

Μέρος 1

Το επιλεγμένο σύνολο δεδομένων της NASA περιλαμβάνει αεροτομές διαφορετικού μεγέθους NACA 0012 σε διάφορες ταχύτητες σήραγγας αέρα και γωνίες επίθεσης. Η επιφάνεια της αεροτομής και η θέση του παρατηρητή ήταν τα ίδια σε όλα τα πειράματα. Περιλαμβάνει 1503 δείγματα και κάθε ένα από αυτά εμφανίζει τα ακόλουθα 6 χαρακτηριστικά:

Αυτό το πρόβλημα έχει τις ακόλουθες εισόδους:

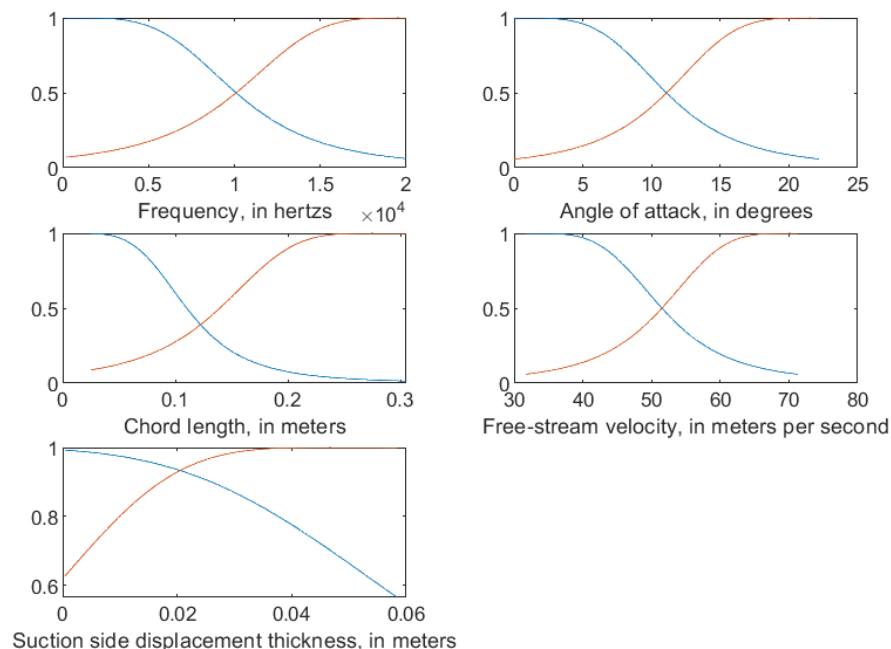
1. Συχνότητα, σε Hertz.
2. Γωνία επίθεσης, σε μοίρες.
3. Μήκος χορδής, σε μέτρα.
4. Ταχύτητα ελεύθερης ροής, σε μέτρα ανά δευτερόλεπτο.
5. Πάχος μετατόπισης πλευρικής αναρρόφησης, σε μέτρα.

Η μόνη έξοδος είναι:

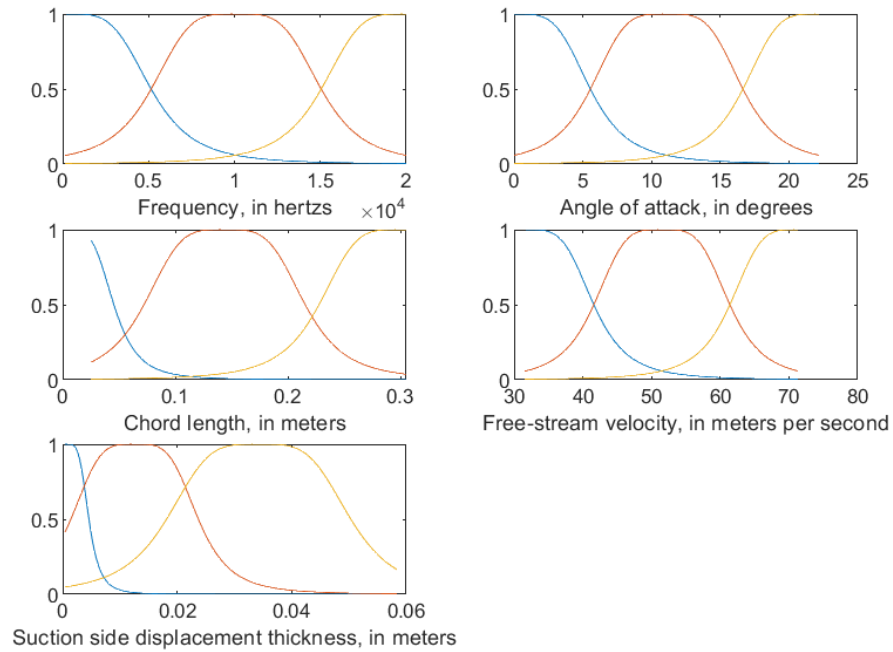
1. Κλίμακα στάθμης ηχητικής πίεσης, σε ντεσιμπέλ

Μετά την αναδιάταξη του συνόλου δεδομένων με τυχαίο τρόπο και το διαμερισμό του σε μη επικαλυπτόμενα σύνολα εκπαίδευσης, επικύρωσης και ελέγχου, όπως ορίζει η εκφώνηση της εργασίας, δημιουργούνται τα τέσσερα μοντέλα TSK που ζητούνται. Μετά την εκπαίδευση των μοντέλων με τη χρήση μια υβριδικής μεθόδου, κατά την οποία οι παράμετροι των συναρτήσεων συμμετοχής και της πολυωνυμικής συνάρτησης βελτιστοποιούνται με τις μεθόδους του back propagation και των ελαχίστων τετραγώνων εξόδου αντίστοιχα, παρουσιάζονται οι συναρτήσεις συμμετοχής των εισόδων που προέκυψαν στα παρακάτω σχήματα.

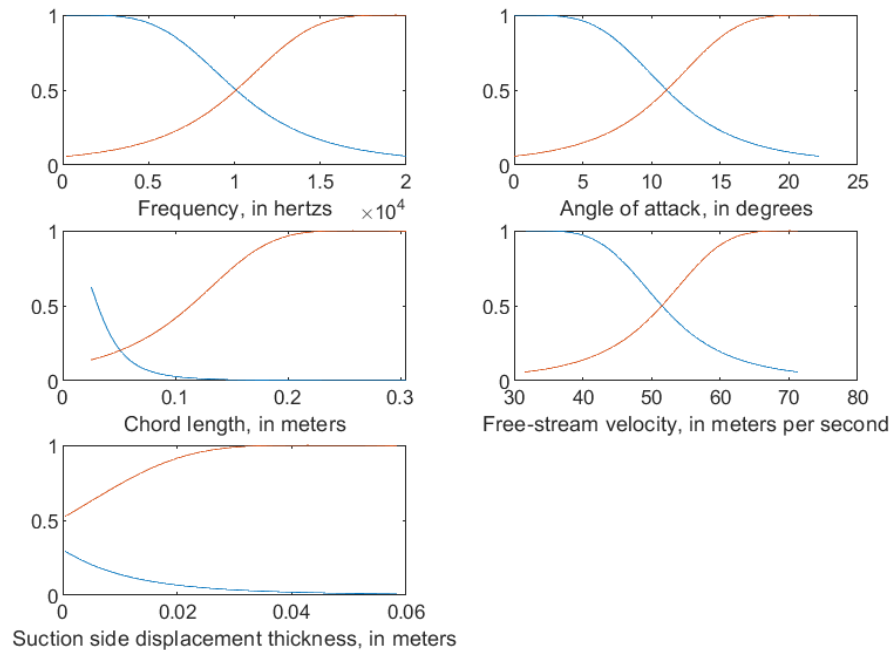
TSK model 1: Trained membership functions



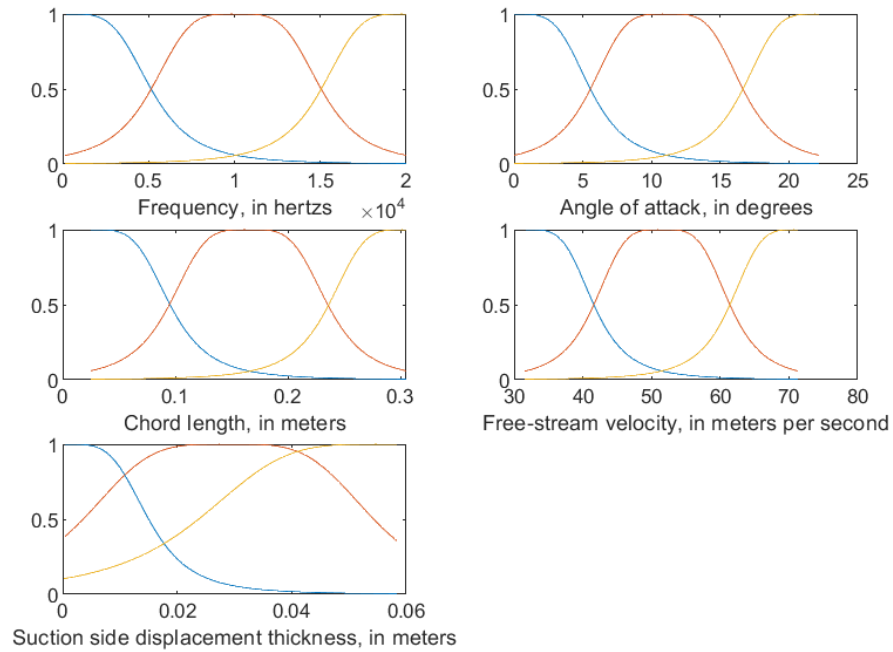
TSK model 2: Trained membership functions



TSK model 3: Trained membership functions

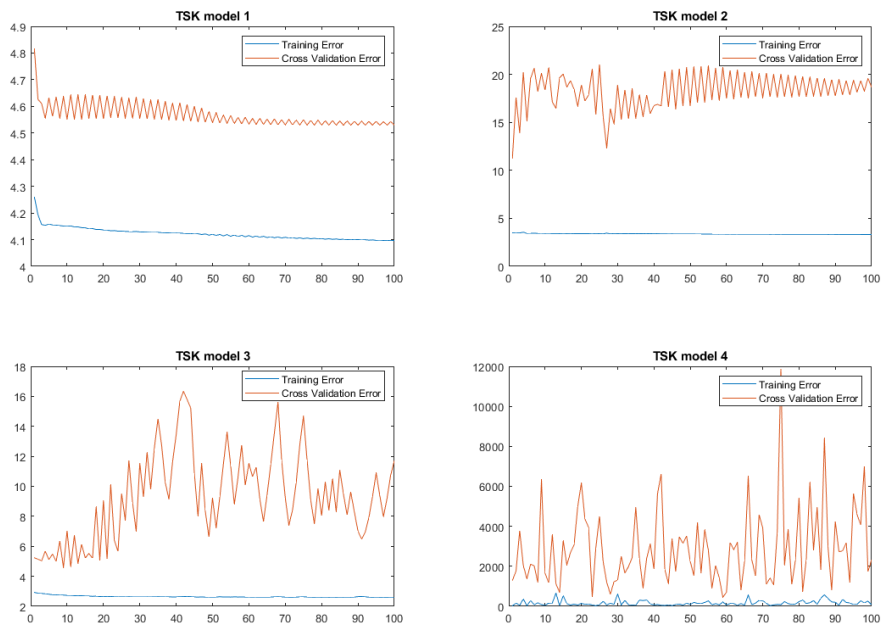


TSK model 4: Trained membership functions

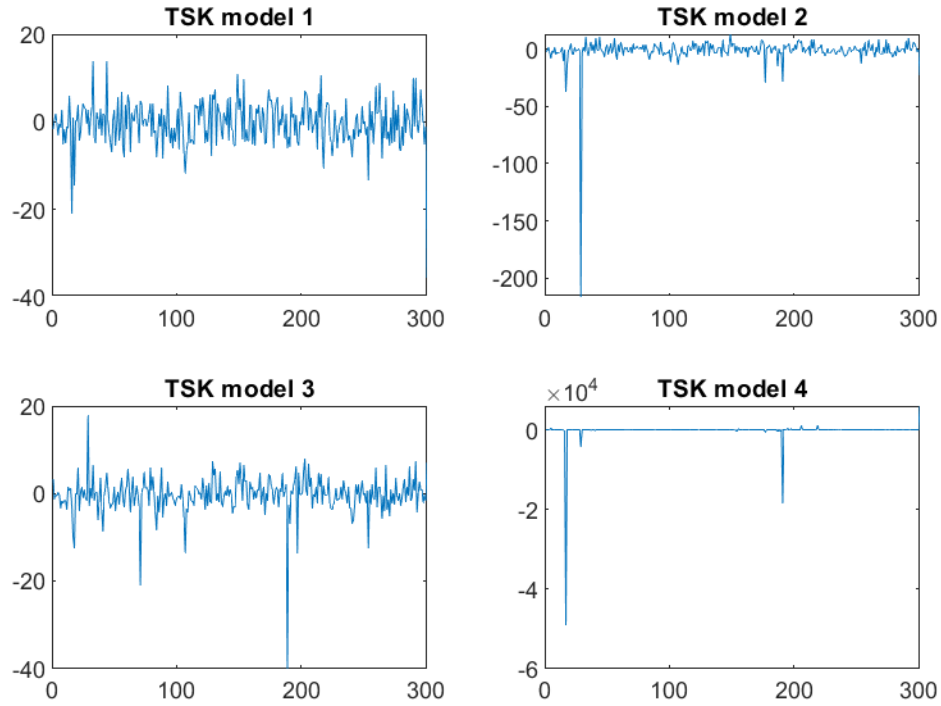


Παρακάτω παρουσιάζονται επίσης οι καμπύλες εκμάθησης και τα σφάλματα πρόβλεψης των τεσσάρων μοντέλων.

Learning Curves



Prediction Errors of models



Μετά το πέρας της διαδικασίας εκπαίδευσης και επικύρωσης, χρησιμοποιούμε τον τελευταίο διαμερισμό του συνόλου δεδομένων για τον έλεγχο της απόδοσης των μοντέλων που παράχθηκαν. Για κάθε ένα από αυτά τα μοντέλα προκύπτουν από το testing οι ζητούμενες παράμετροι εκτίμησης σφάλματος ως δείκτες απόδοσης. Τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα.

	<i>RMSE</i>	<i>NMSE</i>	<i>NDEI</i>	<i>R²</i>
TSK model 1	5.028	0.0016	0.0401	0.9984
TSK model 2	13.6315	0.0118	0.1088	0.9882
TSK model 3	4.3292	0.0012	0.0346	0.9988
TSK model 4	3043.3	590.054	24.291	-589.054

Μέρος 2

Το επιλεγμένο σύνολο δεδομένων είναι το Superconductivity dataset, το οποίο περιλαμβάνει 21263 δείγματα, καθένα από τα οποία περιγράφεται από 81 χαρακτηριστικά. Λόγω της μεγάλης διαστασιμότητας του συνόλου δεδομένων, αρχικά εφαρμόζεται ο αλγόριθμος Relief για την βέλτιστη ταξινόμηση των χαρακτηριστικών, ώστε όταν στη συνέχεια να επιλεγεί μόνο ένας συγκεκριμένος αριθμός χαρακτηριστικών, να χρησιμοποιηθούν εκείνα τα χαρακτηριστικά που συμβάλλουν περισσότερο στην εξαγωγή χρήσιμων συμπερασμάτων.

Πέραν του αριθμού των χαρακτηριστικών που θα επιλεγούν, η δεύτερη ανεξάρτητη μεταβλητή του συγκεκριμένου προβλήματος μοντελοποίησης είναι η ακτίνα των clusters που θα προκύψουν, κάτι που επηρεάζει καθοριστικά και το πλήθος των συναρτήσεων συμμετοχής για το κάθε μοντέλο.

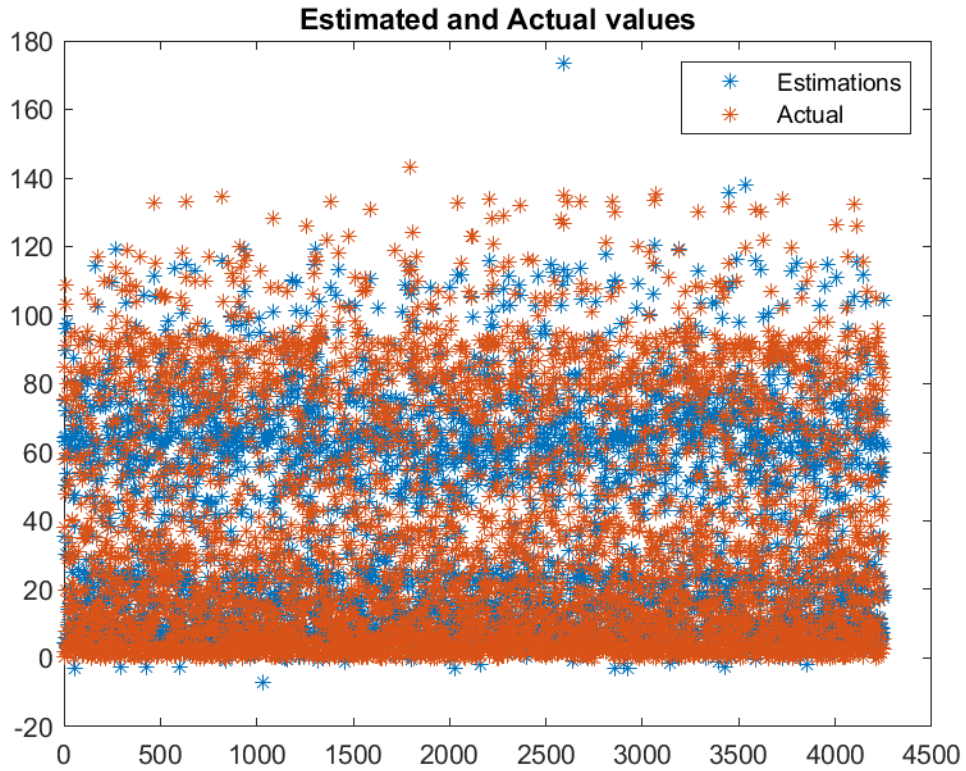
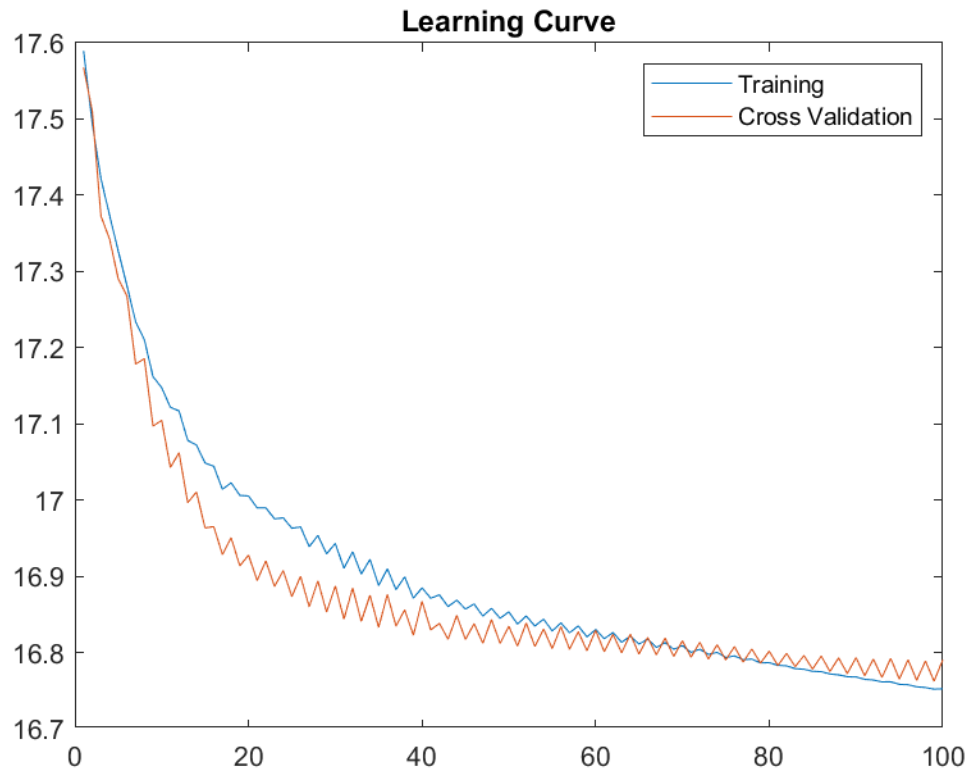
Παράλληλα, γίνεται εφαρμογή subtractive clustering ως μέθοδος ομαδοποίησης για τη δημιουργία των IF-THEN κανόνων.

Στη συνέχεια εφαρμόζεται πενταπτυχής διασταυρωμένη επικύρωση σε κάθε φορά τυχαία διαχωρισμένα σύνολα εκπαίδευσης και επαλήθευσης για κάθε ένα από τα πλήθη κανόνων βάσης που έχουν επιλεγεί. Τα μοντέλα που προκύπτουν στη συνέχεια χρησιμοποιούν τη μέθοδο Fuzzy C-means για να ομαδοποιήσουν τα δεδομένα. Τα αποτελέσματα των μέσων όρων σφαλμάτων που προκύπτουν για τους συνδυασμούς των ανεξάρτητων μεταβλητών που επιλέξαμε παρουσιάζονται στον παρακάτω πίνακα.

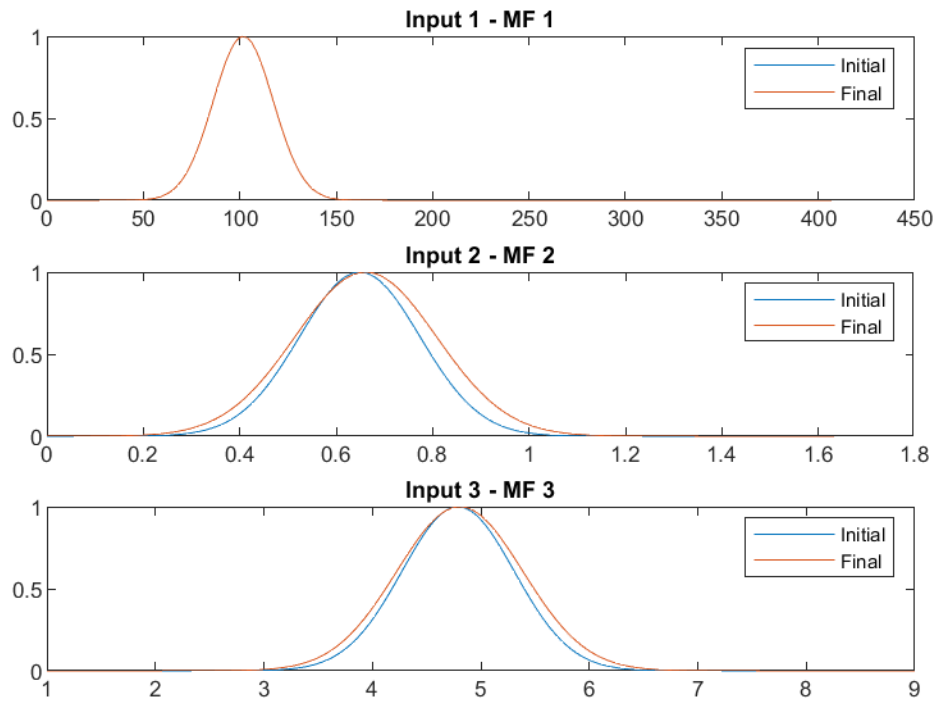
		Number of Features			
		3	6	9	12
Number of Rules	3	21.375	19.1654	18.6624	18.2147
	6	20.644	18.0647	18.1218	17.3649
	9	19.994	17.2127	17.6854	17.1611
	12	19.6074	16.7443	17.1812	16.8894
	15	19.7408	16.8718	17.2482	16.8379

Ο βέλτιστος συνδυασμός αριθμού χαρακτηριστικών και αριθμού συναρτήσεων συμμετοχής είναι εκείνος που επιφέρει το μικρότερο μέσο όρο σφάλματος. Από τον παραπάνω πίνακα είναι προφανές ότι ο συνδυασμός αυτός είναι ο [Number of features, Number of Rules] = [6,12].

Μετά την επιλογή των βέλτιστων παραμέτρων ακολουθεί η εκπαίδευση του μοντέλου και προκύπτουν έτσι η καμπύλη εκμάθησης του μοντέλου, η σύγκριση πραγματικών τιμών και εκτιμήσεων, καθώς και η σύγκριση ορισμένων τυχαίων συναρτήσεων συμμετοχής πριν και μετά τη διαδικασία της εκπαίδευσης και αποτυπώνονται στα παρακάτω σχήματα.



Membership functions



Μετά και τη διαδικασία επαλήθευσης μπορούμε να εκτιμήσουμε και τις κρίσιμες μετρικές σφάλματος του μοντέλου.

<i>RMSE</i>	<i>NMSE</i>	<i>NDEI</i>	<i>R²</i>
17.0039	0.244	0.4939	0.756

Η επιλογή grid partitioning με δύο ή τρία ασαφή σύνολα ανά είσοδο για το ίδιο πλήθος χαρακτηριστικών θα οδηγούσε σε εκθετικά μεγαλύτερο πλήθος κανόνων του ασαφούς συστήματος συμπερασμού, κάτι που θα έκανε τη διαδικασία εκπαίδευσης ενός μοντέλου υπερβολικά πολύπλοκη υπολογιστικά και κοστοβόρα χρονικά.