

Παυλίδης
Μιχαήλ
Ιάσων
ΑΕΜ: 9015
14/10/2021

TSK REGRESSION

Ασαφή Συστήματα 2020 - 2021

3^η Εργασία

Σκοπός Εργασίας

Ο σκοπός αυτής της εργασίας είναι η διερεύνηση της ικανότητας των μοντέλων TSK (Tagaki Sugeno Kang) στη μοντελοποίηση πολυμεταβλητών, μη γραμμικών συναρτήσεων. Ουσιαστικά ενδιαφερόμαστε για την προσέγγιση των τιμών διαφόρων μη γνωστών συναρτήσεων μέσω του προβλήματος της παλινδρόμησης. Επιλέγονται δύο σύνολα δεδομένων. Το πρώτο είναι σχετικά απλό έχοντας μικρό αριθμό χαρακτηριστικών (6), ενώ το δεύτερο είναι πιο πολύπλοκο και περιέχει 81 χαρακτηριστικά.

Πρώτο μέρος

Προετοιμασία Δεδομένων

Στο πρώτο κομμάτι της εργασίας επιλέξαμε ένα σύνολο δεδομένων με μικρό αριθμό χαρακτηριστικών και πιο συγκεκριμένα το *Airfoil Self-Noise* με διαστάσεις 1503 δείγματα \times 6 χαρακτηριστικά. Τα περιεχόμενα του αφορούν μετρήσεις από πειράματα της NASA σε αεροτομές και είναι τα εξής,

Είσοδοι:

1. Frequency [Hz]
2. Angle of attack [degrees]
3. Chord length [m]
4. Free-stream velocity [m/s]
5. Suction side displacement thickness [m]

Έξοδος:

1. Scaled sound pressure level [dB]

Όπως συνηθίζεται στις διαδικασίες εκπαίδευσης ενός μοντέλου με supervised τεχνικές, το αρχικό σύνολο δεδομένων χωρίζεται σε 3 μη επικαλυπτόμενα μέρη. Πιο συγκεκριμένα,

1. Στο σύνολο εκπαίδευσης D_{tr} το οποίο αποτελείται από το 60% των αρχικών δεδομένων. Αυτό το σύνολο χρησιμοποιείται για την εκπαίδευση του μοντέλου και τη ρύθμιση «βασικών» παραμέτρων όπως τα βάρη ενός νευρωνικού δικτύου.
2. Στο σύνολο επικύρωσης D_{val} το οποίο αποτελείται συνήθως από το 20% των αρχικών δεδομένων. Αυτό το σύνολο χρησιμοποιείται για την αξιολόγηση του μοντέλου με σκοπό την απόκτηση καλής ακρίβειας σε συνδυασμό με την αποφυγή του φαινομένου της υπερεκπαίδευσης.

3. Στο σύνολο ελέγχου D_{test} το οποίο επίσης αποτελείται συνήθως από το 20% των αρχικών δεδομένων. Αυτό το σύνολο χρησιμοποιείται για την αξιολόγηση του τελικού μοντέλου, σε δεδομένα που δεν έχουν συμμετάσχει καθόλου στη διαδικασία εκπαίδευσης-επικύρωσης.

Για το διαχωρισμό του συνόλου δεδομένων υλοποιήσαμε τη συνάρτηση `split_dataset()`. Αυτή, χρησιμοποιεί τη `randperm()` η οποία παράγει μια τυχαία μετάθεση αριθμών χωρίς επαναλαμβανόμενα στοιχεία. Βάσει των ποσοστών που αναφέρθηκαν παραπάνω, προκύπτουν τα τρία σύνολα δεδομένων που θα χρησιμοποιηθούν στη συνέχεια.

Προτού υλοποιήσουμε τα μοντέλα μας ακολουθήσαμε ένα ακόμα συνηθισμένο βήμα της προετοιμασίας των δεδομένων. Αυτό της κανονικοποίησης των χαρακτηριστικών. Ανάλογα με τον αλγόριθμο που χρησιμοποιείται, η κανονικοποίηση μπορεί να παίζει από ελάχιστο έως σημαντικό ρόλο στην εκπαίδευση και την ακρίβεια ενός μοντέλου. Στα νευρωνικά δίκτυα, έχει αναφερθεί πως βελτιώνει το χρόνο εκπαίδευσης καθώς και το ποσοστό της ακρίβειας. Επιλέξαμε να εφαρμόσουμε την *min-max* κανονικοποίηση προβάλλοντας τις αρχικές τιμές των χαρακτηριστικών στο διάστημα $[0,1]$ βάσει του τύπου

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Δημιουργία και Εκπαίδευση Μοντέλων

Για τη σύγκριση της συμπεριφοράς και της ακρίβειας των TSK μοντέλων επιλέχθηκαν 4 διαφορετικές αρχιτεκτονικές συστημάτων. Τα χαρακτηριστικά που αλλάζουν είναι ο αριθμός των συναρτήσεων συμμετοχής των εισόδων και ο τύπος της εξόδου ο οποίος μπορεί να είναι είτε *singleton* είτε πολυωνυμικός. Στο ακόλουθο πίνακάκι φαίνονται τα 4 συστήματα που υλοποιήθηκαν.

	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

Σε όλα τα συστήματα η μορφή των συναρτήσεων συμμετοχής είναι *bell-shaped* ενώ η εκπαίδευση έγινε με την υβριδική μέθοδο. Οι παράμετροι των συναρτήσεων συμμετοχής βελτιστοποιήθηκαν μέσω της μεθόδου της οπισθοδιάδοσης (backpropagation) ενώ οι παράμετροι της πολυωνυμικής συνάρτησης εξόδου βελτιστοποιήθηκαν μέσω της μεθόδου των ελαχίστων τετραγώνων.

Για το διαχωρισμό των αρχικών ασαφών συνόλων και τη δημιουργία των κανόνων χρησιμοποιήθηκε η μέθοδος *grid partition* κατά την οποία οι συναρτήσεις συμμετοχής διαχωρίζονται ομοιόμορφα βάσει των

πεδίων τιμών τους και προκύπτει ένας κανόνας για κάθε συνδυασμό των ασαφών συνόλων. Έτσι, αν έχουμε 5 εισόδους όπως στο παράδειγμα μας με 2 συναρτήσεις συμμετοχής για κάθε είσοδο, θα δημιουργηθούν $n = 2 \times 2 \times 2 \times 2 \times 2 = 2^5 = 32$ κανόνες.

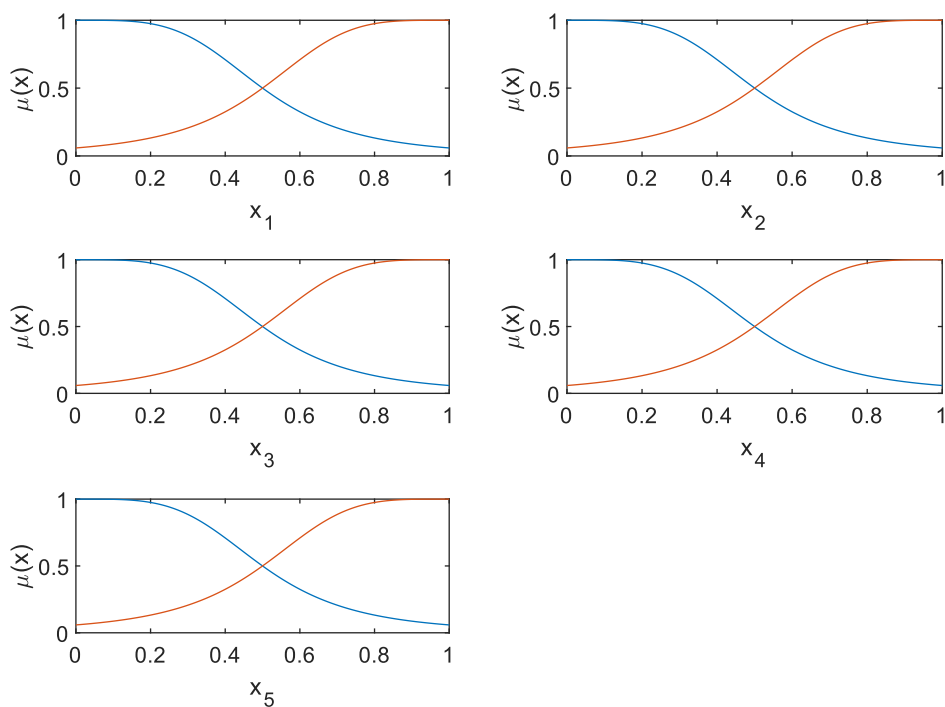
Για την αποτελεσματικότερη εκπαίδευση χρησιμοποιήθηκε όπως αναφέρθηκε και παραπάνω ένα σύνολο δεδομένων επικύρωσης (validation). Έτσι, το εκπαιδευμένο σύστημα που προκύπτει δεν είναι αυτό με το μικρότερο σφάλμα εκπαίδευσης αλλά αυτό με το μικρότερο σφάλμα επικύρωσης. Για να το πετύχουμε αυτό χρειάστηκε να πειραματιστούμε με τον αριθμό των επαναλήψεων εκπαίδευσης ή αλλιώς εποχών. Στα αντίστοιχα διαγράμματα θέλαμε να παρατηρήσουμε το γνωστό «γόνατο», δηλαδή το σημείο στο οποίο το σφάλμα επικύρωσης παύει να μειώνεται και ξεκινά να αυξάνεται. Σε εκείνο το σημείο αρχίζει το ανεπιθύμητο φαινόμενο της υπερεκπαίδευσης.

Στη συνέχεια παρουσιάζονται τα ζητούμενα της εργασίας.

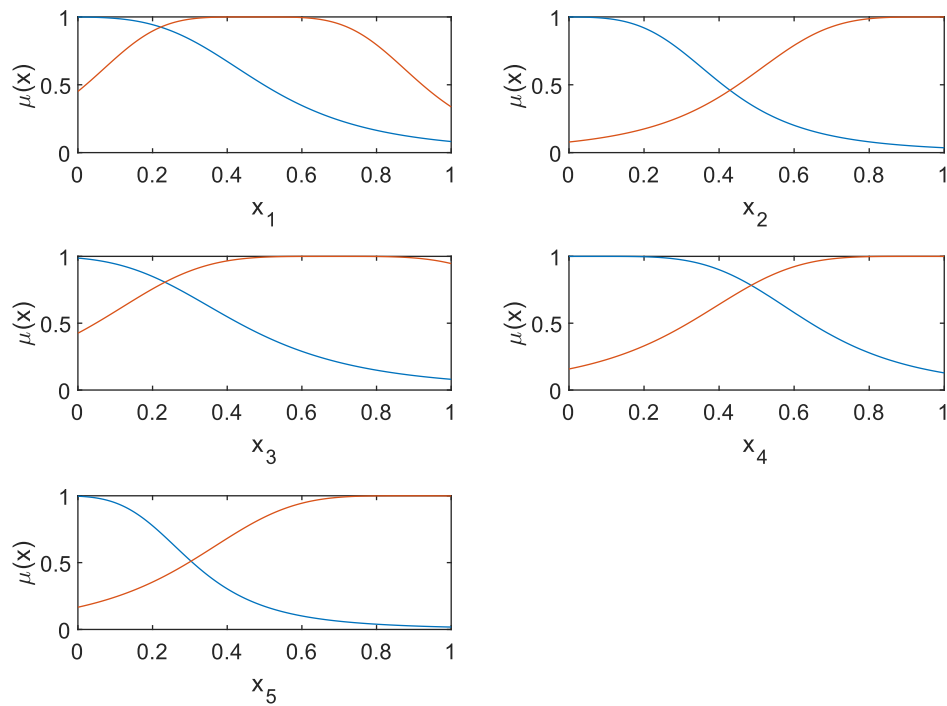
Συναρτήσεις Συμμετοχής Εισόδων

Στα επόμενα διαγράμματα απεικονίζονται οι μορφές των ασαφών συνόλων των πέντε μεταβλητών εισόδου. Παρουσιάζονται οι συναρτήσεις συμμετοχής πριν και μετά τη διαδικασία εκπαίδευσης. Χρησιμοποιήθηκε το μοντέλο που αντιστοιχεί στο ελάχιστο σφάλμα επικύρωσης και όχι αυτό με το ελάχιστο σφάλμα εκπαίδευσης.

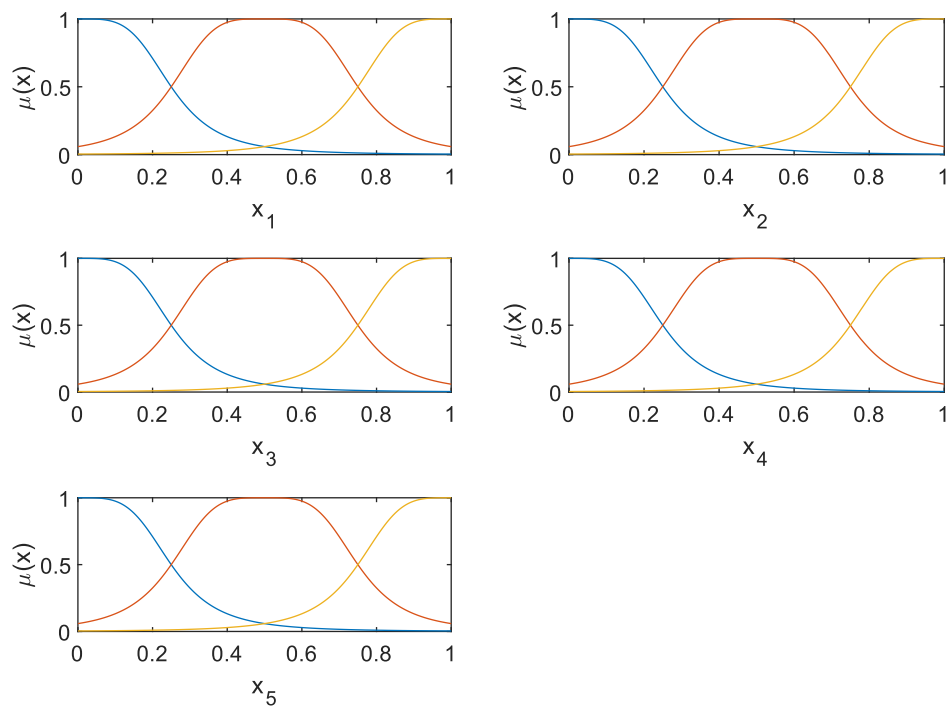
Membership functions - Initial model - TSK model - 1



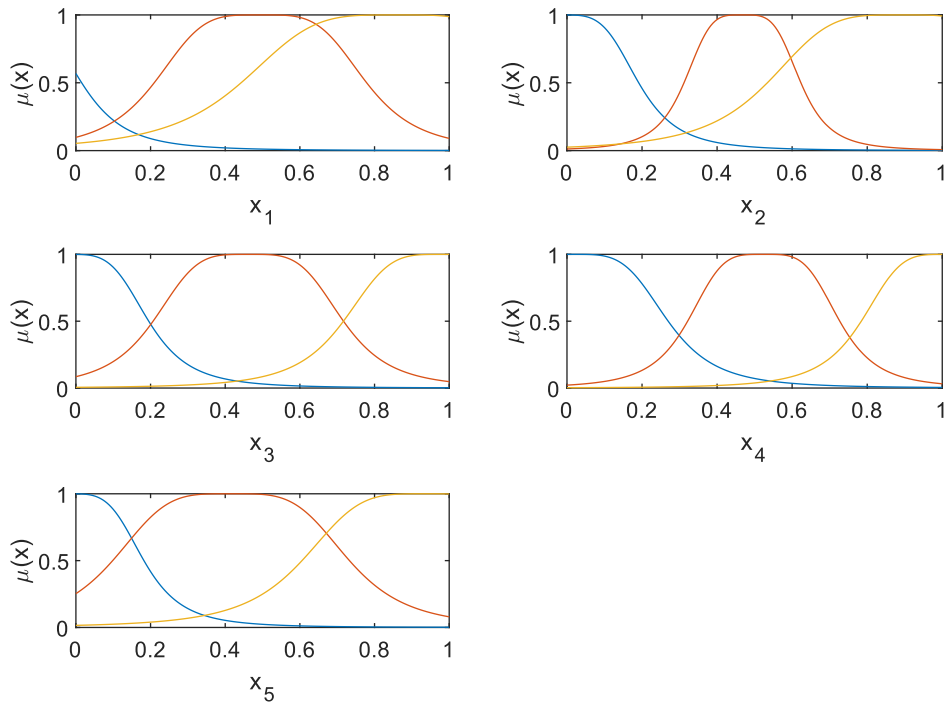
Memberhip functions - Min validation err model - TSK model - 1



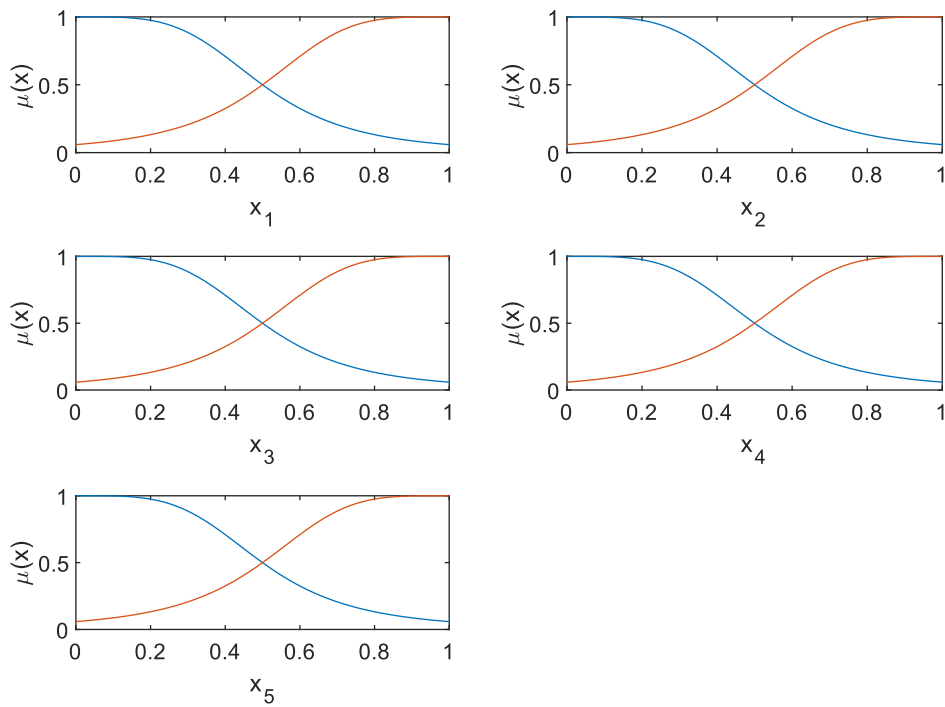
Memberhip functions - Initial model - TSK model - 2



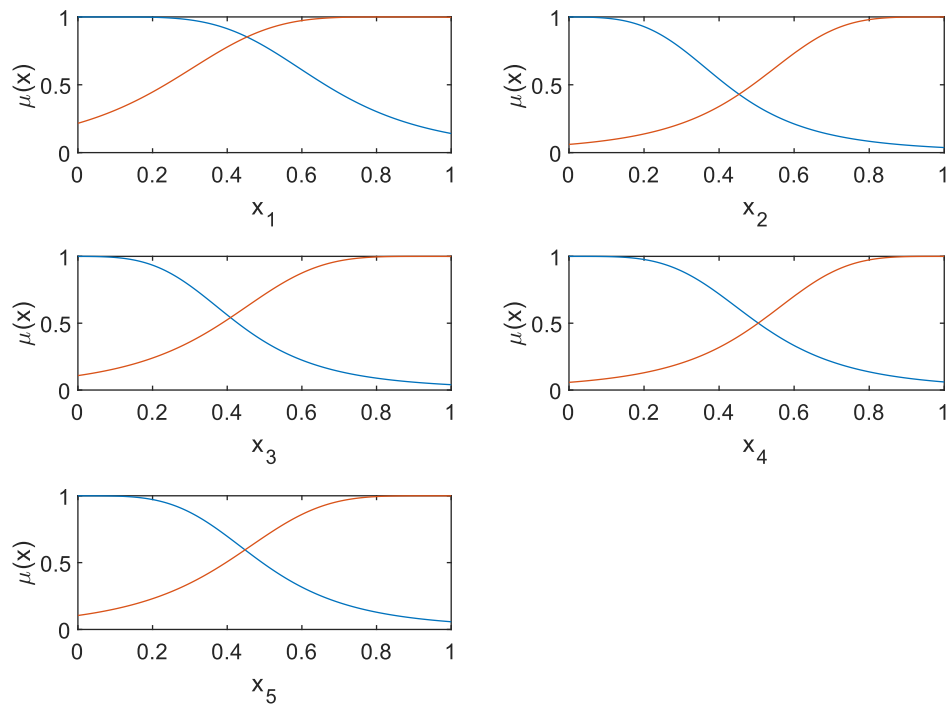
Memberhip functions - Min validation err model - TSK model - 2



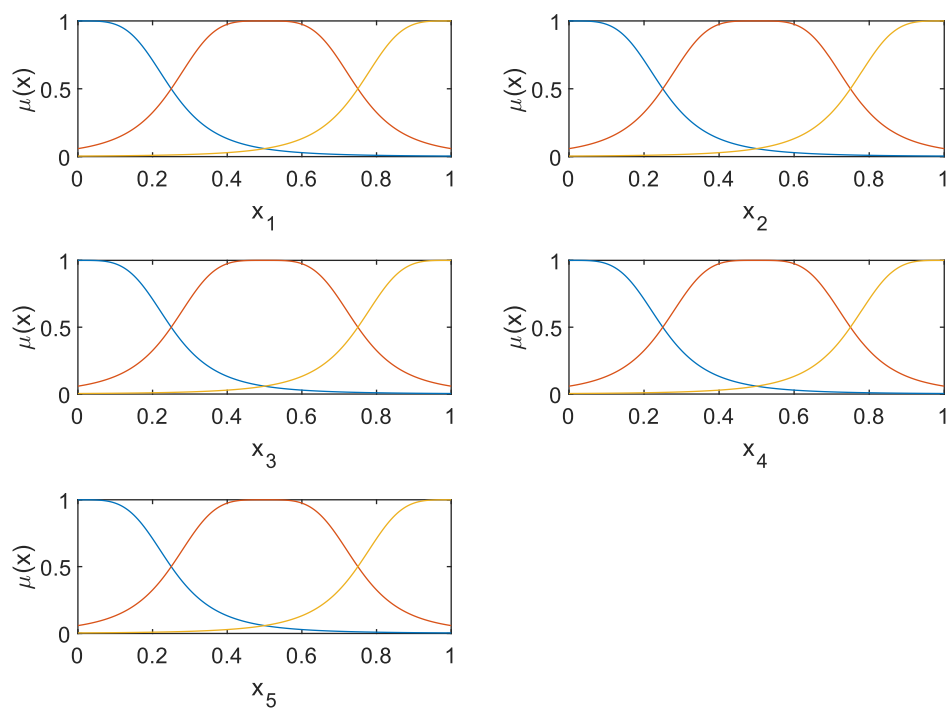
Memberhip functions - Initial model - TSK model - 3



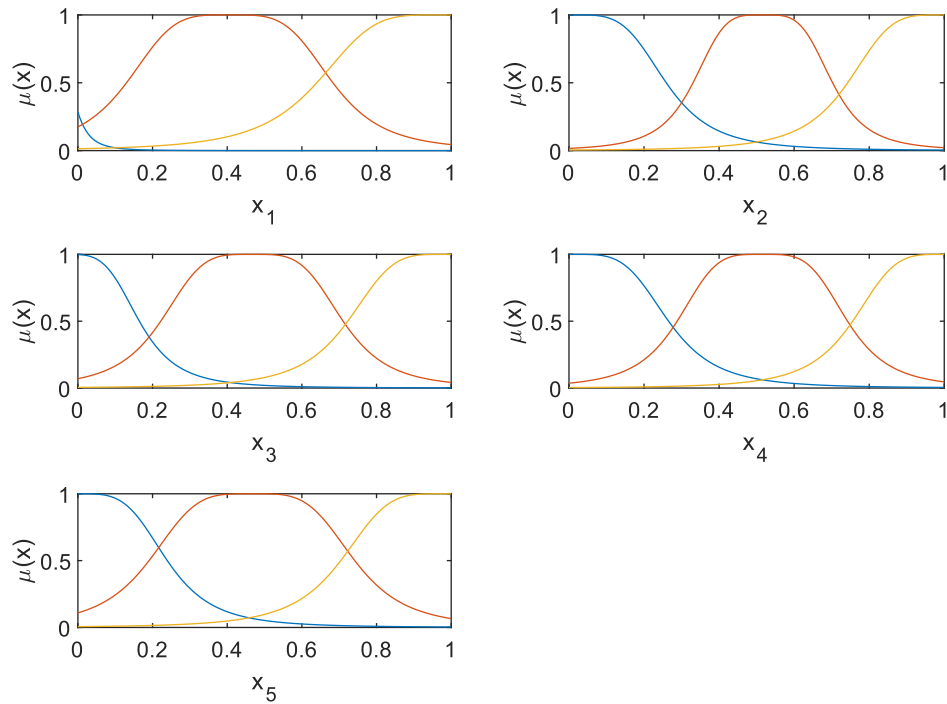
Memberhip functions - Min validation err model - TSK model - 3



Memberhip functions - Initial model - TSK model - 4



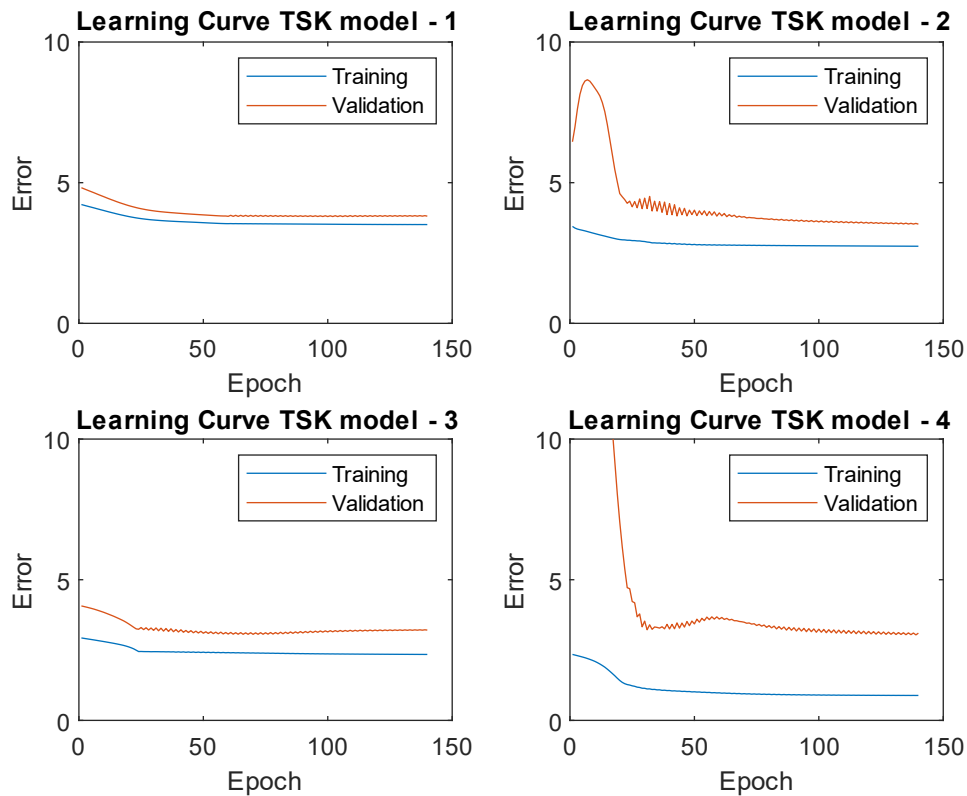
Membership functions - Min validation err model - TSK model - 4



Όπως μπορούμε να δούμε από τα διαγράμματα παραπάνω, οι αρχικές συναρτήσεις συμμετοχής έχουν βαθμό επικάλυψης 0.5 ενώ μετά την εκπαίδευση οι παράμετροι έχουν αλλάξει.

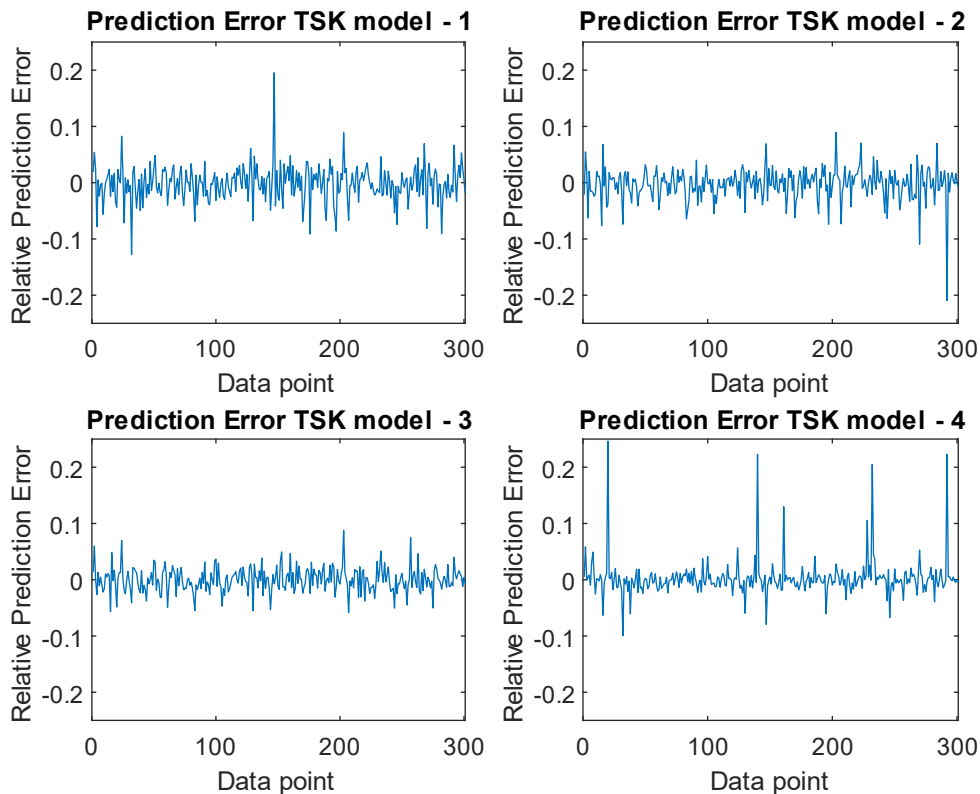
Διαγράμματα εκπαίδευσης

Στα διαγράμματα εκπαίδευσης ή εκμάθησης, μπορούμε να δούμε τις τιμές του σφάλματος εκπαίδευσης και επικύρωσης σε κάθε εποχή. Έτσι, μπορούμε να επιλέξουμε το κατάλληλο σημείο για να σταματήσουμε την εκπαίδευση. Στο επόμενο σχήμα απεικονίζονται τα διαγράμματα εκπαίδευσης των τεσσάρων μοντέλων.



Σφάλμα Πρόβλεψης

Ο σκοπός της εκπαίδευσης είναι η δημιουργία ενός μοντέλου το οποίο μπορεί να προβλέψει σωστά τις τιμές σε δεδομένα που δεν έχει εκπαιδευτεί. Για αυτό το λόγο χρησιμοποιείται το test set με το οποίο ελέγχουμε πόσο καλό είναι το μοντέλο στη γενίκευση. Στο παρακάτω διάγραμμα απεικονίζονται τα σχετικά σφάλματα πρόβλεψης των τεσσάρων μοντέλων.



Αξιολόγηση Μοντέλων

Για την αξιολόγηση της ικανότητας του μοντέλου να προβλέπει σωστά τις τιμές μιας άγνωστης συνάρτησης χρησιμοποιήθηκαν 4 δείκτες αξιολόγησης.

1. **RMSE:** Είναι η τυπική απόκλιση των σφαλμάτων πρόβλεψης του μοντέλου. Υποδηλώνει πόσο συγκεντρωμένες είναι οι τιμές του μοντέλου στις πραγματικές. Η ιδανική τιμή είναι το 0.
2. **R^2 :** Ονομάζεται συντελεστής προσδιορισμού και υποδηλώνει το ποσοστό της συνολικής διασποράς που εξηγείται από το μοντέλο. Η ιδανική τιμή είναι το 1.
3. **NMSE:** Είναι η μέση τιμή του τετραγώνου του σφάλματος κανονικοποιημένη ως προς την ενέργεια του σήματος. Η ιδανική τιμή είναι το 0.
4. **NDEI:** Ορίζεται ως η ρίζα του NMSE.

Στον ακόλουθο πίνακα επισυνάπτονται οι τιμές των παραπάνω δεικτών για τα 4 μοντέλα

Model	RMSE	R^2	NMSE	NDEI
TSK model 1	3.8667	0.6573	0.3427	0.5854
TSK model 2	3.4403	0.7287	0.2713	0.5208
TSK model 3	2.6932	0.8334	0.1662	0.4077
TSK model 4	4.0042	0.6325	0.3675	0.6062

Σχολιασμός Αποτελεσμάτων

Παρατηρώντας τα διαγράμματα εκπαίδευσης και των σφαλμάτων πρόβλεψης μπορούμε να βγάλουμε κάποια χρήσιμα συμπεράσματα για τα μοντέλα που δημιουργήσαμε.

- Συγκρίνοντας τα διαγράμματα εκπαίδευσης των μοντέλων 1 και 2 μπορούμε να παρατηρήσουμε πως η αύξηση των συναρτήσεων συμμετοχής των εισόδων από 2 σε 3 βελτίωσε σημαντικά το σφάλμα εκπαίδευσης. Ανάλογο συμπέρασμα μπορούμε να εξάγουμε και για τα μοντέλα 3 και 4.
- Συγκρίνοντας τα διαγράμματα εκπαίδευσης των μοντέλων 1 και 3 βλέπουμε πως η μετατροπή του τύπου εξόδου από Singleton σε πολυωνυμικό μείωσε τόσο το σφάλμα επικύρωσης όσο και το σφάλμα εκπαίδευσης. Το ίδιο παρατηρείται και για τα μοντέλα 2 και 4.
- Από τα διαγράμματα των σφαλμάτων πρόβλεψης βλέπουμε πως το μοντέλο 3 έχει την καλύτερη διασπορά γύρω από το 0. Ακολουθεί το μοντέλο 2 και μετά το 1. Το 4 που είναι και το πιο πολύπλοκο έχοντας 3 συναρτήσεις συμμετοχής για κάθε είσοδο και πολυωνυμική έξοδο φαίνεται να έχει αρκετά σημεία στα οποία δεν μπόρεσε να κάνει καλή πρόβλεψη. Συνδυάζοντας το γεγονός αυτό με το χαμηλότερο σφάλμα εκπαίδευσης που είχε μπορούμε να συμπεράνουμε πως έχει παρουσιάσει φαινόμενο υπερεκπαίδευσης.
- Από τον πίνακα των δεικτών αξιολόγησης μπορούμε να δούμε πώς όντως το μοντέλο 3 έχει τα καλύτερα χαρακτηριστικά ενώ ακολουθούν τα 2,1 και 4.

Δεύτερο μέρος

Στο δεύτερο μέρος της εργασίας επιλέχθηκε ένα σύνολο δεδομένων το οποίο περιέχει περισσότερα χαρακτηριστικά (81). Αυτό το γεγονός καθιστά αδύνατη τη χρήση του grid partitioning για τον διαχωρισμό των εισόδων. Για αυτό το λόγο χρησιμοποιήσαμε μια τεχνική ομαδοποίησης, αυτή του subtractive clustering. Ακόμα και αυτή η τεχνική όμως εφαρμοσμένη σε 81 χαρακτηριστικά θα ήθελε πολύ σημαντικό χρόνο, για αυτό έπρεπε να γίνει επιλογή χαρακτηριστικών.

Προετοιμασία Δεδομένων

Αρχικά, διαχωρίσαμε το σύνολο δεδομένων με τον ίδιο τρόπο όπως στο πρώτο μέρος της εργασίας. Δηλαδή, 60% να είναι το σύνολο εκπαίδευσης και από 20% να είναι τα σύνολα επικύρωσης και ελέγχου. Στη συνέχεια κανονικοποιήσαμε τις τιμές των χαρακτηριστικών/εισόδων στο διάστημα $[0,1]$ για να βελτιώσουμε τα αποτελέσματα τόσο της ομαδοποίησης όσο και της εκπαίδευσης.

Όπως αναφέραμε, 81 χαρακτηριστικά είναι ένας αρκετά μεγάλος αριθμός για την εκπαίδευση ενός δικτύου. Αν είχε χρησιμοποιηθεί η τεχνική του grid partition με 2 συναρτήσεις συμμετοχής για το διαχωρισμό των εισόδων, θα χρειαζόμασταν 2^{81} κανόνες, κάτι που είναι υπολογιστικά αδύνατο. Επομένως, το επόμενο βήμα ήταν η μείωση των διαστάσεων του προβλήματος. Για αυτό το σκοπό

χρησιμοποιήθηκε ο αλγόριθμος RreliefF. Βάσει αυτού αξιολογήσαμε τη χρησιμότητα του εκάστοτε χαρακτηριστικού.

Επιλογή Παραμέτρων

Οι δύο παράμετροι που έπρεπε να επιλέξουμε για τη δημιουργία του μοντέλου μας ήταν ο αριθμός των χαρακτηριστικών που θα επιλέξουμε και η ακτίνα που θα χρησιμοποιηθεί στο subtractive clustering. Για την επιλογή αυτή χρησιμοποιήθηκε η τεχνική grid search σε συνδυασμό με το k-fold validation. Πιο συγκεκριμένα, επιλέγονται κάποιες τιμές για τις παραμέτρους που θέλουμε να εξετάσουμε. Για κάθε συνδυασμό αυτών δημιουργούμε ένα μοντέλο το οποίο αξιολογούμε βάσει του σφάλματος επικύρωσης του. Για καλύτερα αποτελέσματα αντί για ένα μοντέλο δημιουργούμε k (5 στην περίπτωση μας). Συγχρόνως, χωρίζουμε και τα δεδομένα μας σε 5 διαφορετικά υποσύνολα τα οποία χρησιμοποιούνται για την εκπαίδευση του κάθε ένα μοντέλου. Αποθηκεύεται το μέσο σφάλμα και συνεχίζουμε στον επόμενο συνδυασμό. Όταν τελειώσουν όλοι οι συνδυασμοί επιλέγονται τα χαρακτηριστικά που οδήγησαν στο χαμηλότερο σφάλμα. Μετά την εύρεση των βέλτιστων παραμέτρων δημιουργούμε το τελικό μοντέλο το οποίο εκπαιδεύουμε με το σύνολο των δεδομένων εκπαίδευσης.

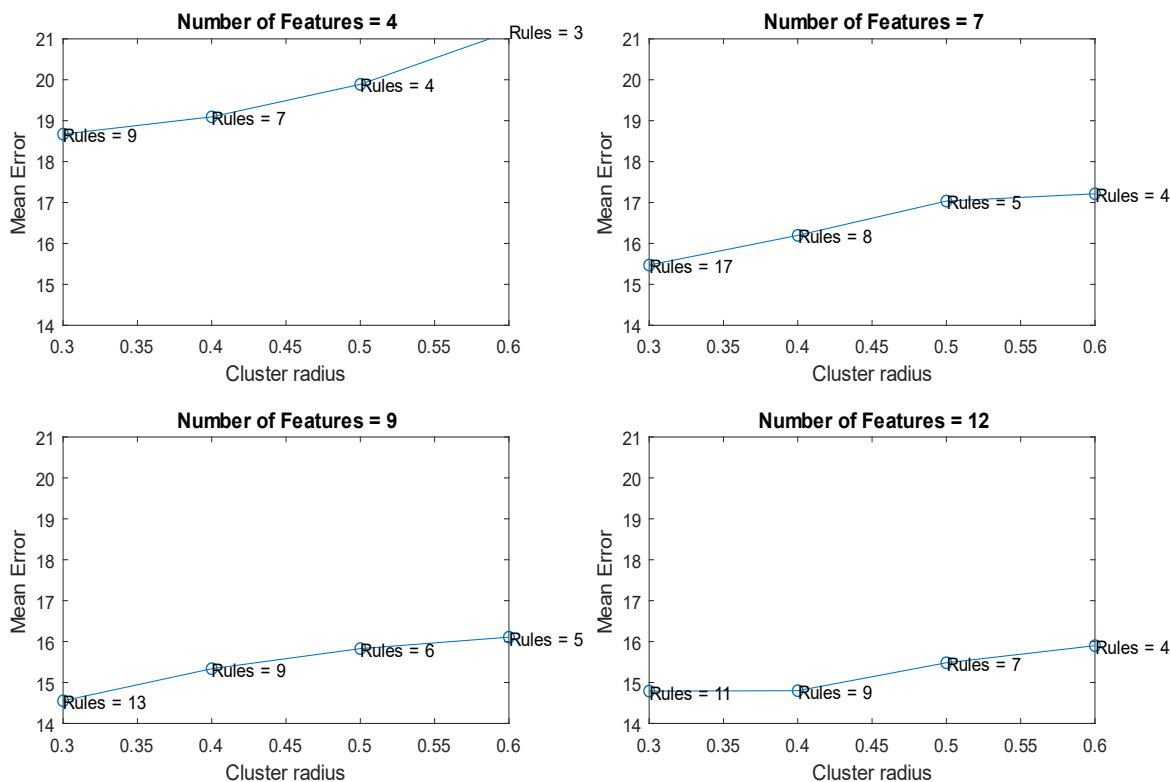
Οι παράμετροι που δοκιμάσαμε ήταν τα [4,7,9,12] χαρακτηριστικά και οι [0.6, 0.5, 0.4, 0.3] ακτίνες ομαδοποίησης. Κάθε ένα από τα 5 μοντέλα που δημιουργήσαμε για τους 16 συνδυασμούς εκπαιδεύτηκε για 100 εποχές. Τα μέσα σφάλματα συναρτήσεως του αριθμού των χαρακτηριστικών και της ακτίνας φαίνονται στον επόμενο πίνακα.

n_feat \ radius	0.6	0.5	0.4	0.3
4	21.1668	19.8890	19.0928	18.6714
7	17.2133	17.0362	16.1974	15.4717
9	16.1083	15.8281	15.3365	14.5526
12	15.8999	15.4823	14.7997	14.7901

Ο αριθμός των κανόνων που δημιουργήθηκαν από τον αλγόριθμο συναρτήσεως του αριθμού των χαρακτηριστικών και της ακτίνας φαίνεται στον επόμενο πίνακα.

n_feat \ radius	0.6	0.5	0.4	0.3
4	3	4	7	9
7	4	5	8	17
9	5	6	9	13
12	4	7	9	11

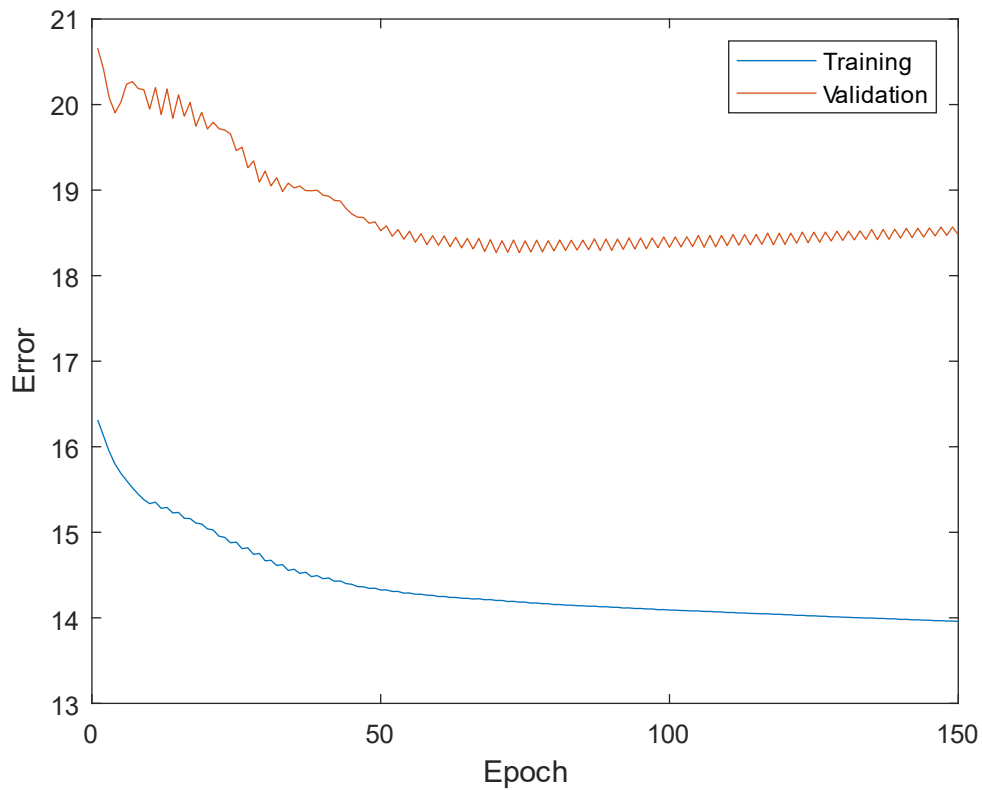
Στα επόμενα διαγράμματα συνοψίζονται οι πληροφορίες των δύο παραπάνω πινάκων.



Μετά την επιλογή των βέλτιστων παραμέτρων,

$$(N, r) = (12, 0.3)$$

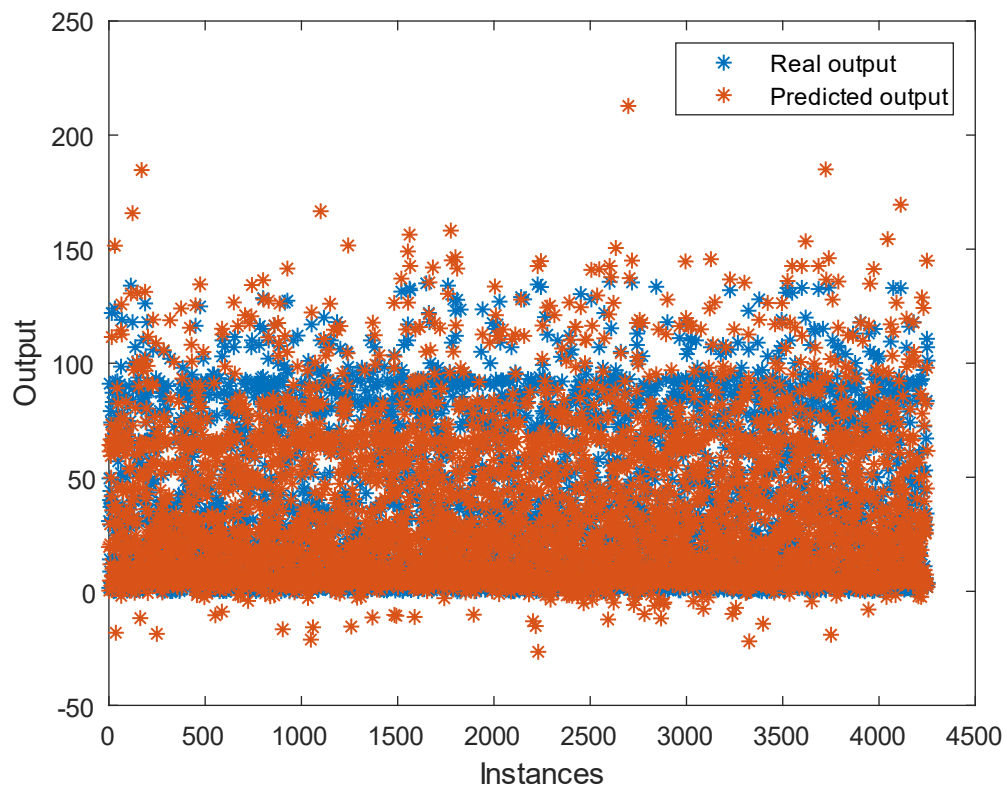
δημιουργήσαμε το τελικό μοντέλο το οποίο εκπαιδεύτηκε για 150 εποχές χρησιμοποιώντας το σύνολο των δεδομένων εκπαίδευσης. Στο επόμενο διάγραμμα παρουσιάζεται το σφάλμα επικύρωσης και εκπαίδευσης ανά εποχή εκπαίδευσης.



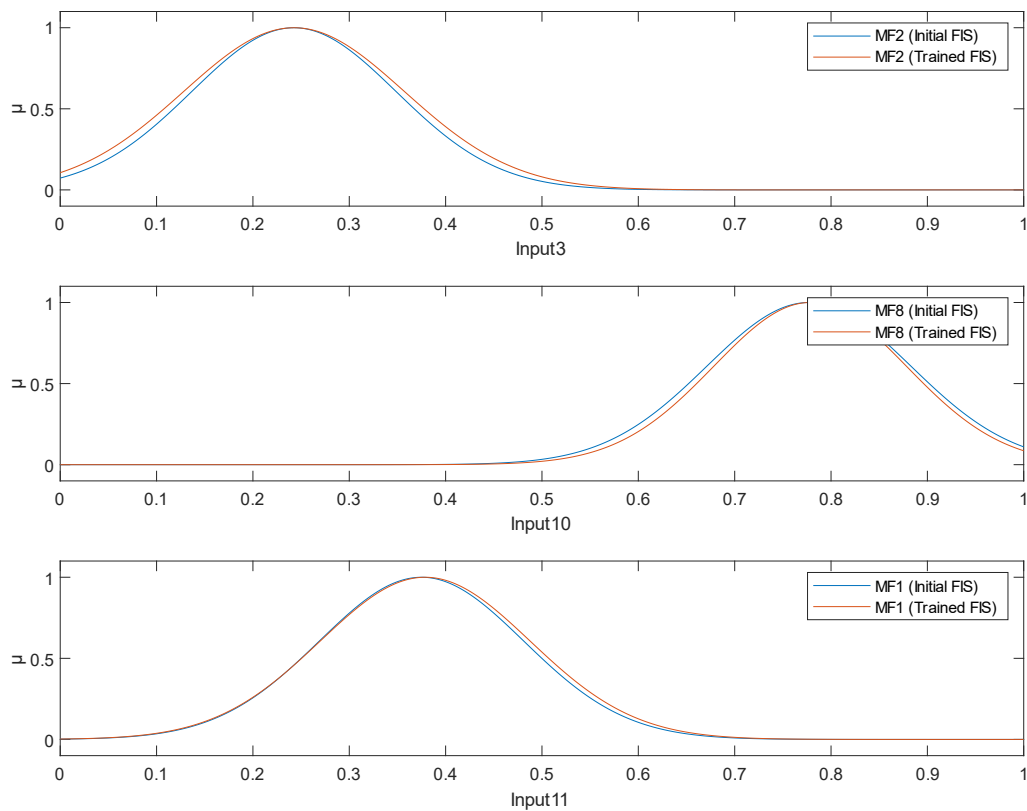
Για την αξιολόγηση του μοντέλου χρησιμοποιήθηκαν οι ίδιες μετρικές όπως στο πρώτο μέρος. Οι τιμές του φαίνονται στον ακόλουθο πίνακα.

RMSE	NMSE	NDEI	R^2
19.6448	0.3297	0.5742	0.6703

Παρατηρούμε υψηλότερο RMSE σε σχέση με το απλούστερο σετ δεδομένων του πρώτου μέρους. Παρόλα αυτά, η χαμηλή σχετικά τιμή του NMSE και το 0.67 του R^2 υποδηλώνουν πως το μοντέλο έχει ικανοποιητική ικανότητα πρόβλεψης. Στο επόμενο διάγραμμα απεικονίζονται οι πραγματικές τιμές του συνόλου ελέγχου και οι τιμές πρόβλεψης του μοντέλου.



Στο επόμενο διάγραμμα απεικονίζονται οι συναρτήσεις συμμετοχής της 3^{ης} , 10^{ης} και 11^{ης} εισόδου του μοντέλου πριν και μετά τη διαδικασία εκπαίδευσης. Για οπτικούς λόγους επιλέξαμε να απεικονίσουμε μια από τις 11 συναρτήσεις συμμετοχής που έχει κάθε είσοδος.



References

- Jang, J.-S. R., "ANFIS: Adaptive-Network-based Fuzzy Inference Systems," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No. 3, May 1993, pp. 665-685.
- Ασαφή Συστήματα, Σημειώσεις Μαθήματος, Θεοχάρης Ιωάννης
- Fuzzy Logic Toolbox User's Guide
- Η συνάρτηση `plot_fuzzy_rules()` μου δόθηκε από τον συμφοιτητή μου Σφέτκο Αχιλλέα.