



ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ
Επίλυση προβλήματος παλινδρόμησης με
χρήση μοντέλων TSK

ΕΙΣΗΓΗΤΗΣ : ΘΕΟΧΑΡΗΣ ΙΩΑΝΝΗΣ
ΚΩΣΤΑΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

kostakonst@ece.auth.gr

AEM : 9209

Ιανουάριος 2022

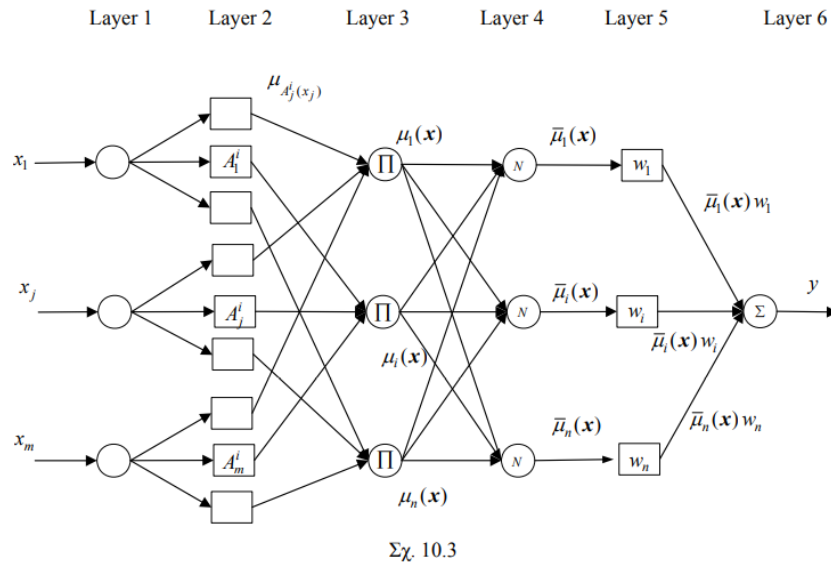
Contents

1	Εισαγωγή	3
2	Εφαρμογή σε απλό dataset	5
2.1	Διαχωρισμός σε σύνολα εκπαίδευσης-επικύρωσης-ελέγχου . . .	5
2.2	Εκπαίδευση TSK μοντέλων με διαφορετικές παραμέτρους	6
2.2.1	TSK_model_1	7
2.2.2	TSK_model_2	8
2.2.3	TSK_model_3	10
2.2.4	TSK_model_4	11
2.3	Αξιολόγηση μοντέλων	12
3	Εφαρμογή σε dataset με υψηλή διαστασιμότητα	14
3.1	Μοντελοποίησης άγνωστης συνάρτησης	14
3.2	Μοντελοποίηση και Βέλτιστη σχεδίαση	14
3.2.1	Διαχωρισμός σε σύνολα εκπαίδευσης- επικύρωσης – ελέγχου:	15
3.2.2	Επιλογή των βέλτιστων παραμέτρων:	15
3.3	Απόδοση τελικού μοντέλου	16
3.4	Συμπέρασμα	20

1 Εισαγωγή

Τα μοντέλα Mamdani(Singleton), που αναπτύχθηκαν στην πρώτη εργασία του μαθήματος μπορούν να περιγράψουν ποιοτικά την συμπεριφορά εισόδου-εξόδου ενός συστήματος.

Η δομή του μοντέλου Singleton περιγράφεται με την βοήθεια έξι στρωμάτων. Με βάση τα έξι αυτά στρώματα, το μοντέλο μπορεί να θεωρηθεί σαν ένα νευρωνικό δίκτυο, και πιο συγκεκριμένα, σαν ένα ασαφές νευρωνικό δίκτυο (Fuzzy Neural Network, FNN). Στην συνέχεια, θα δώσουμε μια συνοπτική περιγραφή των στρωμάτων (σχήμα 10.3), όπως παρουσιάζεται στις σημειώσεις του κ.Θεοχάρη Ι. (Fuz_ch-10):



Στρώμα 1: Οι κόμβοι του στρώματος αυτού λέγονται και κόμβοι εισόδου. Οι κόμβοι αυτοί απλώς περνάνε τα σήματα εισόδου στο επόμενο στρώμα.

Στρώμα 2: Οι κόμβοι στο στρώμα αυτό λέγονται λεκτικοί κόμβοι και ενεργούν σαν συναρτήσεις συμμετοχής. Η έξοδος των κόμβων παρέχει τον βαθμό συμμετοχής με τον οποίο το x_j συμμετέχει στο σύνολο A_j^i .

Στρώμα 3: Οι κόμβοι στο στρώμα αυτό λέγονται κόμβοι κανόνες (rule nodes). Ο κόμβος πολλαπλασιάζει τα εισερχόμενα σήματα $\mu_{A_j^i(x_j)}$ του κανόνα και παρέχει στην έξοδο το γινόμενό τους. Η έξοδος, παράγει τον βαθμό εκπλήρωσης $\mu_i(x)$ του κανόνα ενώ οι λεκτικοί κόμβοι που συνδέονται με τον αυτόν αποτελούν τις συνιστώσες του κανόνα. Αντιστοιχούν δηλαδή, στα ασαφή σύνολα του τμήματος υπόθεσής του.

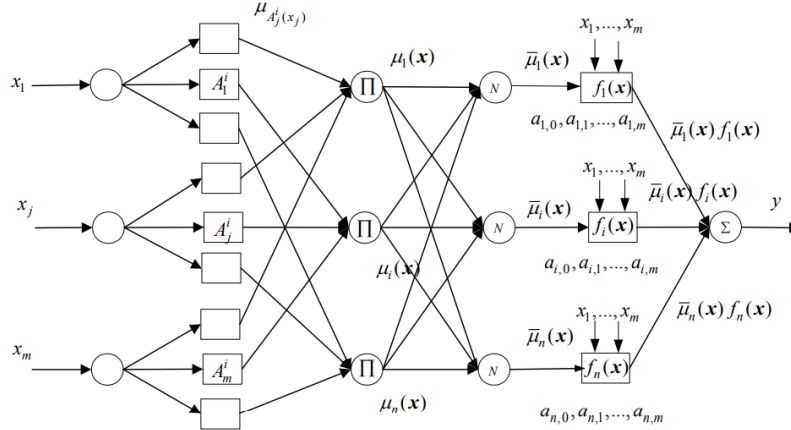
Στρώμα 4: Οι κόμβοι στο στρώμα αυτό έχουν σαν εισόδους τους βαθμούς εκπλήρωσης των κανόνων $\mu_i(x)$ και υπολογίζουν τους κανονικοποιημένους βαθμούς εκπλήρωσης $\bar{\mu}_i(x)$.

Στρώμα 5: Οι κόμβοι στο στρώμα αυτό υπολογίζουν την σταθμισμένη συνεισφορά του τμήματος συμπεράσματος κάθε κανόνα $\bar{\mu}_i(x)w_i$, όπου $\bar{\mu}_i(x)$ είναι η έξοδος του στρώματος 4 και w_i είναι η τιμή του συμπεράσματος του κανόνα $R^{(i)}$.

Στρώμα 6: Το στρώμα αυτό περιλαμβάνει ένα κόμβο, ο οποίος αθροίζει όλα τα εισερχόμενα σήματα από το στρώμα 5 και παρέχει την τελική έξοδο του μοντέλου.

Παρόλα αυτά, τα μοντέλα αυτά δεν παρουσιάζουν ικανοποιητική ακρίβεια προσέγγισης. Για αυτό τον λόγο, έχουν αναπτυχθεί στην βιβλιογραφία τα ασαφή μοντέλα τύπου Takagi-Sugeno-Kang (TSK) τα οποία γενικά παρουσιάζουν πολύ καλή ακρίβεια προσέγγισης με μικρότερο αριθμό κανόνων. Τα μοντέλα αυτά είναι δυνατόν να περιγραφούν και πάλι από μια ιεραρχική δομή 6 στρωμάτων (Adaptive Neuro-Fuzzy Inference System). Το δίκτυο ANFIS είναι ίδιο με αυτό με κανόνες singleton με την μόνη διαφορά να έγκειται στην μορφή των κόμβων του στρώματος 5, οι οποίοι πλέον υλοποιούν πολυωνυμικές συναρτήσεις (σχήμα 10.7).

Και στα δύο αυτά μοντέλα, τα στρώματα 1, 3, 4 και 6 δεν περιέχουν παραμέτρους ελέγχου, γι' αυτό λέγονται και στατικά στρώματα. Αντίθετα, το στρώμα 2 περιλαμβάνει τις παραμέτρους των συναρτήσεων συμμετοχής, δηλαδή τις παραμέτρους του τμήματος υπόθεσης, ενώ το στρώμα 5 περιλαμβάνει τις παραμέτρους του τμήματος συμπερασμού. Οι παράμετροι αυτοί πρέπει να προσδιορισθούν με βάση κάποια μέθοδο εκμάθησης έτσι ώστε το μοντέλο να προσεγγίζει την συμπεριφορά του επιθυμητού συστήματος. Η μέθοδος αυτή εκμάθησης



Σχ.10.7

που ζητείται στην εργασία αυτή είναι η υβριδική που για το μη γραμμικό τμήμα (στρώμα 2) χρησιμοποιεί την τεχνική του back propagation(BP) και για το γραμμικό τμήμα (στρώμα 5) αυτή των ελαχίστων τετραγώνων(LSE).

Στόχος της εργασίας αυτής είναι να διερευνηθεί η ικανότητα των μοντέλων TSK στη μοντελοποίηση πολυμεταβλητών, μη γραμμικών συναρτήσεων. Έτσι, γίνεται εφαρμογή του μοντέλου σε δύο διαφορετικά datasets όπως παρουσιάζεται στην συνέχεια με την χρήση του περιβάλλοντος *matlab*.

2 Εφαρμογή σε απλό dataset

Στην πρώτη φάση της εργασίας, επιλέγεται από το UCI repository το Airfoil Self-Noise dataset, το οποίο περιλαμβάνει 1503 δείγματα (instances) και 6 χαρακτηριστικά (5 εισόδους και 1 έξοδος).

2.1 Διαχωρισμός σε σύνολα εκπαίδευσης-επικύρωσης-ελέγχου

Αρχικά, κρίνεται απαραίτητος ο διαχωρισμός του συνόλου δεδομένων σε τρία μη επικαλυπτόμενα υποσύνολα D_{trn} , D_{val} , D_{chk} (όπου $D_{trn} \cup D_{val} \cup D_{chk} = D$), από τα οποία το πρώτο θα χρησιμοποιηθεί για εκπαίδευση (με βάση το ελάχιστο MSE_{trn}), το δεύτερο για επικύρωση και αποφυγή του φαινομένου υπερεκπαίδευσης (με βάση το ελάχιστο MSE_{val}) και το τελευταίο για τον έλεγχο

της απόδοσης του τελικού μοντέλου(με βάση το ελάχιστο MSE_{chk}). Το MSE(Mean Squared Error) αποτελεί την συνάρτηση σφάλματος που χρησιμοποιείται στο πρόβλημα αυτό βελτιστοποίησης και στην πραγματικότητα είναι ισοδύναμο με την διακύμανση (variance) των σφαλμάτων πρόβλεψης στο σύνολο των προτύπων που εξετάζουμε. Η διαμέριση που χρησιμοποιήθηκε είναι 60% - 20% - 20% για τα τρία σύνολα αντίστοιχα.

2.2 Εκπαίδευση TSK μοντέλων με διαφορετικές παραμέτρους

Ζητείται να εκπαιδευθούν τέσσερα TSK μοντέλα για το παραπάνω dataset με 5 εισόδους και 1 έξοδο το καθένα. Τα χαρακτηριστικά των μοντέλων αυτών δίνονται στον ακόλουθο πίνακα. Οι συναρτήσεις συμμετοχής είναι bell-shaped και η αρχικοποίησή τους γίνεται με τέτοιο τρόπο ώστε τα διαδοχικά ασαφή σύνολα να παρουσιάζουν επικάλυψη περίπου 0.5.

	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

Στις ακόλουθες 4 υποενότητες απεικονίζονται οι τελικές μορφές των ασαφών συνόλων που προέκυψαν μέσω της διαδικασίας εκπαίδευσης, τα διαγράμματα μάθησης (learning curves) και τα διαγράμματα που αποτυπώνονται τα σφάλματα πρόβλεψης του κάθε μοντέλου.

2.2.1 TSK_model_1

Figure 1: Συναρτήσεις Μεταφοράς του TSK_model_1

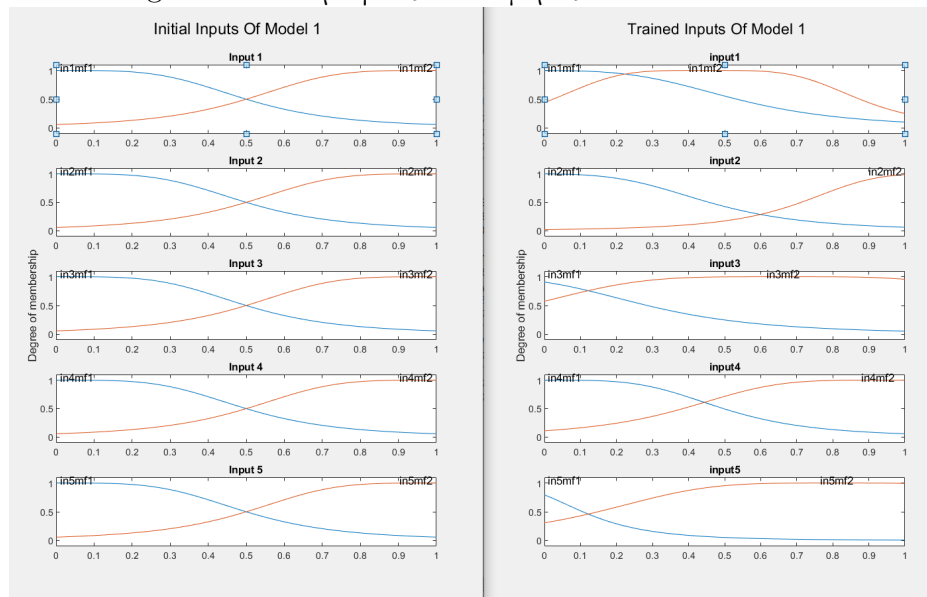


Figure 2: Διάγραμμα μάθησης του TSK_model_1

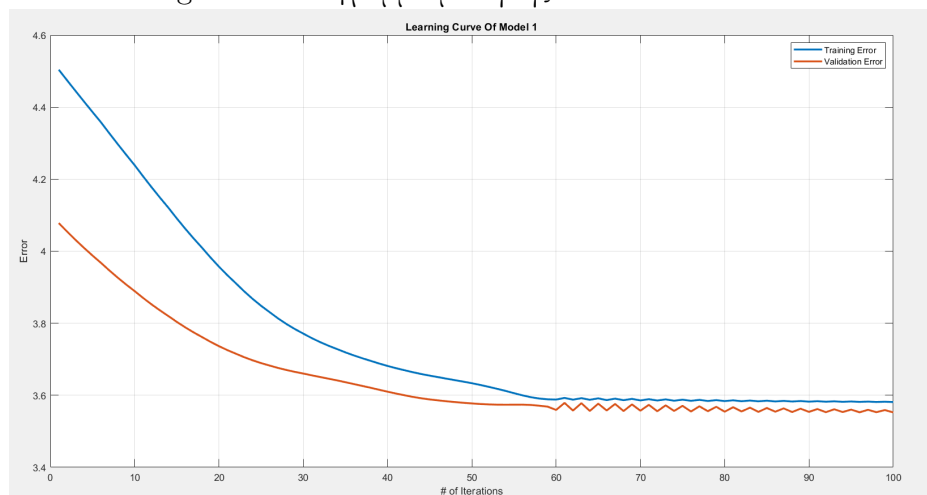
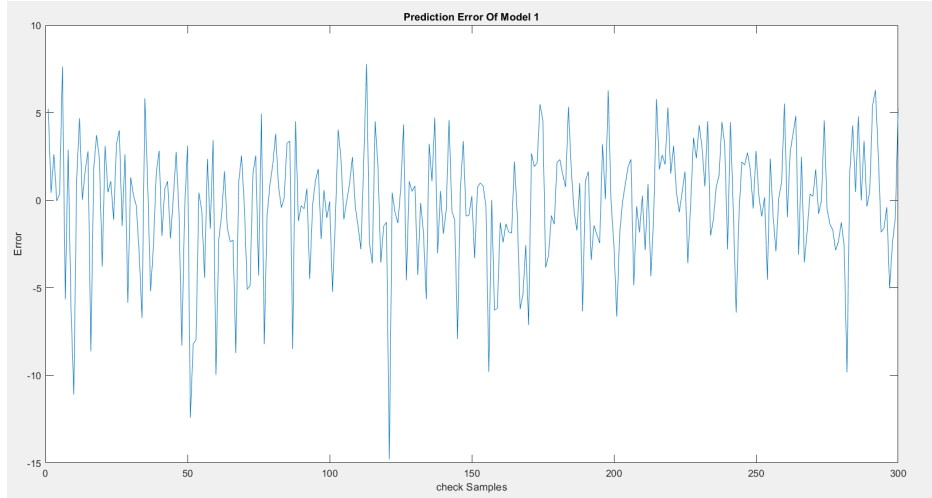


Figure 3: Σφάλμα πρόβλεψης του TSK_model_1



2.2.2 TSK_model_2

Figure 4: Συναρτήσεις Μεταφοράς του TSK_model_2

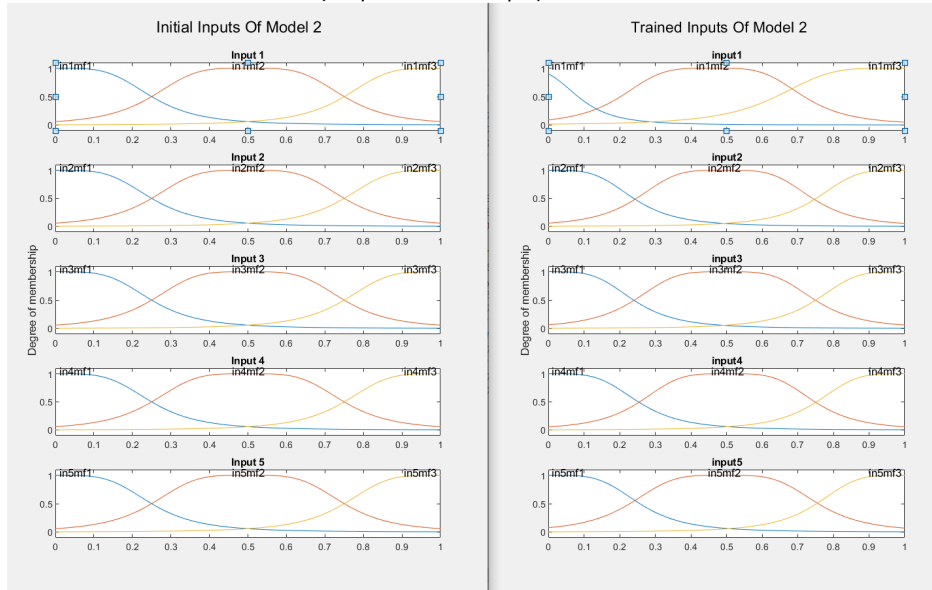


Figure 5: Διάγραμμα μάθησης του TSK_model_2

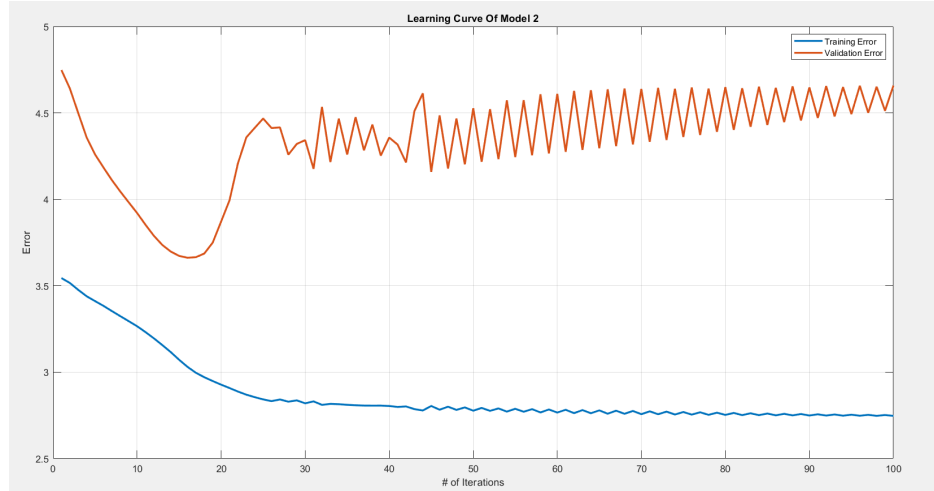
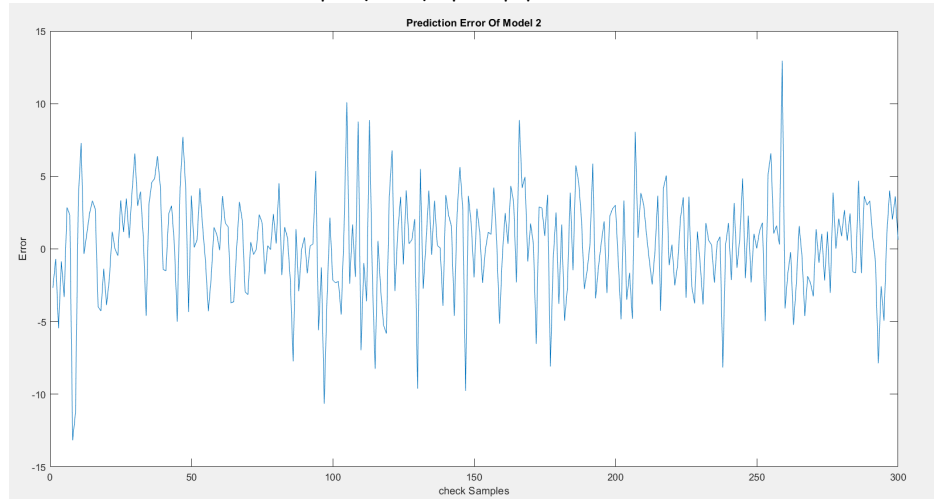


Figure 6: Σφάλμα πρόβλεψης του TSK_model_2



2.2.3 TSK_model_3

Figure 7: Συναρτήσεις Μεταφοράς του TSK_model_3

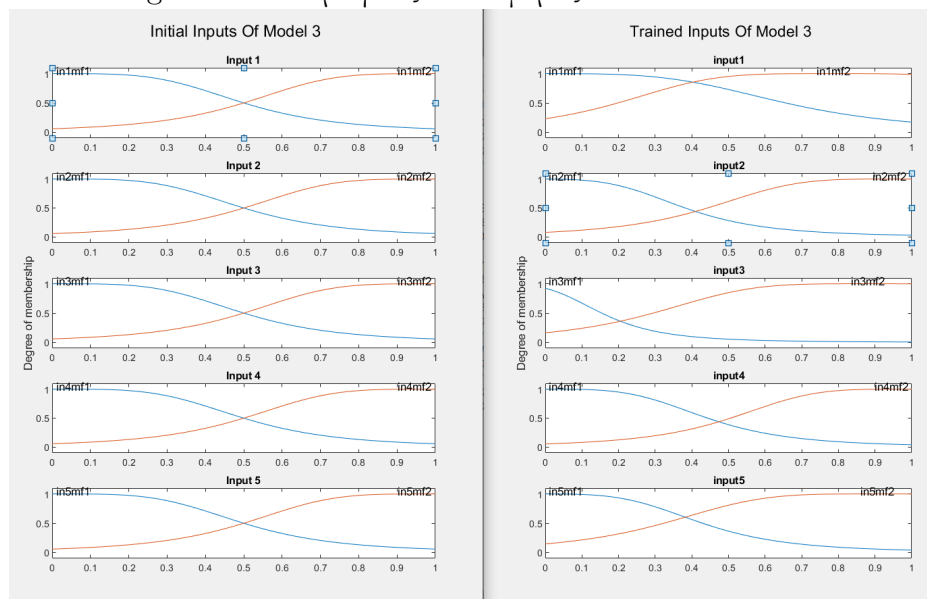


Figure 8: Διάγραμμα μάθησης του TSK_model_3

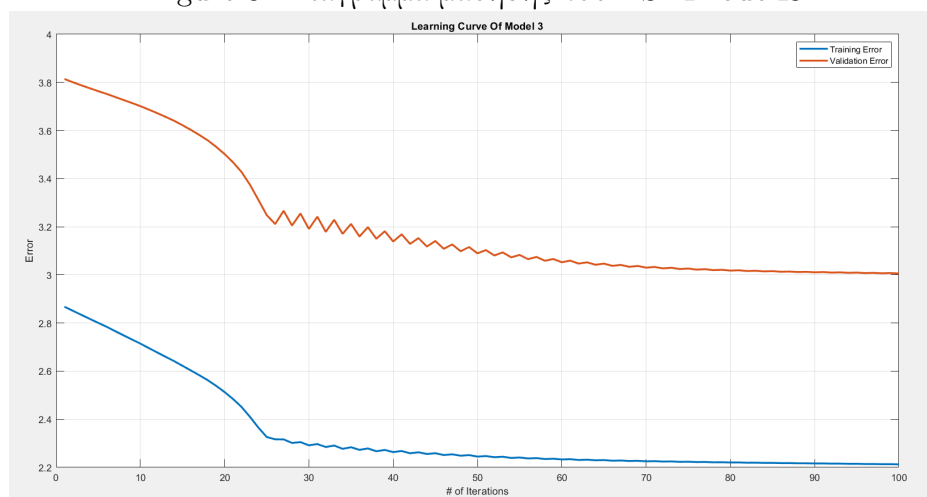
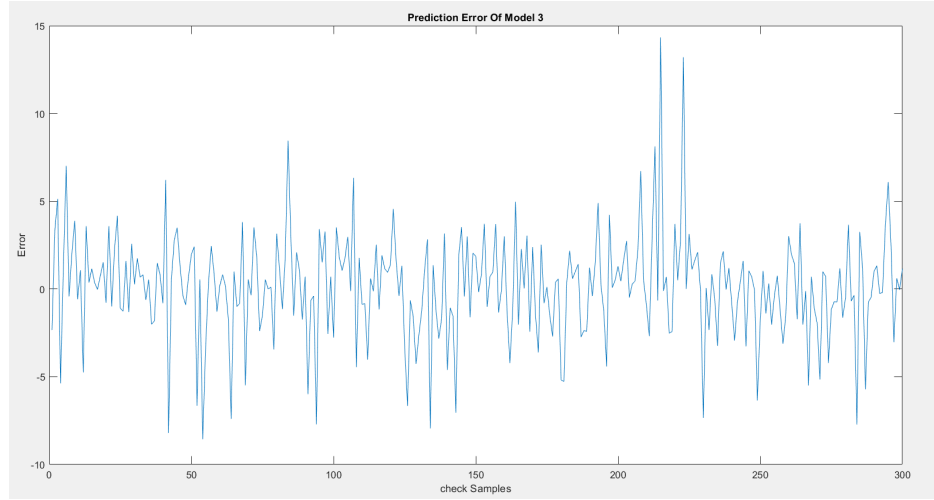


Figure 9: Σφάλμα πρόβλεψης του TSK_model_3



2.2.4 TSK_model_4

Figure 10: Συναρτήσεις Μεταφοράς του TSK_model_4

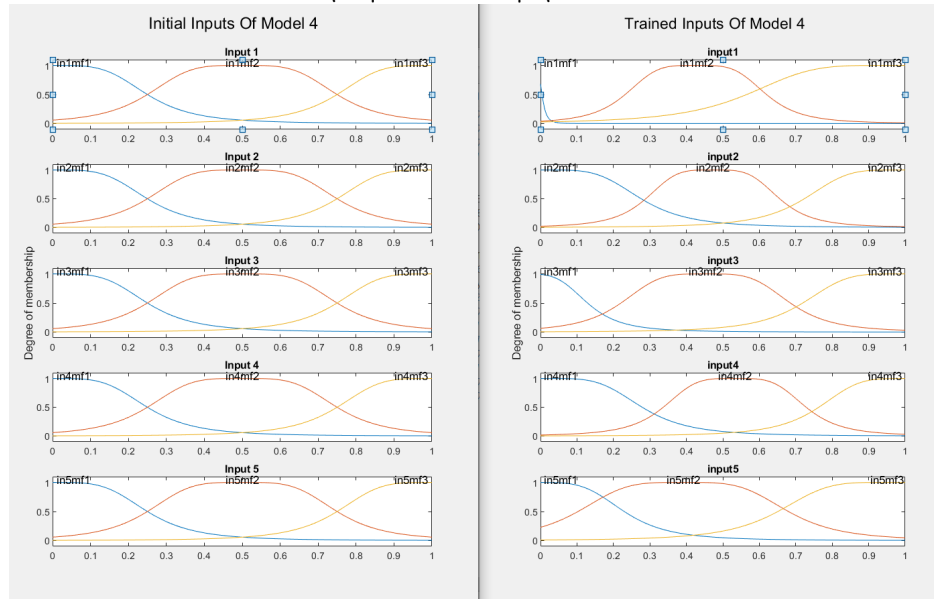


Figure 11: Διάγραμμα μάθησης του TSK_model_4

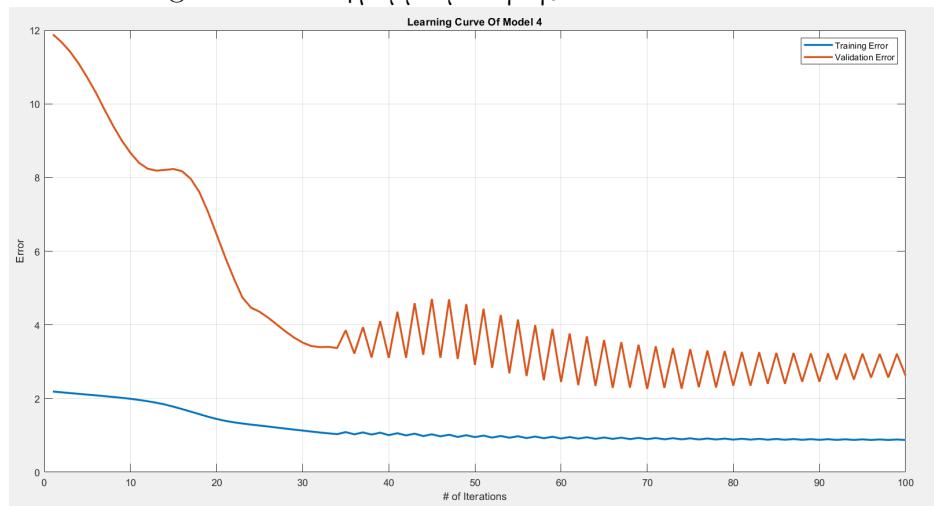
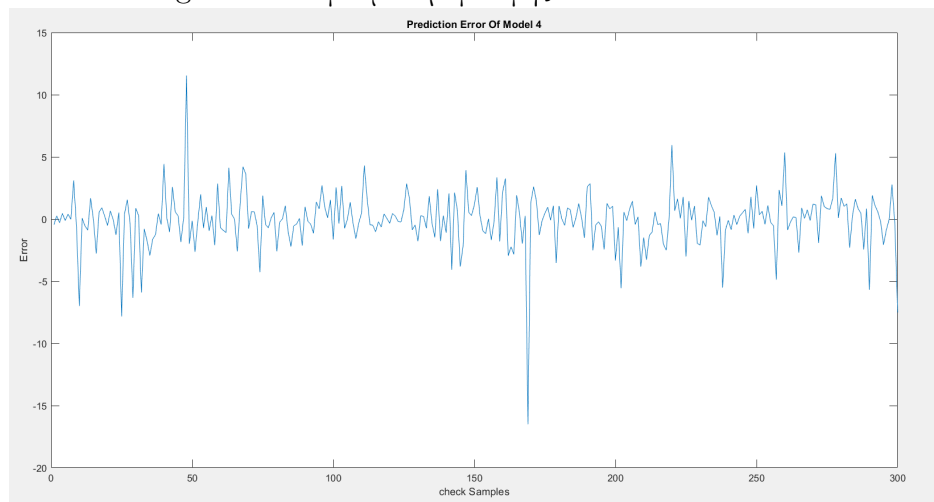


Figure 12: Σφάλμα πρόβλεψης του TSK_model_4



2.3 Αξιολόγηση μοντέλων

	TSK_model_1	TSK_model_2	TSK_model_3	TSK_model_4
RMSE	3.898	3.084	2.814	2.2978
NMSE	0.293	0.1959	0.1737	0.1059
NDEI	0.542	0.442	0.4168	0.3254
R^2	0.706	0.804	0.82639	0.8941

Απο τα αποτελέσματα της προσομοίωσης και τον υπολογισμό των μεταβλητών αξιολόγησης που παρουσιάζονται και στον παραπάνω πίνακα καταλήγουμε ότι το καλύτερο μοντέλο είναι το TSK_model_4 με $R^2 = 89.4\%$ και το χειρότερο το TSK_model_1 με $R^2 = 70.6\%$.

Επιπλέον, η αύξηση του πλήθους των συναρτήσεων συμμετοχής δηλαδή η αύξηση των πιθανών τιμών ανά ασαφή μεταβλητή εισόδου οδηγεί σε καλύτερα αποτελέσματα για ίδια μορφή εξόδου του μοντέλου (διαφορά του μοντέλου TSK_model_1 από το TSK_model_2, και του μοντέλου TSK_model_3 από το TSK_model_4)

Επιπρόσθετα, για ίδιο πλήθος συναρτήσεων συμμετοχής ανά ασαφή μεταβλητή εισόδου, η διατήρηση περισσότερων όρων στην έξοδο του κάθε κανόνα οδηγεί σε καλύτερα αποτελέσματα (διαφορά του μοντέλου TSK_model_1 από το TSK_model_3, και του μοντέλου TSK_model_2 από το TSK_model_4)

Ακόμη, τα μοντέλα (TSK_model_1 και TSK_model_3) παρατηρήθηκε να λειτουργούν με πολύ μεγαλύτερη ταχύτητα και αυτό αποδίδεται στην εκθετική αύξηση των κανόνων με την αύξηση των συναρτήσεων συμμετοχής. Το μοντέλο TSK_model_4 επιβαρύνεται ακόμη περισσότερο με την πολυωνυμική συνάρτηση συμπερασμού.

Τέλος, στο TSK_model_2 παρατηρείται overfitting σε αντίθεση με τα άλλα 3 μοντέλα, γεγονός που φαίνεται από την απόκλιση της καμπύλης μάθησης στο validation set, η οποία δεν εμφανίζει μονότονα φθίνουσα μορφή (όπως εμφανίζει στο training set). Τα μοντέλα με τρεις Membership functions είναι πιο επιρρεπή σε υπερεκπαίδευση, μιας και έχουν περισσότερες συναρτήσεις συμμετοχής που σημαίνει ότι έχουν την δυνατότητα να 'μάθουν' καλύτερα το training dataset. Αυτό φάνηκε και σε αρκετές προσομοιώσεις που το μοντέλο 4 ήταν χειρότερο από το 3 εξαιτίας του φαινομένου Overfitting.

Η προσομοίωση που οδηγεί και στα προαναφερθέντα αποτελέσματα γίνεται με την χρήση του *Matlab* και πιο συγκεκριμένα μέσω του αρχείου *Tsk_reggretion.m* δίνοντας ως είσοδο κατά την εκτέλεση τον αριθμό του μοντέλου που ζητείται η ανάλυση("1","2","3" ή "4"). Το script αυτό για τον διαχωρισμό των δεδομένων εισόδου κάνει χρήση της συνάρτησης *split_data* της οποίας η είσοδος είναι τα δεδομένα και έξοδος τα 3 set δεδομένων [*trnData*, *chkData*, *tstData*]

3 Εφαρμογή σε dataset με υψηλή διαστασιμότητα

3.1 Μοντελοποίησης άγνωστης συνάρτησης

Το dataset που επιλέγεται στο συγκεκριμένο ερώτημα είναι το Superconductivity dataset από το UCI Repository, το οποίο περιλαμβάνει 21263 δείγματα καθένα από τα οποία περιγράφεται από 81 μεταβλητές/χαρακτηριστικά και 1 έξοδο.

Είναι φανερό ότι το dataset καθιστά απαγορευτική μια απλή εφαρμογή ενός TSK μοντέλου εξαιτίας του μεγάλου βαθμού διαστασιμότητας (high-dimensionality).

Ο αριθμός των απαιτούμενων κανόνων με βάση το grid partitioning στο χώρο των εισόδων του μοντέλου θα είναι 2^{81} κανόνες (διαμερίζοντας το χώρο εισόδου κάθε μεταβλητής με δύο ασαφή σύνολα). Για το λόγο αυτό, όπως αναφέρεται και στην εκφώνηση του 2ου μέρους της παρούσας εργασίας, θα χρησιμοποιήσουμε τεχνικές μείωσης της διαστασιμότητας και του απαιτούμενου αριθμού ασαφών(IF-THEN) κανόνων. Οι δύο τεχνικές που θα χρησιμοποιηθούν είναι: α) Ο αλγόριθμός ReliefF για feature subset selection. Αποδίδεται δηλαδή ένα σκορ σημαντικότητας σε κάθε κανόνα.

β) Ο αλγόριθμος Subtractive Clustering για ομαδοποίηση των κανόνων.

Πριν την δημιουργία των κανόνων τα δεδομένα εισόδου προβάλλονται στο χώρο και ομαδοποιούνται. Κάθε ομάδα έχει το δικό της κέντρο που είναι απόρεια μιας συνάρτησης δυναμικού των σημείων της ομάδας. Τέλος, από κάθε κέντρο προκύπτει ένας κανόνας.

Οι δύο αυτές μέθοδοι όμως, παρά τη ελάττωση της πολυπλοκότητας που επιφέρουν, εισάγουν στο πρόβλημα δύο ελεύθερες παραμέτρους. Συγκεκριμένα, τον αριθμό των χαρακτηριστικών(Features) προς επιλογή και το μέγεθος της ακίννας(radius).

3.2 Μοντελοποίηση και Βέλτιστη σχεδίαση

Τα στάδια μοντελοποίησης του προβλήματος αναλύονται στις ακόλουθες υποενότητες. Η προσομοίωση για την συνολική αυτή ανάλυση έγινε στο περιβάλλον του *Matlab* και υπάρχει στο αρχείο *TSK_high_dim.m*. Επιπλέον, εξαιτίας της μεγάλης καθυστέρησης που επιφέρει η αναζήτηση των βέλτιστων τιμών για την μοντελοποίηση έχει προστεθεί στην εργασία ο πίνακας *bestmodel.mat* που

περιέχει τις τιμές ενός section.

3.2.1 Διαχωρισμός σε σύνολα εκπαίδευσης- επικύρωσης – ελέγχου:

Όπως και στο πρώτο κομμάτι της εργασίας, στην αρχή, είναι απαραίτητος ο διαχωρισμός του συνόλου δεδομένων σε τρία υποσύνολα D_{trn} , D_{val} , D_{chk} , με αναλογία 60%-20%-20%.

3.2.2 Επιλογή των βέλτιστων παραμέτρων:

Οι δύο ελεύθερες παράμετροι που εισάγονται πρέπει να οριστούν και για την αποτελεσματική επιλογή τους χρησιμοποιείται η αναζήτηση πλέγματος (grid search) σε συνδυασμό με την αξιολόγηση 5-πτυχης διασταυρωμένης επικύρωσης (5-fold cross validation) στο σύνολο εκπαίδευσης D_{trn} . Ορίζεται δηλαδή ένα δυσδιάστατο πλέγμα με τιμές:

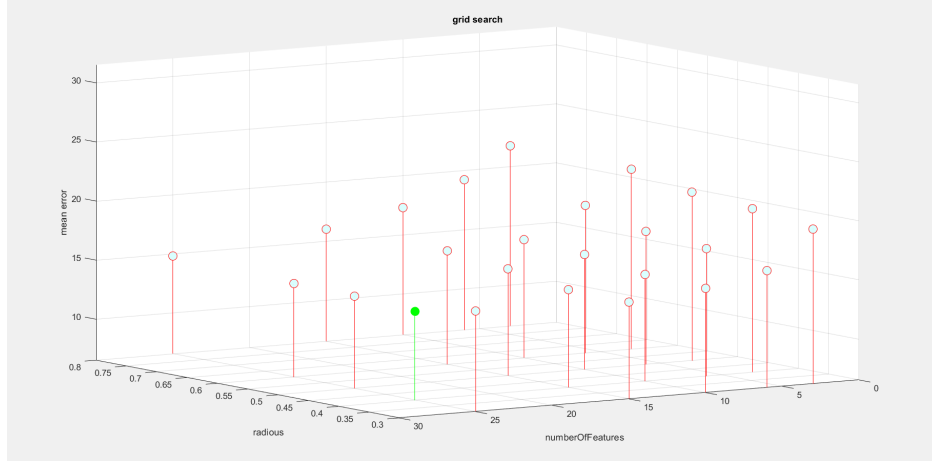
- Αριθμός χαρακτηριστικών: $Nfeatures = [3, 6, 10, 15, 25]$
- Ακτίνα των clusters rad : $rad = [0.3, 0.4, 0.5, 0.6, 0.8]$

Έχοντας λοιπόν τα τελικά validation errors των grid points για κάθε cross-validation run, υπολογίζουμε το μέσο όρο αυτών και κατόπιν επιλέγουμε σαν βέλτιστο αυτό του οποίου ο μέσος όρος είναι ο ελάχιστος. Παρακάτω, παρατίθεται ο μέσος όρος των τελικών validation errors για κάθε grid point:

rad \ Nfeatures	0.3	0.4	0.5	0.6	0.8
3	19.6232	20.3703	20.7930	21.7536	21.7818
6	16.4294	17.3124	17.7938	19.0170	19.2333
10	15.3604	15.5561	16.2705	16.5469	17.2989
15	14.7340	14.8072	15.5916	16.1361	16.0118
25	15.0525	14.0387	14.3469	14.4443	14.8171

Table 1: Μέσο validation error για κάθε grid point ως προς τα 5 folds

Figure 13: grid



Όπως λοιπόν φαίνεται στον πίνακα 1 αλλά και στην τρισδιάστατη απεικόνιση (εικόνα 13) οι τιμές $(Nfeatures, rad) = (25, 0.4)$ επιφέρουν το μικρότερο σφάλμα.

Γενικότερα, η μείωση του αριθμού των χρησιμοποιούμενων χαρακτηριστικών ($Nfeatures$) οδηγεί σε μεγαλύτερα μέσα validation errors, κάτι απολύτως αναμενόμενο εάν λάβουμε υπόψη την υψηλή διαστασιμότητα του dataset. Παρόλα αυτά, μπορεί τα 25 (Features) να είναι η βέλτιστη λύση αλλά αποδεικνύεται ιδιαίτερα χρονοβόρα η διαδικασία κατασκευής της. Θα μπορούσε λοιπόν, με αμελητέα επιβάρυνση του σφάλματος να χρησιμοποιηθεί η $(Nfeatures, rad) = (15, 0.3)$.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
66	63	31	1	23	13	15	71	81	69	37	21	76	14	67	12	26	16	25	36	22	28	57	78	6

Table 2: Τα 25 πιο σημαντικά χαρακτηριστικά (features)

3.3 Απόδοση τελικού μοντέλου

Για την τιμή του $(Nfeatures, rad) = (25, 0.4)$ εκπαιδεύεται το τελικό μοντέλο στο αρχικό split του dataset. Έτσι, προκύπτουν οι καμπύλες μάθησης

(σχήμα 14), το σφάλμα πρόβλεψης(σχήμα 15), και οι προβλέψεις των 500 πρώτων στοιχείων (σχήμα 16). Απο την φθίνουσα πορεία του *ValidationError* συμπεραίνεται οτι δεν υπάρχει το φαινόμενο της υπερεκπαίδευσης (overfitting).

Figure 14: Διάγραμμα μάθησης

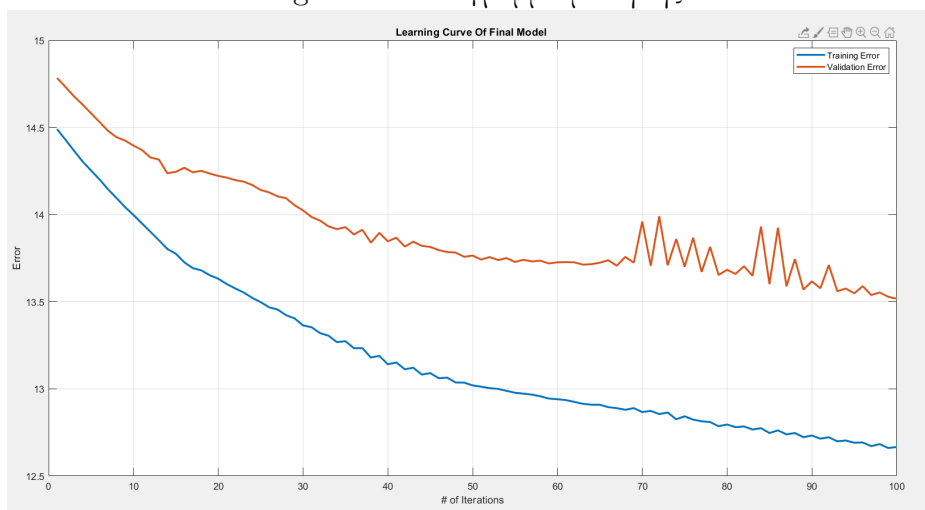


Figure 15: Σφάλμα πρόβλεψης

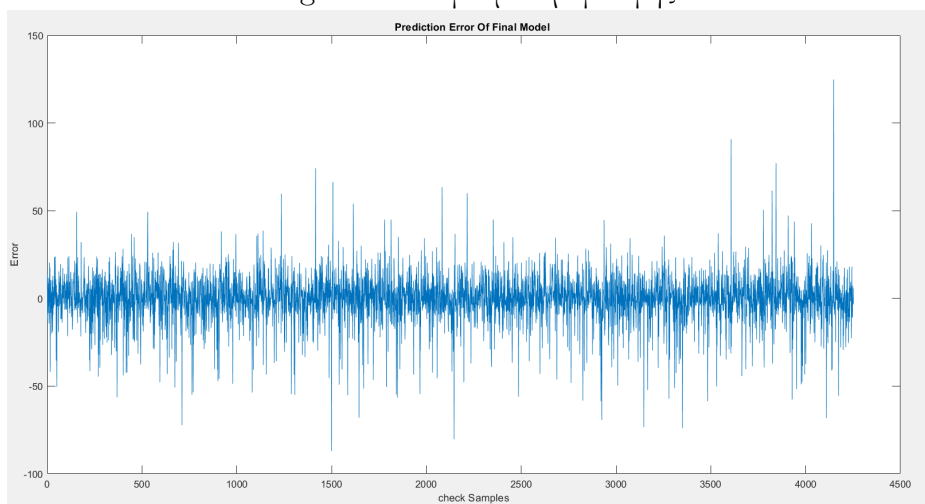
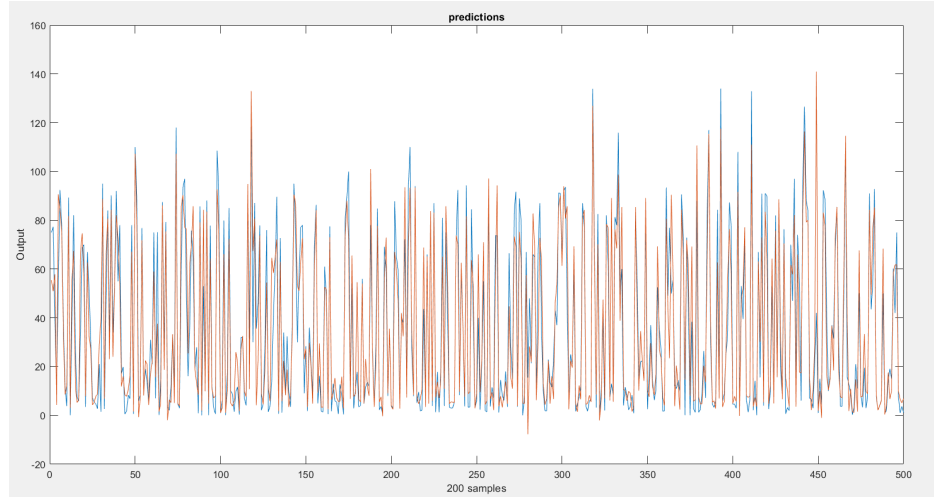


Figure 16: Προβλέψεις των 500 πρώτων στοιχείων



Στην συνέχεια, παρουσιάζονται οι συναρτήσεις συμμετοχής (σχήματα 17-19) πριν και μετά την εκπαίδευση του μοντέλου για τις 25 εισόδους.

Figure 17: Συναρτήσεις συμμετοχής κατά την εκπαίδευση (1-10)

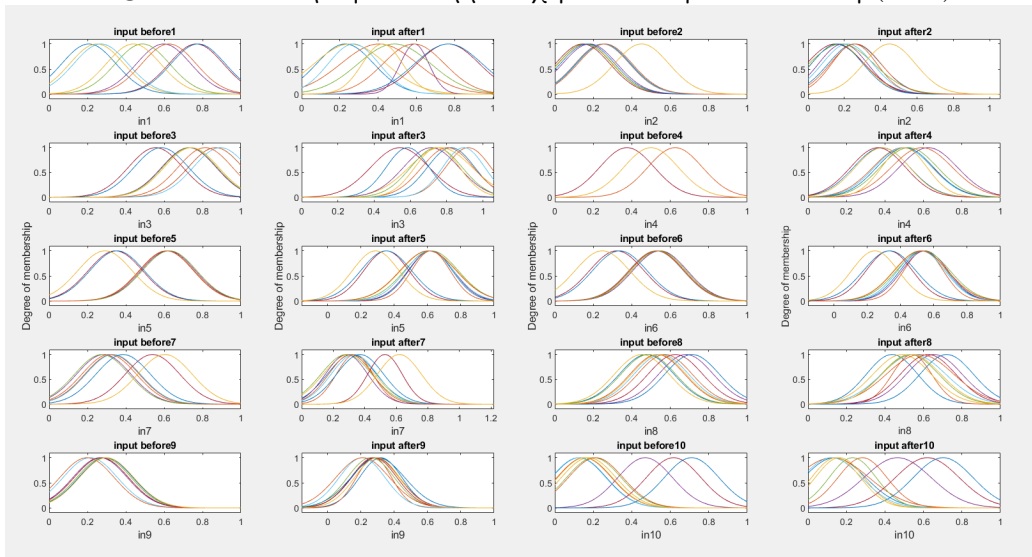


Figure 18: Συναρτήσεις συμμετοχής κατά την εκπαίδευση (11-20)

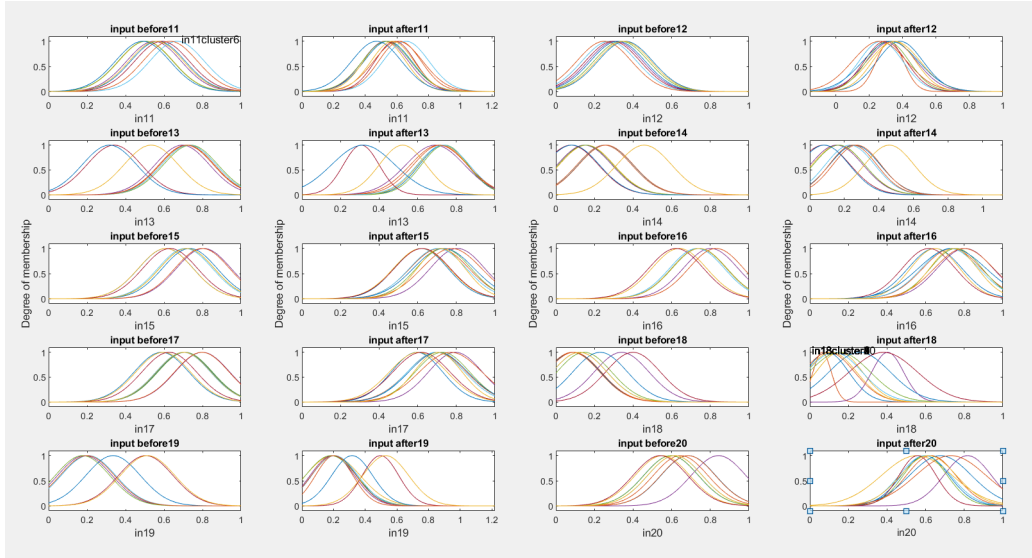
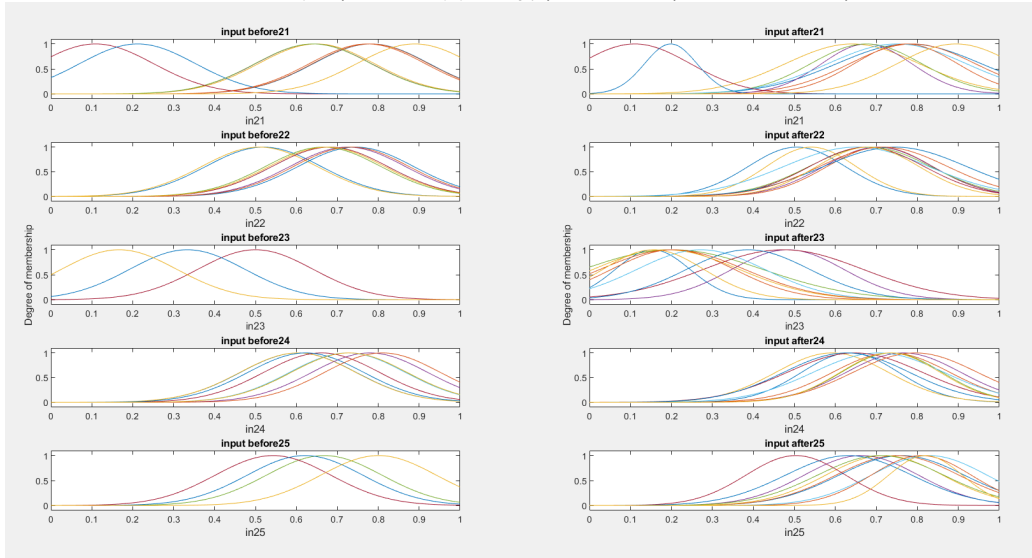


Figure 19: Συναρτήσεις συμμετοχής κατά την εκπαίδευση (20-25)



Όπως φαίνεται το μοντέλο με τον βέλτιστο συνδυασμό εμφανίζει αρκετά ικανοποιητικά αποτελέσματα με $R^2 = 0.843$.

RMSE	13.626
NMSE	0.157
NDEI	0.396
R^2	0.843

Table 3: τιμές των δεικτών απόδοσης

3.4 Συμπέρασμα

Παρατηρούμε ότι για το δοσμένο dataset η ανάπτυξη ενός FNN κρατώντας μόνο τα 25 (σημαντικότερα) από τα συνολικά 81 χαρακτηριστικά οδηγεί σε ικανοποιητικά αποτελέσματα πρόβλεψης. Προφανώς, οι δείκτες απόδοσης αυτοί είναι χειρότεροι σε σύγκριση με τους αντίστοιχους του πρώτου μέρους, κάτι απολύτως αναμενόμενο αν λάβουμε υπόψη το μέγεθος των εισόδων που δεν χρησιμοποιήθηκαν. Τέλος, η χρήση μόνο 13 ασαφών κανόνων σε σύγκριση με τους $2^{25}3^{25}$ που θα απαιτούνταν για grid partitioning , επιβεβαιώνει το πλεονέκτημα της χρήσης ομαδοποίησης στο χώρο των εισόδων (scatter partitioning).