

Υπολογιστική Νοημοσύνη
Εργασία 3
Regression

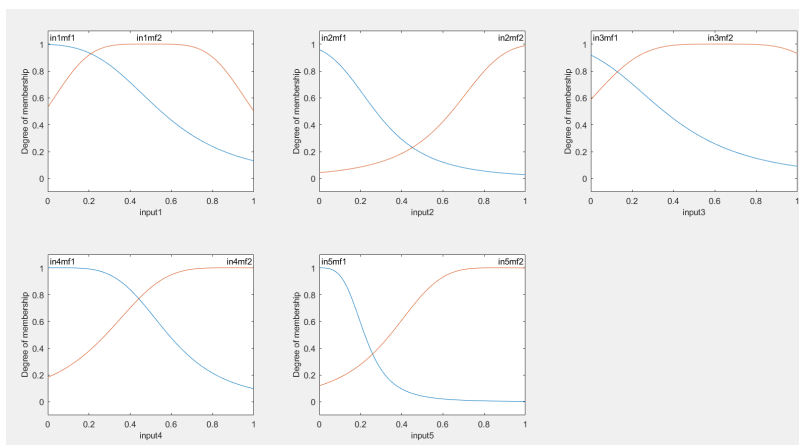
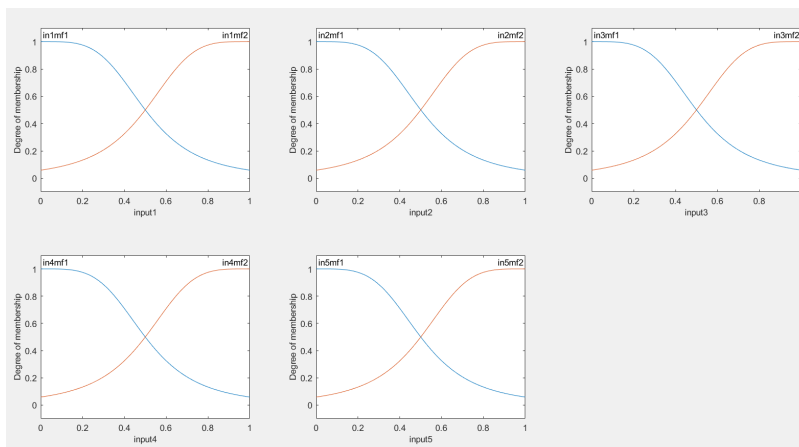
Ηλιάνα Κόγια
(AEM: 10090)
ilianakogia@ece.auth.gr

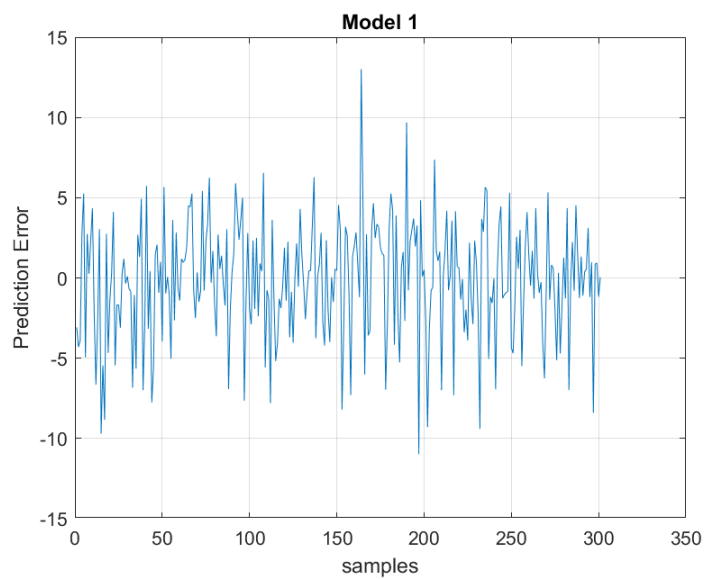
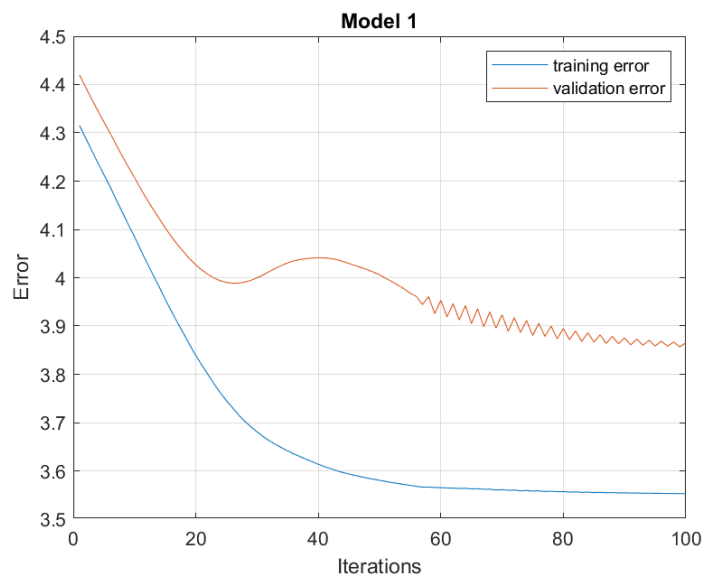
1 Εφαρμογή σε απλό **dataset**

Χωρίζουμε το **dataset** σε train - check - test με αντίστοιχα ποσοστά 60 - 20 -20

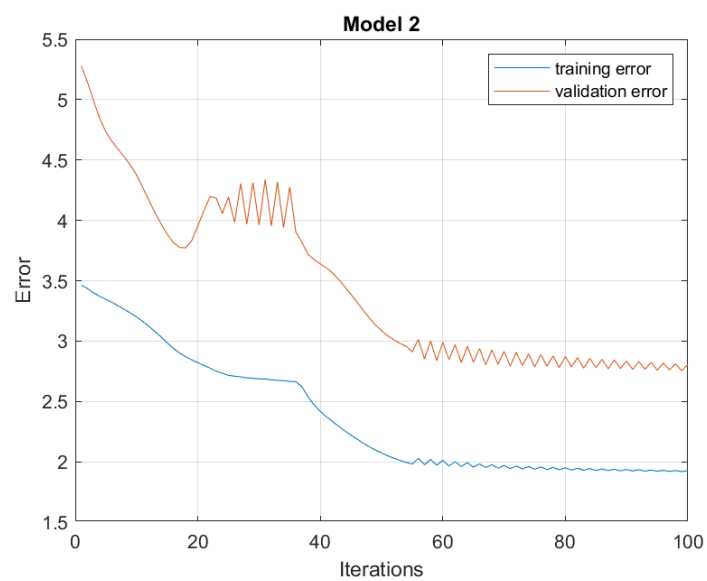
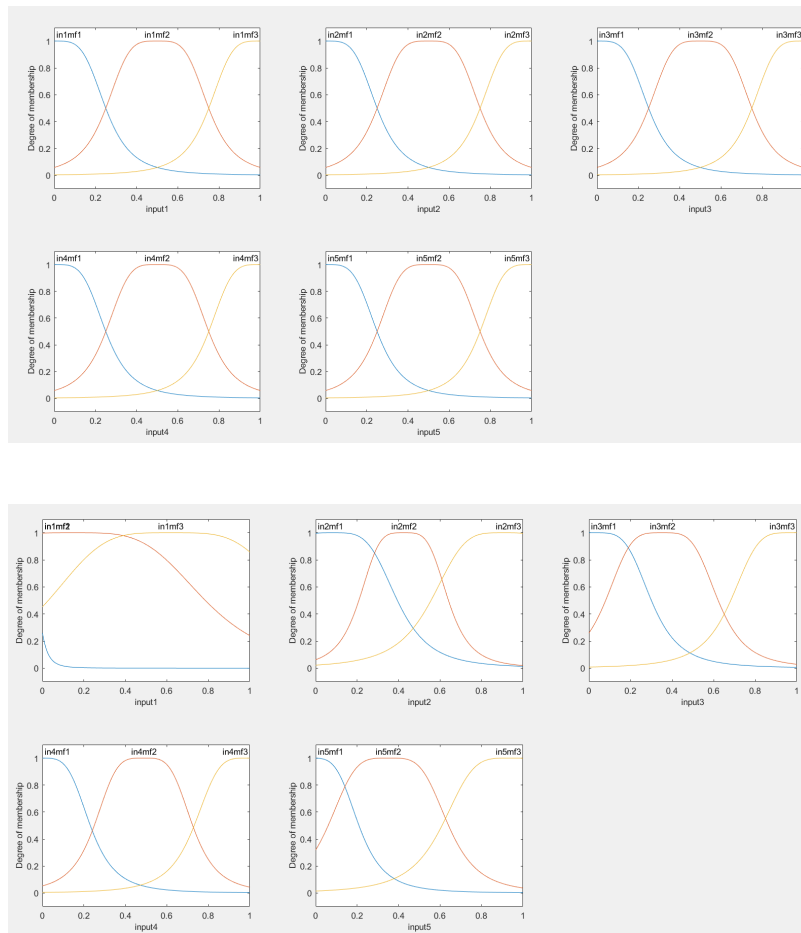
	Πλήθος συναρτήσεων συμμετοχής	Μορφή εξόδου
TSK_model_1	2	Singleton
TSK_model_2	3	Singleton
TSK_model_3	2	Polynomial
TSK_model_4	3	Polynomial

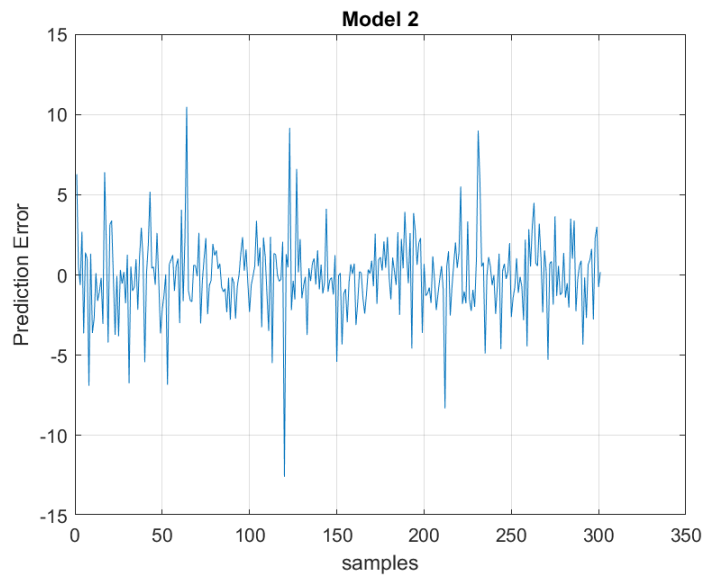
Παρακάτω παρατίθενται τα αποτελέσματα για τα 4 μοντέλα: συναρτήσεις συμμετοχής πριν και μετά την εκπαίδευση, **learning curves** και διαγράμματα των **prediction errors**
Model 1



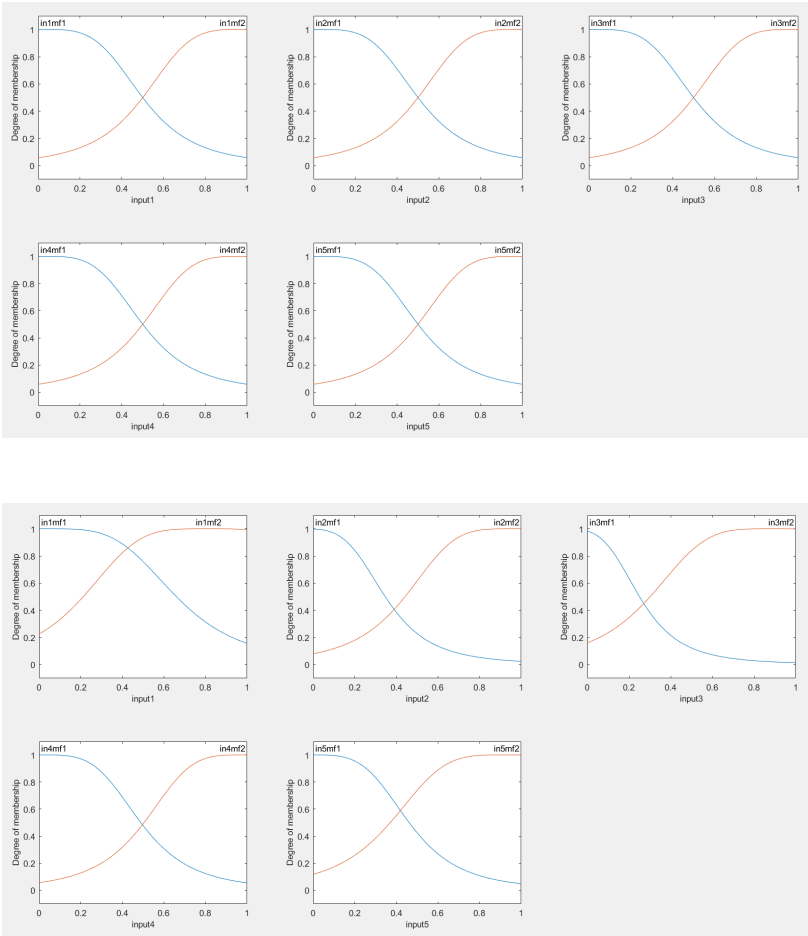


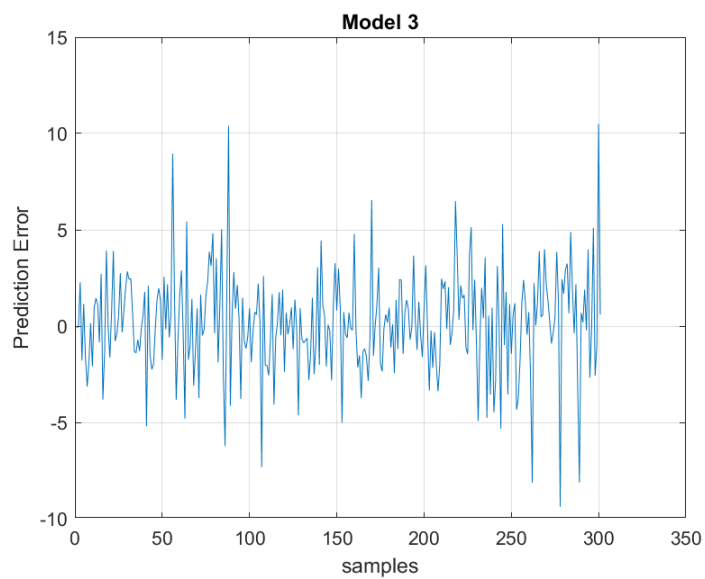
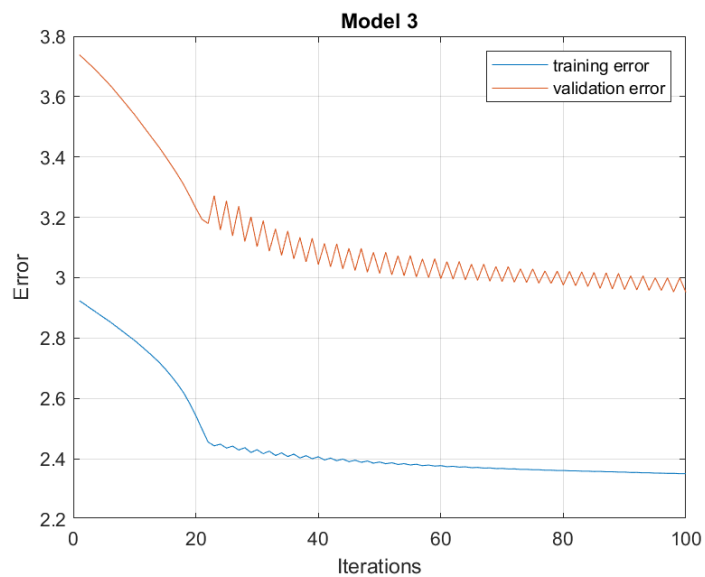
Model 2



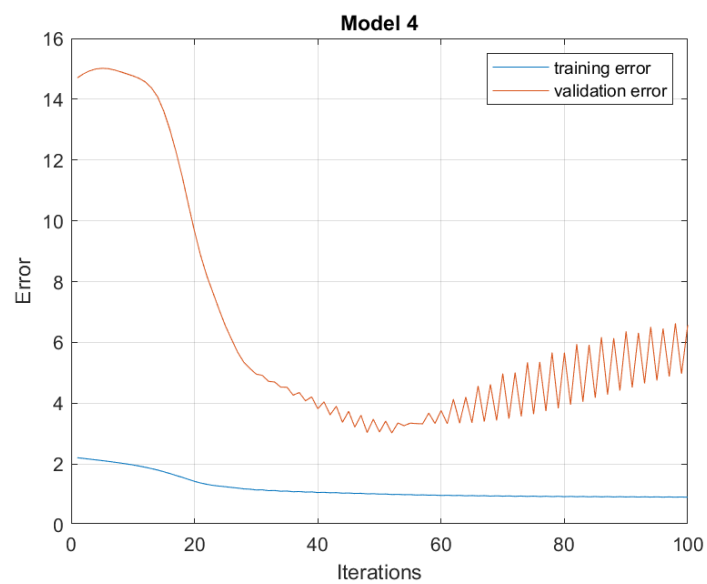
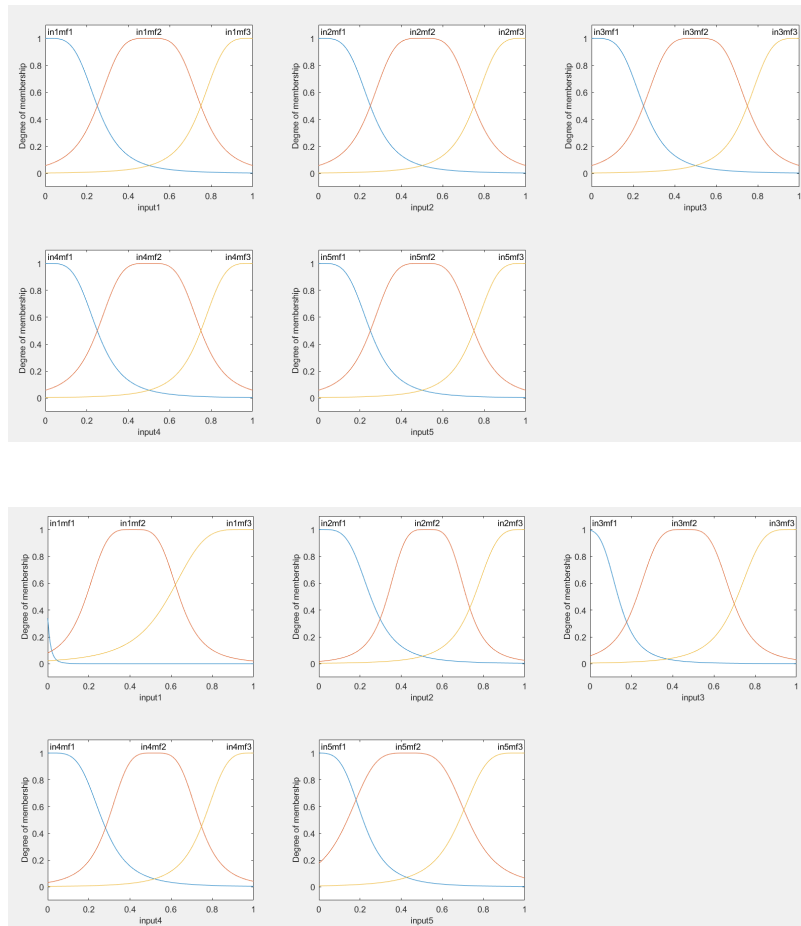


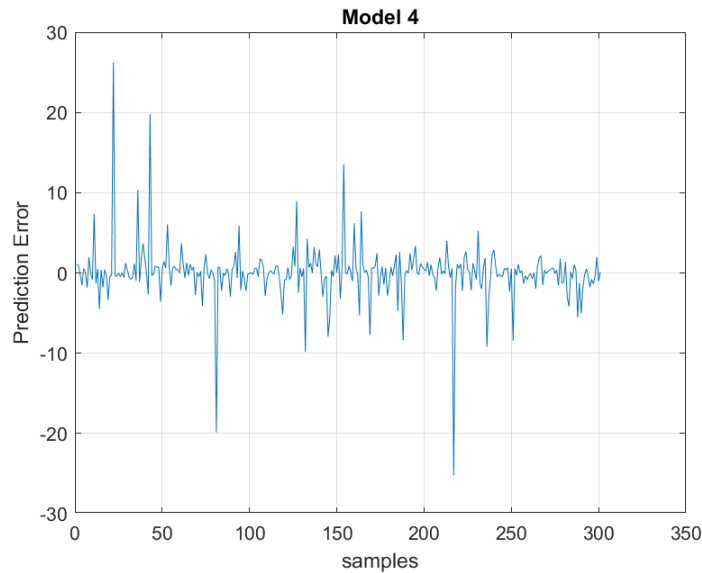
Model 3





Model 4





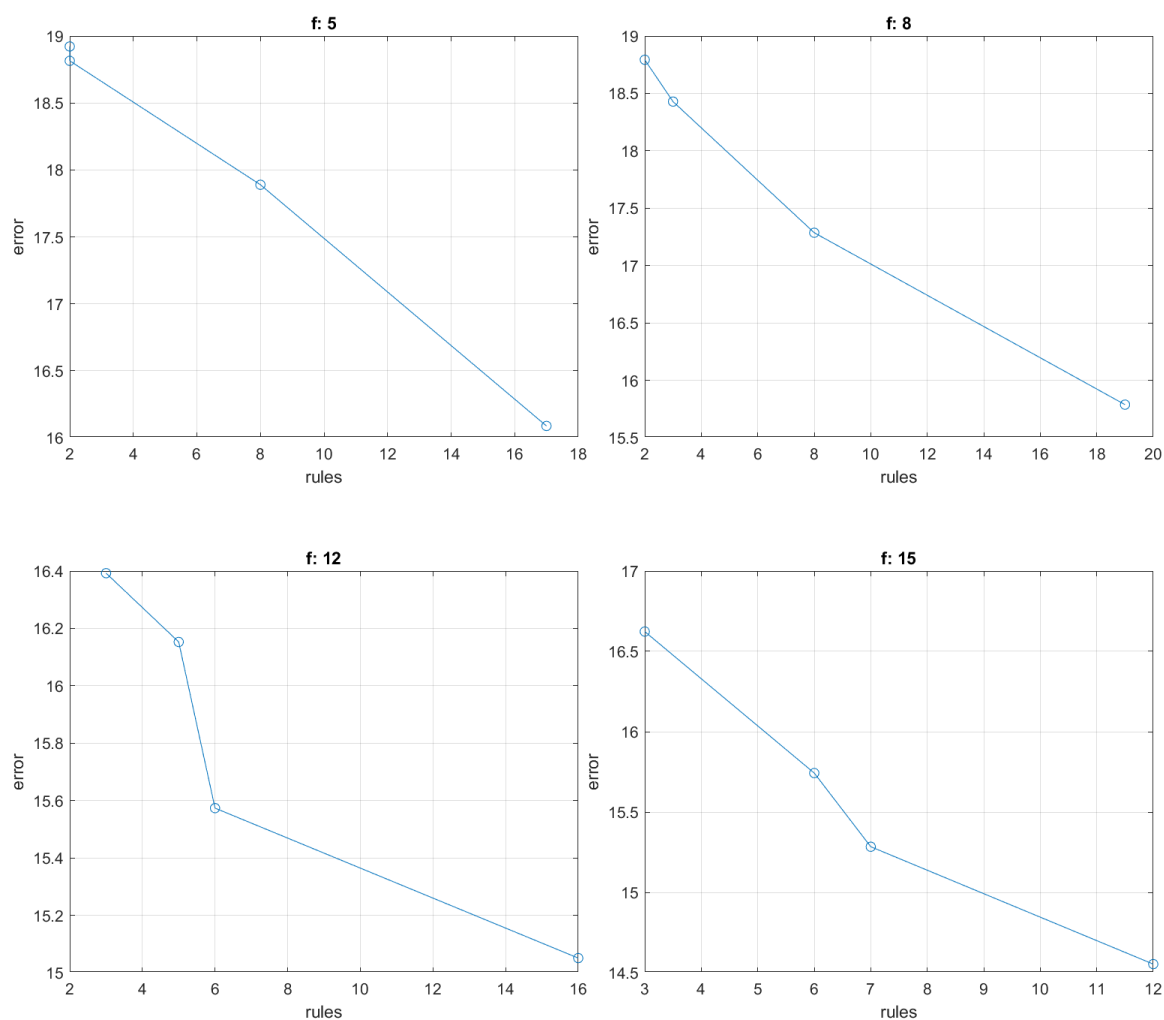
	Model 1		Model 2		Model 3		Model 4	
R2	0.72145	R2	0.85489	R2	0.86781	R2	0.72821	
RMSE	3.6147	RMSE	2.5315	RMSE	2.6332	RMSE	3.5971	
NMSE	0.27855	NMSE	0.14511	NMSE	0.13219	NMSE	0.27179	
NDEI	0.52777	NDEI	0.38093	NDEI	0.36358	NDEI	0.52134	

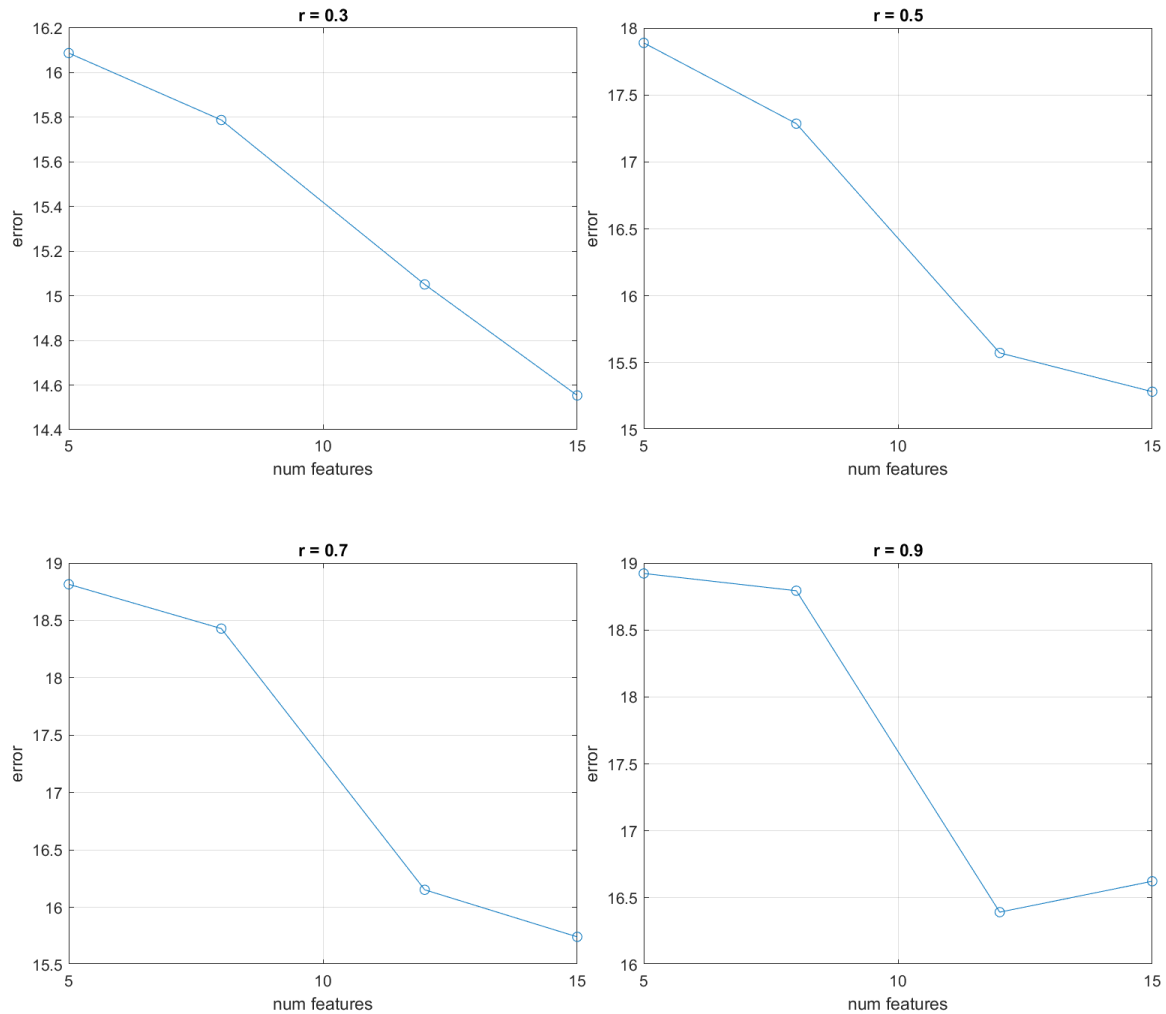
Συμπεραίνουμε από τις παραπάνω μετρικές ότι τα μοντέλο 3 με 2 συναρτήσεις συμμετοχής και πολυωνμική έξοδο παρουσιάζει τα καλύτερα αποτελέσματα για το συγκεκριμένο dataset. Τα μοντέλα 2 και 3 έχουν παρόμοια επίδοση.

2 Εφαρμογή σε **dataset** με υψηλή διαστασιμότητα

Χρησιμοποιούμε **grid search** για την επιλογή του βέλτιστου αριθμού **features** και την ακτίνα των **clusters** / αριθμό των κανόνων. Χωρίζουμε το αρχικό **dataset** σε **60-20-20** όπως πριν και χρησιμοποιώντας το **training set** υλοποιούμε **5-fold cross-validation**. Ελέγχουμε τα μοντέλα για αριθμό **features** = 5, 8, 12, 15 και ακτίνα = 0.3, 0.5, 0.7, 0.9.

Κριτήριο επιλογής είναι ο μέσος όρος του **RMSE** κάθε μοντέλου μετά από τα 5 fold.





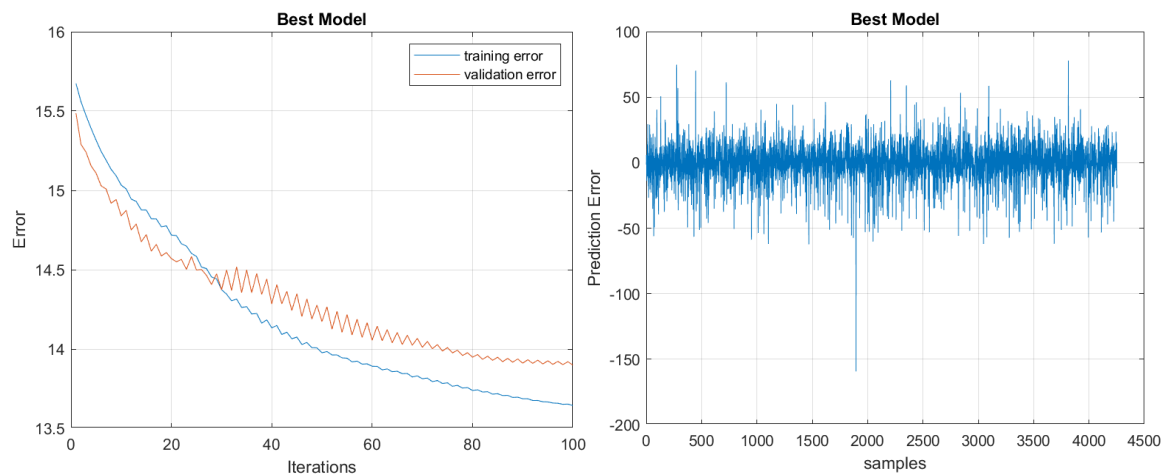
Από τα παραπάνω διαγράμματα βλέπουμε ότι όσο αυξάνεται ο αριθμός των features (με r σταθερό) το μέσο **RMSE** μειώνεται εκτός από $r = 0.9$

Επίσης, παρατηρούμε όσο αυξάνεται ο αριθμός των κανόνων (με αριθμό feat σταθερό) μειώνεται το σφάλμα.

Το καλύτερο από τα παραπάνω μοντέλα είναι για 15 features και ακτίνα 0.3

Εκπαιδεύουμε το τελικό μοντέλο με τις υπερπαραμέτρους που επιλέχθηκαν από την προηγούμενη διαδικασία.

Στο διαγραμμα της καμπύλης μάθησης φαίνεται να έχουμε μικρό overfitting.



Έχουμε παρόμοια στατιστικά για το RMSE και στο test set:

Best Model

R2	0.82861
RMSE	14.106
NMSE	0.17139
NDEI	0.41399