

Υπολογιστική Νοημοσύνη

2η Εργασία στα Ασαφή Συστήματα

Επίλυση προβλήματος παλινδρόμησης με χρήση
μοντέλων TSK

Κωνσταντίνος Κωνσταντινίδης, ΑΕΜ: 9162

Στοιχεία Επικοινωνίας:

- email: konkonstantinidis@ece.auth.gr

Για παράδοση κατά την πτυχιακή εξεταστική του Φεβρουαρίου 2022

1. Εφαρμογή σε απλό dataset

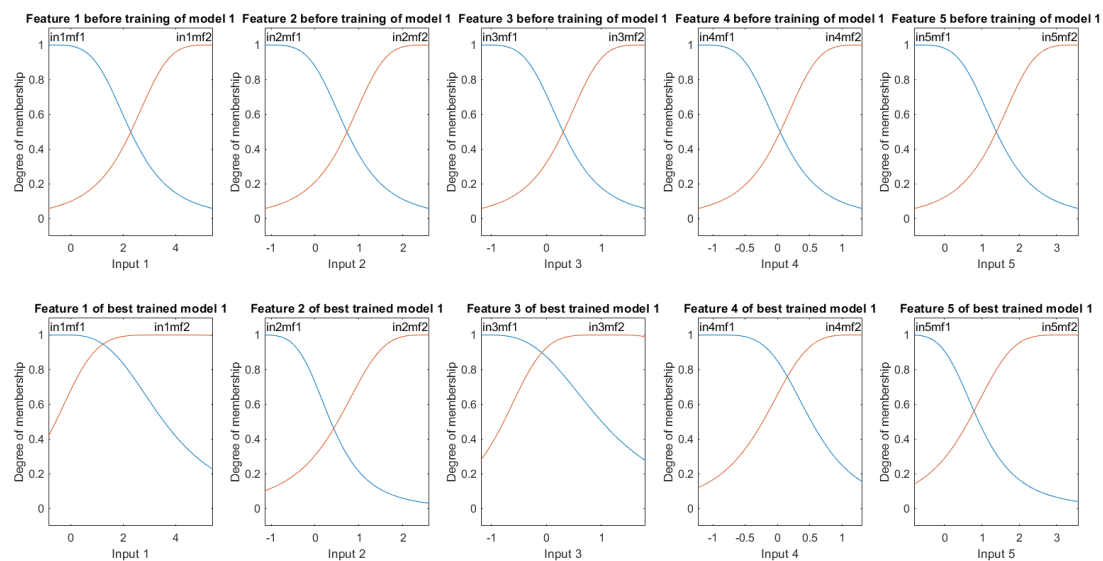
Για την εφαρμογή TSK μοντέλων παλινδρόμησης επάνω στο Airfoil Self-Noise dataset, έχουμε τα εξής:

Τα μοντέλα και οι παράμετροι τους:

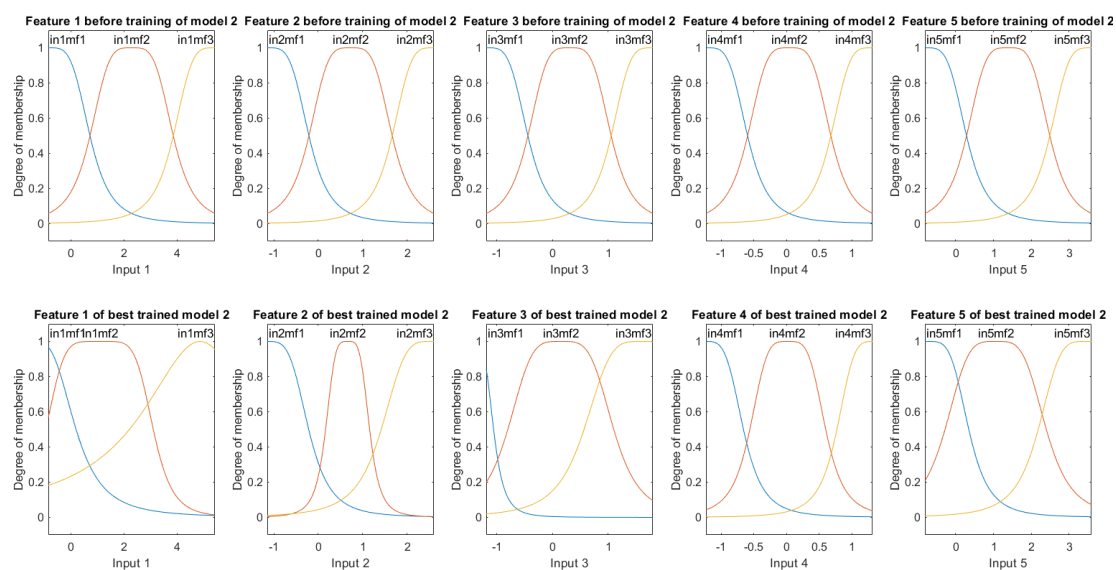
Πλήθος συναρτήσεων συμμετοχής		Μορφή εξόδου
TSK_model_1	2	Singleton
TSK_model_2	3	Singleton
TSK_model_3	2	Polynomial
TSK_model_4	3	Polynomial

Εικόνα 1, Παράμετροι μοντέλων προς εκπαίδευση

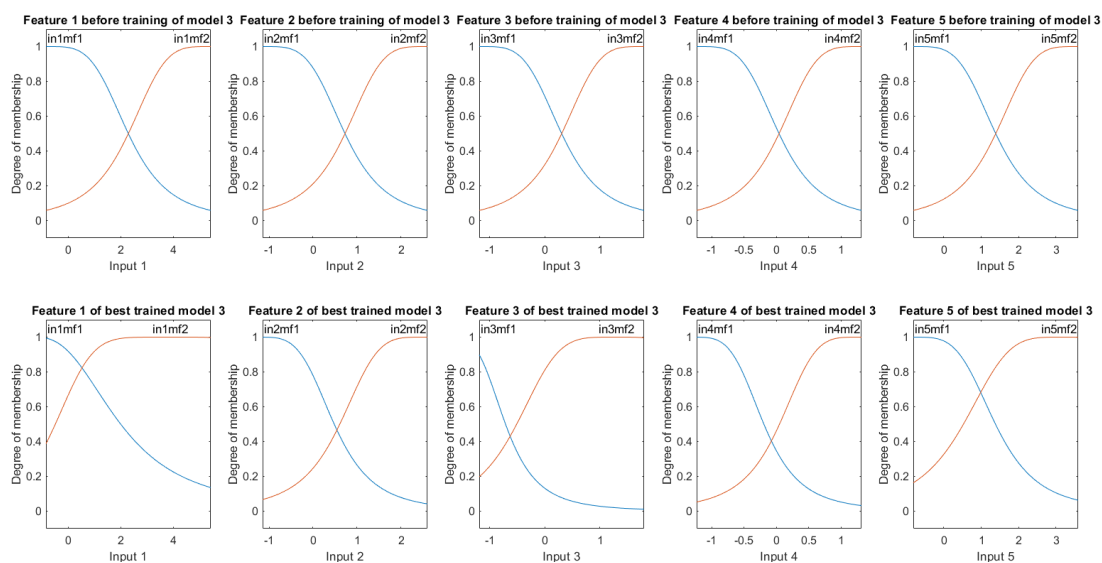
Αρχικά, παρουσιάζονται τα διαγράμματα με τις αρχικές και τελικές μορφές των ασαφών συνόλων, για κάθε ένα από τα 4 μοντέλα:



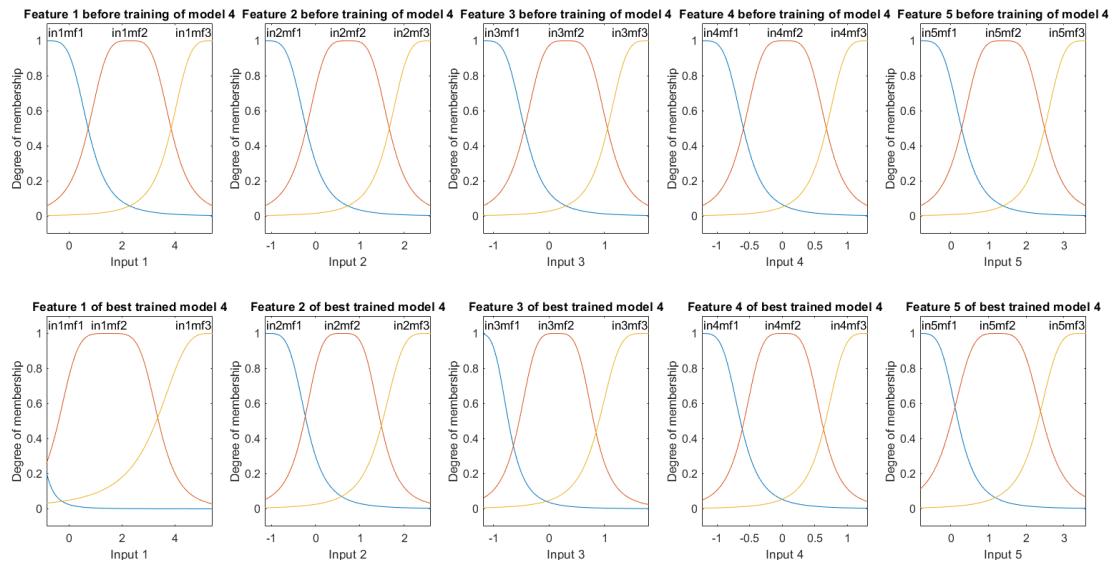
Εικόνα 2, Διαγράμματα αρχικών - τελικών ασαφών συνόλων για το TSK_model_1



Εικόνα 3, Διαγράμματα αρχικών - τελικών ασαφών συνόλων για το TSK_model_2

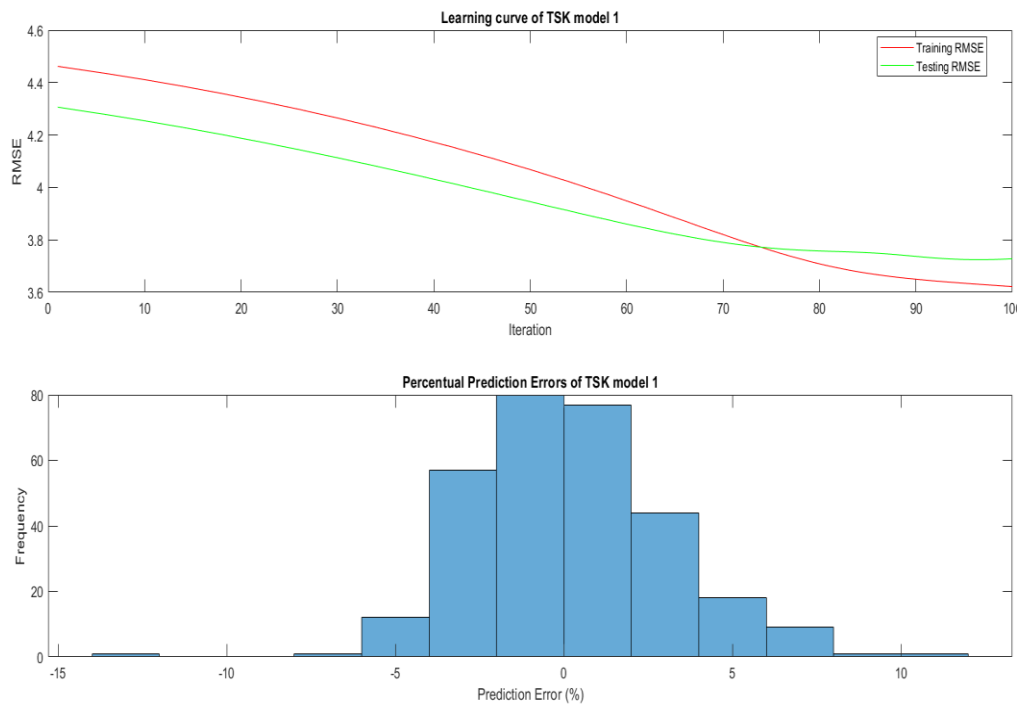


Εικόνα 4, Διαγράμματα αρχικών - τελικών ασαφών συνόλων για το TSK_model_3

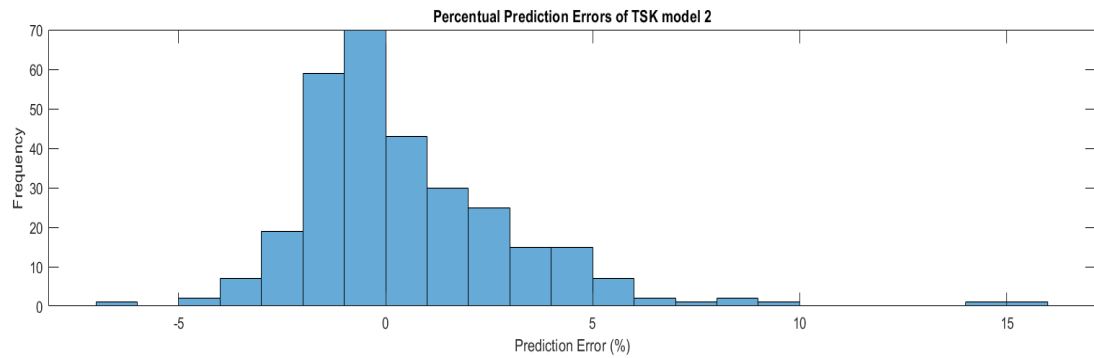
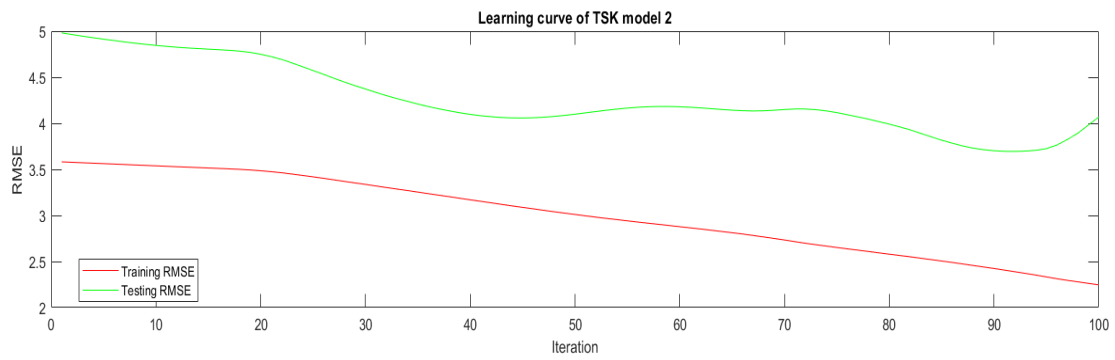


Εικόνα 5, Διαγράμματα αρχικών - τελικών ασαφών συνόλων για το TSK_model_4

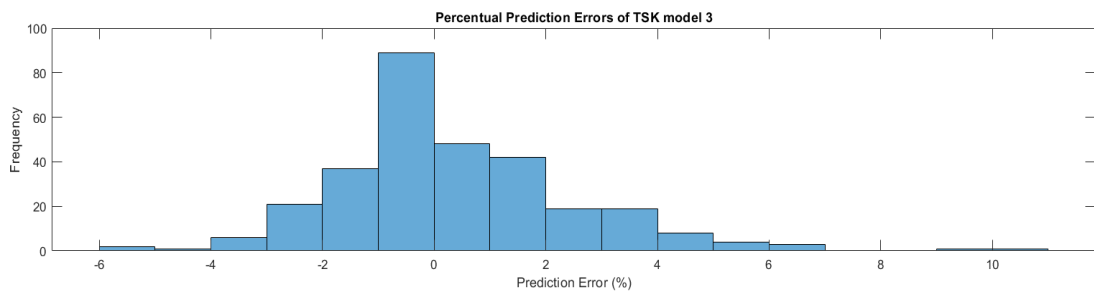
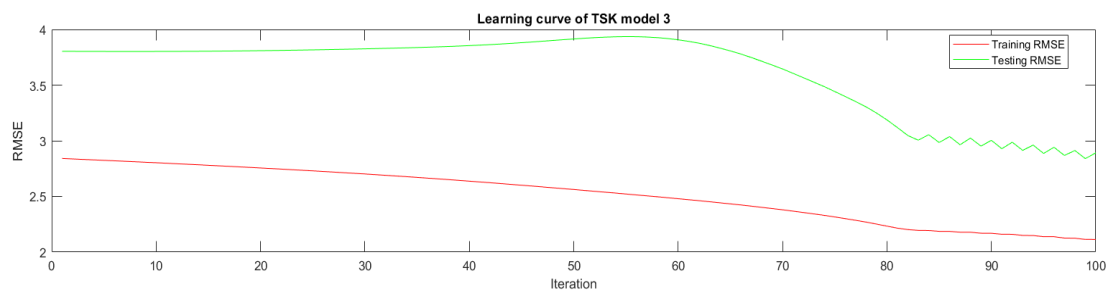
Έπειτα, στην ίδια εικόνα για κάθε μοντέλο απεικονίζονται οι καμπύλες μάθησης για τα training και validation data, καθώς και τα ποσοστιαία σφάλματα πρόβλεψης:



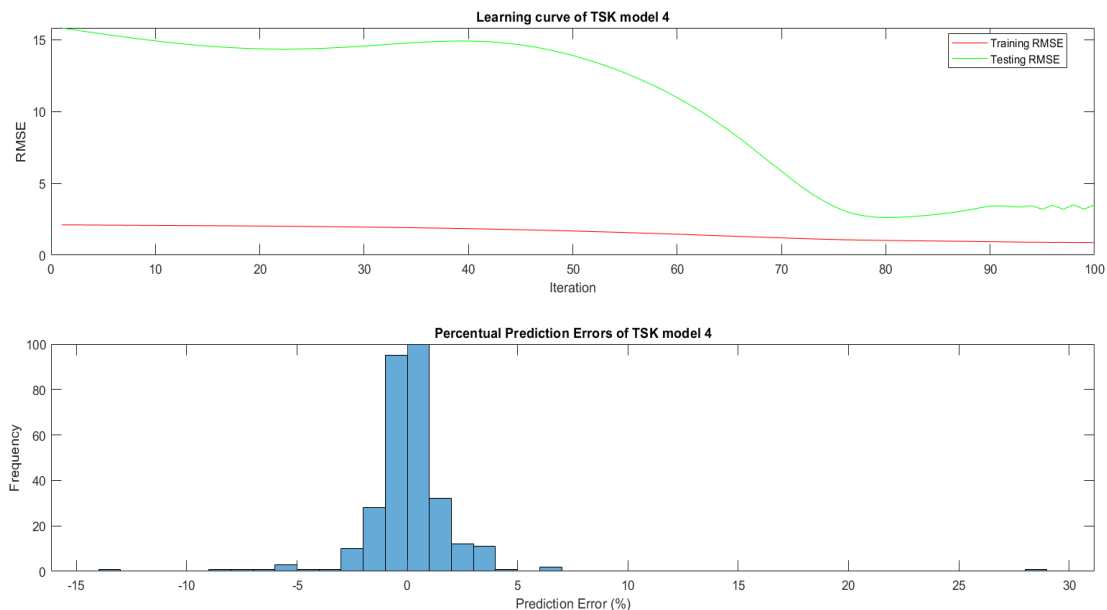
Εικόνα 6, Καμπύλες μάθησης (πάνω) και ποσοστιαία σφάλματα πρόβλεψης (κάτω) του TSK_model_1



Εικόνα 7, Καμπύλες μάθησης (πάνω) και ποσοστιαία σφάλματα πρόβλεψης (κάτω) του TSK_model_2



Εικόνα 8, Καμπύλες μάθησης (πάνω) και ποσοστιαία σφάλματα πρόβλεψης (κάτω) του TSK_model_3



Εικόνα 9, Καμπύλες μάθησης (πάνω) και ποσοστιαία σφάλματα πρόβλεψης (κάτω) του *TSK_model_4*

Τέλος, έχουμε ένα συγκεντρωτικό πίνακα με μετρικές απόδοσης για το κάθε μοντέλο:

	RMSE	NMSE	NDEI	R2
TSK_model_1	3.6137	0.29391	0.54213	0.70609
TSK_model_2	3.3206	0.24816	0.49816	0.75184
TSK_model_3	2.617	0.15414	0.3926	0.84586
TSK_model_4	2.9904	0.20127	0.44863	0.79873

Εικόνα 10, Πίνακας μετρικών απόδοσης των μοντέλων

Παρατηρήσεις / Συμπεράσματα:

Αρχικά, ενώ κάποια μοντέλα τα πήγαν καλύτερα από κάποια άλλα, σε γενικές γραμμές όλα τα μοντέλα έχουν καλή απόδοση (μεγάλα R2, μικρά σφάλματα). Όλα τα μοντέλα έχουν προσαρμόσει τα αρχικά ασαφή σύνολα, σε διαφόρους βαθμούς το καθένα.

Έπειτα, παρατηρείται ότι τα μοντέλα με πολυωνυμική μορφή εξόδου έχουν καλύτερη απόδοση από τα μοντέλα με Singleton μορφή εξόδου, ενώ συγκεκριμένα το μοντέλο με την καλύτερη απόδοση είναι ξεκάθαρα το *TSK_model_3*, το οποίο έχει δύο (2) συναρτήσεις συμμετοχής και πολυωνυμική μορφή εξόδου.

Τέλος, τα μοντέλα με δύο (2) συναρτήσεις συμμετοχής τα πηγαίνουν καλύτερα από τα αντίστοιχα μοντέλα με τρεις (3) συναρτήσεις συμμετοχής (για ίδια μορφή εξόδου). Ο λόγος για αυτό μπορεί εύκολα να φανεί εάν συγκρίνουμε τις καμπύλες μάθησης των μοντέλων αυτών (εικόνες 6 vs 7 και 8 vs 9). Από αυτές, φαίνεται ότι τα μοντέλα με παραπάνω συναρτήσεις συμμετοχής είναι πιο επιρρεπή στο overfitting, αφού για εκείνα το testing (=validation εδώ) error δεν μειώνεται διαρκώς μαζί με το training error ή μένει σταθερό, αλλά από κάποιο σημείο και μετά

θα εμφανίσει ανοδική πορεία ενώ το training error συνεχίζει να μειώνεται, υπονοώντας ότι από αυτό το σημείο και μετά, το μοντέλο περνά περισσότερο σε αποστήθιση του σετ εκπαίδευσης παρά σε εκμάθηση από αυτό, κάτι που είναι πιο εύκολο να γίνει όταν έχουμε παραπάνω συναρτήσεις συμμετοχής.

2. Εφαρμογή σε dataset με υψηλή διαστασιμότητα

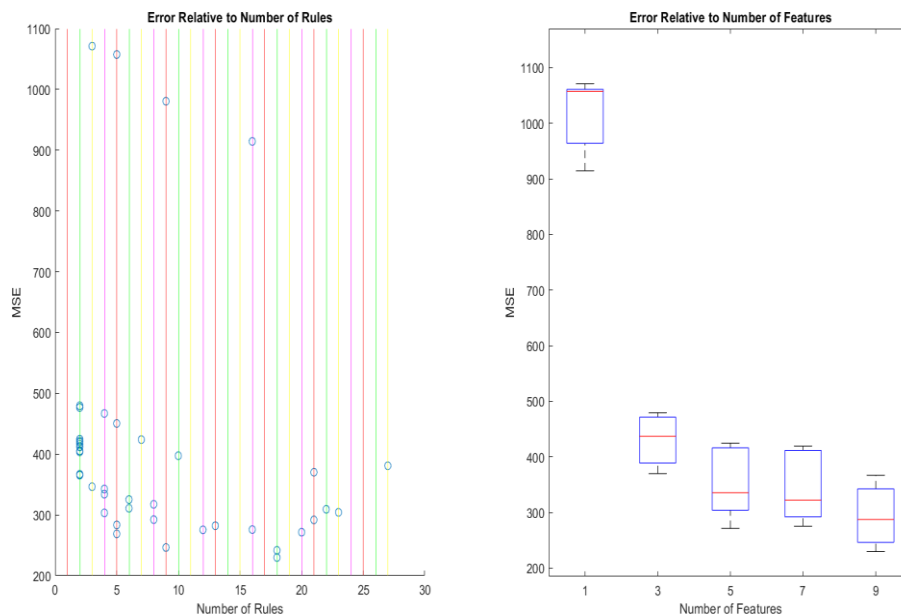
Στην δεύτερη φάση της εργασίας, προχωράμε στο Superconductivity dataset, στο οποίο θα πρέπει αρχικά να γίνει μείωση διαστάσεων για να εκπαιδευτούν επάνω του TSK μοντέλα, καθώς ο αρχικός αριθμός των χαρακτηριστικών (81) είναι απαγορευτικός για την εκπαίδευση TSK μοντέλων λόγω της μεγάλης υπολογιστικής πολυπλοκότητας τους.

Για να γίνει μείωση διαστάσεων, αναλύεται το σύνολο των χαρακτηριστικών με βάση των Relief αλγόριθμο και έπειτα κρατούνται τα $\{1,3,5,7,9\}$ πιο σημαντικά χαρακτηριστικά. Σαν δεύτερη, επίσης μεταβαλλόμενη παράμετρος, θεωρείται η ακτίνα r_a των cluster, που έμμεσα προσδιορίζει τον αριθμό των κανόνων που θα προκύψουν. Η παράμετρος αυτή παίρνει τις τιμές $\{0.1, 0.2, 0.3, 0.4, \dots, 0.9, 1\}$.

Για τους συνδυασμούς των δύο αυτών παραμέτρων λοιπόν, εφαρμόζουμε αναζήτηση πλέγματος και κάθε μοντέλο αξιολογείται με βάση την μετρική Mean Squared Error (MSE), εφαρμόζοντας 5-fold cross validation, σύμφωνα με τις οδηγίες της εκφώνησης.

Για την αναζήτηση πλέγματος, έχουμε τα εξής αποτελέσματα:

Αρχικά, το error (MSE) συναρτήσει των αριθμού των κανόνων και τον αριθμό των επιλεγμένων χαρακτηριστικών:



Εικόνα 11, MSE συναρτήσει αριθμού κανόνων (αριστερά) και αριθμού χαρακτηριστικών (δεξιά)

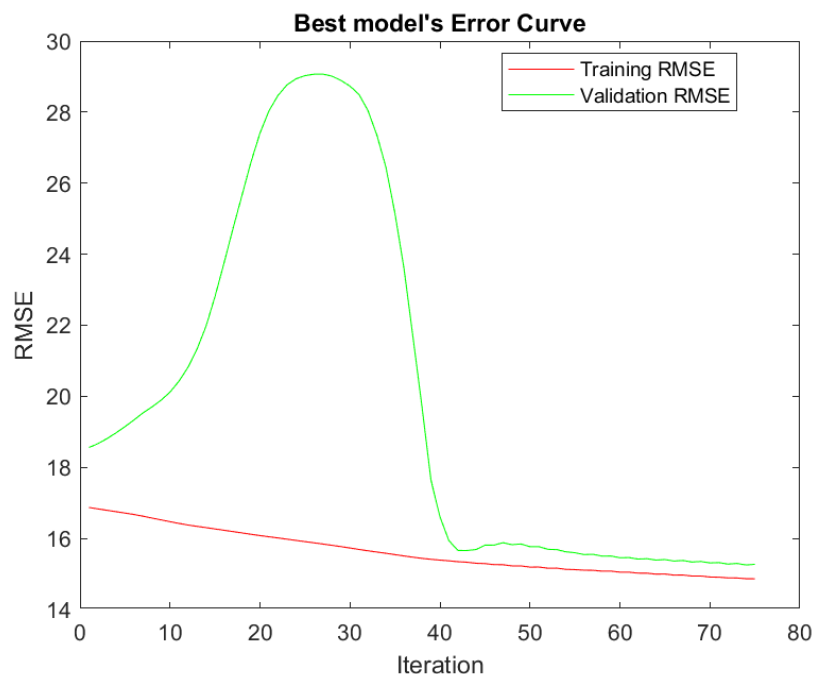
Σύμφωνα με τα αποτελέσματα στην εικόνα 11, δεν παρατηρείται να υπάρχει κάποια συσχέτιση μεταξύ του σφάλματος και του αριθμού των κανόνων που θα έχει το μοντέλο. Από την άλλη, ακόμη και χωρίς την χρήση scatterplot, είναι προφανές ότι όσα παραπάνω χαρακτηριστικά έχει το μοντέλο, τόσο μικρότερες οι μετρικές του σφάλματος. Αυτό ωστόσο ήταν αναμενόμενο, καθώς λιγότερα χαρακτηριστικά σημαίνει λιγότερη πληροφορία, και με βάση τον αλγόριθμο Relief, απλά κρατήσαμε λιγότερα χαρακτηριστικά (=λιγότερες στήλες του dataset), χωρίς να επιχειρήσουμε κάποιο μετασχηματισμό/άλλη μέθοδο μείωσης διαστάσεων ώστε να κρατήσουμε ένα μεγαλύτερο ποσοστό της πληροφορίας, για ίδιο αριθμό στηλών (ICA/PCA).

Με βάση την αναζήτηση πλέγματος, το καλύτερο μοντέλο (με το μικρότερο MSE) είναι εκείνο με εννέα (9) χαρακτηριστικά και ακτίνα $r_a = 0.3$. Για το μοντέλο αυτό, έχουμε τις εξής μετρικές:

RMSE	NMSE	NDEI	R2
15.185	0.19174	0.43788	0.80826

Εικόνα 12, Μετρικές μοντέλου με 9 χαρακτηριστικά και ακτίνα cluster $r = 0.1$

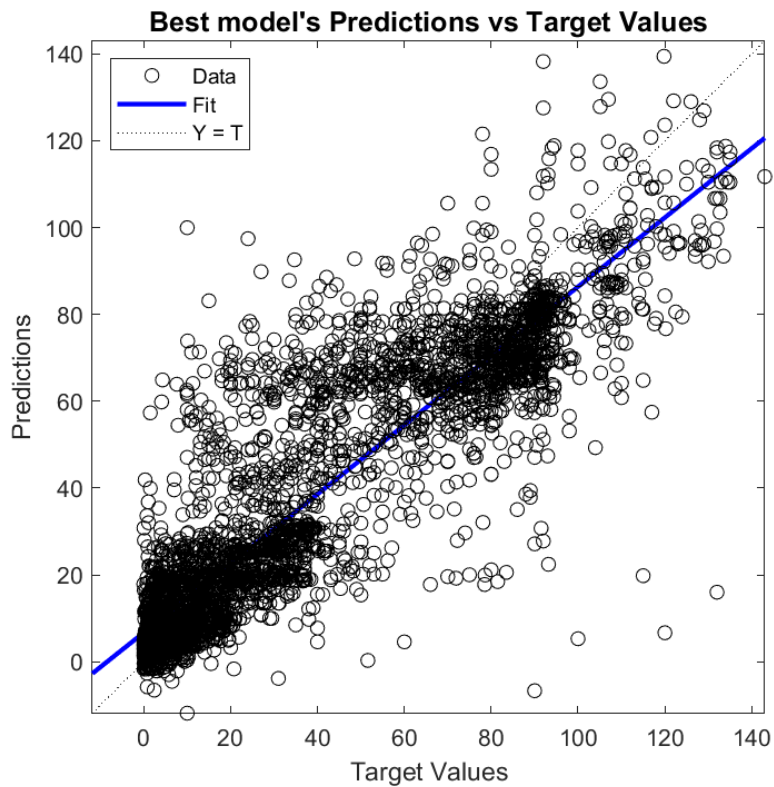
Όσον αφορά την εκπαίδευση του εν λόγω μοντέλου, παρουσιάζονται οι καμπύλες εκμάθησης για τα training και validation data:



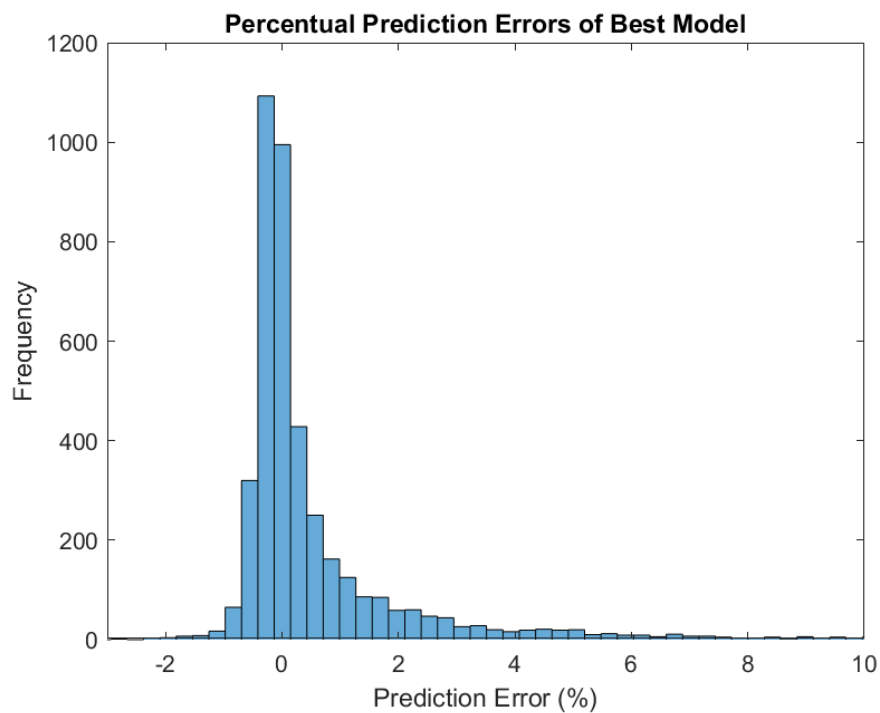
Εικόνα 13, Καμπύλες Εκμάθησης του μοντέλου

Από την εικόνα 13 διασφαλίζεται πως το μοντέλο δεν παρουσιάζει overfitting στο σετ εκπαίδευσης (τελικά).

Έπειτα, για τις προβλέψεις και τα σφάλματα πρόβλεψης του μοντέλου:

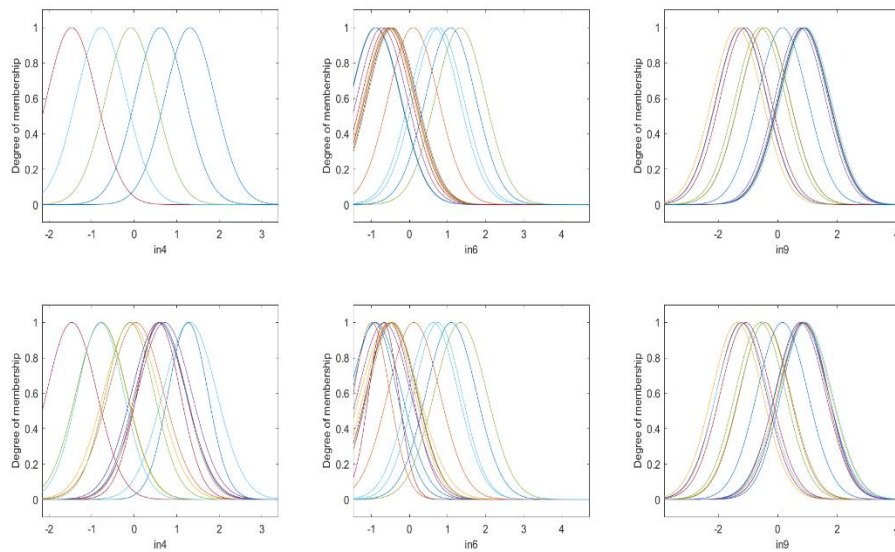


Εικόνα 14, Προβλέψεις του μοντέλου σε σχέση με τις πραγματικές τιμές



Εικόνα 15, Ποσοστιαία σφάλματα πρόβλεψης του μοντέλου

Τέλος, ενδεικτικά παρουσιάζονται οι αρχικές και τελικές καταστάσεις ορισμένων ασαφών συνόλων, πριν και μετά την εκπαίδευση του μοντέλου με τις βέλτιστες παραμέτρους:



Εικόνα 16, Αρχική και τελική μορφή των ασαφή συνόλων των εισόδων 4, 6 και 9

Παρατηρήσεις / Συμπεράσματα:

Το παραπάνω μοντέλο, όπως προέκυψε από την εκπαίδευση επάνω στα δεδομένα, παρέχει αρκετά ικανοποιητικές μετρικές απόδοσης επάνω στο testing σετ, ενώ είχε συνολικά 15 κανόνες. (bestValFis_best.Rules). Από την άλλη, τα μοντέλα του μέρους 1 θα είχαν 2^k ή 3^k κανόνες, όπου k ο αριθμός των χαρακτηριστικών, δηλαδή έως και 2^9 κανόνες για την συγκριμένη εφαρμογή, κάτι που οδηγεί σε απίστευτη υπολογιστική πολυπλοκότητα. Από αυτό φαίνεται λοιπόν η ανωτερότητα του Subtractive Clustering αλγορίθμου έναντι του Grid Partition, όσον αφορά τον αριθμό των εισόδων του μοντέλου.