

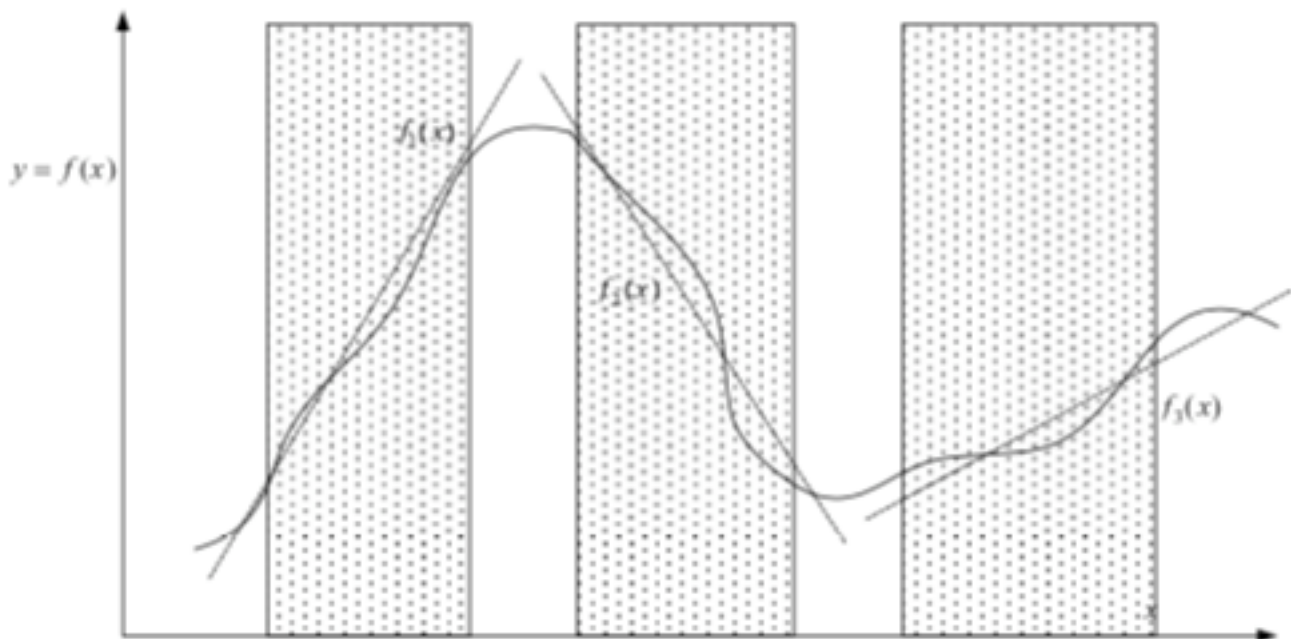
Υπολογιστική Νοημοσύνη

Εργασία 3-Regression

Καράμπελας Σάββας

AEM:9005

savvaska@ece.auth.gr



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Περιγραφή Εργασίας

Στα πλαίσια αυτής της εργασίας διερευνούμε την 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 νευρωνικών μοντέλων. Το πρώτο σύνολο δεδομένων χρησιμοποιείται για μία απλή διερεύνηση 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 χρησιμοποιείται για μια πληρέστερη διαδικασία μοντελοποίησης, η οποία περιλαμβάνει μεταξύ άλλων προεπεξεργαστικά βήματα όπως είναι η επιλογή χαρακτηριστικών(feature selection) καθώς και μεθόδους βελτιστοποίησης των μοντέλων μέσα της διασταυρωμένης επικύρωσης(cross validation).

Εφαρμογή σε απλό dataset

Για το πρώτο κομμάτι της εργασίας, χρησιμοποιούμε από το UCI repository το Airfoil Self-Noise dataset, το οποίο περιλαμβάνει 1503 δείγματα (instances) και 6 χαρακτηριστικά μαζί με αυτό της εξόδου. Το σύνολο δεδομένων αυτό παρέχει πληροφορίες για διάφορες αεροτομές που ελέγχθηκαν από την διαστημική υπηρεσία NASA σε διάφορες συνθήκες σε αεροδυναμικές σήραγγες. Οι είσοδοι είναι η συχνότητα σε Hertz, η γωνία προσβολής(angle of attack)σε γωνίες, το μήκος καλωδίου σε μέτρα, η free stream ταχύτητα σε μέτρα ανά δευτερόλεπτο και το displacement thickness σε μέτρα. Σαν έξοδο έχουμε το επίπεδο ηχητικής πίεσης που ασκείται στην αεροτομή εκφρασμένη σε decibels. Τα βήματα που θα ακολουθήσουμε είναι τα παρακάτω:

- 1) Θα διαχωρίσουμε το dataset σε σύνολα εκπαίδευσης, επικύρωσης και ελέγχου. Είναι απαραίτητο τα τρία αυτά σύνολα να είναι μη επικαλυπτόμενα. Το σύνολο εκπαίδευσης θα είναι τα δεδομένα με βάση τα οποία θα λάβει χώρα η εκπαίδευση, το σύνολο επικύρωσης θα χρησιμοποιηθεί για επικύρωση και αποφυγή του φαινομένου υπερεκπαίδευσης και το τελευταίο χρησιμοποιείται για τον έλεγχο της απόδοσης του τελικού μοντέλου που προκύπτει. Αναλυτικότερα, θα έχουμε το 60% των δειγμάτων ως υποσύνολο εκπαίδευσης και από 20 % του συνόλου των δειγμάτων για κάθε ένα από τα δύο εναπομείναντα υποσύνολα. Ουσιαστικά σε πρώτο στάδιο μας ενδιαφέρουν τα υποσύνολα

εκπαίδευσης και επικύρωσης καθώς κατά την διάρκεια του training επιλέγουμε κάθε φορά το μοντέλο που έχει το χαμηλότερο σφάλμα επικύρωσης. Επίσης, κανονικοποιούμε τα δεδομένα εισόδου στο $[0,1]$.

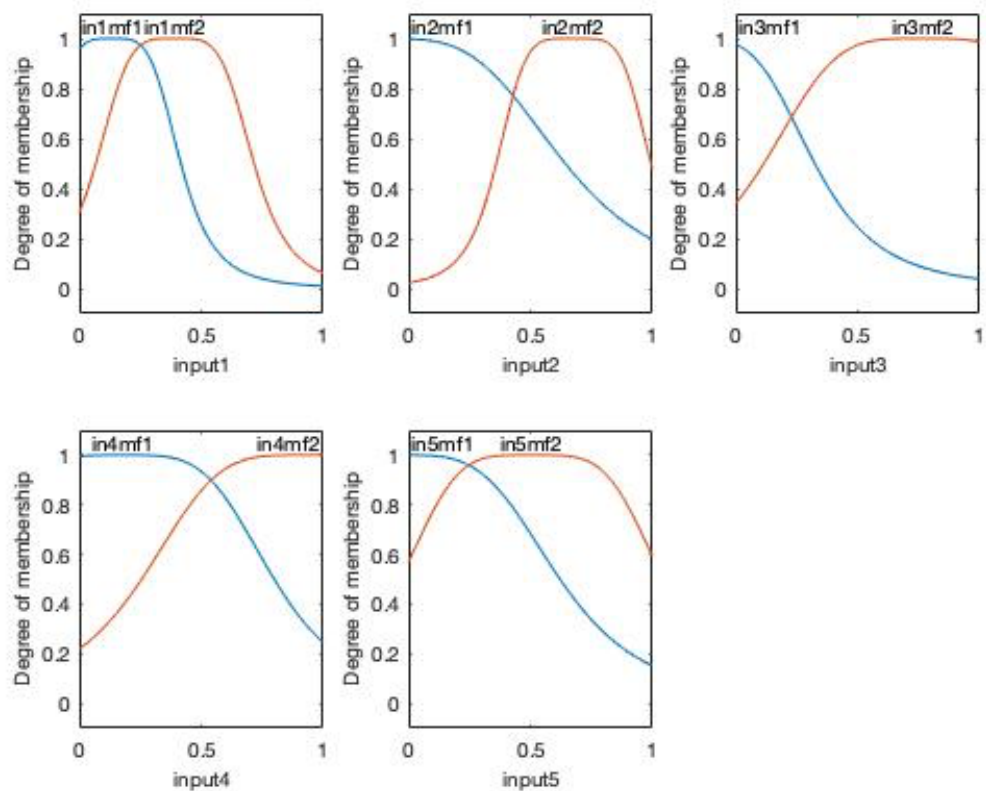
- 2) Θα διεξάγουμε εκπαίδευση για τέσσερα μοντέλα TSK , με διαφορετικές παραμέτρους, και θα εξετάσουμε την απόδοση τους στο σύνολο ελέγχου. Στα μοντέλα αυτά μεταβάλλεται η μορφή της εξόδου καθώς και το πλήθος των συναρτήσεων συμμετοχής για κάθε μεταβλητή εισόδου. Παρακάτω έχει προστεθεί ο πίνακας που περιέχει τις απαραίτητες πληροφορίες για τα μοντέλα.

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

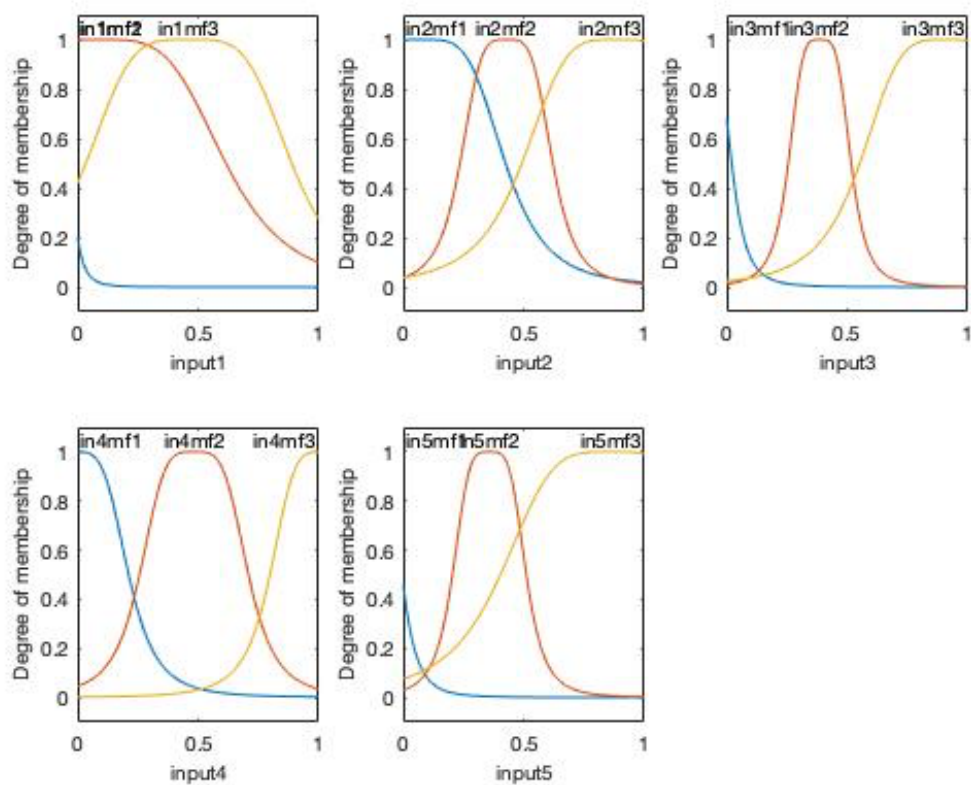
Τέλος, τα μοντέλα θα εκπαιδευτούν με την υβριδική μέθοδο, σύμφωνα με την οποία οι παράμετροι των συναρτήσεων συμμετοχής βελτιστοποιούνται μέσω της μεθόδου της οπισθοδρόμησης(backpropagation algorithm), ενώ οι παράμετροι της πολυωνυμικής συνάρτησης εξόδου βελτιστοποιούνται μέσω της μεθόδου των ελάχιστων τετραγώνων (Least Squares). Οι συναρτήσεις συμμετοχής είναι γκαουσιανές και αρχικοποιούνται έτσι ώστε τα διαδοχικά ασαφή σύνολα να παρουσιάζουν σε κάθε είσοδο, βαθμό επικάλυψης περίπου 0.5.

Διεξάγοντας την εκπαίδευση για τις τέσσερις περιπτώσεις μοντέλων λαμβάνουμε τα παρακάτω αποτελέσματα:

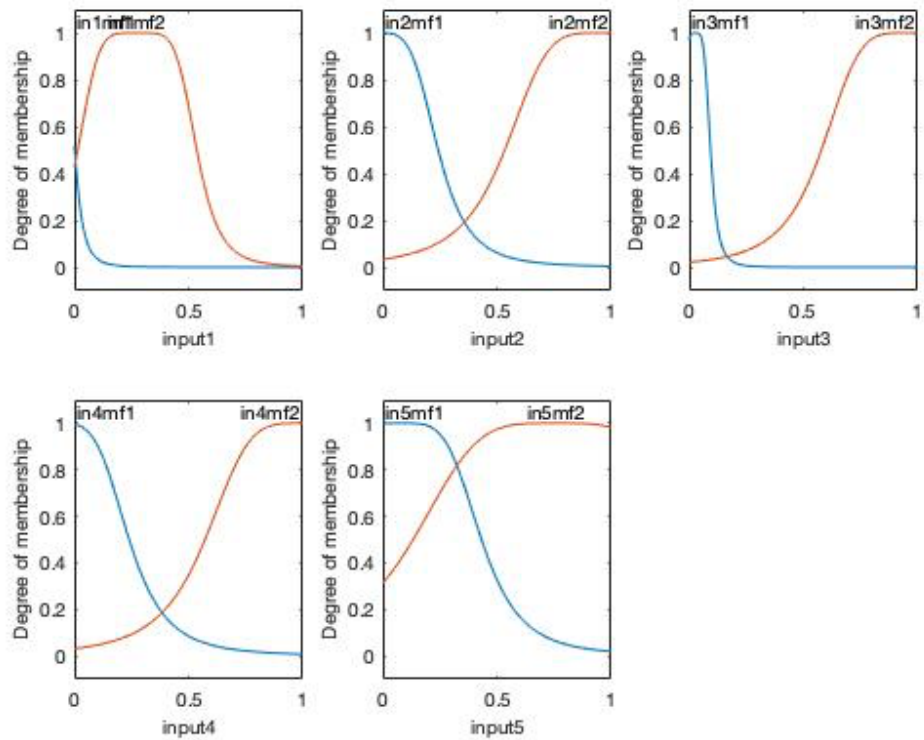
- i. Τα πρώτα αποτελέσματα που παρουσιάζουμε είναι τα διαγράμματα στα οποία απεικονίζονται οι τελικές μορφές των ασαφών συνόλων που προέκυψαν μέσω της διαδικασίας εκπαίδευσης:



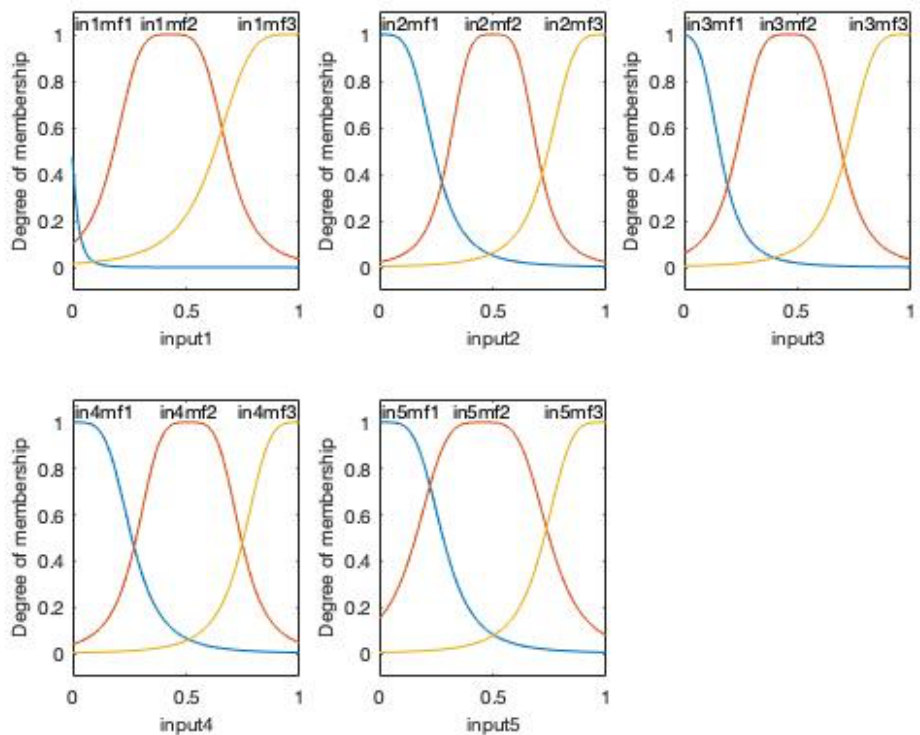
Σχήμα 1: Συναρτήσεις συμμετοχής μεταβλητών TSK μοντέλου 1



Σχήμα 2: Συναρτήσεις συμμετοχής μεταβλητών TSK μοντέλου 2

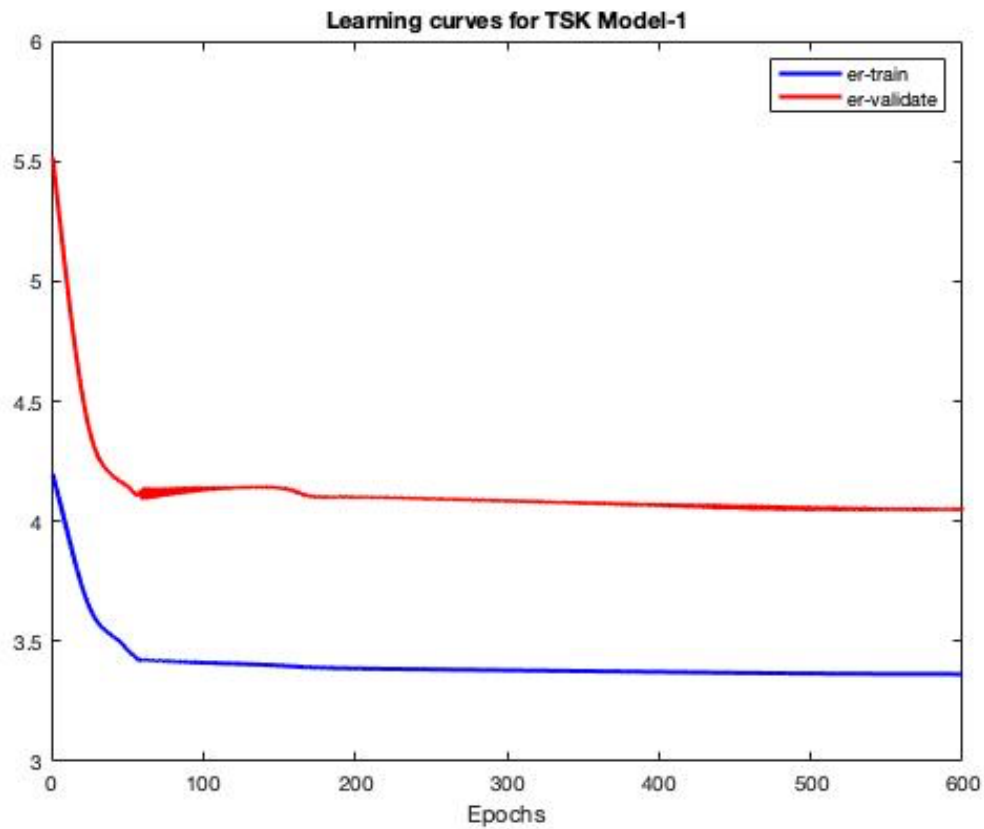


Σχήμα 3: Συναρτήσεις συμμετοχής μεταβλητών TSK μοντέλου 3

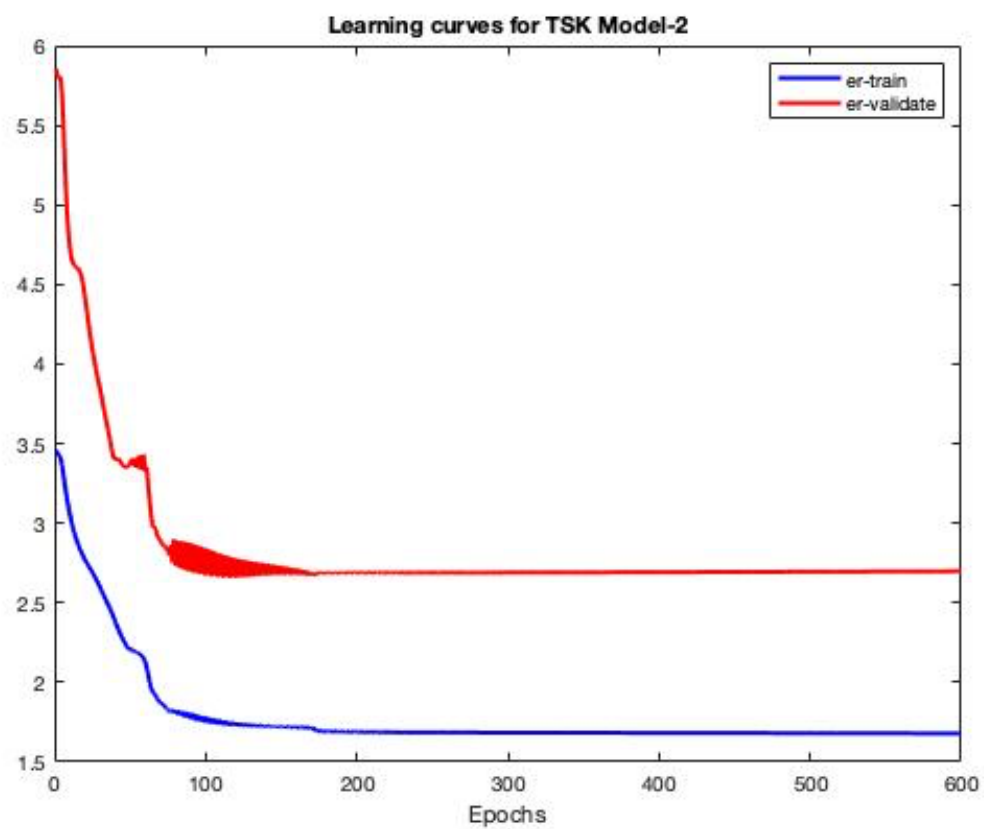


Σχήμα 4: Συναρτήσεις συμμετοχής μεταβλητών TSK μοντέλου 4

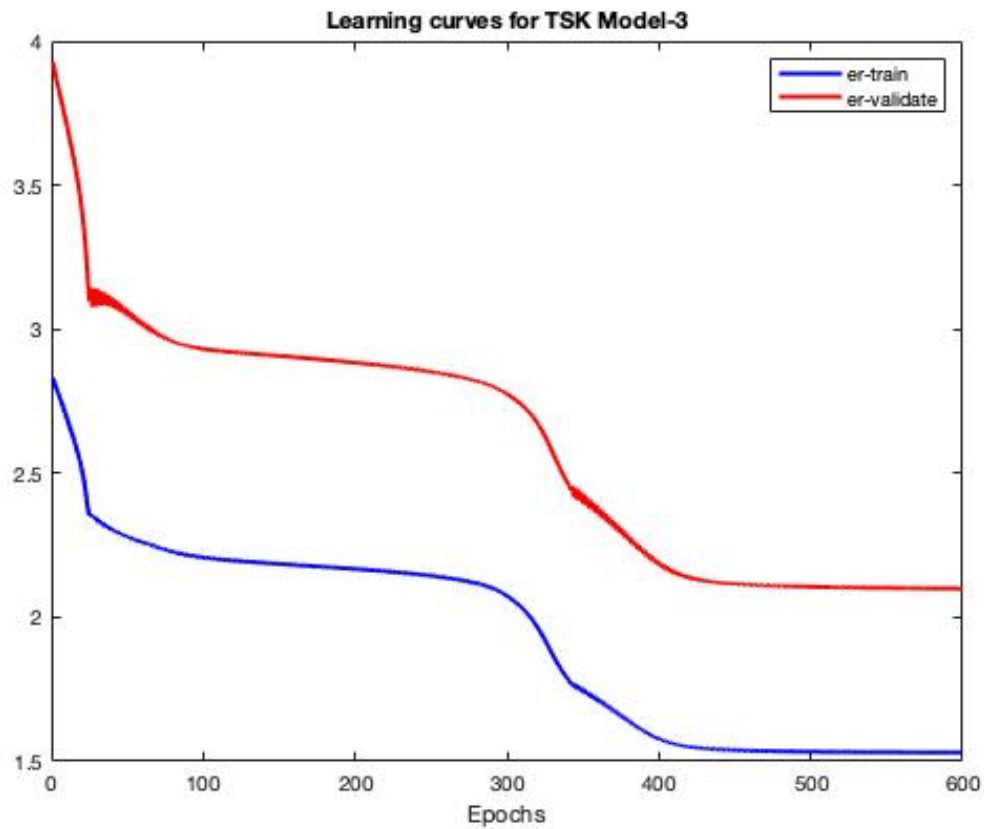
- ii. Παρακάτω δίνονται τα διαγράμματα μάθησης (learning curves) όπου απεικονίζεται το σφάλμα του μοντέλου συναρτήσει του αριθμού των επαναλήψεων(epochs)



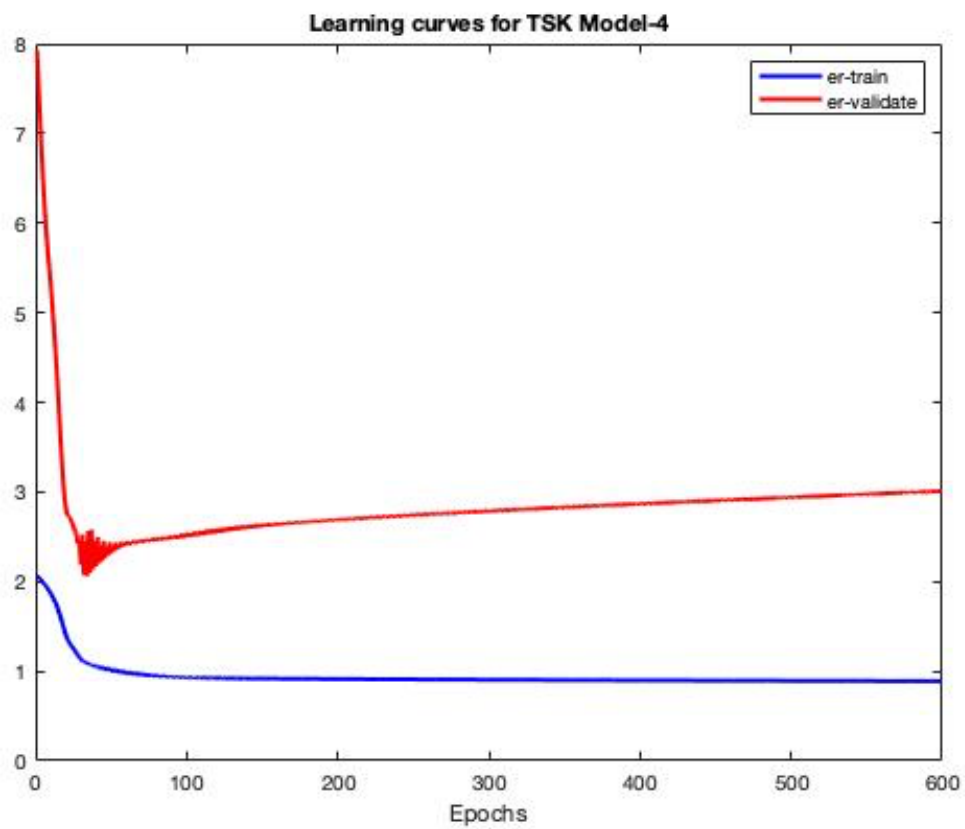
Σχήμα 5: Απεικόνιση RMSE εκπαίδευσης και επικύρωσης του TSK μοντέλου 1



Σχήμα 6: Απεικόνιση RMSE εκπαίδευσης και επικύρωσης του TSK μοντέλου 2

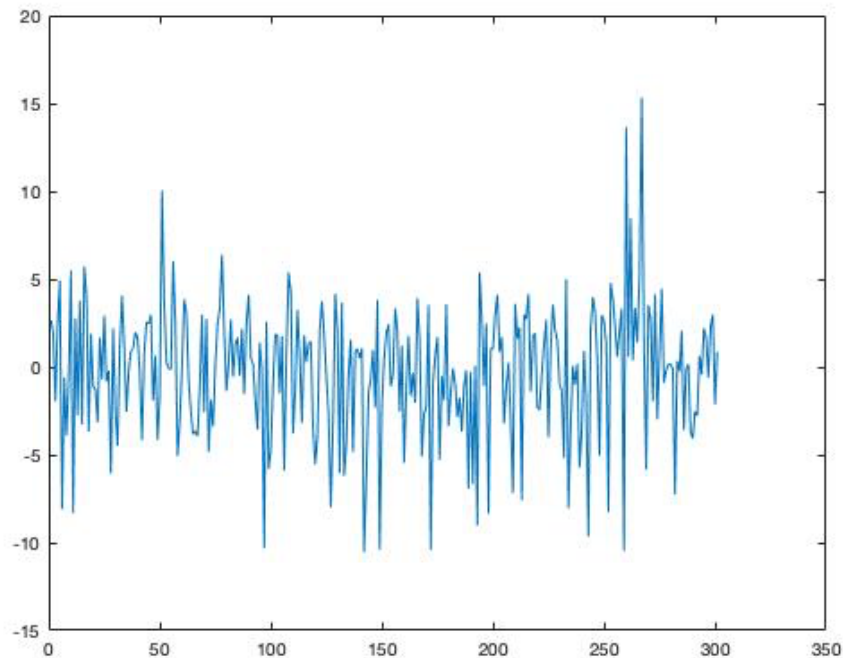


Σχήμα 7: Απεικόνιση RMSE εκπαίδευσης και επικύρωσης του TSK μοντέλου 3

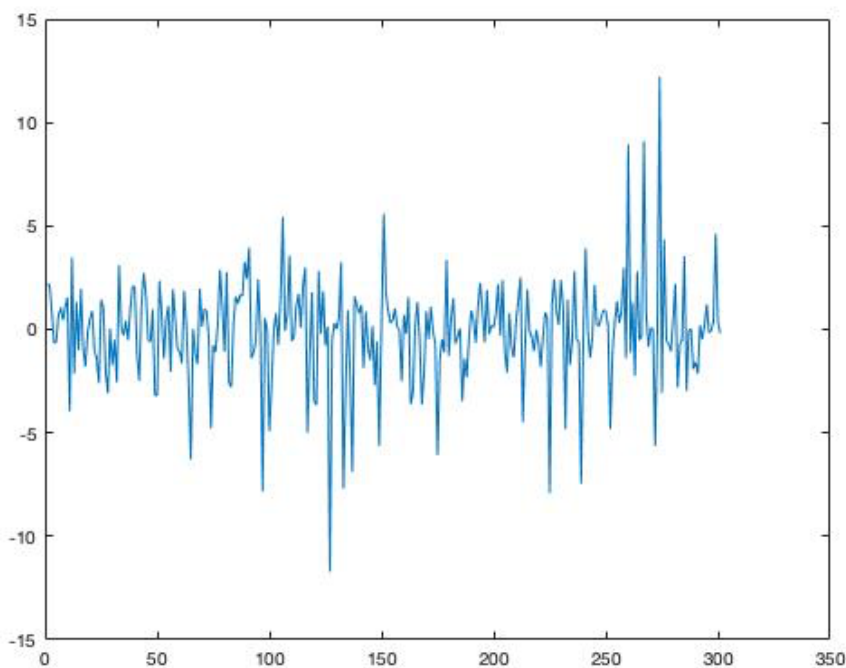


Σχήμα 8: Απεικόνιση RMSE εκπαίδευσης και επικύρωσης του TSK μοντέλου 4

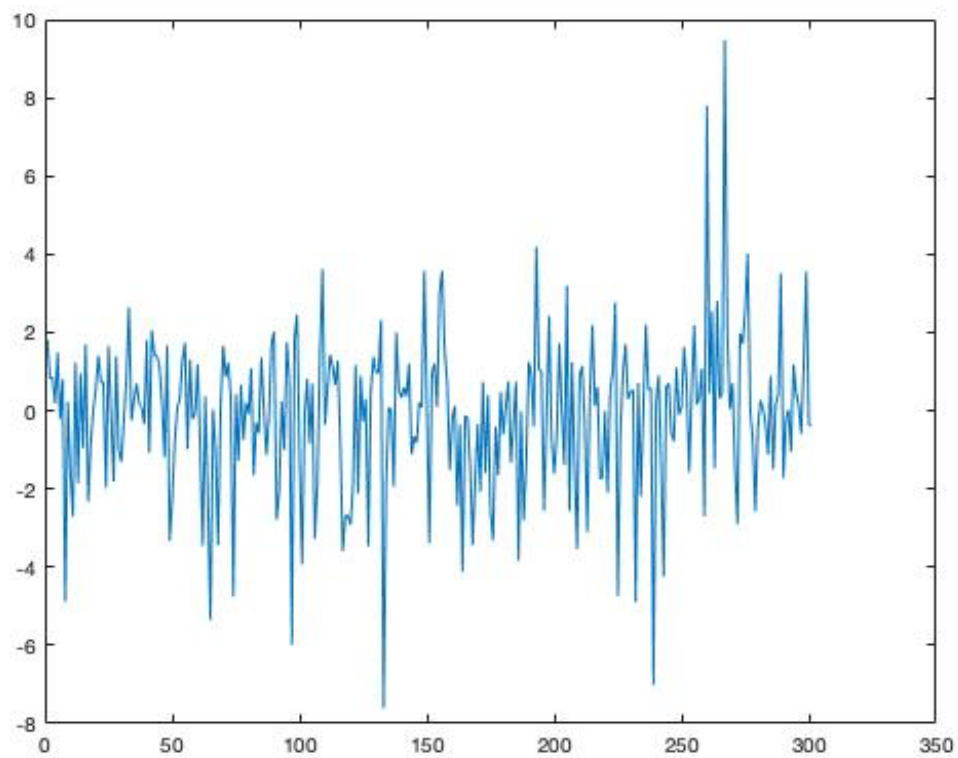
- iii. Παρακάτω δίνονται τα διαγράμματα όπου αποτυπώνονται τα σφάλματα πρόβλεψης. Για να δημιουργηθούν αυτά τα διαγράμματα αντιπαραβάλλουμε κάθε μοντέλο με το υποσύνολο ελέγχου και συγκρίνουμε την έξοδο του μοντέλου που εκπαιδεύσαμε, με την πραγματική έξοδο του υποσυνόλου ελέγχου.



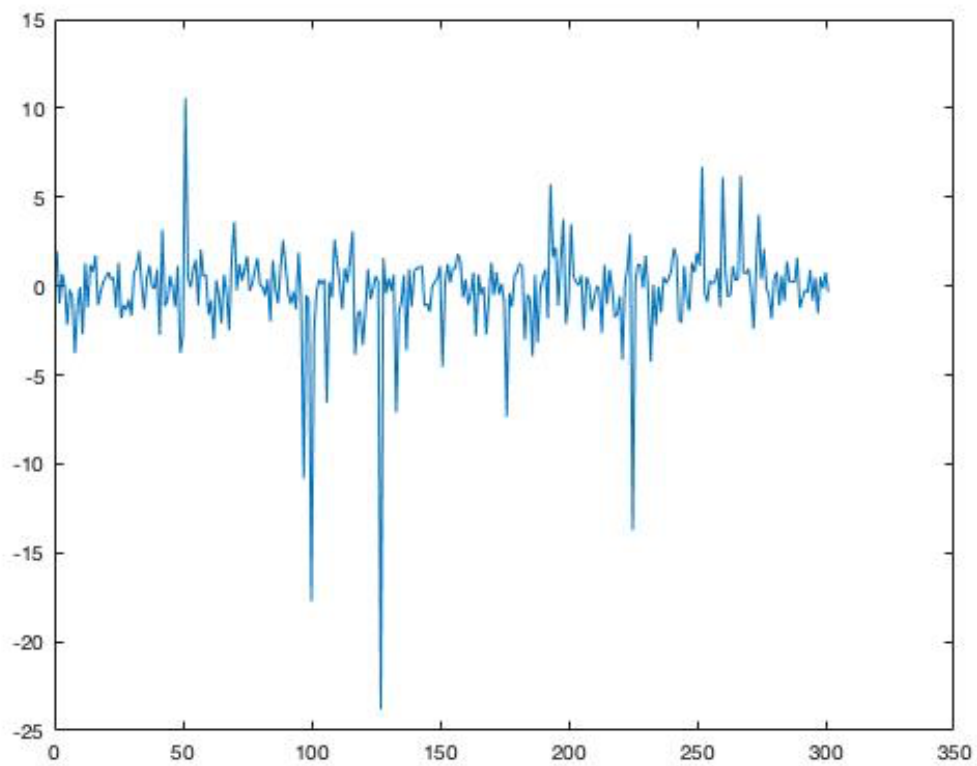
Σχήμα 9: Σφάλμα πρόβλεψης μοντέλου *TSK_Model_1*



Σχήμα 10: Σφάλμα πρόβλεψης μοντέλου *TSK_Model_2*



Σχήμα 11: Σφάλμα πρόβλεψης μοντέλου *TSK_Model_3*



Σχήμα 12: Σφάλμα πρόβλεψης μοντέλου *TSK_Model_4*

3) Σε αυτό το βήμα κάνουμε αξιολόγηση των μοντέλων, δηλαδή θα βρούμε με τι ακρίβεια εκτιμούν τη πραγματική συνάρτηση τα παραπάνω μοντέλα. Για τη διαδικασία αυτή θα χρησιμοποιηθούν οι εξής δείκτες απόδοσης:

- **RMSE**: είναι η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος μεταξύ της εξόδου του μοντέλου και της πραγματικής εξόδου (MSE).
- **R²**: Αυτός συντελεστής προσδιορισμού μας δίνει πληροφορία για το ποσοστό της διακύμανσης της πραγματικής τιμής το οποίο εξάγεται από το μοντέλο μας. Λαμβάνει τιμές στο διάστημα [0,1].
- **NMSE**: ο δείκτης αυτός ορίζεται ως ο λόγος μεταξύ της μεταβλητότητας των σφαλμάτων μοντελοποίησης και της μεταβλητότητας της πραγματικής κυματομορφής. Λαμβάνει τιμές στο διάστημα [0,1].
- **NDEI**: Αυτός ο δείκτης ορίζεται ως η τετραγωνική ρίζα του NMSE και λόγω της ρίζα παίρνει μεγαλύτερες τιμές, και πάλι όμως στο διάστημα [0,1].

Σημαντικό να αναφερθεί ότι επειδή οι δείκτες NMSE και NDEI απαλείφουν την μέση τιμή των υπό εξέταση πραγματικών συστημάτων, αποτελούν πιο ισχυρά και αξιόπιστα εργαλεία για την αξιολόγηση των μοντέλων που προκύπτουν μετά την εκπαίδευση. Όσο πιο μικρή τιμή λαμβάνουν τόσο καλύτερη είναι η προσέγγιση του μοντέλου για τη πραγματική τιμή της εξόδου.

Model	RMSE	R ²	NMSE	NDEI
Singleton 2	3.7208	0.7239	0.2761	0.5255
Singleton 3	2.4621	0.8791	0.1209	0.3477
Polynomial 2	1.9661	0.9229	0.0771	0.2777
Polynomial 3	2.6992	0.8547	0.1453	0.3812

Σχήμα 13: Δείκτες Αξιολόγησης της Απόδοσης των Μοντέλων

Παρατηρήσεις-Σχολιασμός Αποτελεσμάτων

Ξεκινώντας, παρατηρούμε ότι όπως είναι γνωστό από τη θεωρία τα σφάλματα εκπαίδευσης για τα μοντέλα που χρησιμοποίησαν πολυωνυμικές συναρτήσεις συμμετοχής είναι χαμηλότερα από ότι αυτά που χρησιμοποίησαν singleton, όπως φαίνεται αν κοιτάξουμε τον οριζόντιο άξονα στα διαγράμματα των καμπύλων εκμάθησης. Ακόμα, τα μοντέλα που έχουν 3 συναρτήσεις συμμετοχής ανά μεταβλητή όπως είναι φυσικό έχουν μικρότερο σφάλμα εκπαίδευσης σε σχέση με αυτά που χρησιμοποιούν 2 συναρτήσεις συμμετοχής

ανά μεταβλητή όπως φαίνεται αν κοιτάξουμε τον οριζόντιο άξονα στα διαγράμματα των καμπύλων εκμάθησης. Ακόμα, παρατηρούμε ότι το 4^ο μοντέλο που είναι με τρεις πολυωνυμικές συναρτήσεις συμμετοχής ανά είσοδο, παρουσιάζει το φαινόμενο της υπερεκπαίδευσης που φαίνεται και στην αρχή του διαγράμματος. Το σφάλμα επικύρωσης αντί να μειώνεται αυξάνει, και αν και παρουσιάζει μείωση μεταγενέστερα από κάποιες εποχές καταλήγει σε μία γραμμή με ελαφρώς ανοδική πορεία. Χάνει λοιπόν την δυνατότητα γενίκευσης γιατί έχει υπερεκπαιδευτεί.

Επιπλέον, παρατηρώντας τον πίνακα με τους δείκτες αξιολόγησης της απόδοσης των μοντέλων βλέπουμε ότι τη χειρότερη απόδοση παρουσιάζει το singleton 2 και τη καλύτερη το μοντέλο polynomial 2. Όλα τα μοντέλα εκτός του singleton 2 έχουν μικρές αποκλίσεις ως προς τους δείκτες, δηλαδή ακολουθούν παρόμοια συμπεριφορά. Από τα μοντέλα με 3 συναρτήσεις συμμετοχής την καλύτερη απόδοση έχει το singleton 3 με μικρή διαφορά βέβαια. Τέλος, βλέπουμε ότι οι τιμές των λαθών κινούνται από 1.9661 μέχρι 3.7208 και γνωρίζουμε ότι η έξοδος είναι της τάξης 10^2 (έχει τιμή 120-130) και άρα το λάθος είναι της τάξης περίπου 3% άρα είναι απολύτως αποδεκτό. Από τους δείκτες NMSE και NDEI επιβεβαιώνονται τα συμπεράσματα αυτά.

Εφαρμογή σε dataset με Υψηλή Διαστασιμότητα

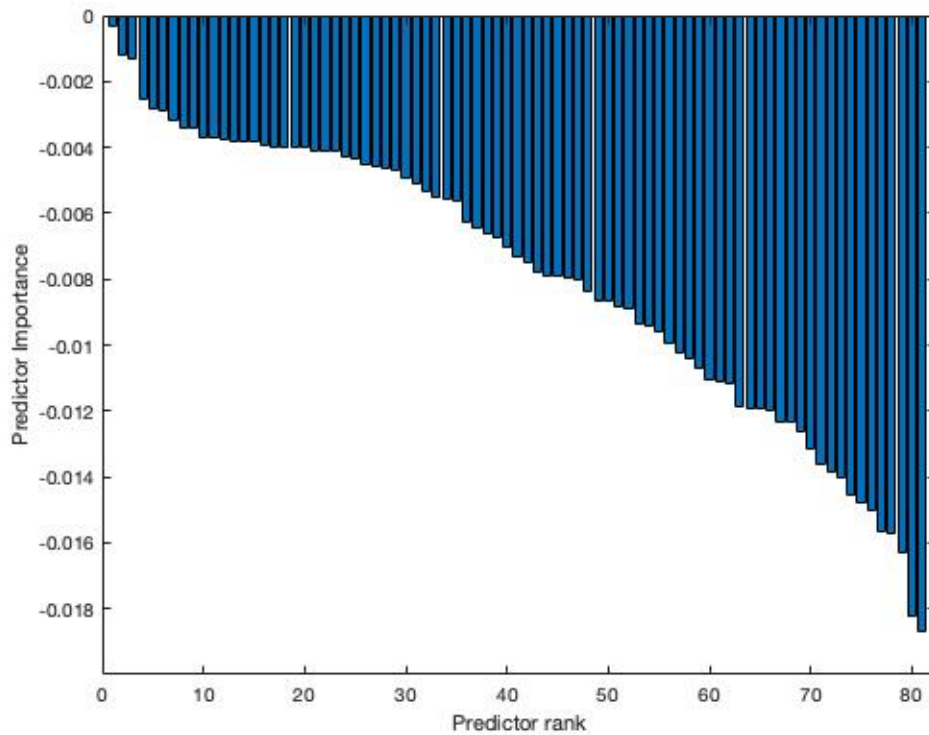
Στη δεύτερη φάση της εργασίας ακολουθούμε μια πιο συστηματική προσέγγιση στο πρόβλημα της μοντελοποίησης μιας άγνωστης συνάρτησης. Για το σκοπό αυτό έχει επιλεγθεί το Superconductivity dataset από το UCI repository, το οποίο έχει υψηλότερο βαθμό διαστασιμότητας. Πιο συγκεκριμένα, περιλαμβάνει 21263 δείγματα καθένα από τα οποία περιγράφεται από 81 μεταβλητές/χαρακτηριστικά. Ένα προφανές πρόβλημα που ανακύπτει είναι η λεγόμενη «έκρηξη» του πλήθους των IF-THEN κανόνων. Από τη θεωρία είναι γνωστό ότι για τη κλασσική περίπτωση του grid partitioning του χώρου εισόδου, ο αριθμός των κανόνων αυξάνεται εκθετικά σε σχέση με το πλήθος των εισόδων, γεγονός που καθιστά πολύ δύσκολη την μοντελοποίηση μέσω ενός TSK μοντέλου ακόμα και για datasets μεσαίας κλίμακας. Είναι αναγκαία λοιπόν η χρήση μεθόδων μείωσης της διαστασιμότητας καθώς και του αριθμού των IF-THEN κανόνων. Ο στόχος αυτός θα επιτευχθεί μέσω της επιλογής χαρακτηριστικών προς επιλογή και της χρήσης διαμέρισης διασκορπισμού. Προφανώς σαν τίμημα

της μείωσης της πολυπλοκότητας, οι μέθοδοι αυτές εισάγουν δύο ελεύθερες παραμέτρους:

- Αριθμός χαρακτηριστικών: Το πλήθος των χαρακτηριστικών που θα χρησιμοποιηθούν στην εκπαίδευση των μοντέλων.
- Ακτίνα των clusters r_α : Η παράμετρος που καθορίζει την ακτίνα επιρροής των clusters και κατ' επέκταση το πλήθος των κανόνων που θα προκύψουν.

Η αναλυτική ,λοιπόν, διαδικασία μοντελοποίησης του προβλήματος ακολουθεί τα παρακάτω βήματα που περιγράφονται :

1. Το σύνολο δεδομένων διαχωρίζεται σε τρία υποσύνολα. Ένα υποσύνολο εκπαίδευσης που περιέχει το 60% των δεδομένων , ένα υποσύνολο επικύρωσης που περιέχει το 20% των δεδομένων και ένα σύνολο ελέγχου που περιέχει το 20% των δεδομένων επίσης. Εφαρμόζουμε κανονικοποίηση στα δεδομένα εισόδου ώστε να βρίσκονται στο διάστημα $[0,1]$, χρησιμοποιώντας κατάλληλη συνάρτηση στο MATLAB.
2. Προκειμένου να διευκολύνουμε την επιλογή των χαρακτηριστικών εφαρμόζουμε τον αλγόριθμο Relieff. Πιο συγκεκριμένα με τον αλγόριθμο αυτό ταξινομούμε τα χαρακτηριστικά σε φθίνουσα σειρά σημαντικότητας. Δηλαδή, το πρώτο είναι το πιο σημαντικό και το τελευταίο το λιγότερο σημαντικό. Αυτό το βήμα είναι μεγάλης σημασίας διότι καθώς θέλουμε να μειώσουμε τη διαστασιμότητα του συνόλου δεδομένων και να το προσεγγίσουμε με ένα ισοδύναμο μικρότερο, πρέπει να επιλέξουμε τα πιο σημαντικά χαρακτηριστικά που συνεισφέρουν το μεγαλύτερο βαθμό στο καθορισμό της εξόδου. Στο σχήμα 14 βλέπουμε ένα ιστόγραμμα που παρουσιάζει το αποτέλεσμα του αλγορίθμου Relieff.



Σχήμα 14: Αποτέλεσμα αλγορίθμου *Relieff* για ταξινόμηση των χαρακτηριστικών ως προς τη σημαντικότητά τους.

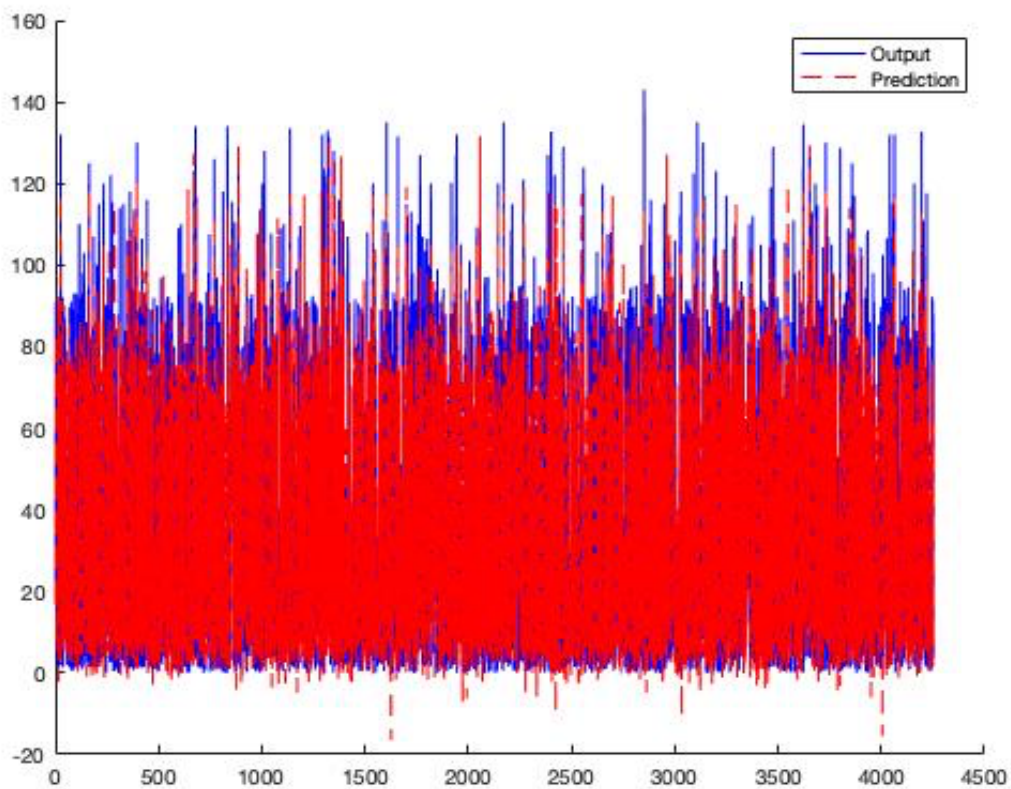
3. Στη συνέχεια προκειμένου να βρούμε τις βέλτιστες τιμές για τις δύο ελεύθερες παραμέτρους του προβλήματος μας εκτελούμε αναζήτηση πλέγματος(grid search) και αξιολόγηση μέσω 5-πτυχης διασταυρωμένης επικύρωσης (5-fold cross validation). Σημειώνεται ότι ως μέθοδος ομαδοποίησης για τη δημιουργία των IF-THEN κανόνων χρησιμοποιήθηκε ο αλγόριθμος Subtractive Clustering, και στην αναζήτηση πλέγματος με διασταυρωμένη επικύρωση τα δεδομένα εκπαίδευσης διαχωρίζονταν σε 80% για εκπαίδευση και 20% για επικύρωση. Τέλος, μετά από διεξαγωγή του grid search για διάφορες τιμές χαρακτηριστικών και τιμών ακτίνας παρουσιάζονται τα αποτελέσματα της αναζήτησης πλέγματος για $NF=[3, 6, 9, 12]$ και $ra=[0.5, 0.4, 0.3, 0.2]$. Το NF είναι τα δυνατά πλήθη χαρακτηριστικών και ra είναι οι δυνατές τιμές της ακτίνας των clusters. Οι τιμές αυτές επιλέχθηκαν διότι για μεγάλα πλήθη χαρακτηριστικών και πολύ μικρές τιμές ακτίνας, παρατηρήθηκαν τα εξής:
 - A) Απαγορευτικός χρόνος υπολογισμού του αποτελέσματος, που καθιστά μη πρακτική τη διαδικασία.
 - B) Για πολύ μικρές τιμές της ακτίνας (κάτω από 0.1) παρατηρήθηκε ότι η συνάρτηση *anfis* επέστρεφε πολλές τιμές σφαλμάτων NaN. Αυτό οδηγούσε όπως γίνεται αντιληπτό σε αδυναμία υπολογισμού της εξόδου μετέπειτα.

Στο σχήμα 15 παρουσιάζεται ο πίνακας που περιέχει τις τιμές του μέσου σφάλματος επικύρωσης για όλους τους συνδυασμούς χαρακτηριστικών και ακτινών. Έχει σημειωθεί ο συνδυασμός που δίνει το χαμηλότερο μέσο σφάλμα επικύρωσης και είναι για 12 χαρακτηριστικά και ακτίνα για τους cluster $ra=0.3$. Η τιμή του χαμηλότερου αυτού σφάλματος είναι 14.5849.

NF\ra	0.5	0.4	0.3	0.2
3	20.9912	20.6312	20.3040	19.4277
6	17.6692	17.2470	16.8265	15.9802
9	16.5432	16.2958	15.2937	15.2724
12	15.8268	15.0818	14.5849	14.7000

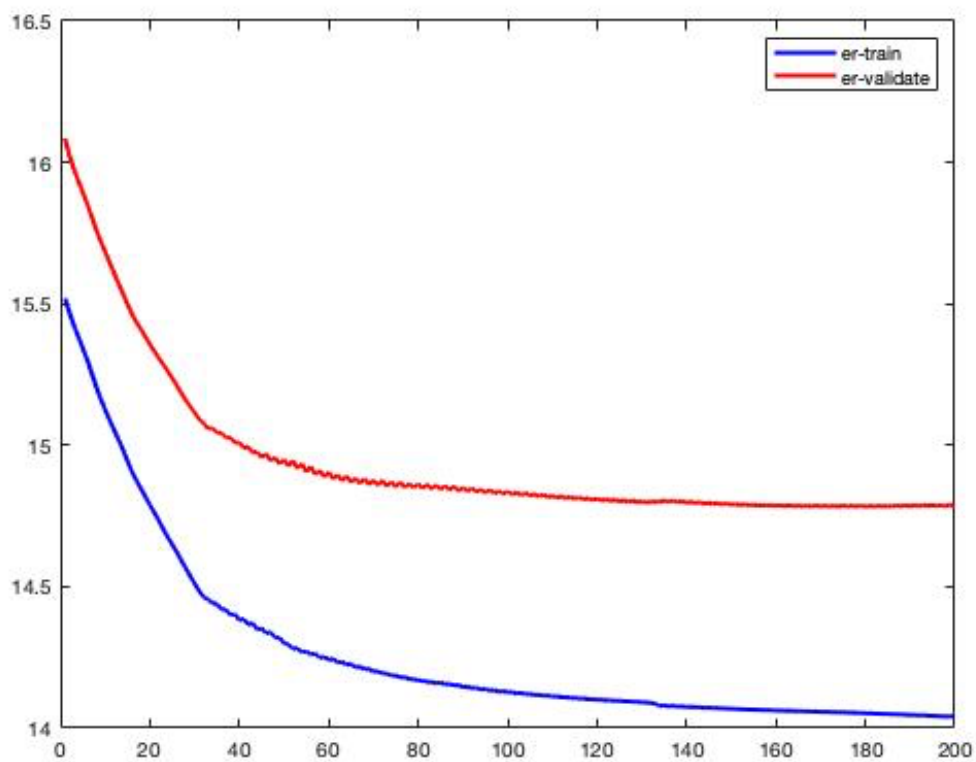
Σχήμα 15: Πίνακας με τα αποτελέσματα της Αναζήτησης Πλέγματος

4. Χρησιμοποιώντας τις βέλτιστες τιμές των ελεύθερων παραμέτρων, που βρήκαμε στο προηγούμενο βήμα θα εκπαιδεύσουμε με τις ίδιες προδιαγραφές όπως και προηγουμένως(SC) το τελικό TSK μοντέλο για το πρόβλημα μας. Σε αυτό το σημείο τονίζουμε ότι σαν σύνολο εκπαίδευσης χρησιμοποιούμε όλο το σετ εκπαίδευσης και αντίστοιχα σαν σύνολο επικύρωσης, όλο το σετ δειγμάτων επικύρωσης. Παρακάτω έχουν προστεθεί τα διάφορα διαγράμματα τα οποία ζητούνται σαν αποτελέσματα:
 - I. Διαγράμματα που αποτυπώνονται οι προβλέψεις του τελικού μοντέλου καθώς και οι πραγματικές τιμές της εξόδου. Για να σχεδιαστεί αυτό το διάγραμμα αντιπαραβάλλουμε το τελικό μας μοντέλο με το υποσύνολο ελέγχου.



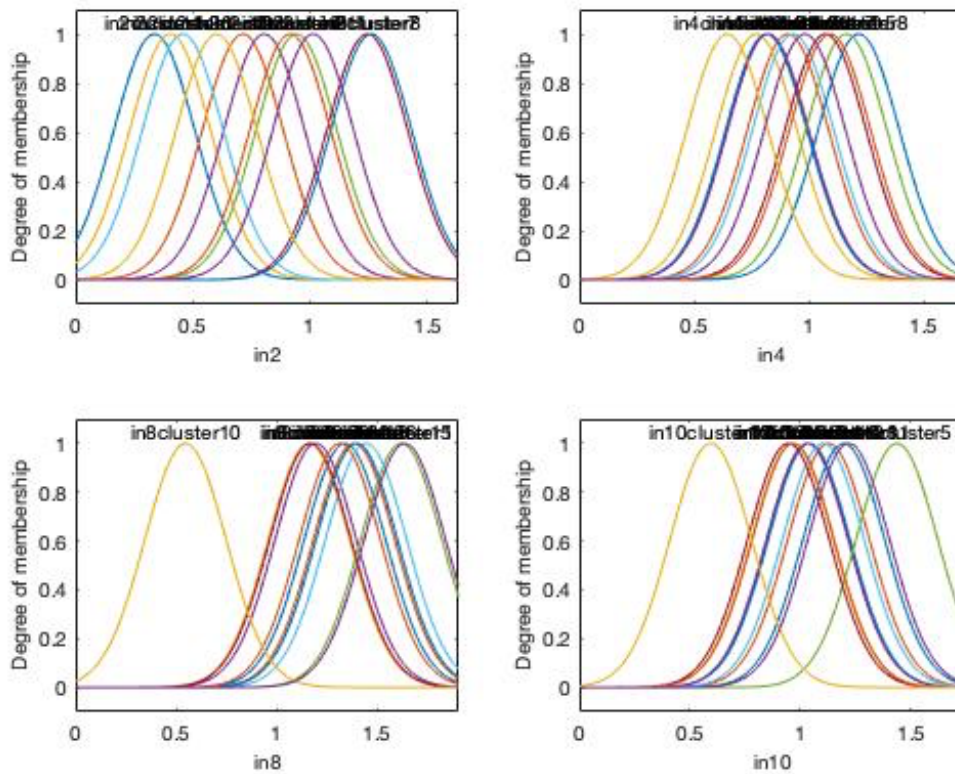
Σχήμα 16: Πρόβλεψη μοντέλου σε σχέση με τη πραγματική έξοδο

- II. Διάγραμμα εκμάθησης όπου απεικονίζεται το σφάλμα συναρτήσε του αριθμού επαναλήψεων:

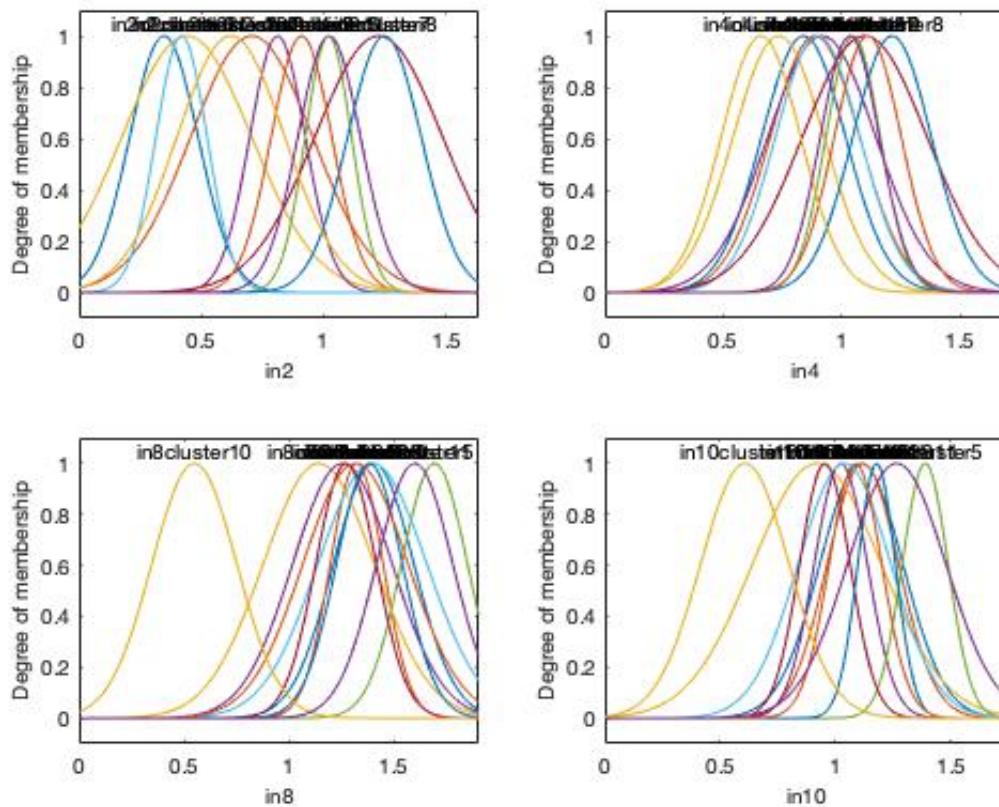


Σχήμα 17: Καμπύλη εκμάθησης τελικού TSK μοντέλου

III. Εδώ έχω επισυνάψει τις αρχικές και τελικές μορφές(μετά την εκπαίδευση) των συναρτήσεων συμμετοχής για τις μεταβλητές εισόδου 2,4,8 και 10.



Σχήμα 18: Συναρτήσεις συμμετοχής πριν το training



Σχήμα 19: Συναρτήσεις συμμετοχής μετά το training

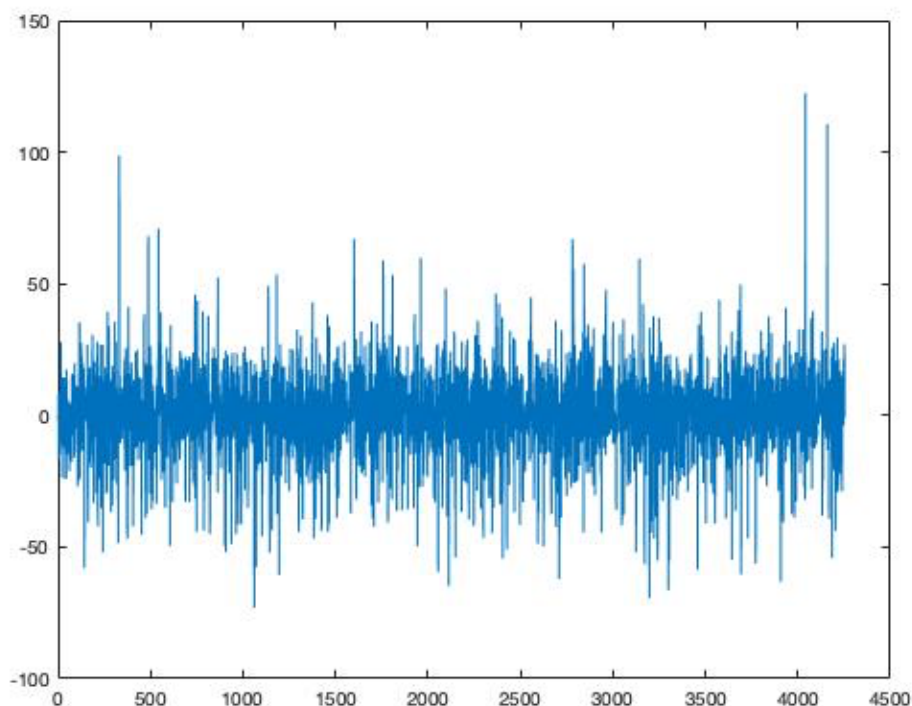
Βλέπουμε ότι στη τελική κατάσταση για τις εισόδους 2 και 4 υπάρχει μια συγκέντρωση των συναρτήσεων συμμετοχής προς το κέντρο. Αντίστοιχα για τις εισόδους 8 και 10 βλέπουμε ότι κάποιες συναρτήσεις έχουν μετατοπιστεί προς τα δεξιά. Έχουμε μεγαλύτερη επικάλυψη στο σημείο που συγκεντρώνονται.

IV. Πίνακας με τις τιμές των δεικτών απόδοσης για το τελικό μοντέλο μετά τη διαδικασία εκπαίδευσης:

Variable	Value
RMSE	14.6755
R^2	0.8209
NMSE	0.1790
NDEI	0.4231

Σχήμα 20: Δείκτες απόδοσης μοντέλου

Από το παραπάνω πίνακα μπορούμε να δούμε ότι το τελικό μας μοντέλο είναι αρκετά ακριβές στην πρόβλεψη της εξόδου. Πράγματι αυτό φαίνεται και στο διάγραμμα σφάλματος πρόβλεψης:



Σχήμα 21: Σφάλμα πρόβλεψης εξόδου

Εάν δούμε βέβαια τον αντίστοιχο πίνακα για τους δείκτες απόδοσης των μοντέλων του πρώτου μέρους των εργασιών, δημιουργείται η αίσθηση ότι ίσως δεν είναι τόσο ακριβές αυτό το μοντέλο σε σχέση με τα προηγούμενα. Ωστόσο, δεν πρέπει να ξεχνάμε ότι το τελικό μας μοντέλο σε αυτό το μέρος της εργασίας εκπαιδεύτηκε με μόλις 12 χαρακτηριστικά από τα 81 που περιγράφουν κάθε δείγμα. Αυτό σε συνδυασμό με το ότι η ακτίνα των cluster $r_a = 0.3$ μας δημιουργεί μόνο 11 ασαφείς κανόνες, μειώνει την ακρίβεια του μοντέλου μας σε ένα βαθμό.

Τέλος, συγκρίνουμε τον αντίστοιχο αριθμό κανόνων που θα είχαμε αν για το ίδιο πλήθος χαρακτηριστικών είχαμε επιλέξει grid partitioning με δύο ή τρία ασαφή σύνολα ανά είσοδο. Με δύο συναρτήσεις συμμετοχής ανά είσοδο θα είχαμε 4096 κανόνες για τα ίδια χαρακτηριστικά ενώ για τρεις συναρτήσεις συμμετοχής ανά είσοδο θα είχαμε 531441 κανόνες! Είναι κατανοητό ότι για τέτοιο πλήθος κανόνων, ο χρόνος της διαδικασίας εκπαίδευσης είναι απαγορευτικός. Εκτός αυτού είναι πολύ πιθανό ότι για τέτοιο πλήθος κανόνων θα είχαμε φαινόμενα υπερεκπαίδευσης με αποτέλεσμα το μοντέλο να χάσει τη δυνατότητα γενίκευσης. Δηλαδή, το μοντέλο χάνει την ικανότητα της επιτυχούς πρόβλεψης της εξόδου για άγνωστες εισόδους.