ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών

Εργασία 3:
«Επίλυση Προβλήματος Παλινδρόμησης με χρήση μοντέλων TSK»

Υπολογιστική Νοημοσύνη

Διδάσκων: Θεοχάρης Ιωάννης

Γιάννης Τατσόγλου, 9568

Σεπτέμβριος, 2023

Περιεχόμενα

Μέρος 1 – Εφαρμογή σε μικρό dataset

Για το πρώτο μέρος της εργασίας, επιλέγεται το σύνολο δεδομένων που υπάρχει στο αρχείο «airfoil_self_noise.dat». Το αρχείο περιέχει συνολικά 1503 δείγματα και 6 χαρακτηριστικά, 5 εκ των οποίων είσοδοι και 1 έξοδο.

Σύμφωνα με τα ζητούμενα της εργασίας, εκπαιδεύτηκαν τέσσερα TSK μοντέλα με τις προδιαγραφές που φαίνονται στην παρακάτω εικόνα.

	Πλήθος συναρτήσεων συμμετοχής	Μορφή εξόδου
TSK_model_1	2	Singleton
TSK_model_2	3	Singleton
TSK_model_3	2	Polynomial
TSK_model_4	3	Polynomial

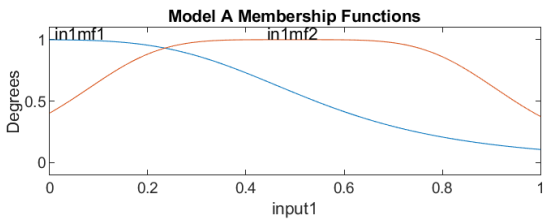
Αρχικά το dataset χωρίστηκε σε 3 μέρη, training, validation και test. Το training κομμάτι έχει μέγεθος το 60% του συνόλου των δειγμάτων και χρησιμοποιήθηκε για την αρχική εκπαίδευση του μοντέλου, το validation έχει μέγεθος το 20% του συνόλου των δειγμάτων και χρησιμοποιήθηκε για την επικύρωση και αποφυγή υπερεκπαίδευσης του μοντέλου, και το test έχει μέγεθος το 20% του συνόλου των δειγμάτων και χρησιμοποιήθηκε για τον έλεγχο της απόδοσης του μοντέλου.

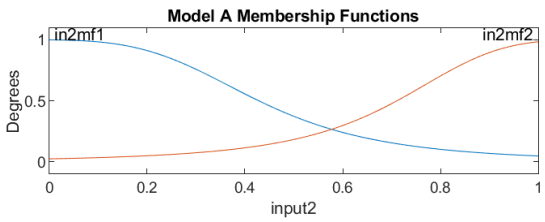
Για τον διαχωρισμό των δεδομένων στα επιμέρους υποσύνολα χρησιμοποιήθηκε η συνάρτηση split scale, η οποία εκτός από τον διαχωρισμό, κανονικοποιεί επιπλέον τα δεδομένα. Τα αποτελέσματα με την χρήση της παραπάνω συνάρτησης είναι πολύ καλύτερα από αυτά που προκύπταν με manual διαχωρισμό.

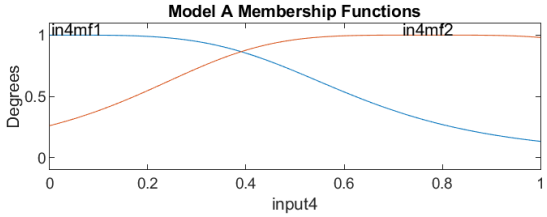
Ασαφή Σύνολα

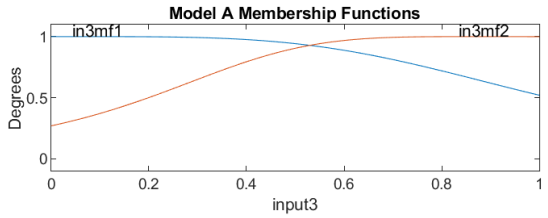
Για την εκπαίδευση των μοντέλων χρησιμοποιήθηκαν 200 εποχές. Παρακάτω παρουσιάζονται οι συναρτήσεις συμμετοχής που προέκυψαν για κάθε μοντέλο.

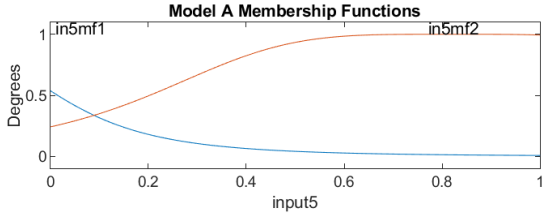
Μοντέλο 1 – Συναρτήσεις Συμμετοχής



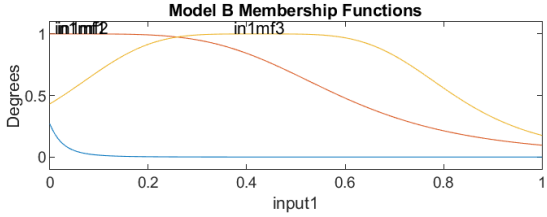


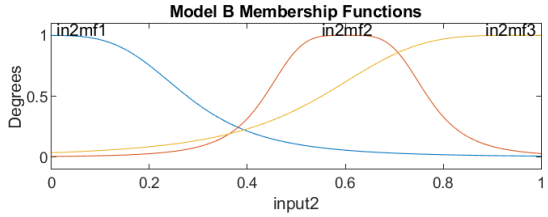


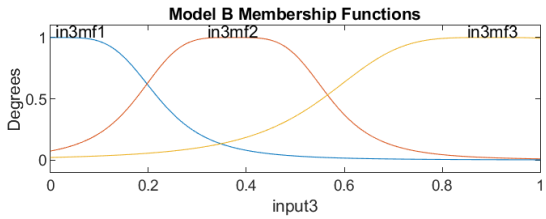


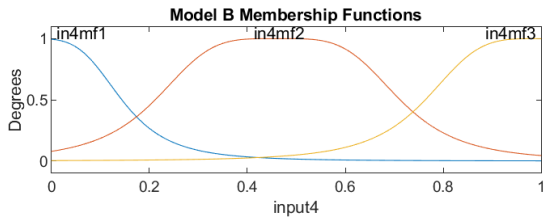


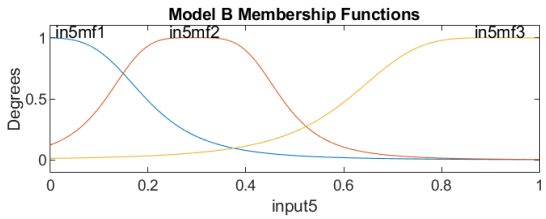
Μοντέλο 2 – Συναρτήσεις Συμμετοχής



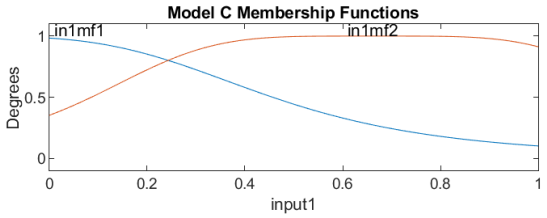


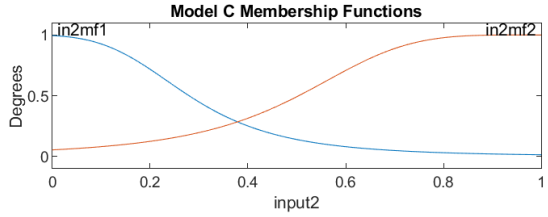


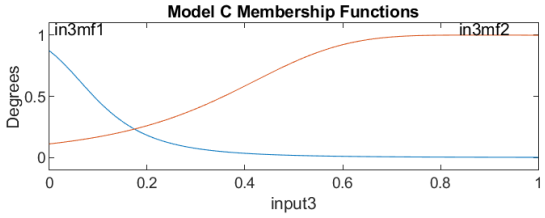


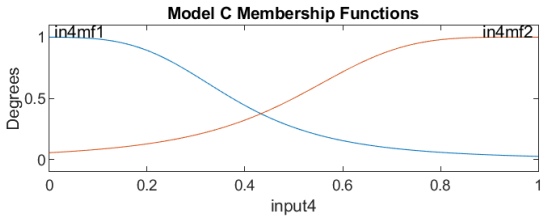


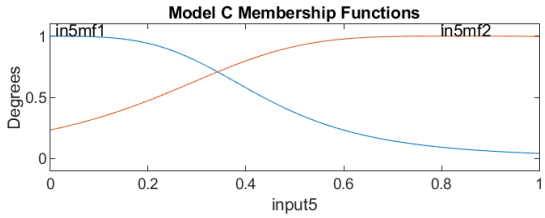
Μοντέλο 3 – Συναρτήσεις Συμμετοχής



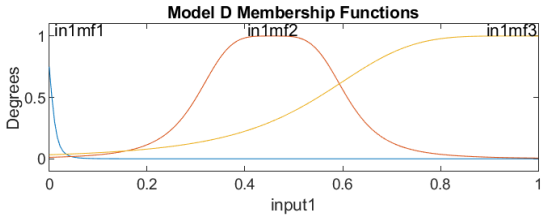


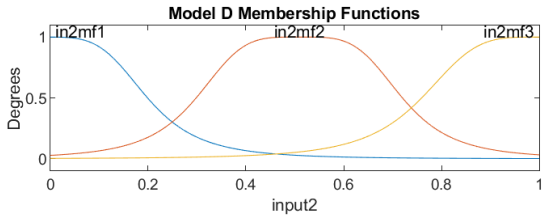


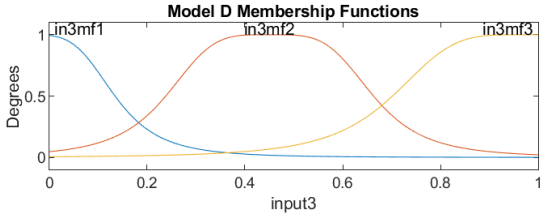


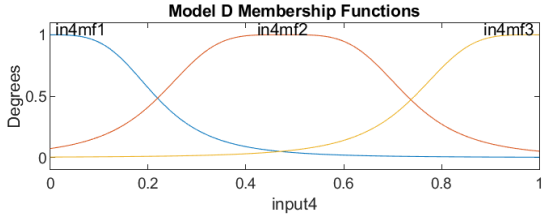


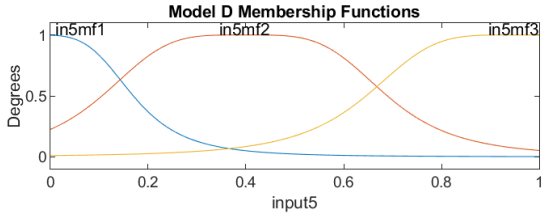
Μοντέλο 4 – Συναρτήσεις Συμμετοχής







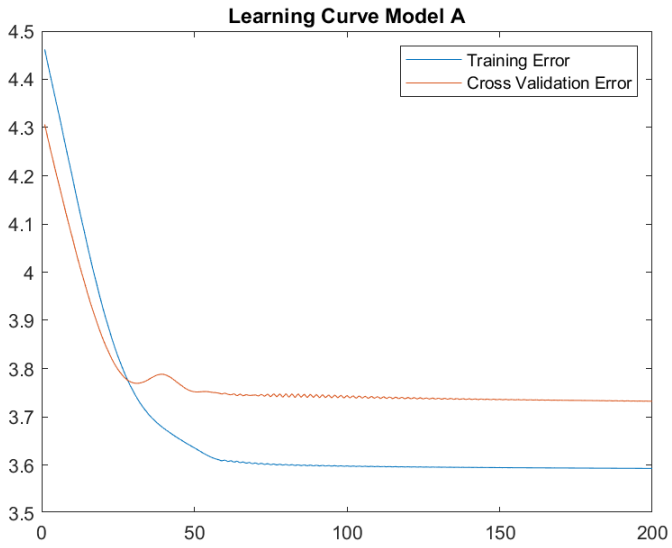




Καμπύλες Μάθησης

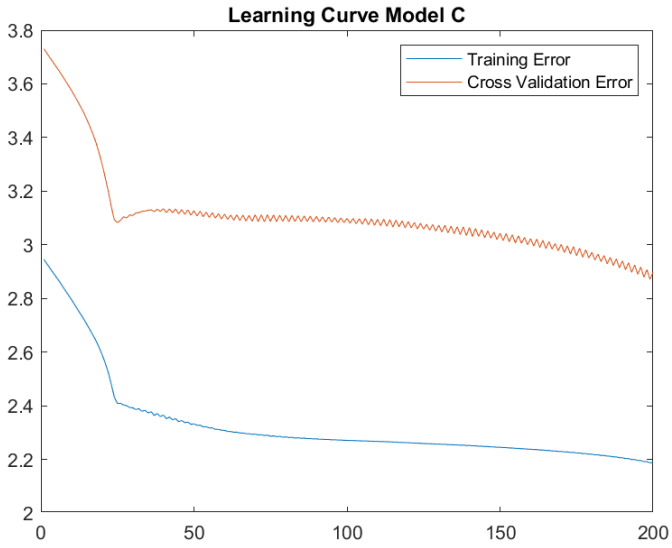
Παρακάτω ακολουθούν οι καμπύλες μάθησης για κάθε μοντέλο.

Μοντέλο 1 – Καμπύλη Μάθησης

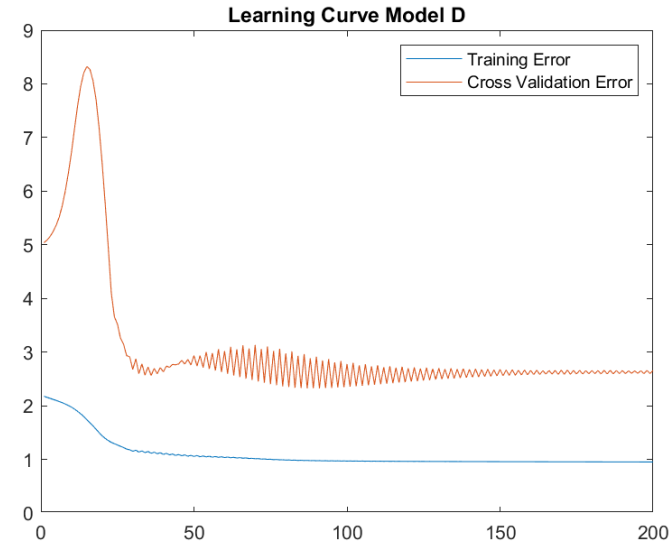


Μοντέλο 2 – Καμπύλη Μάθησης

Μοντέλο 3 – Καμπύλη Μάθησης



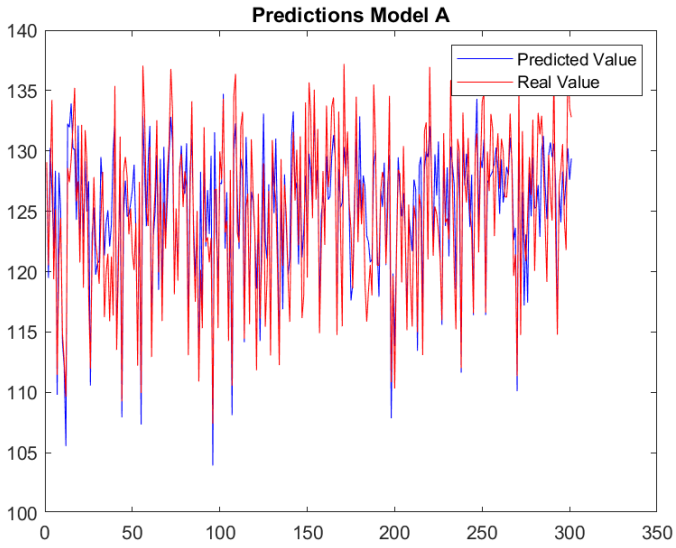
Μοντέλο 4 – Καμπύλη Μάθησης



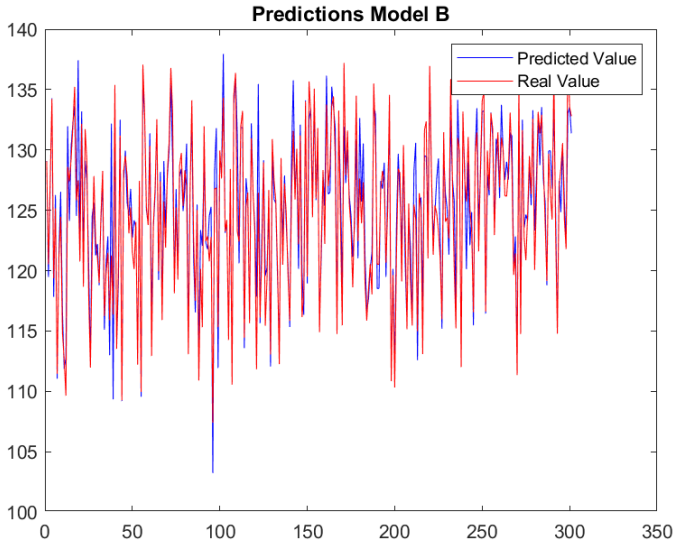
Σφάλματα Πρόβλεψης

Στα διαγράμματα που ακολουθούν, αποτυπώνονται για κάθε μοντέλο τα σφάλματα πρόβλεψης στο test dataset. Με την μπλε γραμμή αποτυπώνεται η πρόβλεψη του μοντέλου ενώ με την κόκκινη γραμμή αποτυπώνεται η πραγματική τιμή της εξόδου.

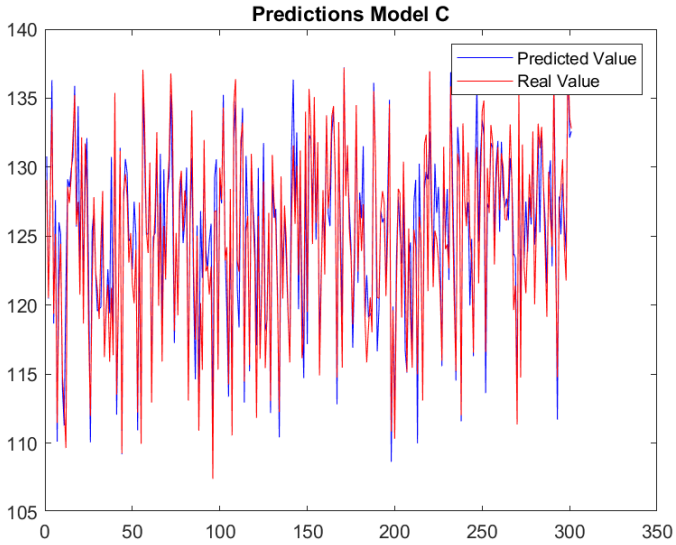
Μοντέλο 1 – Σφάλματα Πρόβλεψης



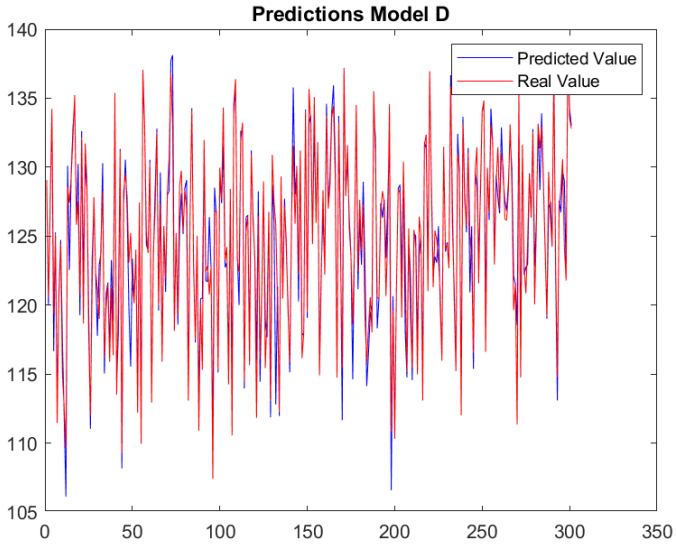
Μοντέλο 2 – Σφάλματα Πρόβλεψης



Μοντέλο 3 – Σφάλματα Πρόβλεψης



Μοντέλο 4 – Σφάλματα Πρόβλεψης



Μετρικές

Στον πίνακα που ακολουθεί παρατίθενται οι τιμές των μετρικών που προκύπτουν για τα μοντέλα που μελετήσαμε παραπάνω.

Μοντέλο	Συναρτήσεις Συμμετοχής	Τύπος Εξόδου	MSE	RMSE	NMSE	NDEI	R^2
1	2	Singleton	12,8414	3,5835	0,28902	0,5376	0,71098
2	3	Singleton	5,521	2,3497	0,12426	0,3525	0,87574
3	2	Polynomial	6,8628	2,6197	0,15446	0,39302	0,84554
4	3	Polynomial	3,7272	1,9306	0,083887	0,28963	0,91611

Σχολιασμός Αποτελεσμάτων

Αξιολογώντας τον πίνακα με τις μετρικές παραπάνω εξάγουμε τα εξής συμπεράσματα.

Αρχικά συγκρίνουμε τα μοντέλα 1 και 2, τα οποία έχουν έξοδο singleton και διαφέρουν ως προς τον αριθμό των συναρτήσεων συμμετοχής με 2 και 3 συναρτήσεις εισόδου αντίστοιχα. Παρατηρούμε ότι το μοντέλο 2 υπερτερεί σε απόδοση καθώς όλες οι μετρικές σφάλματός του (MSE, RMSE, NMSE, NDEI), έχουν μικρότερη τιμή ενώ ο συντελεστής προσδιορισμού R^2 είναι μεγαλύτερος.

Όμοια το ίδιο συμπέρασμα προκύπτει αν συγκρίνουμε τα μοντέλα 3 και 4 μεταξύ τους. Τα μοντέλα αυτά έχουν έξοδο πολυωνυμική και αριθμό συναρτήσεων συμμετοχής 2 και 3 αντίστοιχα. Πάλι παρατηρούμε ότι το μοντέλο με τον μεγαλύτερο αριθμό συναρτήσεων εισόδου (4) έχει καλύτερη απόδοση.

Στην συνέχεια συγκρίνουμε τα ζευγάρια μοντέλων 1,3 και 2,4 τα οποία έχουν τον ίδιο αριθμό συναρτήσεων εισόδου 2 και 3 αντίστοιχα και διαφέρουν ως προς τον τύπο της εξόδου ο οποίος είναι singleton για τα μοντέλα 1,2 και πολυωνυμικός για τα μοντέλα 3,4. Και στις δύο περιπτώσεις παρατηρούμε ότι το μοντέλο με πολυωνυμική έξοδο (3,4) υπερτερεί του μοντέλου με singleton έξοδο (1,2).

Συνοψίζοντας, σύμφωνα με τα στοιχεία του παραπάνω πίνακα και με τις συγκρίσεις που περιεγράφηκαν στις προηγούμενες παραγράφους, συμπεραίνουμε ότι τα μοντέλα με πολυωνυμική έξοδο υπερτερούν των μοντέλων με singleton έξοδο, και επιπλέον ο μεγαλύτερος αριθμός συναρτήσεων εισόδου βελτιώνει την απόδοση του εκάστοτε μοντέλου.

Τέλος, παρατηρούμε ότι στα μοντέλα 1 και 2 το cross-validation error συγκλίνει σε μια σταθερή τιμή με την πάροδο των εποχών εκπαίδευσης, στο μοντέλο 3 μειώνεται διαρκώς και ενδεχομένως να υπήρχε περαιτέρω βελτίωση στην απόδοση αν είχε εκπαιδευτεί για περισσότερες εποχές, ενώ το μοντέλο 4 είναι το μόνο για το οποίο μπορούμε να μιλήσουμε για υπερεκπαίδευση αφού η καμπύλη του cross-validation error παρουσιάζει μια πάρα πολύ ελαφριά αυξητική πορεία.

Μέρος 2 – Εφαρμογή σε μεγάλο dataset

Το dataset για το δεύτερο μέρος της εργασίας είναι το «superconduct.csv», το οποίο έχει υψηλό βαθμό διαστασιμότητας αποτελούμενο από 21.263 δείγματα και 81 χαρακτηριστικά. Συνεπώς δεν είναι δυνατόν να χρησιμοποιηθεί η προσέγγιση του πρώτου μέρους, καθώς ο αριθμός κανόνων θα αυξανόταν σε πάρα πολύ μεγάλο βαθμό.

Αντί αυτού, χρησιμοποιούμε την μέθοδο της επιλογής χαρακτηριστικών και διαμέρισης του χώρου εισάγοντας δυο ελεύθερες μεταβλητές, τον αριθμό χαρακτηριστικών προς επιλογή και τον αριθμό ομάδων που δημιουργούνται. Τις παραπάνω ελεύθερες μεταβλητές θα προσεγγίσουμε με την μέθοδο αναζήτησης πλέγματος (grid search).

Προσέγγιση Ελεύθερων Μεταβλητών

Το dataset χωρίζεται κατά 60%, 20% και 20% για τα training, validation και test αντίστοιχα. Για τον αριθμό των χαρακτηριστικών και την ακτίνα των clusters χρησιμοποίησα τις τιμές των παρακάτω πινάκων.

- Features_number = [3, 5, 10, 15]
- Cluster_radius = [0.55, 0.45, 0.35, 0.25]

Μέσω της 5-fold cross validation και για κάθε τιμή του αριθμού χαρακτηριστικών και ακτίνας cluster, εκπαιδεύεται ένα μοντέλο για 50 epochs και συγκρίνεται ο μέσος όρος σφάλματός του με τα υπόλοιπα. Στο τέλος επιλέγουμε τον συνδυασμό που παράγει το μοντέλο με το μικρότερο σφάλμα, το εκπαιδεύουμε και υπολογίζουμε τις διάφορες μετρικές.

Στον παρακάτω πίνακα απεικονίζονται τα σφάλματα που προκύπτουν από τους διάφορους συνδυασμούς.

NF / CR	3	5	10	15
0.55	23.9004	23.7499	23.7012	22.9684
0.45	22.3682	22.0686	21.1667	20.3218
0.35	20.3304	19.9733	19.0555	19.5926
0.25	19.0520	18.3424	17.6395	18.0139

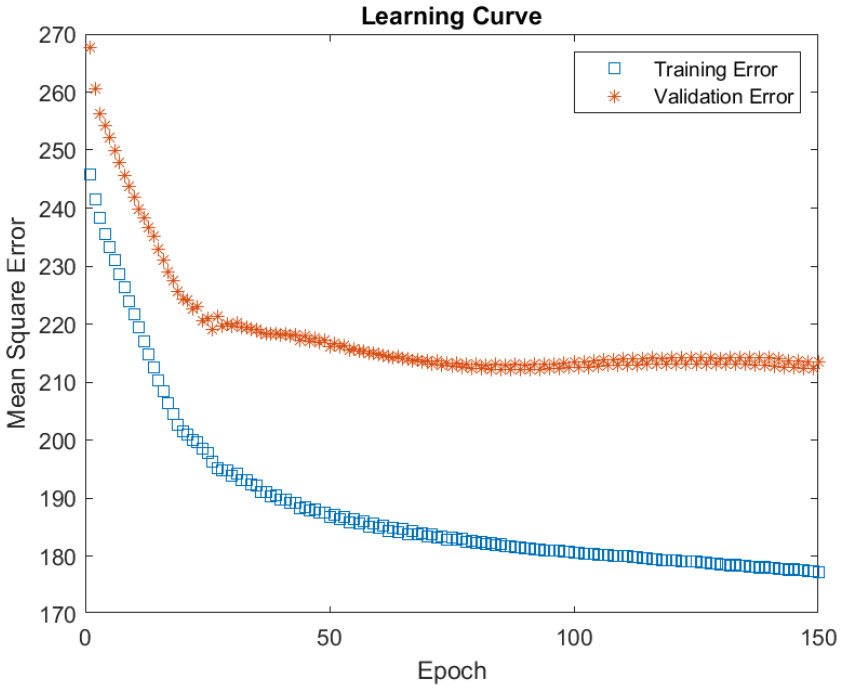
Είναι εμφανές ότι ο βέλτιστος συνδυασμός που προκύπτει είναι για αριθμό χαρακτηριστικών 10 και για ακτίνα cluster 0.25

Εκπαίδευση Μοντέλου – Χαρακτηριστικά

Στην συνέχεια εκπαιδεύτηκε ξανά το μοντέλο με τα βέλτιστα χαρακτηριστικά, αυτή την φορά για 150 epochs.

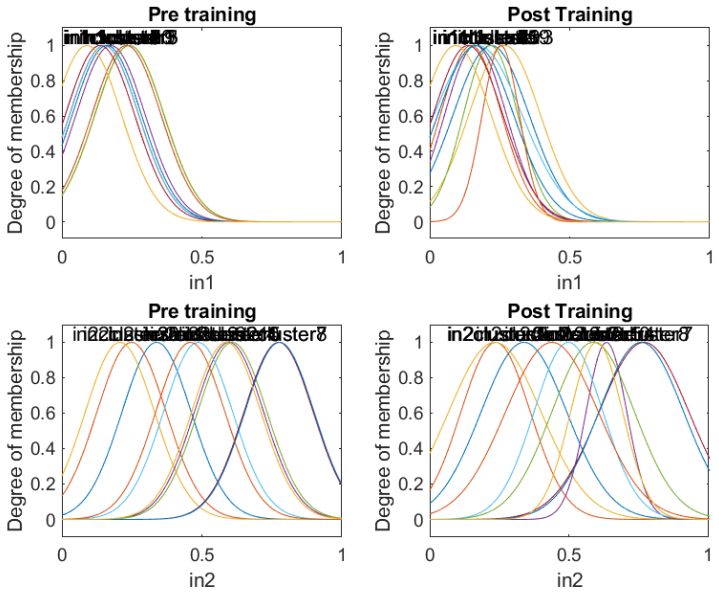
Στα ακόλουθα διαγράμματα παρουσιάζονται τα χαρακτηριστικά του μοντέλου.

Καμπύλη Μάθησης.



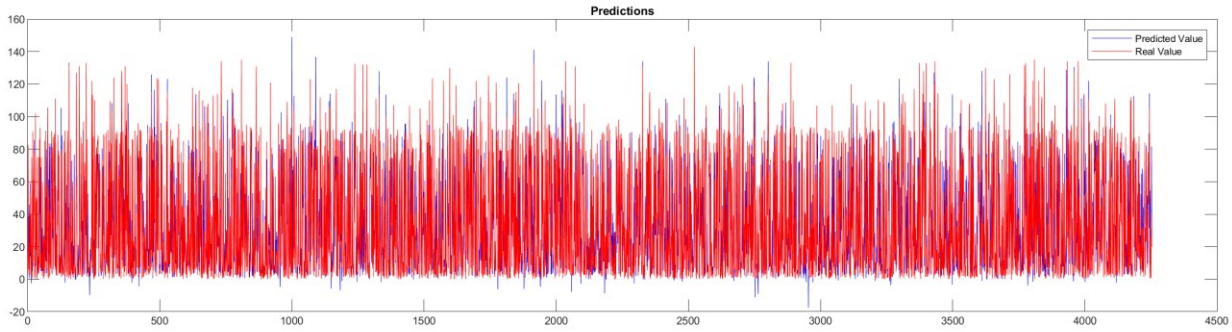
Συναρτήσεις Συμμετοχής.

Πριν και μετά την βελτιστοποίηση



Σφάλματα Πρόβλεψης

Με μπλε χρώμα απεικονίζεται η τιμή που δίνει το μοντέλο ενώ με κόκκινο η πραγματική τιμή.



Μετρικές

Στον παρακάτω πίνακα φαίνονται οι μετρικές που προκύπτουν για το βέλτιστο μοντέλο.

RMSE	NMSE	NDEI	R^2
14.3165	0.1725	0.41534	0.8275

Σχολιασμός Αποτελεσμάτων

Παρατηρούμε από τον πίνακα με τις μετρικές ότι ο δείκτης RMSE είναι σχετικά μεγάλος, οι υπόλοιπες μετρικές όμως έχουν μικρές και σε ικανοποιητικά επίπεδα τιμές. Συγκρίνοντας με τα αποτελέσματα του πρώτου μέρους, θα λέγαμε ότι το μοντέλο που προκύπτει δεν είναι τόσο ακριβές, γεγονός που οφείλεται στο ότι δεν χρησιμοποιούμε όλες τις μεταβλητές εισόδου για την εξαγωγή του αποτελέσματος αλλά μέρος αυτών.

Σε αντίθεση με τα μοντέλα του πρώτου μέρους όμως, το μοντέλο που προέκυψε για αυτό το μέρος χρειάστηκε αισθητά λιγότερο χρόνο και οι προβλέψεις του είναι αρκετά ικανοποιητικές.

Αν είχαμε επιλέξει την μέθοδο του grid partitioning με 81 χαρακτηριστικά θα έπρεπε να χρησιμοποιηθούν 2^{81} η 3^{81} για 2 και 3 εισόδους αντίστοιχα, πράγμα που θα έκανε την υλοποίηση απαγορευτική σε χρόνο και πόρους. Στην δική μας περίπτωση παρήχθησαν 10 κανόνες.