

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Υπολογιστική Νοημοσύνη

Εργασία 2

Επίλυση προβλήματος παλινδρόμησης με χρήση μοντέλων TSK

Κούτση Χριστίνα

AEM: 9871

email: cvkoutsis@ece.auth.gr

Οκτώβριος, 2022

1 Εφαρμογή σε απλό **Dataset**

Στο πρώτο ερώτημα της εργασίας μας ζητείται η δημιουργία τεσσάρων TSK μοντέλων για την επίλυση του προβλήματος παλινδρόμησης του **Airfoil Self-Noise dataset** του **UCI repository**. Τα 4 TSK μοντέλα θα έχουν τα εξής χαρακτηριστικά:

	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

Πίνακας 1: Ταξινόμηση μοντέλων προς εκπαίδευση.

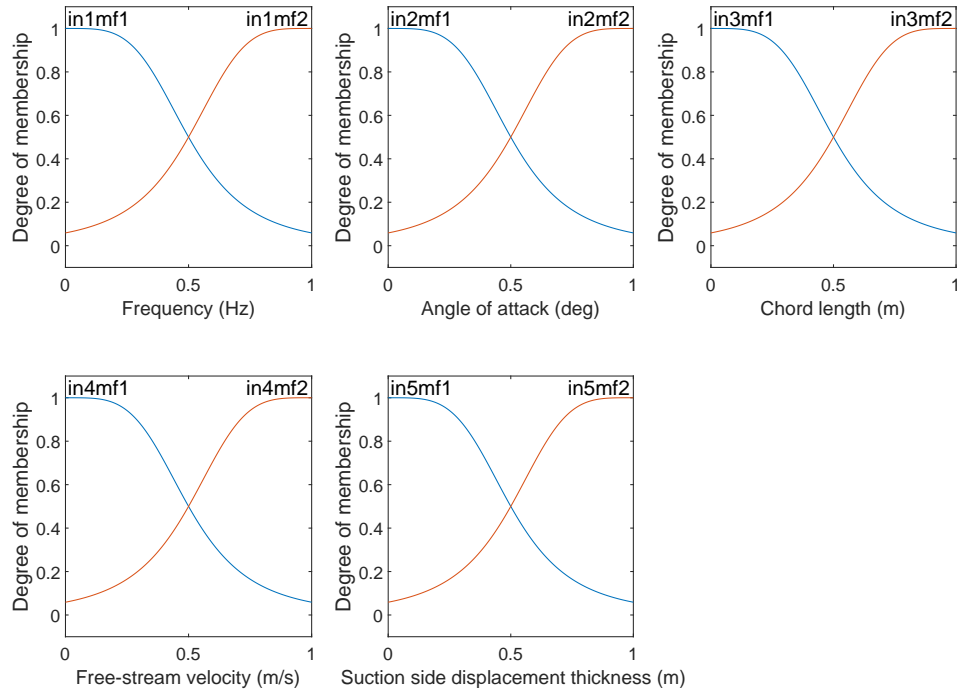
Για κάθε μοντέλο θα παρουσιάζονται:

1. Τα ασαφή σύνολα πριν και μετά την διαδικασία εκπαίδευσης
2. Το διάγραμμα μάθησης του μοντέλου
3. Το διάγραμμα σφάλματος πρόβλεψης του μοντέλου
4. οι τιμές των μετρικών RMSE, NMSE, NDEI, R^2

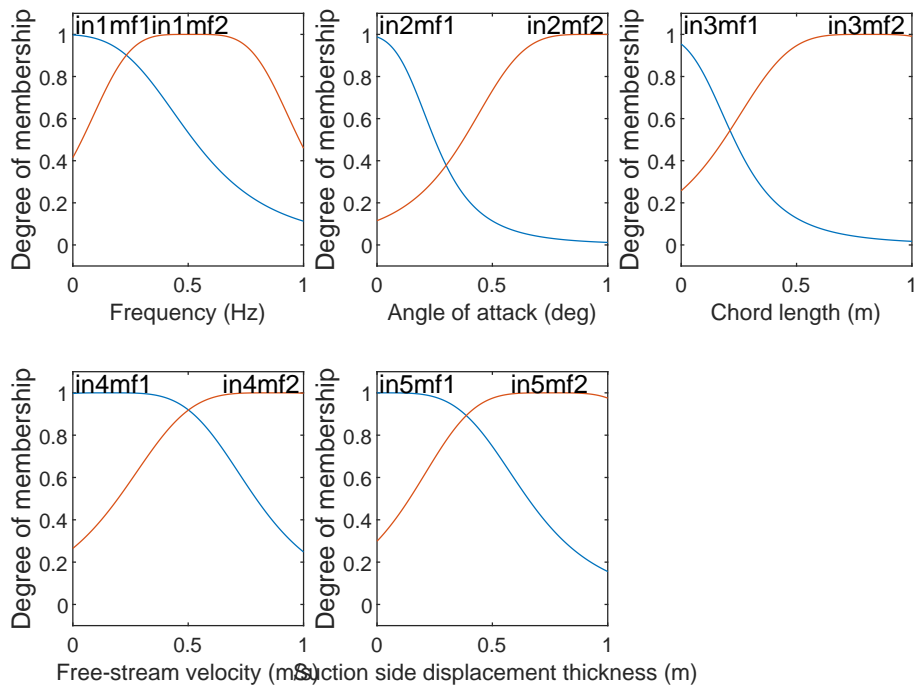
1.1 Μοντέλο 1

1. Ασαφή σύνολα πριν και μετά την διαδικασία εκπαίδευσης

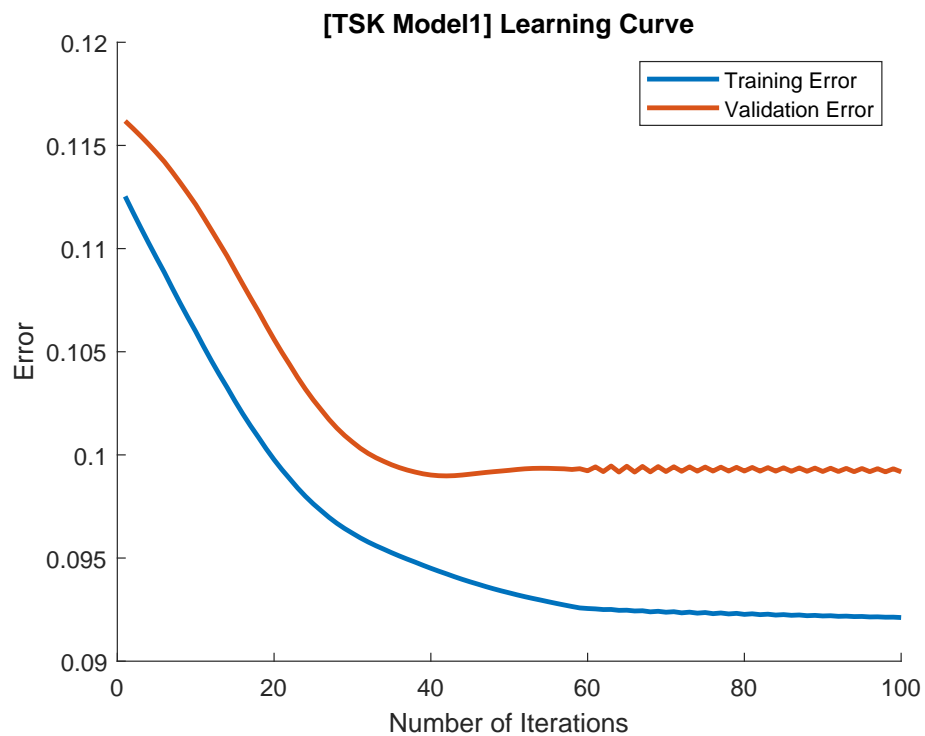
[TSK Model1] Membership Functions before training



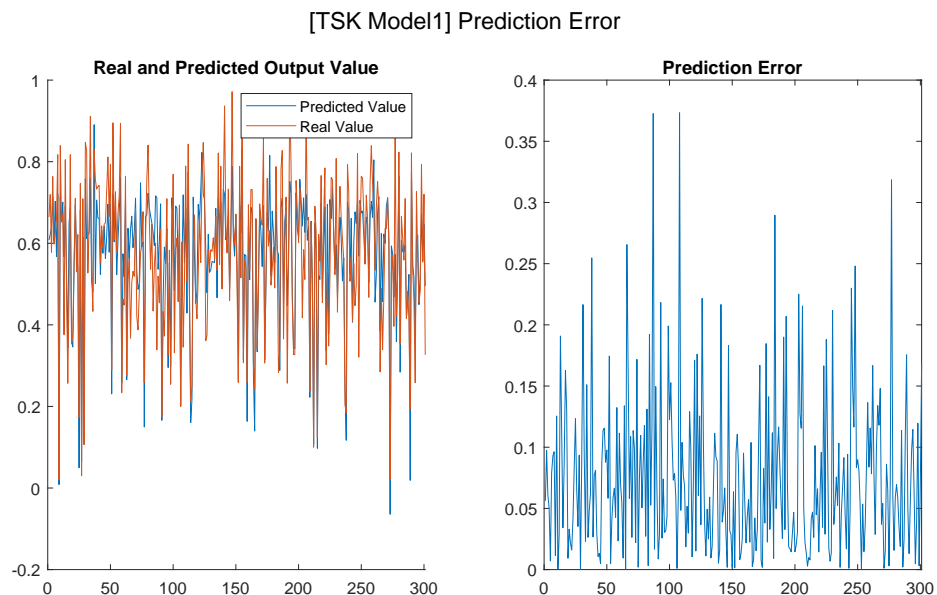
[TSK Model1] Membership Functions after training



2. Διάγραμμα μάθησης του μοντέλου



3. Διάγραμμα σφάλματος πρόβλεψης του μοντέλου



4. Μετρικές RMSE, NMSE, NDEI, R^2

```

-----TSK MODEL 1-----
RMSE = 1.7173
NMSE = 0.28235
NDEI = 0.53137
R2 = 0.71765

```

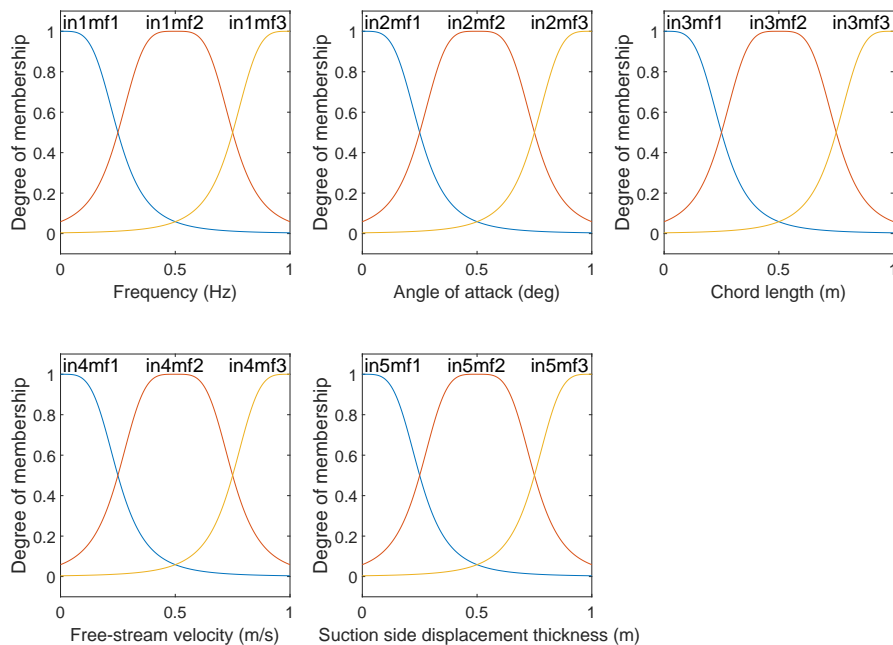
Παρατηρήσεις

- Το σφάλμα του **training set** μετά το πέρας των εποχών παρατηρώ ότι σταθεροποιείται σε μια αρκετά μικρή τιμή, ενώ το σφάλμα του **validation set** σταθεροποιείται γύρω από το 0.1 και ταλαντώνεται γύρω από αυτή τη τιμή κάτι που μπορεί να οφείλεται σε **overfitting**
- Το σφάλμα πρόβλεψης του μοντέλου παρατηρώ ότι έχει μεγάλη διακύμανση

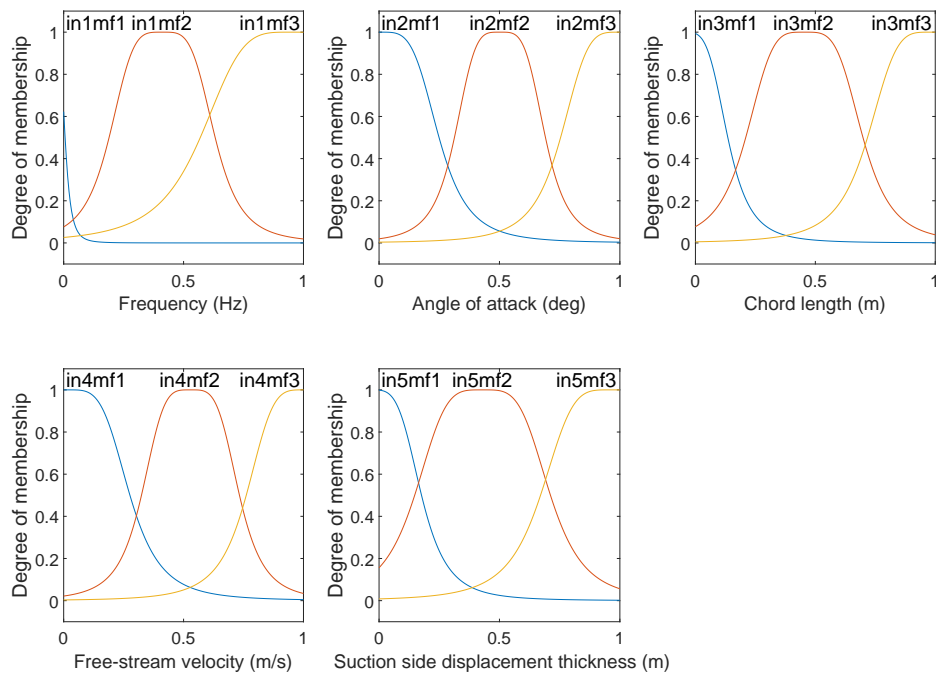
1.2 Μοντέλο 2

1. Ασαφή σύνολα πριν και μετά την διαδικασία εκπαίδευσης

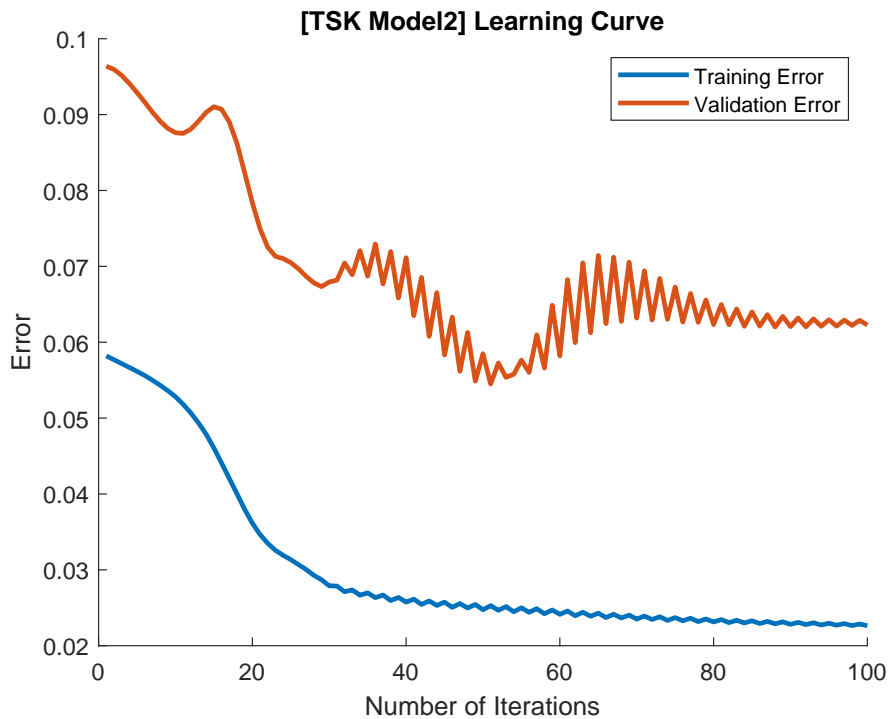
[TSK Model2] Membership Functions before training



[TSK Model2] Membership Functions after training

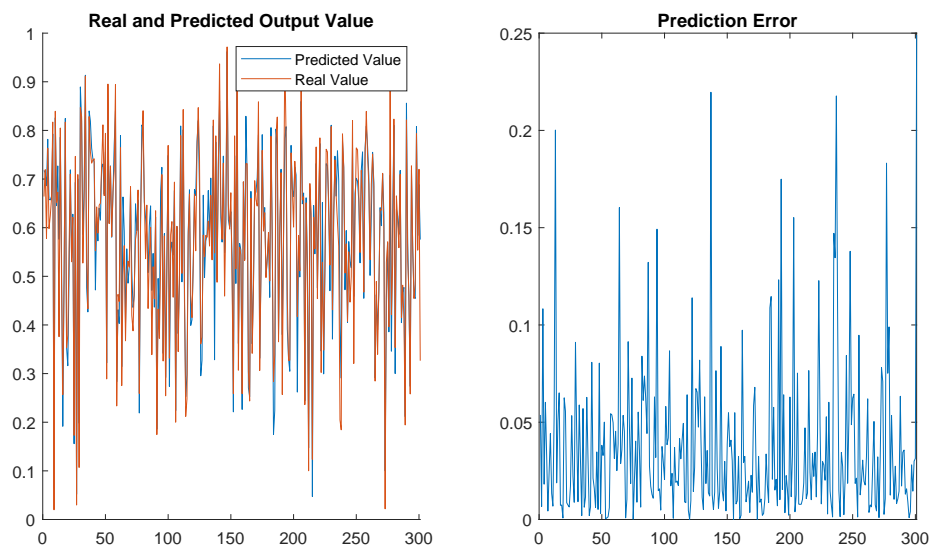


2. Διάγραμμα μάθησης του μοντέλου



3. Διάγραμμα σφάλματος πρόβλεψης του μοντέλου

[TSK Model2] Prediction Error



4. Μετρικές RMSE, NMSE, NDEI, R^2

```
-----TSK MODEL 2-----
RMSE = 0.94542
NMSE = 0.085579
NDEI = 0.29254
R2 = 0.91442
```

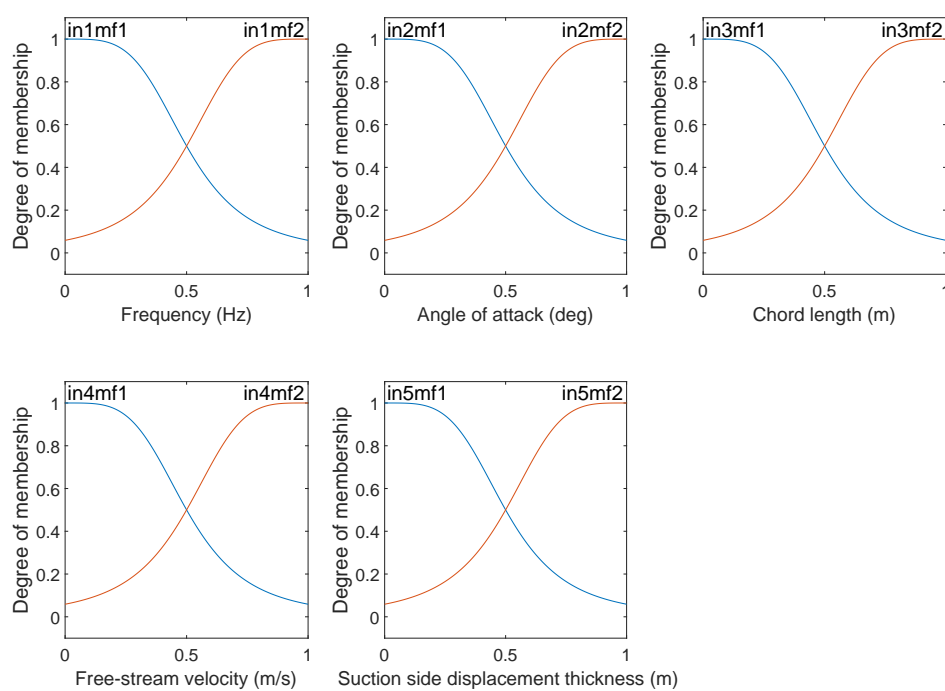
Παρατηρήσεις

- Το σφάλμα του **training set** παρατηρώ ότι σταθεροποιείται συγκλίνει στο μηδέν, ενώ το σφάλμα του **validation set** ταλαντώνεται γύρω από **0.07** κάτι που μπορεί να οφείλεται σε **overfitting**
- Το σφάλμα πρόβλεψης του μοντέλου παρατηρώ ότι έχει μεγάλη διακύμανση

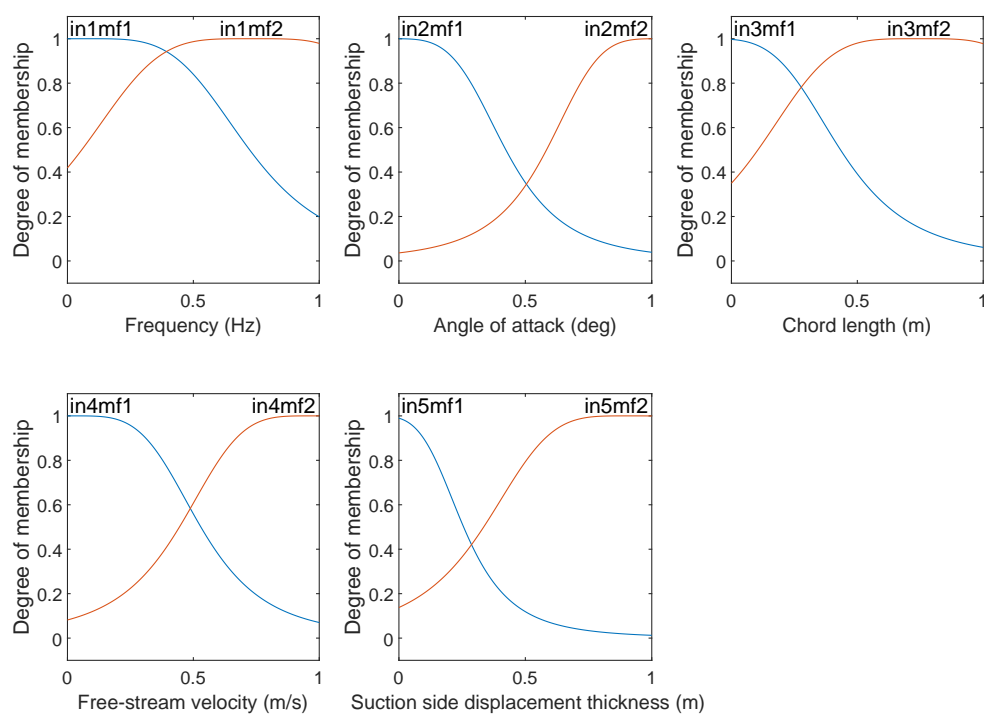
1.3 Μοντέλο 3

1. Ασαφή σύνολα πριν και μετά την διαδικασία εκπαίδευσης

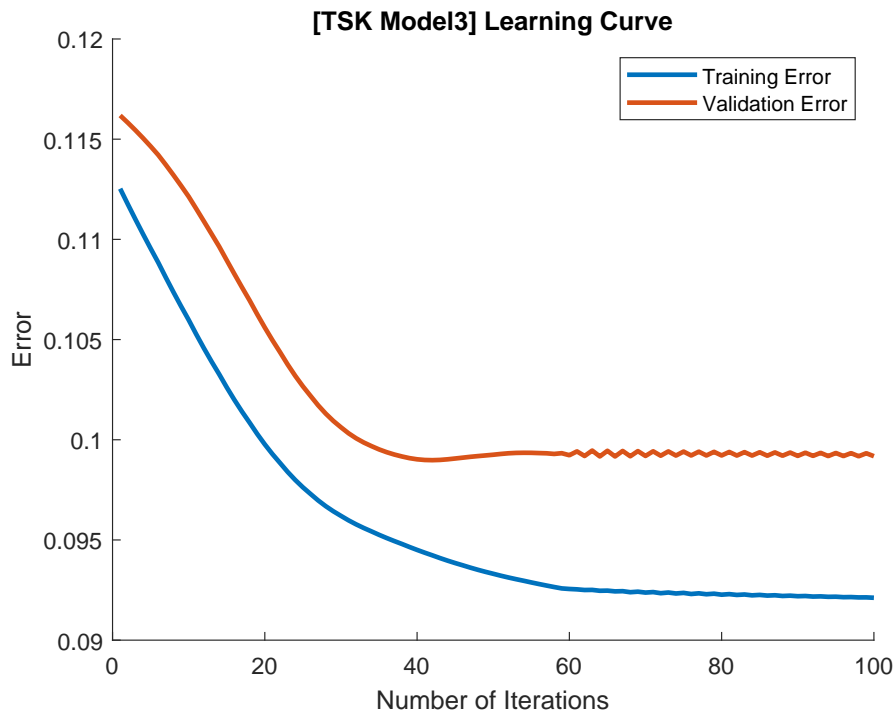
[TSK Model3] Membership Functions before training



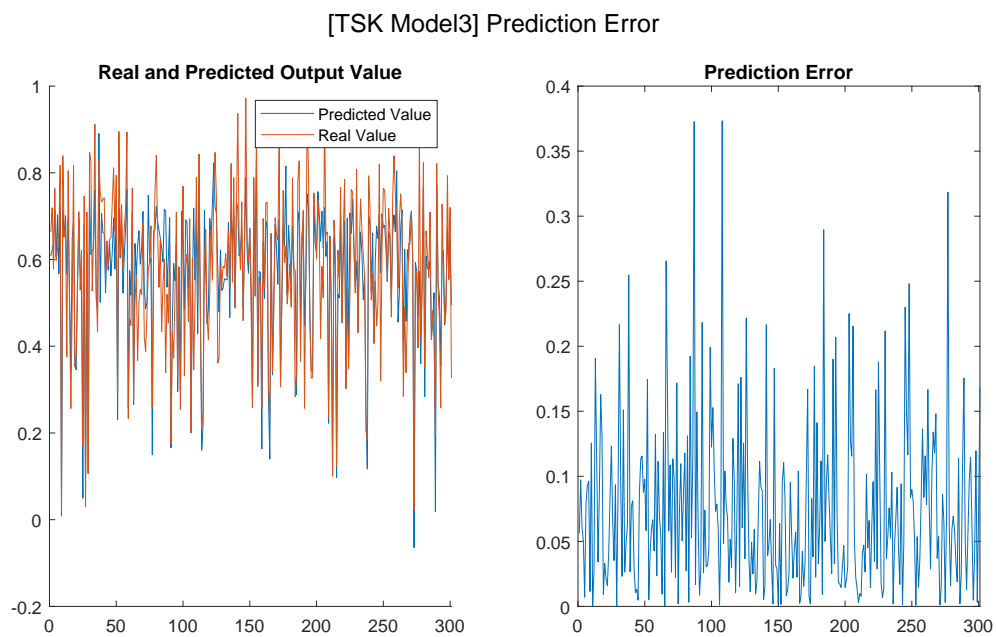
[TSK Model3] Membership Functions after training



2. Διάγραμμα μάθησης του μοντέλου



3. Διάγραμμα σφάλματος πρόβλεψης του μοντέλου



4. Μετρικές RMSE, NMSE, NDEI, R^2

-----TSK MODEL 3-----

RMSE = 1.7173
 NMSE = 0.28235
 NDEI = 0.53137
 R2 = 0.71765

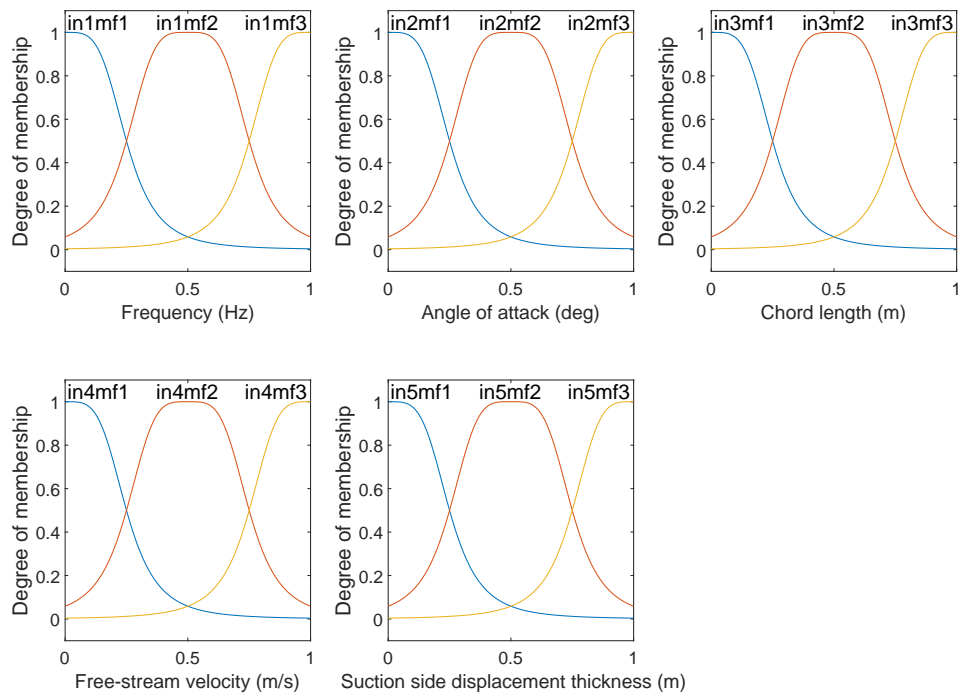
Παρατηρήσεις

- Το σφάλμα του **training set** μετά το πέρας των εποχών παρατηρώ ότι σταθεροποιείται σε μια αρκετά μικρή τιμή, ενώ το σφάλμα του **validation set** σταθεροποιείται γύρω από το 0.1 και ταλαντώνεται γύρω από αυτή τη τιμή κάτι που μπορεί να οφείλεται σε **overfitting**. Παρατηρώ ότι τα σφάλματα έχουν παρόμοια πορεία με αυτά του μοντέλου 1.
- Το σφάλμα πρόβλεψης του μοντέλου παρατηρώ ότι έχει μεγάλη διακύμανση

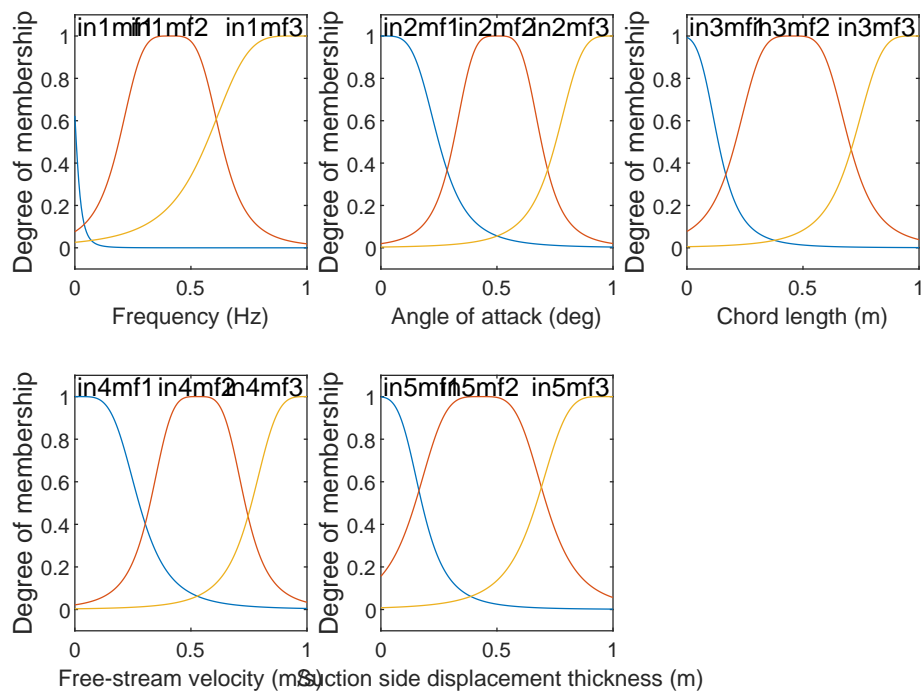
1.4 Μοντέλο 4

1. Ασαφή σύνολα πριν και μετά την διαδικασία εκπαίδευσης

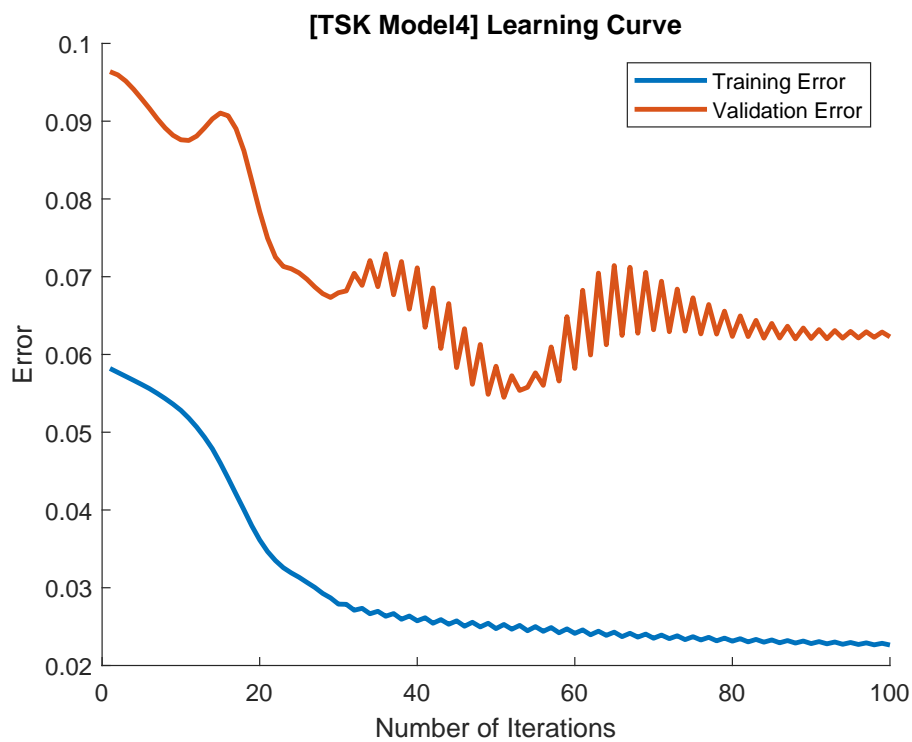
[TSK Model4] Membership Functions before training



[TSK Model4] Membership Functions after training

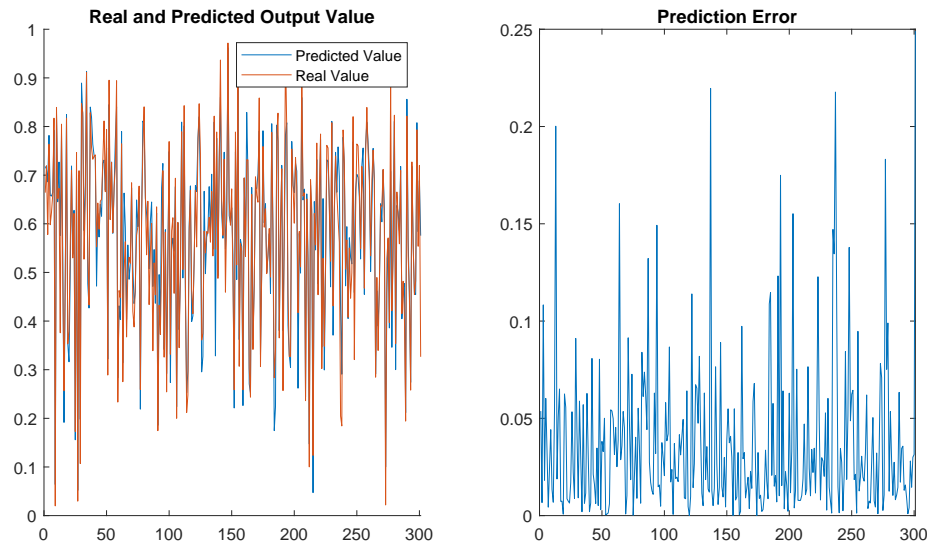


2. Διάγραμμα μάθησης του μοντέλου



3. Διάγραμμα σφάλματος πρόβλεψης του μοντέλου

[TSK Model4] Prediction Error



4. Μετρικές RMSE, NMSE, NDEI, R^2

```
-----TSK MODEL 4-----
RMSE = 0.94542
NMSE = 0.085579
NDEI = 0.29254
R2 = 0.91442
```

Παρατηρήσεις

- Το σφάλμα του **training set** παρατηρώ ότι σταθεροποιείται συγκλίνει στο μηδέν, ενώ το σφάλμα του **validation set** ταλαντώνεται γύρω από 0.07 κάτι που μπορεί να οφείλεται σε **overfitting**. Παρατηρώ ότι το σφάλμα του μοντέλου 4 ακολουθεί παρόμοια πορεία με αυτό του μοντέλου 2.
- Το σφάλμα πρόβλεψης του μοντέλου παρατηρώ ότι έχει μεγάλη διακύμανση

Συμπέρασμα: Η χρήση τριών συναρτήσεων συμμετοχής για τις μεταβλητές εισόδου οδηγεί σε μικρότερο σφάλμα στη έξοδο, ενώ όταν έχουμε μορφή εξόδου **Singleton** το σφάλμα μειώνεται περεταίρω. Επομένως επιλέγω το Μοντέλο 2.

2 Εφαρμογή **Dataset** με υψηλή διαστασιμότητα

Λόγω της μεγάλης διάστασης του **dataset**, εμφανίζονται τα προβλήματα έκρηξης κανόνων και πολλών χαρακτηριστικών. Για να αντιμετωπίσουμε το πρόβλημα αυτό, εφαρμόζουμε μεθόδους επιλογής χαρακτηριστικών και διαμέρισης διασκορπισμού. Οι

μέθοδοι αυτοί εισάγουν δύο ελεύθερες μεταβλητές, τον αριθμό των χαρακτηριστικών προς επιλογή και τον αριθμό των ομάδων που θα δημιουργηθούν (ακτίνα των **clusters ra**). Για να επιλέξουμε τις δύο αυτές μεταβλητές εφαρμόζουμε **grid partitioning**.

Αφού τρέξουμε τον αλγόριθμο λοιπόν του **grid partitioning** παίρνουμε στην έξοδο το μέσο σφάλμα που παρουσίασε κάθε μοντέλο. Στην συγκεκριμένη εφαρμογή εφαρμόζω **grid partitioning** για $ra = [0.2, 0.4, 0.6, 0.8]$ και για αριθμό χαρακτηριστικών = $[5\ 10\ 15\ 20]$. Φτιάχνω δηλαδή ένα 4x4 πλέγμα.