

3^η Εργασία στην Υπολογιστική Νοημοσύνη-Regression

Δημήτριος Τικβίνας

AEM 9998

Σεπτέμβριος 2023

dtikvina@ece.auth.gr

Εισαγωγή

Σκοπός της εργασίας αυτής είναι να διερευνηθεί η ικανότητα των μοντέλων TSK (Takagi – Sugeno - Kang), στην μοντελοποίηση πολυμεταβλητών, μη γραμμικών συναρτήσεων. Επιλέγουμε δύο σύνολα δεδομένων, από τα οποία το πρώτο θα χρησιμοποιηθεί για μια απλή διερεύνηση της διαδικασίας εκπαίδευσης και αξιολόγησης μοντέλων TSK, και το δεύτερο πολυπλοκότερο σύνολο θα χρησιμοποιηθεί για μια πληρέστερη διαδικασία μοντελοποίησης.

Εφαρμογή σε απλό dataset

Τα μοντέλα που καλούμαστε να υλοποιήσουμε είναι τα εξής

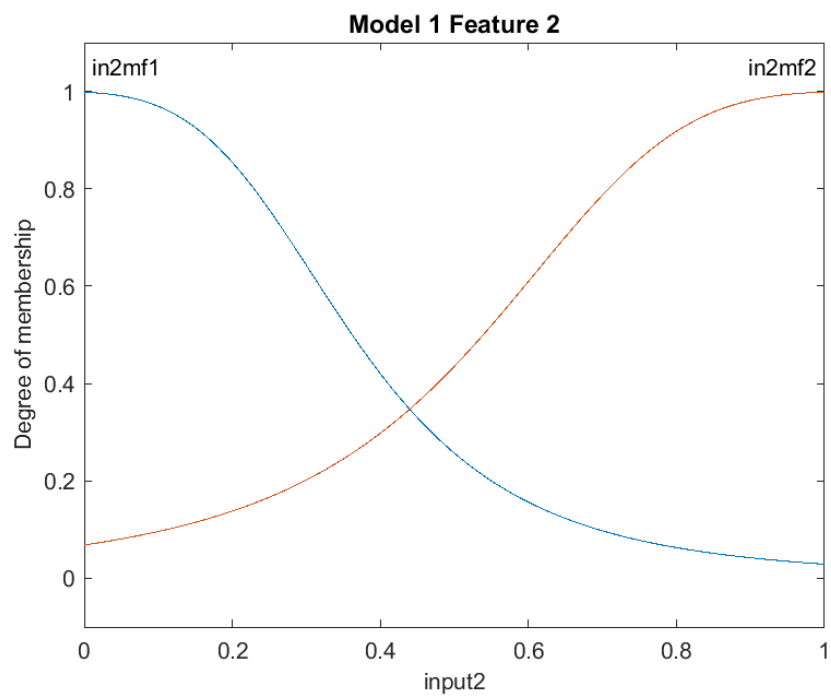
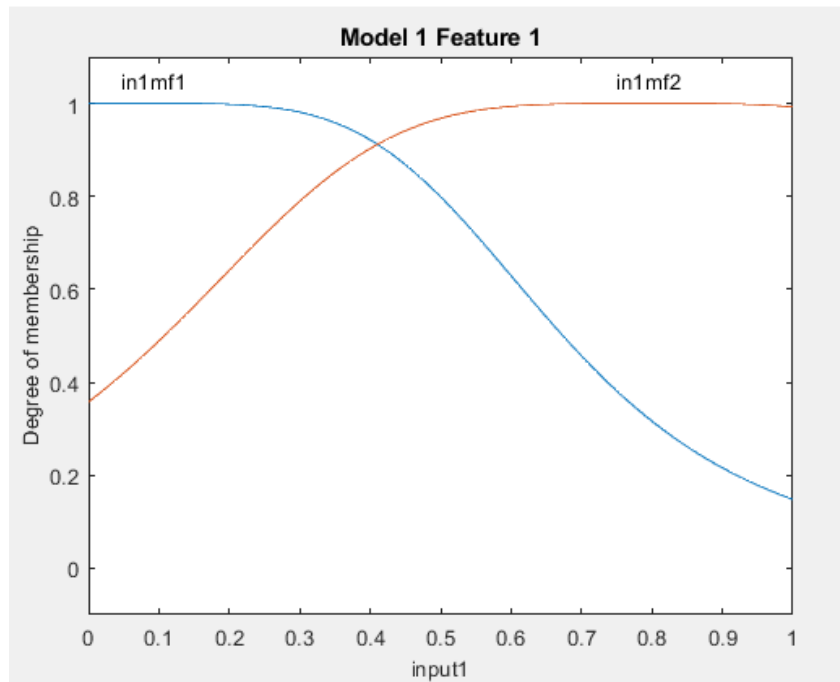
Μοντέλο	Πλήθος mfs	Μορφή εξόδου
TSK_model_1	2	Singleton
TSK_model_2	3	Singleton
TSK_model_3	2	Polynomial
TSK_model_4	3	Polynomial

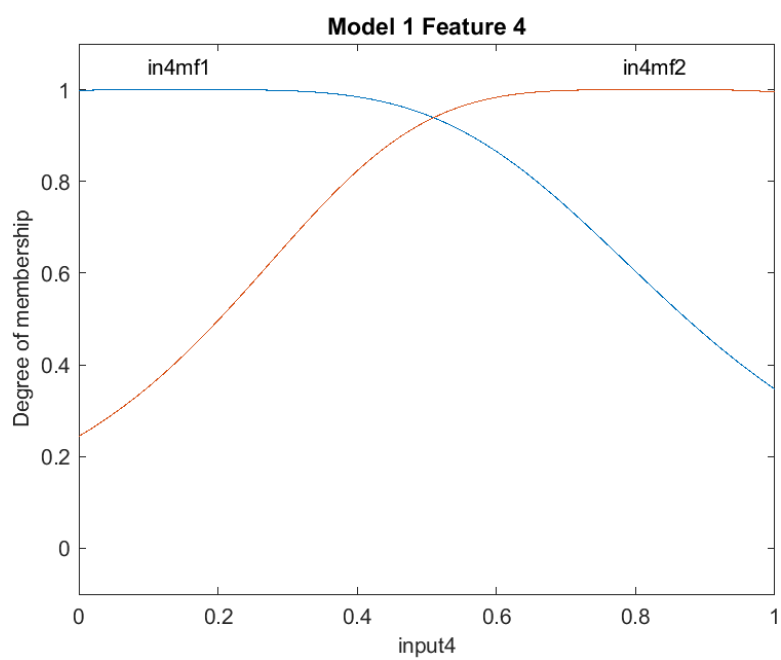
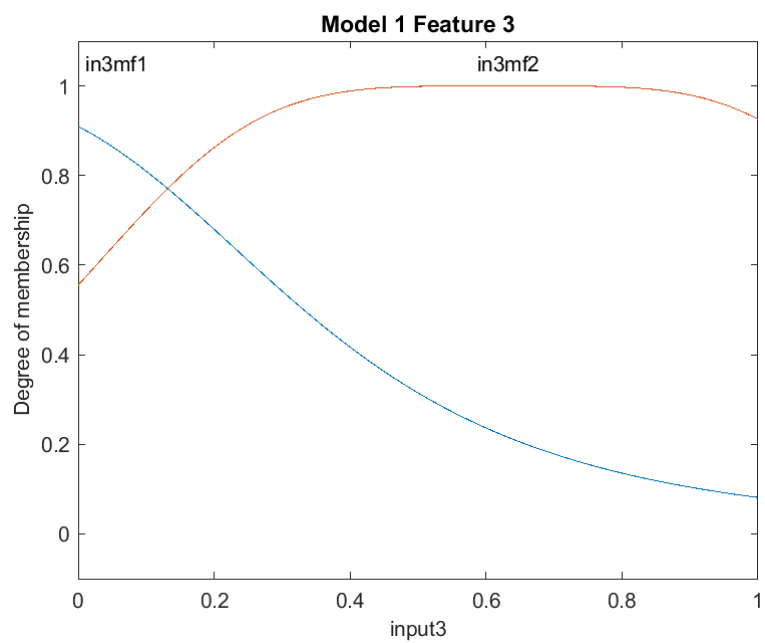
Ως dataset χρησιμοποιούμε από το UCI repository το **Airfoil Self-Noise** dataset το οποίο περιλαμβάνει 1503 δείγματα και 6 features. Η 6η στήλη είναι και η εξαρτημένη στήλη ή αλλιώς, Target feature. Στη συνέχεια, για κάθε ένα από τα παραπάνω μοντέλα υπολογίζουμε τα **R²**, **RMSE**, **NMSE** και **NDEI**. Μαζί με τους παραπάνω υπολογισμούς, δημιουργήθηκαν για κάθε μοντέλο τα εξής:

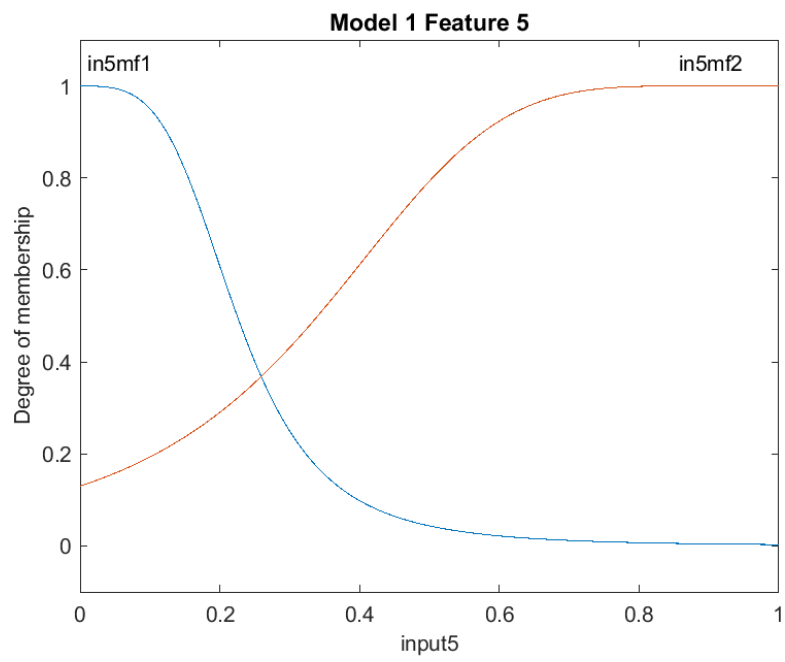
- Διάγραμμα με τις τελικές μορφές των ασαφών συνόλων
- Διάγραμμα με το learning curve
- Διάγραμμα με τα σφάλματα πρόβλεψης

1^ο Μοντέλο TSK

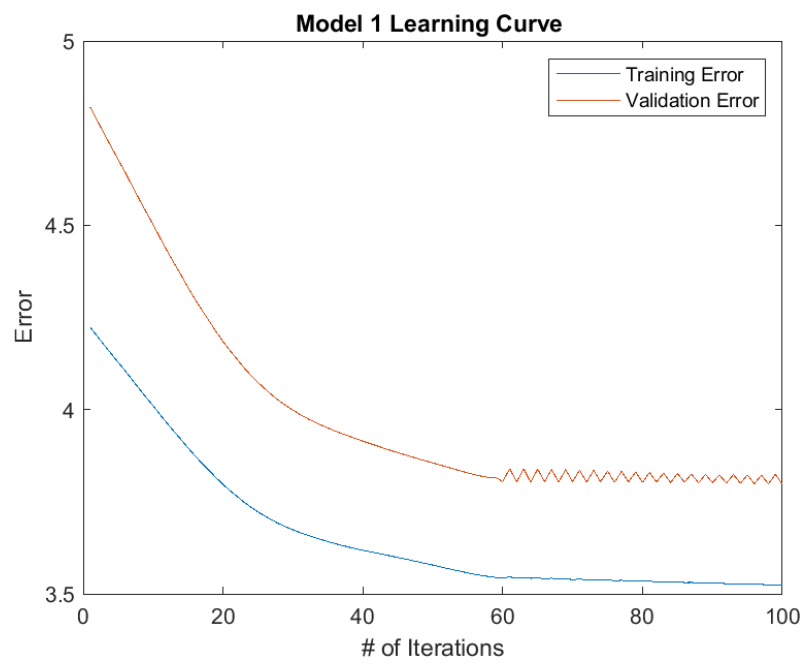
- Membership functions



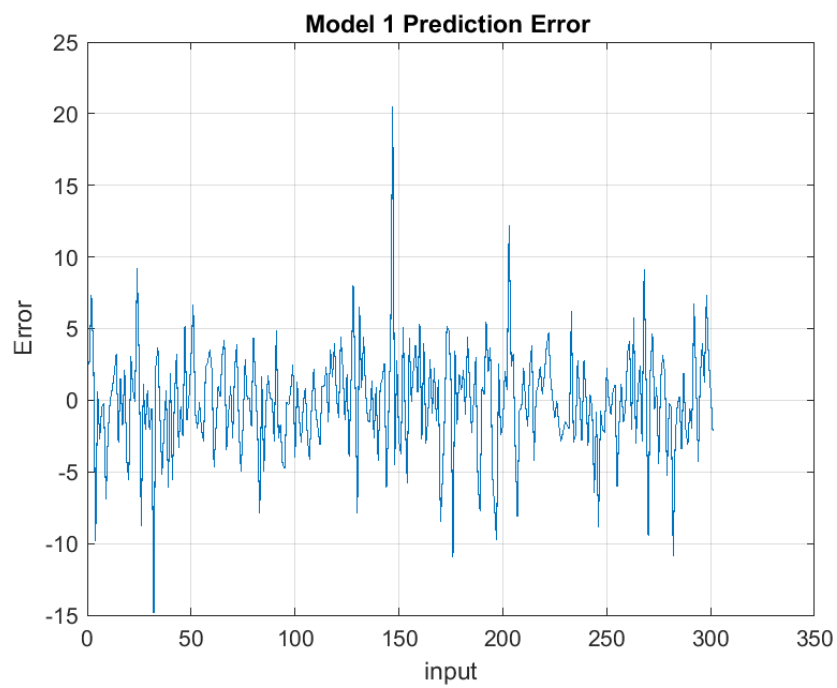




- Learning curve

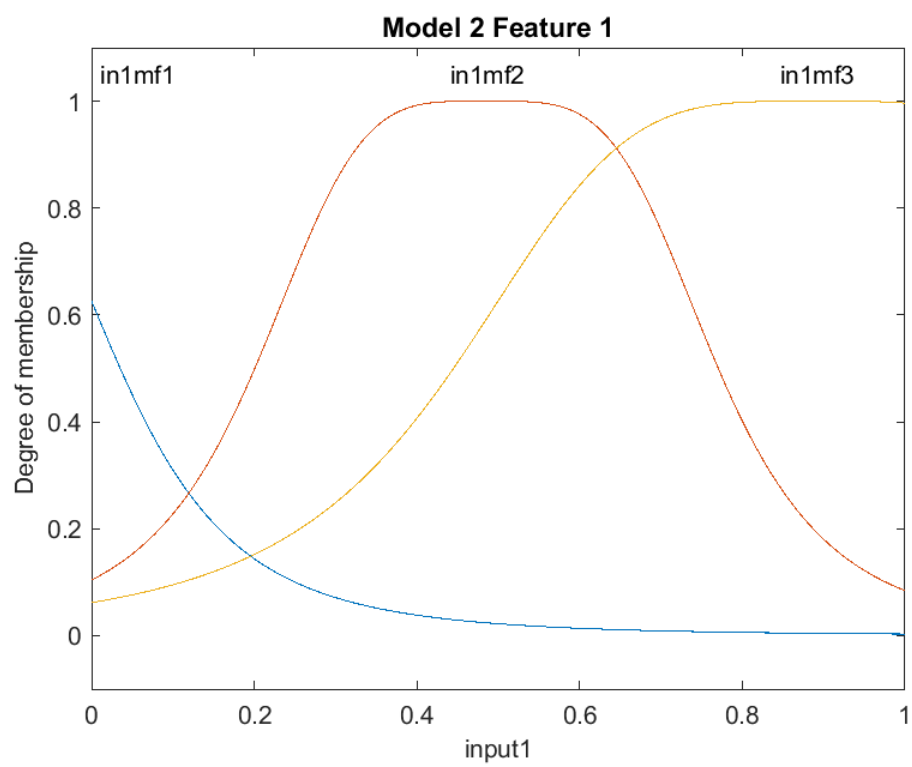


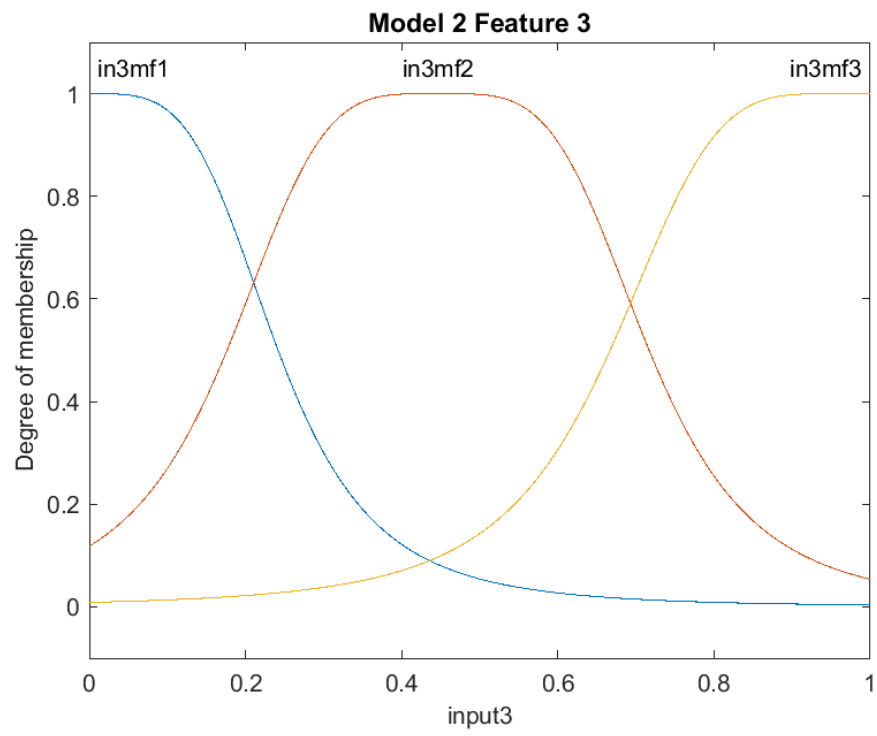
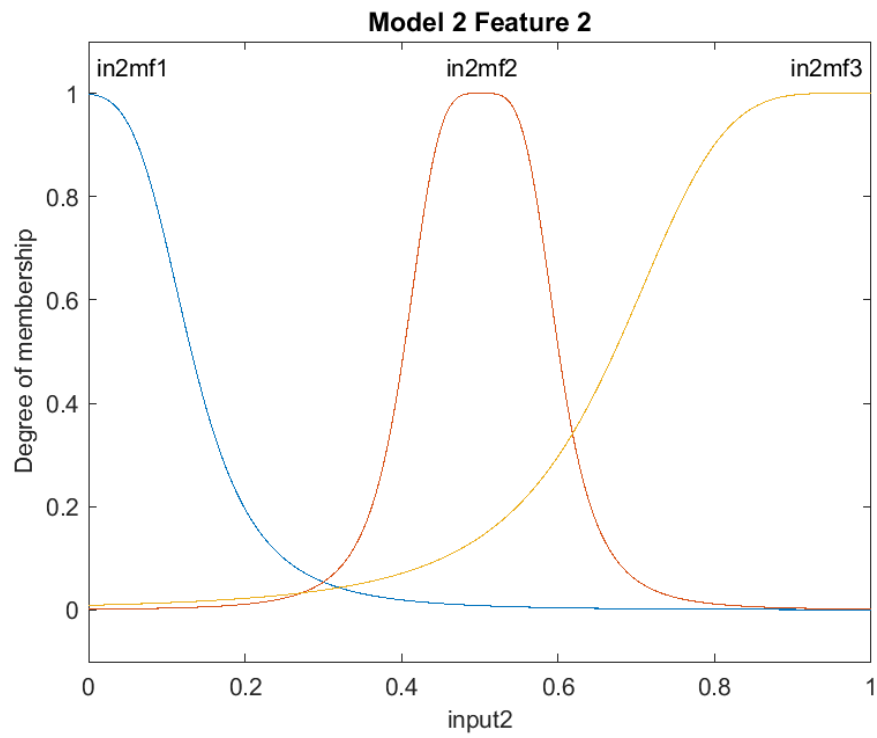
- Prediction error

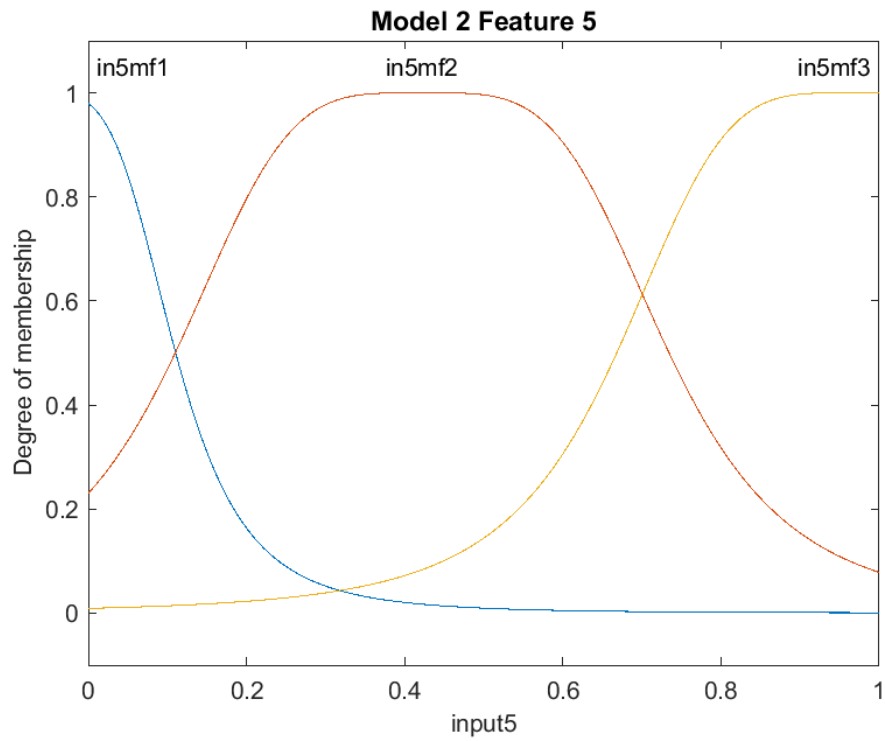
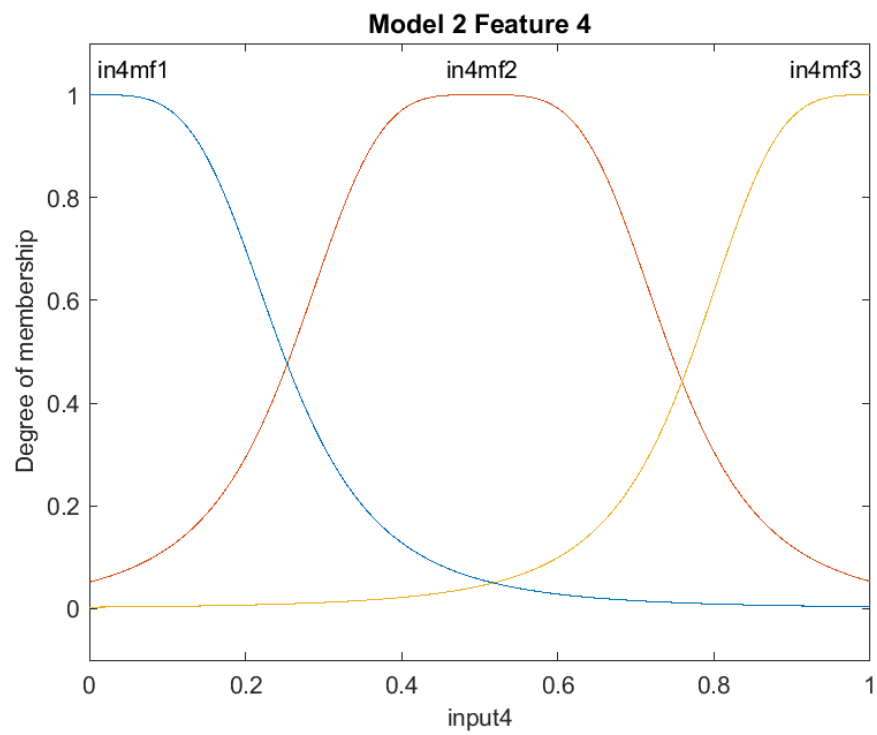


2^ο Μοντέλο TSK

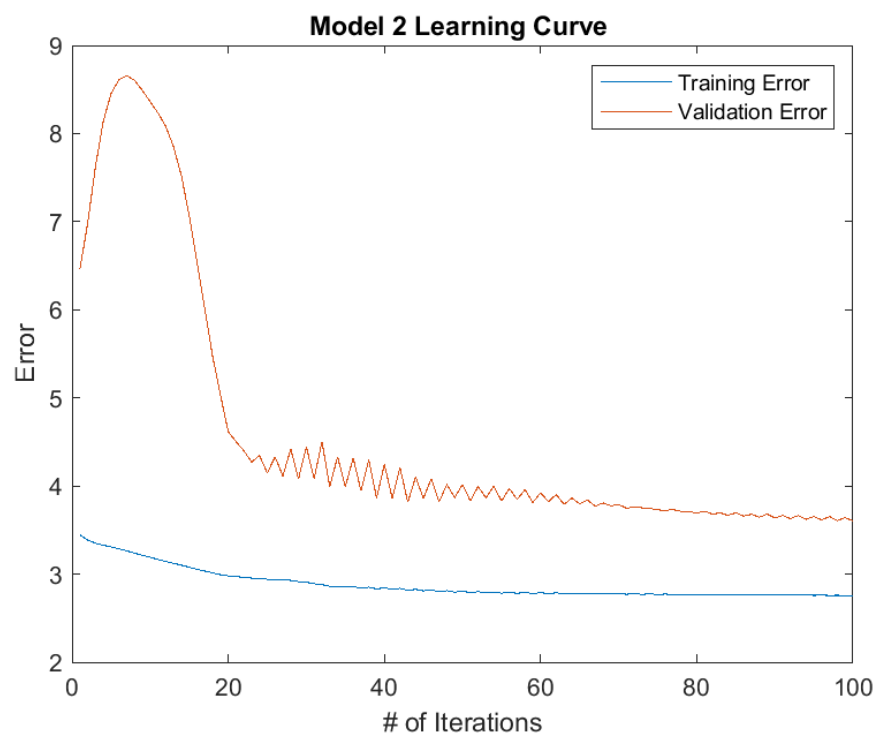
- Membership functions



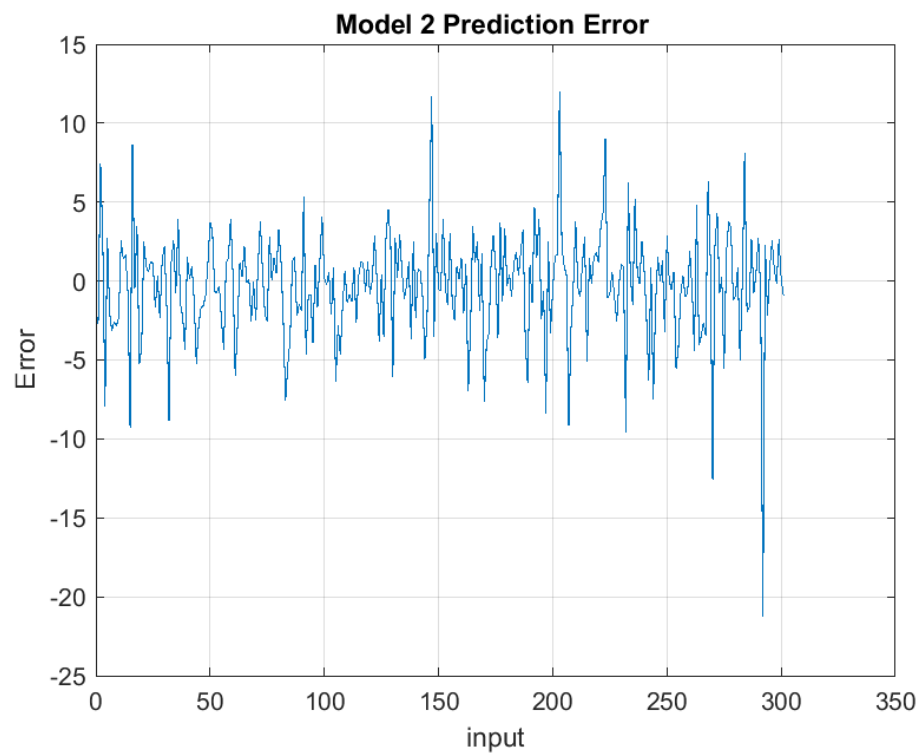




- Learning curve

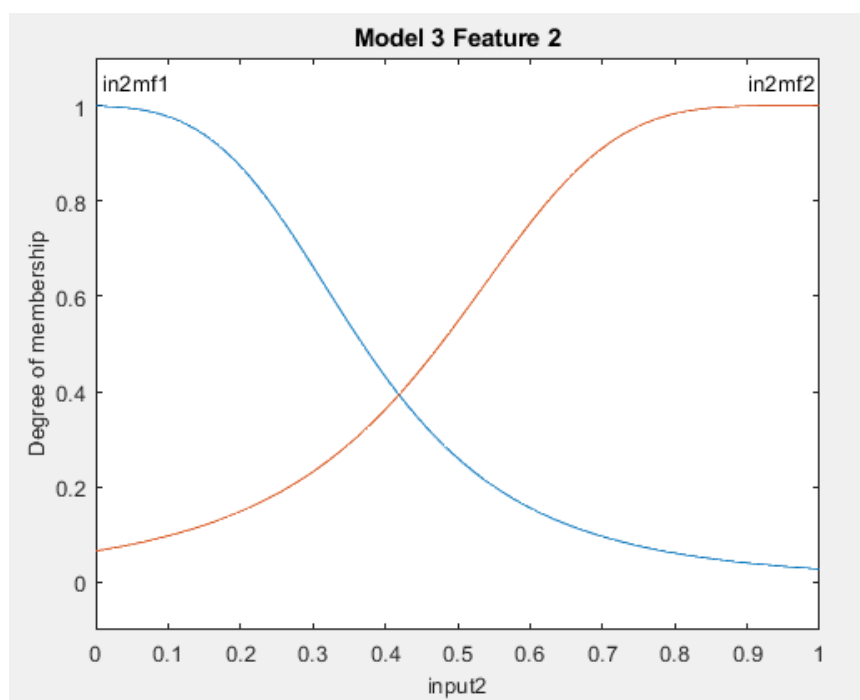
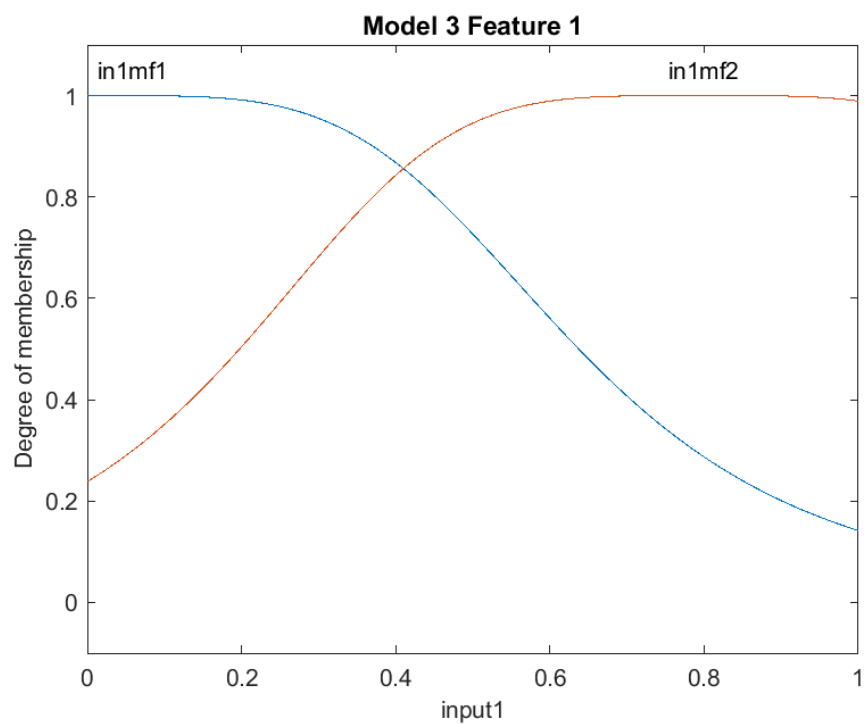


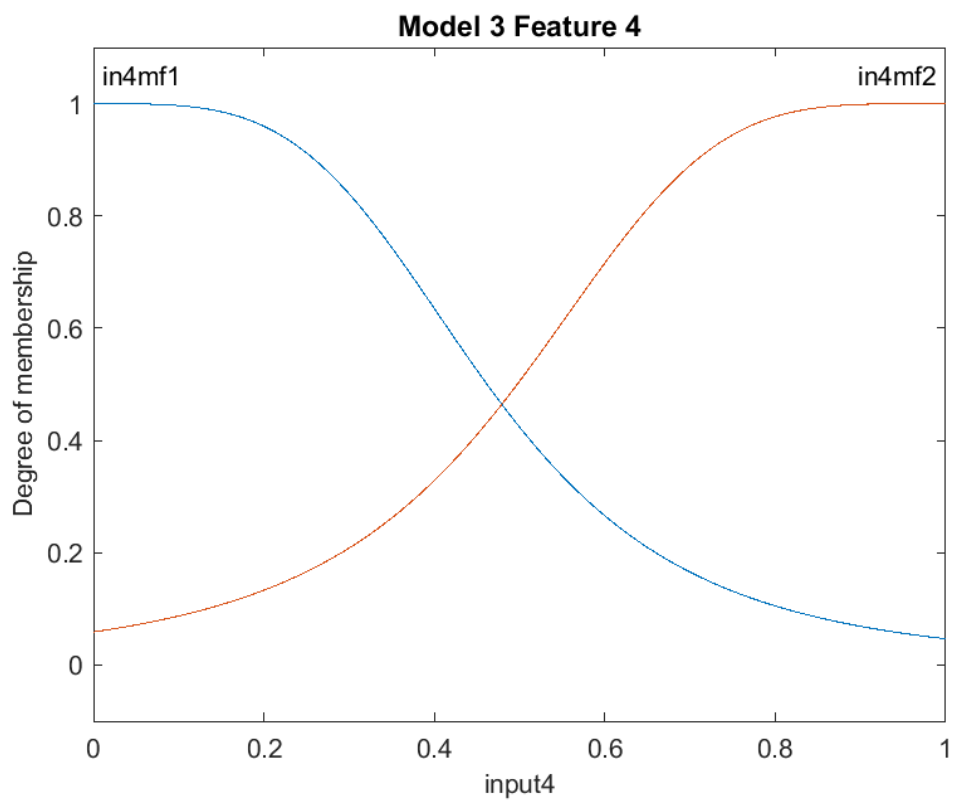
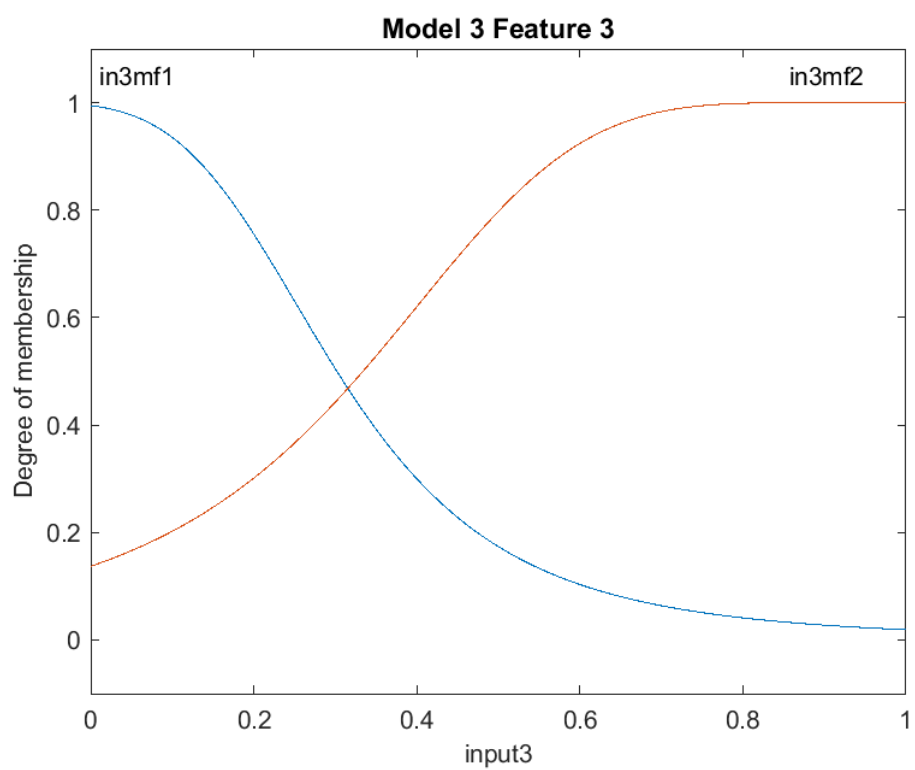
- Prediction error

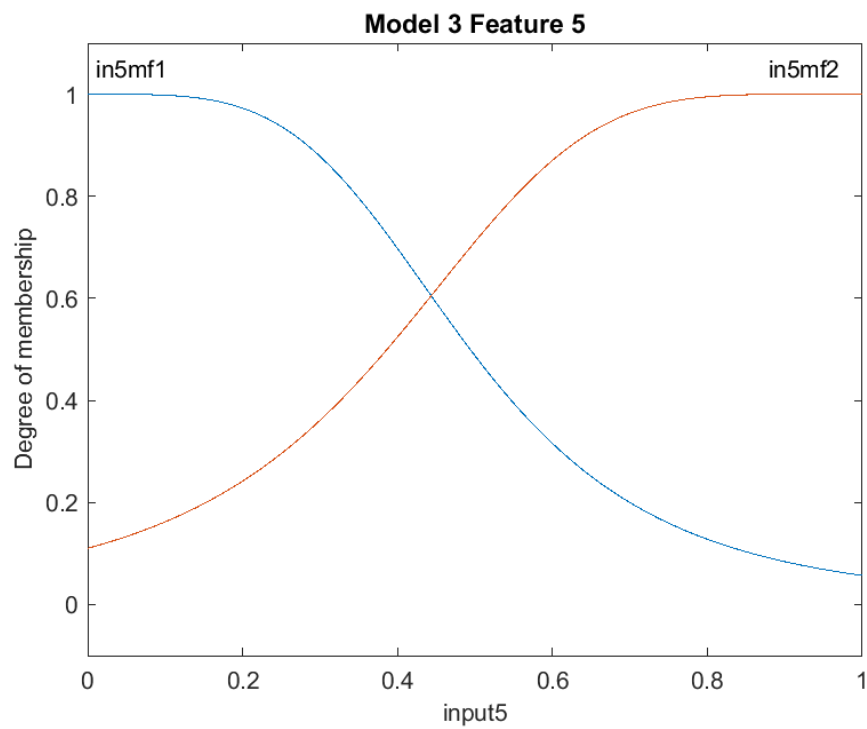


3^ο Μοντέλο TSK

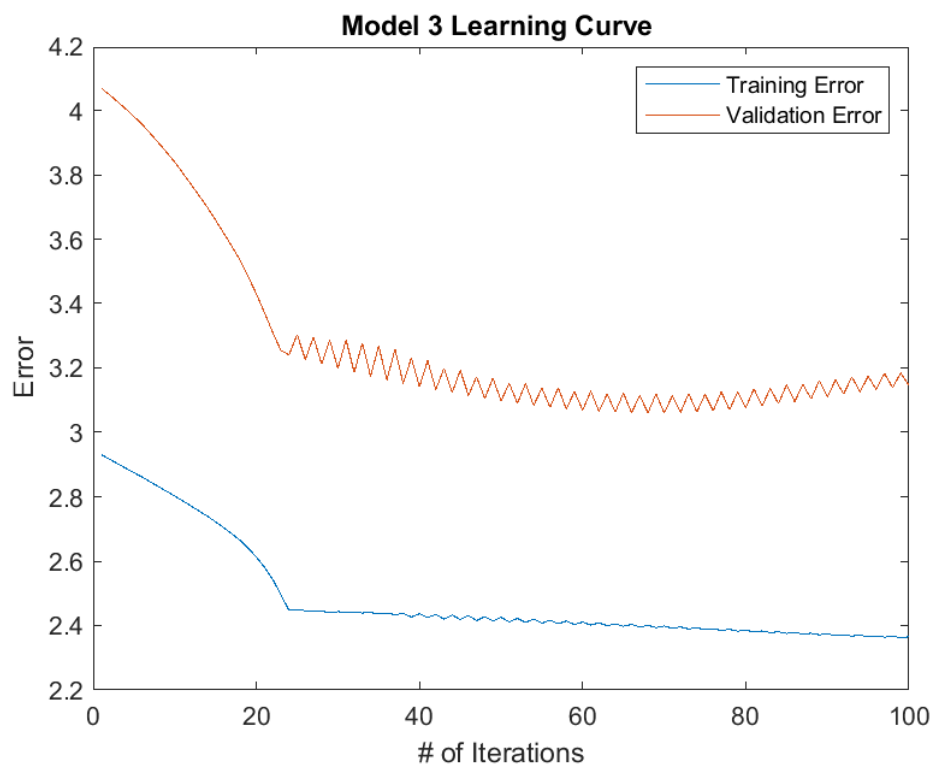
- Membership functions



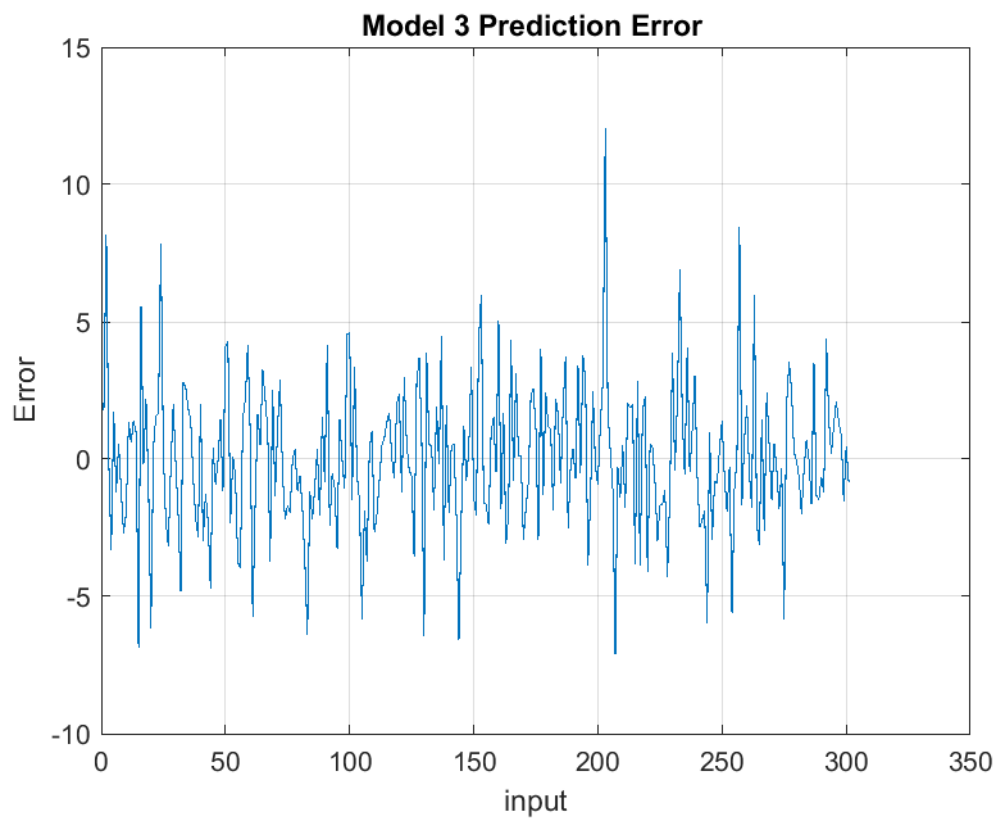




- Learning curve

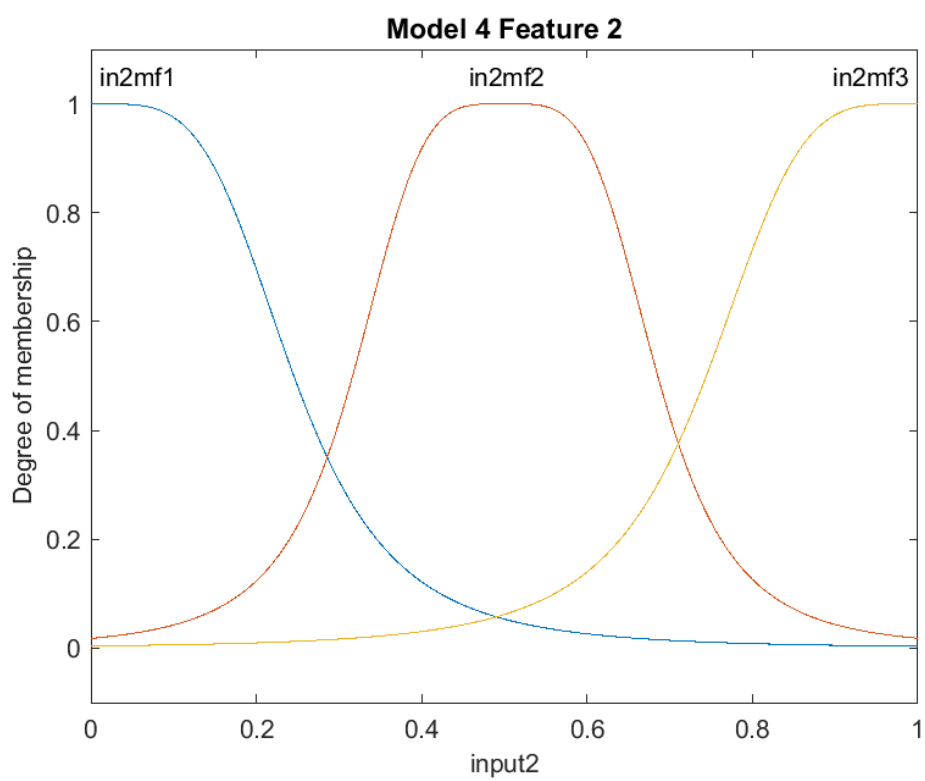
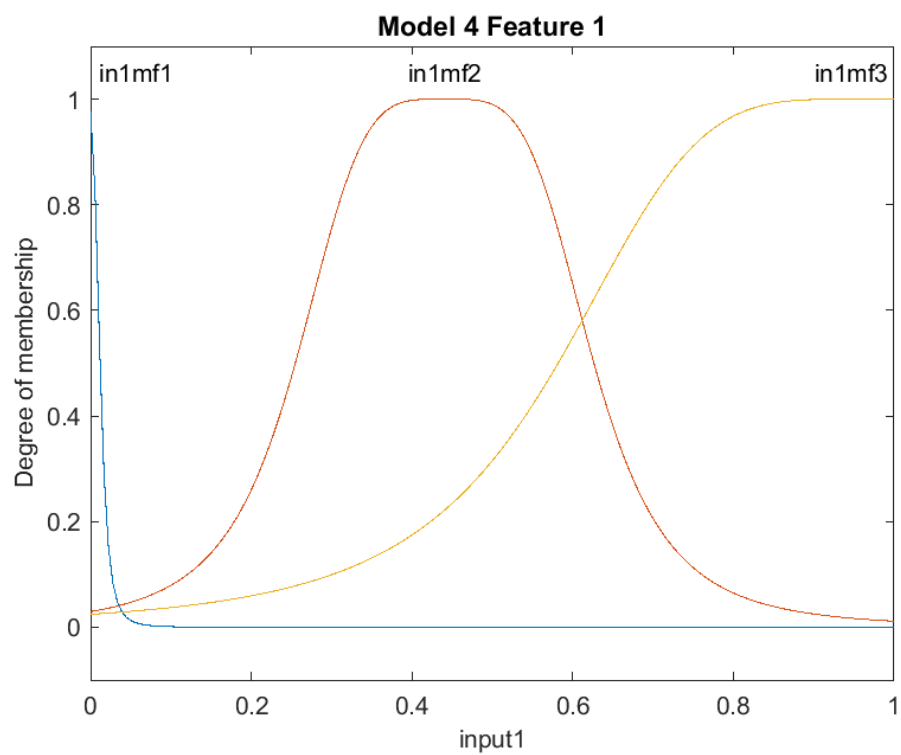


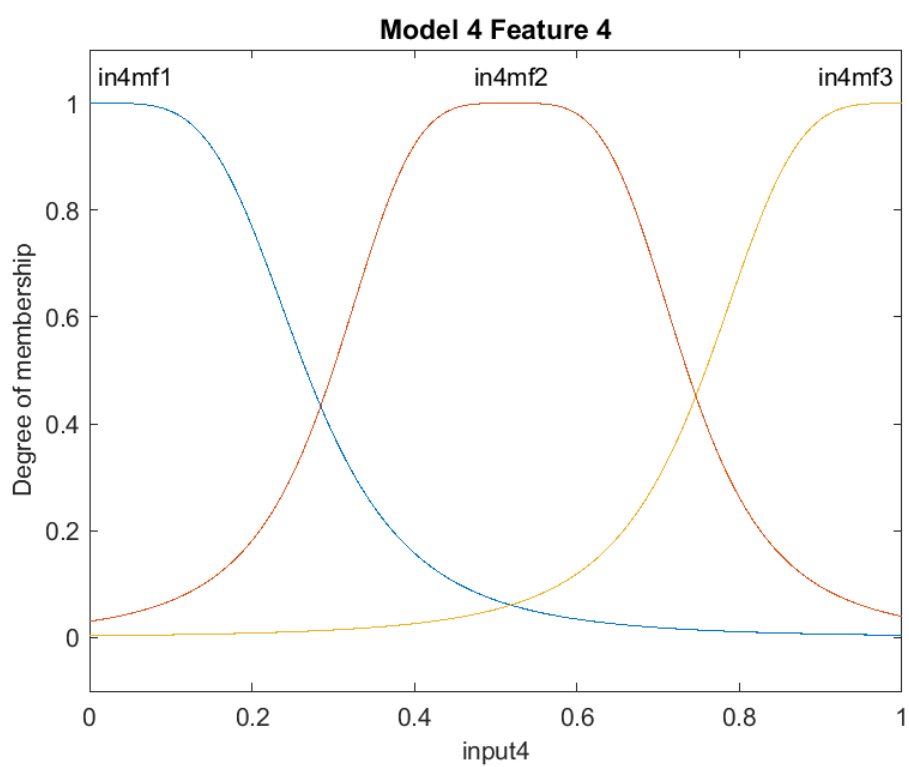
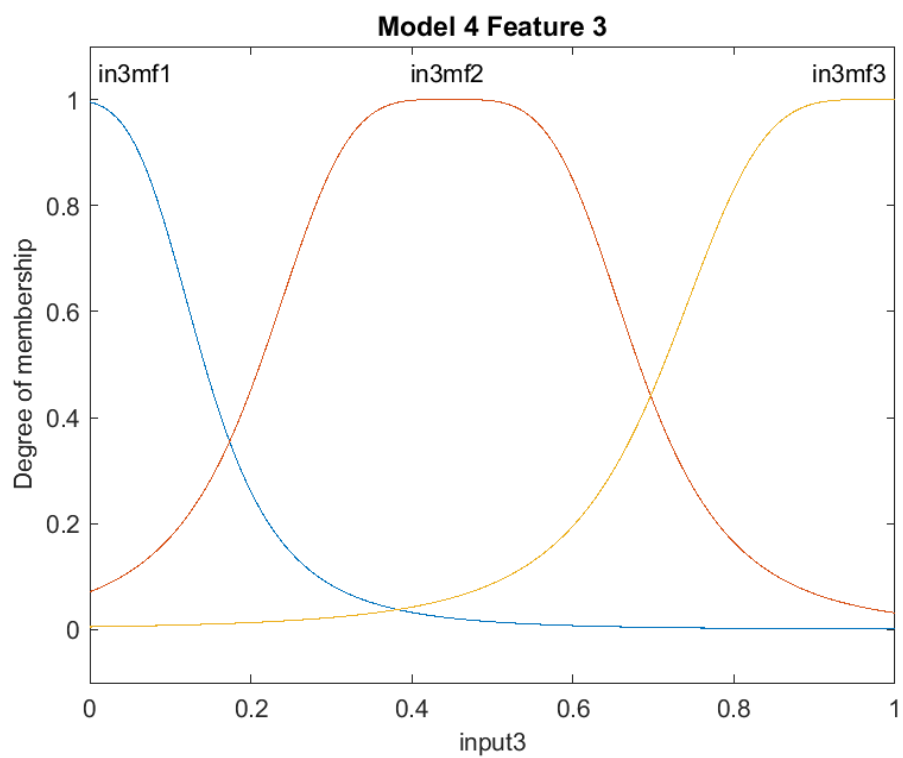
- Prediction error

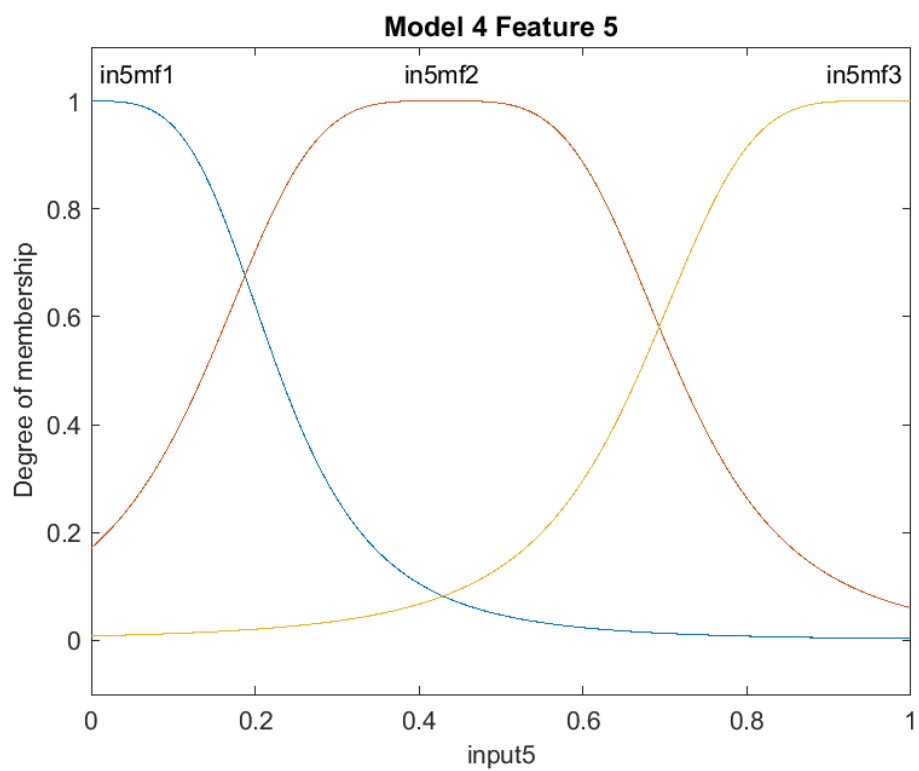


4^ο Μοντέλο TSK

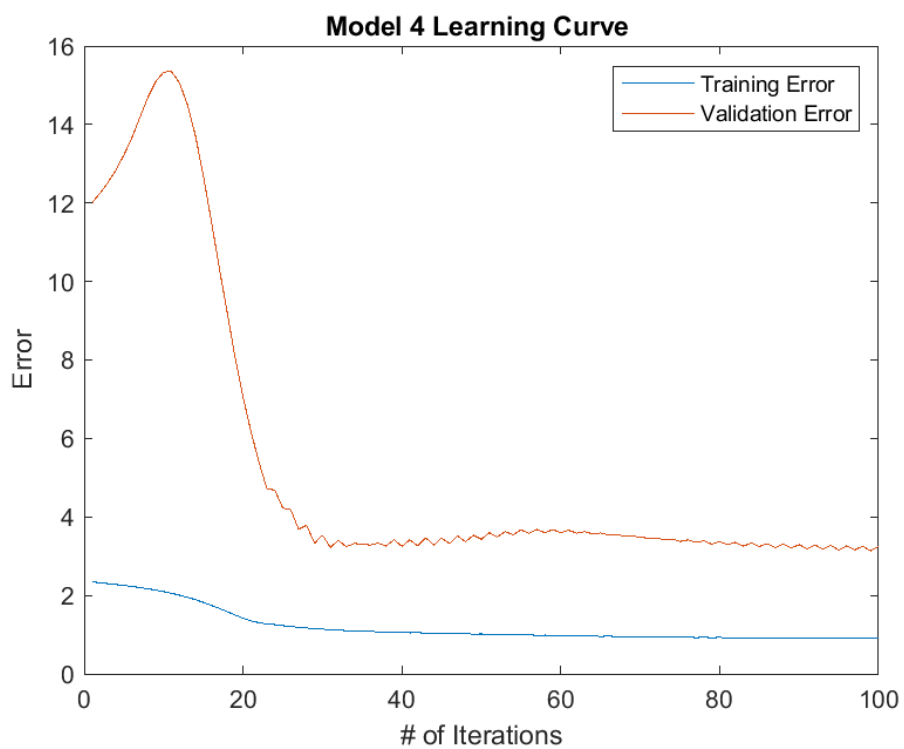
- Membership functions



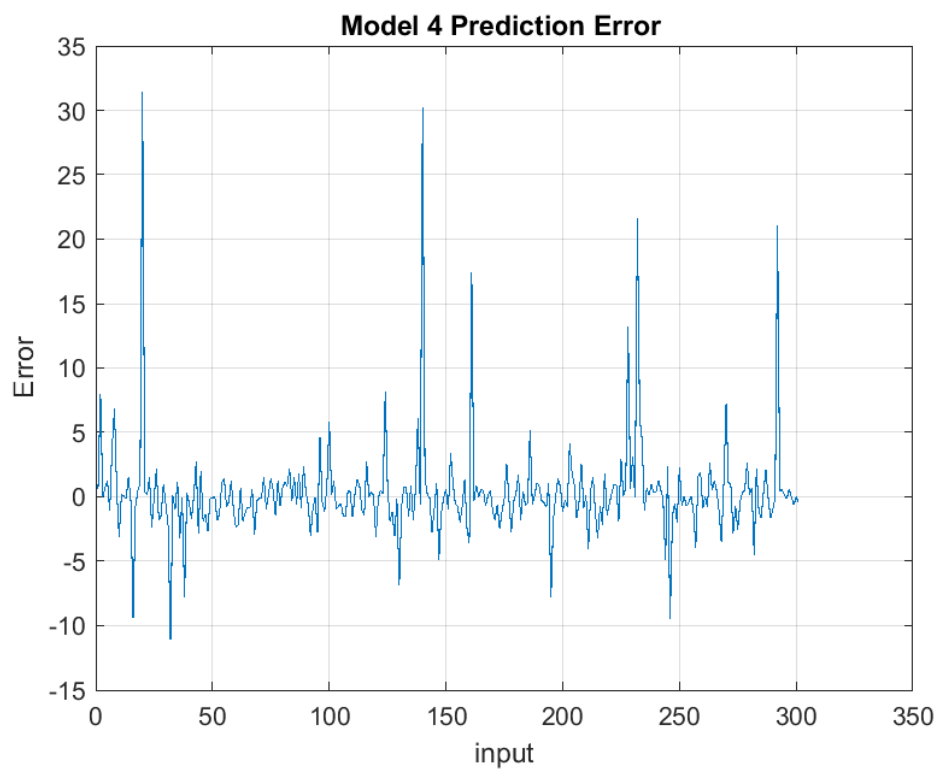




- Learning curve



- Prediction error



Μετρικές Μοντέλων

	1° Μοντέλο TSK	2° Μοντέλο TSK	3° Μοντέλο TSK	4° Μοντέλο TSK
R2	0.66211	0.71916	0.83377	0.63402
RMSE	3.8397	3.5005	2.6932	3.9961
NMSE	0.33789	0.28084	0.16623	0.36598
NDEI	0.58129	0.52994	0.40771	0.60496

Συμπεράσματα

Από τα παραπάνω διαγράμματα και τις μετρικές των μοντέλων, εξάγουμε τα εξής συμπεράσματα:

- ✓ Το **καλύτερο** μοντέλο είναι το Μοντέλο 3
- ✓ Το **χειρότερο** μοντέλο είναι το Μοντέλο 4
- ✓ Τα μοντέλα με Polynomial έξοδο παρουσιάζουν πιο ακραία Prediction errors
- ✓ Παρατηρούμε στα learning curves ότι τα μοντέλα με 2 συναρτήσεις συμμετοχής παρουσιάζουν μικρότερη απόκλιση ανάμεσα στο training και στο validation error

Εφαρμογή σε dataset με υψηλή διαστασιμότητα

Στο δεύτερο μέρος της εργασίας, θα χρησιμοποιήσουμε ένα dataset με υψηλότερο βαθμό διαστασιμότητας. Το dataset που χρησιμοποιήθηκε είναι το **Superconductivity** dataset από το UCI Repository, το οποίο περιέχει 21263 δείγματα και 81 μεταβλητές / features. Είναι προφανές ότι πρέπει να διαλέξουμε έναν μικρό αριθμό features που θα κρατήσουμε, λόγω της έκρηξης των κανόνων IF - THEN , ώστε το μοντέλο TSM να μπορεί να είναι λειτουργικό. Συνεπώς, το πρώτο πράγμα που κάνουμε σε αυτό το κομμάτι της εργασίας είναι να τρέξουμε ένα σκριπτάκι (Regression_Part_2.m) ώστε να μπορέσουμε να βρούμε τις βέλτιστες τιμές, των κρατημένων features και της ακτίνας r των clusters. Για την παραπάνω διαδικασία χρησιμοποιούμε τη μέθοδο του grid search. Ο μέγιστος αριθμός των κρατημένων features που διαλέξαμε είναι 12. Επίσης, στα features των data κάναμε normalization με την εντολή *normalize()*, η μέθοδος ομαδοποίησης για τη δημιουργία των IF – THEN κανόνων επιλέχθηκε να είναι ο αλγόριθμος *Subtractive Clustering (SC)* και η επιλογή των χαρακτηριστικών έγινε με τον αλγόριθμο *Relief*.

Από το **Grid Search** προέκυψαν οι εξής τιμές των μετρικών:

➤ RMSE

Kept_features\ r	0.2	0.4	0.6	0.8	1
3	19.1872	20.4173	21.6675	21.8687	25.0884
6	15.8451	16.2437	17.2211	19.1606	19.2684
9	16.1890	15.4427	16.4611	17.8754	18.2170
12	15.6289	14.8676	15.5020	16.7082	16.9965

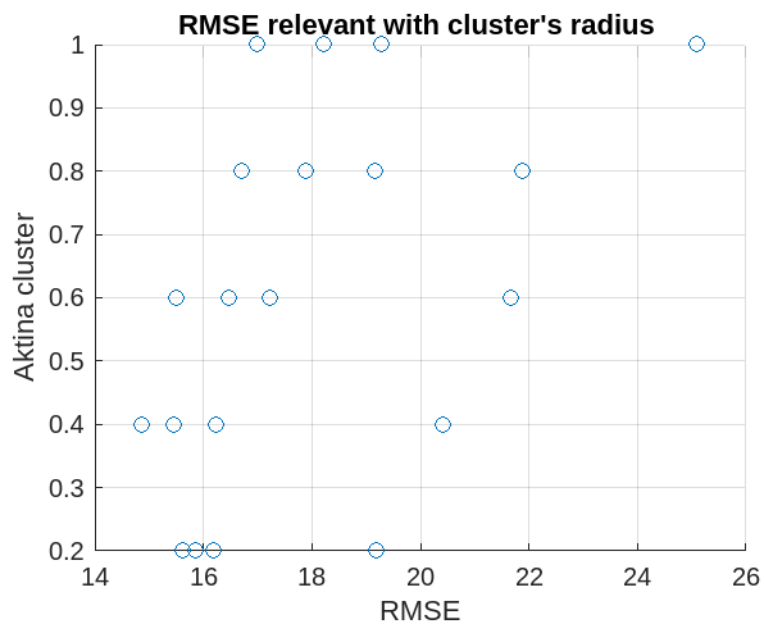
➤ R2

Kept_features\ r	0.2	0.4	0.6	0.8	1
3	0.7739	0.7636	0.7077	0.6529	0.6633
6	0.7829	0.7928	0.7648	0.7233	0.7194
9	0.7944	0.8128	0.7933	0.7621	0.7487
12	0.7981	0.8183	0.8053	0.8011	0.7818

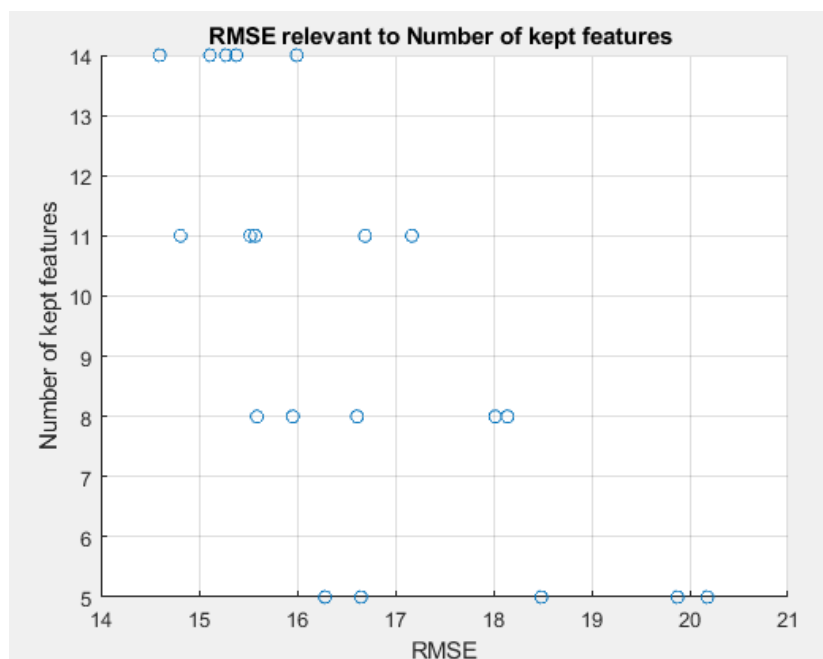
Το κάθε μοντέλο εκπαιδεύτηκε για 100 epoches και προέκυψαν οι παραπάνω μετρικές με την βοήθεια της μεθόδου του 5-fold cross validation. Καταλήγουμε στο ότι το μικρότερο RMSE σφάλμα προκύπτει για **κρατημένα features = 12** και **ακτίνα των clusters = 0.4**.

Επίσης, παραθέτονται scatter plots όπου παρουσιάζεται το σφάλμα RMSE συναρτήσει της ακτίνας r , των κρατημένων features και τον αριθμό των fuzzy rules αντίστοιχα.

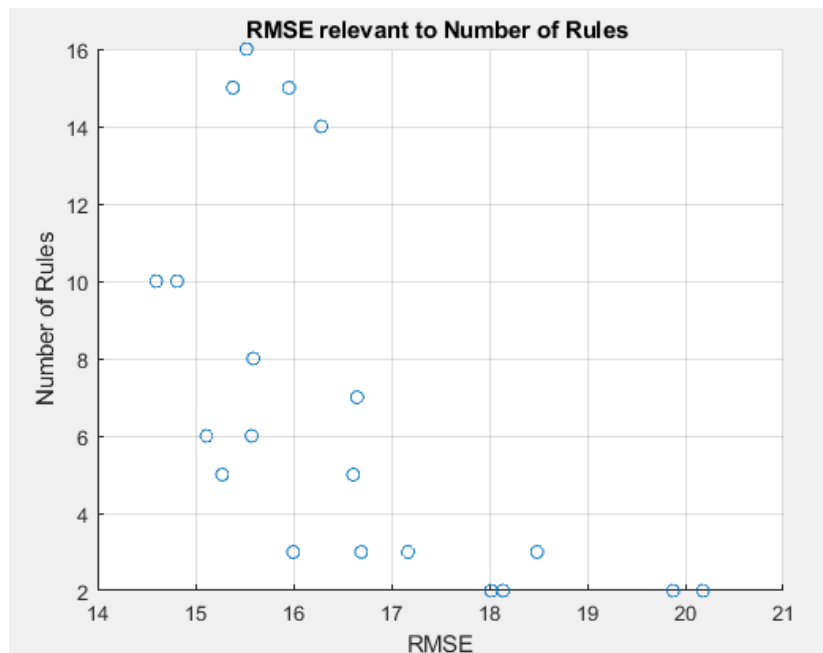
RMSE και ακτίνα r



RMSE και κρατημένα features

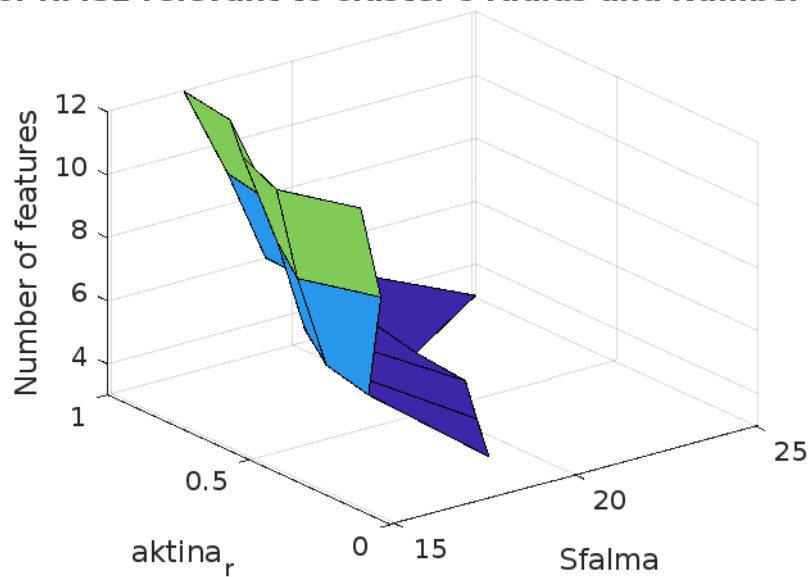


✚ RMSE και αριθμός κανόνων



✚ Surf plot για τον οπτικό εντοπισμό της ελαχιστοποίησης του RMSE

of RMSE relevant to cluster's radius and Number of



Τελικά συμπεράσματα

Το μικρότερο RMSE σφάλμα προκύπτει για **κρατημένα features = 12** και **ακτίνα των clusters = 0.4**, όπως ήδη αναφέραμε και δείξαμε. Για σταθερό αριθμό κρατημένων features, βλέπουμε πως το RMSE αυξάνεται όσο αυξάνεται η ακτίνα r των clusters. Από την άλλη, κρατώντας σταθερή την ακτίνα, παρατηρούμε ότι αυξάνοντας τον αριθμό των κρατημένων features, τόσο μειώνεται το RMSE. Αυτό συμβαίνει, διότι, όσο περισσότερα features κρατάμε, τόσο λιγότερη πληροφορία χάνεται από το dataset. Δεν μπορούμε, όμως, να αυξάνουμε συνεχώς την παράμετρο αυτή, καθώς έτσι ο χρόνος εκπαίδευσης αυξάνεται εκθετικά. Επιπλέον, μικρότερη ακτίνα r των clusters σημαίνει πως το μοντέλο μαθαίνει καλύτερα το dataset, με αποτέλεσμα να πραγματοποιεί καλύτερες προβλέψεις. Η συνεχής όμως μείωση της ακτίνας οδηγεί το μοντέλο σε overfitting, με απότοκο της εξάλειψη της γενικότητας. Γι' αυτό και το RMSE είναι καλύτερο για $r = 0.4$ και όχι για $r = 0.2$ με αριθμό κρατημένων features = 12.

Βέλτιστο Μοντέλο TSK με kept_features = 12 και radius = 0.4

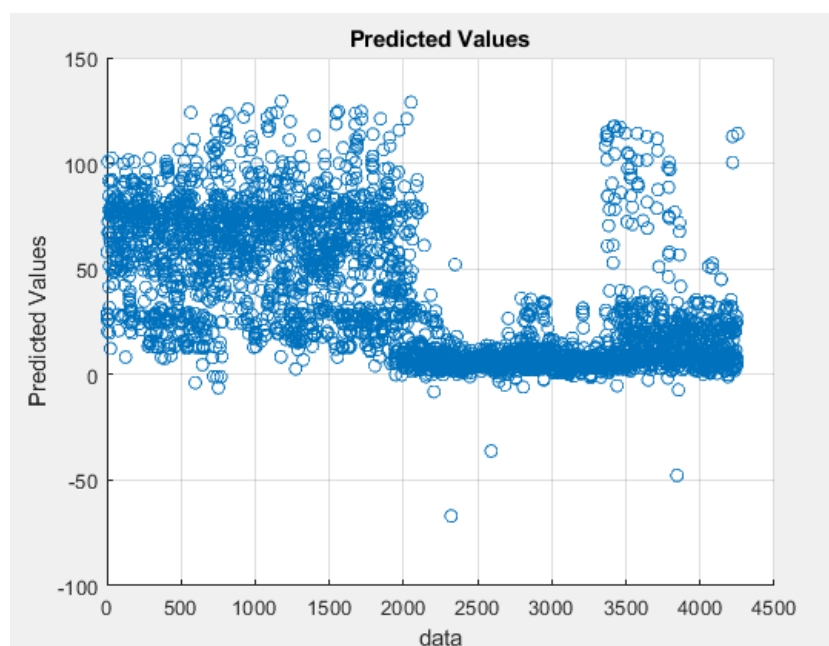
Η εκπαίδευση του μοντέλου αυτού πραγματοποιήθηκε για 120 epoches αντί για 100, με απώτερο σκοπό την καλύτερη του απόδοση μετά το training.

Ακολουθούν τα αποτελέσματα παρακάτω:

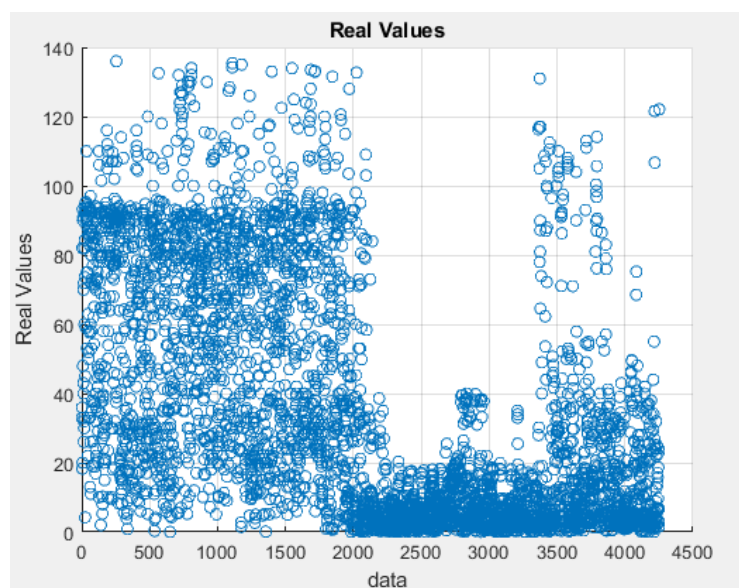
- Πίνακας δεικτών απόδοσης

	R2	RMSE	NMSE	NDEI
Βέλτιστο Μοντέλο TSK	0.8165	14.6706	0.1835	0.4283

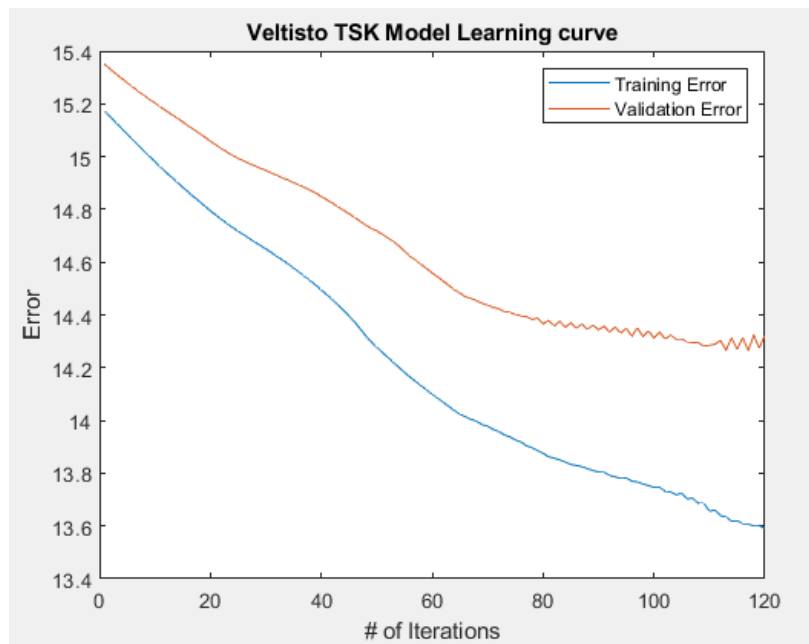
- Προβλέψεις τελικού μοντέλου



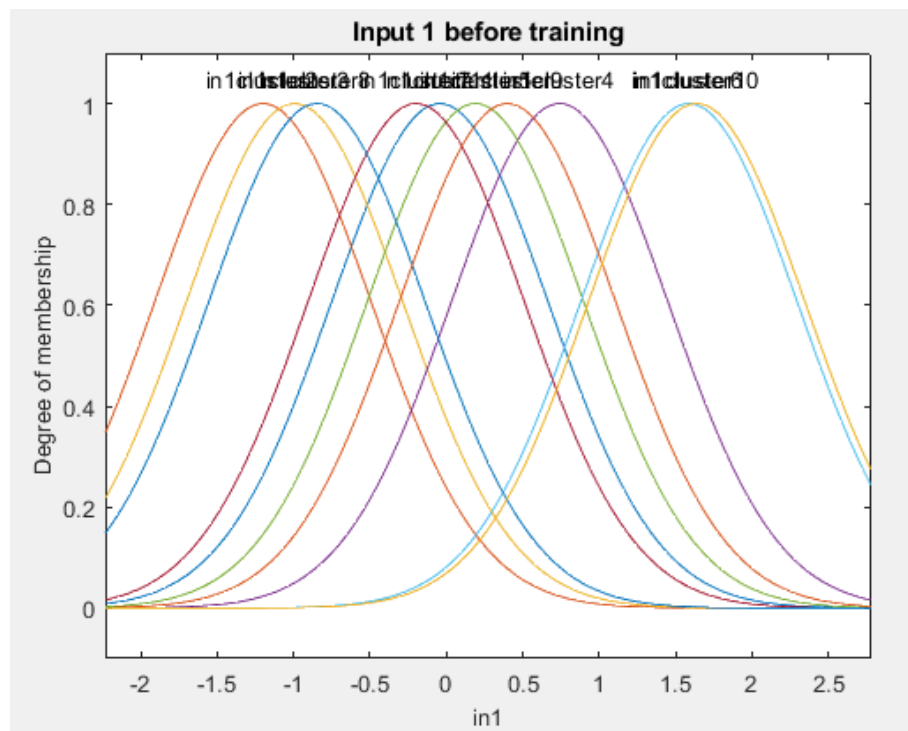
- Πραγματικές τιμές Target του dataset

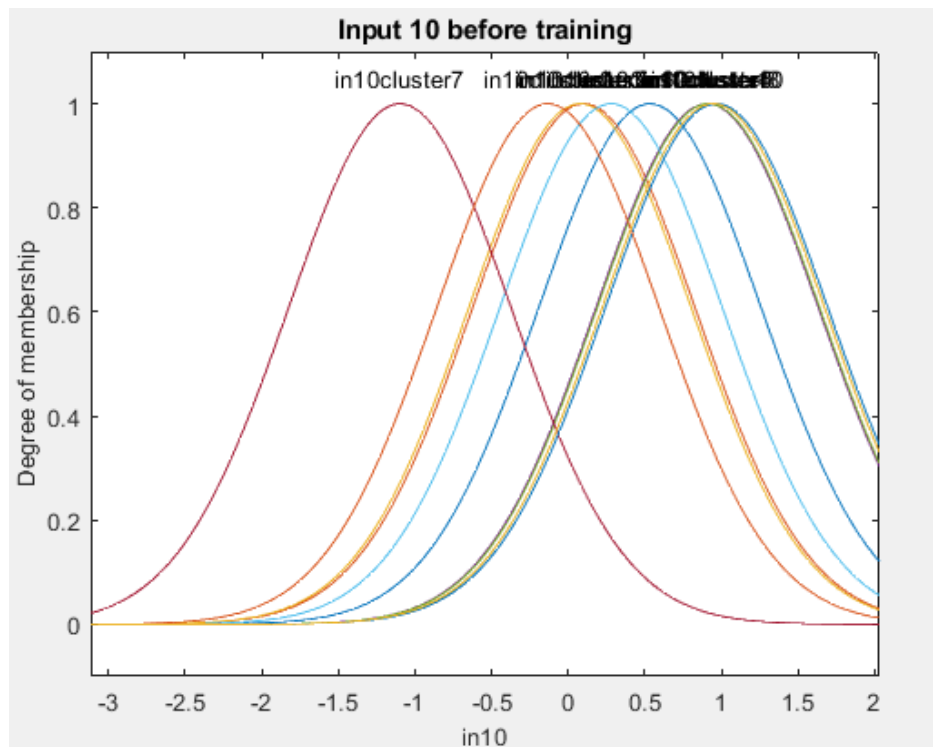


- Learning curve

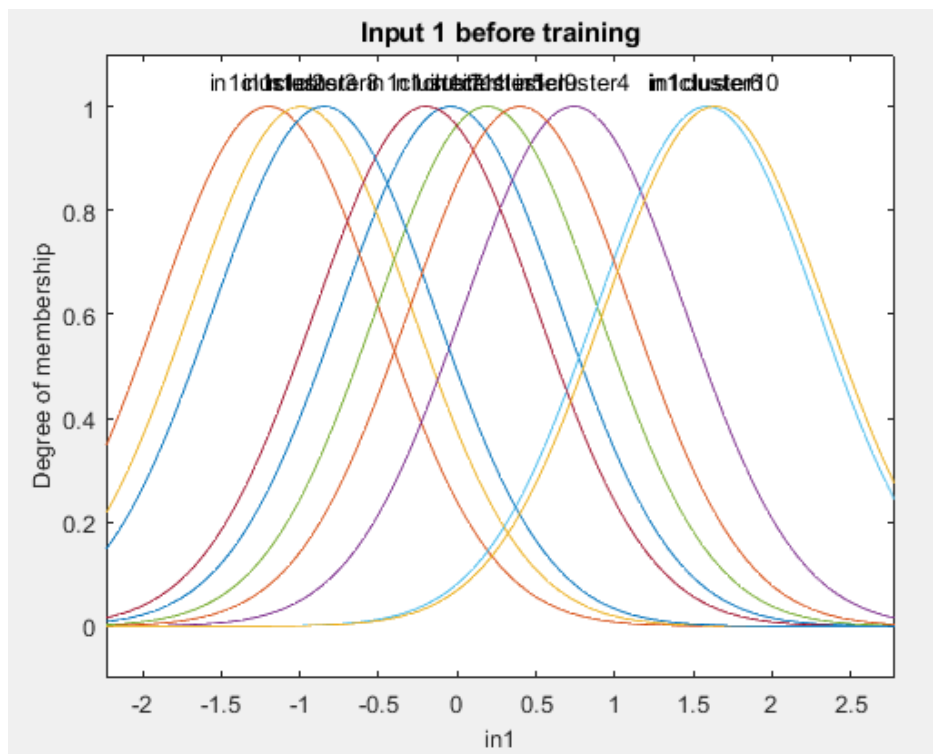


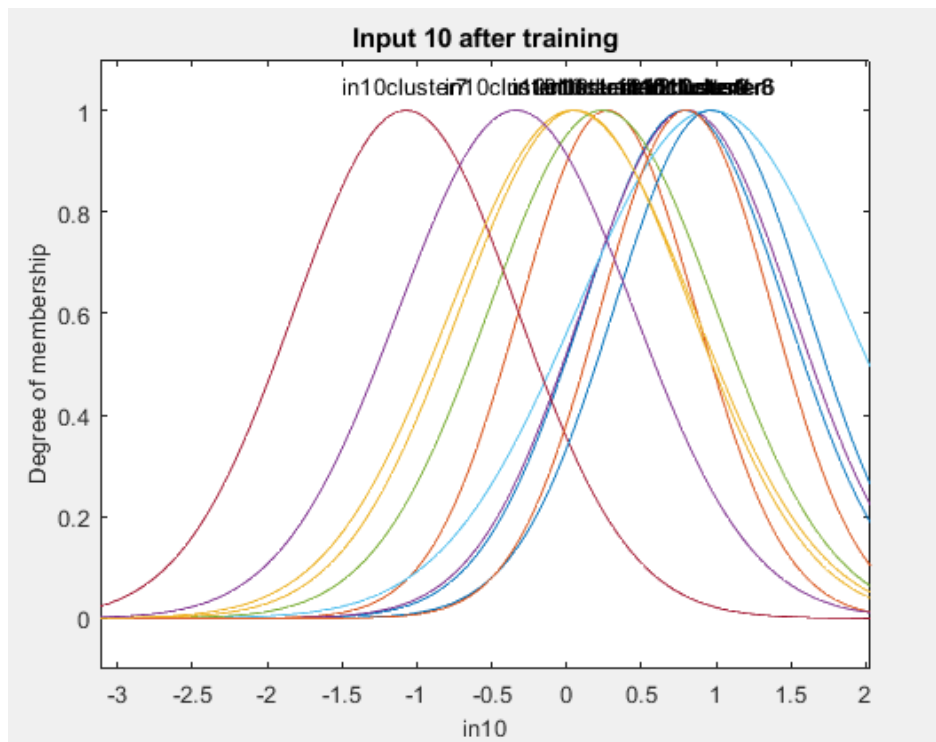
- Ασαφή σύνολα στην αρχική τους μορφή



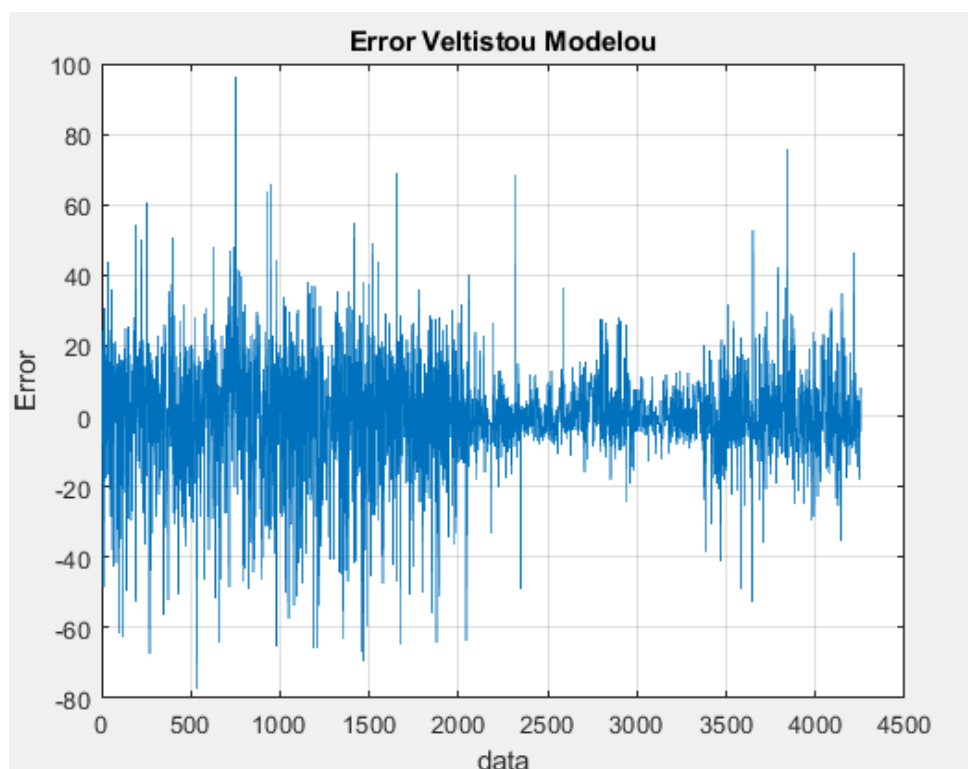


- Ασαφή σύνολα στην τελική τους μορφή





- Σφάλμα προβλέψεων στα testing data



Τελικά συμπεράσματα για το βέλτιστο μοντέλο

Βλέποντας τα διαγράμματα των Predicted και Real Values, συμπεραίνουμε ότι το μοντέλο προβλέπει αρκετά καλά το testing dataset. Επίσης, το πλήθος των fuzzy rules του βέλτιστου μοντέλου είναι **11**. Για τον ίδιο αριθμό κρατημένων χαρακτηριστικών με την μέθοδο του *grid partitioning*, αν για κάθε είσοδο είχαμε 2 ή 3 ασαφή σύνολα, θα είχαμε συνολικά 2^{14} ή 3^{14} κανόνες αντίστοιχα. Καταδεικνύεται έτσι το σημαντικό πλεονέκτημα επιφέρει η χρήση του *Subtractive clustering* έναντι του grid partitioning.