

Ασαφή Συστήματα

Υπολογιστική Νοημοσύνη

Εργασία 3-Επίλυση προβλήματος
παλινδρόμησης με χρήση μοντέλων TSK

Παναγιώτης Σαββίδης
8094
11.7.2021

Σκοπός της εργασίας αυτής είναι να διερευνηθεί η ικανότητα των μοντέλων TSK στη μοντελοποίηση πολυμεταβλητών, μη γραμμικών συναρτήσεων. Η εργασία αυτή έχει 2 κομμάτια. Στο πρώτο (*main_3_1.m*) γίνεται εκπαίδευση 4 TSK μοντέλων με την χρήση ενός μικρού μεγέθους dataset χωρίς κάποια προ επεξεργασία του. Στο δεύτερο (*main_3_2.m*) , το dataset είναι πολύ μεγαλύτερο και γι' αυτό τον λόγο χρειάζεται να χωρίσουμε το dataset για να μειώσουμε τον χρόνο εκτέλεσης. Για να επιτύχω αυτό τον διαχωρισμό, χρησιμοποίησα την συνάρτηση `split_scale.m` που υπάρχει στο `elearning`.

Μέρος 1^ο :

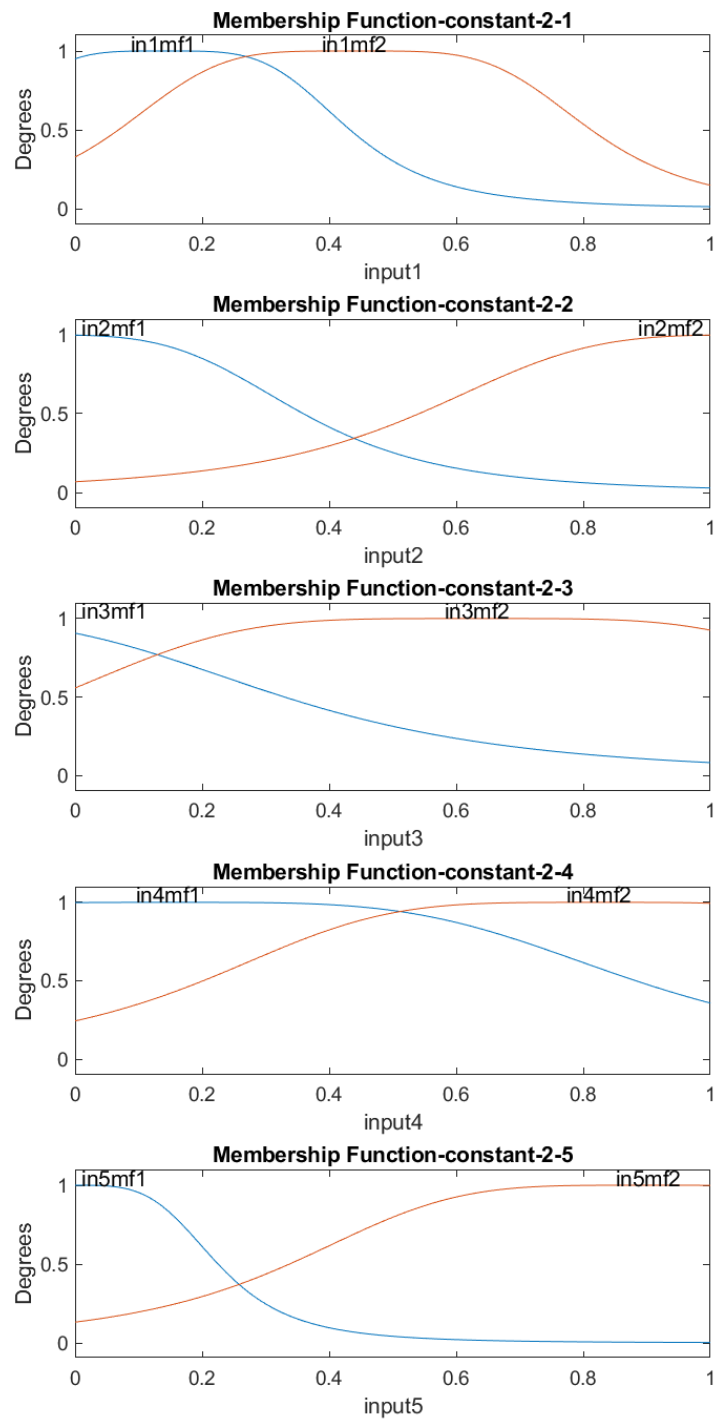
Το dataset που μας δόθηκε για το 1^ο κομμάτι της 3^{ης} εργασίας είναι το Airfoil Self-Noise Data Set και από αυτό, σύμφωνα με την εκφώνηση, εκπαιδεύτηκαν 4 TSK μοντέλα με τις παρακάτω προδιαγραφές. Σε όλα τα TSK μοντέλα χρησιμοποίησα 200 εποχές

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

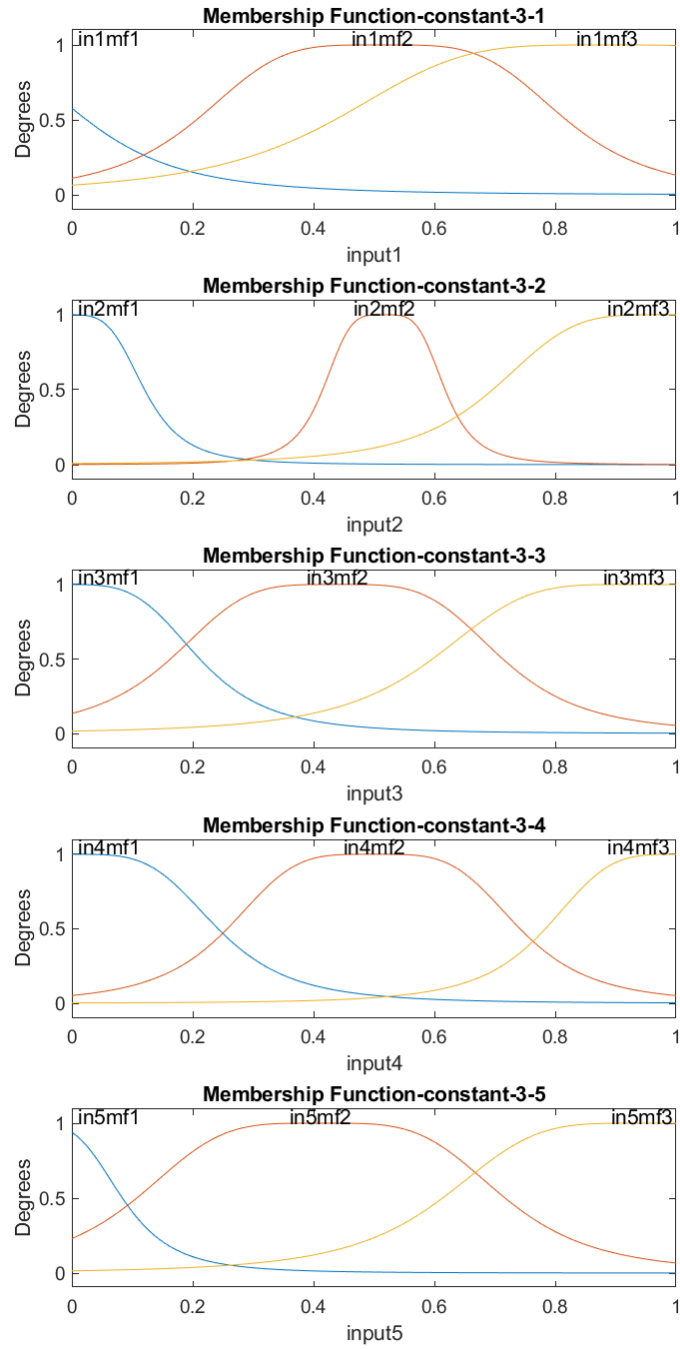
Τα αποτελέσματα της εκτέλεσης φαίνονται στα παρακάτω διαγράμματα:

Membership Functions:

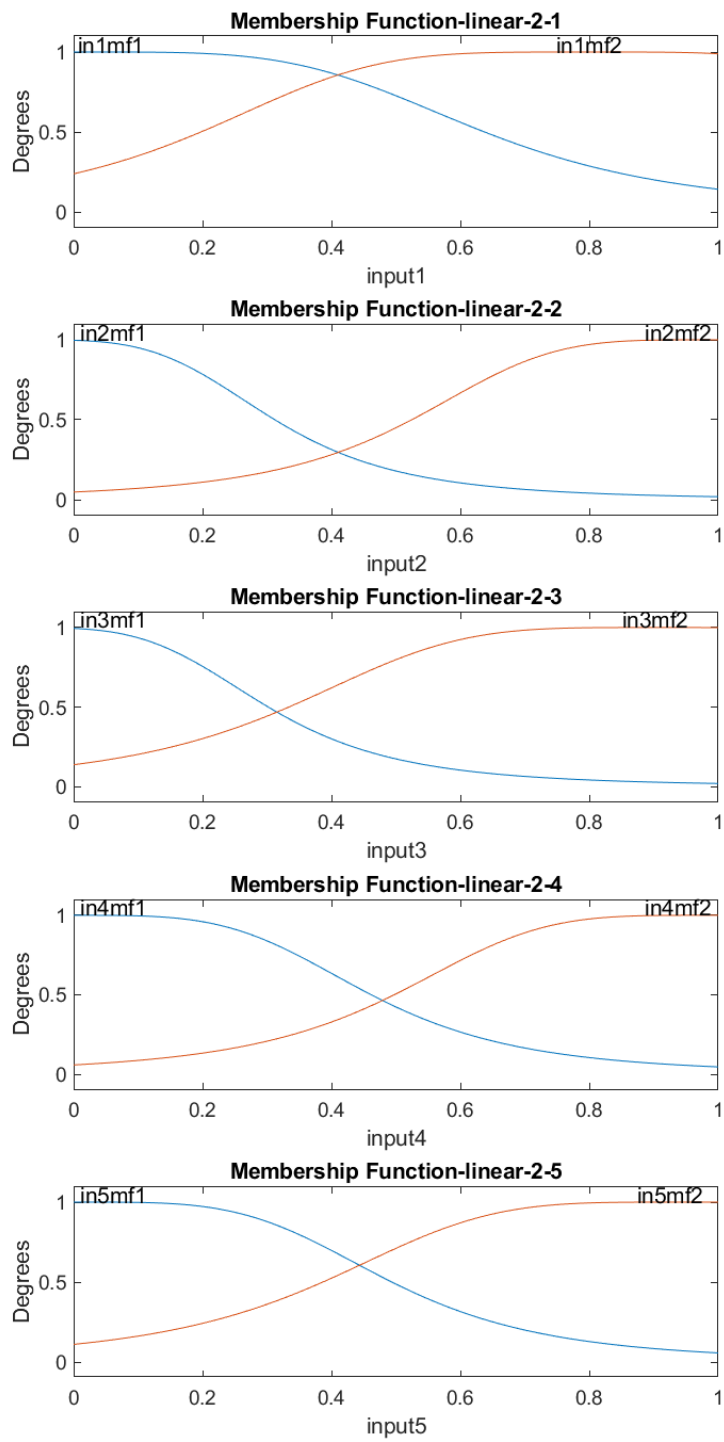
TSK model 1



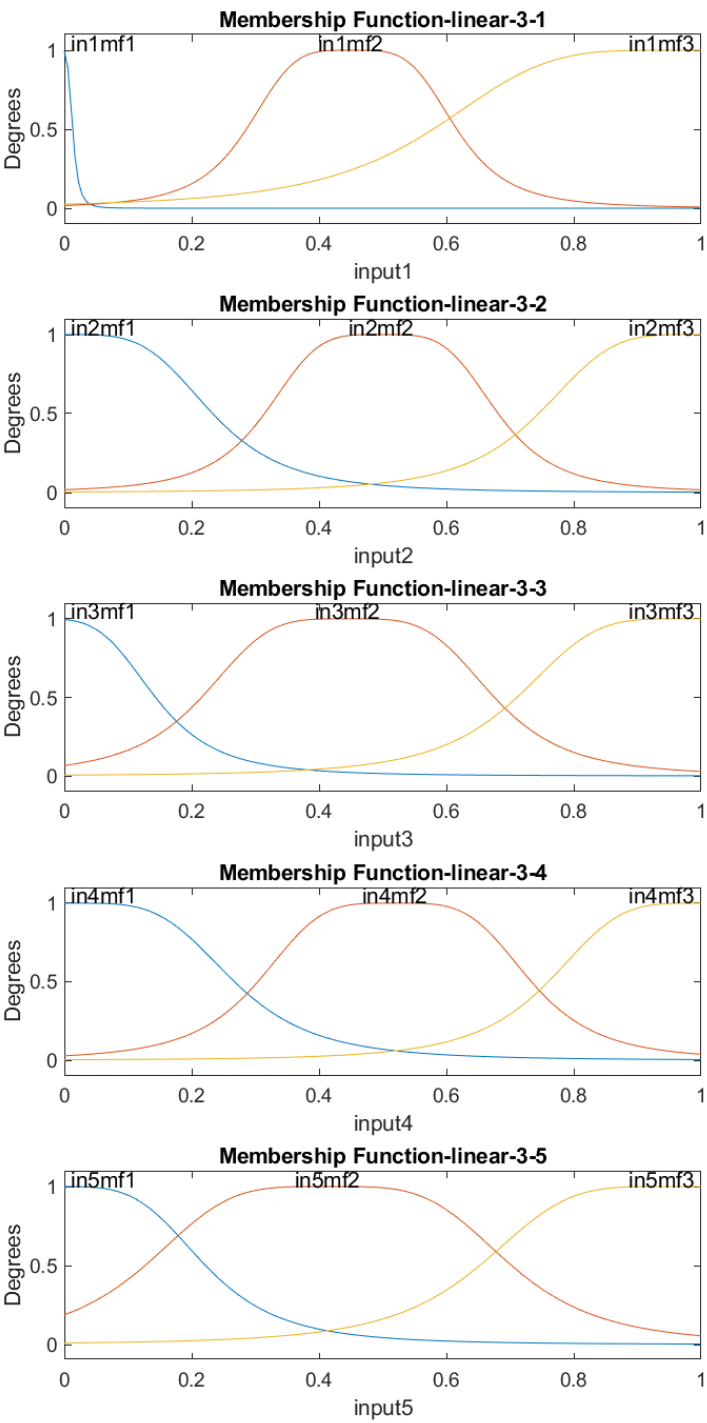
TSK model 2



TSK model 3

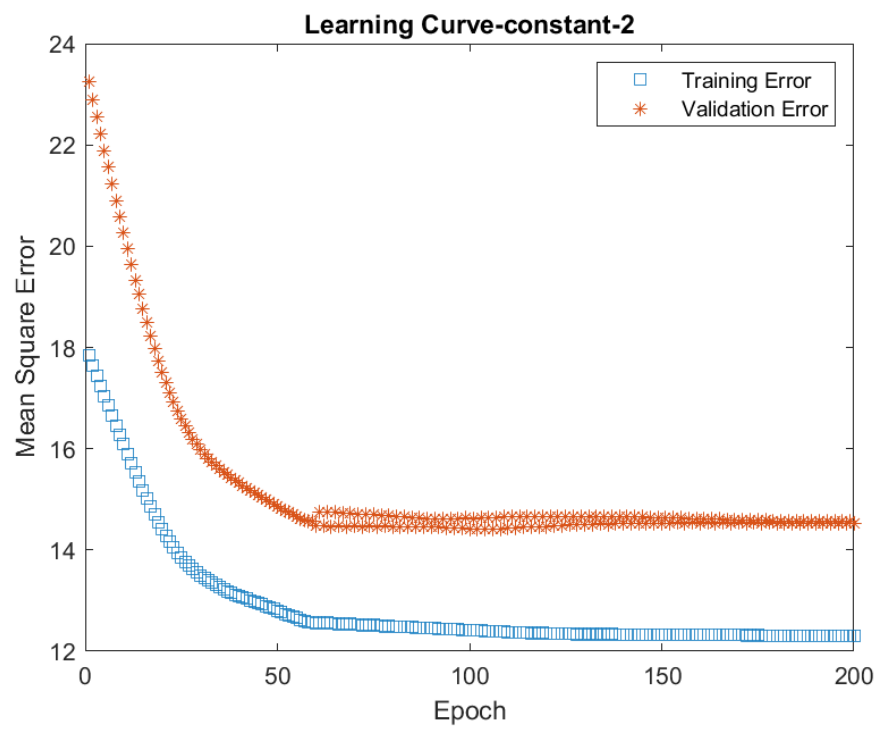


TSK model 4

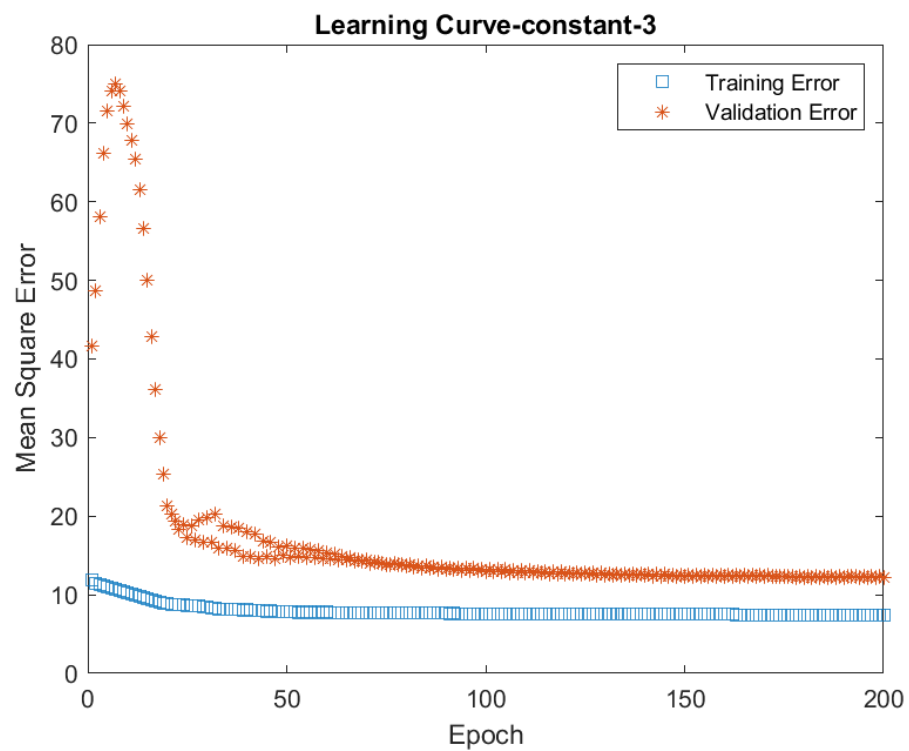


Learning Curves:

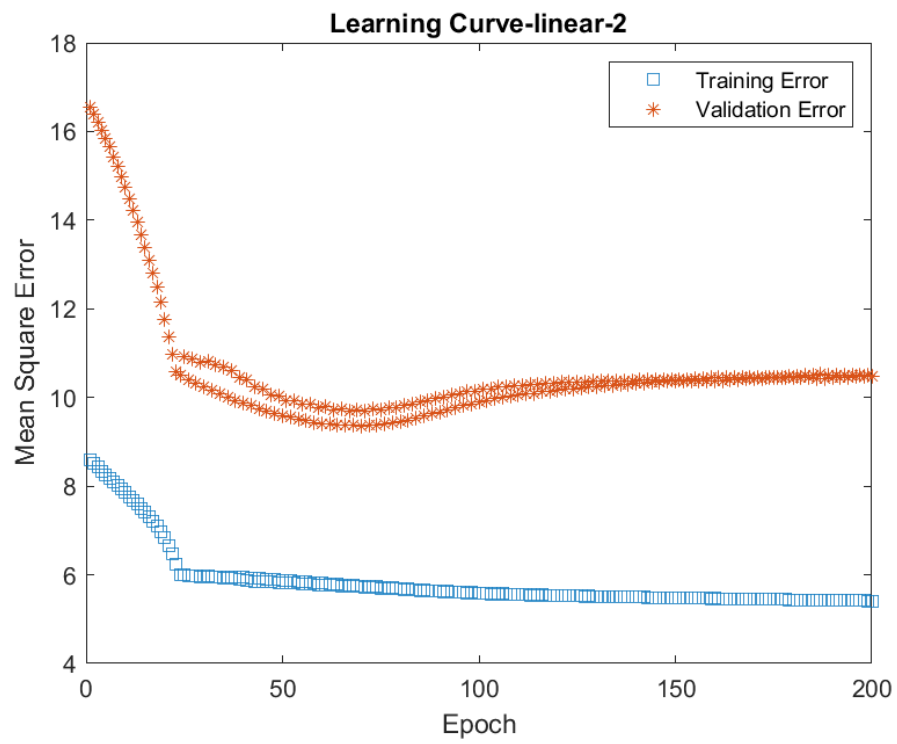
TSK model 1



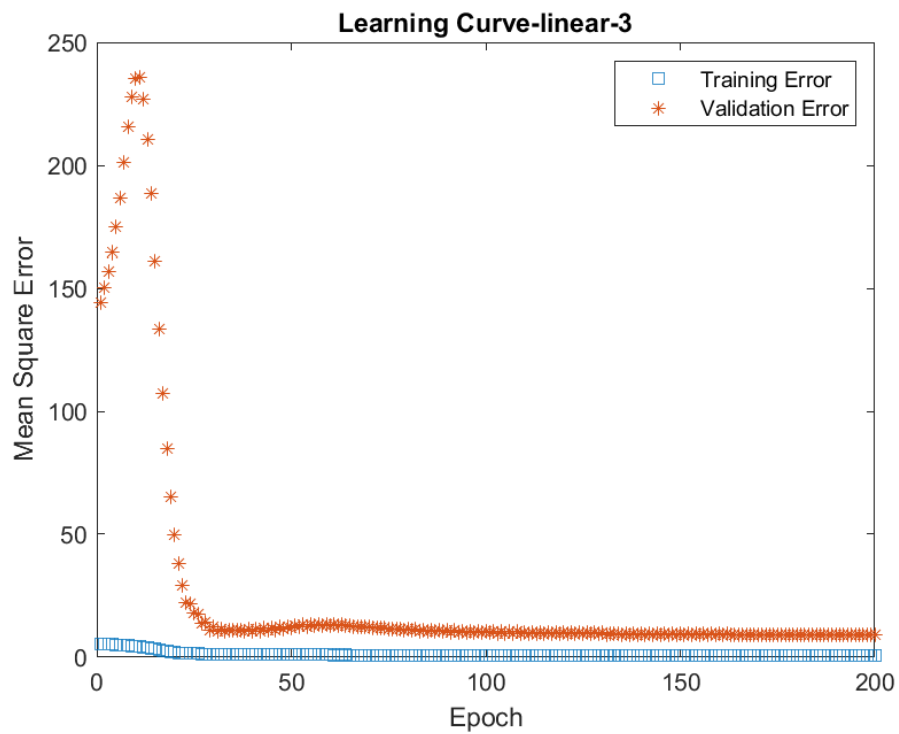
TSK model 2



TSK model 3

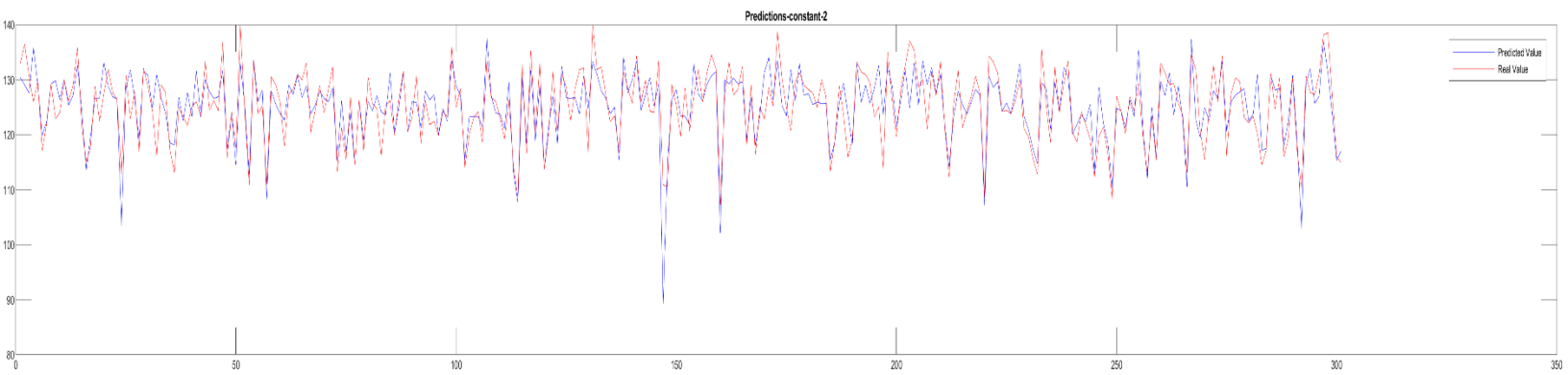


TSK model 4

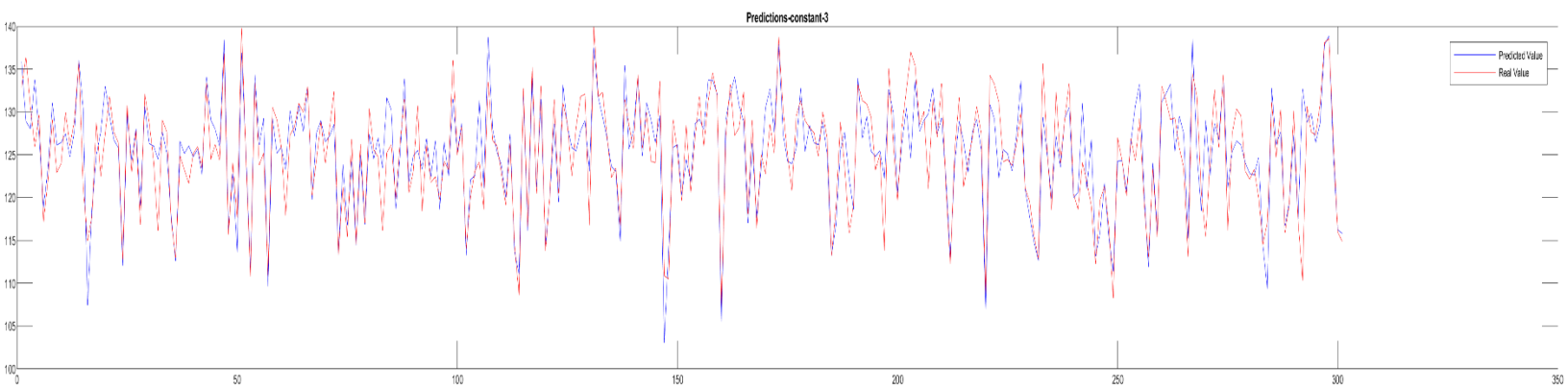


Prediction Errors:

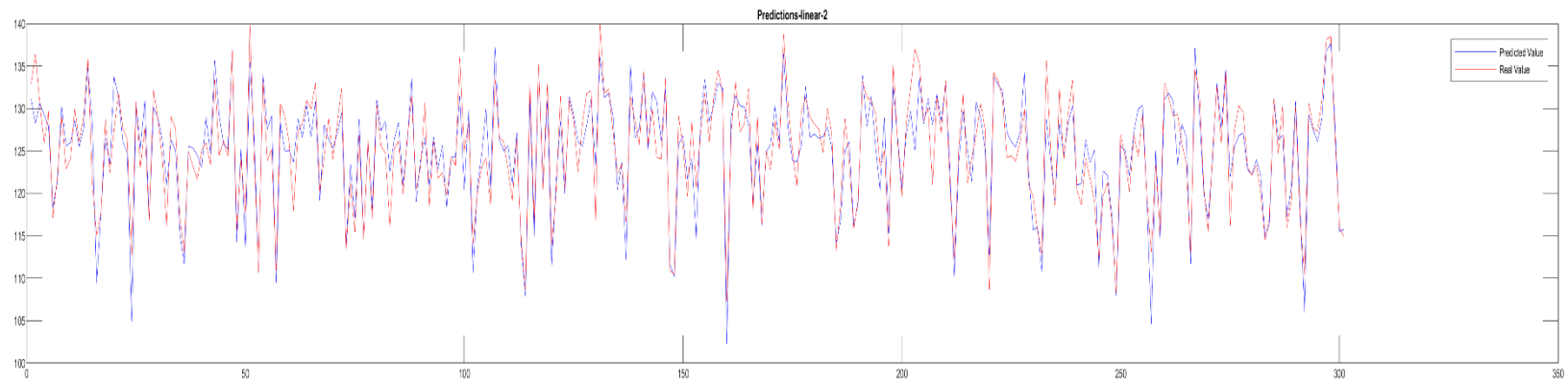
TSK model 1



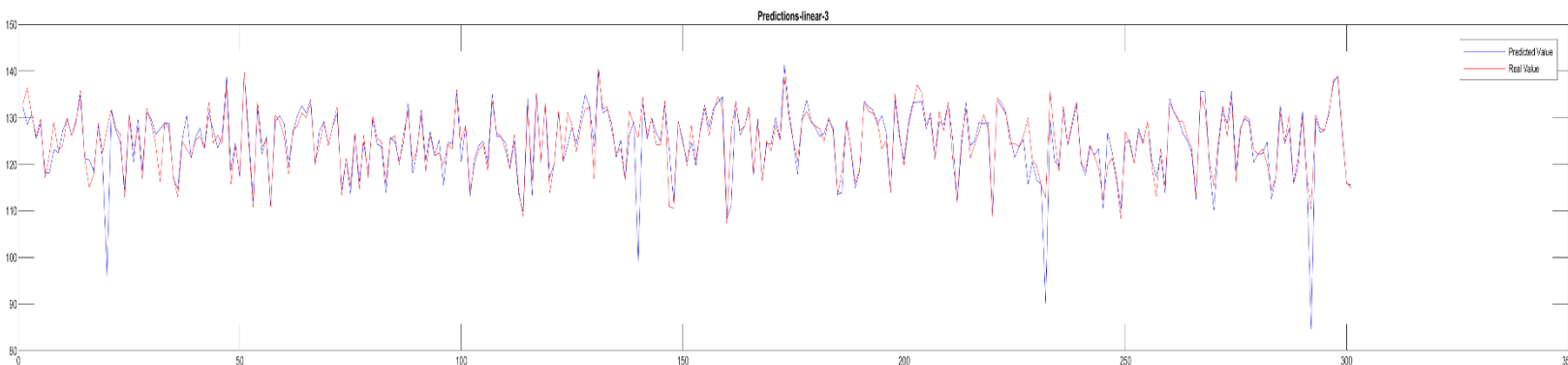
TSK model 2



TSK model 3



TSK model 4



Model no	MFs	Output	RMSE	NMSE	NDEI	R^2
Model 1	2	Singleton	3.8667	0.34266	0.58537	0.65734
Model 2	3	Singleton	3.2497	0.26958	0.51921	0.73042
Model 3	2	Polynomial	2.6932	0.16623	0.40771	0.83377
Model 4	3	Polynomial	3.9939	0.36557	0.60462	0.63443

Στις εικόνες των membership Functions παρατηρούμε ότι υπάρχει μεγάλη επικάλυψη ανάμεσα στα διαδοχικά ασαφή σύνολα, που δικαιολογείται από την συνάρτηση την οποία χρησιμοποιήσα.

Οι Learning curves έχουν την μορφή την οποία περιμέναμε να έχουν. Παρατηρείται μάλιστα, ένα φαινόμενο «ψαλιδισμού», που είναι φυσιολογικό και αναμενόμενο σε τέτοιες περιπτώσεις. Επίσης, οι τιμές του MSE, τόσο στο Validation όσο και στο Training Error, μειώνεται συνεχώς στα διαδοχικά μοντέλα και μάλιστα στο Model 4 το Training Error έχει σχεδόν μηδενιστεί, ενώ το Validation Error είναι σταθερά πολύ χαμηλό.

Ένα σημείο που πρέπει να σταθούμε είναι η αρχική αύξηση του σφάλματος στα μοντέλα 2 και 4 πριν σταθεροποιηθεί στις χαμηλές του τιμές. Αυτό το φαινόμενο επηρεάζεται από τον αριθμό των εισόδων καθώς και από την έξοδο του συστήματος. Έτσι στο μοντέλο 4 που έχουμε 3 εισόδους και πολυωνμική έξοδο, το σφάλμα στις πρώτες περίπου 25 εποχές αυξάνεται αρκετά πριν φτάσει στην σταθερότητα.

Οι προβλεπόμενες τιμές στα prediction error διαγράμματα, ακολουθούν ικανοποιητικά τις πραγματικές τιμές. Μόνο στο 1^ο μοντέλο μπορούμε εύκολα να αναγνωρίσουμε εύκολα τις διαφορές που υπάρχουν στις δυο καμπύλες.

Τέλος από τον πίνακα παρατηρούμε ότι το πιο αποδοτικό μοντέλο είναι το 3^ο καθώς η τιμή του R^2 είναι μεγαλύτερη και οι τιμές των σφαλμάτων μικρότερες από τα υπόλοιπα μοντέλα. Επίσης φαίνεται πως τόσο ο αριθμός των ασαφών συνόλων ανά είσοδο όσο και η έξοδος του συστήματος επηρεάζουν το μοντέλο. Συγκεκριμένα το μοντέλο 2 που έχει 3 MF φαίνεται βελτιωμένο σε σχέση με το μοντέλο 1, αλλά και το μοντέλο 3, που έχει πολυωνυμική έξοδο, σε σχέση με το μοντέλο 1. Όμως βλέπουμε πως ο συνδυασμός των 3 MF με πολυωνυμική έξοδο χειροτερεύουν το μοντέλο, αφού το μοντέλο 4 είναι οριακά χειρότερο από το 2 που έχουν ίδιο αριθμό MF και αρκετά χειρότερο από το 3 που έχουν την ίδια μορφή εξόδου.

Μέρος 2^ο :

Το dataset που έπρεπε να χρησιμοποιήσω στο μέρος 2 της 3^{ης} εργασίας είναι το Superconductivity dataset, το οποίο έχει πολύ περισσότερα δείγματα και με πολλά περισσότερα χαρακτηριστικά στο καθένα, από το αντίστοιχο dataset του πρώτου μέρους. Όπως μπορεί να αντιληφθεί κανείς, το να δουλέψουμε όπως στο 1^ο μέρος της εργασίας είναι αδύνατο, καθώς οι κανόνες που θα προκύψουν θα ήταν πάρα πολλοί κάνοντας τον χρόνο εκτέλεσης απαγορευτικά μεγάλο.

Έτσι, για να μειώσω τον χρόνο εκτέλεσης, έτρεξα τον αλγόριθμο relief ο οποίος επιλέγει την βαρύτητα κάθε χαρακτηριστικού και στην συνέχεια επιλέγω τον αριθμό των χαρακτηριστικών που θα επιλεγούν και την ακτίνα των διαφορετικών cluster. Όπως φαίνεται και στην συνάρτηση main_3_2.m οι προδιαγραφές που επέλεξα είναι οι εξής:

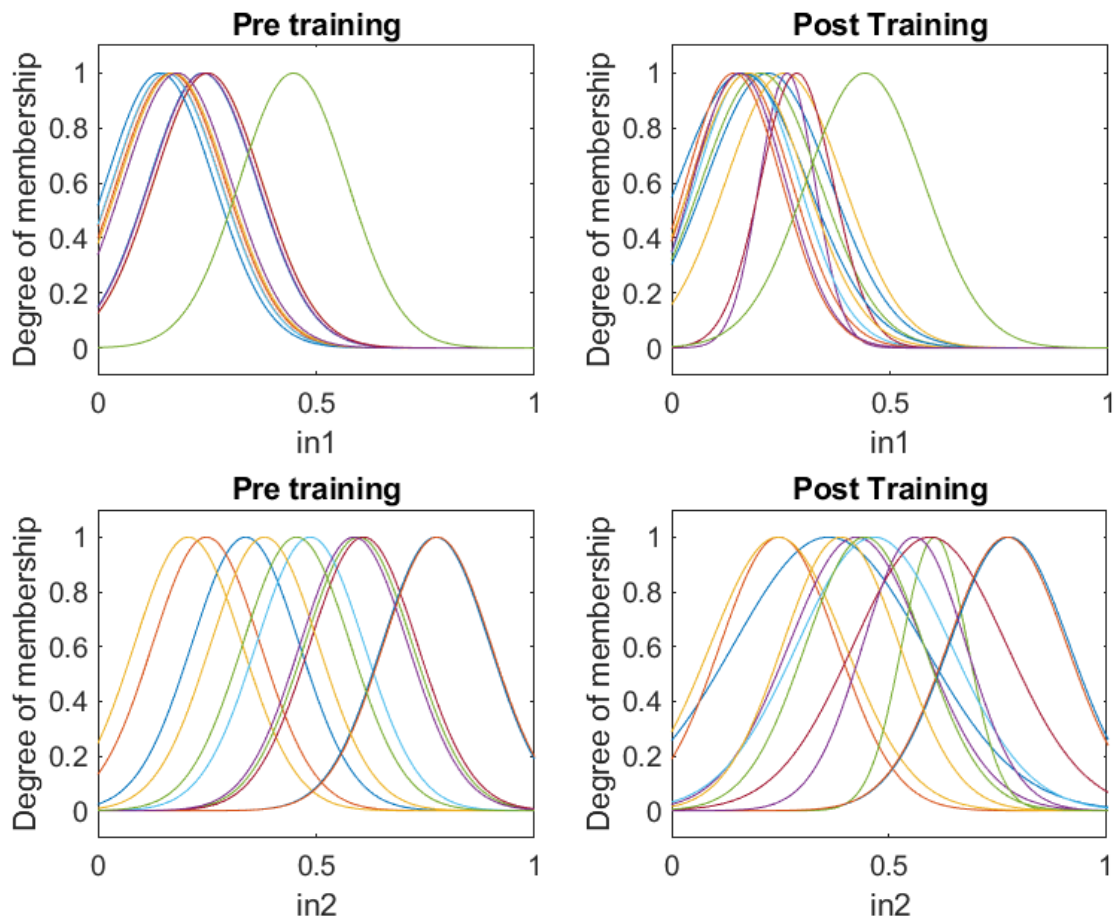
```
%Random selected characteristics
characteristics = [1 5 10 15];
%Rads
clust_rad = [0.55 0.45 0.35 0.25];
```

Οι αριθμοί αυτοί προέκυψαν από δοκιμές και είναι αυθαίρετοι αλλά παράλληλα και συγκεκριμένοι για να ελέγξουμε τις «δυνατότητες» του μοντέλου μας. Η επιλογή του 1 χαρακτηριστικού δεν ενδείκνυται και αυτό ακριβώς ελέγχω σε αυτή την περίπτωση, ενώ τιμές μεγαλύτερες του 17-18 καθυστερούσαν πολύ τον χρόνο εκτέλεσης. Η τιμή 0.55 της ακτίνας όπως θα φανεί και στην συνέχεια είναι οριακή καθώς για μεγαλύτερες τιμές, οι κανόνες που προκύπτουν είναι πάρα πολύ λίγοι.

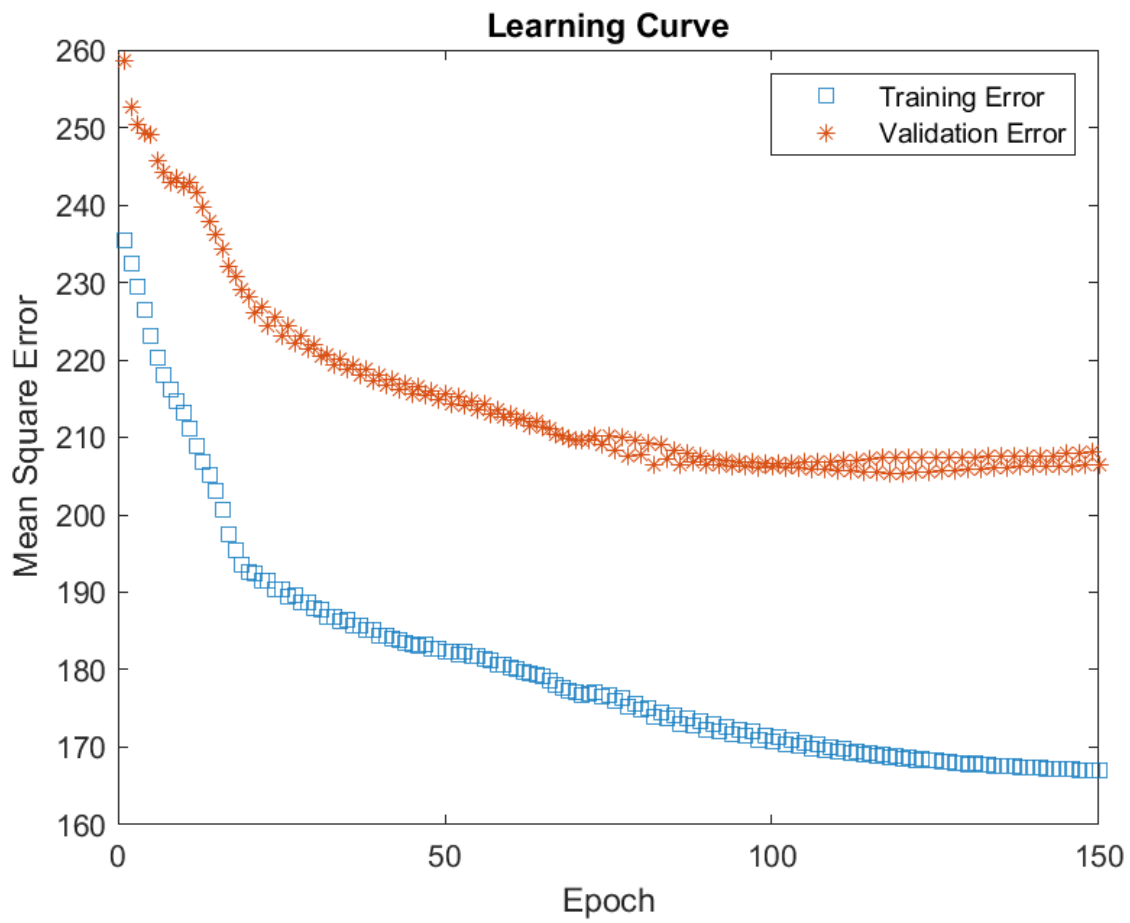
Αφού τελείωσα την προ επεξεργασία των δεδομένων, έπρεπε να εφαρμόσω 5-fold cross validation για 50 εποχές. Ο σκοπός μου είναι να βρω ποιος συνδυασμός από αριθμό χαρακτηριστικών και ακτίνας οδηγεί στο μικρότερο σφάλμα και με αυτό τον συνδυασμό τιμών

να εκπαιδεύσω το τελικό μου μοντέλο. Και ο «μαγικός» συνδυασμός είναι μας δίνει τα παρακάτω διαγράμματα

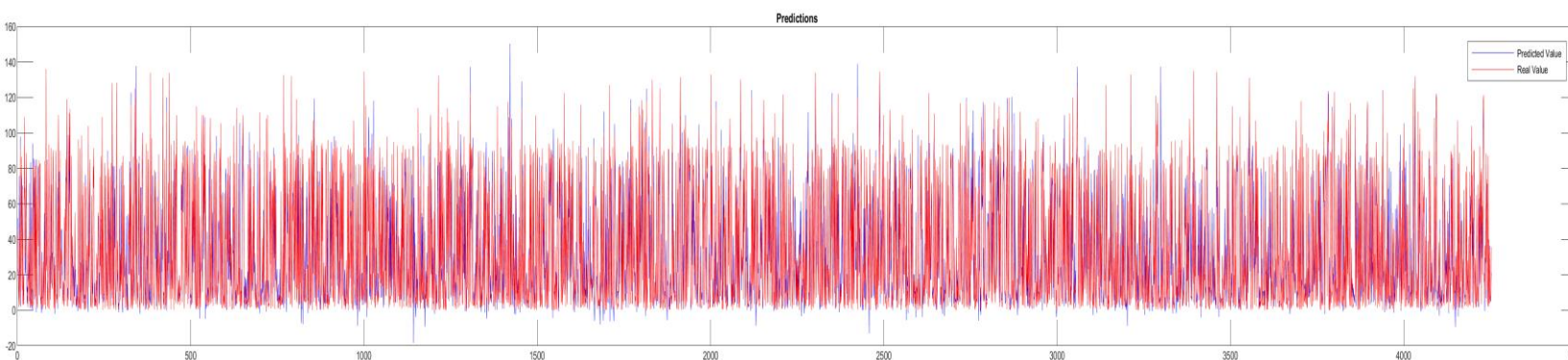
Membership functions:



Learning Curves:



Prediction Errors:



	RMSE	NMSE	NDEI	R^2
Μοντέλο	13.426	0.15377	0.39215	0.84623

Όπως και στο πρώτο μέρος της εργασίας, υπάρχει μεγάλη επικάλυψη στις membership functions με ορατή μια πολύ μικρή βελτίωση. Η διαφορά με το πρώτο μέρος είναι ότι από τις learning curves φαίνεται ότι υπάρχει μεγάλο σφάλμα και μάλιστα μετά τις 150 εποχές φαίνεται να σταθεροποιείται σε αυτή την μεγάλη τιμή σφάλματος.

Αυτό μολονότι δεν είναι ιδανικό, είναι δυστυχώς αναπόφευκτο, καθώς το dataset ήταν εξ αρχής αρκετά προβληματικό και αν μπορούσαμε να συγκρίνουμε τις τιμές που θα προκύπταν χωρίς την διαδικασία που ακολουθήσαμε, θα βλέπαμε ότι υπήρξε τεράστια βελτίωση. Το αρχικό dataset περιέχει πολύ θόρυβο και αυτό φάνηκε και στο grid search, που μετά από κάποιες τιμές το σφάλμα αυξήθηκε πάρα πολύ.

Από το διάγραμμα του prediction error βλέπουμε ότι είναι διακριτές κάποιες μπλε γραμμές παρόλο ανάμεσα στις πολύ πυκνές κόκκινες, που σημαίνει ότι έγιναν κάποια σημαντικά λάθη στα predictions.

Μετά το grid searching επιλέξαμε τις τιμές εκείνες που ελαχιστοποιούν το error, αλλά και πάλι οι τιμές δεν είναι οι αναμενόμενες, αφού το μέσο σφάλμα παραμένει αρκετά υψηλό. Αυτές οι τιμές όμως ίσως να οφείλονται και στο πόσα λίγα είναι τα χαρακτηριστικά που επέλεξα, σε σχέση με τα αρχικά χαρακτηριστικά που είχε το dataset.

Τέλος, εφόσον μιλάμε για μέγιστο 15 χαρακτηριστικά, οι κανόνες που θα μπορούσαν να προκύψουν αν δεν κάναμε όλες τις παραπάνω διαδικασίες, θα ήταν 2^{15} και 3^{15} αντίστοιχα, που όπως εύκολα αντιλαμβάνεται κάποιος ο χρόνος εκτέλεσης θα ήταν πάρα πολύ μεγαλύτερος.