



## "Προηγμένα Συστήματα Υπολογιστών & Επικοινωνιών"

Διαπανεπιστημιακό Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών

- Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών
- Ιατρικής • Μουσικών Σπουδών • Δημοσιογραφίας και ΜΜΕ
- Λογιστικής και Χρηματοοικονομικής του Πανεπιστημίου Μακεδονίας



## Επίλυση προβλημάτων (i) παλινδρόμησης & (ii) ταξινόμησης με χρήση μοντέλων TSK

Εξαμηνιαία Εργασία Μαθήματος  
«Υπολογιστική νοημοσύνη – Συστήματα  
εμπνευσμένα από τη Βιολογία: Ασαφή Συστήματα»

**Ομάδα**

Τσαλίδης Γεώργιος 472  
Ντόμι Αρμάντο 445

Θεσσαλονίκη, Ιούλιος 2020

# Εισαγωγή

Στόχος της εργασίας αυτής είναι να διερευνηθεί η ικανότητα των μοντέλων TSK στη μοντελοποίηση πολυμεταβλητών μη γραμμικών συναρτήσεων τόσο σε προβλήματα ταξινόμησης όσο και για προβλήματα παλινδρόμησης. Επιλέγονται τέσσερα σύνολα δεδομένων, όπου τα πρώτα δύο σύνολα θα χρησιμοποιηθεί για μια απλή διερεύνηση της διαδικασίας εκπαίδευσης και αξιολόγησης μοντέλων αυτού του είδους καθώς και μια επίδειξη τρόπων ανάλυσης και ερμηνείας των αποτελεσμάτων. Τα δύο υπόλοιπα θα χρησιμοποιηθούν για μια πληρέστερη διαδικασία μοντελοποίησης, η οποία θα περιλαμβάνει και προ επεξεργαστικά βήματα.

## I. Παλινδρόμηση

### 1. Εφαρμογή σε απλό dataset

Το dataset που θα χρησιμοποιηθεί περιλαμβάνει 1503 δείγματα και 6 χαρακτηριστικά από τα οποία θα προβλέψουμε αυτά της τελευταίας στήλης. Τα αποτελέσματα των τεσσάρων (4) μοντέλων TSK που δημιουργήθηκαν παρουσιάζονται παρακάτω:

- TSK\_model\_1

Πλήθος συναρτήσεων συμμετοχής: 2

Μορφή εξόδου: Singleton

Αριθμός κανόνων: 32

Μερικές συναρτήσεις συμμετοχής πριν την εκπαίδευση

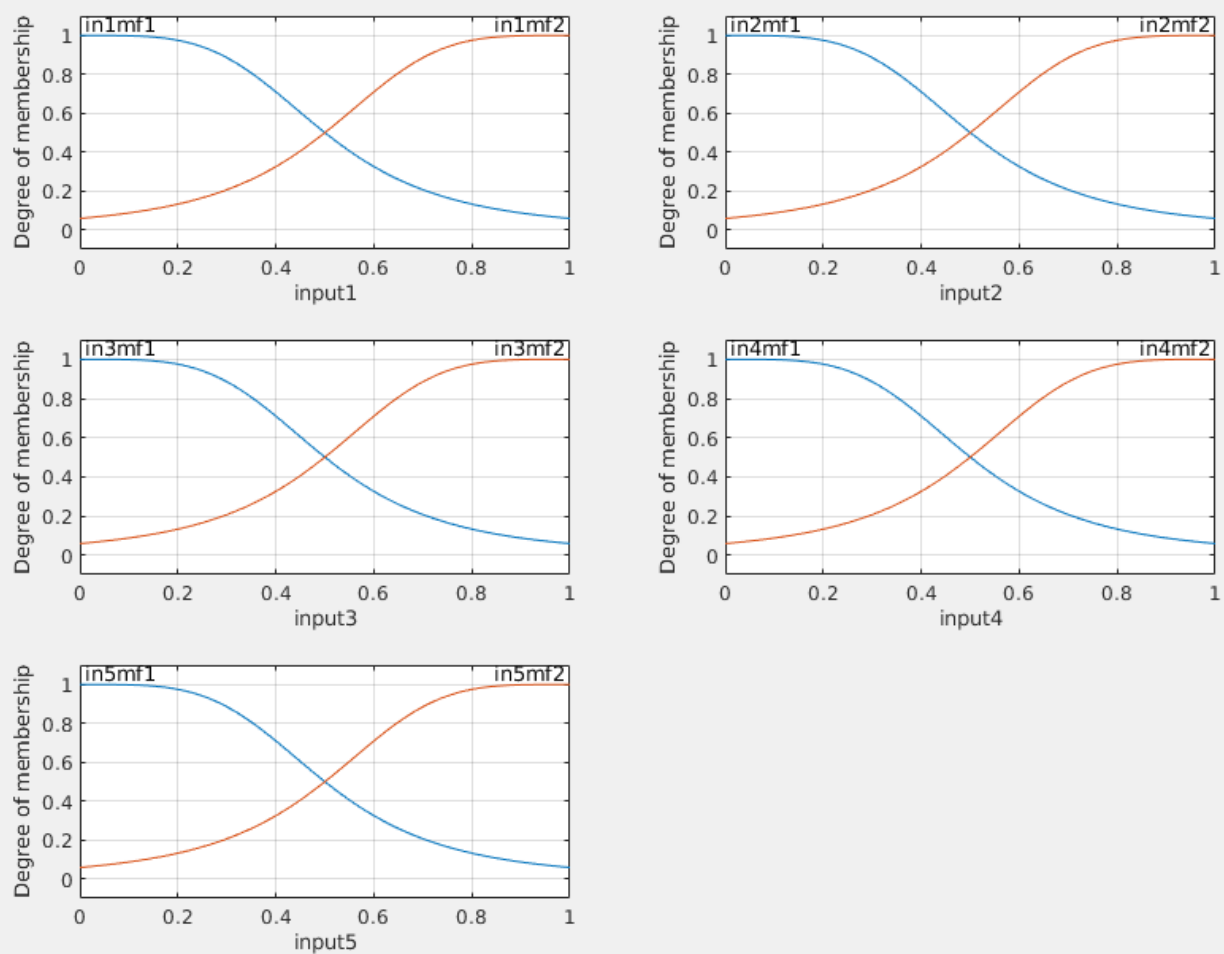


Figure 1: TSK\_model\_1 initial state.

Μερικές συναρτήσεις συμμετοχής μετά την εκπαίδευση

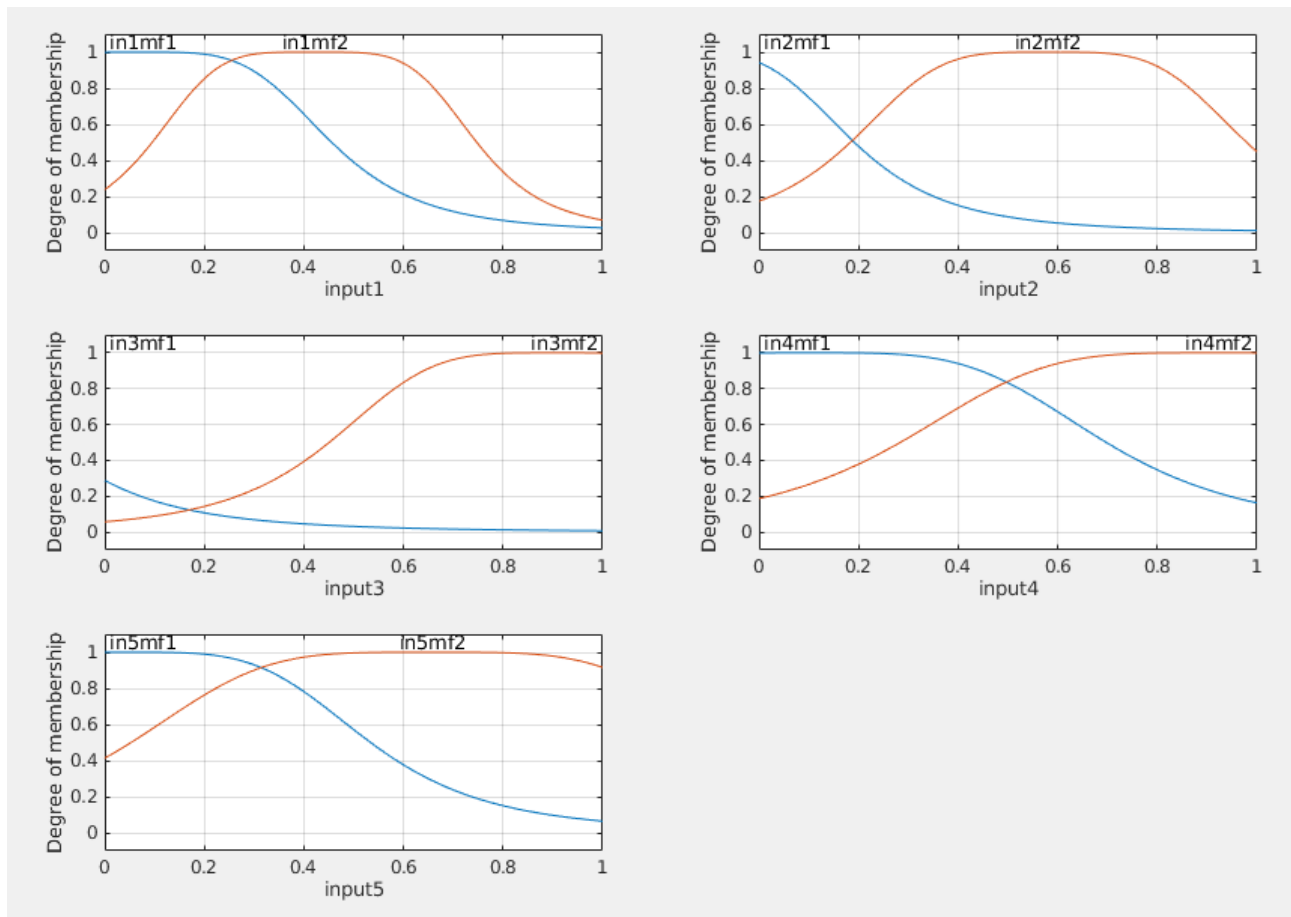


Figure 2: TSK\_model\_1 final state.

Διαγράμματα μάθησης (learning curves) όπου να απεικονίζεται το σφάλμα του μοντέλου συναρτήσει του αριθμού των επαναλήψεων (iterations)

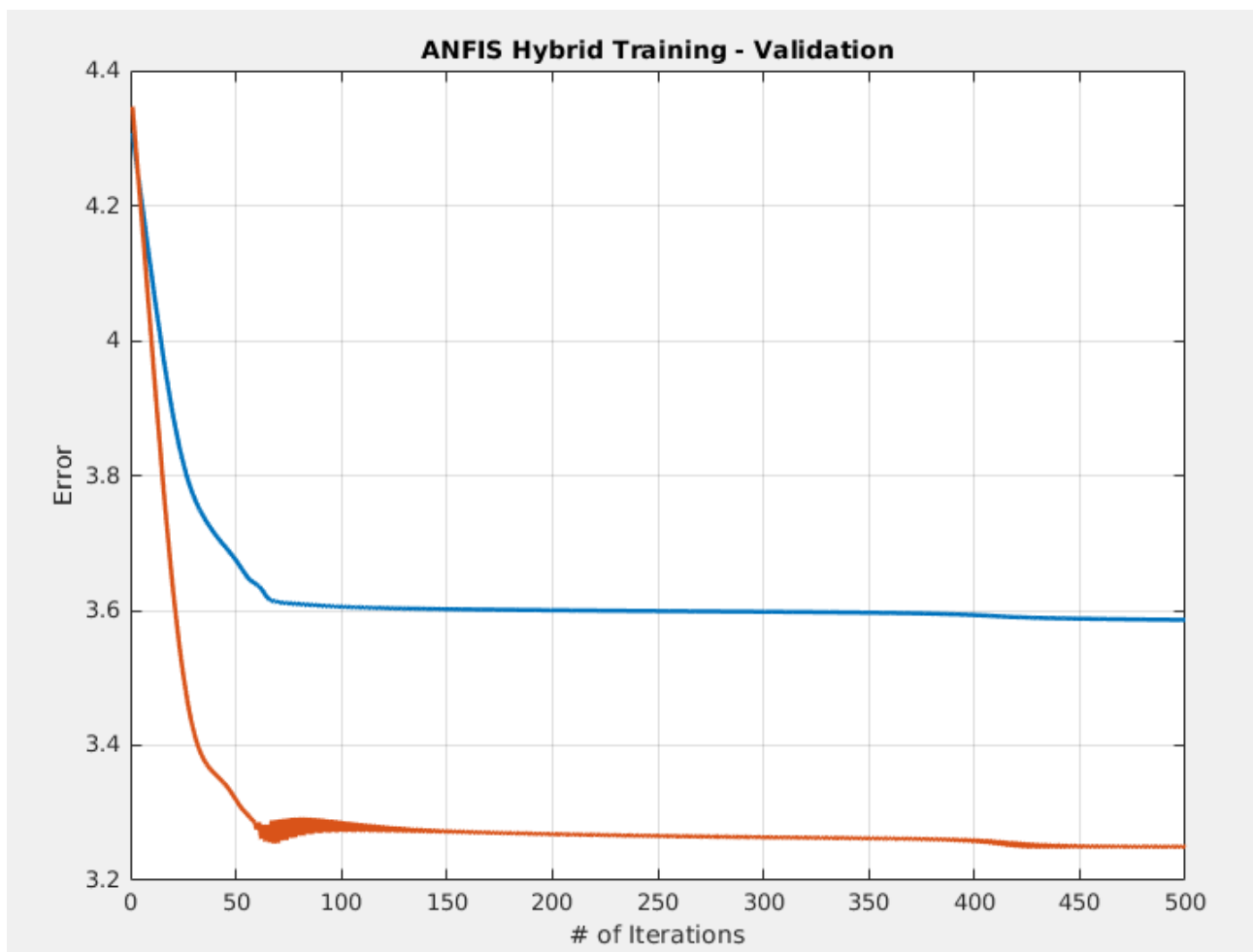


Figure 3: TSK\_model\_1 training and testing error.

- TSK\_model\_2

Πλήθος συναρτήσεων συμμετοχής: 3

Μορφή εξόδου: Singleton

Αριθμός κανόνων: 243

Μερικές συναρτήσεις συμμετοχής πριν την εκπαίδευση

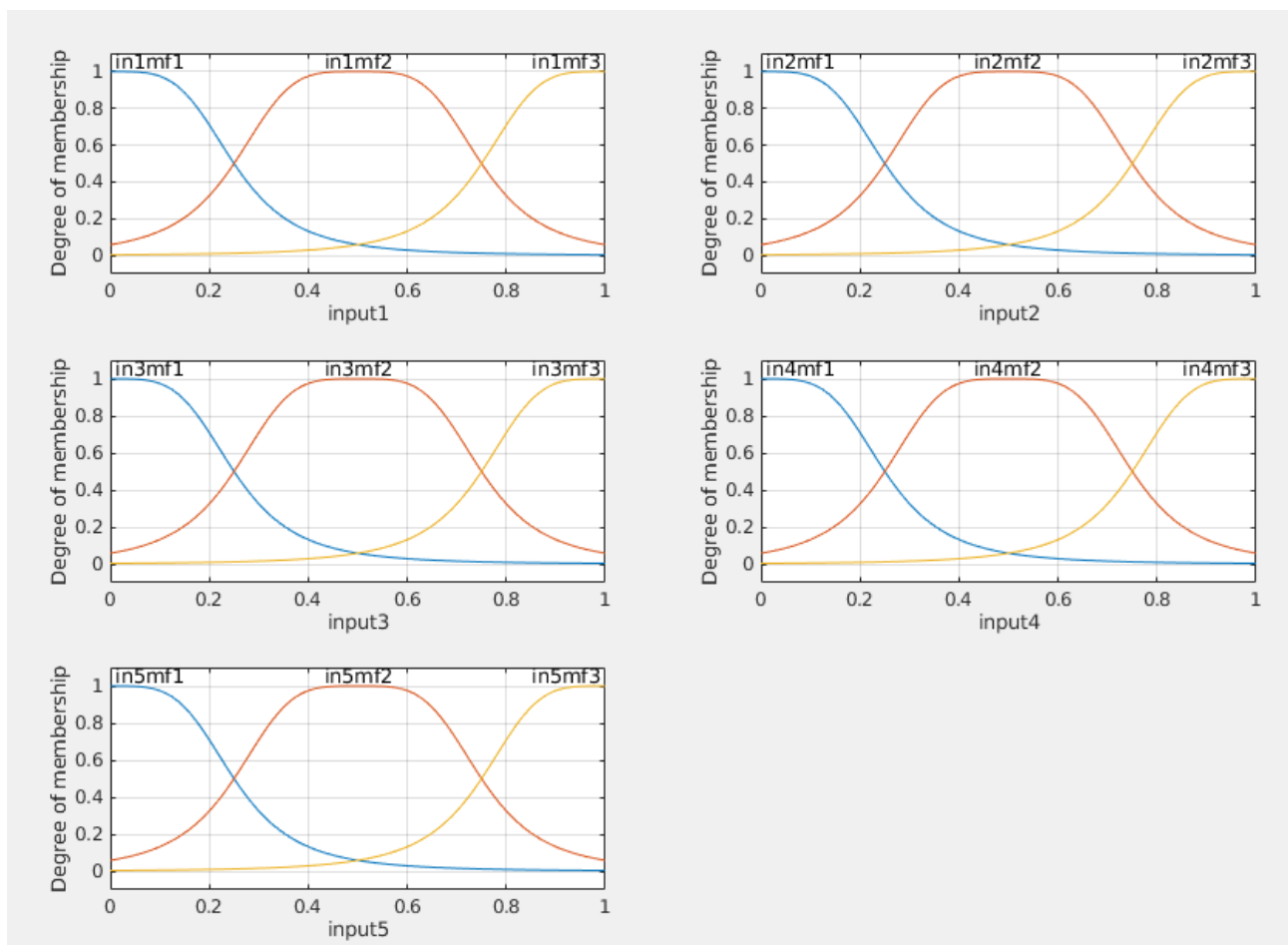


Figure 4: TSK\_model\_2 initial state.

Μερικές συναρτήσεις συμμετοχής μετά την εκπαίδευση

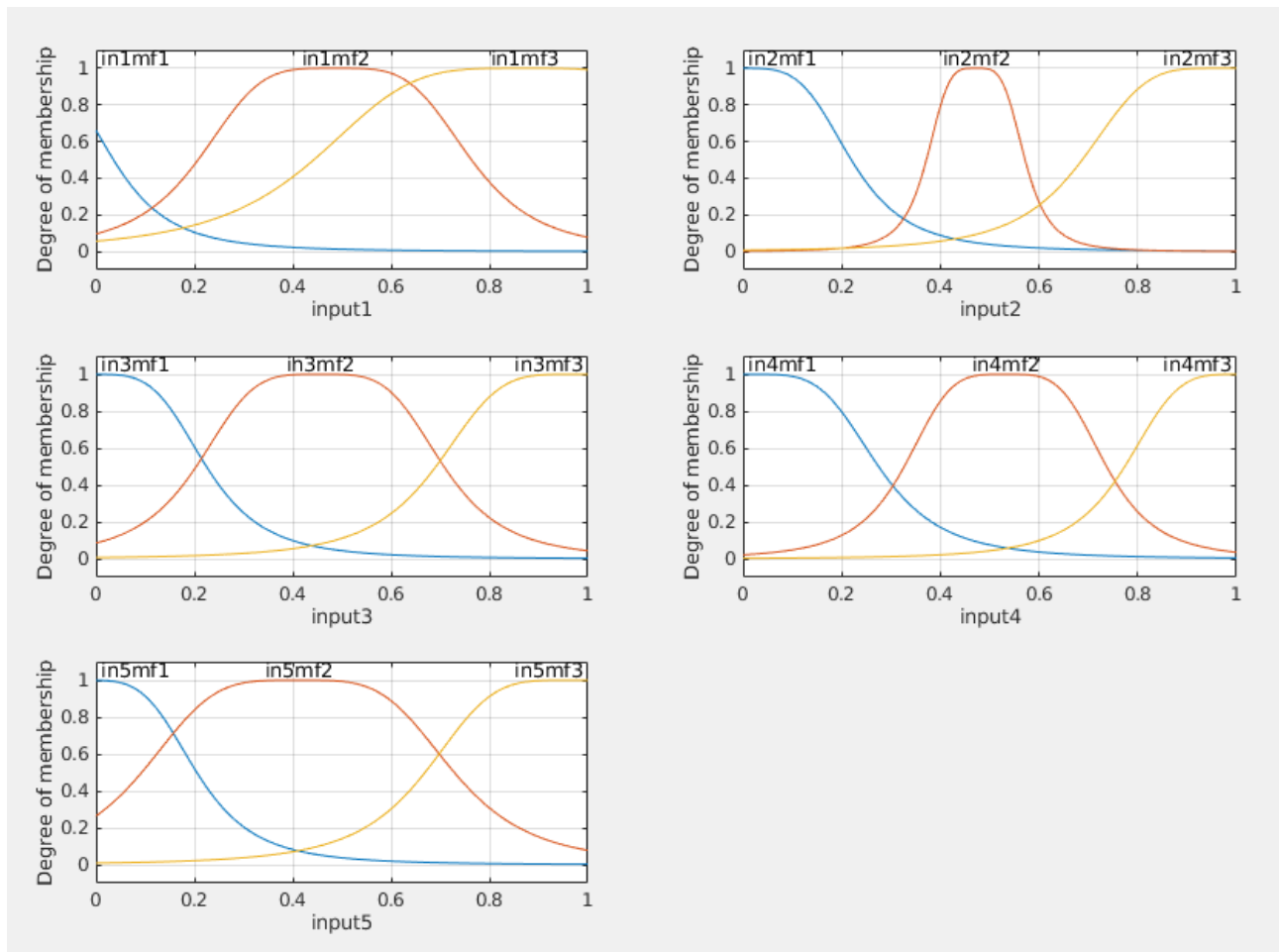


Figure 5: TSK\_model\_2 final state.

Διαγράμματα μάθησης (learning curves) όπου να απεικονίζεται το σφάλμα του μοντέλου συναρτήσει του αριθμού των επαναλήψεων (iterations)

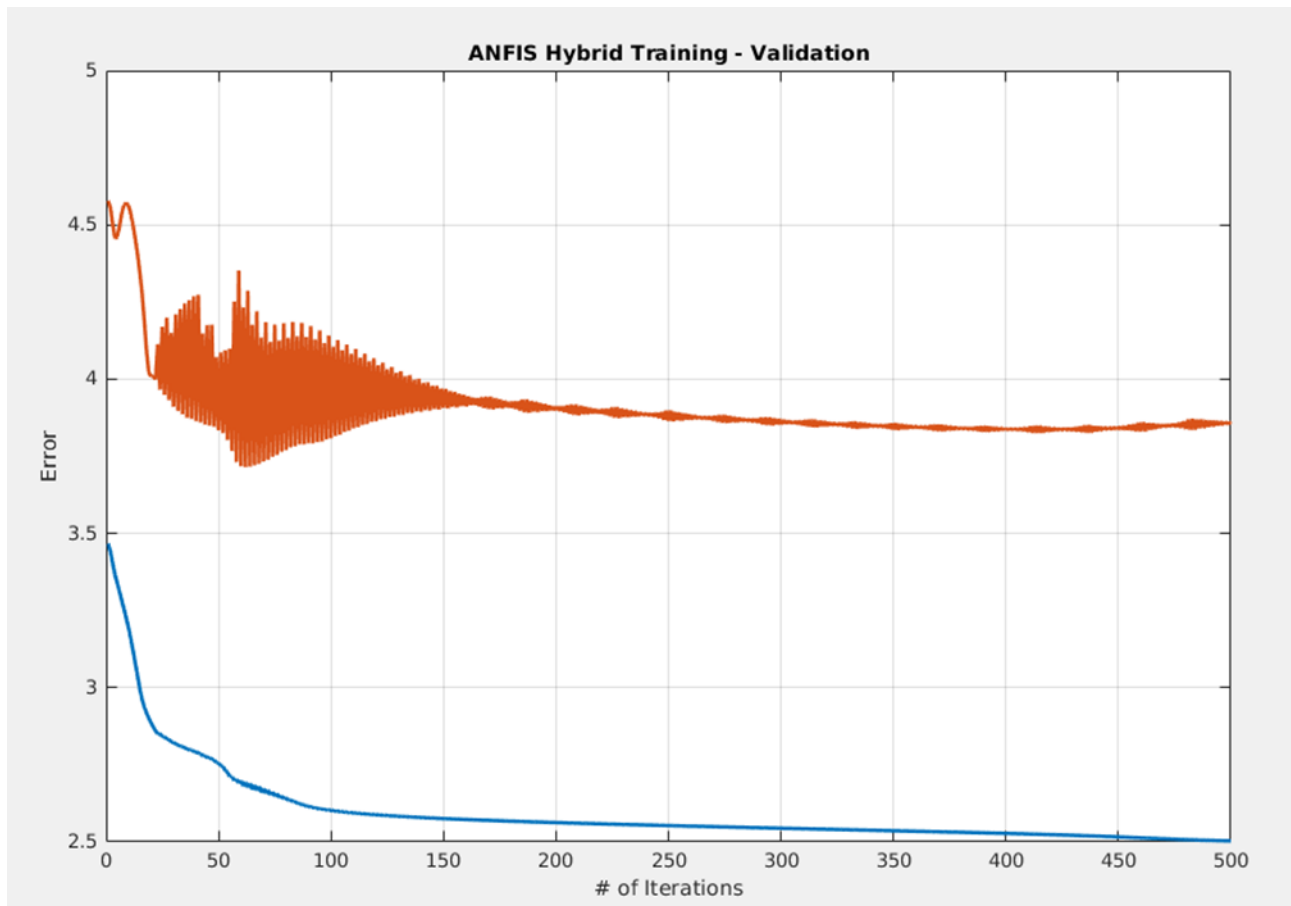


Figure 6: TSK\_model\_2 training and testing error.

- TSK\_model\_3

Πλήθος συναρτήσεων συμμετοχής: 2

Μορφή εξόδου: Polynomial

Αριθμός κανόνων: 32

Μερικές συναρτήσεις συμμετοχής πριν την εκπαίδευση



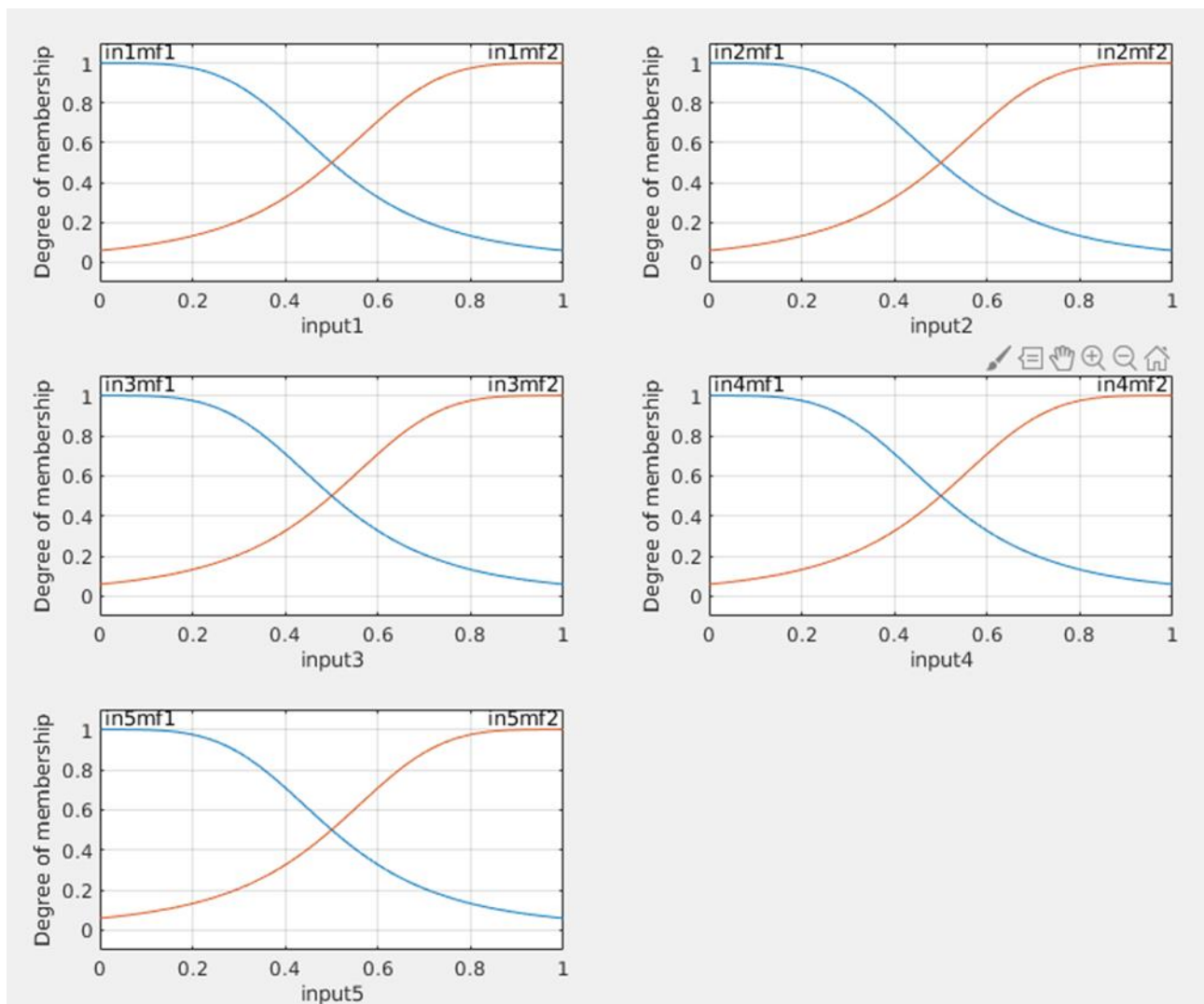


Figure 7: TSK\_model\_3 initial state.

Μερικές συναρτήσεις συμμετοχής μετά την εκπαίδευση

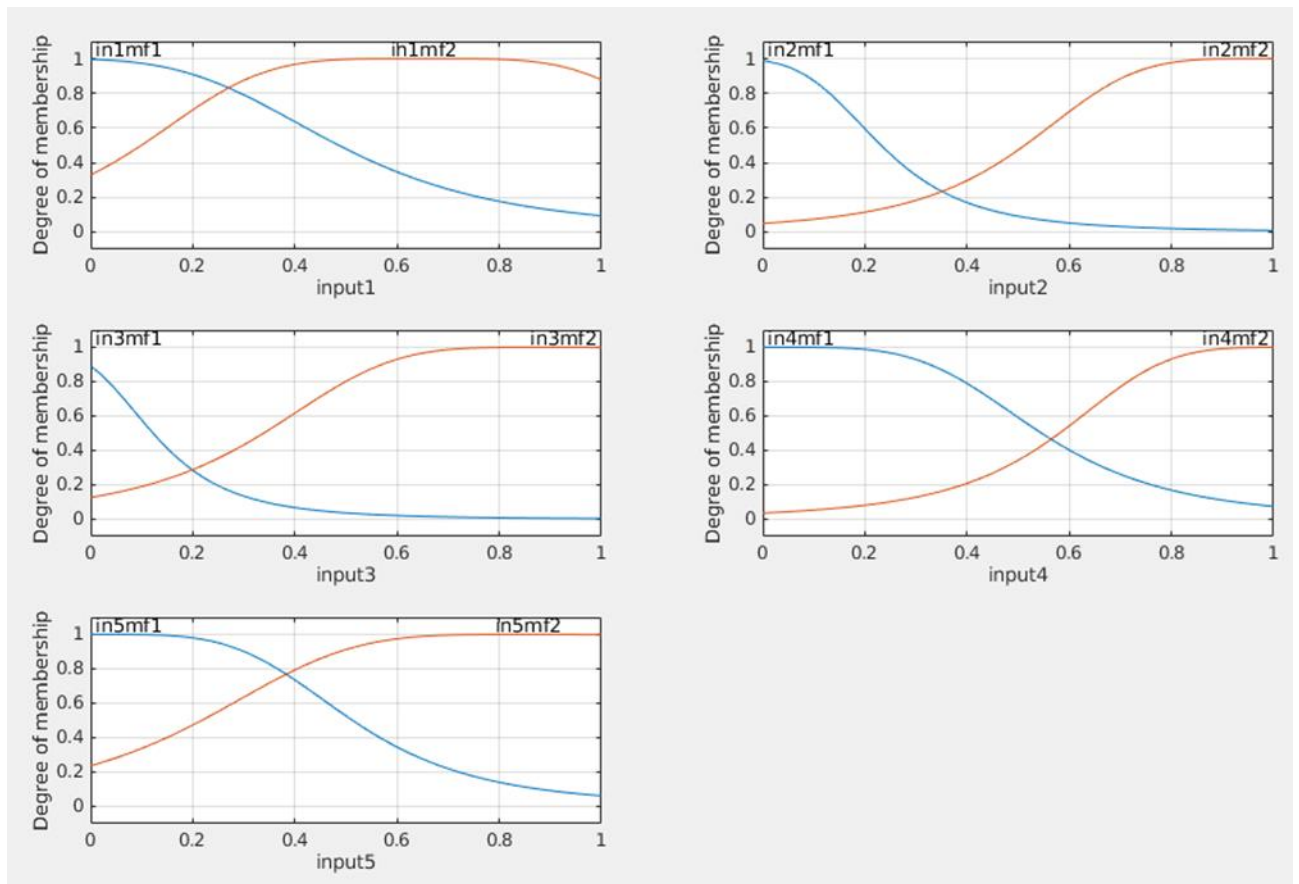


Figure 8: TSK\_model\_3 final state.

Διαγράμματα μάθησης (learning curves) όπου να απεικονίζεται το σφάλμα του μοντέλου συναρτήσει του αριθμού των επαναλήψεων (iterations)

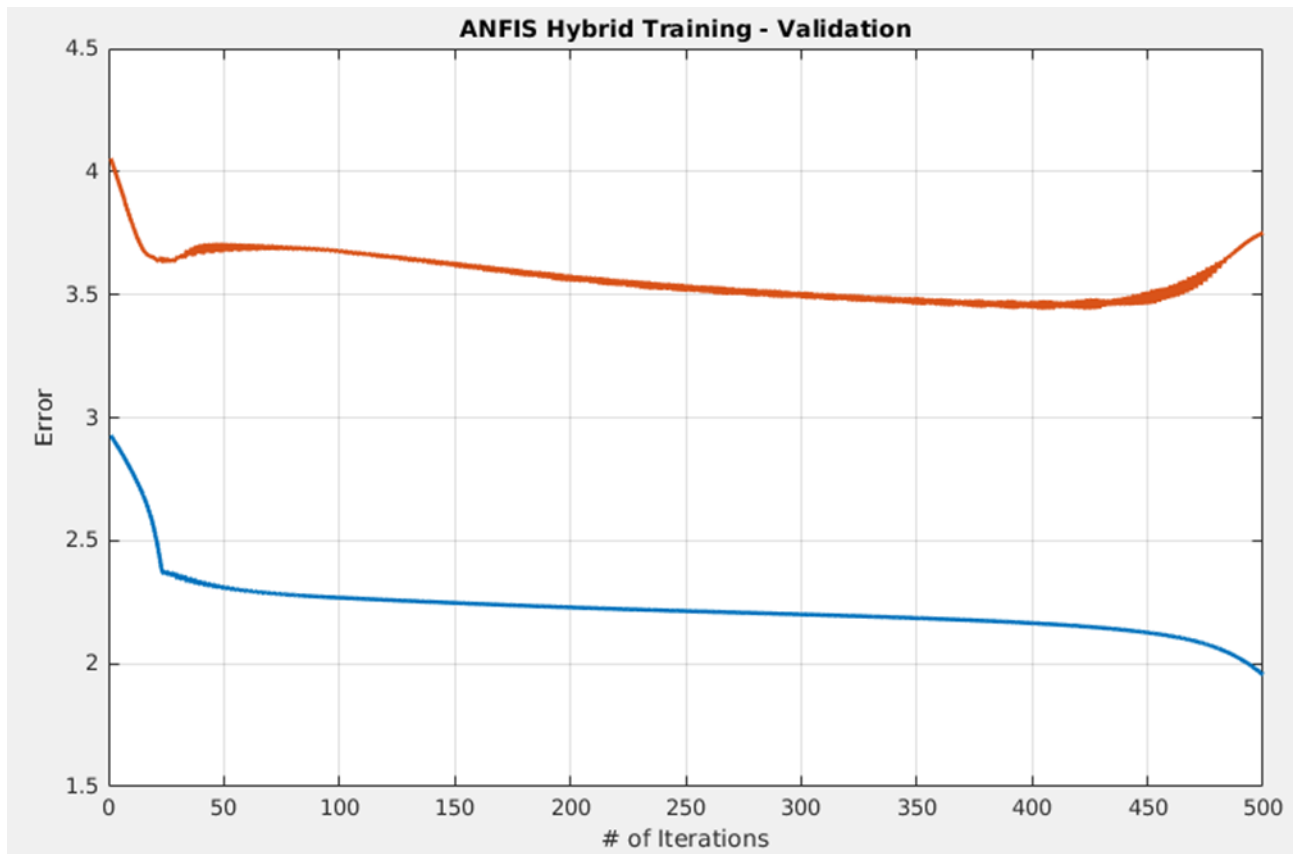


Figure 9: TSK\_model\_3 training and testing error.

- TSK\_model\_4

Πλήθος συναρτήσεων συμμετοχής: 3

Μορφή εξόδου: Polynomial

Αριθμός κανόνων: 243

Μερικές συναρτήσεις συμμετοχής πριν την εκπαίδευση

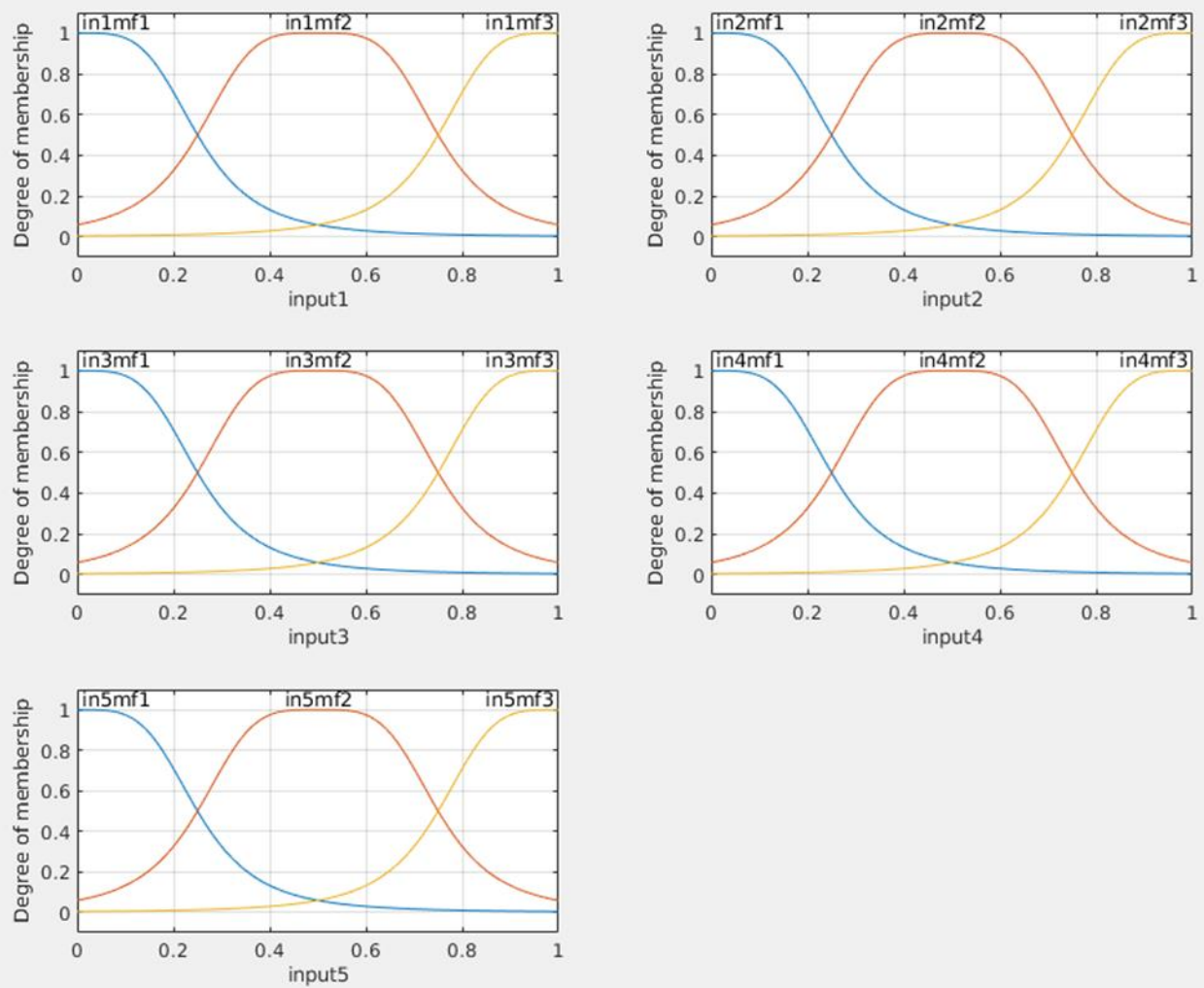


Figure 10: TSK\_model\_4 initial state.

Μερικές συναρτήσεις συμμετοχής μετά την εκπαίδευση

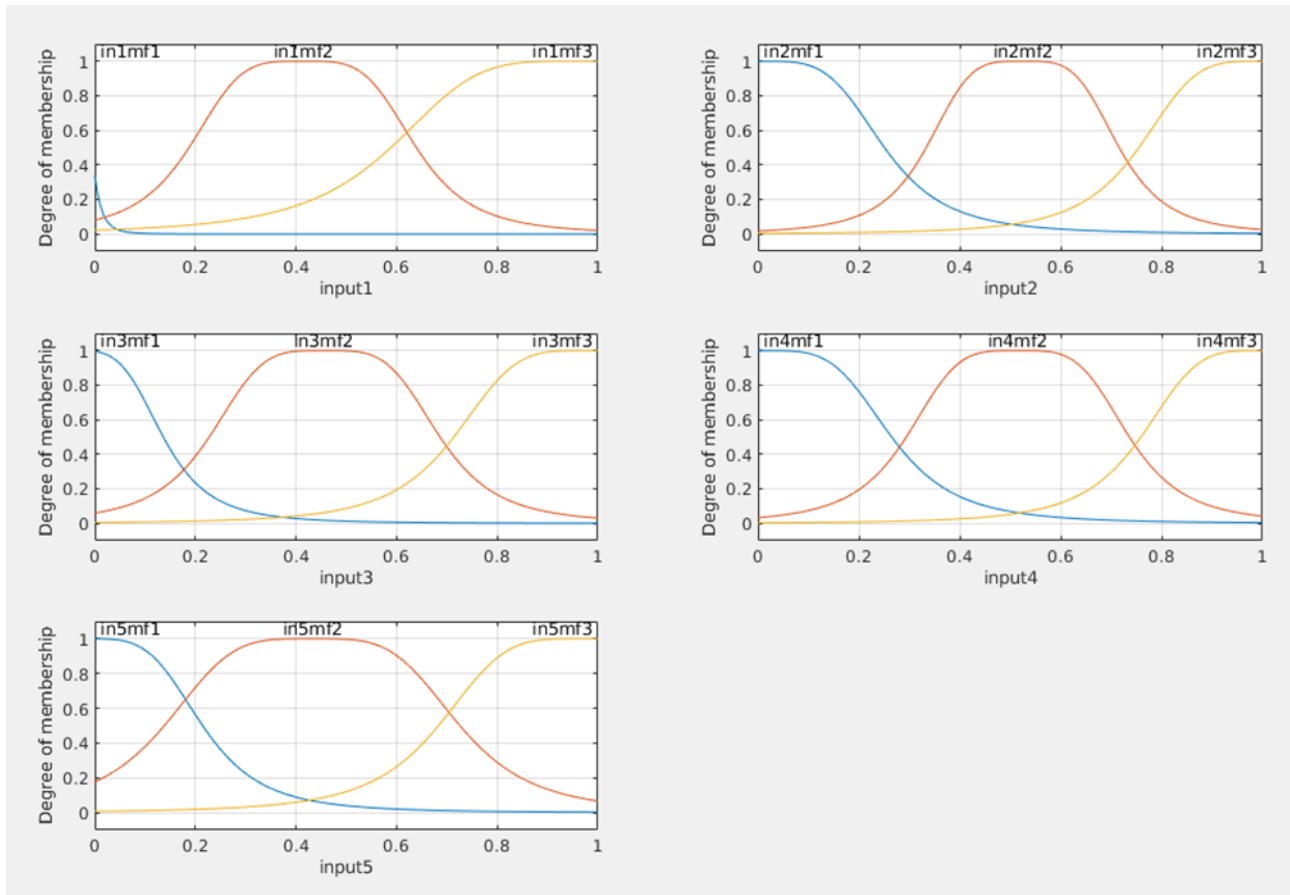


Figure 11: TSK\_model\_4 final state

Διαγράμματα μάθησης (learning curves) όπου να απεικονίζεται το σφάλμα του μοντέλου συναρτήσει του αριθμού των επαναλήψεων (iterations)

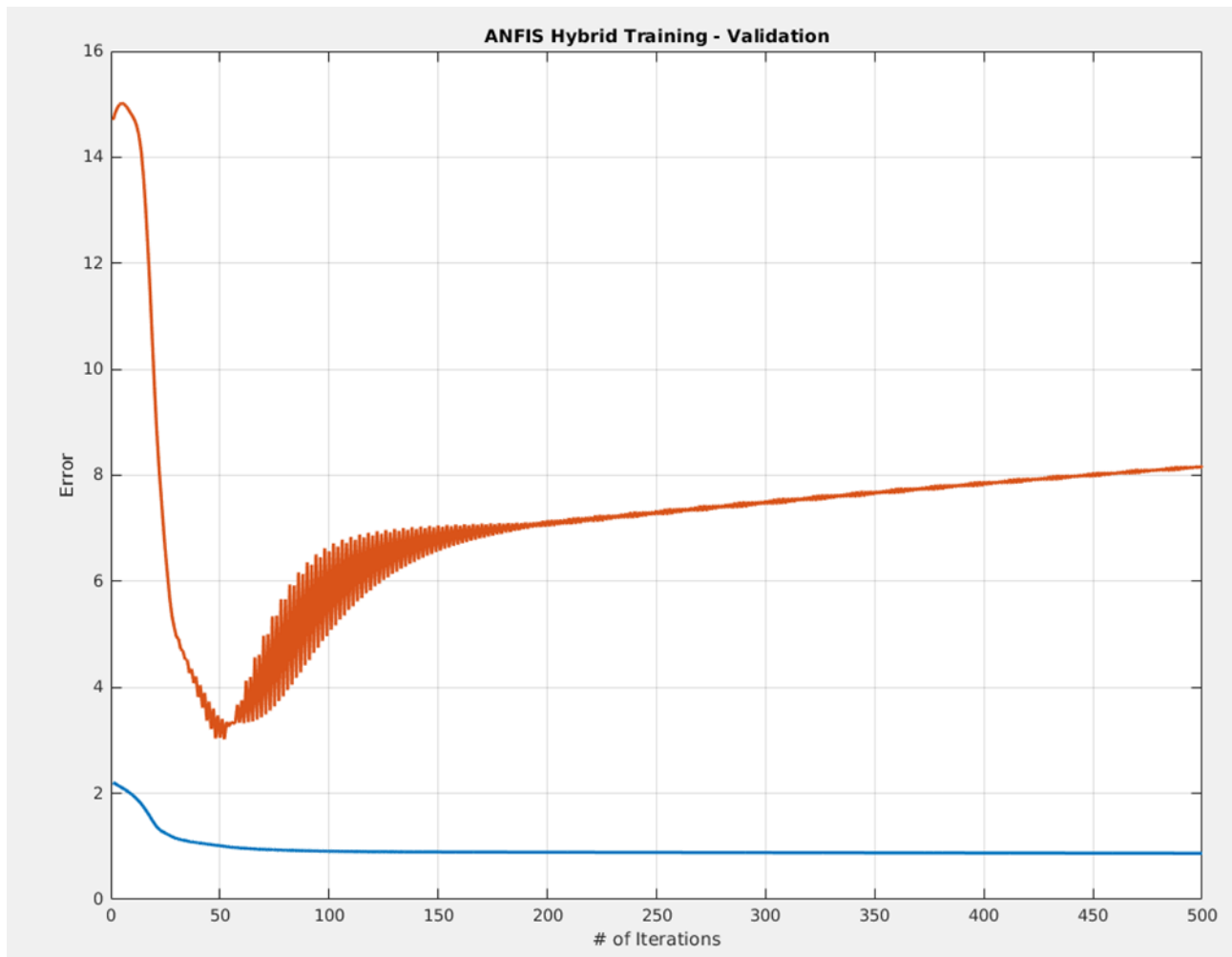


Figure 12: TSK\_model\_4 training and testing error.

Τέλος σύγκριση των μοντέλων.

	<b>R-squared</b>	<b>RMSE</b>	<b>NMSE</b>	<b>NDEI</b>
TSK_model_1	0.6708	3.8323	0.3292	0.5738
TSK_model_2	0.7526	3.3008	0.2474	0.4974
TSK_model_3	0.7983	3.1061	0.2017	0.4491
TSK_model_4	0.7282	3.5971	0.2718	0.5213

Table 1: Σύγκριση μοντέλων.

## Συμπεράσματα

Ο αριθμός των κανόνων που προκύπτουν εξαρτάται από τον αριθμό των χαρακτηριστικών του μοντέλου. Συγκεκριμένα, για 2 membership functions προκύπτουν (25) 32 κανόνες ενώ για 3 membership functions προκύπτουν (35) 243 κανόνες, οι δυνατοί συνδυασμοί, δηλαδή, των μεταβλητών εισόδου. Το τρίτο μοντέλο (TSK\_model\_3) παρουσιάζει την καλύτερη overall απόδοση.

## 2. Εφαρμογή σε dataset με υψηλή διαστασιμότητα

Το dataset που θα χρησιμοποιηθεί περιλαμβάνει 21263 δείγματα και 81 μεταβλητές/χαρακτηριστικά από τα οποία θα προβλέψουμε αυτά της τελευταίας στήλης. Αν διαμερίζαμε το χώρο εισόδου κάθε μεταβλητής (81) με δύο ασαφή σύνολα, θα καταλήγαμε με 281 κανόνες. Για το λόγο αυτό θα μειώσουμε τη διαστασιμότητα και τον αριθμό των if-then κανόνων μέσω της επιλογής χαρακτηριστικών (feature selection) και της χρήσης διαμέρισης διασκορπισμού. Ακολουθώντας τη μέθοδο αυτή προκύπτουν δύο (2) ελεύθερες παράμετροι, το πλήθος των χαρακτηριστικών που θα ληφθούν υπ' όψη στην εκπαίδευση και το πλήθος των ομάδων που θα δημιουργηθούν, οι οποίες εξαρτώνται από το βαθμό επικάλυψης των συναρτήσεων εισόδου. Η επιλογή των τιμών των δύο παραμέτρων θα γίνει μέσω της αναζήτησης πλέγματος (grid search).

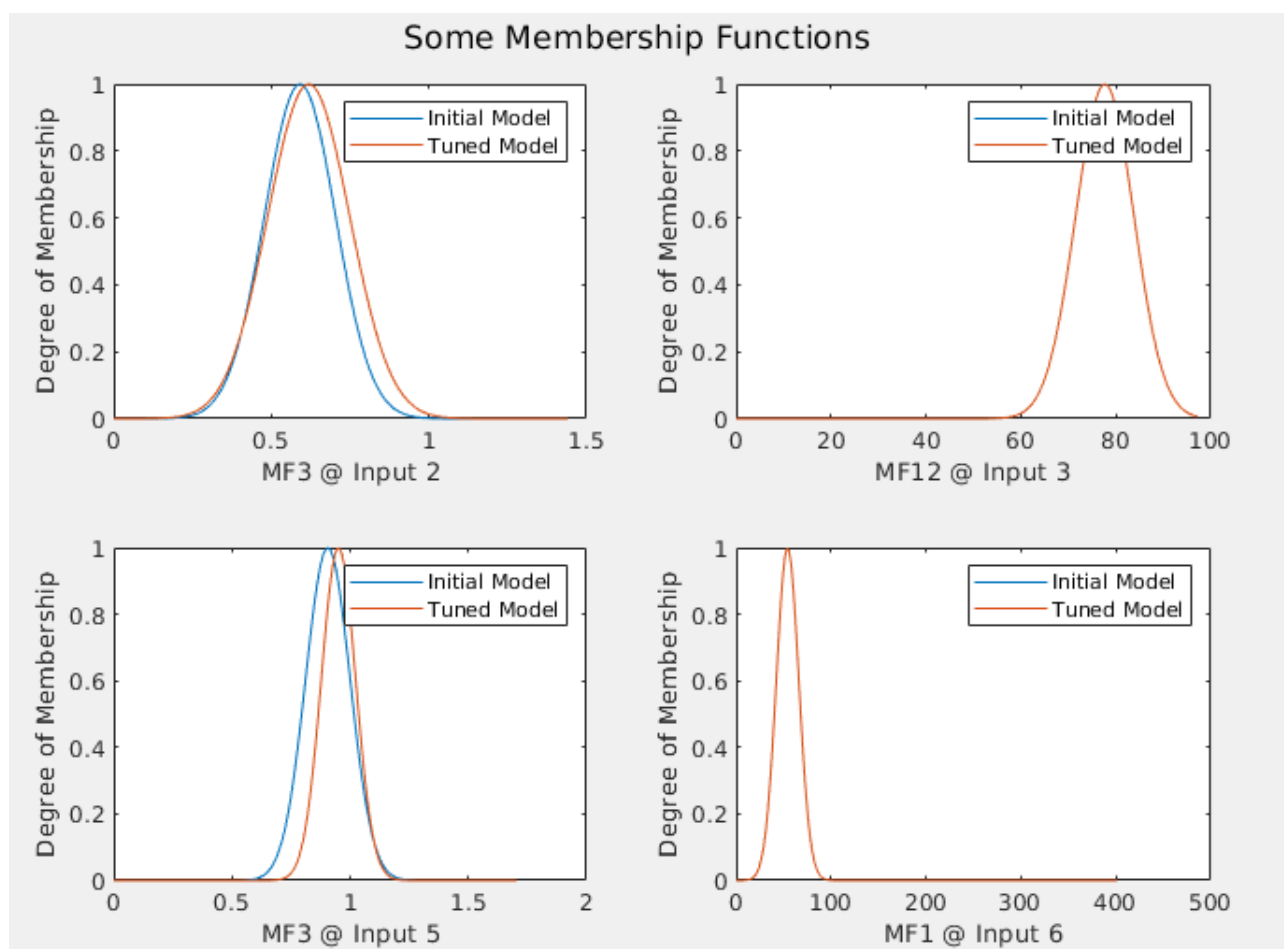


Figure13: Some membership of initial and final state of model.

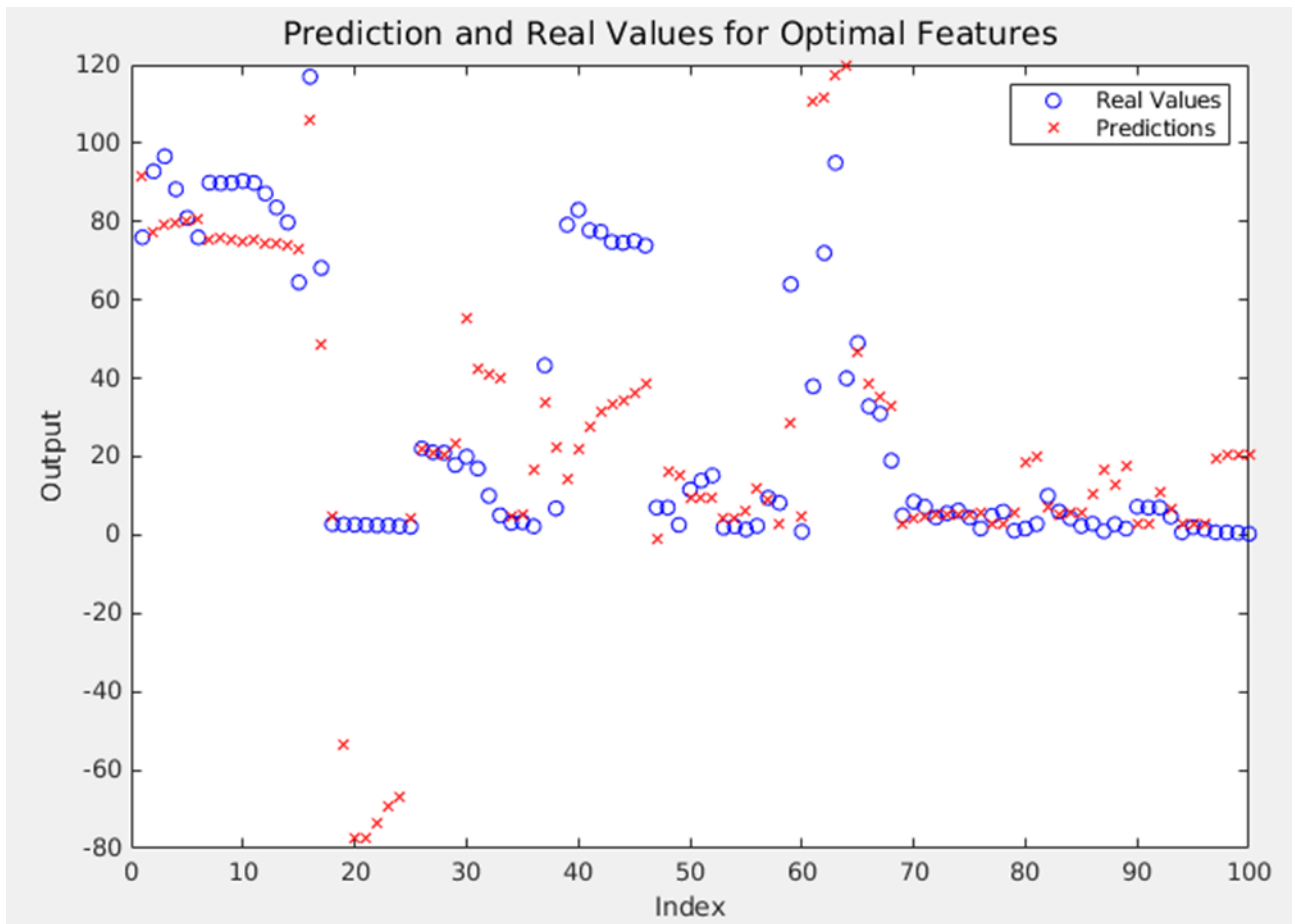


Figure 14: Real and prediction values.



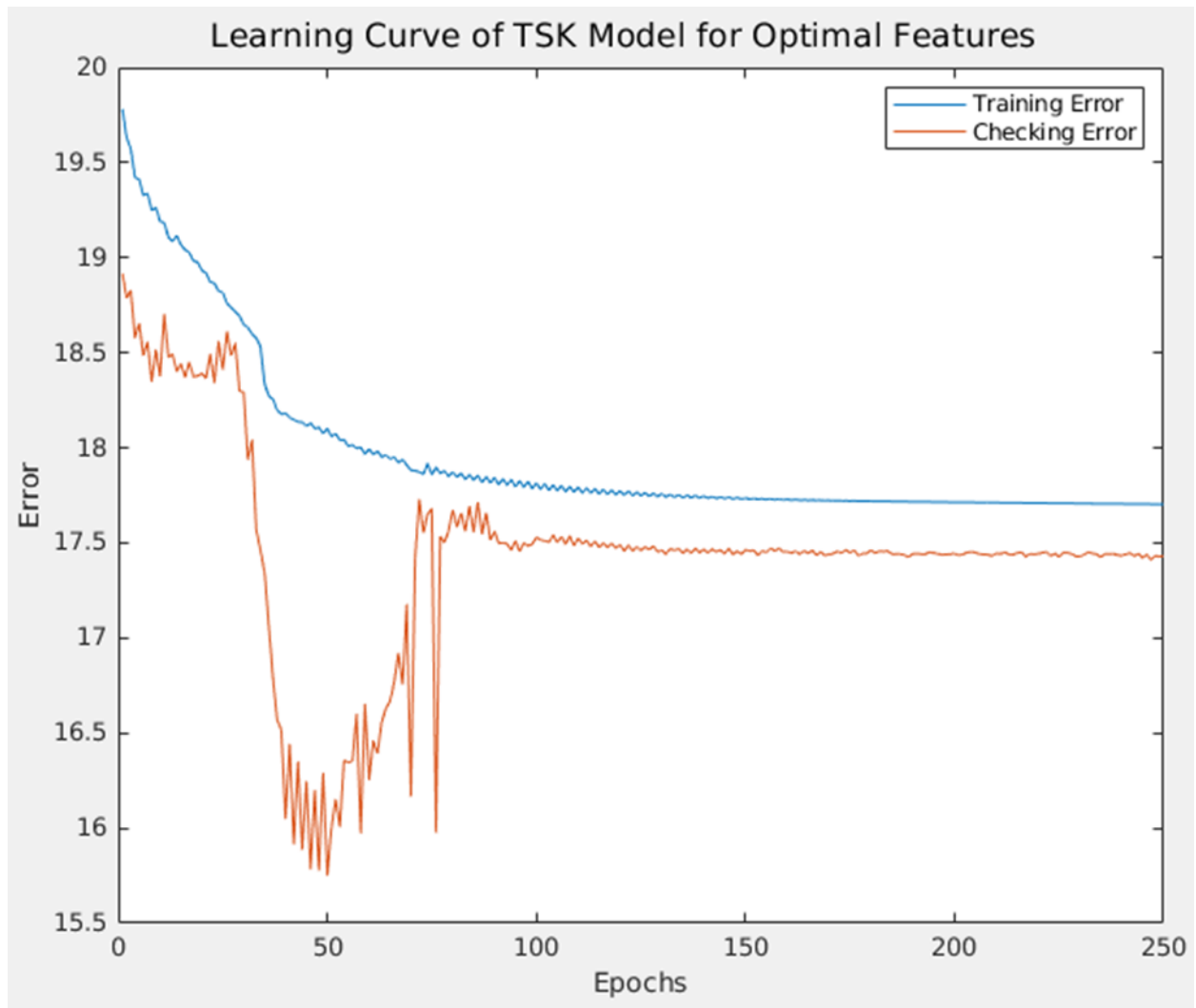


Figure 15: TSK\_model training and testing error.

Mean Square Error (MSE): 360.082399

Root Mean Square Error (RMSE): 18.975837

$R^2$ : 0.329646

Normalised Mean Square Error (NMSE): 0.670196

NDEI: 0.818655

## II. Ταξινόμηση

### 1. Εφαρμογή σε απλό dataset

Στην συγκεκριμένη εργασία καλούμαστε να επιλύσουμε ένα πρόβλημα ταξινόμησης με χρήση TSK μοντέλων. Το dataset που χρησιμοποιήθηκε αποτελείται από 3 χαρακτηριστικά, ενώ για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε η μέθοδος του Subtractive Clustering. Για την βέλτιστη επίλυση του προβλήματος η εκπαίδευση του μοντέλου πραγματοποιήθηκε για δύο διαφορετικές τιμές της ακτίνας των clusters (0.7 και 0.9). Επίσης, για κάθε τιμή της ακτίνας των clusters πραγματοποιήθηκαν 2 πειράματα, ένα με διαχωρισμό των κλάσεων και ένα χωρίς. Επιπλέον αξίζει να σημειωθεί ότι για όλα τα πειράματα πραγματοποιήθηκε κανονικοποίηση της τιμής εξόδου (αποτέλεσμα του ασαφούς νευρωνικού δικτύου) στο διάστημα  $[1, 2]$  ώστε να μας επιστραφεί η τιμή της κλάσης. Τέλος, για την εκπαίδευση του νευρωνικού δικτύου έτρεξε για 100 εποχές.

Για τα πειράματα που εκτελέστηκαν για τιμή ακτίνας κάθε cluster ίση με 0.7 έχουμε τα εξής αποτελέσματα και για τις δύο περιπτώσεις (Class Dependent, Class Independent):

- Class Independent (ακτίνα: 0.7)

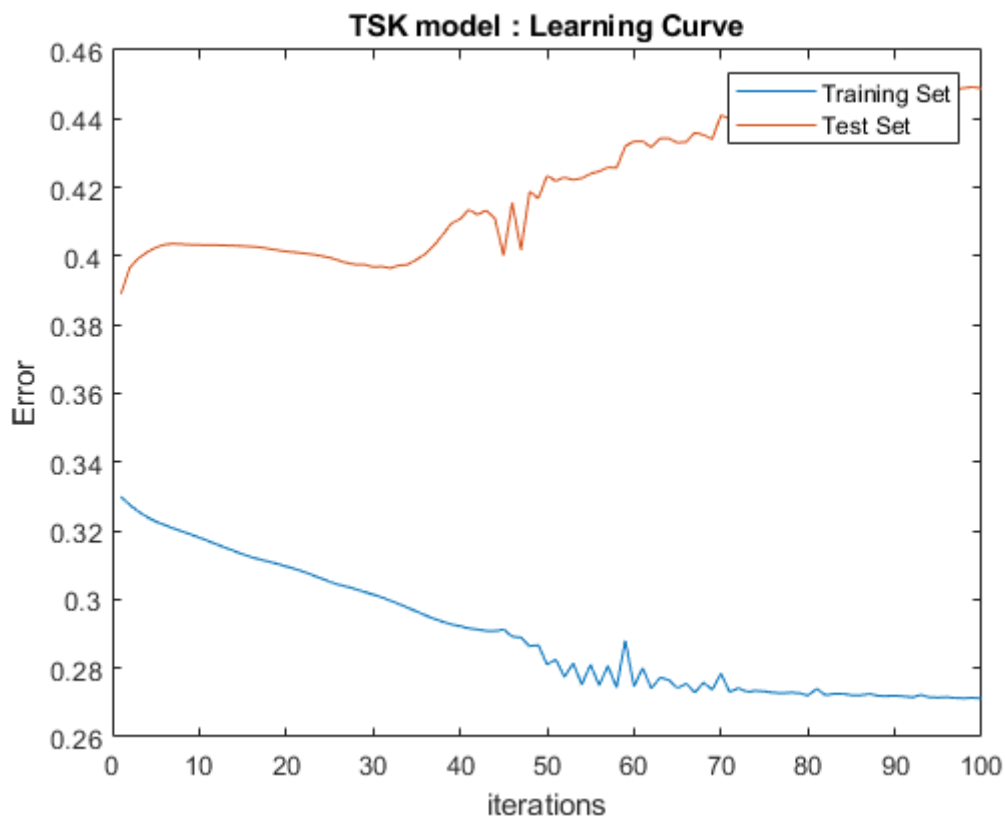


Figure 16: TSK\_model\_1 training and testing error.

Το συνολικό accuracy του μοντέλου είναι:  
Overall Accuracy = 84.3%

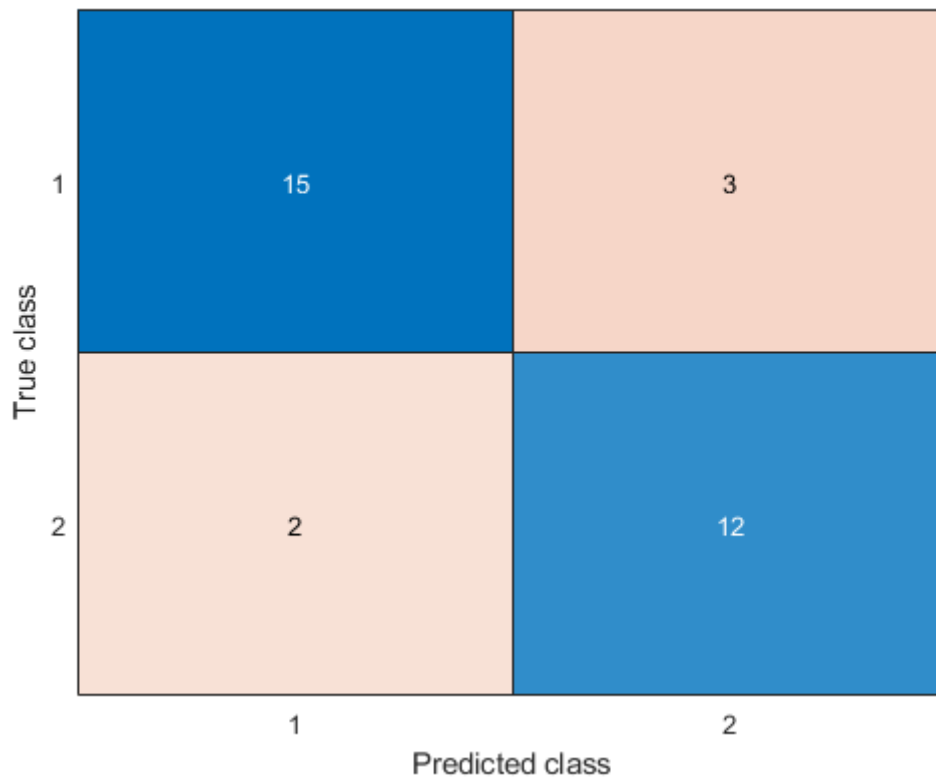
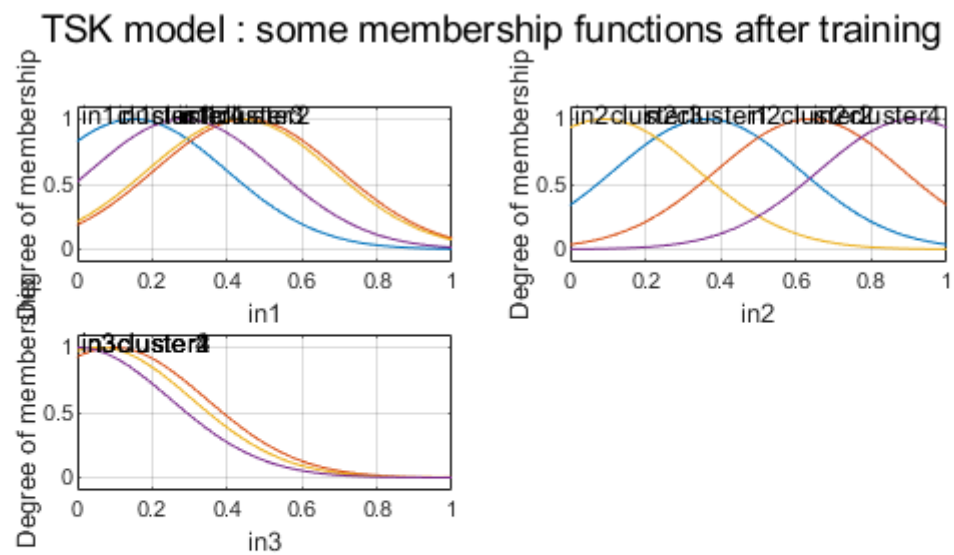


Figure 17: Confusion Matrix, Independent (ακτίνα: 0.7).

Μερικές συναρτήσεις συμμετοχής πριν την εκπαίδευση



Εικόνα 18: TSK\_model\_1 initial state.

Μερικές συναρτήσεις συμμετοχής μετά την εκπαίδευση

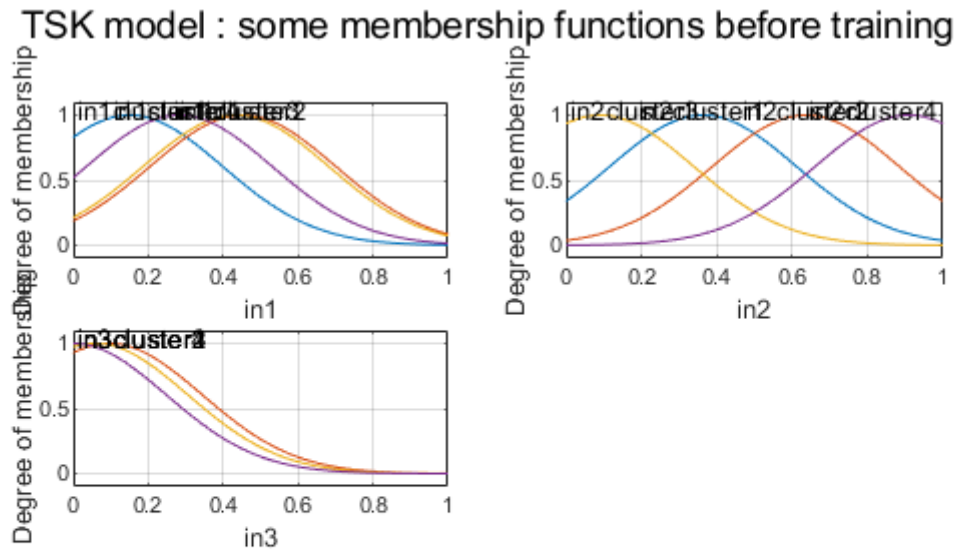


Figure 19: TSK\_model\_1 final state.

Παρακάτω ο πίνακας Producer's accuracy – User's accuracy

Class	UA (%)	PA (%)
1	0.8333	0.8824
2	0.8571	0.8

Table 2: User's και Producer's Accuracy.

Τέλος το μέγεθος κ είναι 0.6850

- Class Dependent (ακτίνα: 0.7)

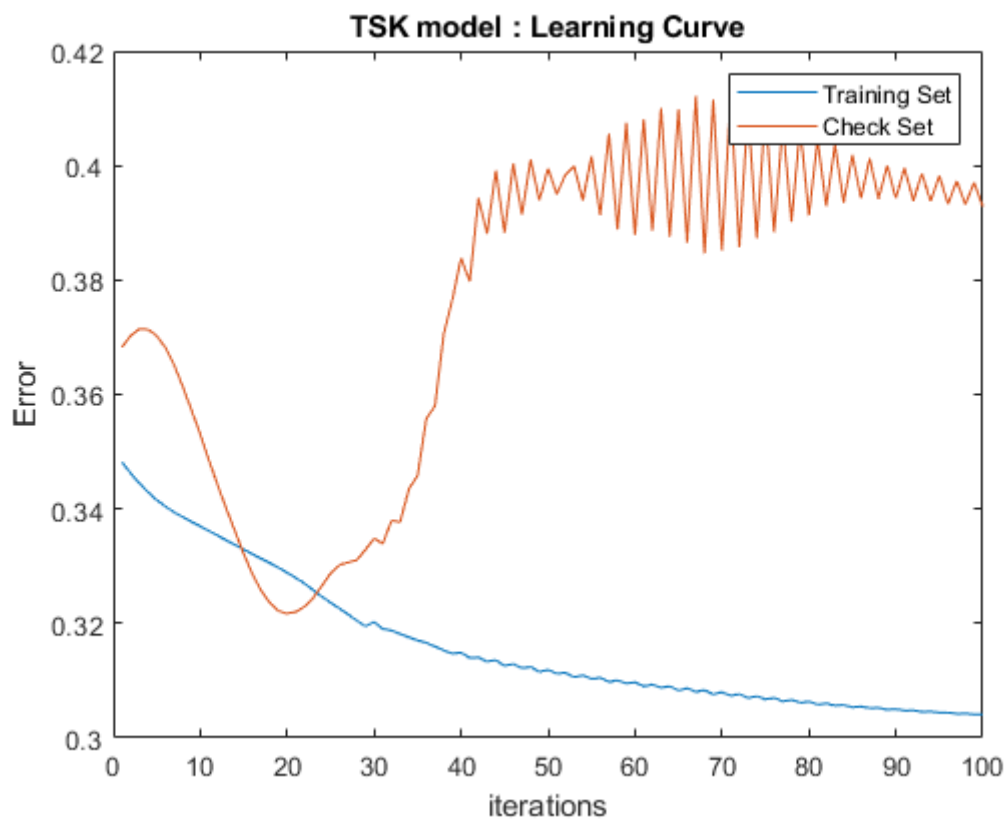


Figure 20: TSK\_model \_2 training and testing error.

Το συνολικό accuracy του μοντέλου είναι:  
Overall Accuracy = 87.5%

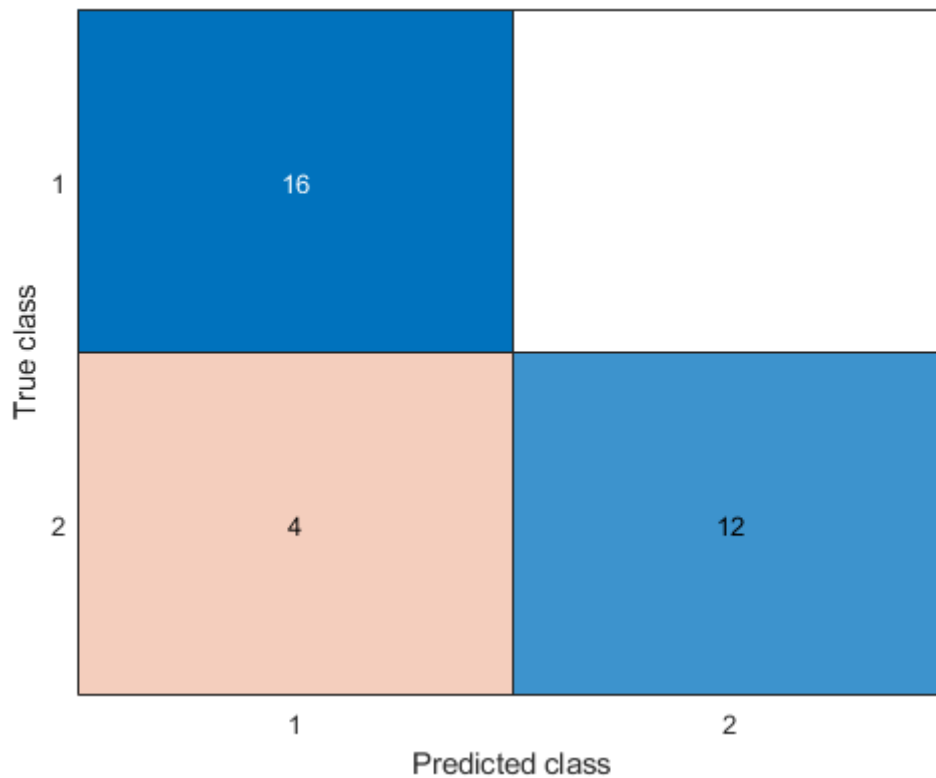


Figure 21: Confusion Matrix, Dependent (ακρίβια: 0.7).

Μερικές συναρτήσεις συμμετοχής πριν την εκπαίδευση

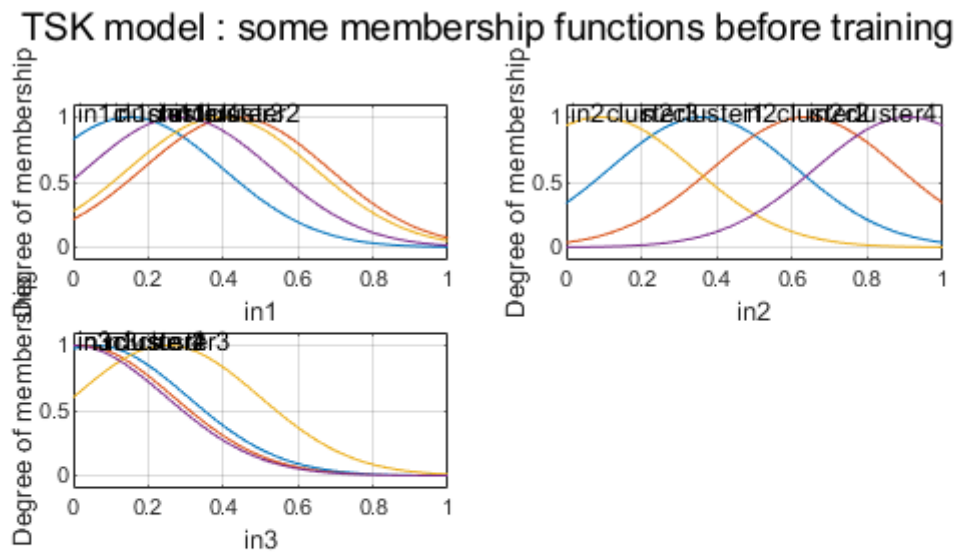


Figure 22: TSK\_model\_2 initial state.

Μερικές συναρτήσεις συμμετοχής μετά την εκπαίδευση

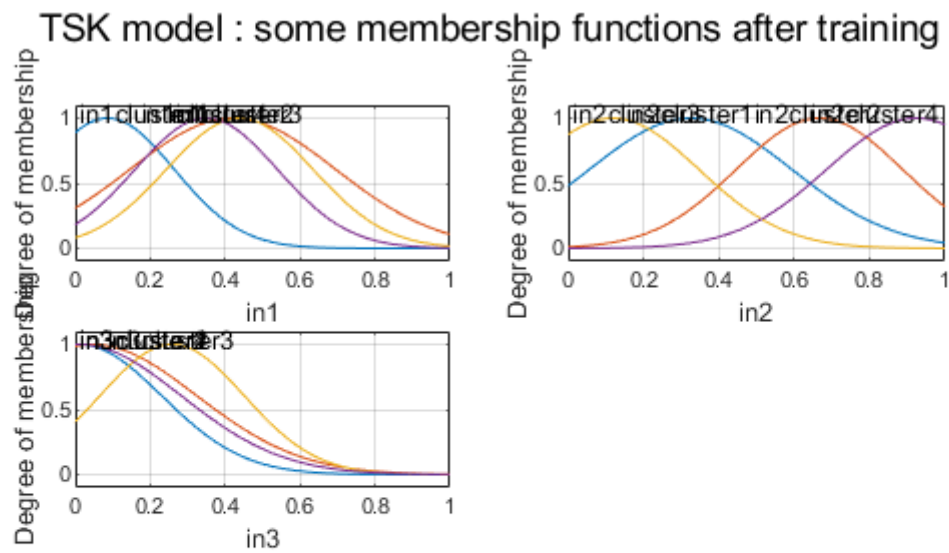


Figure 23: TSK\_model\_2 final state.

Παρακάτω ο πίνακας Producer's accuracy – User's accuracy

Class	UA (%)	PA (%)
1	1	0.8
2	0.75	1

Table 3: User's και Producer's Accuracy.

Τέλος το μέγεθος κ είναι 0.6850

- Class Independent (ακτίνα: 0.9)

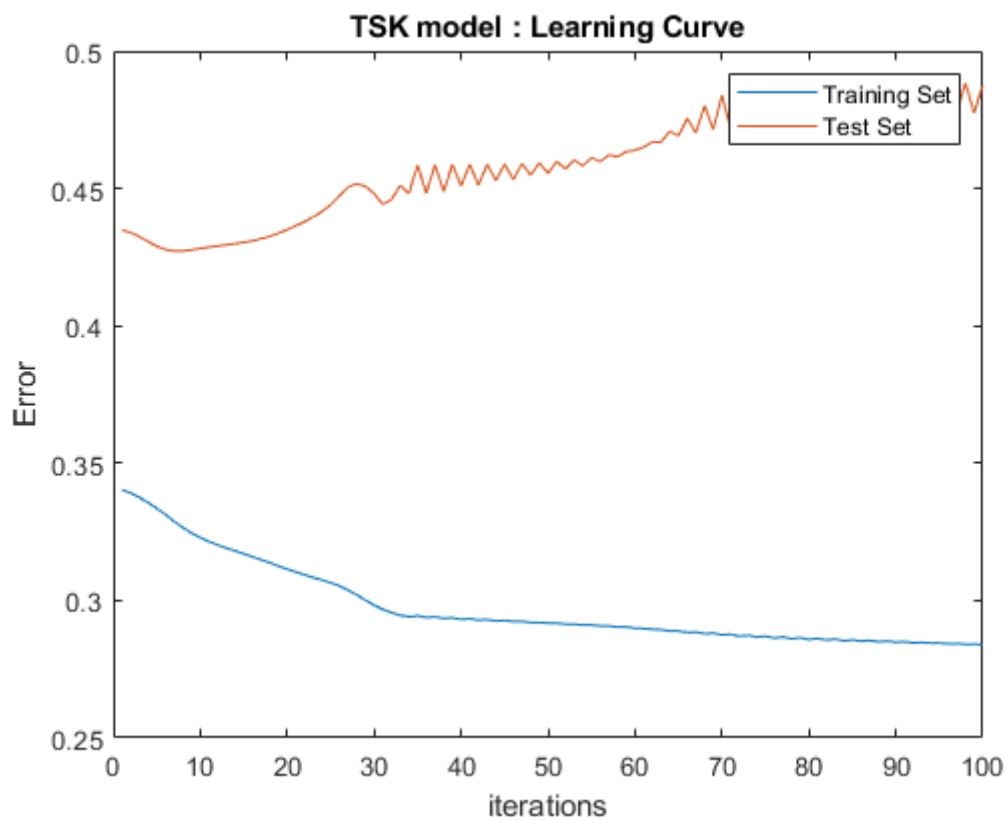


Figure 24: TSK\_model\_3 training and testing error.

Το συνολικό accuracy του μοντέλου είναι:  
Overall Accuracy = 84.3%

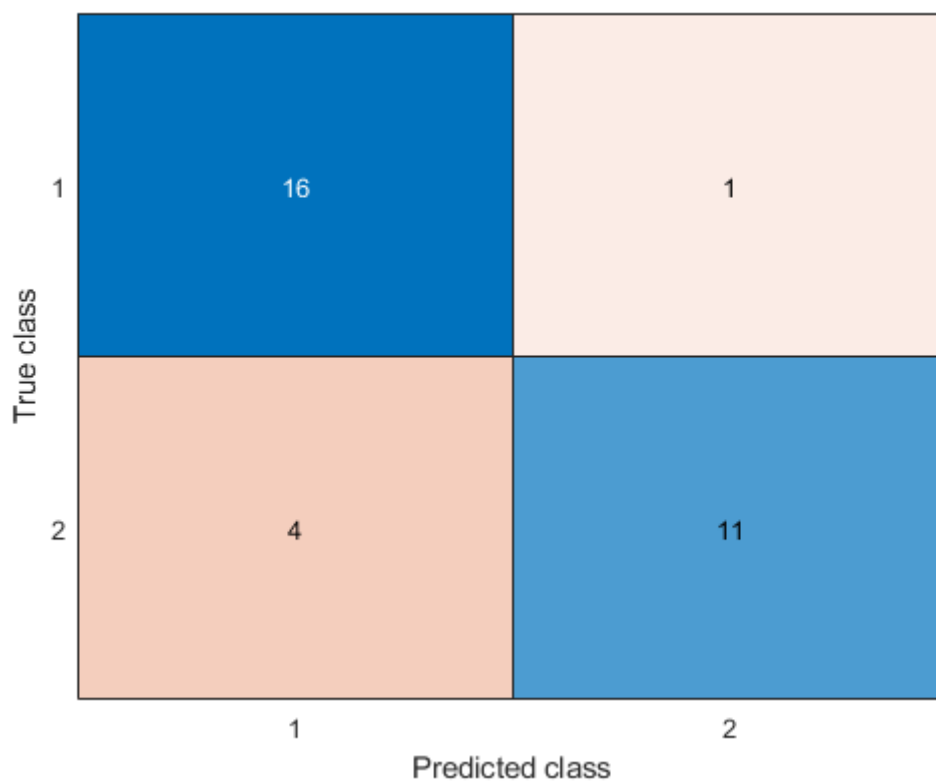


Figure 25: Confusion Matrix, Independent (ακτίνα: 0.9).



Μερικές συναρτήσεις συμμετοχής πριν την εκπαίδευση

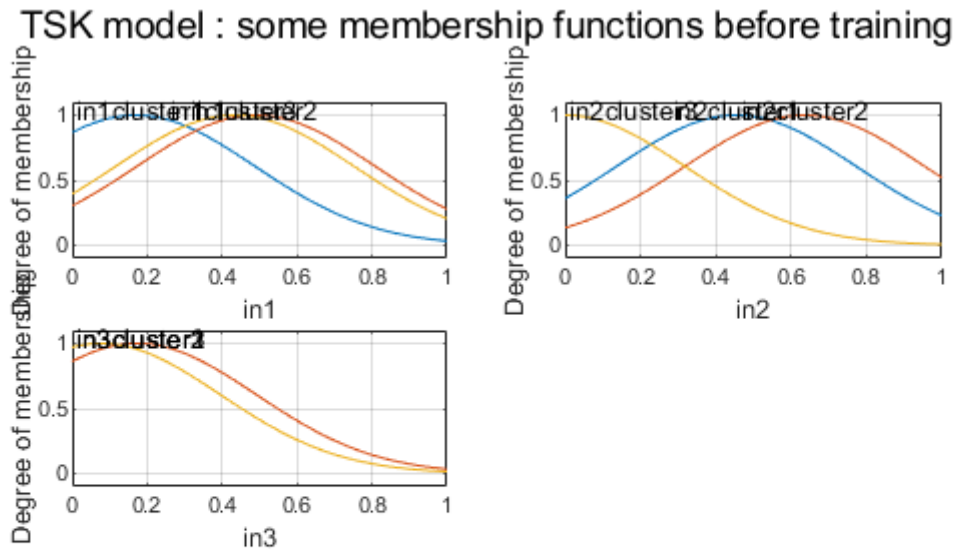


Figure 26: TSK\_model\_3 initial state.

Μερικές συναρτήσεις συμμετοχής μετά την εκπαίδευση

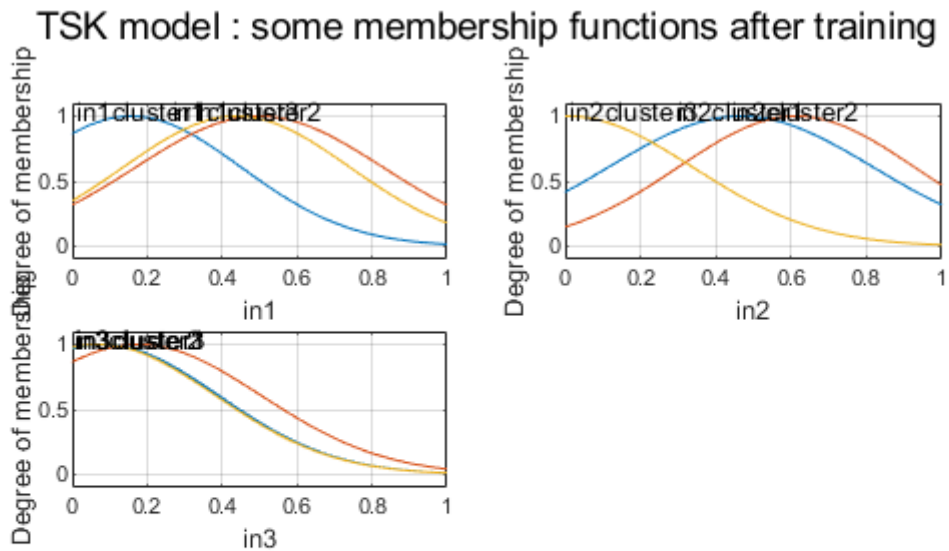


Figure 27: TSK\_model\_4 final state.

Παρακάτω ο πίνακας Producer's accuracy – User's accuracy

Class	UA (%)	PA (%)
1	0.94	0.8
2	0.7333	0.9167

Πίνακας 4: User's και Producer's Accuracy.

Τέλος το μέγεθος κ είναι 0.75

- Class Dependent (ακτίνα: 0.9)

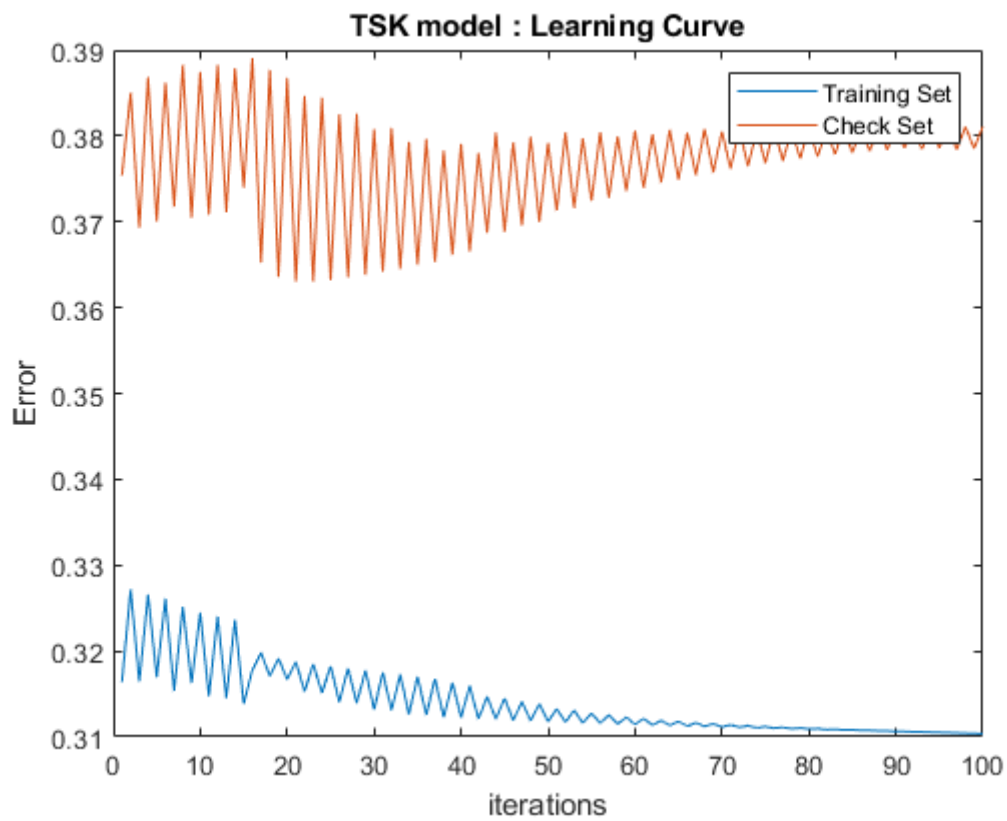


Figure 28: TSK\_model\_4 training and testing error.

Το συνολικό accuracy του μοντέλου είναι:

Overall Accuracy = 81.25%

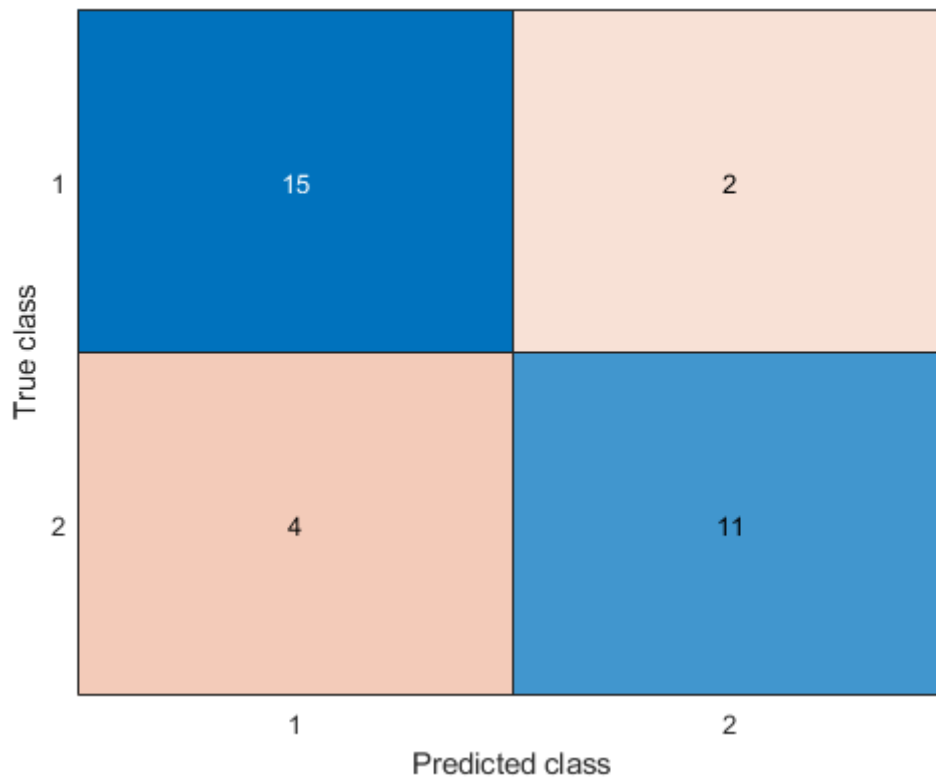


Figure 29: Confusion Matrix, Dependent (ακρίβια: 0.9).

Μερικές συναρτήσεις συμμετοχής πριν την εκπαίδευση

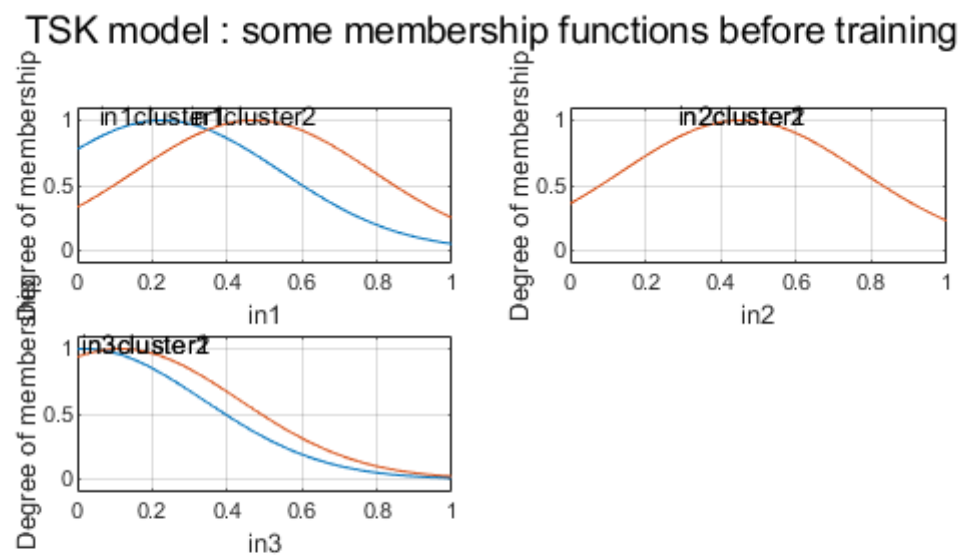
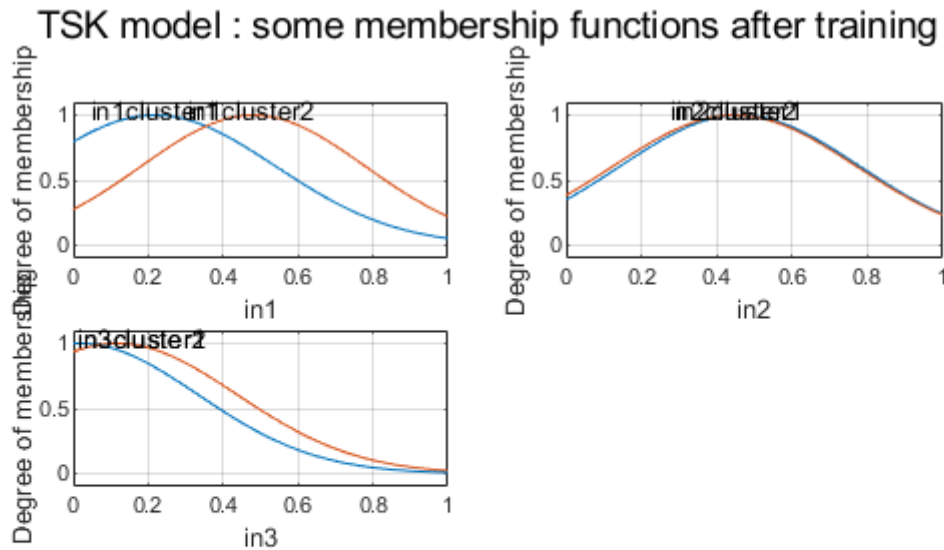


Figure 30: TSK\_model\_4 initial state.

Μερικές συναρτήσεις συμμετοχής μετά την εκπαίδευση



Εικόνα 31: TSK\_model\_4 final state.

Παρακάτω ο πίνακας Producer's accuracy – User's accuracy

Class	UA (%)	PA (%)
1	0.8824	0.7895
2	0.7333	0.8462

Table 5: User's και Producer's Accuracy.

Τέλος το μέγεθος  $\kappa$  είναι 0.6206

## 2. Εφαρμογή σε dataset με υψηλή διαστασιμότητα

Στο δεύτερο κομμάτι της συγκεκριμένης εργασίας καλούμαστε να υλοποιήσουμε έναν ταξινομητή. Ωστόσο, το dataset που δίνεται για την υλοποίηση του συγκεκριμένου ταξινομητή αποτελείται από 11500 δείγματα με 179 χαρακτηριστικά ανά δείγμα. Είναι φανερό ότι η υπολογιστική ισχύς που απαιτείται για την επίλυση του συγκεκριμένου προβλήματος με απλή εφαρμογή ασαφούς νευρωνικού δικτύου είναι εξαιρετικά μεγάλη. Για τον παραπάνω λόγο, υλοποιήθηκε αναζήτηση πλέγματος (grid search) ώστε να οριστεί ο βέλτιστος συνδυασμός της τιμής της ακτίνας των clusters και ο αριθμός των χαρακτηριστικών που θα χρησιμοποιηθούν.

Μέσω της αναζήτησης πλέγματος πραγματοποιήθηκαν οι εξής έλεγχοι:

- [10,15,19,20] για τον αριθμό των χαρακτηριστικών που θα χρησιμοποιηθούν.
- [0.3,0.5,0.7,0.9] για το μέγεθος της ακτίνας των clusters.

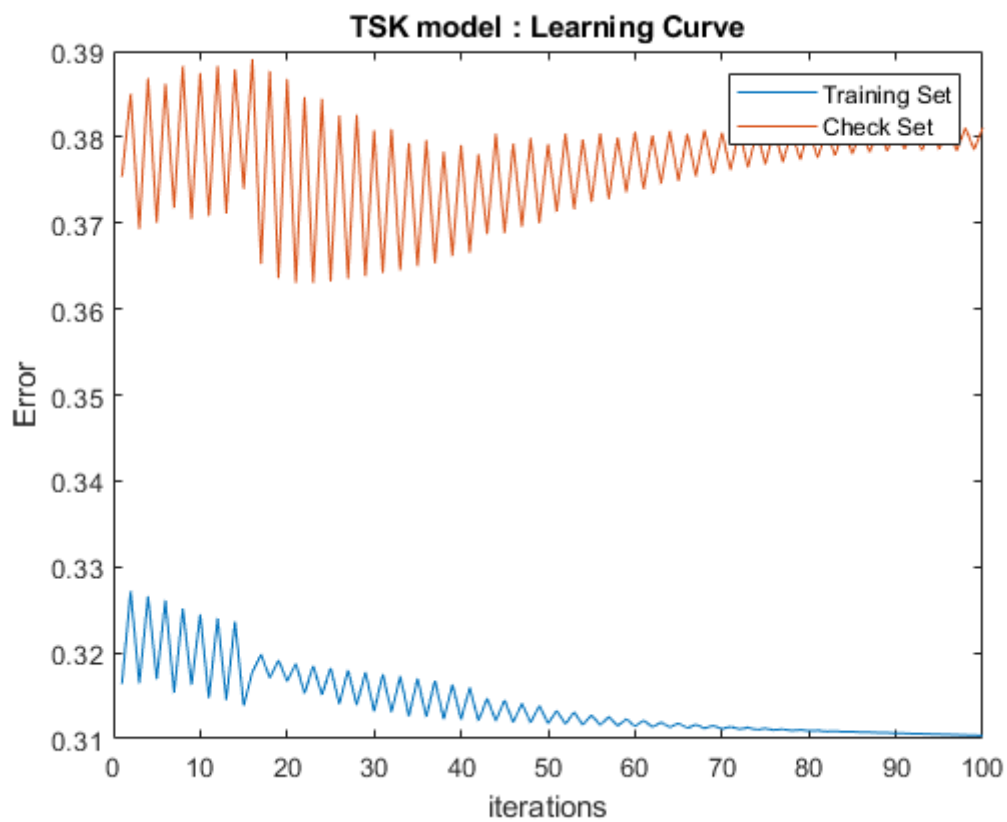


Figure 32: TSK\_model training and testing error.

Το συνολικό accuracy του μοντέλου είναι:  
Overall Accuracy = 42.1304%

Μερικές συναρτήσεις συμμετοχής πριν την εκπαίδευση

## Final TSK model : some membership functions before training

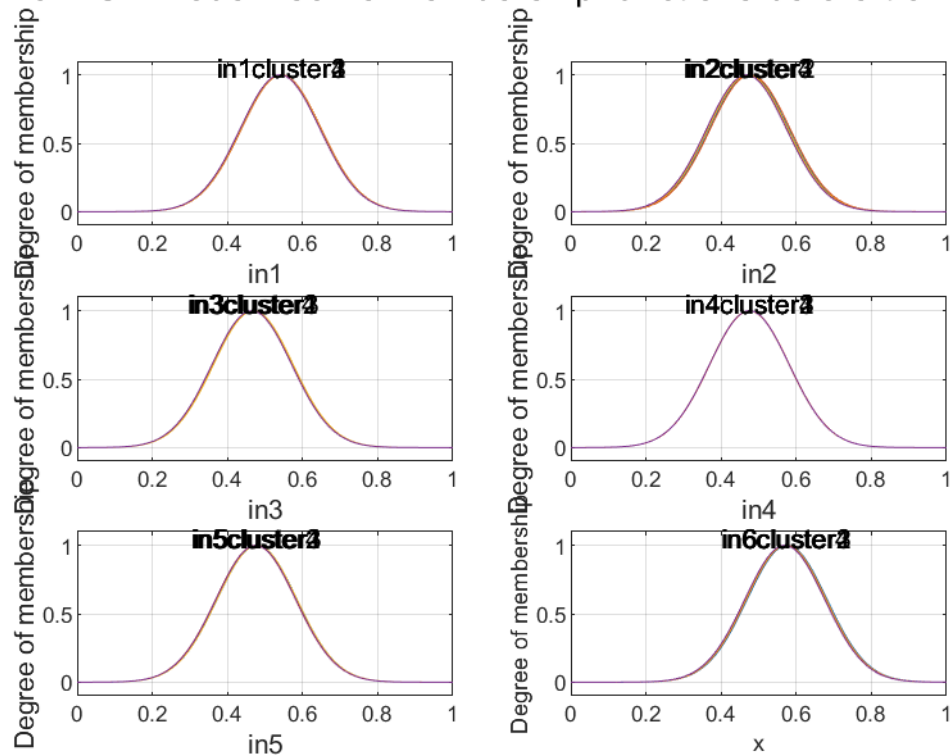


Figure 33: TSK\_model initial state.

Μερικές συναρτήσεις συμμετοχής μετά την εκπαίδευση

## Final TSK model : some membership functions after training

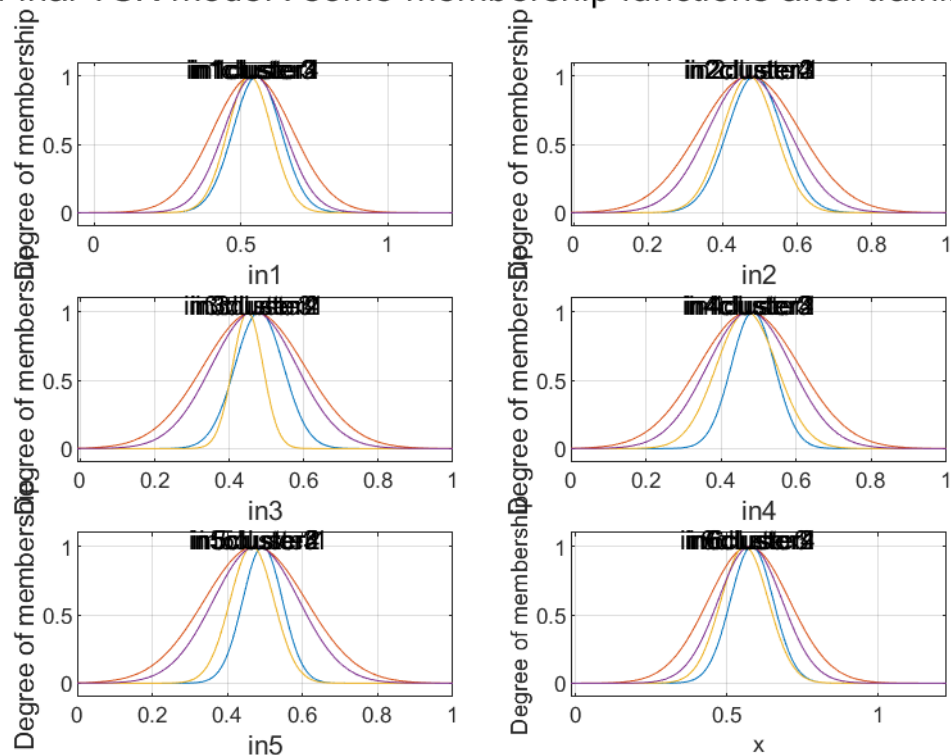


Figure 34: TSK\_model final state.

Παρακάτω ο πίνακας Producer's accuracy – User's accuracy

Class	UA (%)	PA (%)
1	0	0
2	0.7759	0.9397

Table 6: User's και Producer's Accuracy.

Τέλος το μέγεθος κ είναι 0.1311

## Συμπεράσματα

Τα διάφορα μοντέλα που αναφέρονται παραπάνω παρατηρείται ότι έχουν σχετικά καλό ποσοστό επιτυχίας πρόβλεψης εκτός από την τελευταία περίπτωση. Για την μείωση των διαφόρων μετρικών σφάλματος όπως και την αύξηση του ποσοστού επιτυχίας θα μπορούσαμε να τρέξουμε διάφορα πειράματα για περισσότερες εποχές, αρκεί βέβαια να μην εμφανιστεί φαινόμενο υπερ-εκπαίδευσης. Επιπλέον, κατά την υλοποίηση των διαφόρων μοντέλων παρατηρήθηκε ότι όσο περισσότερο αυξάνεται ο αριθμός των κανόνων, τόσο αυξάνεται και η προ-απαιτούμενη υπολογιστική ισχύς, ωστόσο, παρατηρήθηκε ραγδαία αύξηση της προβλεπτικής ικανότητας των μοντέλων.