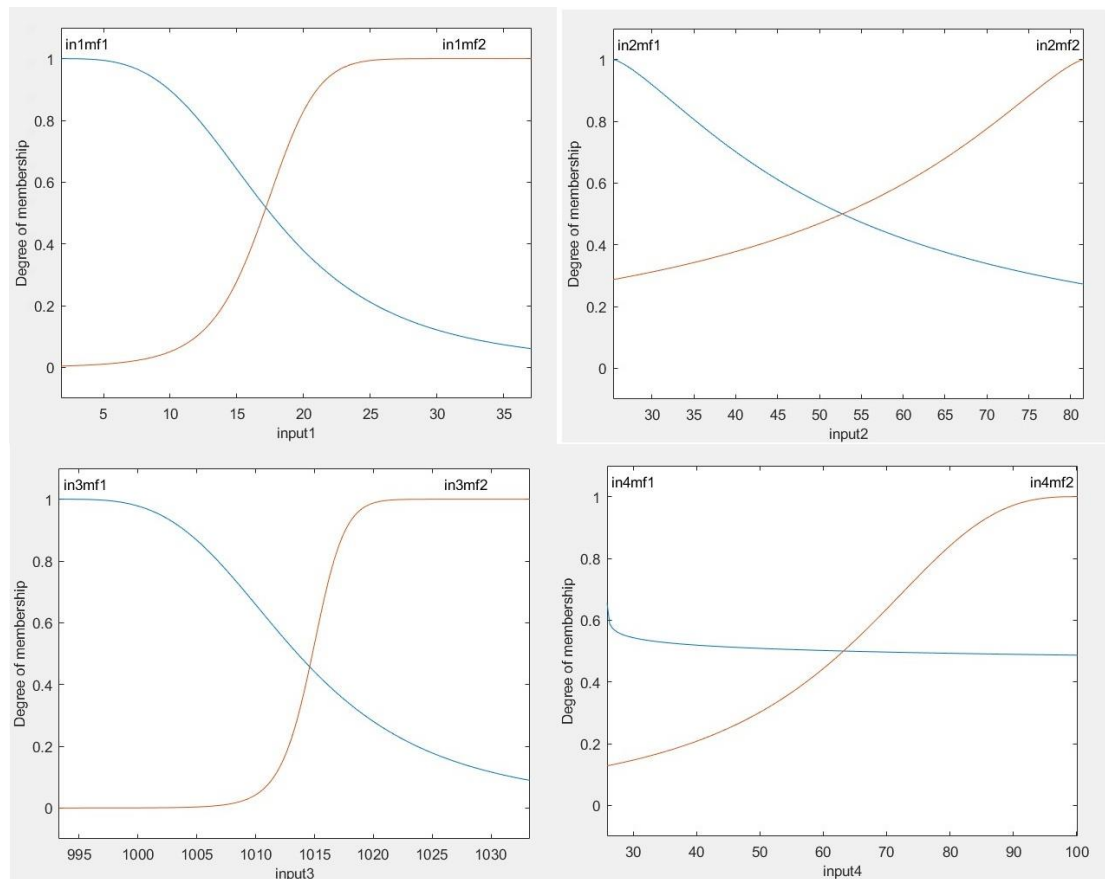
Σχήμα 9: Σφάλματα Πρόβλεψης Μοντέλο 2

TSK Μοντέλο 3

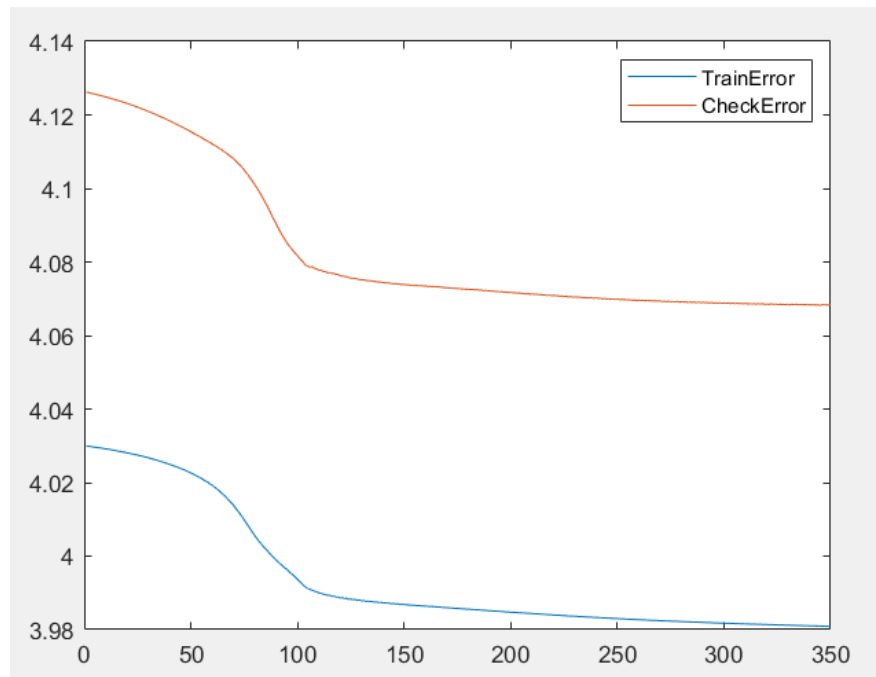
Το μοντέλο 3 έχει δύο συναρτήσεις συμμετοχής τύπου Bell-Shaped με επικάλυψη 0.5 για κάθε μεταβλητή εισόδου, ενώ η έξοδός του είναι μορφής Polynomial. Οι συναρτήσεις συμμετοχής φαίνονται πριν την εκπαίδευση στο Σχ.10 και μετά την εκπαίδευση στο Σχ.11.



Σχήμα 10: Αρχικές Συναρτήσεις Συμμετοχής Μοντέλο 3

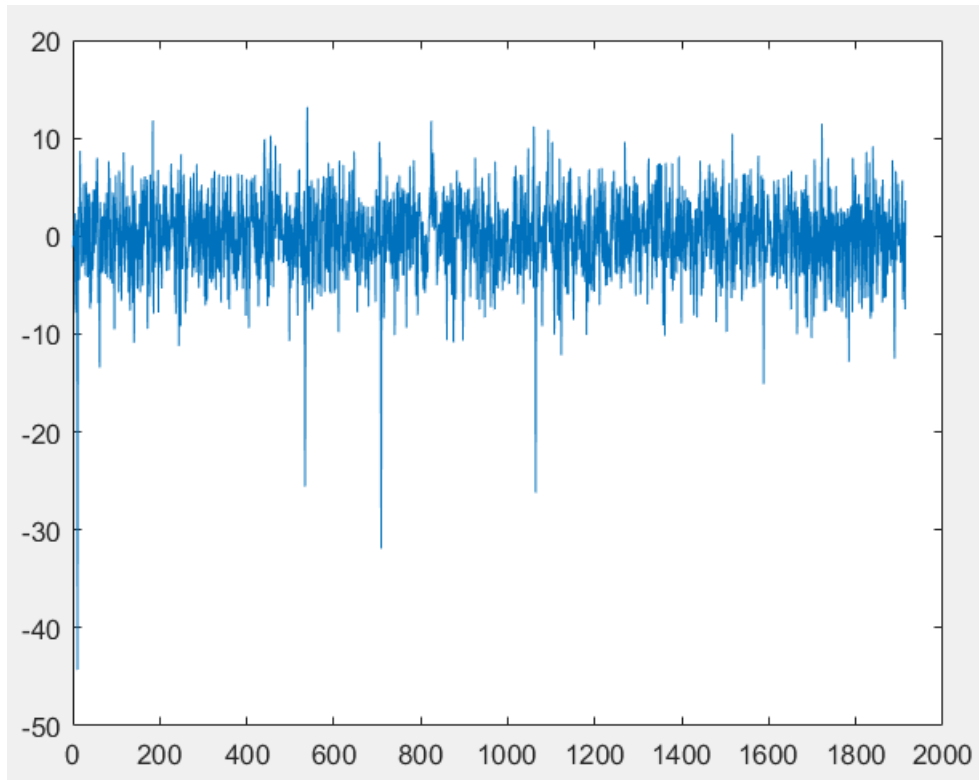


Σχήμα 11: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση Μοντέλο 3
Παρακάτω παρουσιάζονται οι καμπύλες εκμάθησης (TrainError, CheckError).



Σχήμα 12: Καμπύλες εκμάθησης Μοντέλο 3

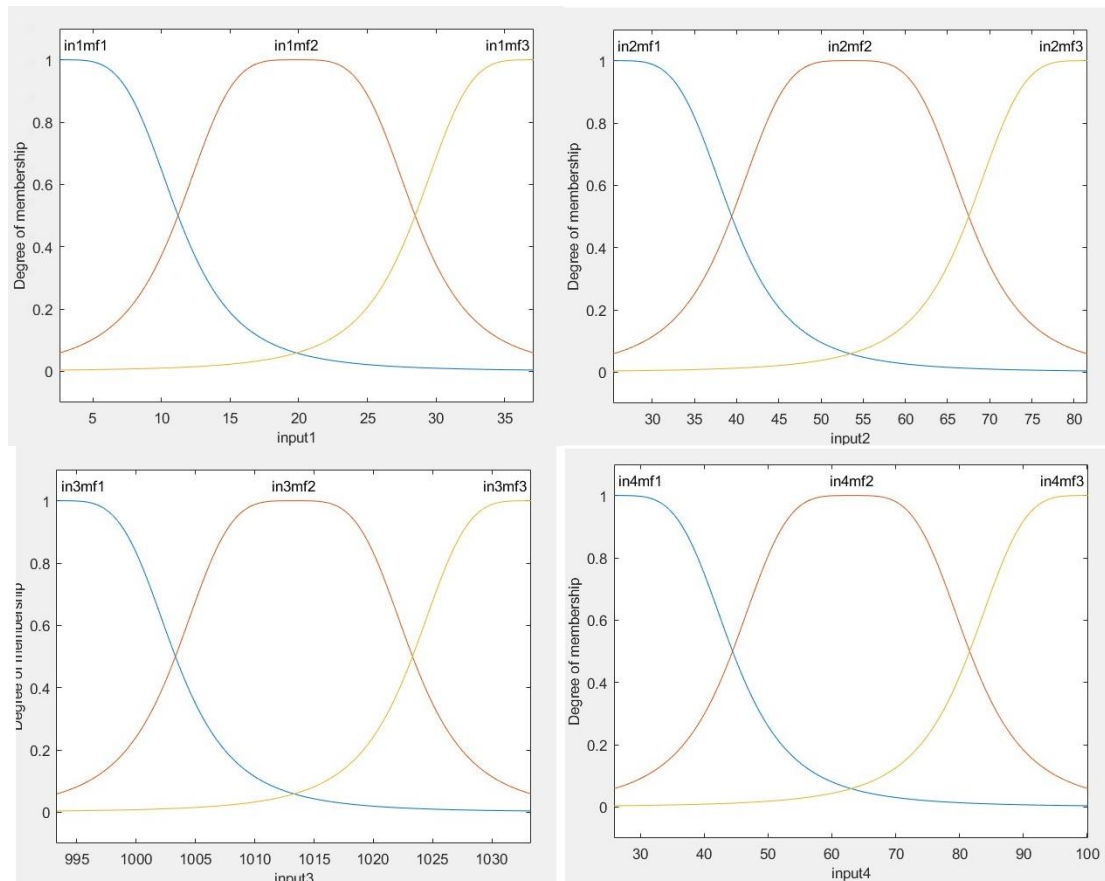
Στη συνέχεια, παρουσιάζεται το διάγραμμα, στο οποίο αποτυπώνεται το σφάλμα πρόβλεψης.



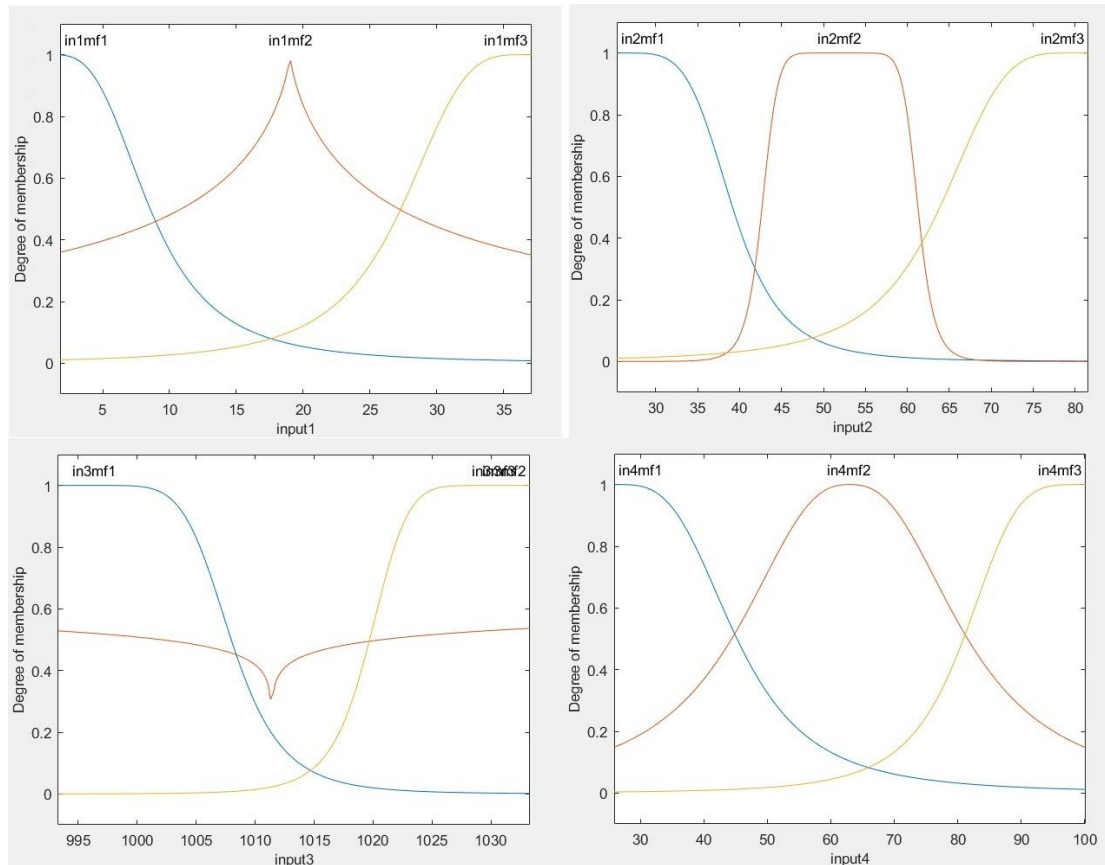
Σχήμα 13: Σφάλματα Πρόβλεψης Μοντέλο 3

TSK Μοντέλο 4

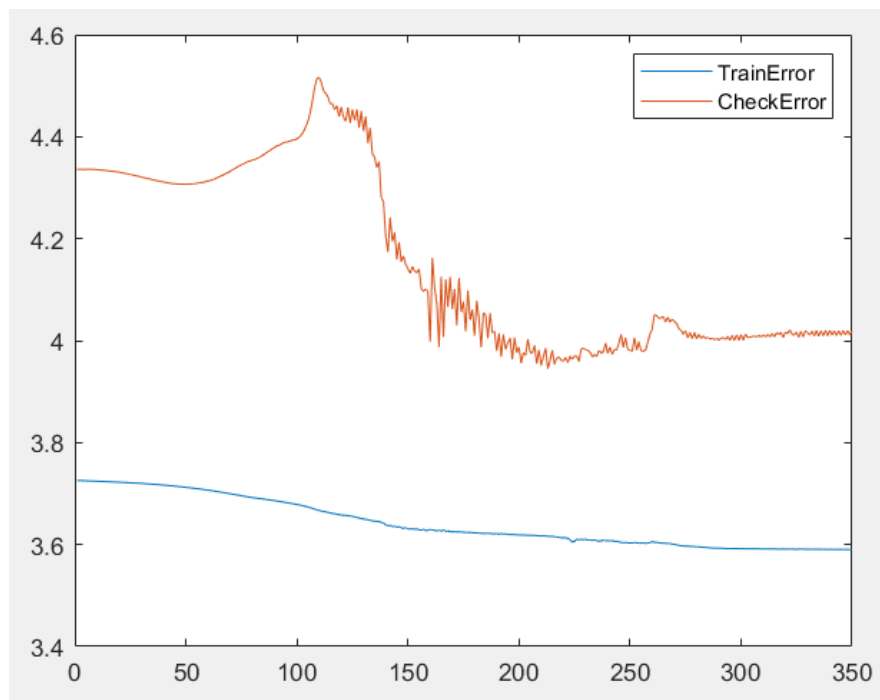
Το μοντέλο 4 έχει τρεις συναρτήσεις συμμετοχής τύπου Bell-Shaped με επικάλυψη 0.5 για κάθε μεταβλητή εισόδου, ενώ η έξοδός του είναι μορφής Polynomial. Οι συναρτήσεις συμμετοχής φαίνονται πριν την εκπαίδευση στο Σχ.14 και μετά την εκπαίδευση στο Σχ.15.



Σχήμα 14: Αρχικές Συναρτήσεις Συμμετοχής Μοντέλο 4

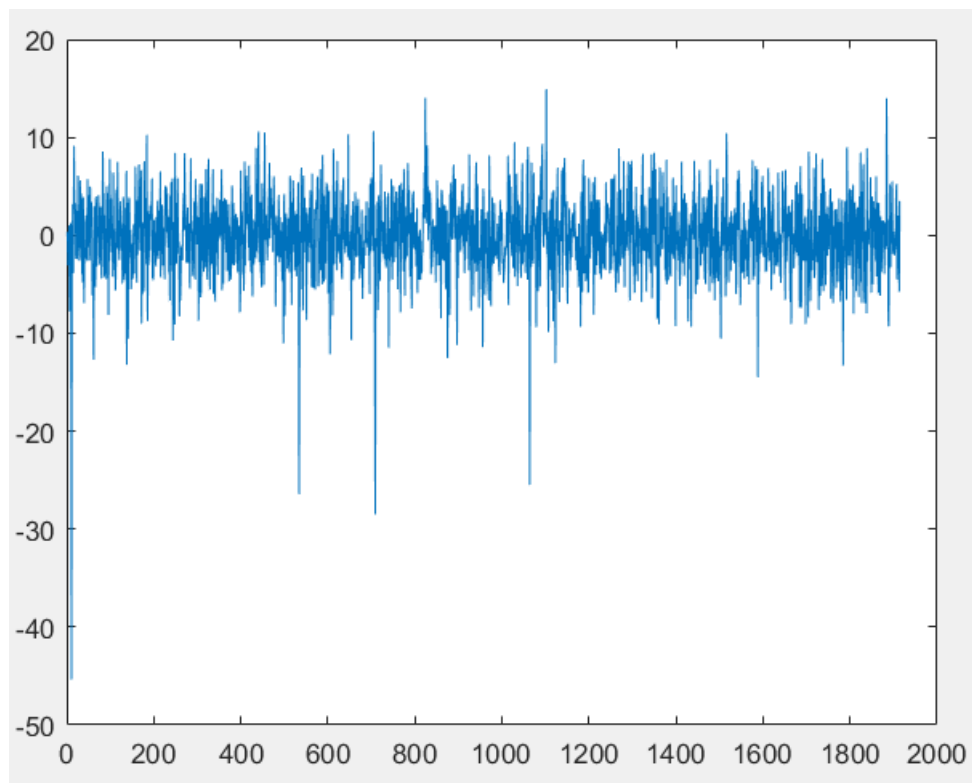


Σχήμα 15: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση Μοντέλο 4
 Παρακάτω παρουσιάζονται οι καμπύλες εκμάθησης (TrainError, CheckError).



Σχήμα 16: Καμπύλες εκμάθησης Μοντέλο 4

Στη συνέχεια, παρουσιάζεται το διάγραμμα, στο οποίο αποτυπώνεται το σφάλμα πρόβλεψης.



Σχήμα 17: Σφάλματα Πρόβλεψης Μοντέλο 4

Αξιολόγηση μοντέλων

Στον παρακάτω πίνακα παρουσιάζονται συγκεντρωμένες όλες οι μετρικές αξιολόγησης για το κάθε μοντέλο.

Evaluation	Model 1	Model 2	Model 3	Model 4
MSE	18.6574	17.0319	17.6430	16.5601
RMSE	4.3194	4.1270	4.2004	4.0694
R^2	0.9354	0.9410	0.9389	0.9427
NMSE	0.0646	0.0590	0.0611	0.0573
NDEI	0.2542	0.2429	0.2472	0.2395

Πίνακας 1: Μετρικές Αξιολόγησης Μοντέλων

Με βάση τον παραπάνω πίνακα βλέπουμε ότι και τα τέσσερα μοντέλα έχουν παρόμοιες επιδόσεις. Πιο συγκεκριμένα, το μοντέλο 4 (τρεις συναρτήσεις συμμετοχής και πολυωνυμική μορφή εξόδου) είναι το βέλτιστο μεταξύ των τεσσάρων. Βέβαια, για την επιλογή του καλύτερου μοντέλου πρέπει να λάβουμε υπόψη και άλλους παράγοντες όπως το χρόνο εκπαίδευσης και την προδιάθεση για υπερεκπαίδευση. Το μοντέλο 4 έχει πολύ μεγαλύτερο χρόνο εκπαίδευσης σε σχέση με τα άλλα μοντέλα. Επίσης, παρατηρώντας πιο προσεκτικά την καμπύλη εκμάθησης του μοντέλου 4, βλέπουμε ότι η καμπύλη του σφάλματος επικύρωσης έχει ανοδική πορεία που μας προϊδεάζει ότι ίσως οδηγηθούμε σε υπερεκπαίδευση.

Συμπερασματικά, από τον παραπάνω πίνακα μπορούμε να εξάγουμε το πόρισμα ότι η χρήση μεγαλύτερου αριθμού συναρτήσεων συμμετοχής βελτιώνει τα αποτελέσματα. Επίσης, η χρήση της εξόδου Singleton μειώνει σημαντικά το χρόνο εκπαίδευσης, αλλά έχει ως αποτέλεσμα μικρότερη ακρίβεια σε σχέση με τη χρήση της εξόδου Polynomial.

Εφαρμογή σε dataset με υψηλή διαστασιμότητα

Στη δεύτερη φάση της εργασίας θα ακολουθηθεί μια πιο συστηματική προσέγγιση στο πρόβλημα μοντελοποίησης. Για το σκοπό αυτό θα επιλεγεί ένα dataset με υψηλότερο βαθμό διαδραστικότητας. Ένα προφανές πρόβλημα που ανακύπτει από την επιλογή αυτή, είναι η λεγόμενη "έκρηξη" του πλήθους IF-THEN κανόνων (rule explosion). Όπως είναι γνωστό από τη θεωρία, για την κλασική περίπτωση του grid partitioning του χώρου εισόδου, ο αριθμός των κανόνων αυξάνεται εκθετικά σε σχέση με το πλήθος εισόδων, γεγονός που καθιστά πολύ δύσκολη την μοντελοποίηση μέσω ενός TSK μοντέλου ακόμα και για datasets μεσαίας κλίμακας.

Σύμφωνα με τα παραπάνω, μια αρκετά προφανής προσέγγιση θα ήταν να προσπαθήσουμε να μειώσουμε ταυτόχρονα και το πλήθος των εισόδων και το πλήθος των κανόνων. Οι δύο τεχνικές που παρουσιάζονται παρακάτω στοχεύουν σε αυτόν ακριβώς το διττό σκοπό.

Διαχωρισμός του dataset και μείωση διαστάσεων και κανόνων

Αρχικά, πραγματοποιούμε τον διαχωρισμό του dataset σε τρία μη επικαλυπτόμενα υποσύνολα $\{D_{trn}, D_{val}, D_{chk}\}$ ως εξής:

- $D_{trn} = 60\%$ του αρχικού σετ (Σετ το οποίο θα χρησιμοποιηθεί κατά την εκπαίδευση του μοντέλου).
- $D_{val} = 20\%$ του αρχικού σετ (Σετ το οποίο θα χρησιμοποιηθεί για την επικύρωση και αποφυγή του φαινομένου της υπερεκπαίδευσης).
- $D_{chk} = 20\%$ του αρχικού σετ (Σετ που θα χρησιμοποιηθεί για τον έλεγχο της απόδοσης του τελικού μοντέλου).

Στη συνέχεια, προκειμένου να μειώσουμε τις διαστάσεις της κάθε εγγραφής του σετ δεδομένων, εφαρμόζουμε τον αλγόριθμο Relief. Ο παραπάνω αλγόριθμος επιστρέφει τα πιο σημαντικά χαρακτηριστικά (από το πιο σημαντικό στο λιγότερο).

Έπειτα, προκειμένου να βρούμε το χαρακτηριστικά του μοντέλου που ανταποκρίνεται καλύτερα στο πρόβλημα εφαρμόζουμε την μέθοδο Grid Search σε συνδυασμό με τον K-fold Cross Validation (όπου $k = 5$).

Κατά την διαδικασία του Grid Search και του K-fold Cross Validation, αρχικά διαχωρίζουμε το σετ εκπαίδευσης σε 5 διαφορετικά σετ. Σε κάθε μία από τις 5 επαναλήψεις (λόγω του $k = 5$), τα 4 από τα 5 μέρη του σετ χρησιμοποιούνται για την εκπαίδευση του μοντέλου και το 5^ο για την επικύρωση του. Στη συνέχεια, υπολογίζεται η μέση τιμή του μέσου τετραγωνικού σφάλματος (MSE), ώστε αργότερα να γίνει η σύγκριση των διαφορετικών μοντέλων (διαφορετικά μοντέλα θεωρούνται εκείνα που έχουν διαφορετικό αριθμό κανόνων και χρησιμοποιούν διαφορετικό αριθμό χαρακτηριστικών του dataset) και να επιλεγεί το καλύτερο.

Για την ομαδοποίηση και την δημιουργία των κανόνων, αρχικά χρησιμοποιήθηκε η μέθοδος FCM (Fuzzy C-Means). Όμως, επειδή πέρα από τον μεγάλο χρόνο εκτέλεσης του Grid Search (λόγω του αριθμού των features και του αριθμού των rules $NF \times NR = \{3,9,15,21\} \times \{4,8,12,16,20\}$) υπήρχαν και πολλές NaN τιμές κατά τον υπολογισμό του CheckError, αρχικά πραγματοποιήθηκε αλλαγή των τιμών μήπως λυθεί το πρόβλημα. Οι νέες τιμές που χρησιμοποιήθηκαν ήταν: $NF \times NR = \{3,9,10,16\} \times \{4,8,9,11,15\}$. Παρόλο που αυτή η αλλαγή μείωσε σε μεγάλο βαθμό το πρόβλημα, δεν έγινε απαλοιφή όλων των NaN τιμών.

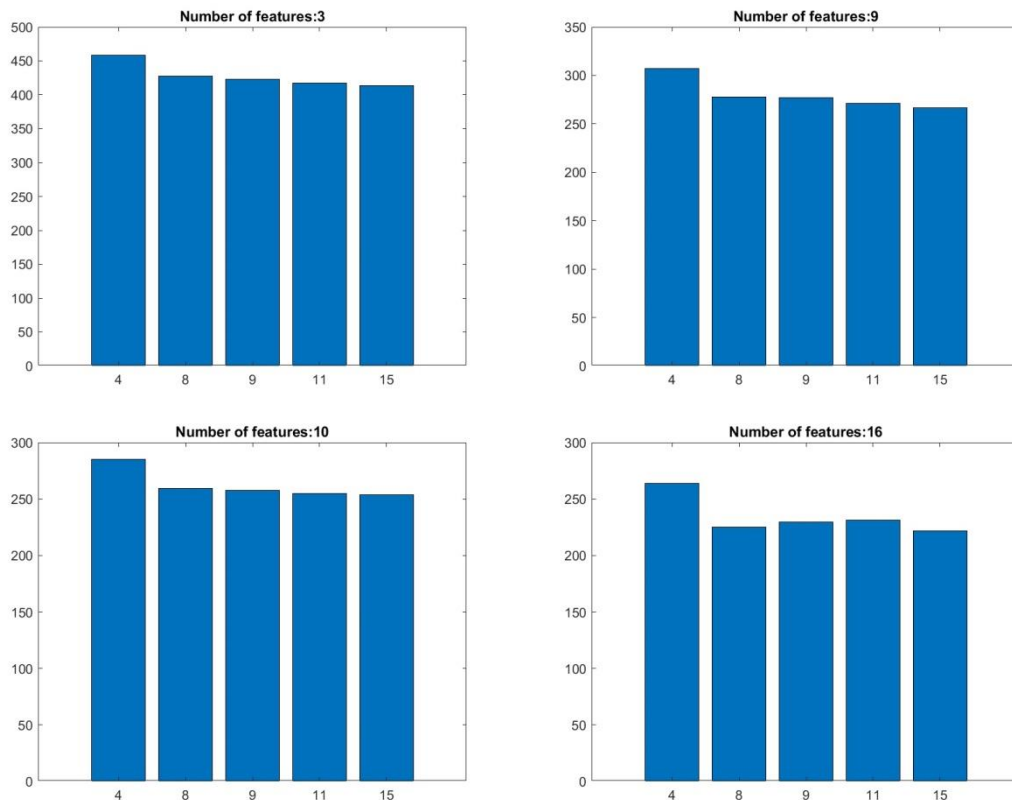
Λόγω του προβλήματος με τις NaN τιμές ακολουθήθηκε μία διαφορετική προσέγγιση για την δημιουργία των κανόνων. Συνεπώς, αντί για τη μέθοδο FCM έγινε χρήση της μεθόδου Subtractive Clustering. Κατά τη χρήση αυτής της μεθόδου ένα πρόβλημα που εμφανίστηκε ήταν ότι ο αριθμός των clusters δεν ήταν σταθερός λόγω της διαρκούς αλλαγής του σετ εκπαίδευσης κατά της διαδικασίας του Grid Search. Για την αντιμετώπιση του παραπάνω προβλήματος έγινε χρήση μίας συνάρτησης κατά την οποία μέσω της μεθόδου της διχοτόμησης γίνεται επιλογή του κατάλληλου Squash και του κατάλληλου InfluenceRange σε κάθε επανάληψη.

Στον παρακάτω πίνακα παρουσιάζεται η μέση τιμή του MSE των μοντέλων.

Rules Feature	4	8	9	11	15
3	458.0199	427.9584	423.3448	417.3948	413.6602
9	307.0308	277.6732	276.7962	271.1595	266.8616
10	285.0050	259.1341	257.8666	254.7029	253.6791
16	263.7303	225.1300	229.5658	231.4244	221.8779

Πίνακας 2: Μέση τιμή MSE μοντέλων

Στο παρακάτω σχήμα παρουσιάζονται η μέση τιμή των MSE σε γραφική μορφή.

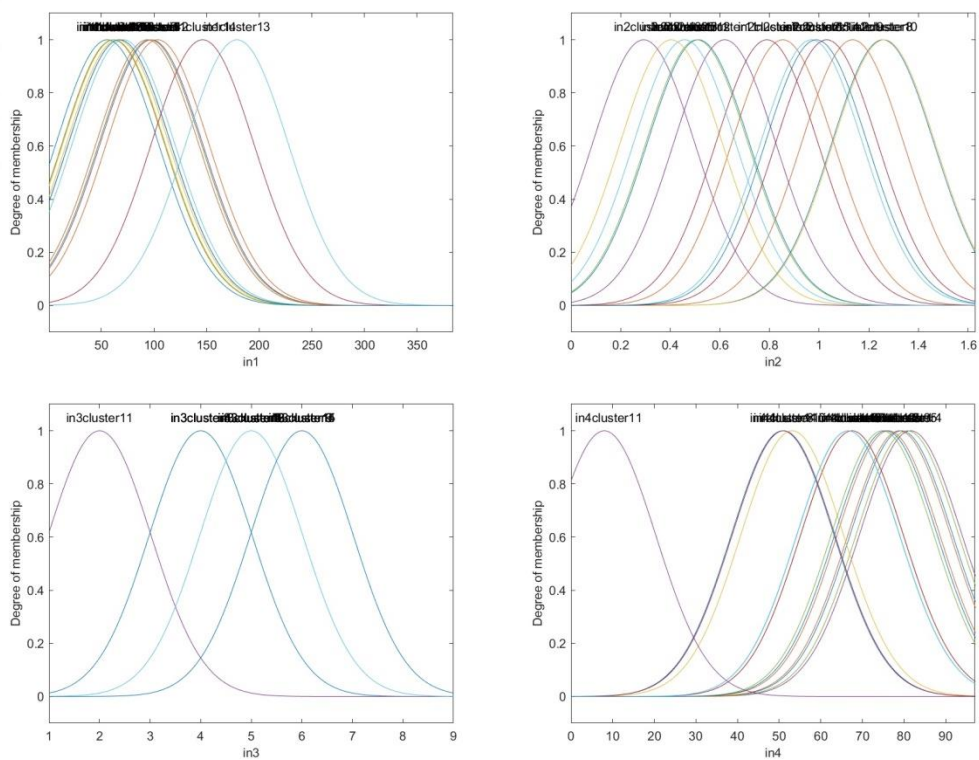


Σχήμα 18: Ιστόγραμμα των MSE των μοντέλων

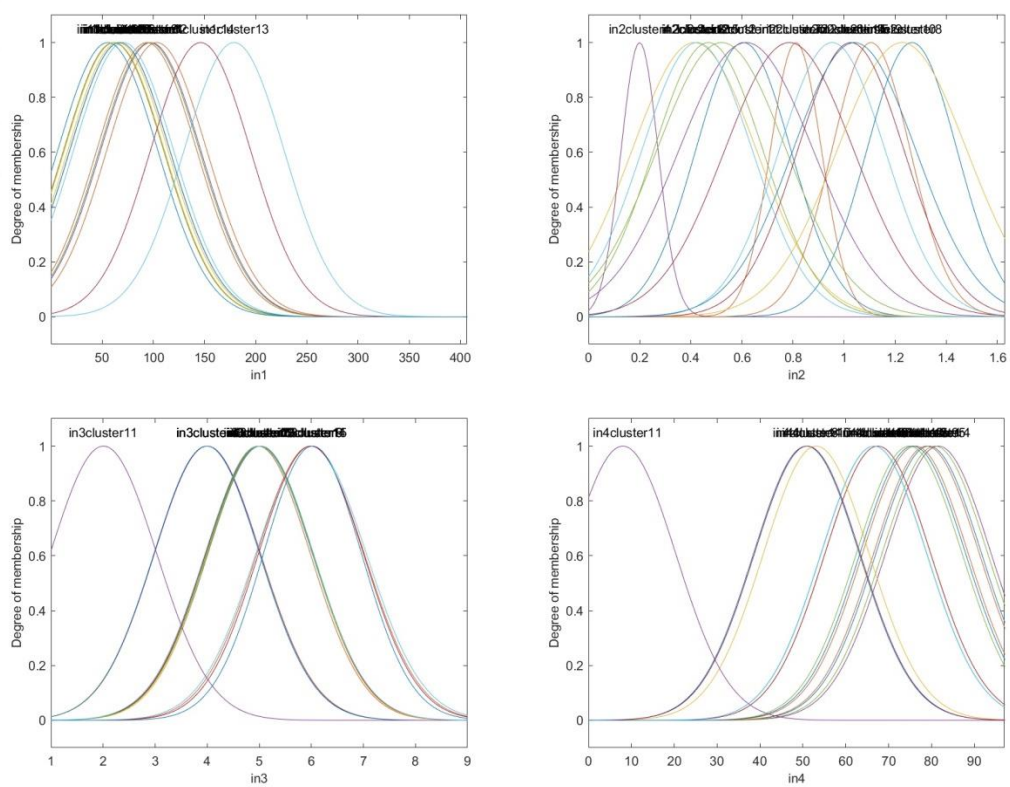
Βέλτιστο Μοντέλων (16 features, 15 rules)

Από τον πίνακα 2, βλέπουμε ότι το βέλτιστο μοντέλο είναι εκείνο που εκπαιδεύεται με 16 χαρακτηριστικά, από τα 81, και έχει 15 IF THEN κανόνες.

Στα παρακάτω διαγράμματα παρουσιάζονται ενδεικτικά μερικά ασαφή σύνολα στην αρχική και τελική τους μορφή.

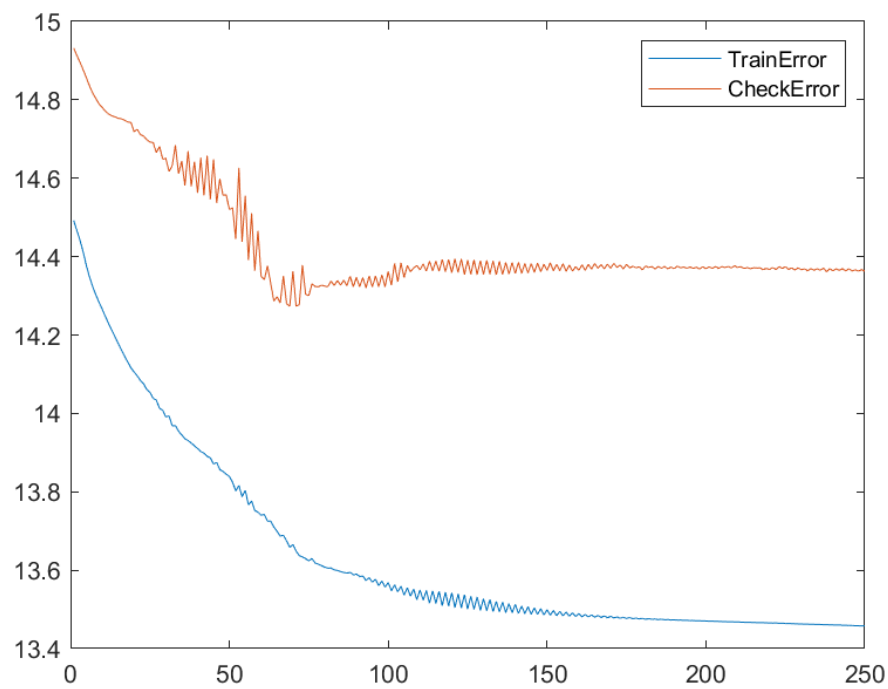


Σχήμα 19: Ενδεικτικές συναρτήσεις συμμετοχής πριν την εκπαίδευση

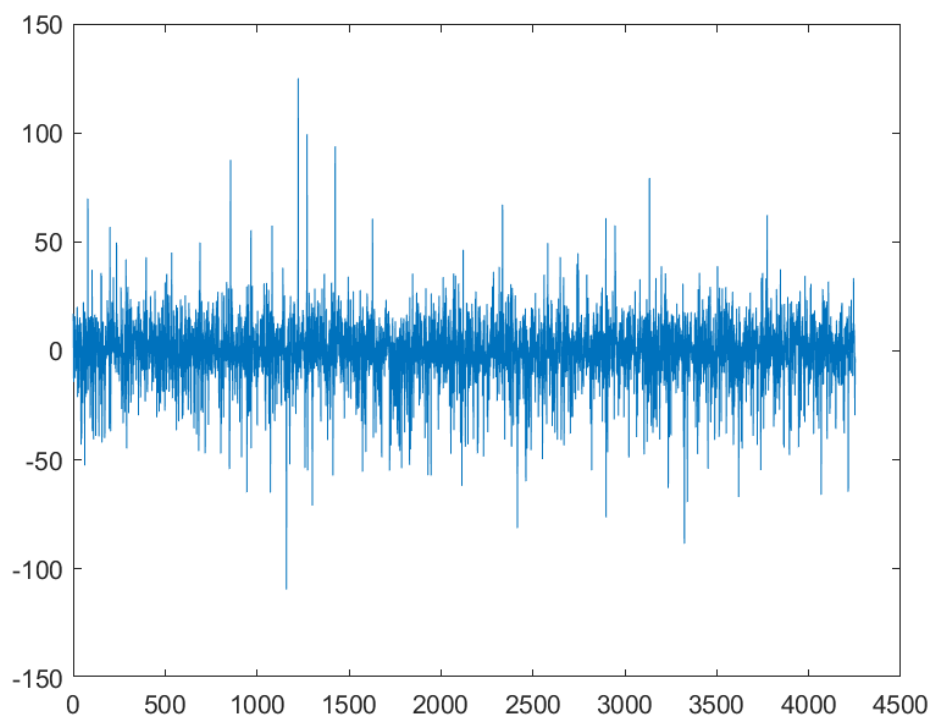


Σχήμα 20: Ενδεικτικές συναρτήσεις συμμετοχής μετά την εκπαίδευση

Στη συνέχεια, παρουσιάζονται οι καμπύλες εκμάθησης (learning curves) και το διάγραμμα που αποτυπώνεται το σφάλμα πρόβλεψης.



Σχήμα 21: Καμπύλες εκμάθησης



Σχήμα 22: Σφάλμα Πρόβλεψης

Αξιολόγηση μοντέλων

Στον παρακάτω πίνακα παρουσιάζονται συγκεντρωμένες όλες οι μετρικές αξιολόγησης για το βέλτιστο μοντέλο.

Evaluation	MSE	RMSE	R^2	NMSE	NDEI
	224.5196	14.9840	0.8038	0.1962	0.4430

Πίνακας 3: Μετρικές Αξιολόγησης Μοντέλων

Μέσω των μετρικών αξιολόγησης βλέπουμε ότι το μοντέλο είναι σχετικά ικανοποιητικό. Παρότι, η μετρική MSE μας προϊδεάζει σε ένα όχι και τόσο καλό αποτέλεσμα, οι μετρικές R^2 , NMSE, NDEI έχουν σχετικά καλές τιμές. Αυτό συμβαίνει διότι, όσο πιο κοντά στην μονάδα είναι η R^2 και όσο πιο μικρές είναι οι NMSE, NDEI τόσο καλύτερο είναι το μοντέλο. Επιπλέον, στο διάγραμμα των καμπυλών εκμάθησης (Σχήμα 21) παρατηρούμε ότι το CheckError έχει σταθεροποιηθεί, πράγμα που σημαίνει ότι αν αυξήσουμε της εποχές εκπαίδευσης ίσως οδηγηθούμε στην υπερεκπαίδευση.

Στην περίπτωση που είχε γίνει χρήση της μεθόδου Grid Partitioning ο χρόνος δημιουργίας και εκπαίδευσης των μοντέλων θα ήταν υπερβολικά μεγάλος. Είναι γνωστό από την θεωρία με τη χρήση αυτής της κλασικής μεθόδου, ο αριθμός των κανόνων αυξάνεται σε σχέση με το πλήθος των εισόδων (ένα φαινόμενο που ονομάζεται «rule explosion»). Συνεπώς, ο αριθμός των κανόνων που τελικά θα δημιουργούνταν θα ήταν $2^{16} = 65536$ ή $3^{16} = 43046721$.

Αρχεία

- `ccrrp.m`: Επιλογή μοντέλου προς εκπαίδευση για το πρώτο dataset και υπολογισμός μετρικών αξιολόγησης.
- `plotmfin_CCPP.m`: Δημιουργία των διαγραμμάτων των συναρτήσεων συμμετοχής των εισόδων.
- `gridSearch_SubtractClust.m`: Υλοποίηση του Grid Search για την εύρεση του βέλτιστου μοντέλου.
- `infr_sq.m`: Εύρεση του κατάλληλου InfluenceRange και Squash για το κάθε μοντέλο.
- `optimum_model.m`: Εκπαίδευση του βέλτιστου μοντέλου και υπολογισμός των μετρικών.
- `plotmfin.m`: Δημιουργία των διαγραμμάτων των συναρτήσεων συμμετοχής των εισόδων.