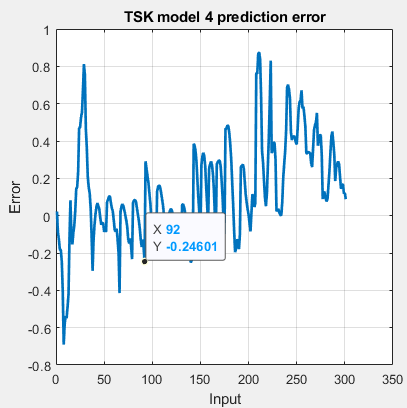
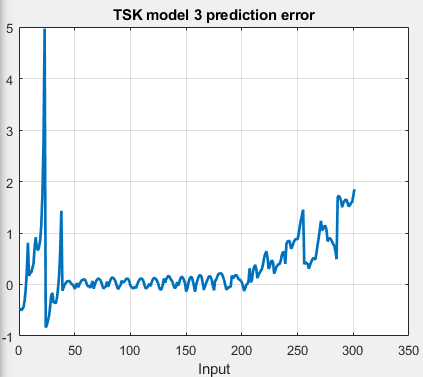
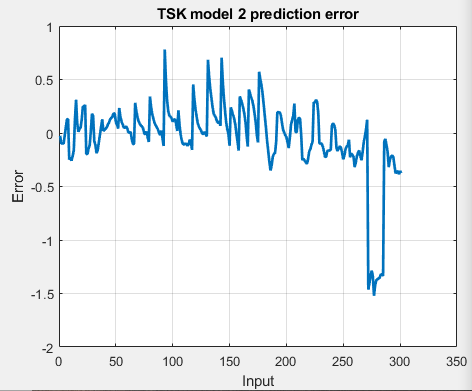
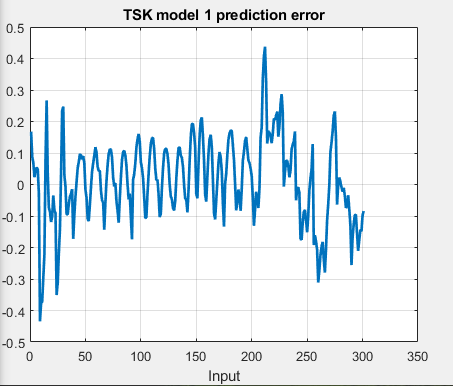
**Επίλυση προβλήματος ταξινόμησης με χρήση μοντέλων TSK.**

*Τριφηνόπουλος Χρήστος*

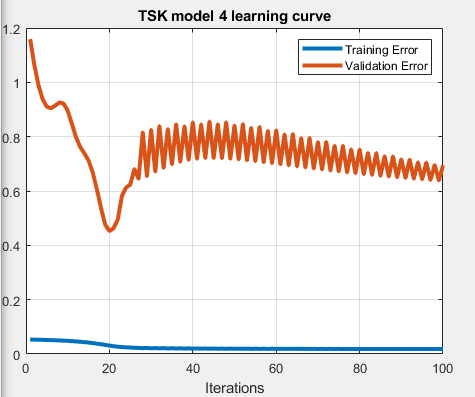
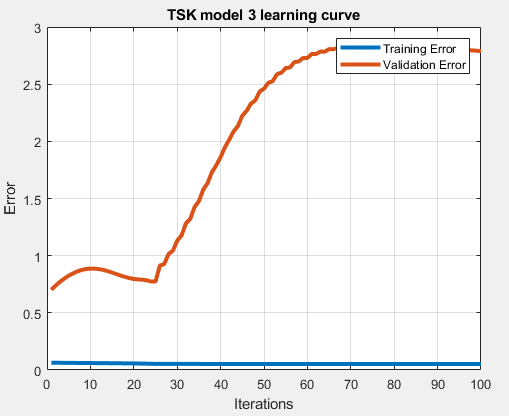
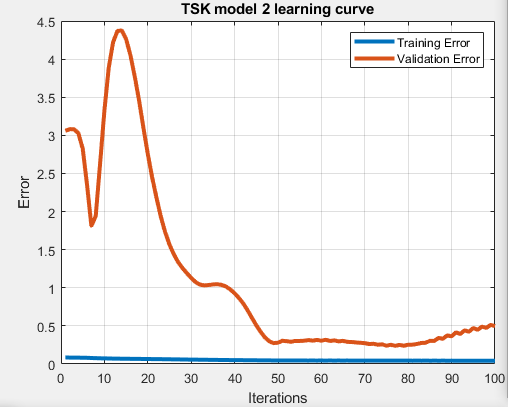
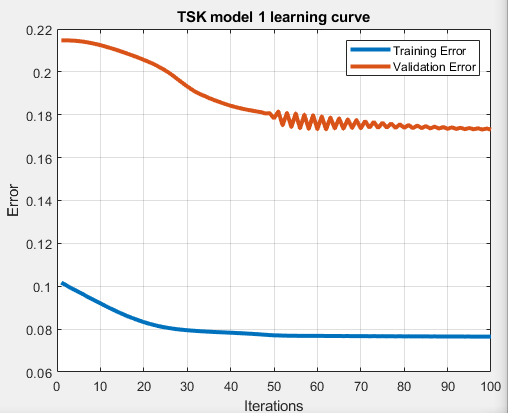
**Μέρος πρώτο: Απλό dataset**

Για αυτό το κομμάτι χρησιμοποιήθηκε το airfoil self-noise dataset από το UCI repository που περιλαμβάνει 1503 observations και 6 features. Κάθε στήλη του dataset αντιπροσωπεύει ένα feature, οι πέντε πρώτες είναι predictor variables και η τελευταία είναι η μεταβλητή-στόχος που προσπαθούμε να προβλέψουμε με τα μοντέλα μας. Εδώ μελετάμε 4 διαφορετικά μοντέλα TSK και τα συγκρίνουμε ως προς τα σφάλματα που δίνουν.

**Σφάλμα μοντέλων για όλα τα inputs**



**Learning curves των μοντέλων**



**Μέρος δεύτερο: Complex dataset**

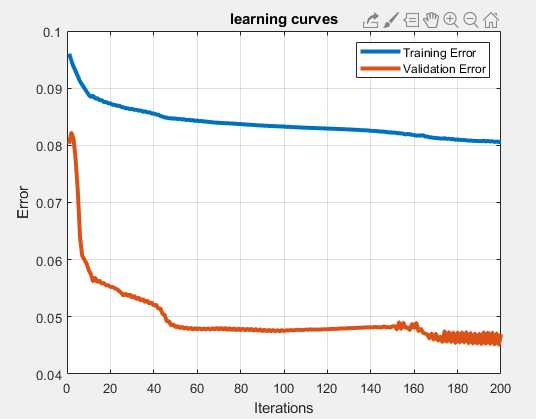
Το μέσο σφάλμα για κάθε συνδυασμό από παραμέτρους προέκυψε από το 5-fold cross validation:

|  |  |  |  |
| --- | --- | --- | --- |
| Features/radius | 0.2 | 0.4 | 0.8 |
| 10 | 0.0681 | 0.0680 | 0.0746 |
| 15 | 0.0683 | 0.0677 | 0.0707 |
| 20 | 0.0672 | 0.0637 | 0.0687 |

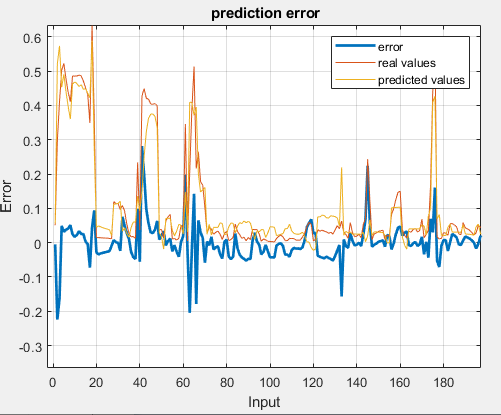
Οι βέλτιστες τιμες που προέκυψαν είναι:

* The minimum error among the models is **0.063696**
* Optimal radius for clusters is **0.4**
* Optimal amount of predictor variables is **20**

Για το βέλτιστο μοντέλο με εκπαίδευση σε 200 epochs παίρνουμε:



Σφάλμα testing για τα πρώτα 200 inputs.



**Μετρικές για το τελικό μοντέλο:**

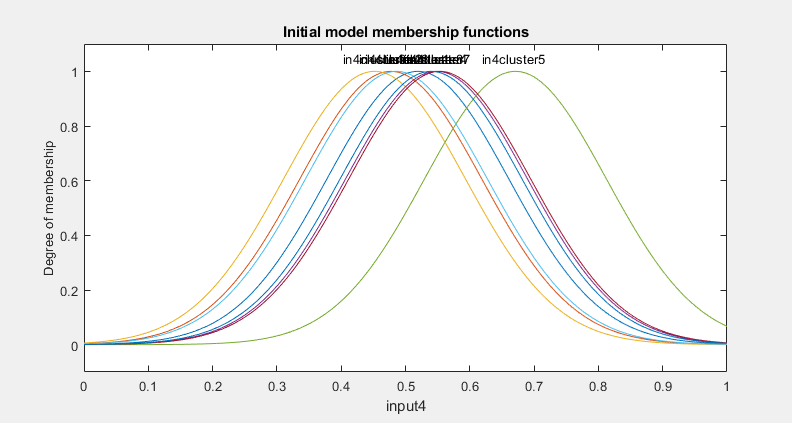
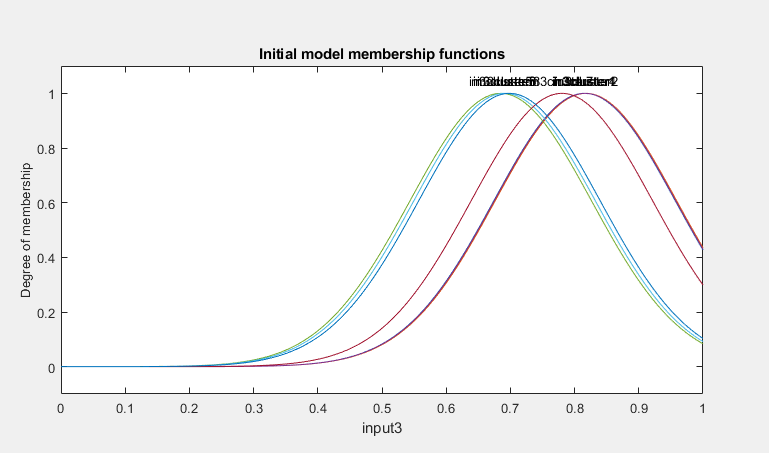
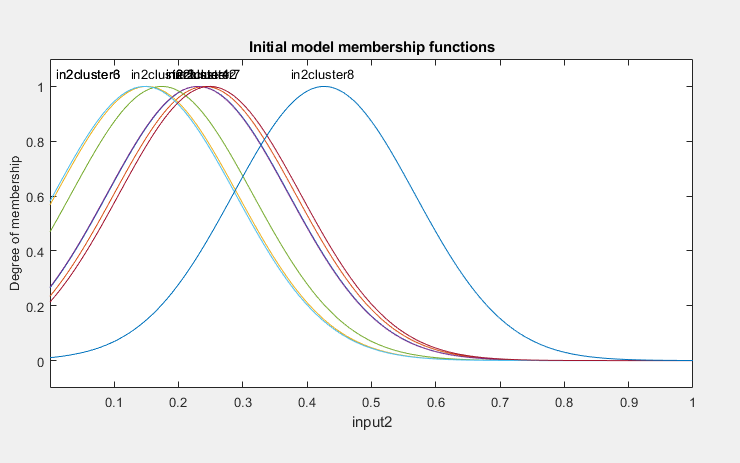
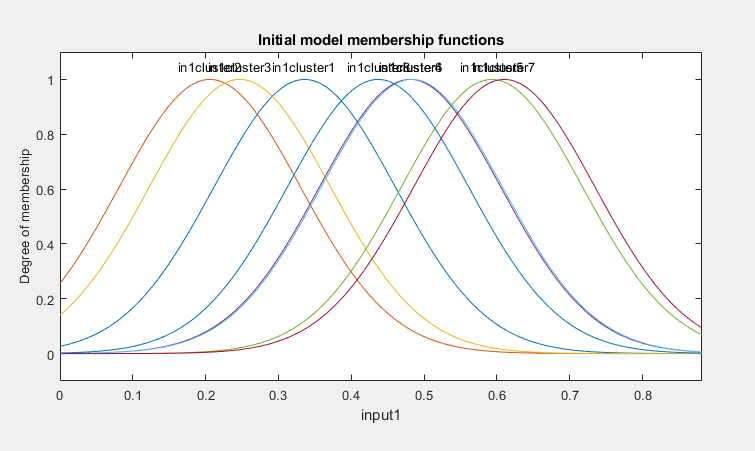
*Rsquare is* ***0.485833***

*RMSE is* ***0.089838***

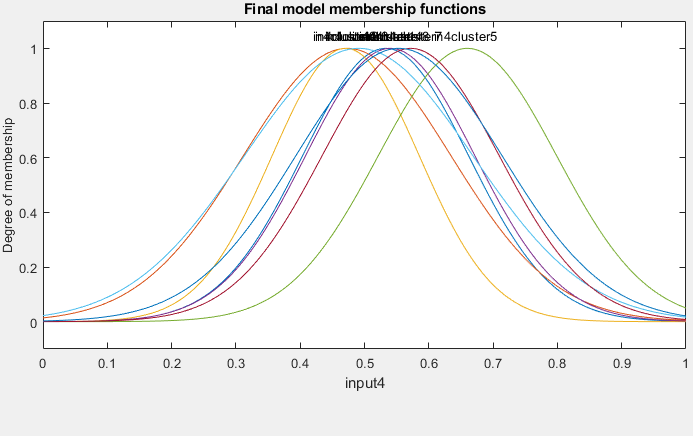
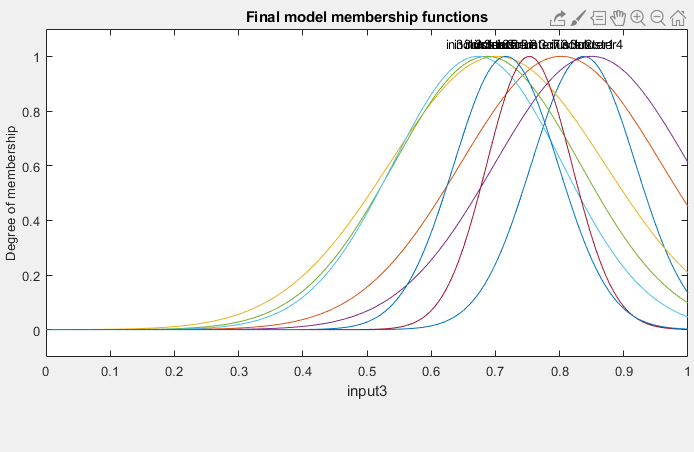
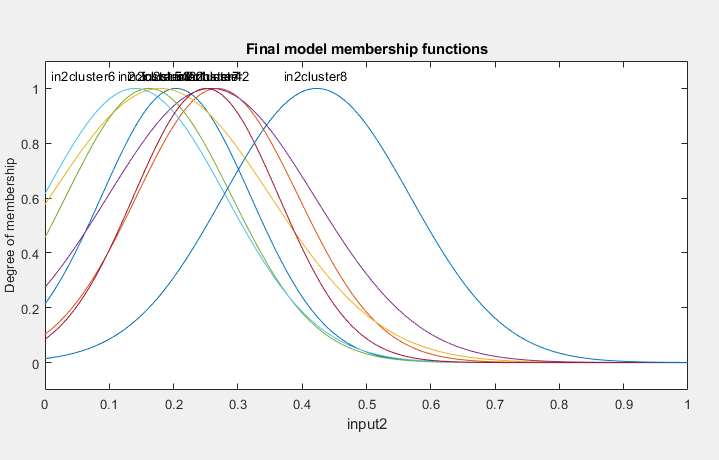
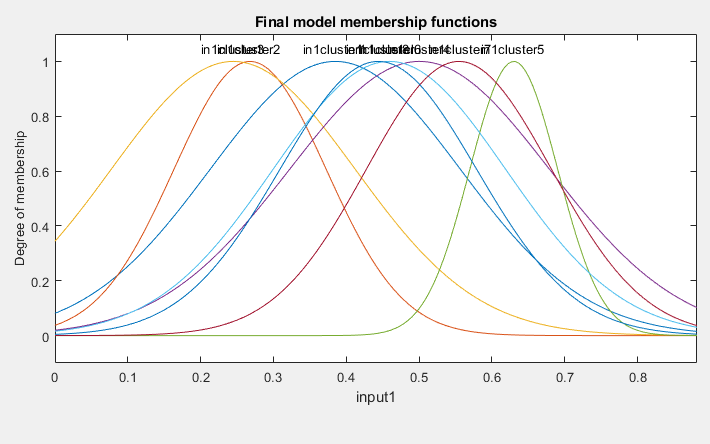
*NMSE is* ***0.514167***

*NDEI is* ***0.717054***

**Συναρτήσεις συμμετοχής για το αρχικό μοντέλο.**



**Συναρτήσεις συμμετοχής για το τελικό εκπαιδευμένο μοντέλο.**



Βλέπουμε ότι το βέλτιστο μοντέλο περιέχει 8 κανόνες.

Η διαφορά μεταξύ grid partitioning και subtractive clustering ως προς τους κανόνες που παράγονται είναι ότι στο πρώτο έχουμε έναν κανόνα για κάθε συνδυασμό από συναρτήσεις συμμετοχής εισόδου ενώ στο δεύτερο έχουμε έναν κανόνα για κάθε cluster. Αυτό σημαίνει ότι για grid partitioning θα είχαμε 20 εισόδους με 2 ή 3 ασαφή σύνολα για την κάθε μία και στην συνέχεια θα παίρναμε τους συνδυασμούς αυτών για να παράξουμε τους fuzzy κανόνες οι οποίοι θα ήταν πολύ περισσότεροι από 8 (2^20 ή 3^20 αντίστοιχα αφού για κάθε μία από τις 20 εισόδους έχουμε 3 πιθανές τιμές και πρέπει να πάρουμε όλους τους πιθανούς συνδυασμούς αυτών). Προφανώς κάτι τέτοιο δεν θα ήταν πρακτικό για μεγάλους αριθμούς εισόδων.