

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρονικής και Υπολογιστών



Εργασία στο Μάθημα Ασαφή Συστήματα

Regression

Ομάδα 1

Εργασία 3

Όνομα: Στέφανος Γανωτάκης

AEM: 7664

Email: sganotak@auth.gr

ΙΟΥΝΙΟΣ 2019

Περιεχόμενα

1	Εισαγωγή	3
1.1	Αντικείμενο Εργασίας	3
1.2	Scripts Εργασίας	3
2	CCPP Dataset	4
2.1	TSK Model 1: 2 membership functions - Singleton output	4
2.2	TSK Model 2: 3 membership functions - Singleton output	5
2.3	TSK Model 3: 2 membership functions – Polynomial output	6
2.4	TSK Model 4: 3 membership functions – Polynomial output	7
2.5	Σχολιασμός μοντέλων και σύγκριση δεικτών	8
3	Superconductivity Dataset	9
3.1	Grid Search & 5-fold Cross validation	9
3.2	Τελικό TSK μοντέλο	10
3.2.1	Σύγκριση με μοντέλο 3 χαρακτηριστικών	12

1 Εισαγωγή

1.1 Αντικείμενο Εργασίας

Στην εργασία επιλύθηκε ένα πρόβλημα παλινδρόμησης (regression) με την εκπαίδευση ασαφών νευρωνικών δικτύων. Στο πρώτο κομμάτι της εργασίας εκπαιδεύτηκαν και αναλύθηκαν 4 διαφορετικά TSK μοντέλα για το CCPP Dataset του UCI Repository. Στο δεύτερο κομμάτι της εργασίας, εκπαιδεύτηκε ένα αρκετά πιο σύνθετο dataset, στο οποίο βρήκαμε τον βέλτιστο συνδυασμό χαρακτηριστικών και κανόνων, με τη χρήση του αλγορίθμου relief και εφαρμόζοντας grid partitioning με 5 fold cross validation στο dataset.

1.2 Scripts Εργασίας

Για την εκπόνηση της εργασίας χρησιμοποιήθηκαν τα παρακάτω scripts σε Matlab

ccpp_tsk_model_1.m: Εκπαίδευση του 1^{ου} μοντέλου για το πρώτο κομμάτι της εργασίας

ccpp_tsk_model_2.m: Εκπαίδευση του 2^{ου} μοντέλου για το πρώτο κομμάτι της εργασίας

ccpp_tsk_model_3.m: Εκπαίδευση του 3^{ου} μοντέλου για το πρώτο κομμάτι της εργασίας

ccpp_tsk_model_4.m: Εκπαίδευση του 4^{ου} μοντέλου για το πρώτο κομμάτι της εργασίας

grid_search.m: Script που πραγματοποιεί αναζήτηση πλέγματος για την εύρεση του βέλτιστου συνδυασμού κανόνων και χαρακτηριστικών για το 2^ο κομμάτι της εργασίας

FinalModel.m: Εκπαίδευση του τελικού TSK μοντέλου με τον συνδυασμό κανόνων και χαρακτηριστικών που βρήκε ο αλγόριθμος αναζήτησης πλέγματος

2 CCPP Dataset

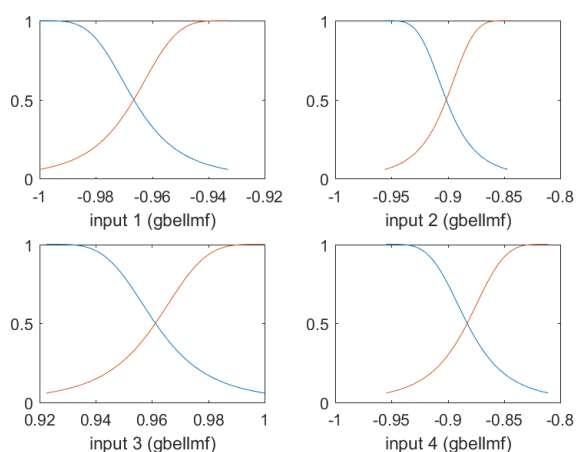
Στο πρώτο κομμάτι της εργασίας μας ζητήθηκε να εκπαιδεύσουμε 4 TSK μοντέλα με διαφορετικό πλήθος συναρτήσεων συμμετοχής (2 ή 3) και διαφορετικό τύπο εξόδου (Singleton ή Polynomial). Το εν λόγω πρώτο κομμάτι βρίσκεται στο φάκελο Project 3/matlab/part1.

Σύμφωνα με την εκφώνηση χωρίσαμε το dataset σε set εκπαίδευσης (60%), set επικύρωσης (20%) και set ελέγχου (20%). Επιπρόσθετα λόγω του διαφορετικού εύρους των τιμών των features του dataset έγινε κανονικοποίηση των δεδομένων στο διάστημα $[-1,1]$ με τη χρήση της συνάρτησης `mat2gray` (image processing toolbox) έτσι ώστε να βελτιστοποιήσουμε την εκπαίδευση του νευρωνικού δικτύου. Στη συνέχεια, με τη χρήση της συνάρτησης `genfis1` εκπαιδεύσαμε τα 4 TSK μοντέλα που μας ζητήθηκαν, για 100 εποχές το καθένα, αφού διαπιστώθηκε ότι δεν υπήρχε βελτίωση του training error για μεγαλύτερο αριθμό εποχών. Τέλος υπολογίσθηκαν οι ζητούμενοι δείκτες απόδοσης προκειμένου να συγκρίνουμε τα 4 μοντέλα και να αποφανθούμε για το ποιο μοντέλο είναι το καλύτερο.

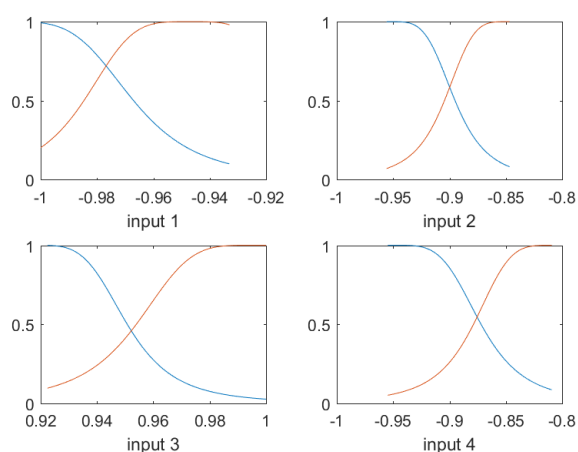
2.1 TSK Model 1: 2 membership functions - Singleton output

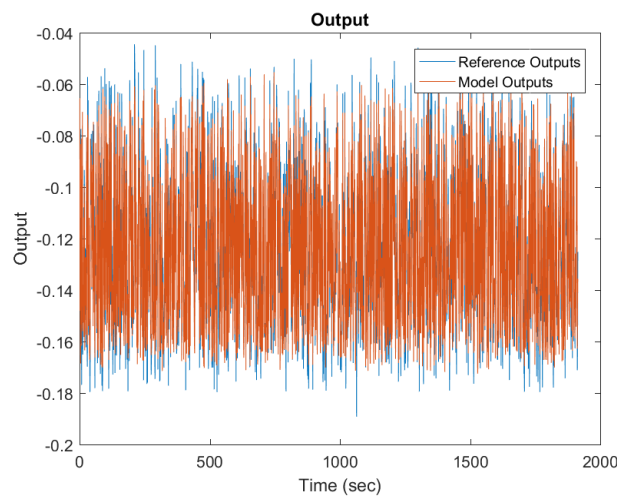
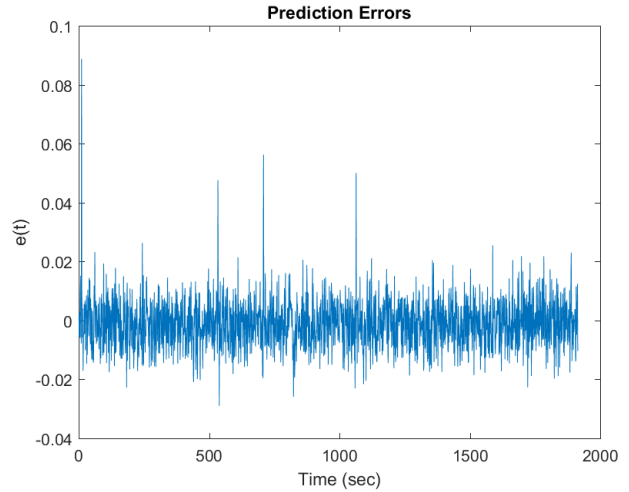
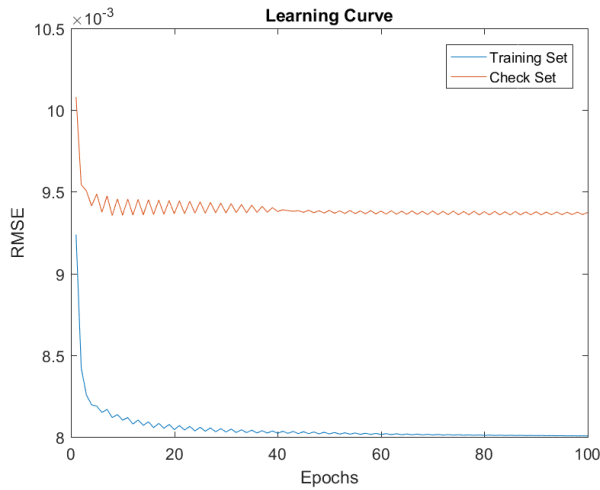
Παρακάτω παρουσιάζονται οι συναρτήσεις συμμετοχής που προέκυψαν πριν και μετά την εκπαίδευση του μοντέλου, η καμπύλη μάθησης, η καμπύλη σφάλματος πρόβλεψης καθώς και η καμπύλη εξόδου του μοντέλου.

TSK model 1 : membership functions before training



TSK model 1 : membership functions after training

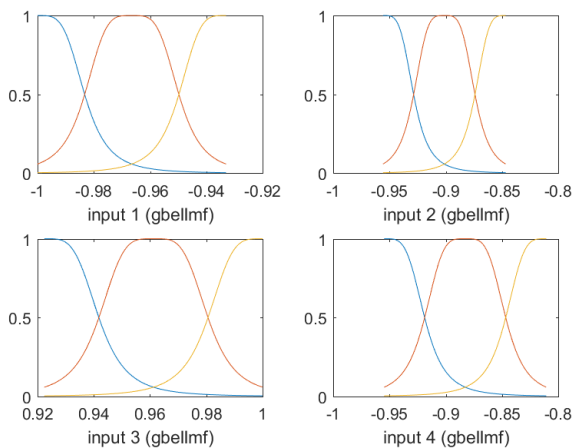




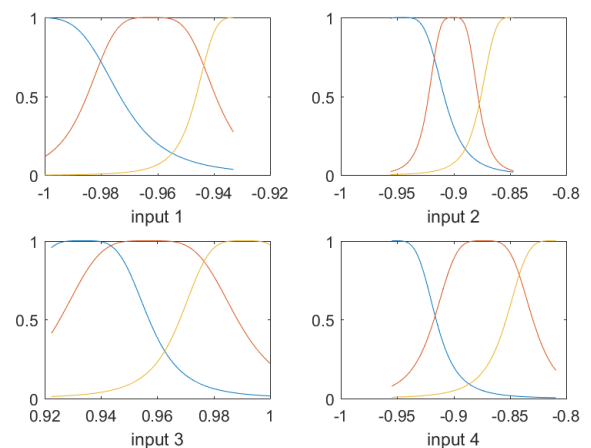
2.2 TSK Model 2: 3 membership functions - Singleton output

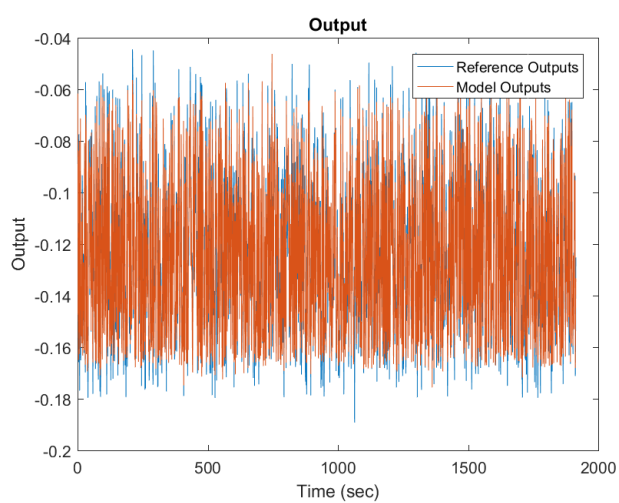
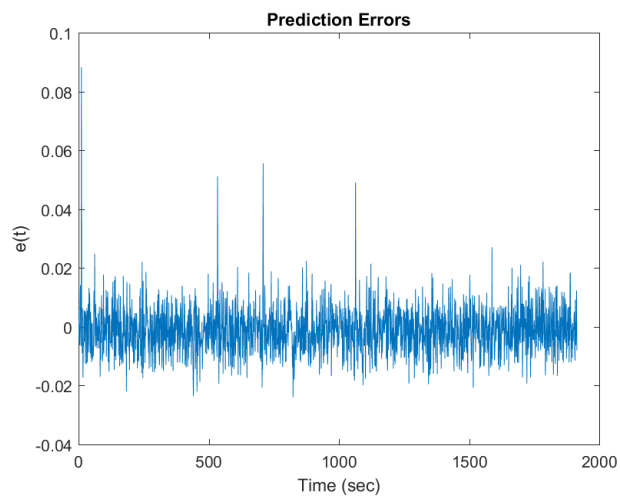
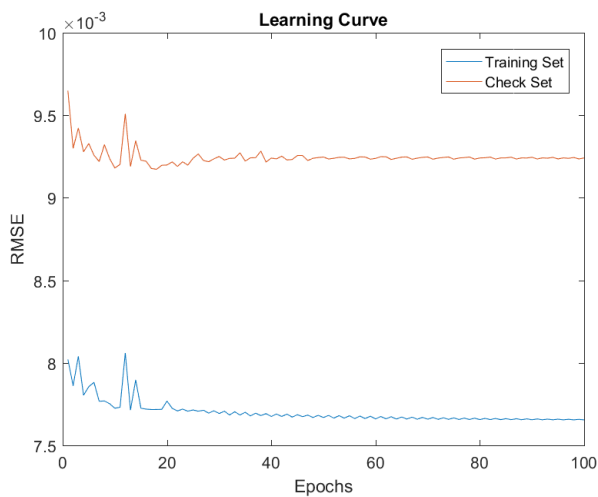
Παρακάτω παρουσιάζονται οι συναρτήσεις συμμετοχής που προέκυψαν πριν και μετά την εκπαίδευση του μοντέλου, η καμπύλη μάθησης, η καμπύλη σφάλματος πρόβλεψης καθώς και η καμπύλη εξόδου του μοντέλου.

TSK model 2 : membership functions before training



TSK model 2 : membership functions after training

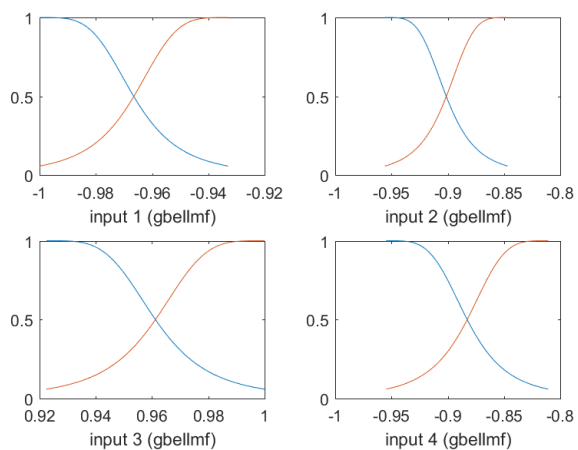




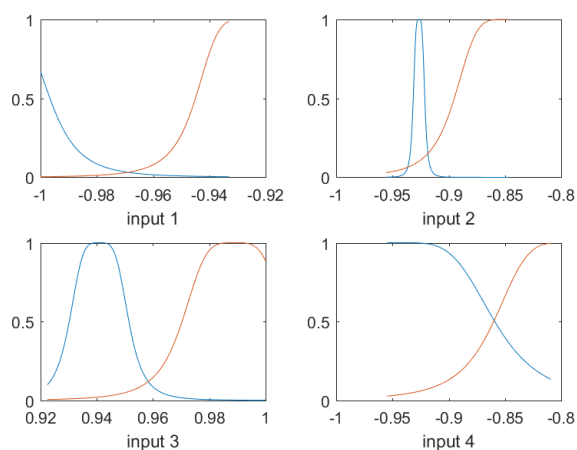
2.3 TSK Model 3: 2 membership functions – Polynomial output

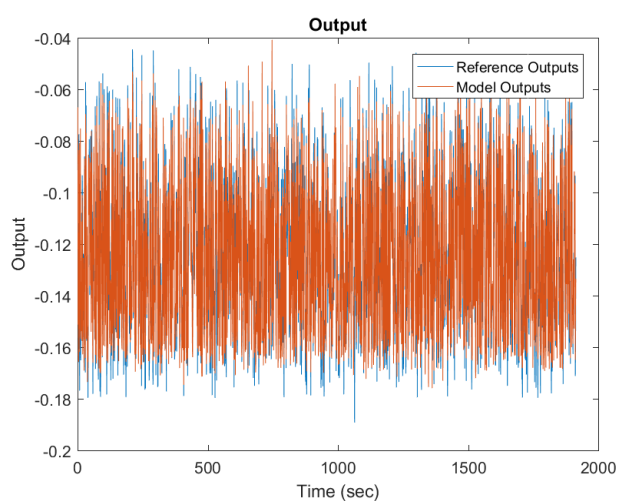
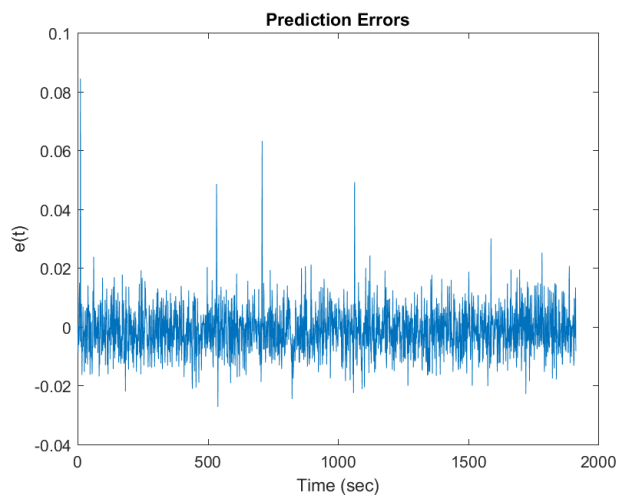
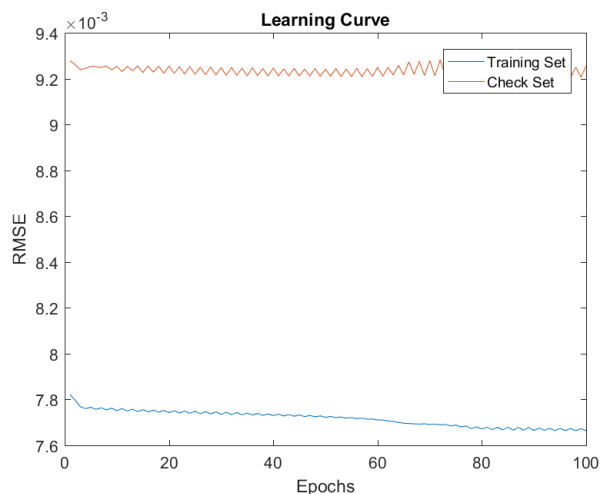
Παρακάτω παρουσιάζονται οι συναρτήσεις συμμετοχής που προέκυψαν πριν και μετά την εκπαίδευση του μοντέλου, η καμπύλη μάθησης, η καμπύλη σφάλματος πρόβλεψης καθώς και η καμπύλη εξόδου του μοντέλου.

TSK model 3 : membership functions before training



TSK model 3 : membership functions after training

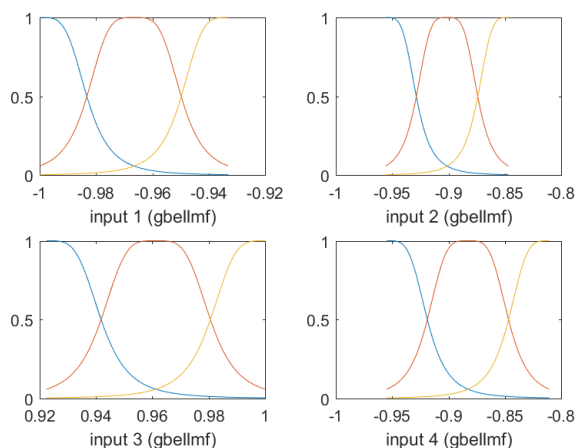




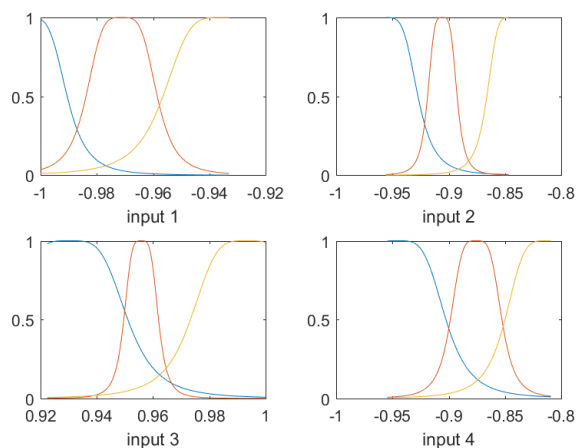
2.4 TSK Model 4: 3 membership functions – Polynomial output

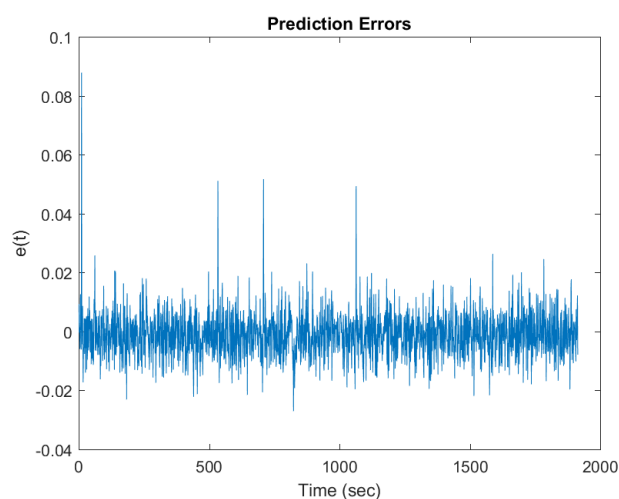
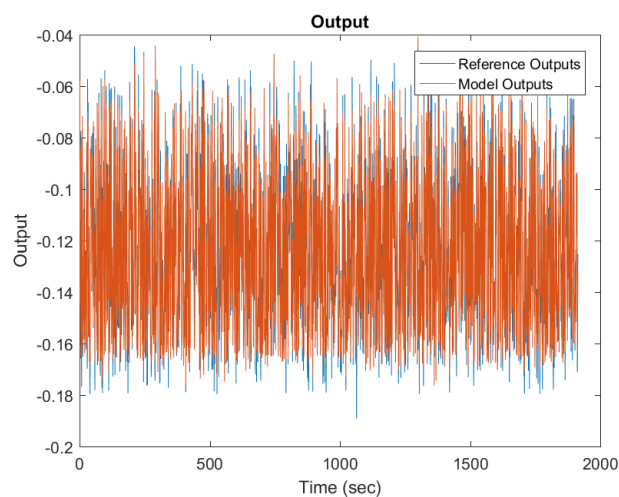
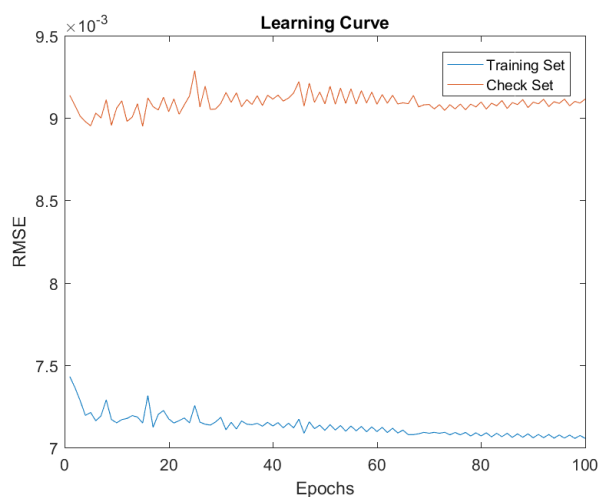
Παρακάτω παρουσιάζονται οι συναρτήσεις συμμετοχής που προέκυψαν πριν και μετά την εκπαίδευση του μοντέλου, η καμπύλη μάθησης, η καμπύλη σφάλματος πρόβλεψης καθώς και η καμπύλη εξόδου του μοντέλου.

TSK model 4 : membership functions before training



TSK model 4 : membership functions after training





2.5 Σχολιασμός μοντέλων και σύγκριση δεικτών

Στον παρακάτω πίνακα παρουσιάζονται οι δείκτες $MSE, RMSE$, $R^2, NMSE, NDEI$ για κάθε μοντέλο

	MSE	$RMSE$	R^2	$NMSE$	$NDEI$
Model 1	0.000073	0.008520	0.933198	0.065596	0.256117
Model 2	0.000067	0.008178	0.938462	0.060279	0.245518
Model 3	0.000068	0.008218	0.937858	0.060803	0.246583
Model 4	0.000063	0.007924	0.942224	0.056685	0.238085

Όπως φαίνεται στο πίνακα και τα σχήματα και τα 4 μοντέλα παρακολουθούν την έξοδο αρκετά καλά και οι διαφορές μεταξύ των αποδόσεων τους είναι μικρές. Ωστόσο το μοντέλο 4 (3 mf- Polynomial output) έχει το μικρότερο μέσο τετραγωνικό σφάλμα και ταυτόχρονα το μεγαλύτερο συντελεστή προσδιορισμού άρα είναι και το καλύτερο

3 Superconductivity Dataset

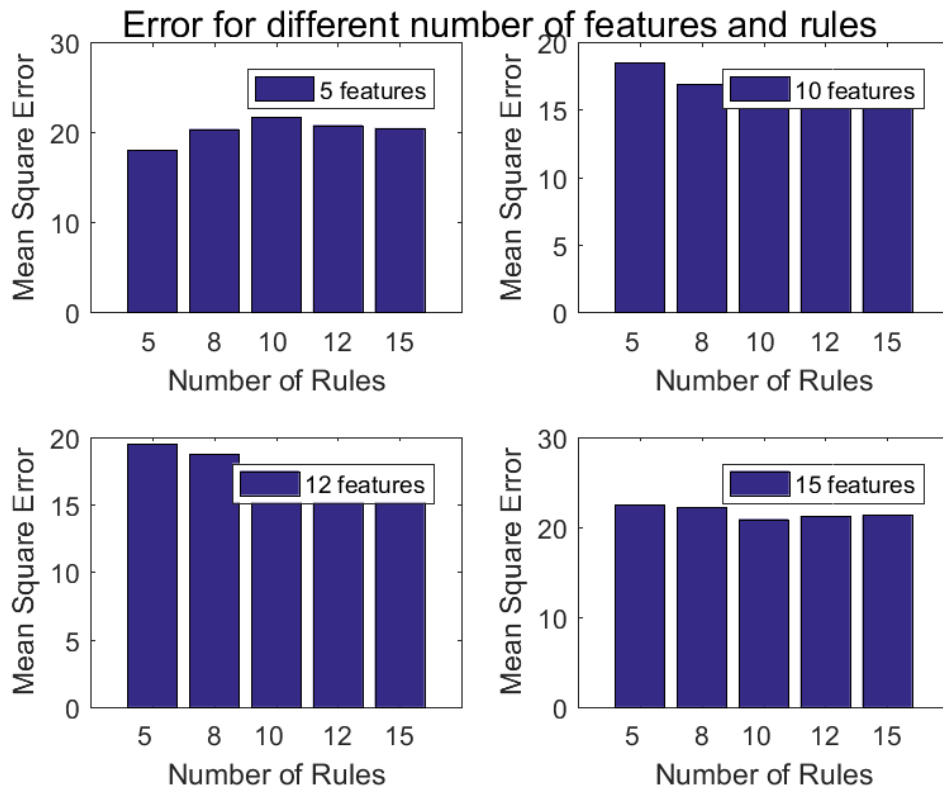
Λόγω της υψηλής διασπασιμότητας του 2^{ου} dataset ήταν αδύνατον να εκπαιδεύσουμε ένα TSK μοντέλο όπως κάναμε προηγουμένως. Για αυτόν τον λόγο αρχικά χρησιμοποιήσαμε τον αλγόριθμο Relief ώστε να κατατάξουμε τα χαρακτηριστικά με βάση την χρησιμότητα τους. Έπειτα εφαρμόσαμε αναζήτηση πλέγματος και 5-fold cross validation για την επιλογή του βέλτιστου συνδυασμού χαρακτηριστικών και κανόνων. Ως μέθοδος ομαδοποίησης για την δημιουργία των ασαφών κανόνων επιλέχθηκε ο αλγόριθμος Fuzzy C-Means. Η διαδικασία που ακολουθήσαμε για την επιλογή του βέλτιστου σετ παραμέτρων καθώς και τα αποτελέσματα που πήραμε από την εκπαίδευση του τελικού TSK μοντέλου περιγράφονται παρακάτω

3.1 Grid Search & 5-fold Cross validation

Σύμφωνα με την εκφώνηση τα σετ παραμέτρων με τα οποία έπρεπε να εργαστούμε ήταν τα NF=[5 10 15 20] και NR= [5 10 15 20 25]. Ωστόσο κατά το cross validation για τιμές NF>= 15 και NR>=15 αντιμετωπίσαμε το πρόβλημα του Rule Explosion και ήταν αδύνατο να εκπαιδεύσουμε μοντέλο και να εξάγουμε τους δείκτες για να αξιολογήσουμε την αποτελεσματικότητα του. Τελικά, έγινε αναπροσαρμογή των σετ παραμέτρων σε NF=[5 10 12 15] και NR=[5 8 10 12 15] για να πάρουμε πιο έγκυρα αποτελέσματα. Ο αλγόριθμος που ακολουθήσαμε για να βρούμε τον βέλτιστο συνδυασμό παραμέτρων είναι ο ακόλουθος. Η εκπαίδευση των TSK μοντέλων έγινε για 5 εποχές, καθώς ο χρόνος εκτέλεσης για μεγαλύτερο αριθμό εποχών ήταν απαγορευτικός και δεν παρατηρήθηκε σημαντική βελτίωση του σφάλματος

```
1 Define sets of model parameter values to evaluate
2 for each parameter set do
3     for each resampling iteration do
4         Hold-out specific samples
5         [Optional] Pre-process the data
6         Fit the model on the remainder
7         Predict the hold-out samples
8     end
9     Calculate the average performance across hold-out predictions
10 end
11 Determine the optimal parameter set
12 Fit the final model to all the training data using the optimal parameter set
```

Οι τιμές του μέσου σφάλματος για τις διαφορετικές τιμές του grid παρουσιάζονται στο παρακάτω διάγραμμα.

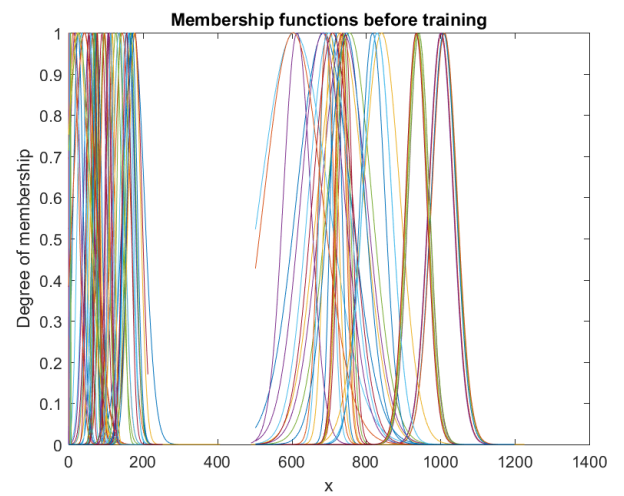
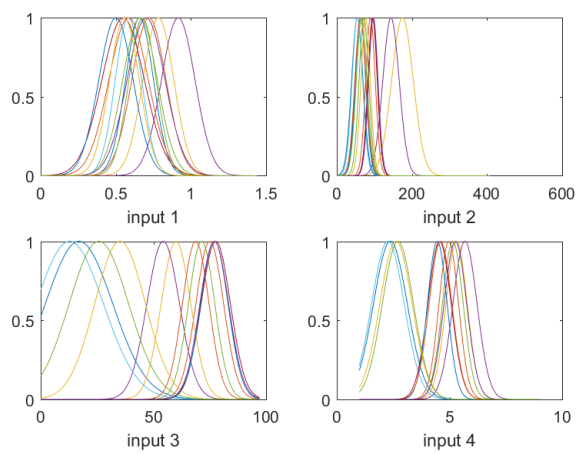


Όπως βλέπουμε, ο βέλτιστος συνδυασμός κανόνων και χαρακτηριστικών είναι 10 χαρακτηριστικά και 12 κανόνες.

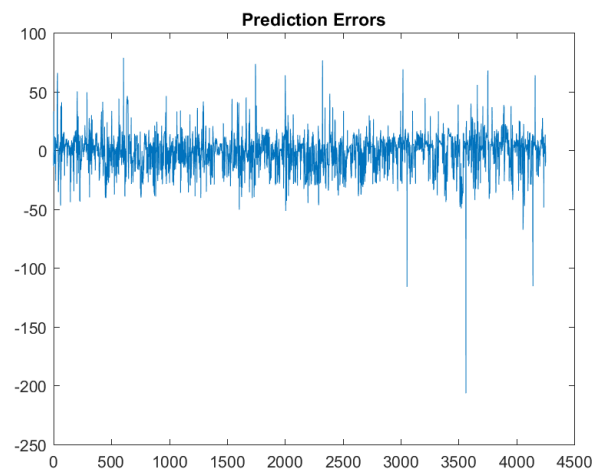
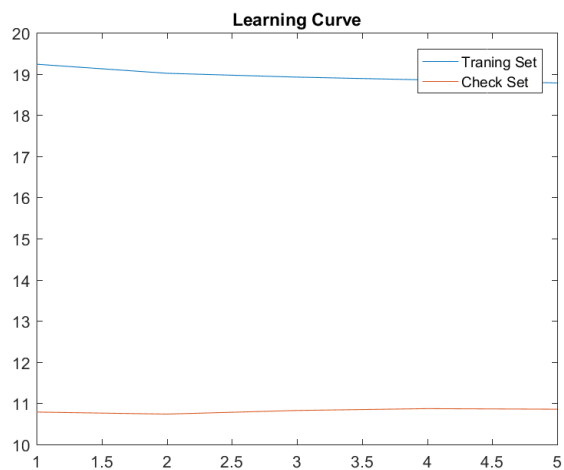
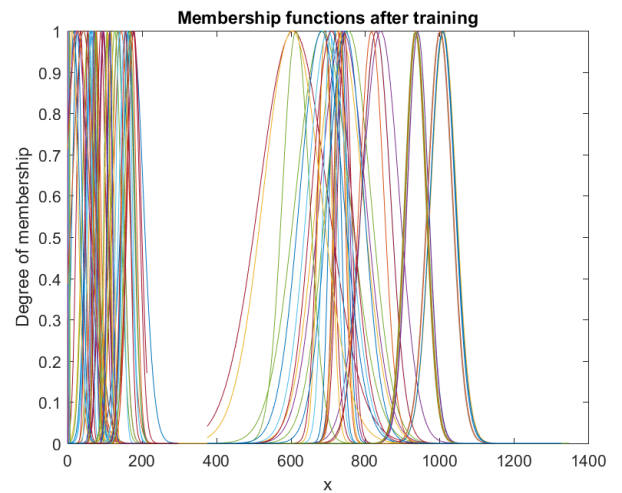
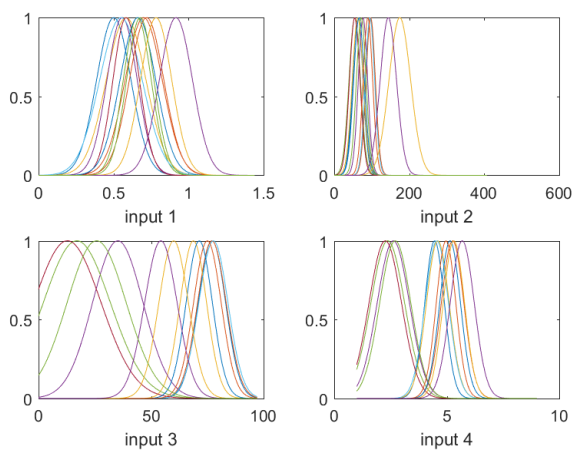
3.2 Τελικό TSK μοντέλο

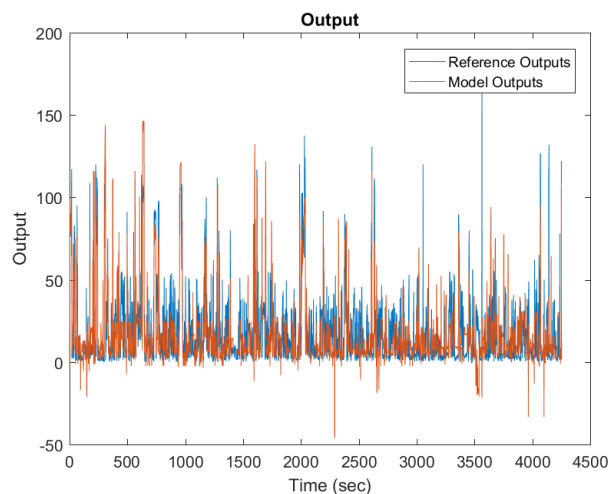
Παρακάτω παρουσιάζονται οι συναρτήσεις συμμετοχής που προέκυψαν πριν και μετά την εκπαίδευση του μοντέλου, η καμπύλη μάθησης, η καμπύλη σφάλματος πρόβλεψης καθώς και η καμπύλη εξόδου του τελικού TSK μοντέλου που εκπαιδεύτηκε για 10 χαρακτηριστικά και 12 κανόνες

Final TSK model : some membership functions before training



Final TSK model : some membership functions after training



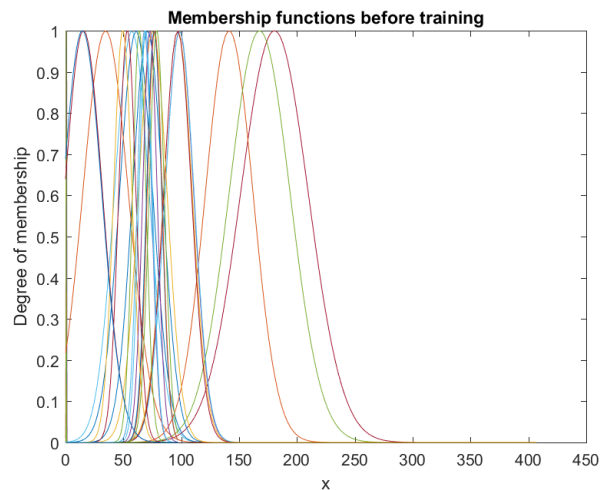
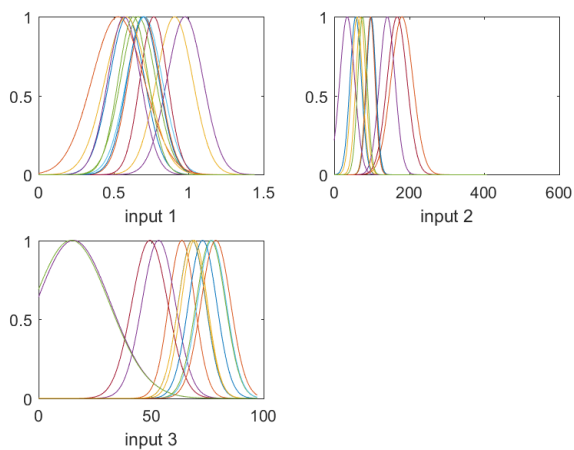


MSE	243.943762
$RMSE$	15.618699
R^2	0.545924
$NMSE$	0.449782
$NDEI$	0.670658

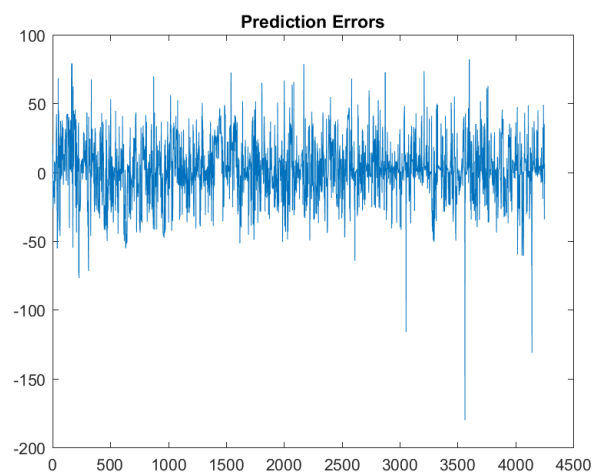
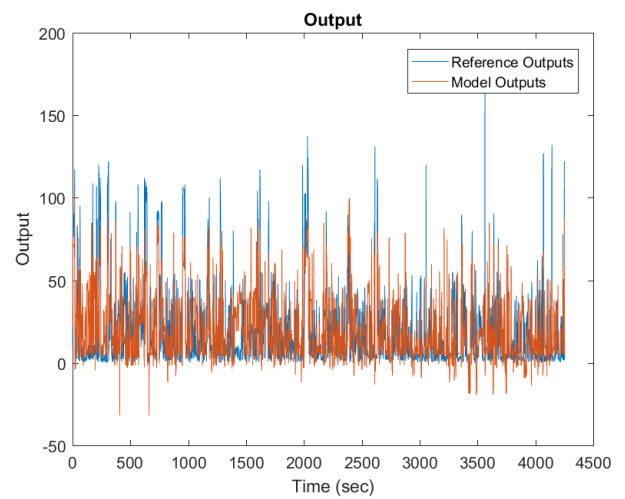
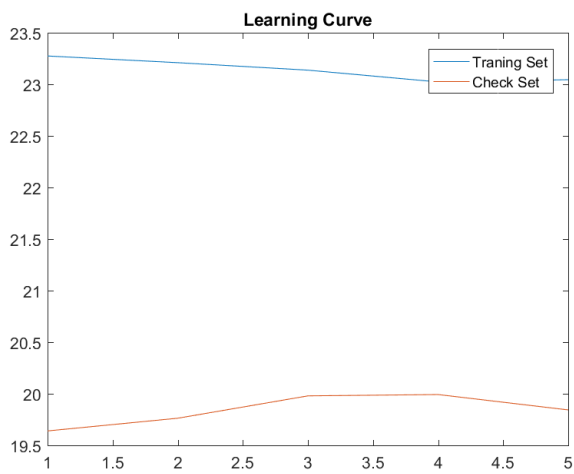
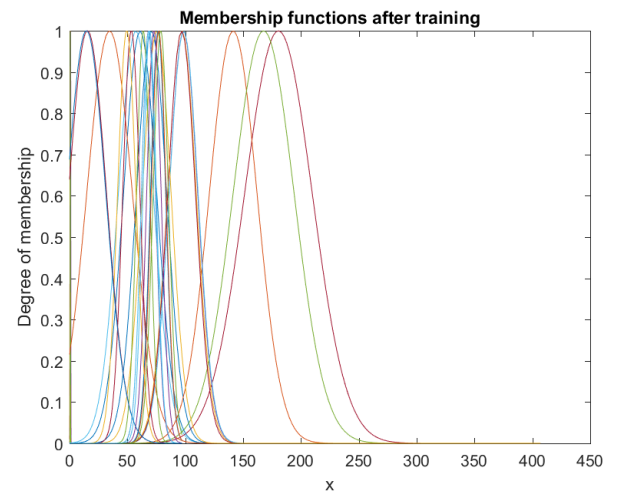
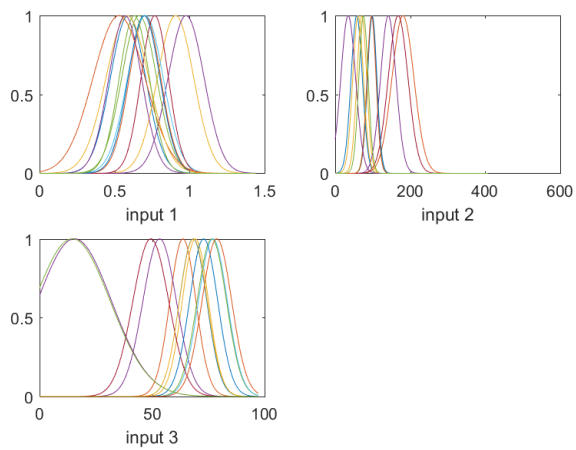
3.2.1 Σύγκριση με μοντέλο 3 χαρακτηριστικών

Εκπαιδεύουμε ένα μοντέλο με 3 χαρακτηριστικά και 12 κανόνες ώστε να το συγκρίνουμε με το βέλτιστο μοντέλο που εκπαιδεύτηκε προηγουμένως. Παρακάτω παρουσιάζονται οι καμπύλες και οι δείκτες του μοντέλου

Final TSK model : some membership functions before training



Final TSK model : some membership functions after training



<i>MSE</i>	433.773338
<i>RMSE</i>	20.827226
<i>R²</i>	0.192576
<i>NMSE</i>	0.802690
<i>NDEI</i>	0.895930

Όπως βλέπουμε, παρόλο που το βέλτιστο μοντέλο θα μπορούσε να χαρακτηριστεί underfit καθώς έχουμε *RMSE* της τάξης του 8,5% περίπου και συντελεστή προσδιορισμού 0,54, εντέλει αποδίδει αρκετά καλύτερα από το μοντέλο των 3 χαρακτηριστικών και 12 κανόνων. Συνεπώς ο αλγόριθμος αναζήτησης πλέγματος δούλεψε ικανοποιητικά.