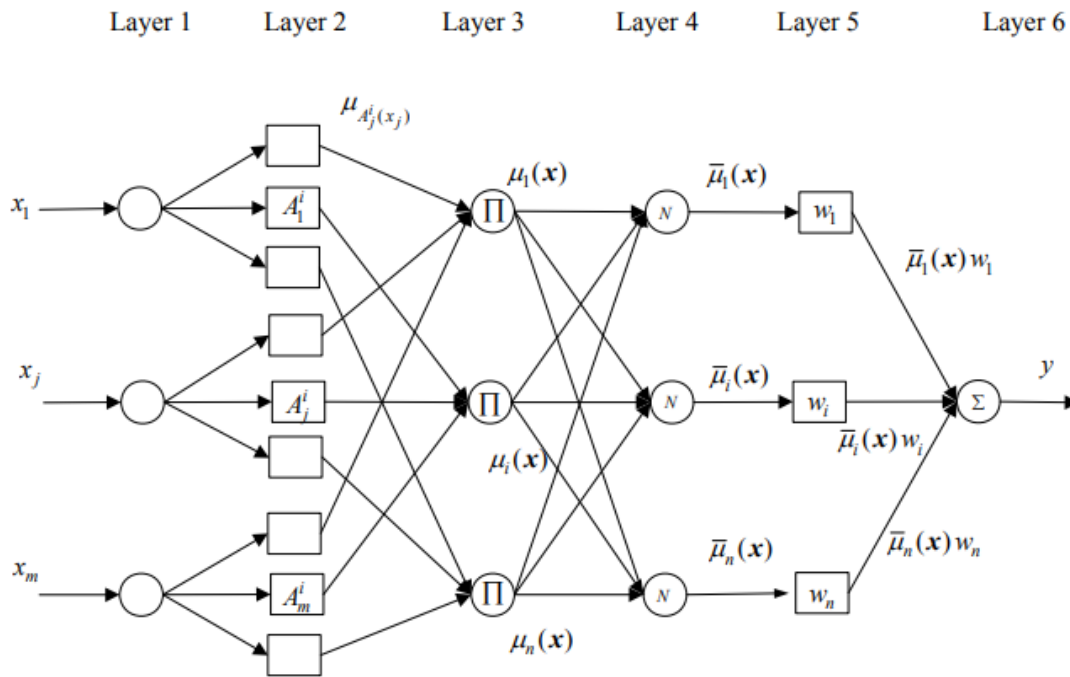


ΑΣΑΦΗ ΣΥΣΤΗΜΑΤΑ



Εργασία 3 σειρά 4

Επίλυση προβλήματος παλινδρόμησης με χρήση μοντέλων TSK

Παρασκευαΐδης Κωνσταντίνος

AEM: 7754

konstapf@auth.gr

ΕΙΣΑΓΩΓΗ

Στόχος της εργασίας αυτής είναι να διερευνηθεί 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 νευρωνικών μοντέλων. Το πρώτο σύνολο δεδομένων θα χρησιμοποιηθεί για μια 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 θα χρησιμοποιηθεί για μια πληρέστερη διαδικασία μοντελοποίησης.

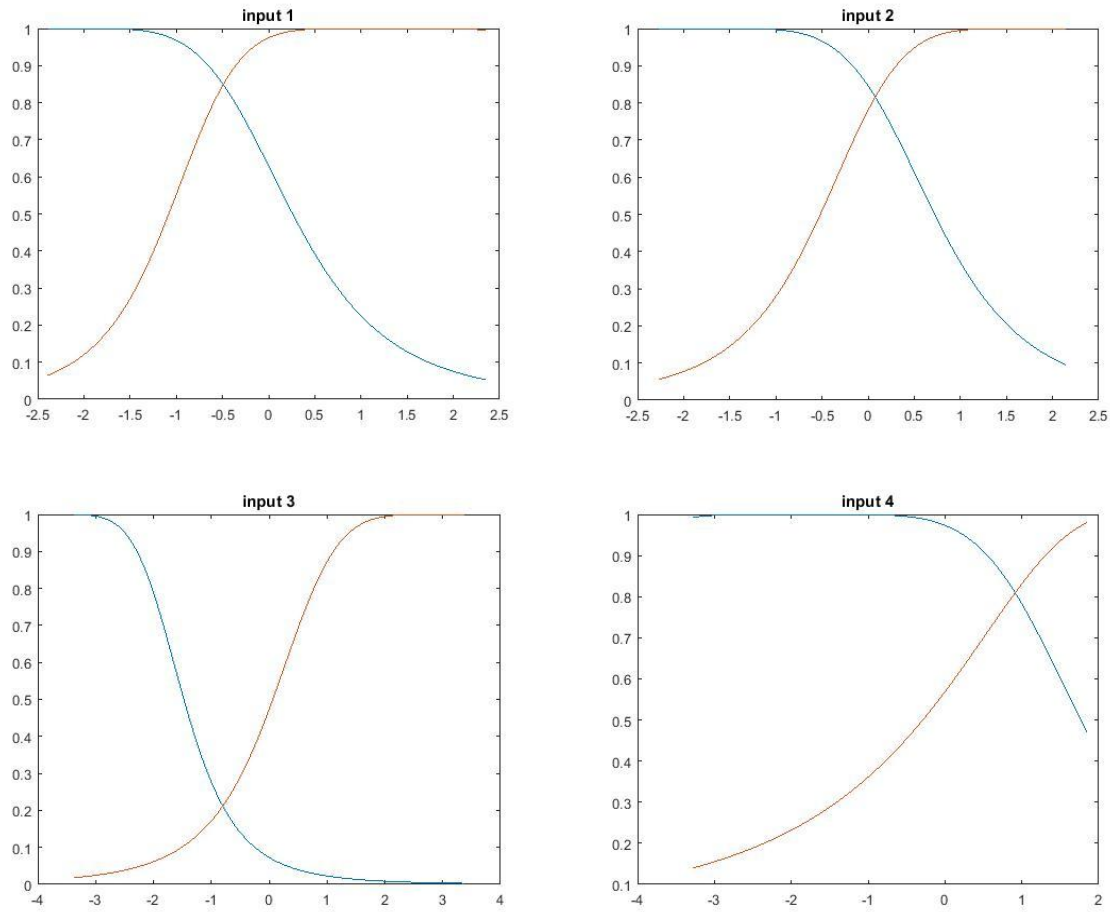
1 CCPP dataset

Αρχικά θα εξεταστούν διάφορα μοντέλα 4 TSK όσον αφορά την απόδοσή τους στο σύνολο ελέγχου, στα οποία θα μεταβάλλονται η μορφή της εξόδου καθώς και το πλήθος των συναρτήσεων συμμετοχής για κάθε μεταβλητή εισόδου. Και τα 4 μοντέλα θα εκπαιδευτούν με την υβριδική μέθοδο. Για την εκπαίδευση, επικύρωση, έλεγχο των μοντέλων χρησιμοποιήθηκε το CCPP dataset.

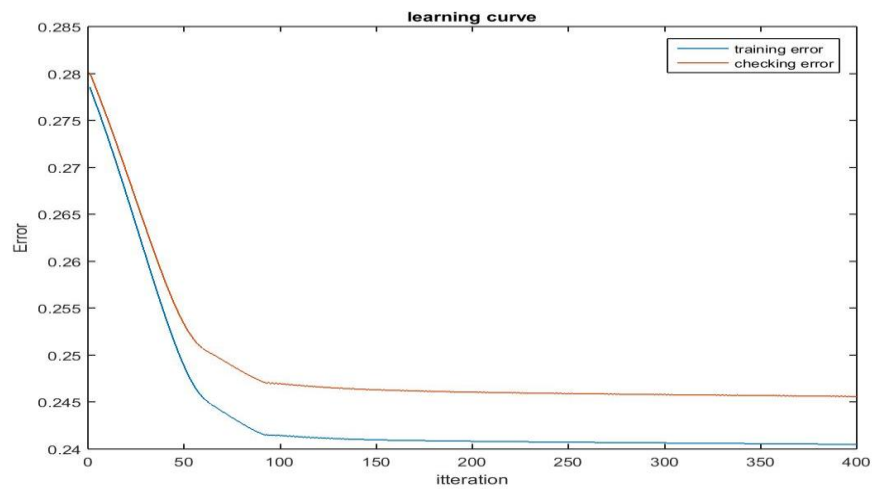
	Πλήθος συναρτήσεων συμμετοχής	Μορφή εξόδου
TSK_Model_1	2	Singleton
TSK_Model_2	3	Singleton
TSK_Model_3	2	Polynomial
TSK_Model_4	3	Polynomial

1.1 TSK_Model_1

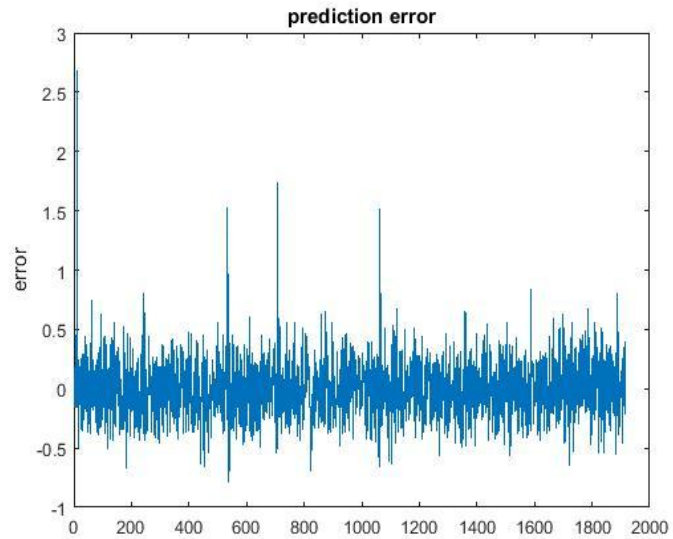
Τελικές μορφές ασαφών συνόλων



Learning curve



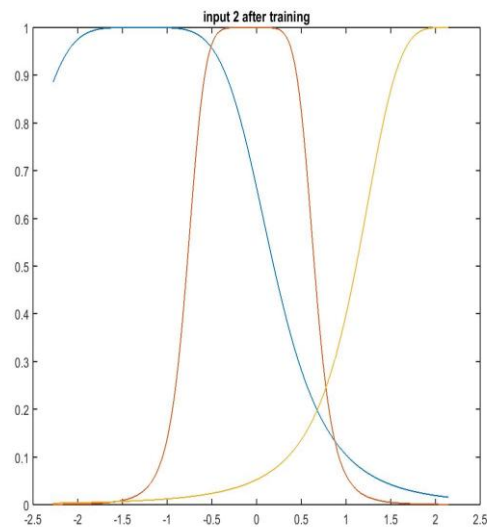
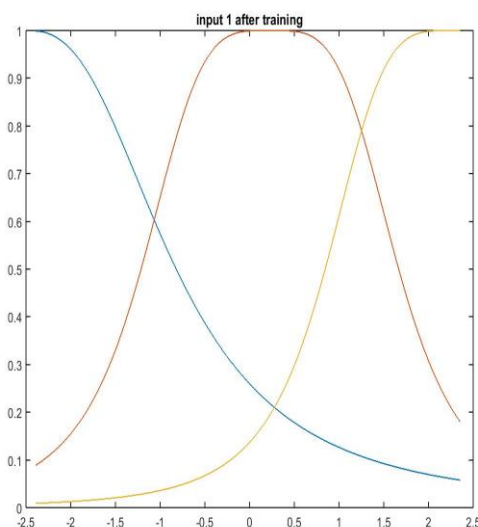
Από τις μετρικές παρατηρούμε ότι το πρώτο μοντέλο έχει μικρά σφάλματα εξόδου. Επομένως το μοντέλο είναι αξιόπιστο. Επιπλέον από το διάγραμμα εκμάθησης παρατηρούμε ότι χρειάζεται λιγότερες από 100 επαναλήψεις ώστε να συγκλίνει το σφάλμα. Από τις μετρικές αξιολόγησης για τα σετ εκπαίδευσης και επικύρωσης αλλά και από την το διάγραμμα εκμάθησης παρατηρούμε ότι δεν υπάρχει υπερεκπαίδευση.

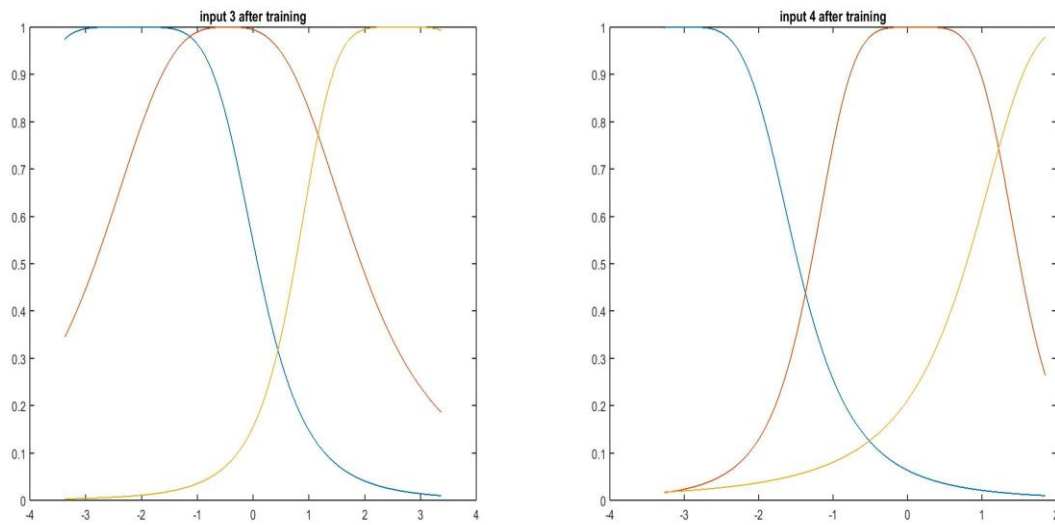


Data	RMSE	NMSE	NDEI	R ²
Training_set	0.1116	0.0042	0.0644	0.9958
Validation_set	0.1138	0.0130	0.1140	0.9870
Test_set	0.2512	0.0638	0.2527	0.9362

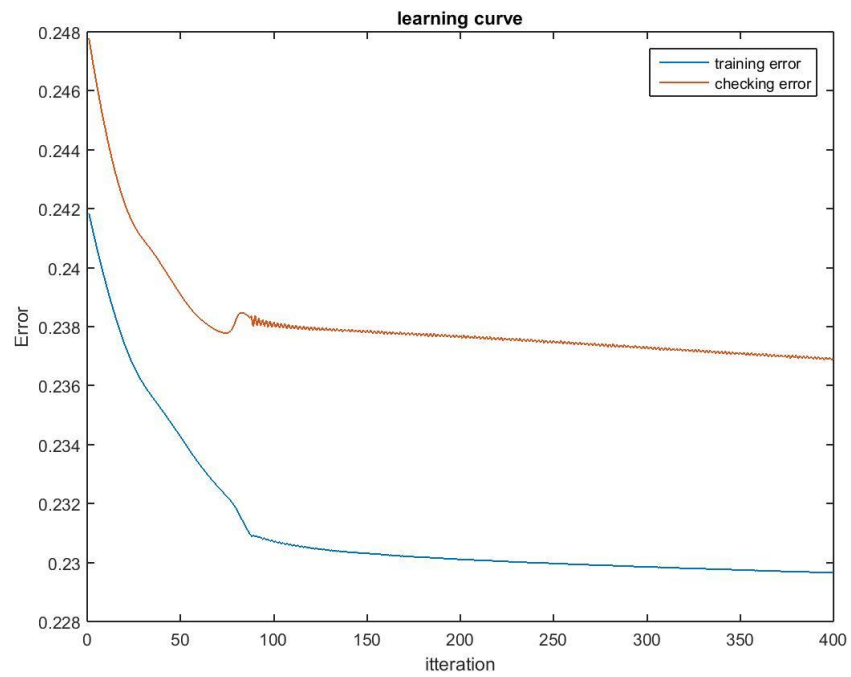
1.2 TSK_Model_2

Τελικές μορφές ασαφών συνόλων

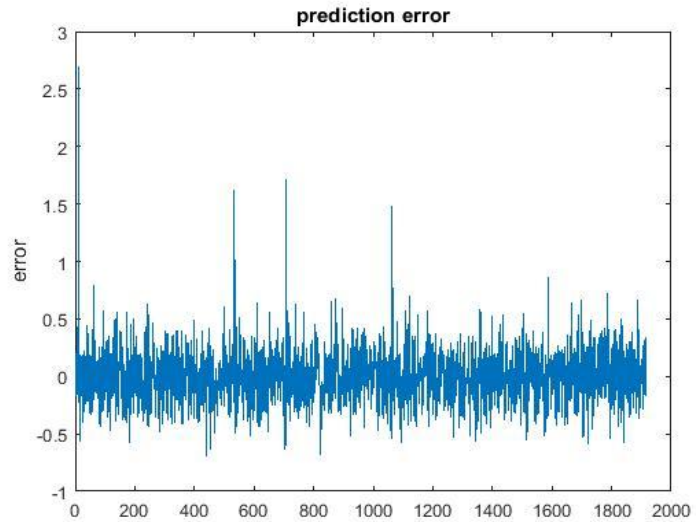




Learning curve



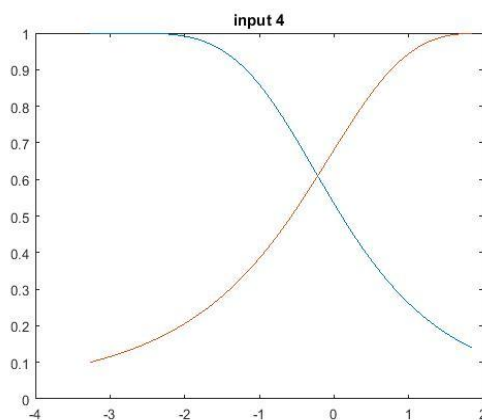
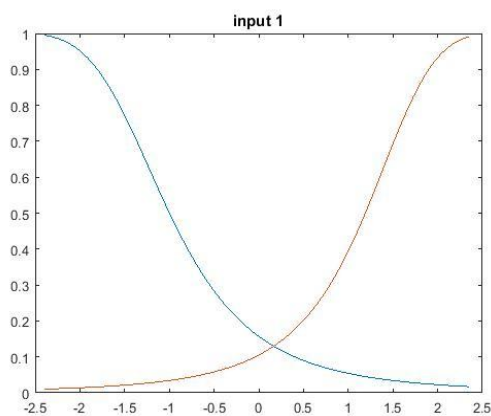
Με την προσθήκη μια συνάρτησης συμμετοχής για τις μεταβλητές εισόδου παρατηρούμε ότι το σφάλμα μειώνεται επομένως έχουμε καλύτερες προβλέψεις. Το σφάλμα συγκλίνει πριν τις 100 επαναλήψεις. Επίσης ούτε σε αυτήν την περίπτωση έχουμε υπερεκπαίδευση .

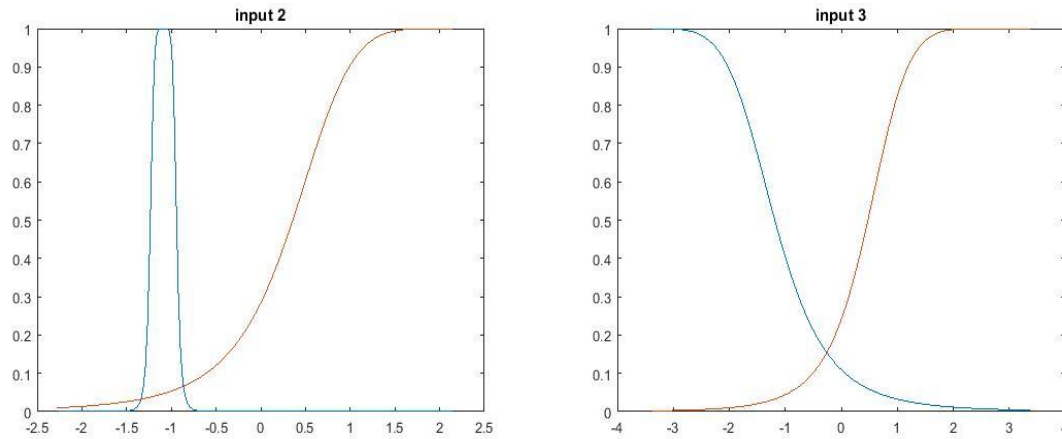


Data	RMSE	NMSE	NDEI	R ²
Training_set	0.1057	0.0037	0.0610	0.9963
Validation_set	0.1089	0.0119	0.1090	0.9881
Test_set	0.2412	0.0588	0.2426	0.9412

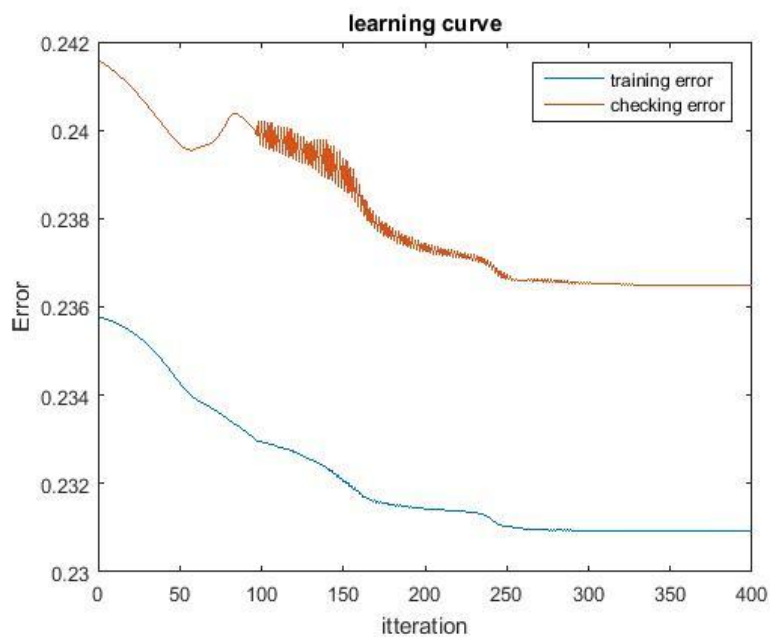
1.3 TSK_Model_3

Τελικές μορφές ασαφών συνόλων

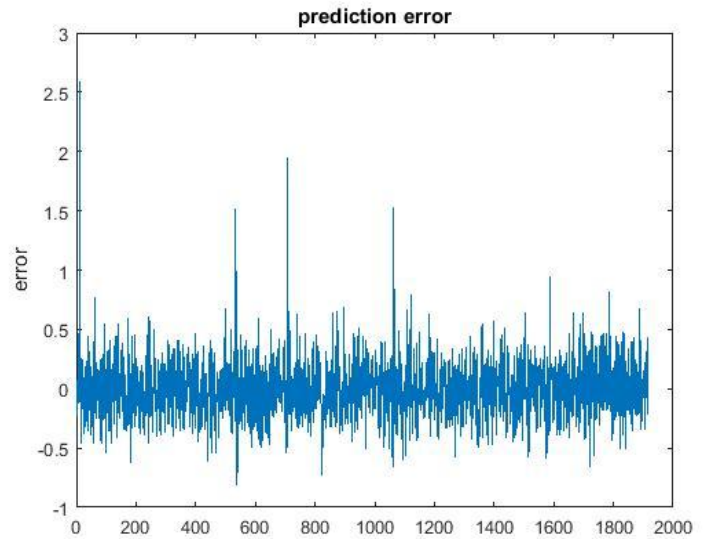




Learning curve



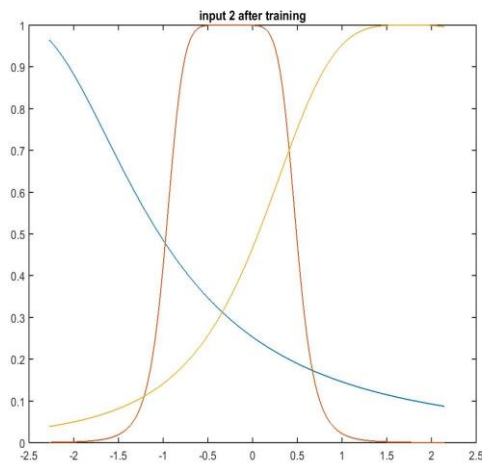
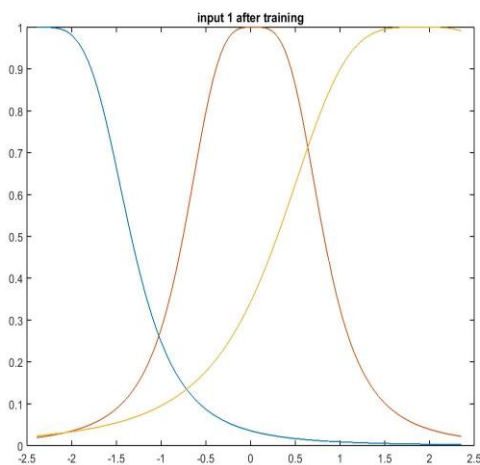
Και αυτά το μοντέλο έχει μικρά σφάλματα . Παρατηρούμε ότι οι μετρικές αξιολόγησης για το σετ έλεγχου είναι καλύτερες από το αντίστοιχο μοντέλο Singleton με 2 συναρτήσεις συμμετοχής . Το αποτέλεσμα αυτό είναι αναμενόμενο αφού τα TSK μοντέλα χρησιμοποιούν περισσότερες παραμέτρους στο τμήμα συμπεράσματος. Αυτό βέβαια αυξάνει την πολυπλοκότητα του μοντέλου και το κάνει πιο αργό Όπως φαίνεται και στο διάγραμμα εκμάθησης το σφάλμα θέλει πάνω από 200 επαναλήψεις ώστε να συγκλίνει . Επιπλέον παρατηρούμε ότι γύρο στις 100 επαναλήψεις το σφάλμα αξιολόγησης αυξάνεται όμως στη συνέχεια μειώνεται.

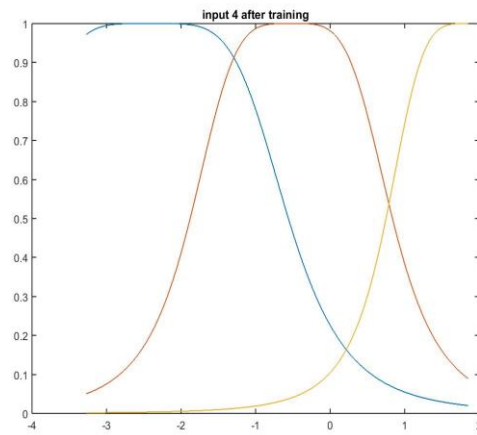
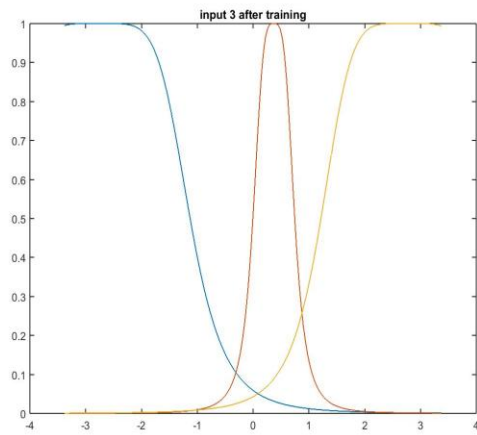


Data	RMSE	NMSE	NDEI	R ²
Training_set	0.1061	0.0038	0.0613	0.9962
Validation_set	0.1088	0.0119	0.109	0.9881
Test_set	0.245	0.0607	0.2464	0.9393

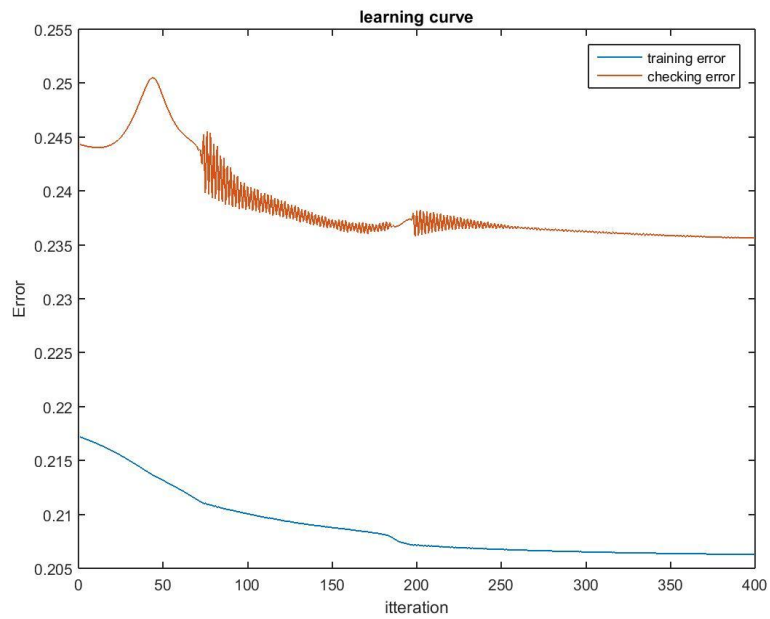
1.4 TSK_Model_4

Τελικές μορφές ασαφών συνόλων

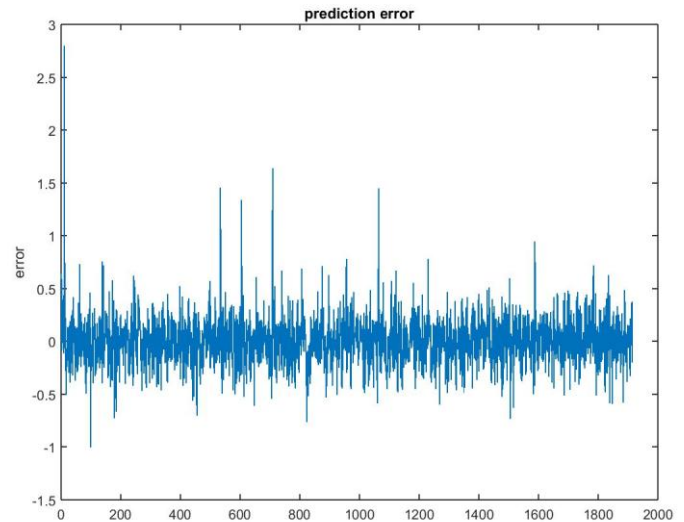




Learning curve



Με την προσθήκη μια συνάρτησης συμμετοχής για τις μεταβλητές εισόδου παρατηρούμε ότι το σφάλμα μειώνεται. Και αυτό το μοντέλο έχει λίγο μικρότερο σφάλμα από αντίστοιχο μοντέλο Singleton με 3 συναρτήσεις συμμετοχής. Από το διάγραμμα εκμάθησης παρατηρούμε ότι το σφάλμα συγκλίνει λίγο πιο γρήγορα σε σχέση με το προηγούμενο μοντέλο με τις 2 συναρτήσεις συμμετοχής αλλά και πάλι είναι πιο αργό από αντίστοιχο μοντέλο Singleton.



Data	RMSE	NMSE	NDEI	R ²
Training_set	0.083	0.0023	0.0479	0.9977
Validation_set	0.0949	0.009	0.095	0.9910
Test_set	0.2414	0.0589	0.2428	0.9411

Παρατηρήσεις

- 1) Το μοντέλο με το μικρότερο σφάλμα στο σετ αξιολόγησης είναι το TSK με 3 συναρτήσεις συμμετοχής. Γενικά τα TSK μοντέλα είχαν μικρότερα σφάλματα σε σχέση με τα αντίστοιχα Singleton όπως αναμενόταν , αλλά το σφάλμα αργεί περισσότερο να συγκλίνει .
- 2) Η προσθήκη μιας συνάρτησης συμμετοχής μείωνε το σφάλμα στα μοντέλα . Όπως παρατηρείτε από τα διαγράμματα εκμάθησης παρότι το σφάλμα αξιολόγησης σε ορισμένες περιπτώσεις αποκλίνει για λίγο και παρουσιάζει διακυμάνσεις στα μοντέλα με 3 συναρτήσεις μετά από ένα αριθμό επαναλήψεων μειώνεται . Επομένως το φαινόμενο της υπέρ-εκπαίδευσης δεν είναι σημαντικό σε αυτές τις περιπτώσεις .
- 3) Οι συναρτήσεις συμμετοχής να είναι bell shaped και κατά την αρχικοποίηση έχουν βαθμό επικάλυψης περίπου 0.5. Οι τελικές μορφές όμως των συναρτήσεων έχουν διαφορετικό βαθμό επικάλυψης κάτι που οφείλεται στη μέθοδο backpropagation που χρησιμοποιείτε για την βελτιστοποίηση των συναρτήσεων συμμετοχής.

2 Εφαρμογή σε dataset με υψηλή διαστασιμότητα

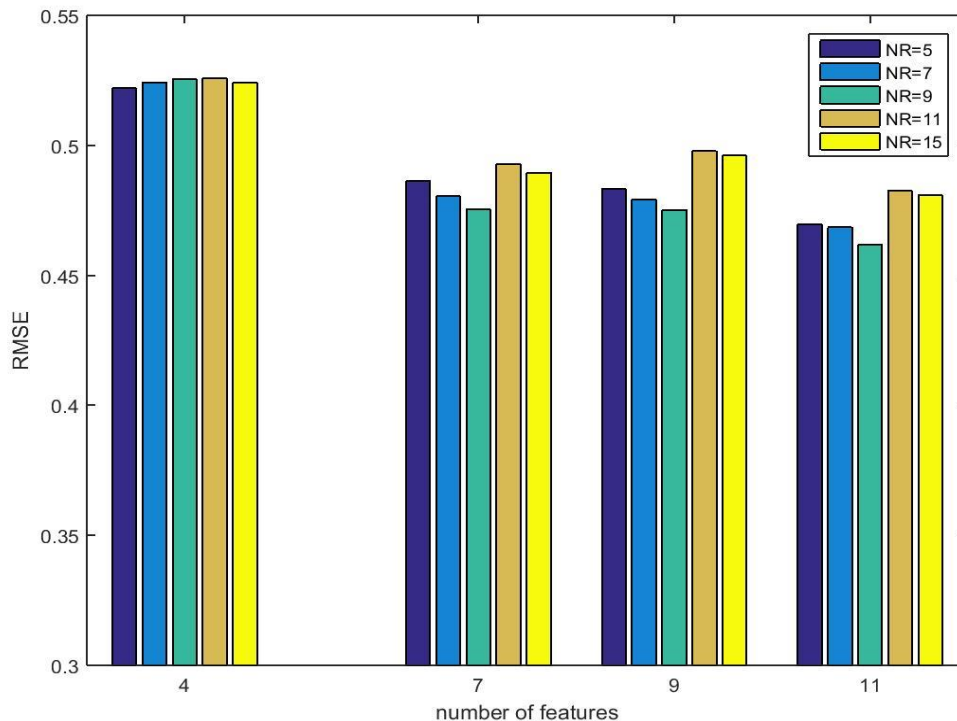
Στη συνέχεια έγινε μια προσέγγιση στο πρόβλημα της μοντελοποίησης μιας άγνωστης συνάρτησης σε dataset με υψηλή διαστασιμότητα. Λόγο της έκρηξης του πλήθους των IF-THEN κανόνων σε τέτοια dataset η μοντελοποίηση με χρήση TSK καθίστατε δύσκολη . Για αυτό τον λόγο πρέπει να μειώσουμε το πλήθος των κανόνων και το πλήθος των εισόδων .

2.1 Επιλογή παραμέτρων

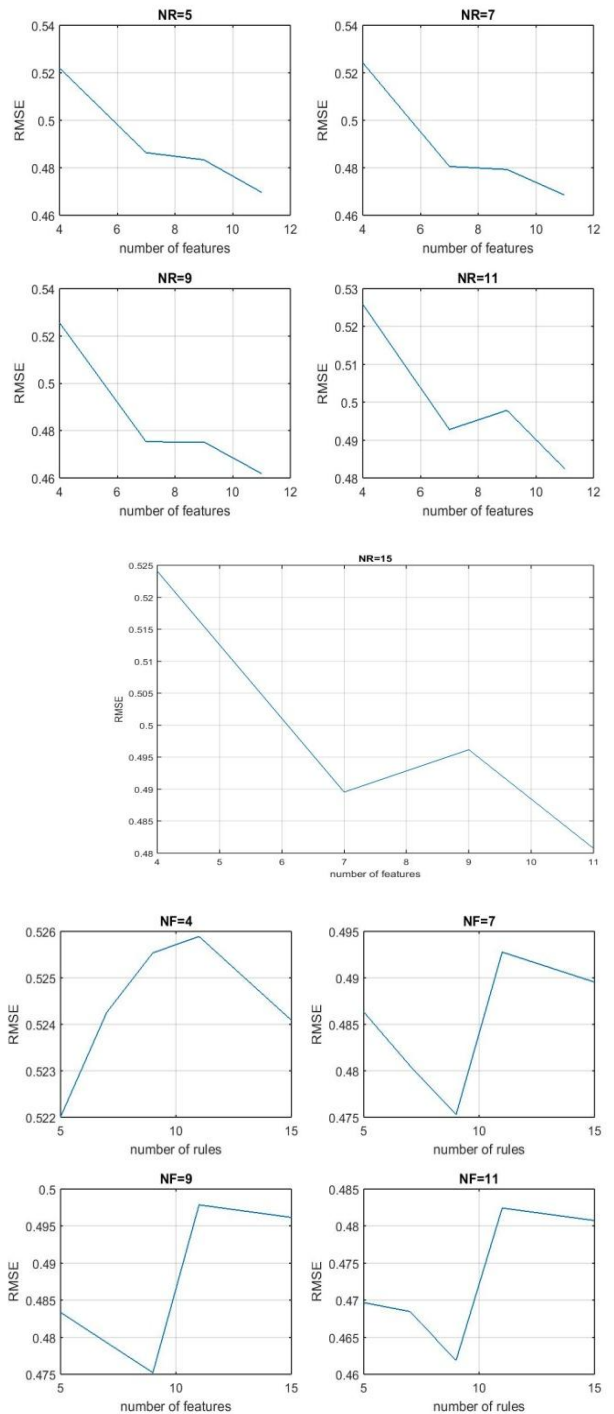
Αρχικά εκτελεστική αναζήτηση πλέγματος (grid search) και αξιολόγηση μέσω 5-πτυχης διασταυρωμένης επικύρωσης (5-fold cross validation) για την επιλογή των βέλτιστων τιμών των παραμέτρων. Ως μέθοδος ομαδοποίησης για τη δημιουργία των IF-THEN κανόνων επιλέγεται ο αλγόριθμος Fuzzy C-Means και ως μέθοδος επιλογής χαρακτηριστικών επιλέγεται ο αλγόριθμος Relief . η αναζήτηση έγινε για τιμές χαρακτηριστούν από το σύνολο $NF=\{4,7,8,11\}$ και για αριθμό κανόνων από το σύνολο $NR=\{5,7,9,11,15\}$. Οι τιμές που επιλέχτηκαν είναι αυτές που είχαν το μικρότερο μέσο σφάλμα , οι οποίες είναι :

- ❖ Αριθμός χαρακτηριστικών $NF=11$
- ❖ Αριθμός κανόνων $NR=9$

Στο επόμενο διάγραμμα φαίνεται το μέσο σφάλμα για όλους τους συνδυασμούς χαρακτηριστικών.

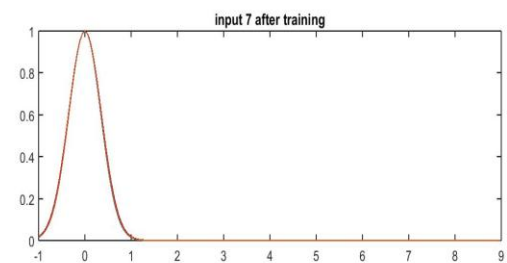
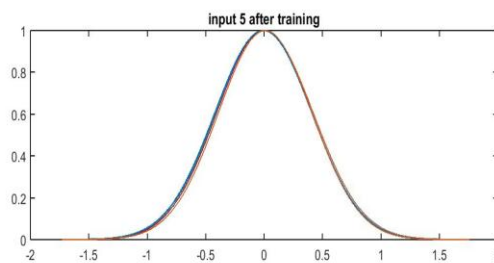
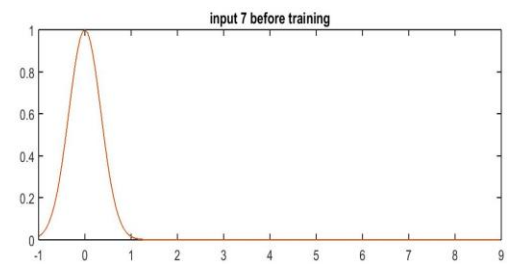
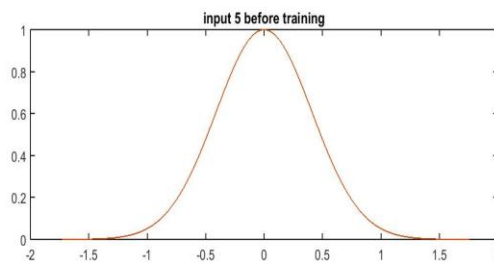
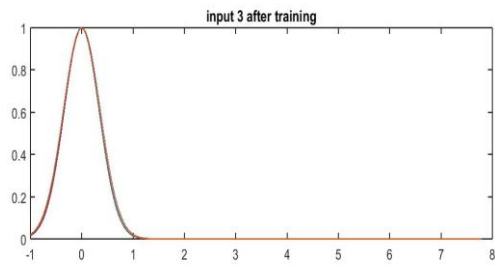
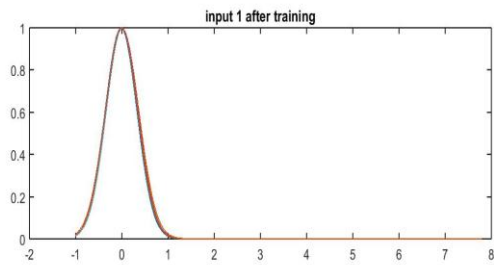
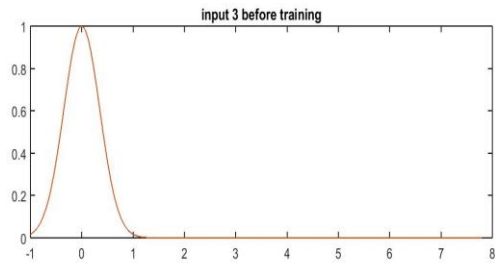
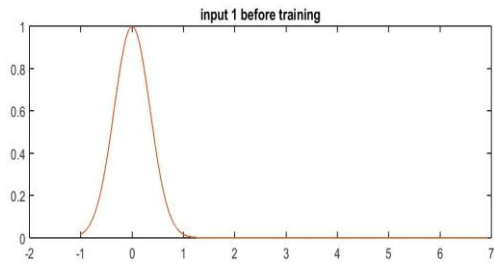


Για σταθερό αριθμό κανόνων παρατηρούμε από τα διπλανά διάγραμμα ότι το σφάλμα μειώνεται όσο αυξάνει ο αριθμός των χαρακτηριστικών . Ενώ παρατηρούμε ότι για σταθερό αριθμό χαρακτηριστικών το μικρότερο σφάλμα συνήθως αρχικά μειώνεται καθώς αυξάνονται ο αριθμός μέχρι τους 9 κανόνες και στην συνέχεια αυξάνεται . Επομένως όσο περισσότερα χαρακτηριστικά χρησιμοποιήσουμε το μοντέλο θα έχει καλύτερα αποτελέσματα το ίδιο όμως δεν ισχύει και για τον αριθμό των κανόνων . Αρκετά μικρός ή μεγάλος αριθμός κανόνων αυξάνει το σφάλμα του μοντέλου .

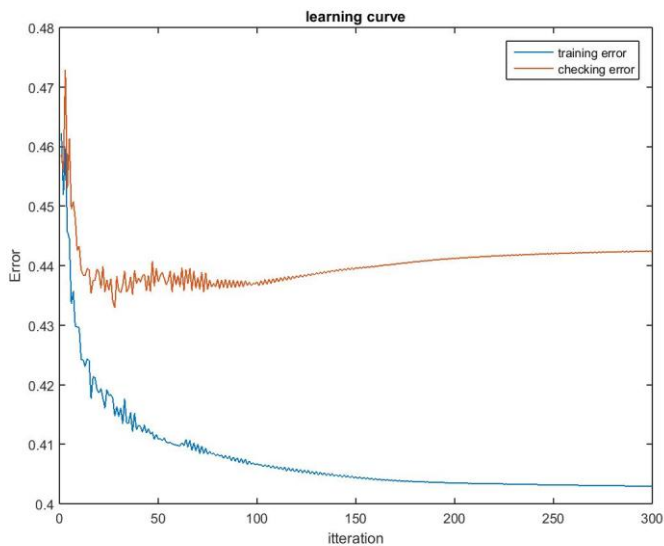


2.2 Εκπαίδευση μοντέλου

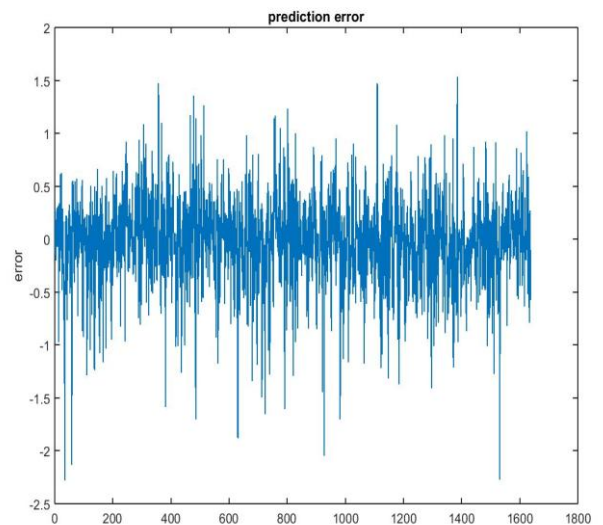
Τελικές μορφές ασαφών συνόλων



Learning curve



Prediction error



Μετρικές αξιολόγησης

Data	RMSE	NMSE	NDEI	R ²
Training_set	0.4065	0.116	0.0134	0.9866
Validation_set	0.4408	0.2158	0.0466	0.9534
Test_set	0.4414	0.4512	0.2036	0.7964

Οι μετρικές αξιολογήσεις δεν είναι τόσο καλές όσο στα προηγούμενα μοντέλα . Επίσης παρατηρείται μάγκο φαινόμενο υπέρ-εκπαίδευσης μια στο διάγραμμα εκμαθείσης μετά από της 100 επαναλήψεις το σφάλμα αξιολόγησης αποκλίνει ελαφρός . Όμως ο με χρήση grid partitioning ο αριθμός των κανόνων και τον χαρακτηρίστηκαν αυξάνεται εκθετικά, κάτι που οδηγεί στην αύξηση της πολυπλοκότητας του συστήματος και θα δυσκόλευε αρκετά την μοντελοποίηση και θα την καθιστούσε πολύ αργή .