

Υπολογιστική Νοημοσύνη

**Εργασία Αερινού Εξαμήνου
2019-2020**

**3^η Εργασία
Regression**

ΝΙΚΗΦΟΡΙΔΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ 9084

nikifori@ece.auth.gr

Θεσσαλονίκη, Ιανουάριος 2021

Αναφορά

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

Παρακάτω φαίνονται τα μοντέλα τα οποία καλούμαστε να υλοποιήσουμε.

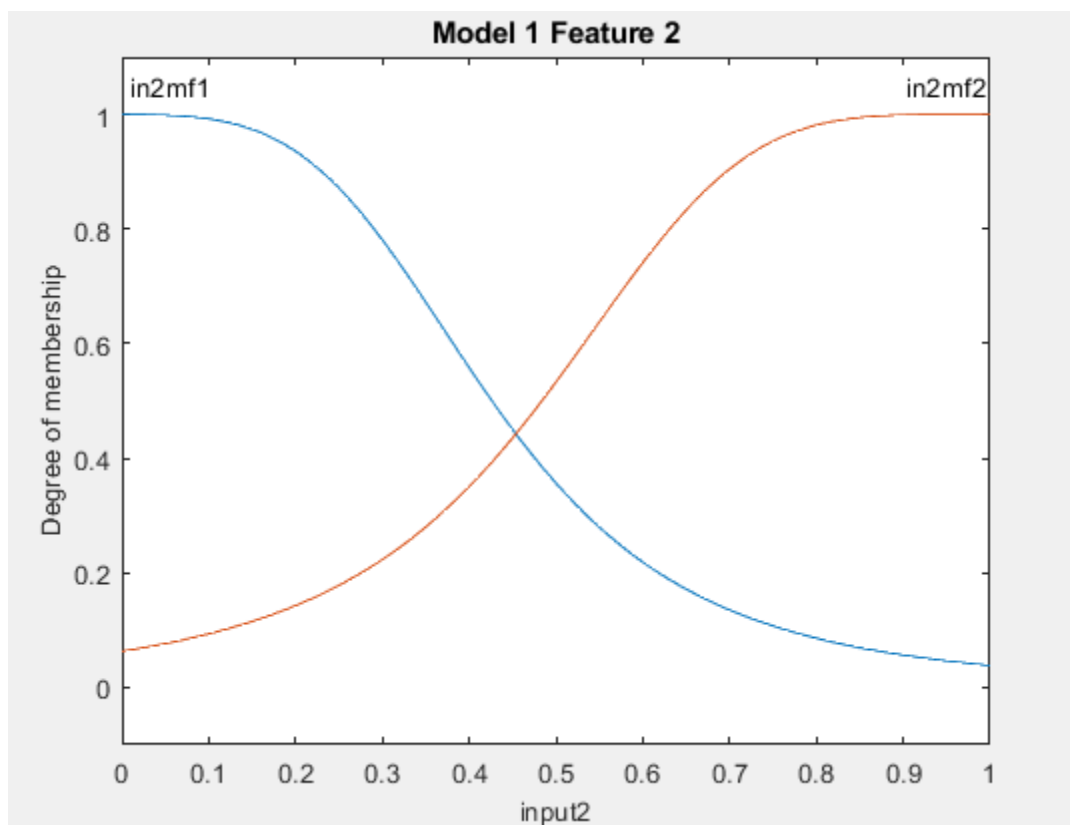
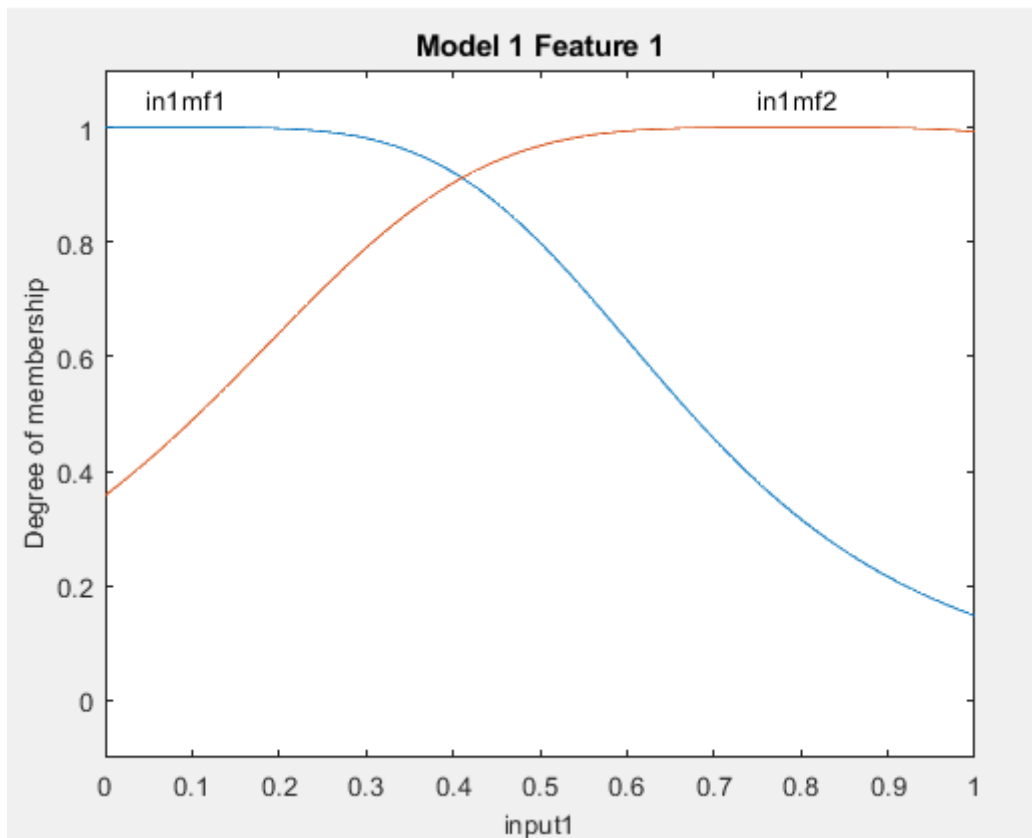
Πλήθος συναρτήσεων συμμετοχής		Μορφή εξόδου
TSK_model_1	2	Singleton
TSK_model_2	3	Singleton
TSK_model_3	2	Polynomial
TSK_model_4	3	Polynomial

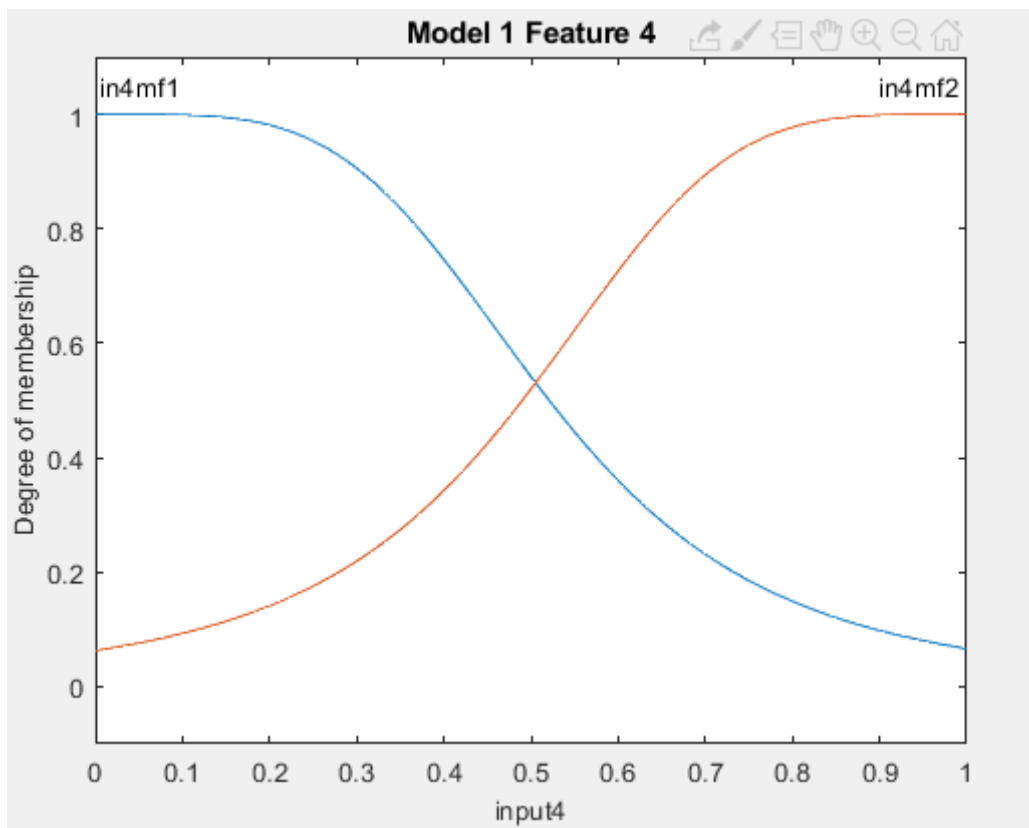
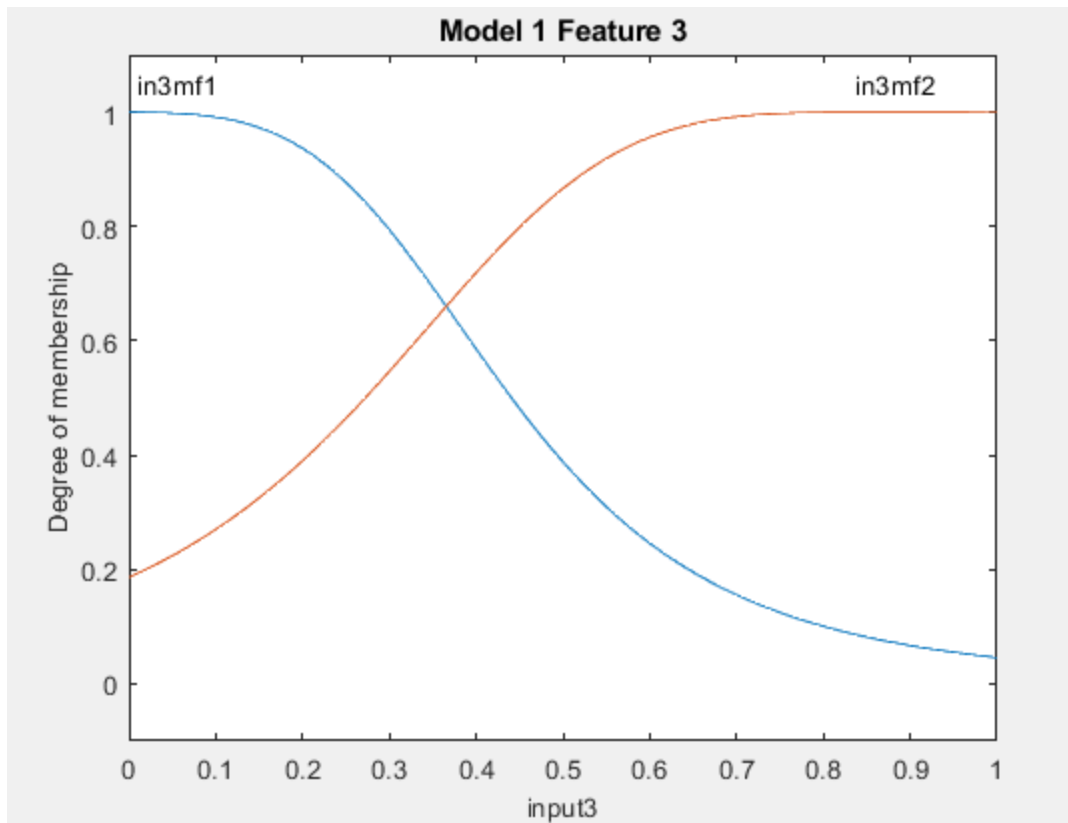
Ως dataset χρησιμοποιούμε από το UCI repository το Airfoil Self-Noise dataset το οποίο περιλαμβάνει 1503 δείγματα και 6 features. Η 6^η στήλη είναι και η εξαρτημένη στήλη ή αλλιώς, Target feature. Στη συνέχεια, για κάθε ένα από τα παραπάνω μοντέλα υπολογίζουμε τα R2, RMSE, NMSE και NDEI. Μαζί με τους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

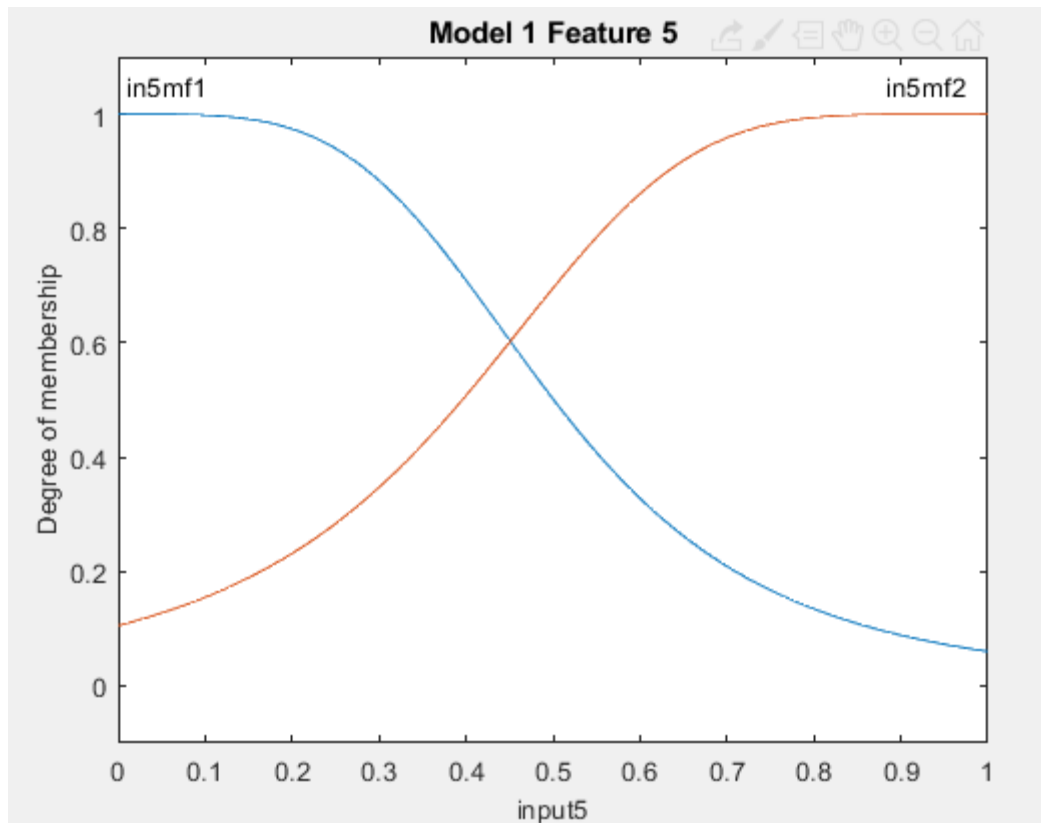
- Διάγραμμα με τις τελικές μορφές των ασαφών συνόλων
- Διάγραμμα με το learning curve
- Διάγραμμα με τα σφάλματα πρόβλεψης

1^ο Μοντέλο TSK

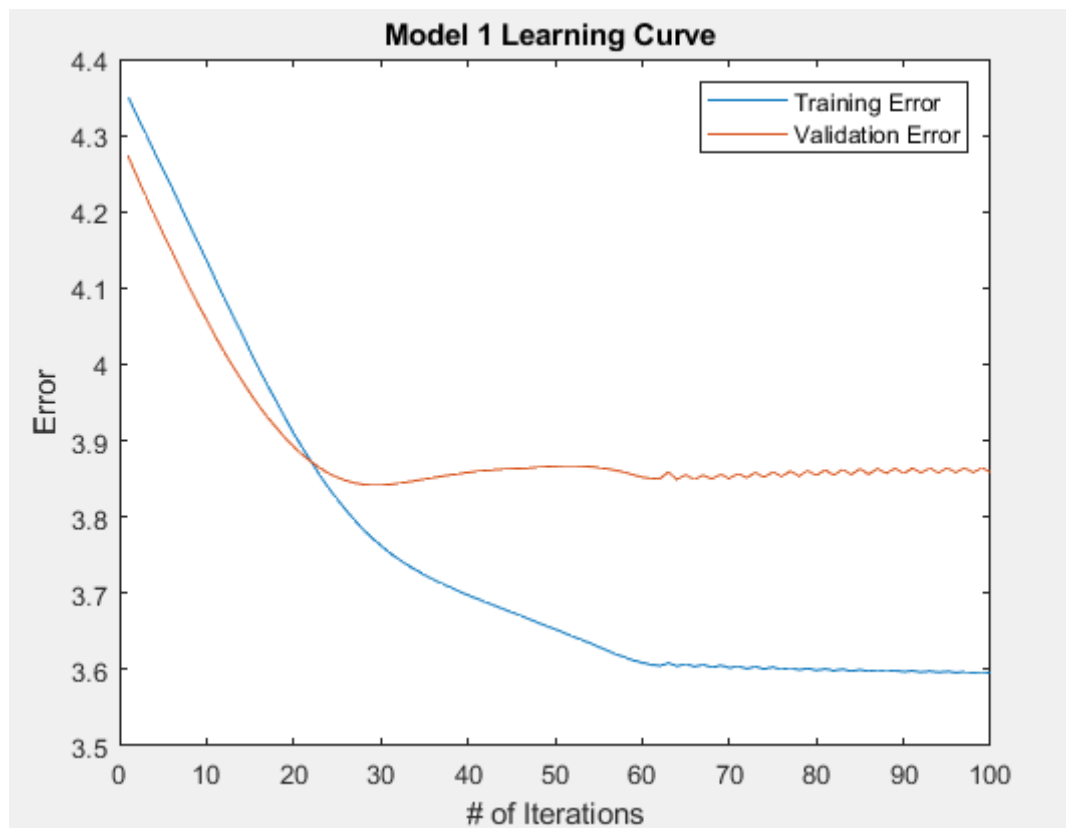
- Membership functions



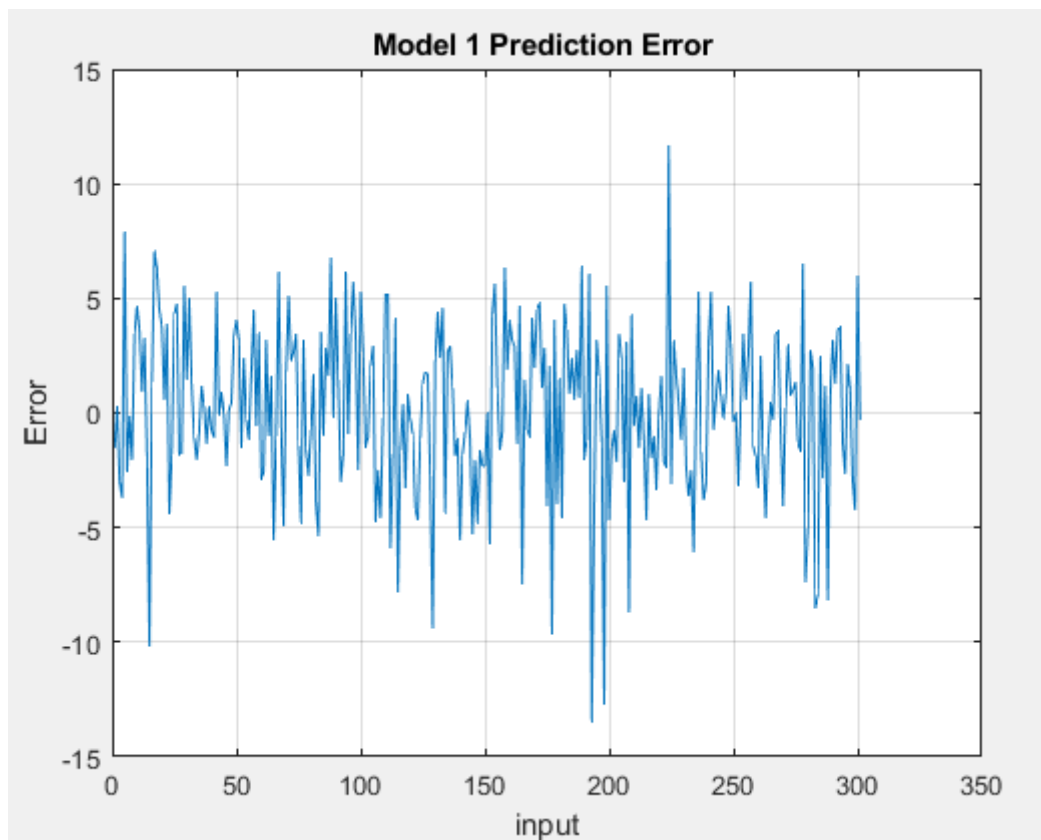




- Learning curve

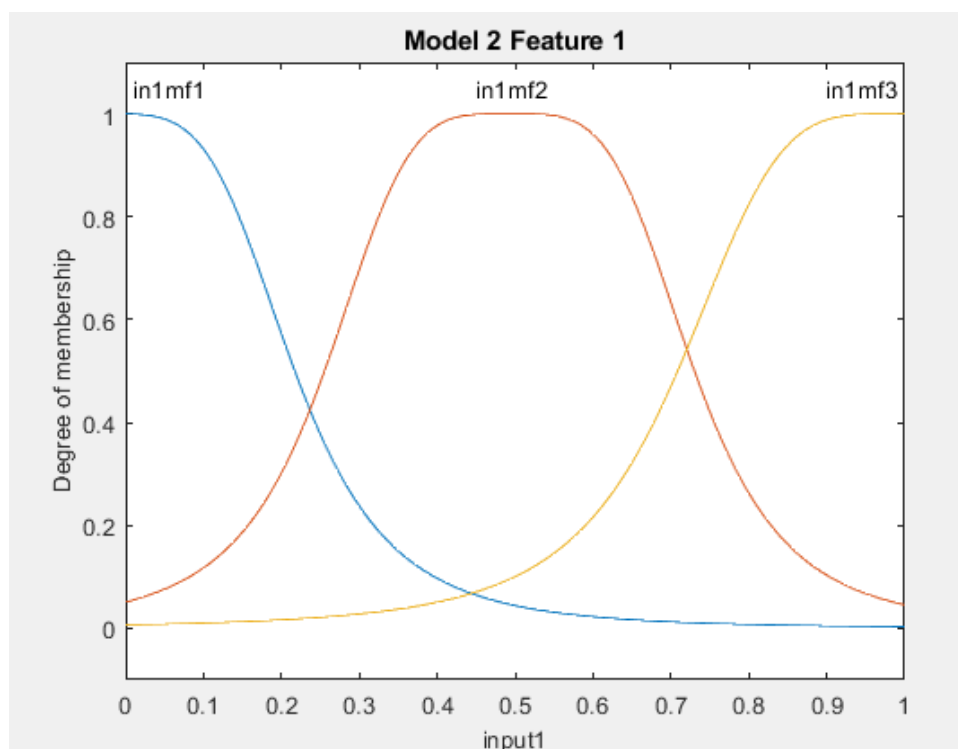


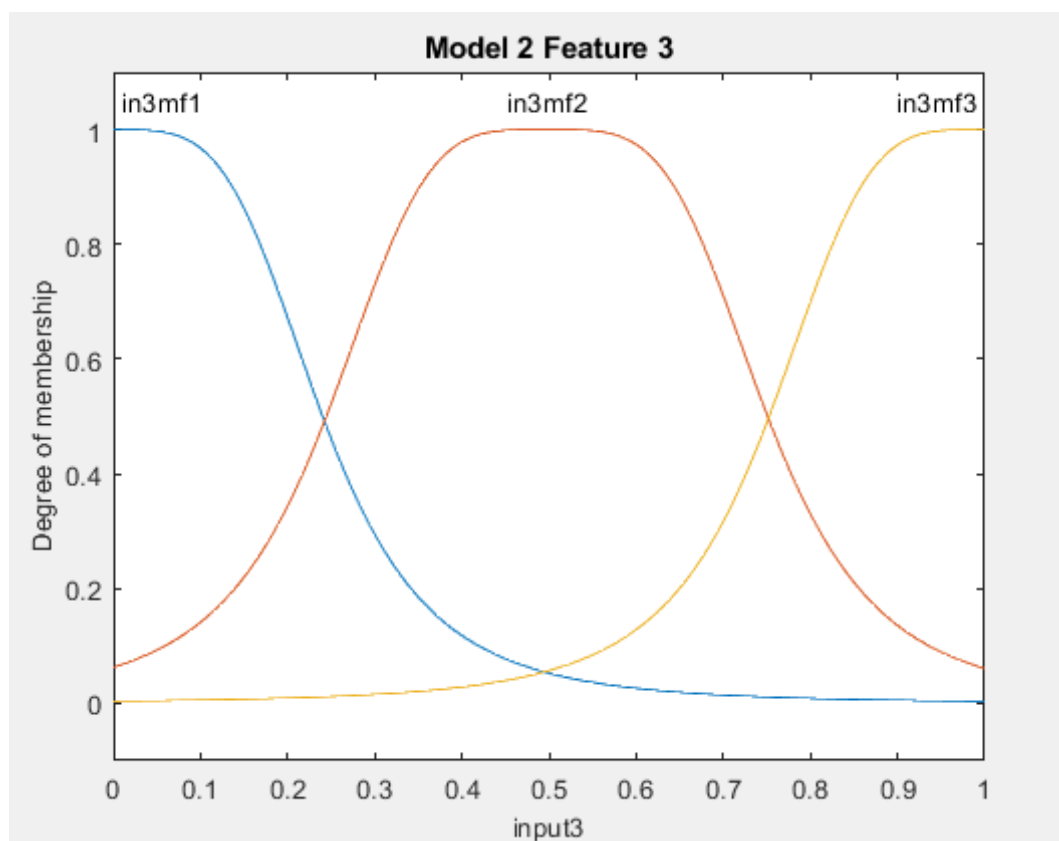
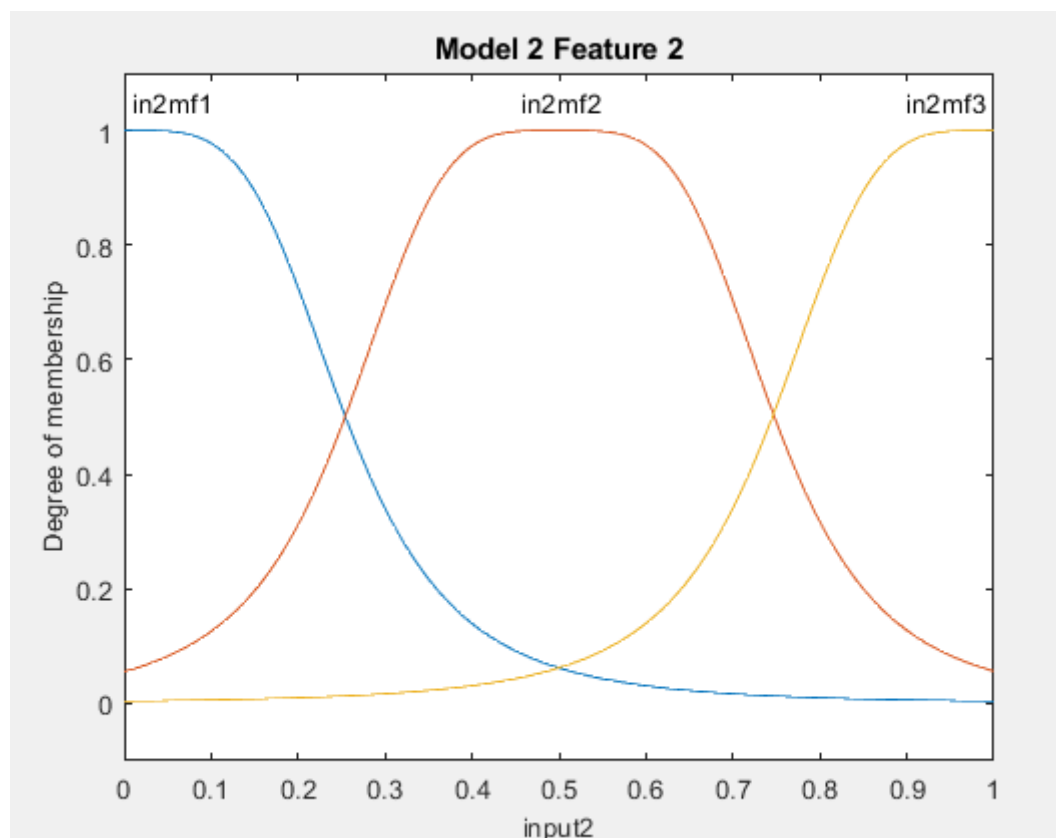
- Prediction error

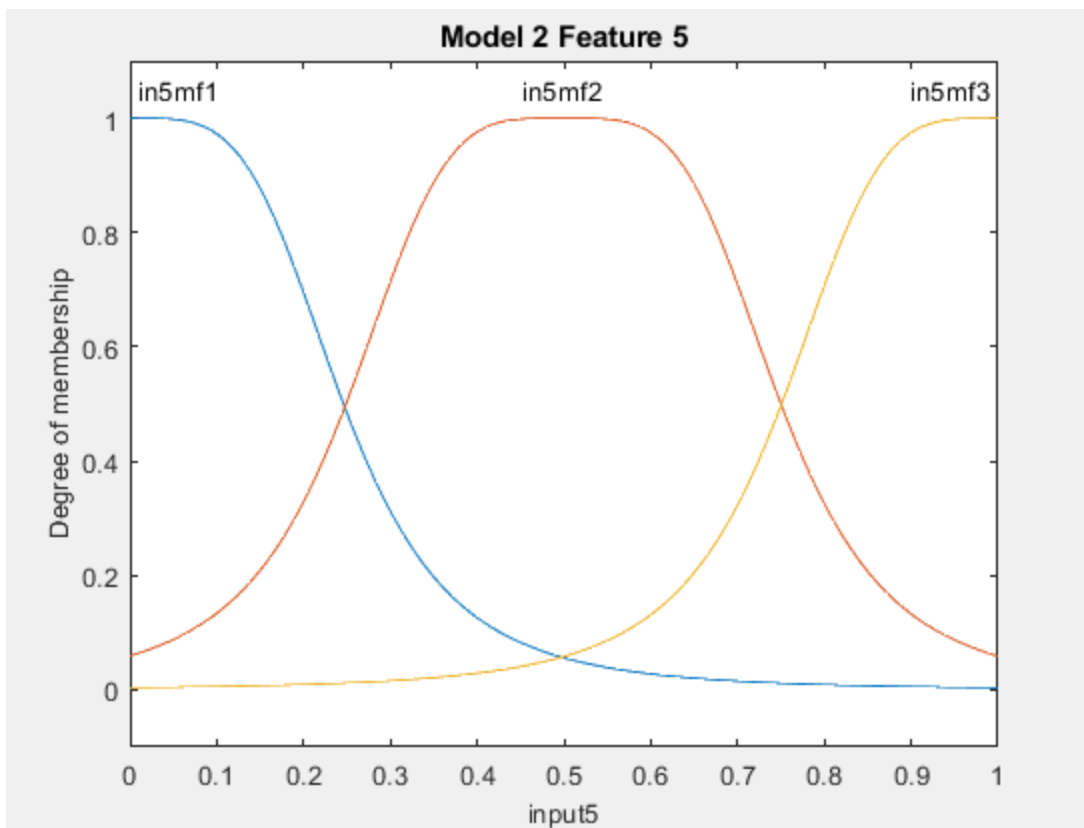
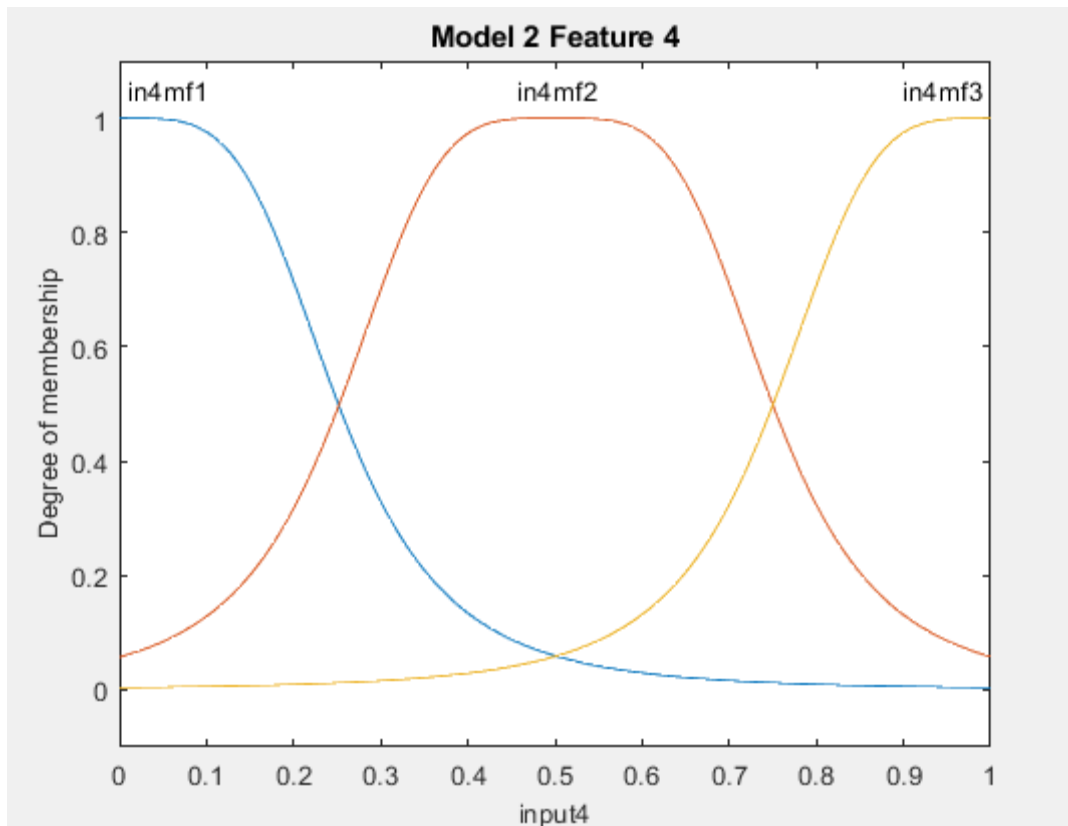


2° Μοντέλο TSK

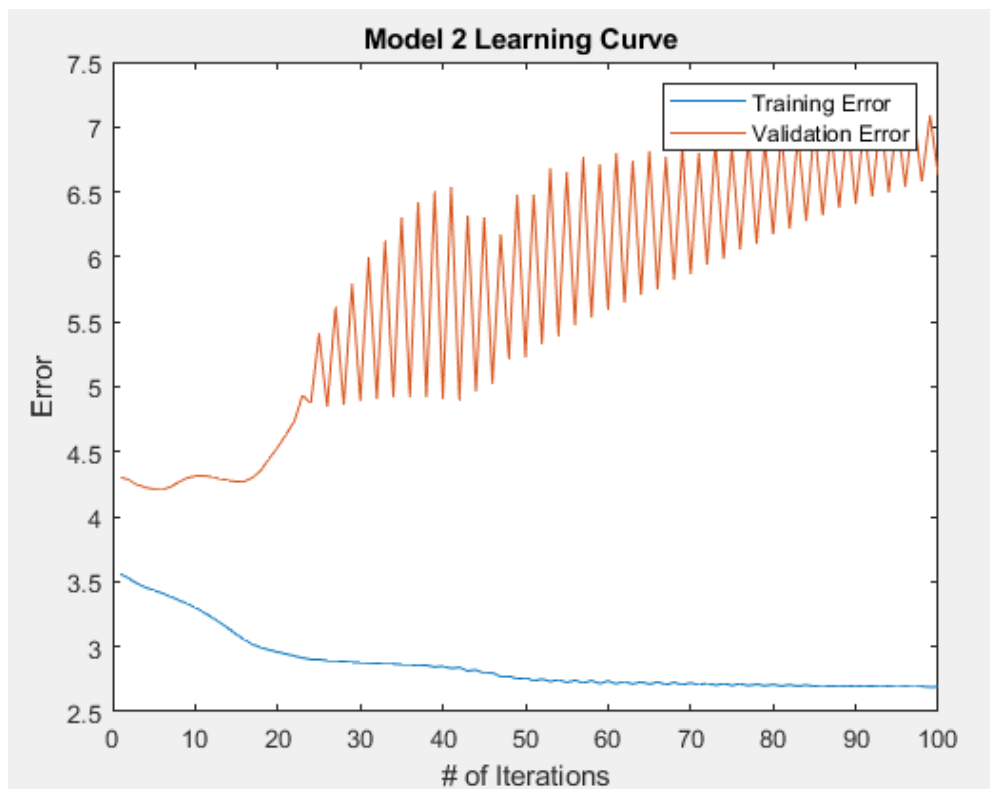
- Membership functions



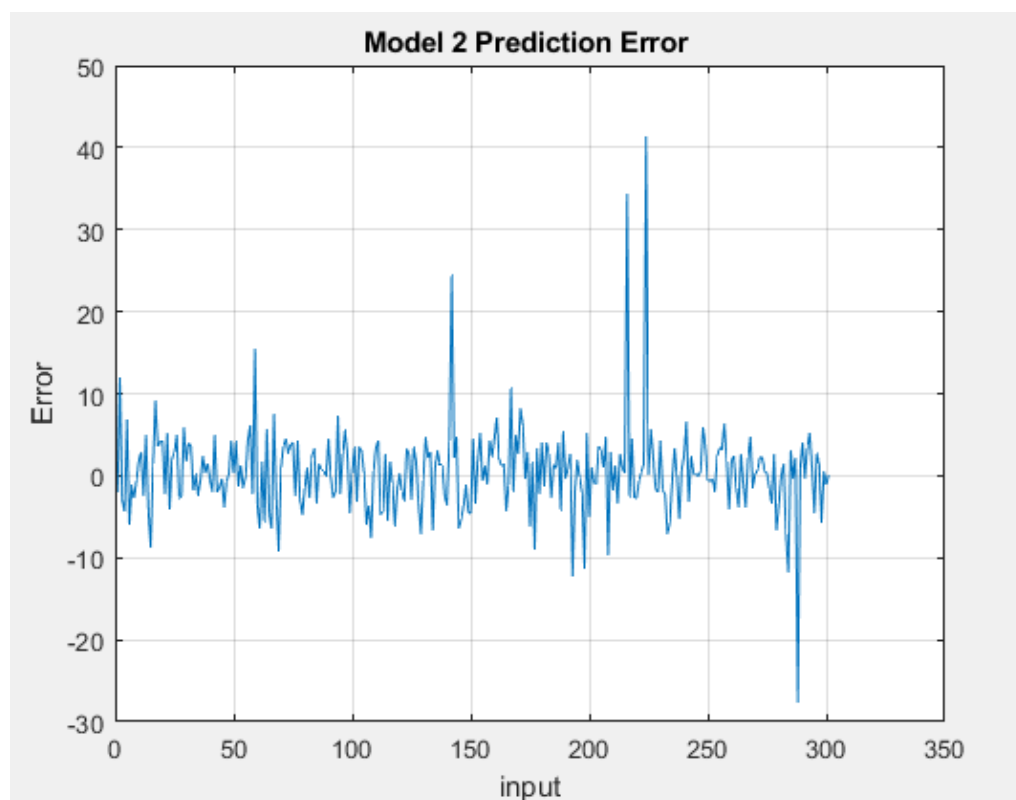




- Learning curve

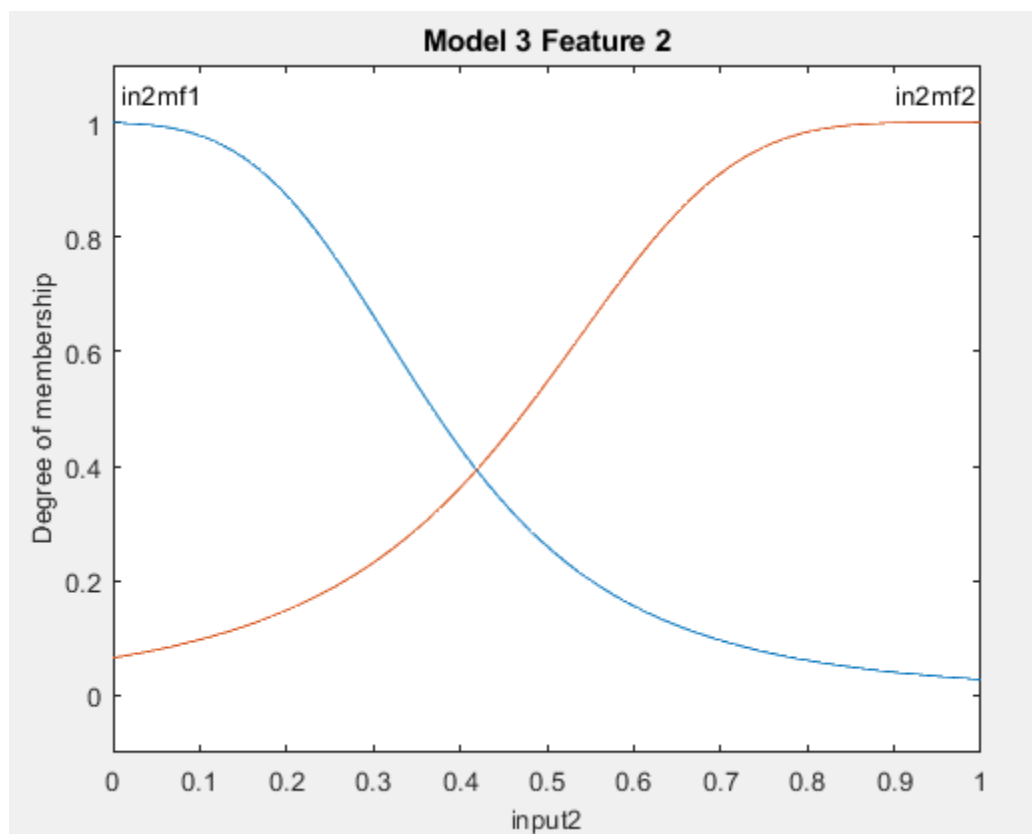
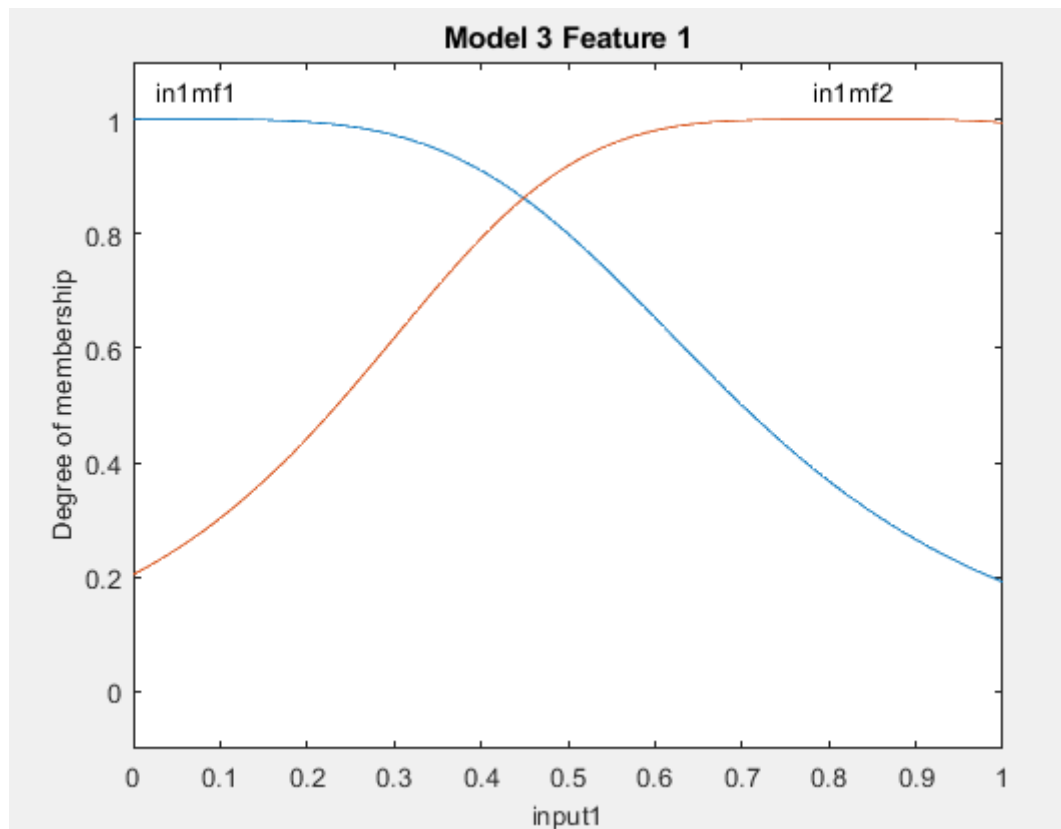


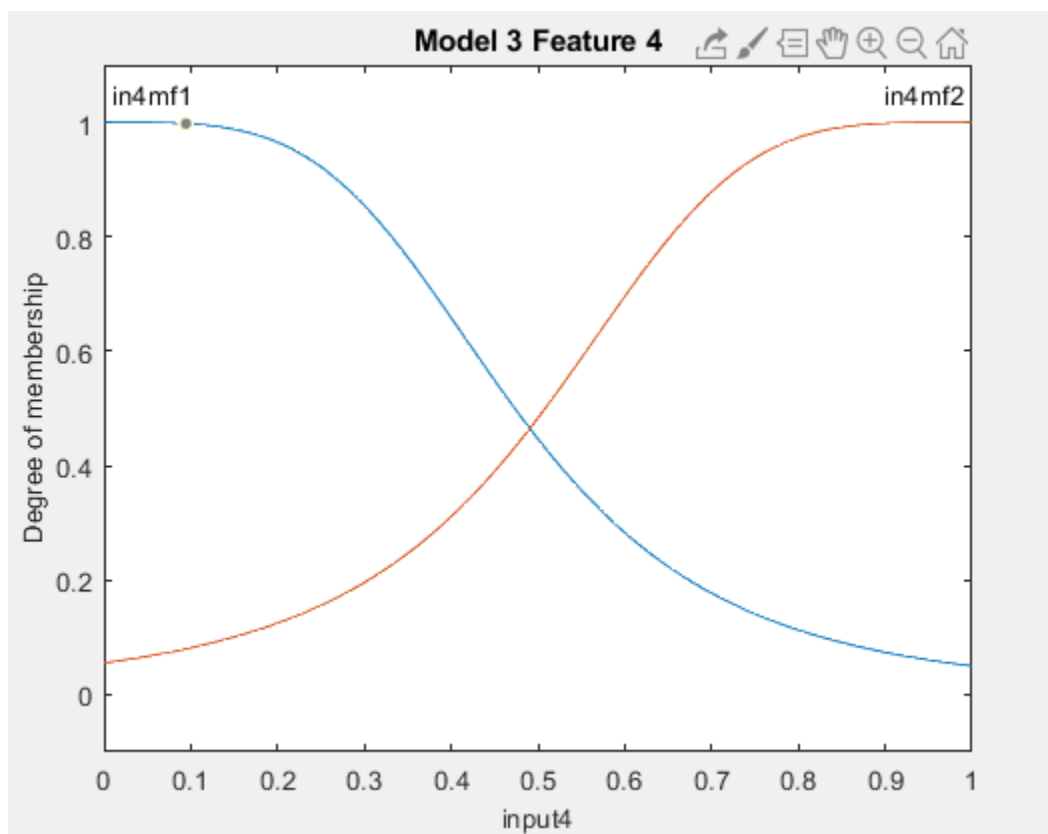
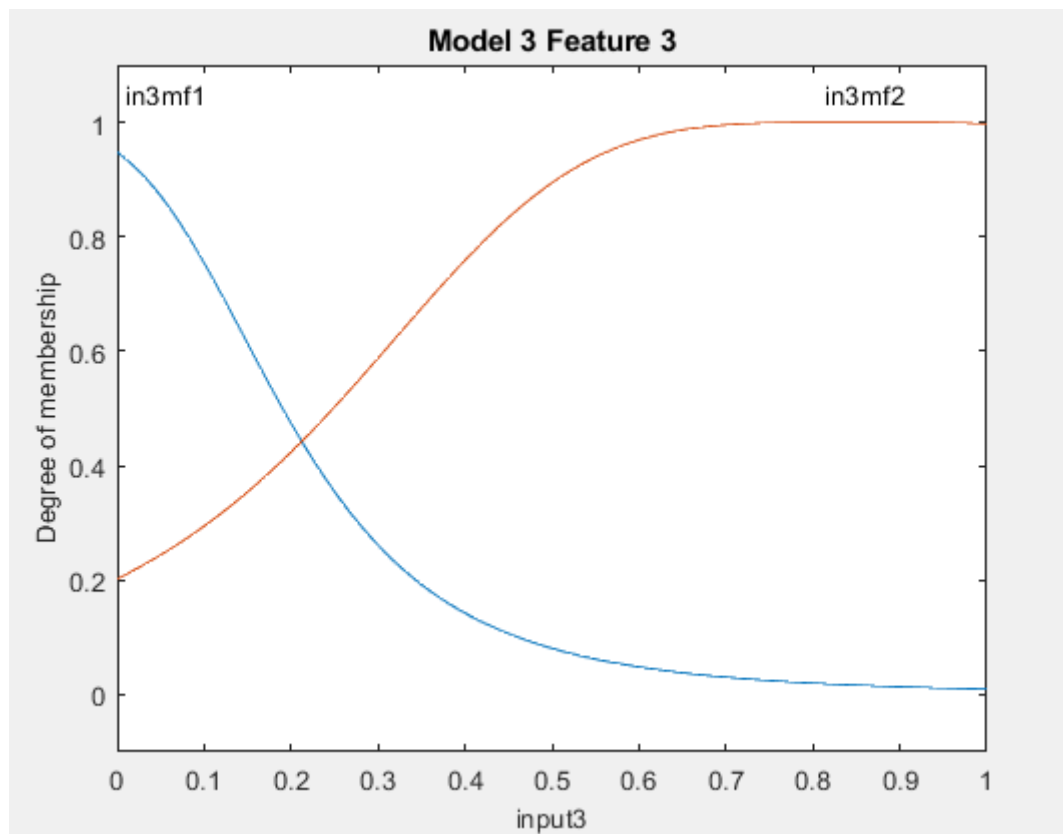
- Prediction error

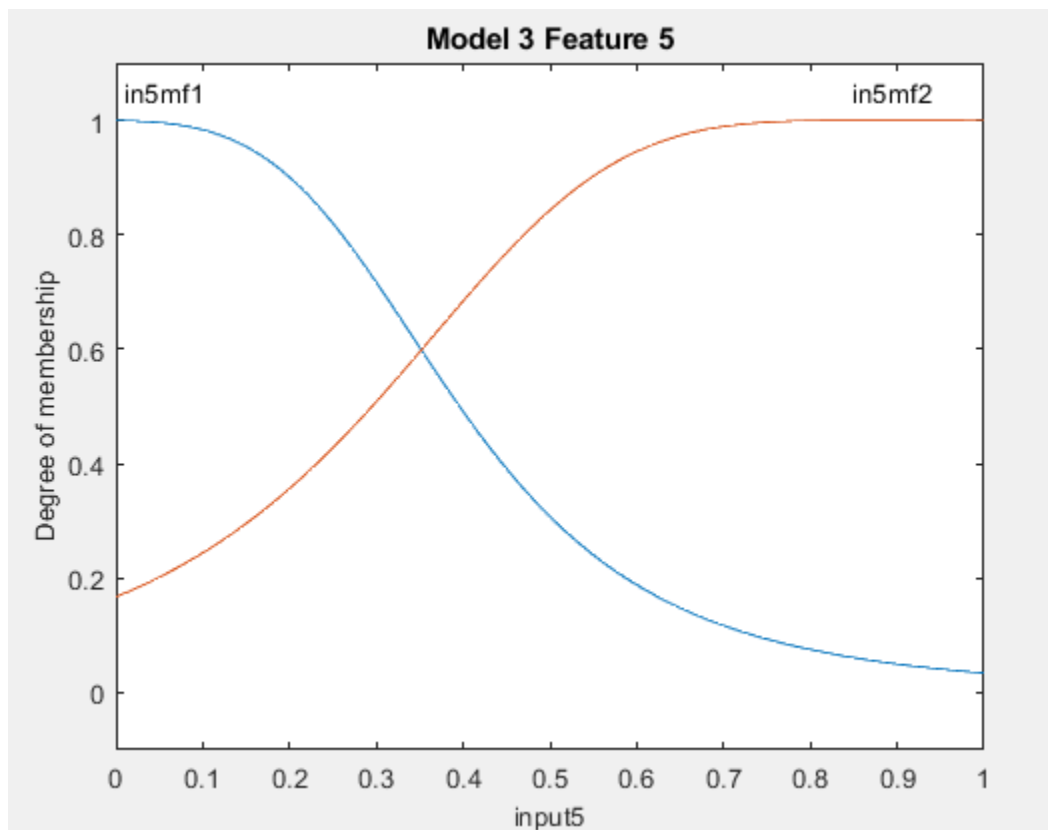


3^ο Μοντέλο TSK

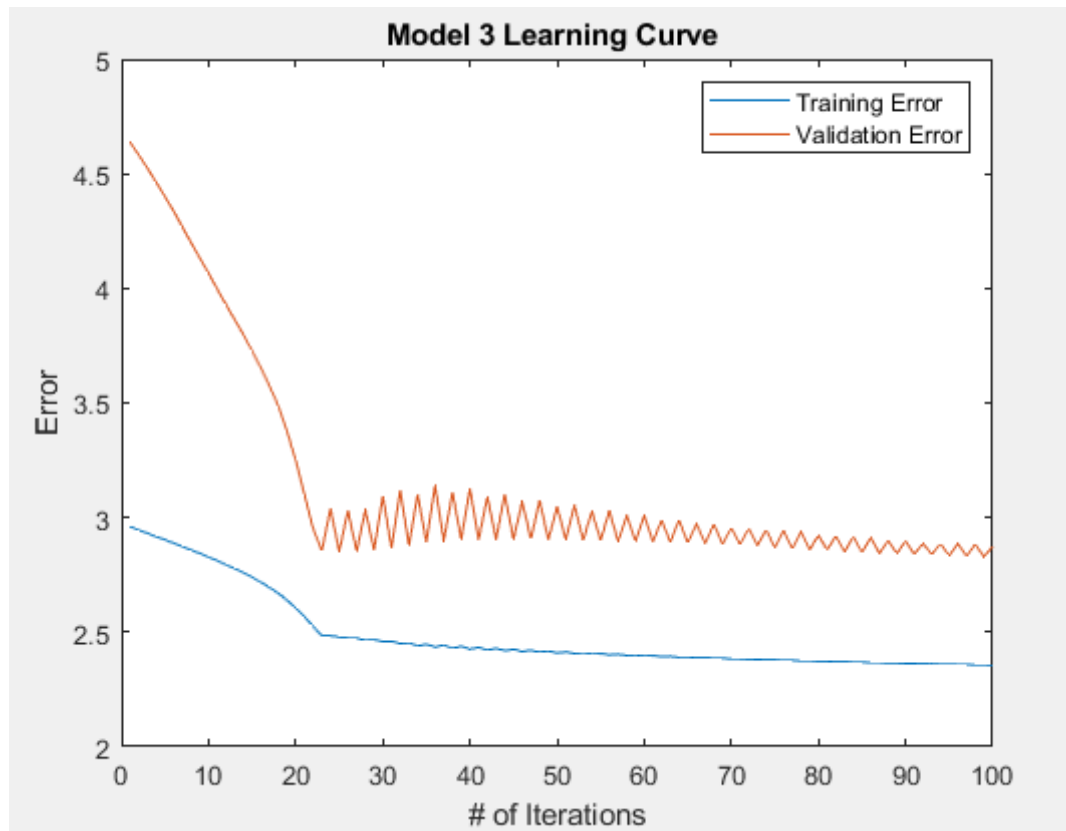
- Membership functions



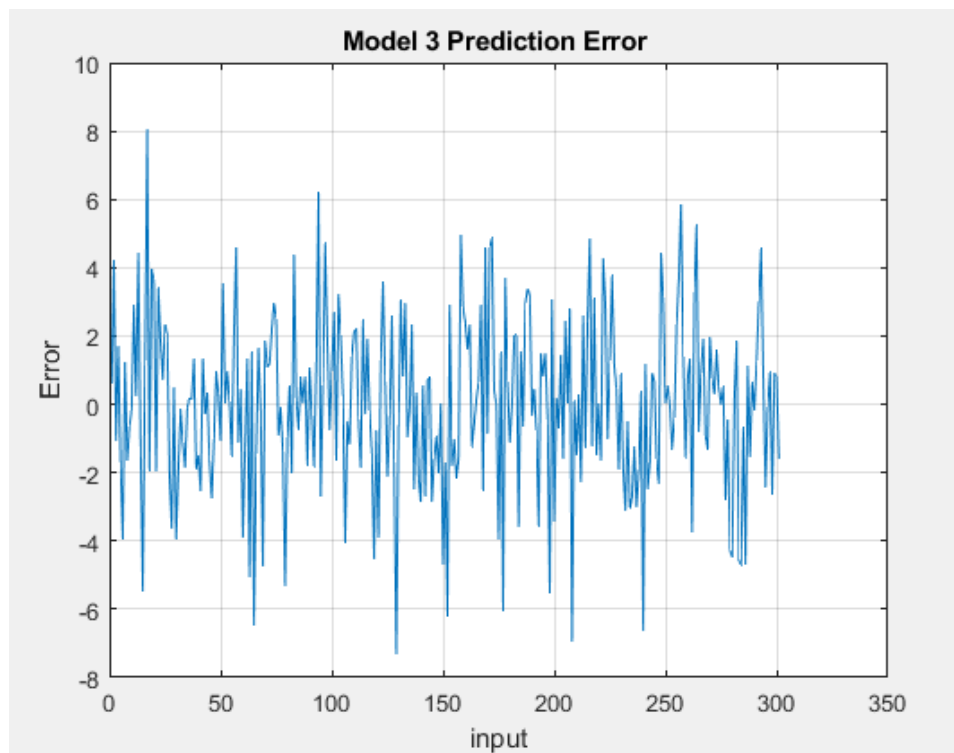




- Learning curve

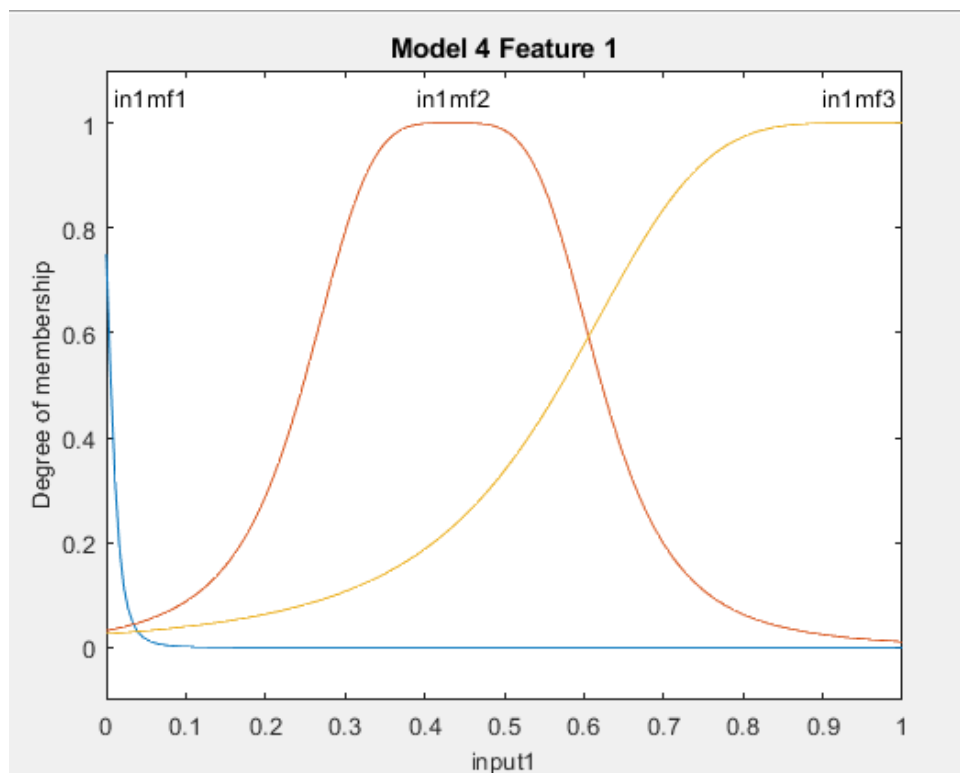


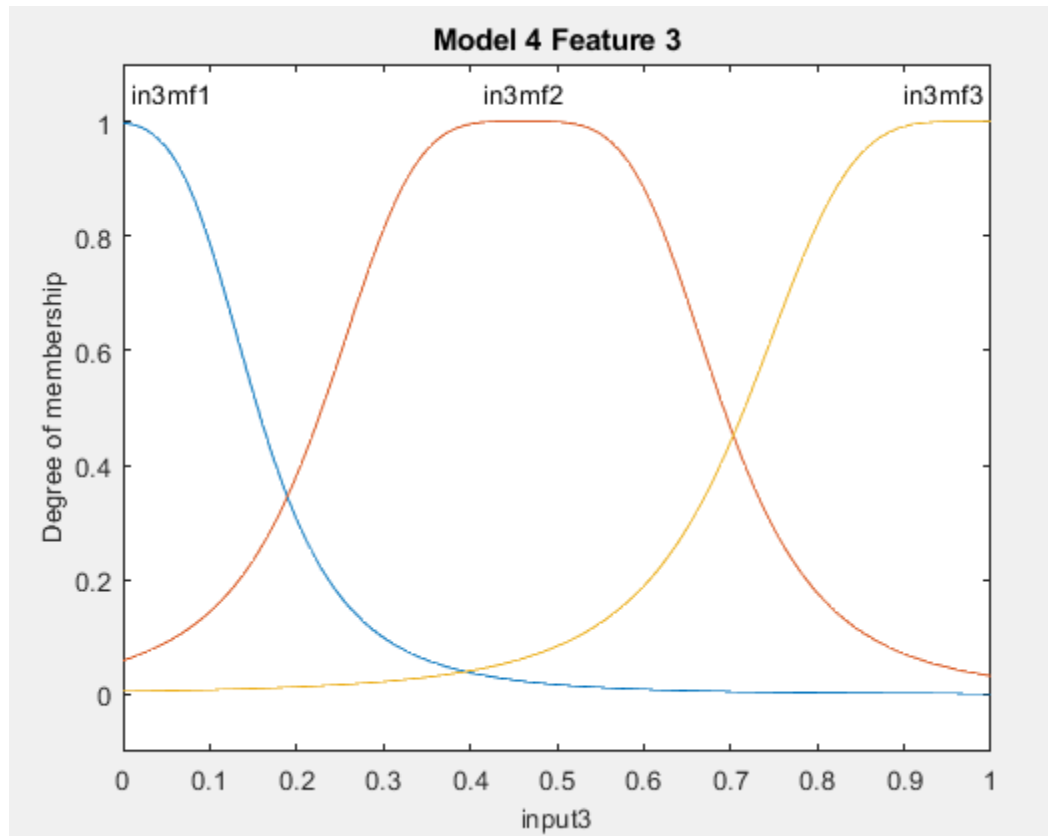
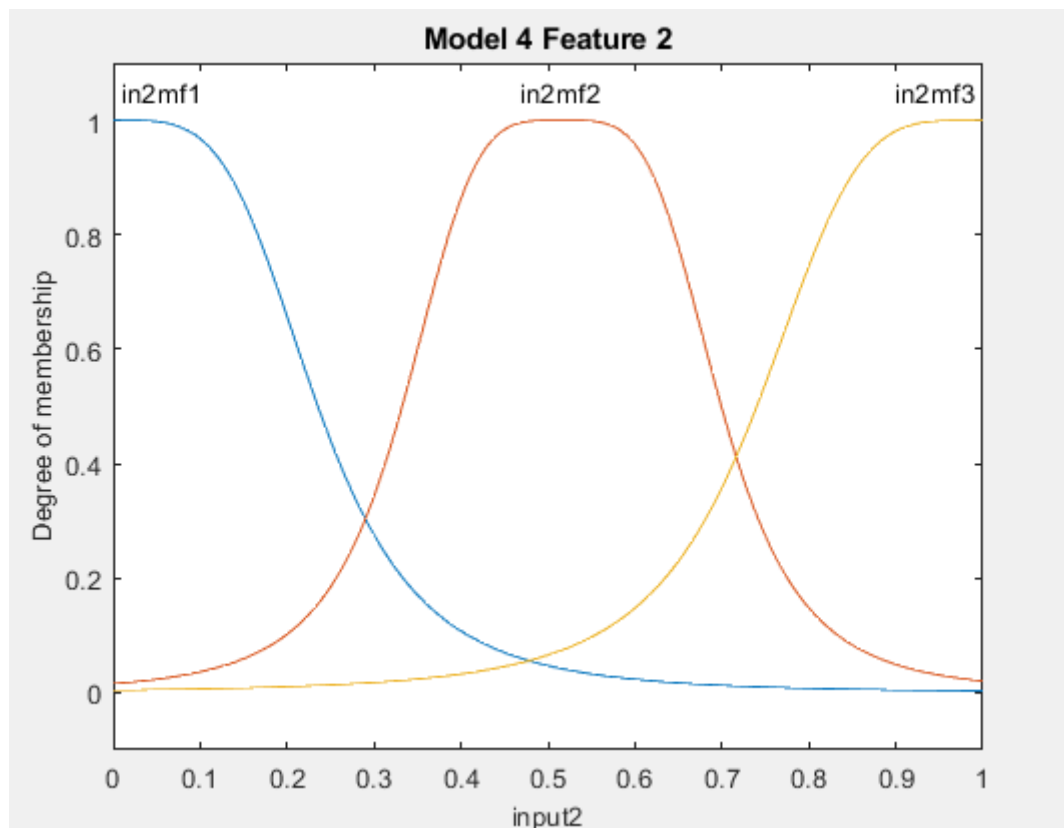
- Prediction error

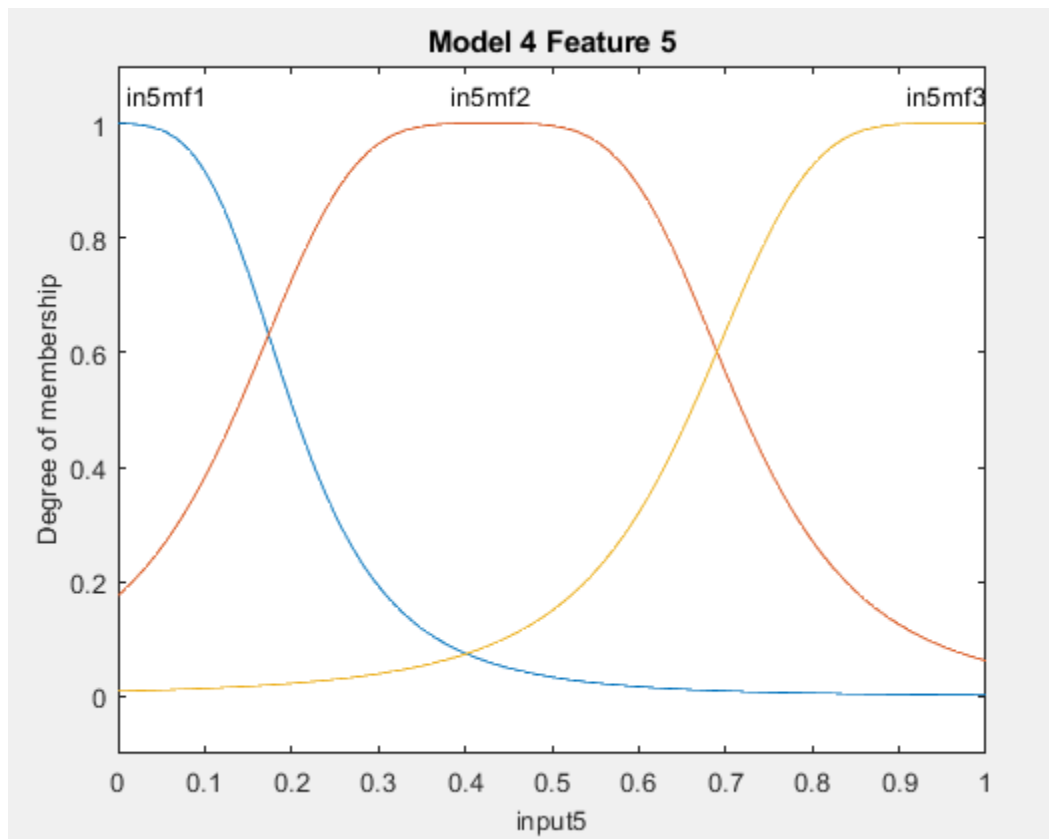
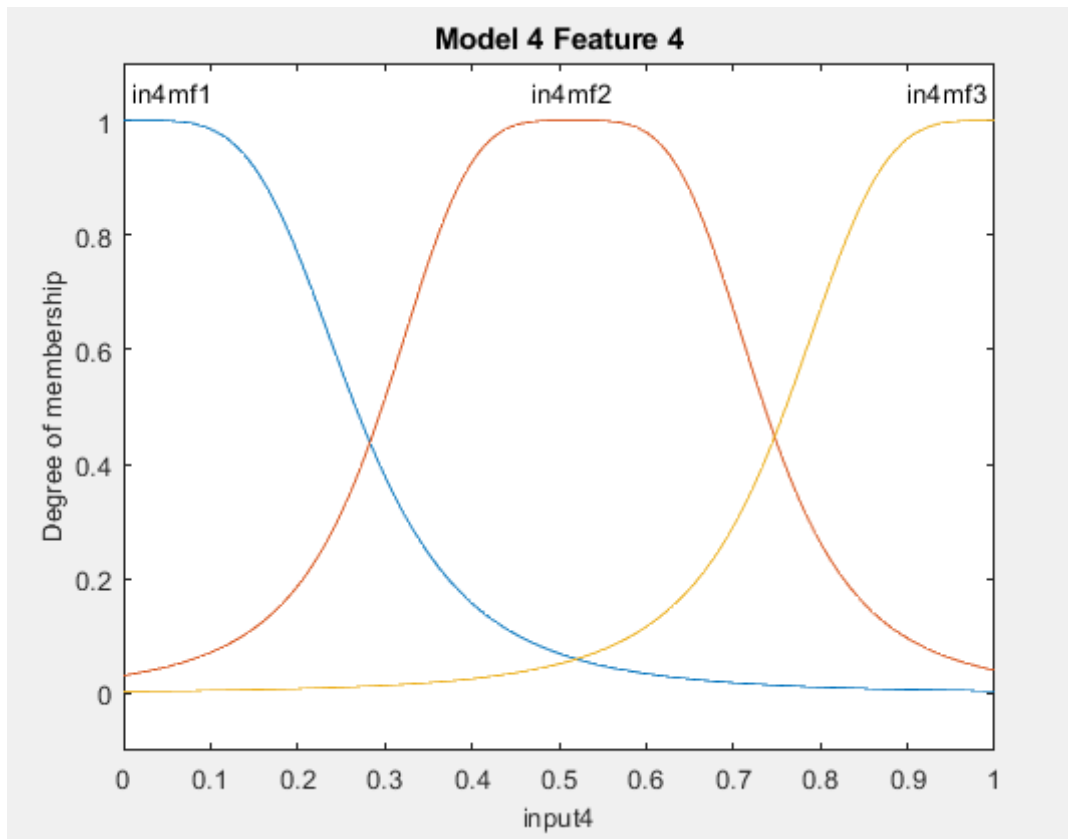


4^ο Μοντέλο TSK

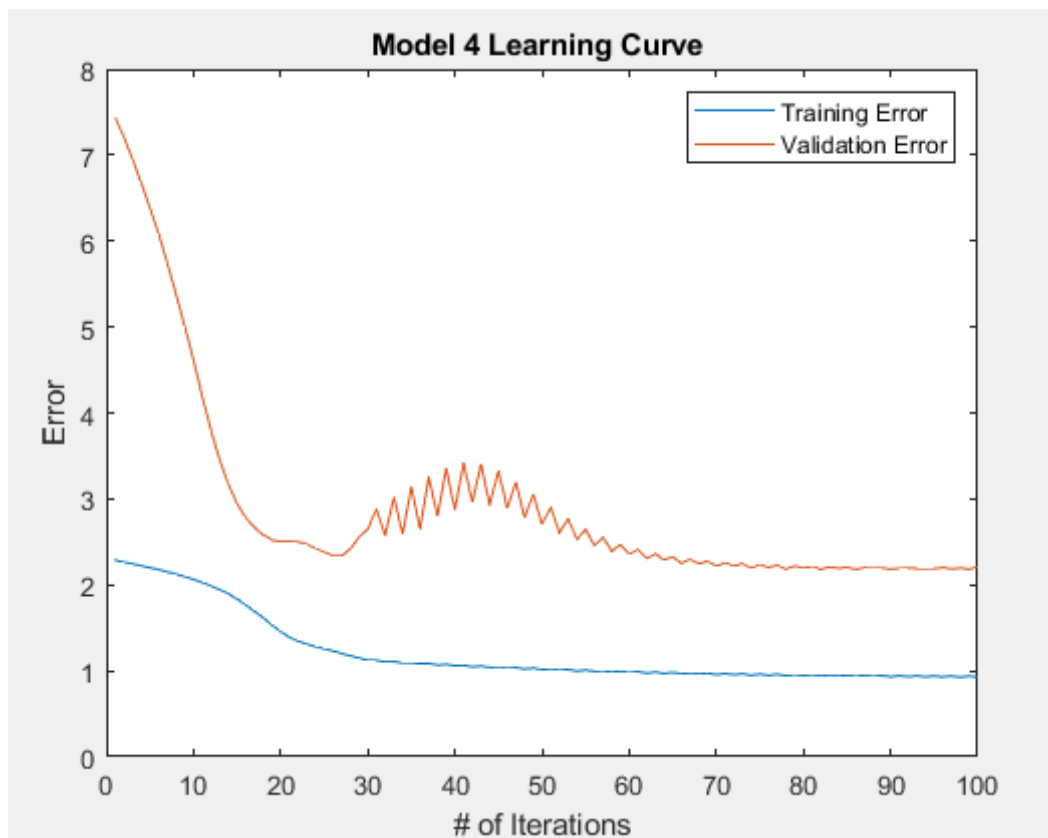
- Membership functions



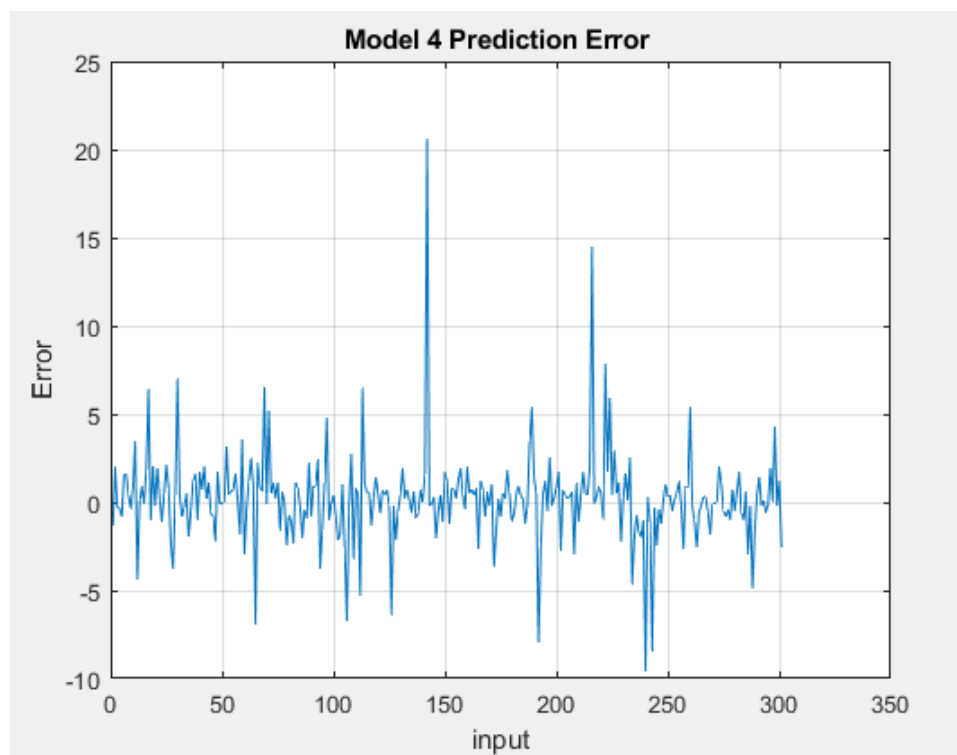




- Learning curve



- Prediction error



Μετρικές Μοντέλων

	1° Μοντέλο TSK	2° Μοντέλο TSK	3° Μοντέλο TSK	4° Μοντέλο TSK
R2	0.74223	0.43021	0.87719	0.87325
RMSE	3.6475	5.4229	2.5177	2.5577
NMSE	0.25777	0.56979	0.12281	0.12675
NDEI	0.50771	0.75484	0.35045	0.35603

Τελικά Συμπεράσματα

Από όλα τα παραπάνω, βγαίνουν τα εξής συμπεράσματα:

- Το καλύτερο μοντέλο είναι το Μοντέλο 3.
- Το χειρότερο μοντέλο είναι το Μοντέλο 2.
- Τα μοντέλα με πολυωνυμική έξοδο παρουσιάζουν καλύτερη συμπεριφορά, με μικρότερα RMSE.
- Τα μοντέλα με πολυωνυμική έξοδο παρουσιάζουν μεμονωμένα, πιο ακραία Prediction Errors.
- Το 2° Μοντέλο TSK με βάση το Learning curve, περίπου στο 20° Iteration ξεκινάει να κάνει υπερεκπαίδευση, όπου αυτό σημαίνει ότι το Validation Error ξεκινάει να αυξάνεται. Το γεγονός αυτό φαίνεται και από την μετρική του RMSE η οποία είναι και η μεγαλύτερη σχετικά με τα υπόλοιπα μοντέλα.
- Το 4° Μοντέλο, περίπου στο 30° Iteration πήγε να κάνει και αυτό υπερεκπαίδευση, αλλά στη συνέχεια το σφάλμα σταθεροποιήθηκε σε 'φυσιολογική' τιμή.
- Γενικότερα, στα μοντέλα με δύο Membership functions έχουμε μικρότερη απόκλιση μεταξύ του Training Error και του Validation Error, όπως φαίνεται ξεκάθαρα στα διαγράμματα των Learning curves. Για παράδειγμα, στο 3° μοντέλο, έχω απόκλιση $3.9 - 2.4 = 0.5$ ενώ στο 4° μοντέλο, έχω απόκλιση $2.2 - 0.9 = 1.3$.

- Τα μοντέλα με τρεις Membership functions είναι πιο επιρρεπή σε υπερεκπαίδευση, μιας και επειδή έχουν περισσότερες συναρτήσεις συμμετοχής έχουν την δυνατότητα να 'μάθουν' καλύτερα το training dataset. Το γεγονός αυτό, φάνηκε όταν τρέξαμε το σκριπτάκι στο Matlab για αρκετές φορές, κάτι το οποίο έπρεπε να γίνει μιας και η όλη παραπάνω διαδικασία έχει στοχαστικά χαρακτηριστικά.

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 (Ergasia3_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 της αναζήτησης πλέγματος. Ο μέγιστος αριθμός των κρατημένων features που διαλέξαμε είναι **δεκατέσσερα**, διότι πέρα από αυτή την τιμή ο χρόνος που χρειάζεται το Matlab για να κάνει τους υπολογισμούς, ξεφεύγει εκθετικά. Επίσης, στα features των data κάναμε κανονικοποίηση 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.

Αποτελέσματα της αναζήτησης πλέγματος

- RMSE

r Kep_features	0.2	0.4	0.6	0.8	1
5	16.2848	16.6507	18.4870	20.1784	19.8743
8	15.9559	15.5907	16.6101	18.0179	18.1420
11	15.5212	14.8123	15.5734	16.6931	17.1698
14	15.3814	14.5988	15.1127	15.2741	15.9982

- R2

r Kep_features	0.2	0.4	0.6	0.8	1
5	0.7739	0.7636	0.7077	0.6529	0.6633
8	0.7829	0.7928	0.7648	0.7233	0.7194
11	0.7944	0.8128	0.7933	0.7621	0.7487
14	0.7981	0.8183	0.8053	0.8011	0.7818

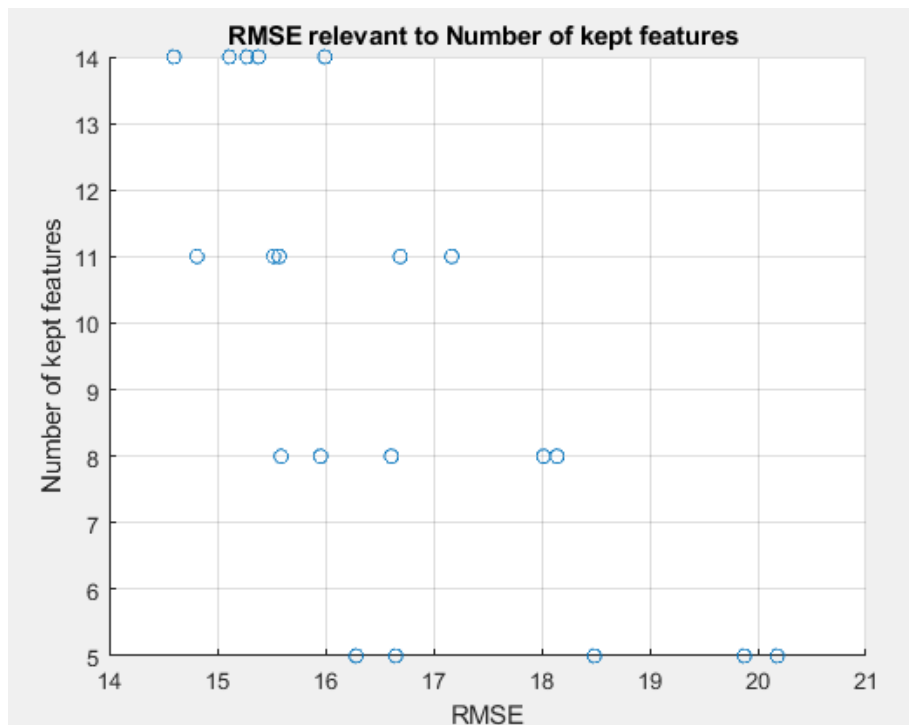
Τα παραπάνω αποτελέσματα βγήκαν μετά από τη διαδικασία της αναζήτησης πλέγματος για 5, 8, 11, 14 features και 0.2, 0.4, 0.6, 0.8, 1 ακτίνα r . Επίσης, το κάθε μοντέλο εκπαιδεύτηκε για 100 epoches και τα αποτελέσματα των μετρικών, βγήκαν με τη βοήθεια της μεθόδου του 5-fold cross validation. Από τους παραπάνω πίνακες, βλέπουμε ότι το μικρότερο σφάλμα RMSE παρουσιάζεται για τον συνδυασμό: κρατημένα features = 14 και ακτίνα $r = 0.4$.

Στη συνέχεια παραθέτονται scatters όπου φαίνονται το σφάλμα RMSE σε σχέση με την ακτίνα r , τα κρατημένα features και τον αριθμό των κανόνων αντίστοιχα.

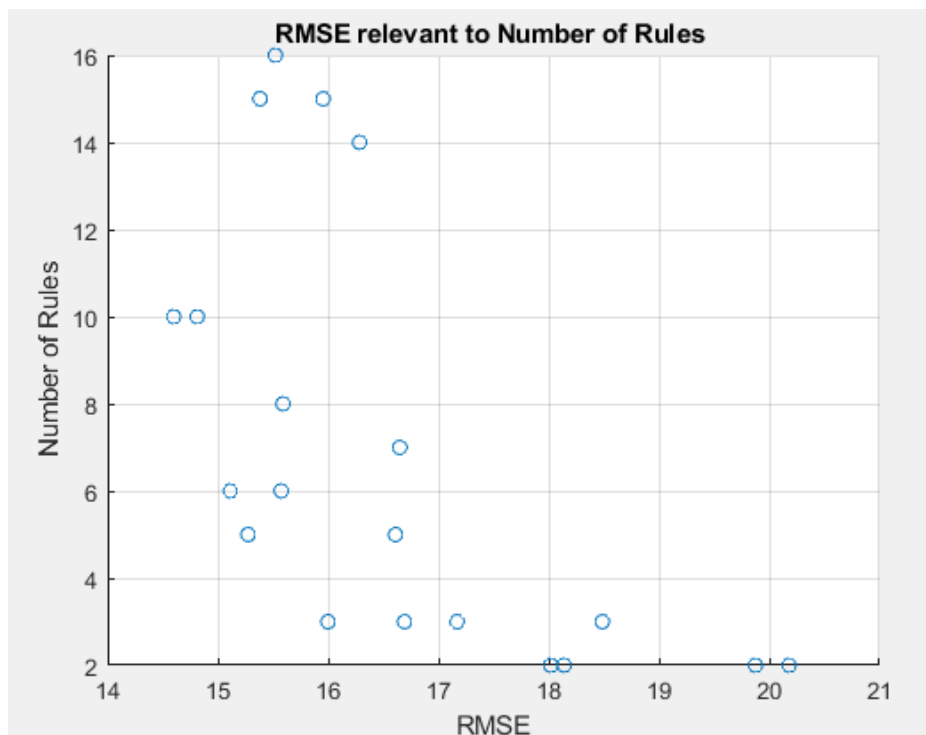
- RMSE και ακτίνα r



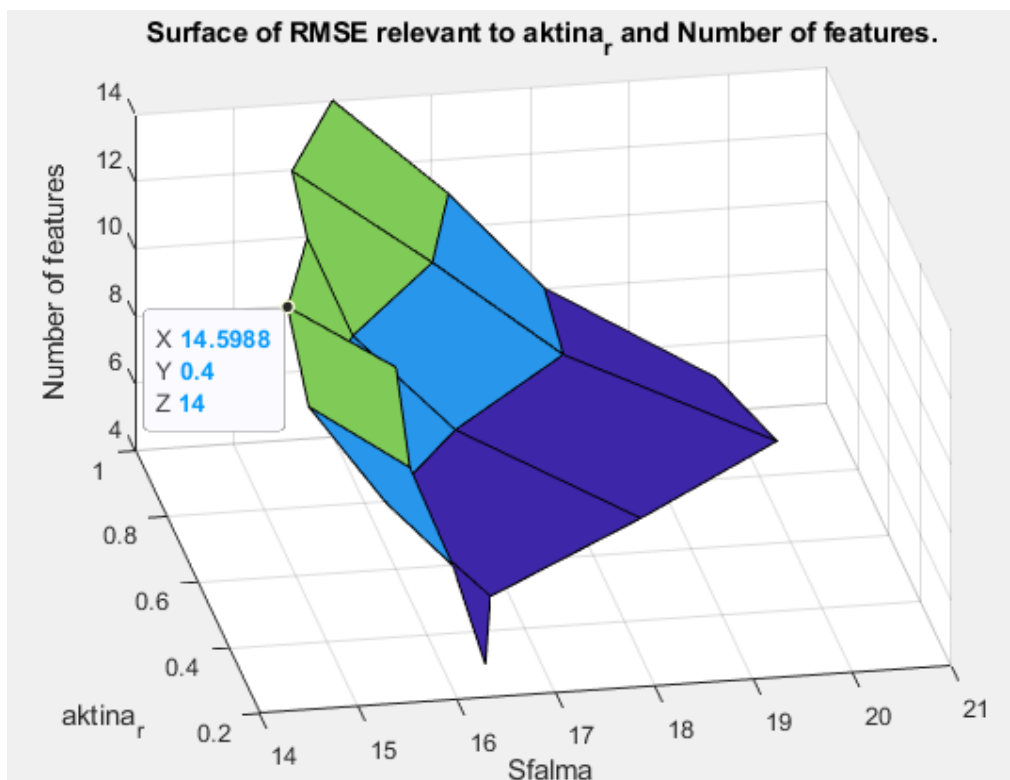
- RMSE και κρατημένα features



- RMSE και αριθμός κανόνων



Τέλος, στην παρακάτω εικόνα φαίνεται η επιφάνεια που δείχνει τα αποτελέσματα του πίνακα με τα RMSE τις κάθε δοκιμής.



Τελικά συμπεράσματα

Από όλα τα παραπάνω συμπεραίνουμε ότι το βέλτιστο μοντέλο με βάση τη μετρική του RMSE είναι το μοντέλο με `kept_features = 14` και `aktina_r = 0.4`. Επίσης, βλέπουμε ότι όσο μεγαλώνει η `aktina_r` στις δοκιμές μας, αυξάνεται και το RMSE, αν υποθέσουμε σταθερό αριθμό κρατημένων features (βλέπω γραμμές). Παρόμοια με το προηγούμενο σκεπτικό αν υποθέσουμε ότι έχουμε σταθερή `aktina_r` (κοιτάμε τις στήλες), βλέπουμε ότι όσο αυξάνονται τα κρατημένα features τόσο μικραίνει και το RMSE. Η εξήγηση για το παραπάνω γεγονός είναι ότι όσο ο αριθμός των κρατημένων features αυξάνεται, τόσο λιγότερη πληροφορία από το dataset μας, χάνεται. Συνεπώς και το RMSE των προβλέψεων του μοντέλου θα είναι μικρότερο. Δεν μπορούμε όμως να κρατήσουμε πολλά features διότι ο χρόνος εκπαίδευσης ξεφεύγει εκθετικά, όπως και ο αριθμός των κανόνων του μοντέλου. Επιπλέον, σε ότι αφορά την ακτίνα r των clusters, ισχύει ότι, όσο μικρότερη ακτίνα έχω, τόσο καλύτερα το μοντέλο μαθαίνει το dataset, συνεπώς τόσο καλύτερες προβλέψεις κάνει. Αυτό το γεγονός όμως ισχύει μέχρι ένα σημείο, διότι πέρα από ένα συγκεκριμένο κατώφλι της ακτίνας r , το μοντέλο κάνει overfitting, δηλαδή μαθαίνει πάρα πολύ καλά το training data και στη συνέχεια το μοντέλο δε μπορεί να γενικεύσει αυτά που 'έμαθε'. Για παράδειγμα, στη δικιά μας περίπτωση βλέπουμε ότι για 14 κρατημένα features, με ακτίνα r των cluster στα 0.2, το RMSE είναι 15.3814, που είναι χειρότερο από το 14.5988, το οποίο RMSE το έχω για 14 κρατημένα features και ακτίνα r των cluster στα 0.4.

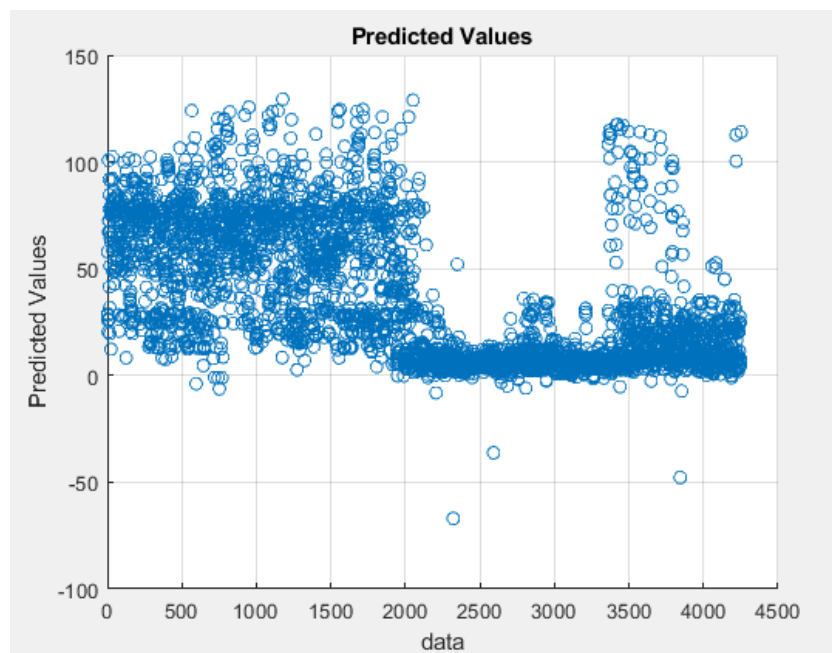
Βέλτιστο Μοντέλο TSM με `kept_features = 14` και `aktina_r = 0.4`

Το βέλτιστο μοντέλο εκπαιδεύτηκε ακριβώς με τον ίδιο τρόπο όπως και στο σκριπτάκι 'Ergasia3_2', με την μόνη διαφορά ότι η εκπαίδευση έγινε για 120 epoches αντί για 100, προκειμένου να πάρουμε ένα καλύτερο μοντέλο μετά την εκπαίδευση. Ακολουθούν τα αποτελέσματα παρακάτω.

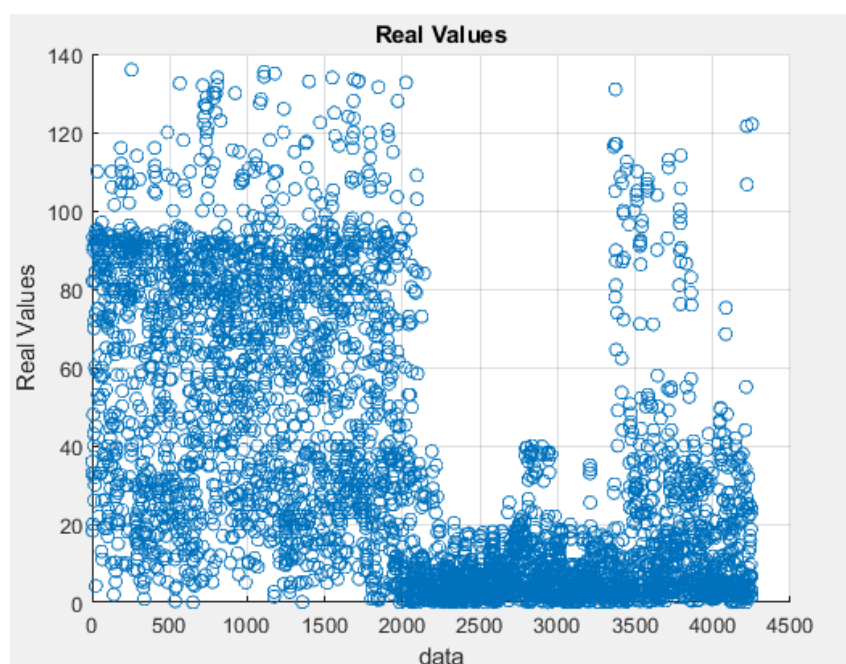
- Πίνακας δεικτών απόδοσης

	R2	RMSE	NMSE	NDEI
Βέλτιστο Μοντέλο TSK	0.8228	14.4174	0.1772	0.4209

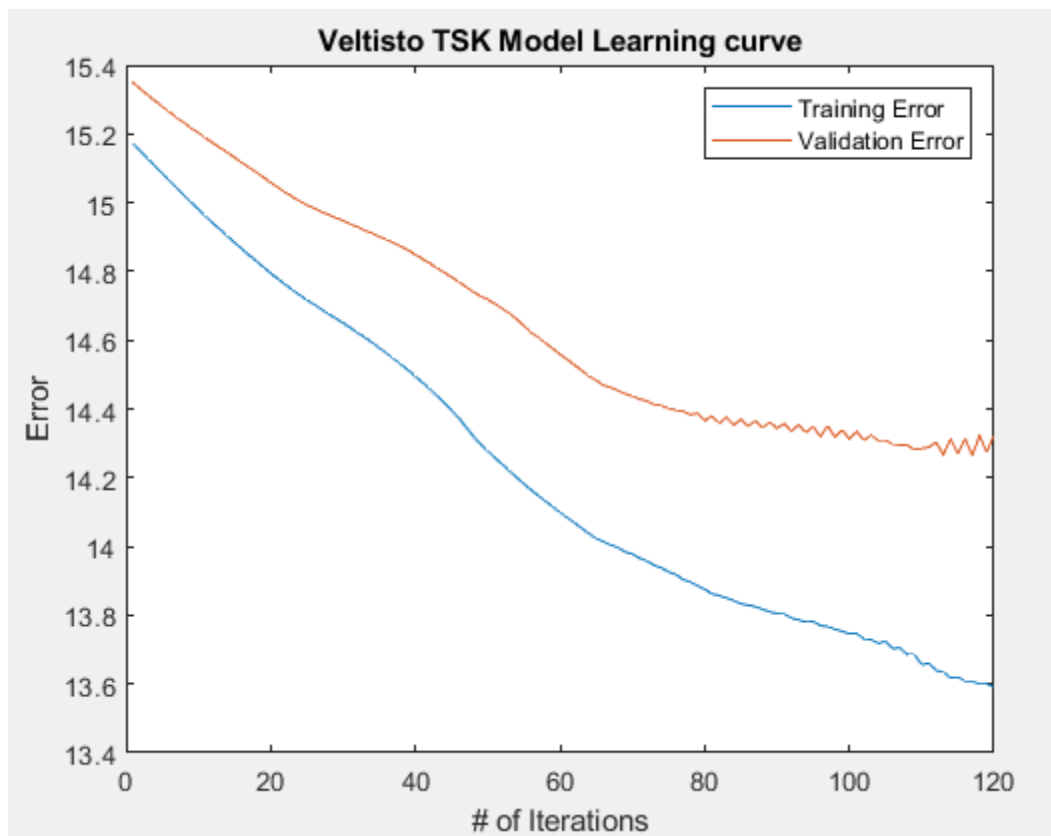
- Προβλέψεις βέλτιστου μοντέλου



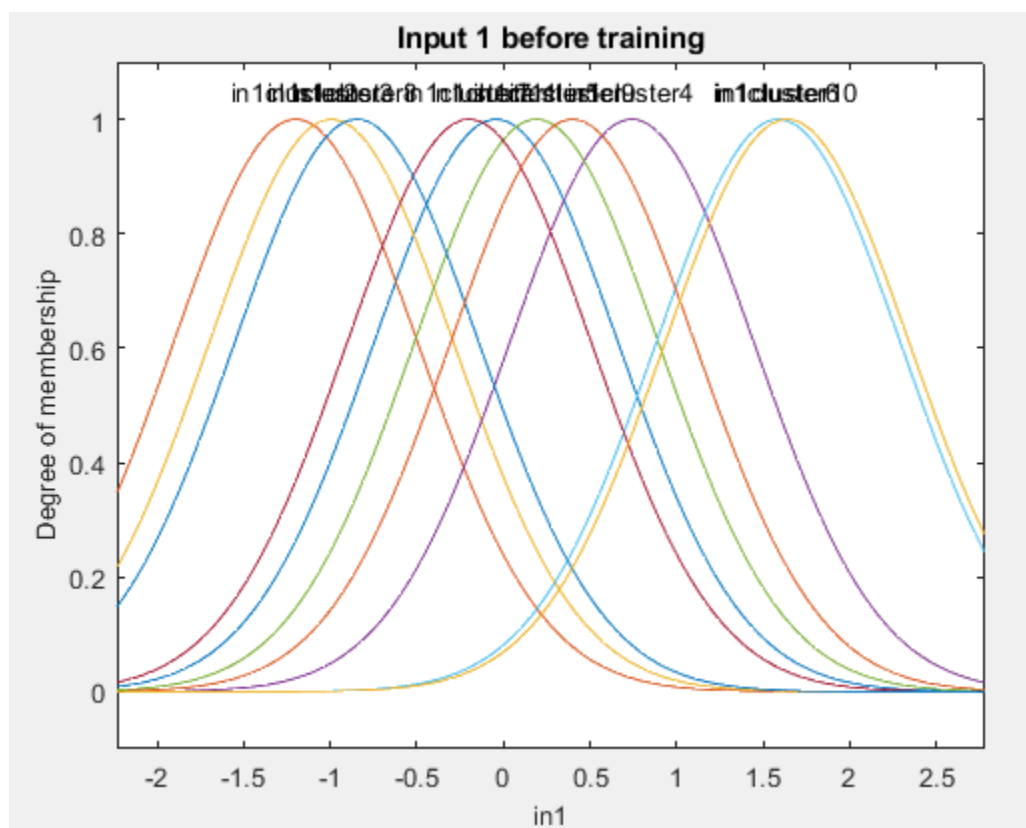
- Πραγματικές τιμές Target του dataset

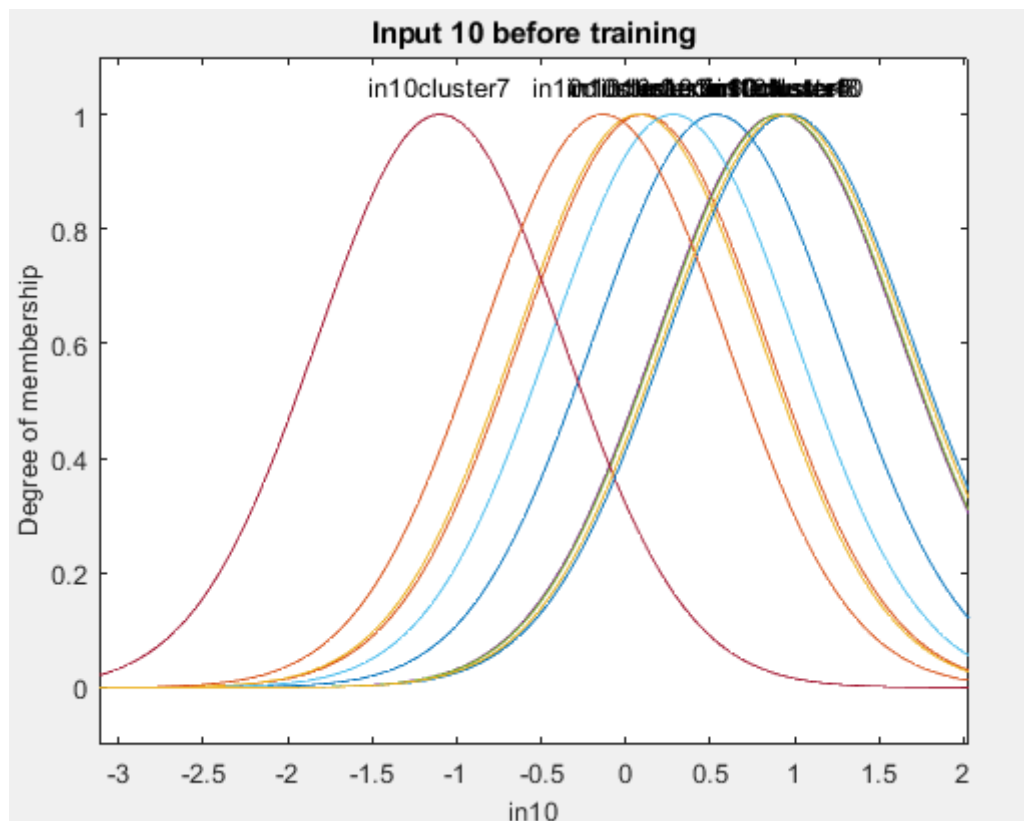


- Learning curve

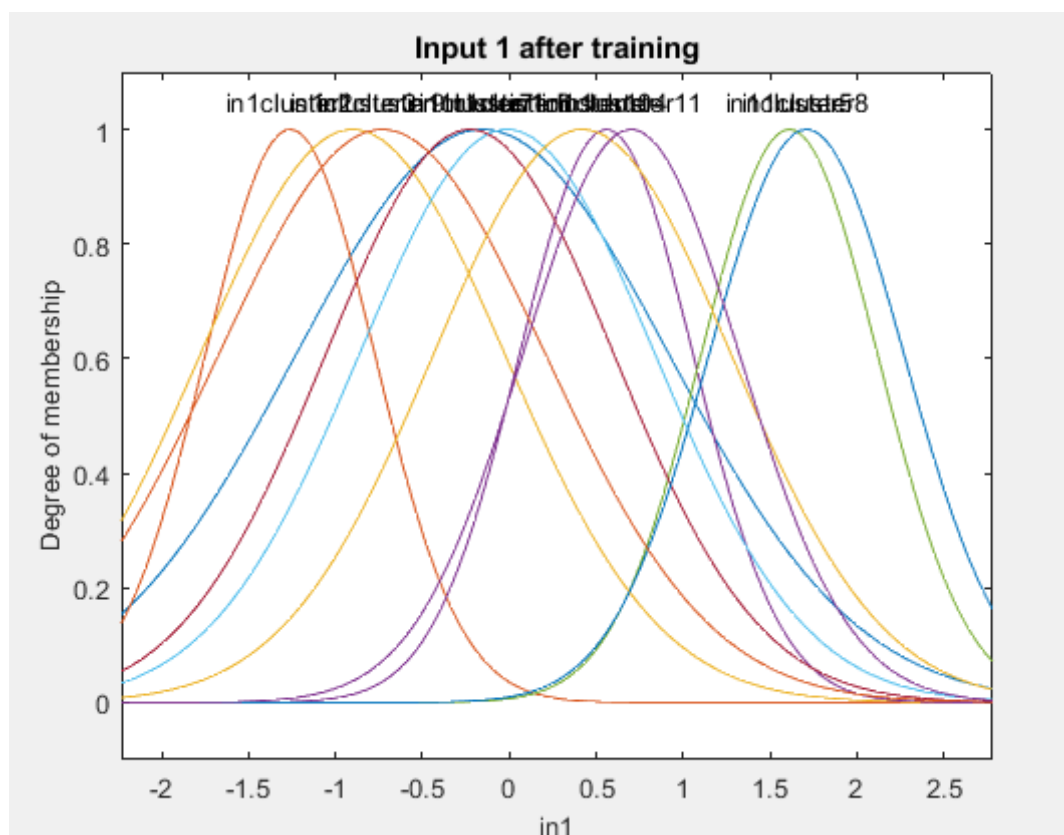


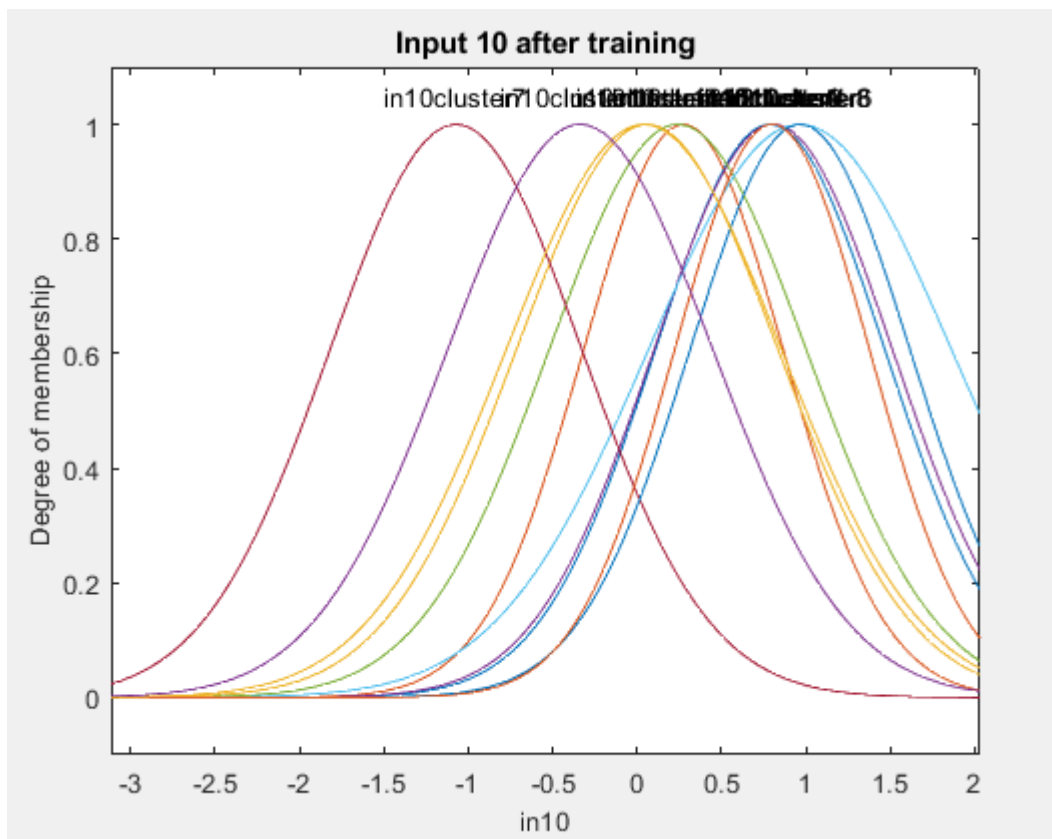
- Ασαφή σύνολα στην αρχική τους μορφή



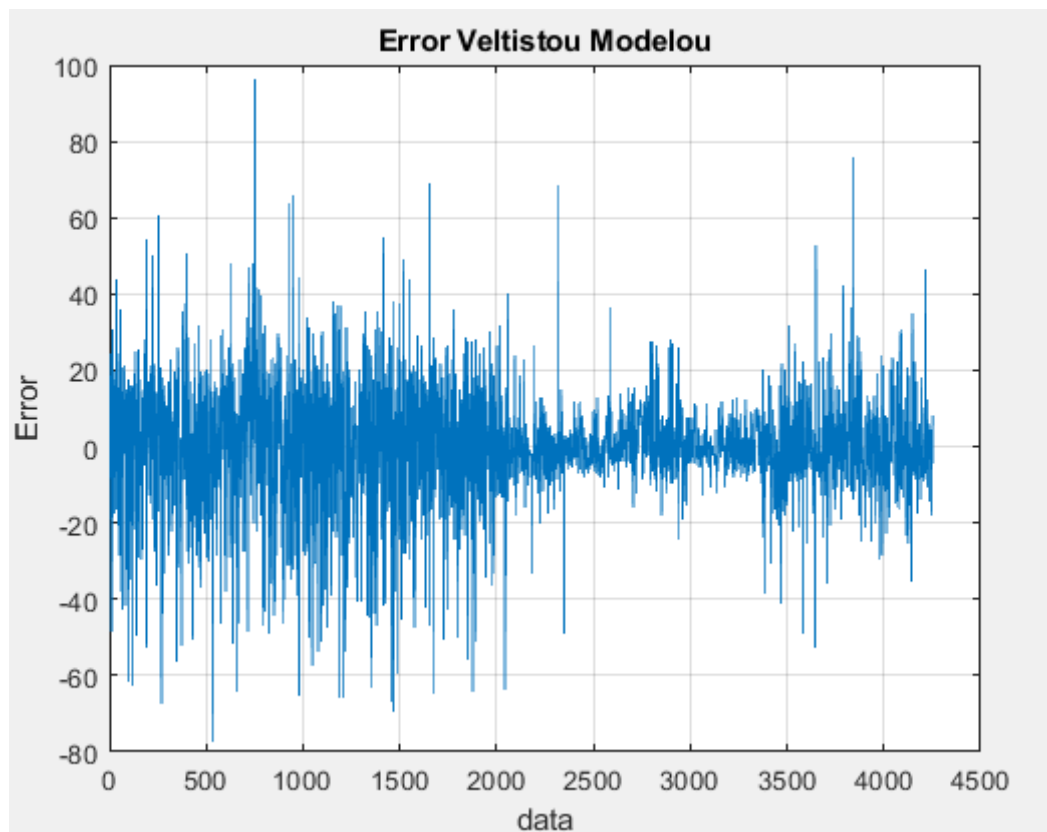


- Ασαφή σύνολα στην τελική τους μορφή





- Σφάλμα προβλέψεων στα testing data



Τελικά συμπεράσματα για το βέλτιστο μοντέλο

Βλέποντας, τα διαγράμματα των Predicted Values και των Real Values, βλέπουμε ότι το μοντέλο προβλέπει αρκετά καλά το testing dataset. Επιπλέον, το πλήθος των κανόνων του βέλτιστου μοντέλου είναι 11 κανόνες και αυτό το βλέπουμε από το valFis.Rules. 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. Τα δύο τελευταία νούμερα, σε σχέση με το 11 έχουν τεράστια διαφορά και μας επιβεβαιώνουν πόσο μη πρακτικό και χρονοβόρο στη φάση του training, θα ήταν το μοντέλο μας, αν αντί για την μέθοδο του subtractive clustering χρησιμοποιούσαμε τη μέθοδο του grid partitioning.