



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ
ΠΛΗΡΟΦΟΡΙΚΗΣ

ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ

Διδάσκοντες: Σ. Λυκοθανάσης, Δ. Κουτσομητρόπουλος

Ακαδημαϊκό Έτος 2019-2020

Εργαστηριακή Άσκηση
Μέρος Α΄

ΠΑΝΑΓΙΩΤΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ
ΑΘΑΝΑΣΟΠΟΥΛΟΣ
ΑΜ:1058112

A1)

α) Επειδή κάποιοι χρήστες έχουν βαθμολογήσει συστηματικά με χαμηλούς βαθμούς και άλλοι χρήστες έχουν βαθμολογήσει συστηματικά με υψηλούς βαθμούς κρίνεται απαραίτητο να εφαρμοσθεί η μέθοδος του κεντραρίσματος ώστε να μετριαστεί ο βαθμός πόλωσης κάποιων αξιολογήσεων των χρηστών. Το νέο διάστημα στο οποίο κινούνται οι τιμές των αξιολογήσεων των χρηστών είναι το $[-3.724, 3.508]$.

β)

i) Αν συμπληρώσουμε τις ελλιπείς τιμές με μια τυχαία επιλεγμένη τιμή στο κατάλληλο εύρος, το νέο διάστημα στο οποίο κινούνται οι τιμές παραμένει το ίδιο , δηλαδή $[-3.724, 3.508]$.

ii) Αν συμπληρώσουμε τις ελλιπείς τιμές με μηδέν το νέο διάστημα στο οποίο κινούνται οι τιμές παραμένει το ίδιο , δηλαδή $[-3.724, 3.508]$.

iii) Αν συμπληρώσουμε τις ελλυπείς τιμές με τον μέσο όρο του διανύσματος αξιολόγησης το νέο διάστημα στο οποίο κινούνται οι τιμές παραμένει το ίδιο , δηλαδή [-3.724,3.508].

A2)

α)

Για κάθε πρότυπο παρέχεται ένα σύνολο στατιστικών δεικτών που αποτελεί την τετραγωνική ρίζα της μέσης τιμής του τετραγώνου των σφαλμάτων(RMSE). Αποτελεί μία από τις πιο συχνά χρησιμοποιούμενες μετρικές, καθώς λαμβάνει υπόψιν της τα μεγάλα σφάλματα. Η ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) μετράει την απόκλιση του σφάλματος υψωμένο στο τετράγωνο για όλο το σύνολο της πρόβλεψης. Εξαιτίας του τετραγώνου τα σφάλματα αντίθετου πρόσημου δεν ακυρώνονται μεταξύ τους. Ένα άλλο χαρακτηριστικό του RMSE είναι ότι «τιμωρεί» τα μεγάλα σφάλματα περισσότερο από τα μικρά. Σημειώνεται ωστόσο, ότι δεν είναι εύκολα ερμηνεύσιμο, αλλά θεωρείται καλό μέτρο όσον αναφορά τα σφάλματα.

Το μέσο τετραγωνικό σφάλμα (MAE) μετράει τη μέση απόλυτη απόκλιση των προβλεπόμενων τιμών από τις πραγματικές για όλο το σύνολο της πρόβλεψης. Λόγω της απόλυτης τιμής του οι θετικές τιμές δεν αναιρούνται από τις αρνητικές. Επίσης το σφάλμα είναι στην ίδια κλίμακα με τα δεδομένα.

Οι μετρικές RMSE και MAE είναι παρόμοιες, με τη πρώτη να δίνει μεγαλύτερο βάρος στη μεγαλύτερη απόκλιση. Ως προς τις τιμές που λαμβάνουν οι μετρικές, όσο μικρότερες είναι οι τιμές τους, τόσο καλύτερη είναι η αξιοπιστία του αλγορίθμου.

Μετά την επιλογή του κατάλληλου μοντέλου και των παραμέτρων αυτού, ακολουθεί η αξιολόγηση της ακρίβειας της πρόβλεψης που προκύπτει. Αυτό επιτυγχάνεται με το διαχωρισμό των δεδομένων σε σετ εκπαίδευσης (training set) και σετ δοκιμών (test set). Το training set χρησιμοποιείται για το την εκπαίδευση του μοντέλου και τον υπολογισμό των προβλεπόμενων τιμών και το test set για την αξιολόγηση της ακρίβειας.

Για να προσδιοριστεί η ακρίβεια μιας πρόβλεψης χρησιμοποιούνται διάφορα μέτρα αξιολόγησης. Γενικά το σφάλμα πρόβλεψης θεωρείται η διαφορά μεταξύ της πραγματικής και της προβλεπόμενης τιμής και εκφράζεται ως: $et = Y_t - \hat{Y}_t$ όπου Y_t η πραγματική τιμή και \hat{Y}_t η προβλεπόμενη για τη χρονική στιγμή t .

Με άλλα λόγια, στο πείραμά μας θα εισάγεται ένας χρήστης στην είσοδο του νευρωνικού δικτύου. Με την μετρική RMSE η είσοδος ,δηλαδή η προβλεπόμενη αξιολόγηση του χρήστη , που παίρνουμε αφαιρείται από την πραγματική τιμή που είχε στο data set. Αυτό γίνεται για κάθε χρήστη και υπολογίζεται η ρίζα της μέσης τιμής. Με την μετρική MAE η είσοδος , δηλαδή η προβλεπόμενη αξιολόγηση του χρήστη , που παίρνουμε αφαιρείται από την πραγματική τιμή που είχε στο data set και υπολογίζεται η απόλυτη τιμή. Αυτό γίνεται για κάθε χρήστη και υπολογίζεται η μέση τιμή όλων των χρηστών

β) Θα χρειαστούμε 946 εισόδους, όσοι δηλαδή είναι και οι χρήστες.

Η κωδικοποίηση one hot encoding χρησιμοποιείται συχνά για την ένδειξη της κατάστασης της μηχανής. Όταν χρησιμοποιείται binary encoding ή gray encoding απαιτείται ένας

αποκωδικοποιητής για τον προσδιορισμό της κατάστασης. Μια one-hot κατάσταση δεν χρειάζεται αποκωδικοποιητή αφού η κατάσταση μηχανής βρίσκεται στην νιοστή κατάσταση αν και μόνο αν το νιοστό bit είναι 1.

γ) Θα χρειαστούμε 1 νευρώνα που στην έξοδο θα βγάξει κάθε φορά την αξιολόγηση ενός χρήστη για μια ταινία.

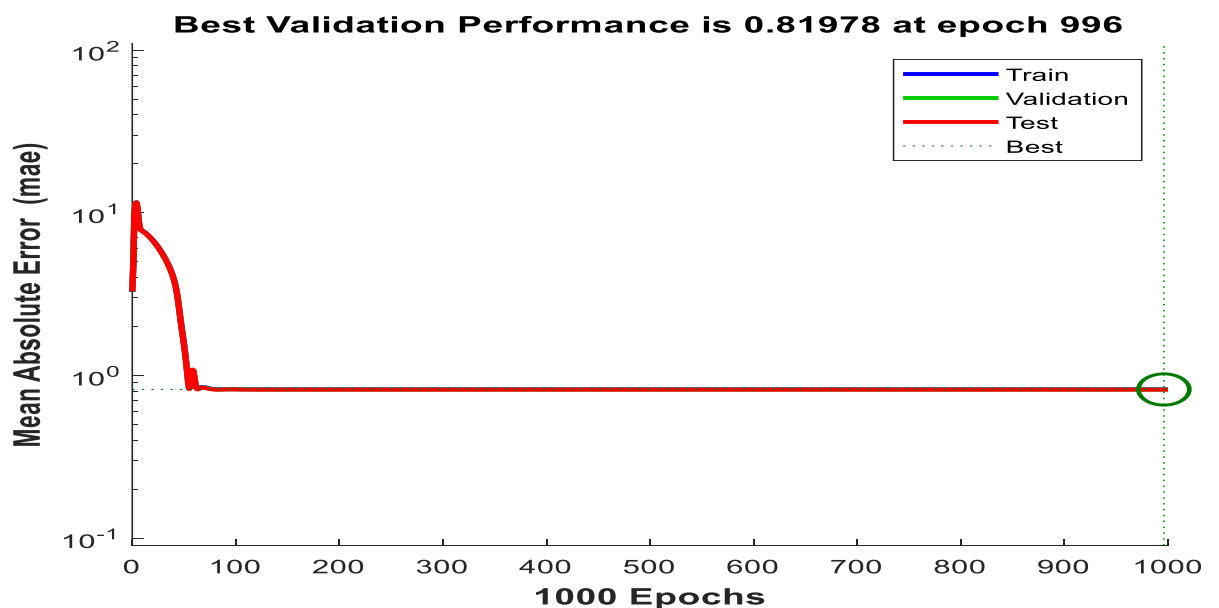
δ) Για το κρυφό επίπεδο θα χρησιμοποιήσουμε τη σιγμοειδή συνάρτηση. Το προτέρημα της είναι ότι επιτρέπει τη μείωση ακραίων ή άτυπων τιμών σε έγκυρα δεδομένα χωρίς να τα εξαλείψει. Μετατρέπει ανεξάρτητες μεταβλητές σχεδόν άπειρου εύρους σε απλές πιθανότητες μεταξύ 0 και 1. Το μεγαλύτερο μέρος της εξόδους της είναι πολύ κοντά στο 0 ή 1.

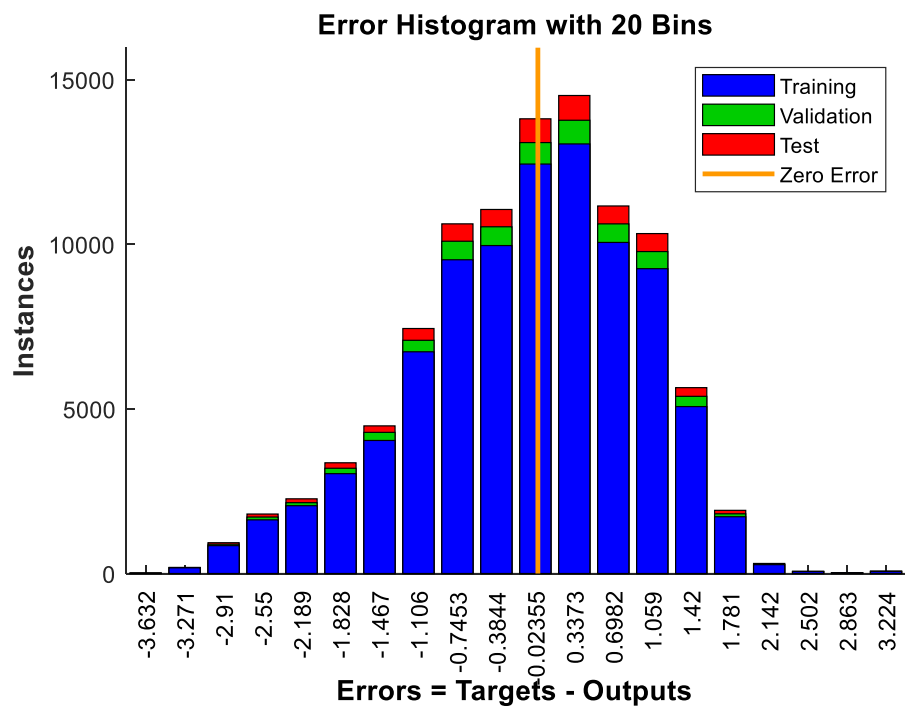
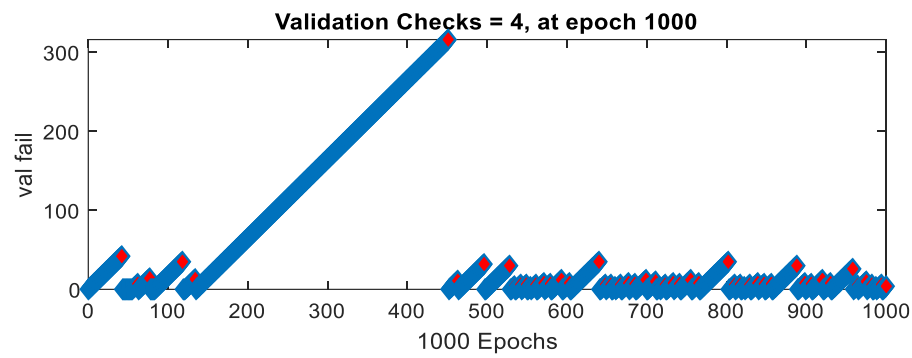
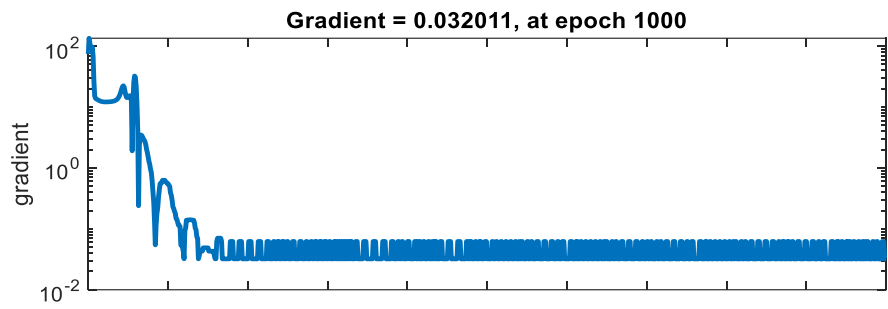
ε) Για το επίπεδο εξόδου θα χρησιμοποιήσουμε την γραμμική συνάρτηση ενεργοποίησης μιας και το dataset μας έχει τιμές με μεγάλο εύρος. Η γραμμική συνάρτηση ενεργοποίησης είναι ουσιαστικά η συνάρτηση με την οποία το σήμα δεν αλλάζει.

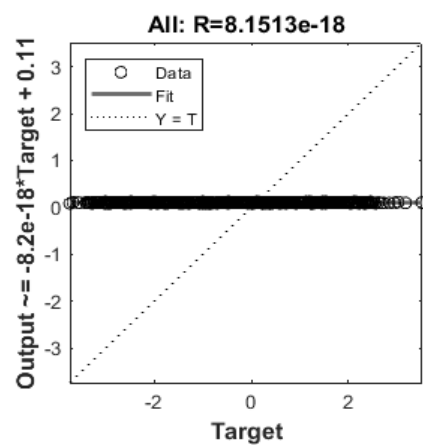
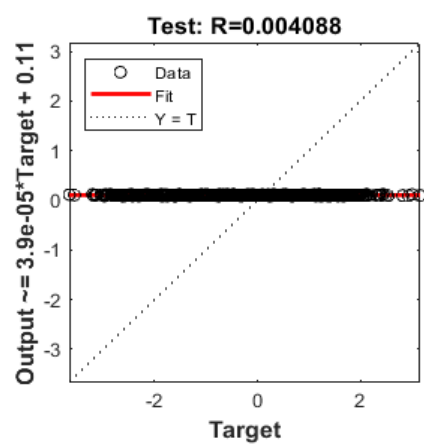
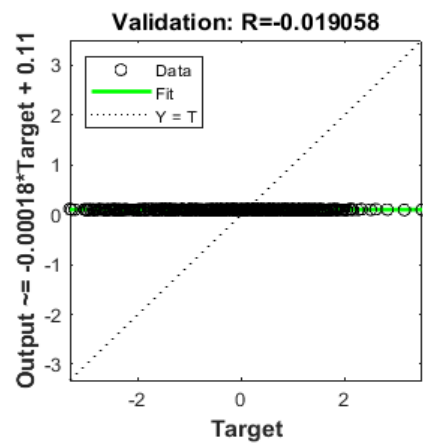
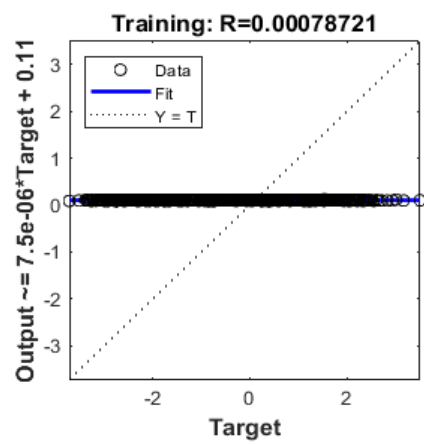
στ) Για $H=10$ νευρώνες στο κρυφό επίπεδο

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $rmse = 1.0362$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $mae = 0.8262$



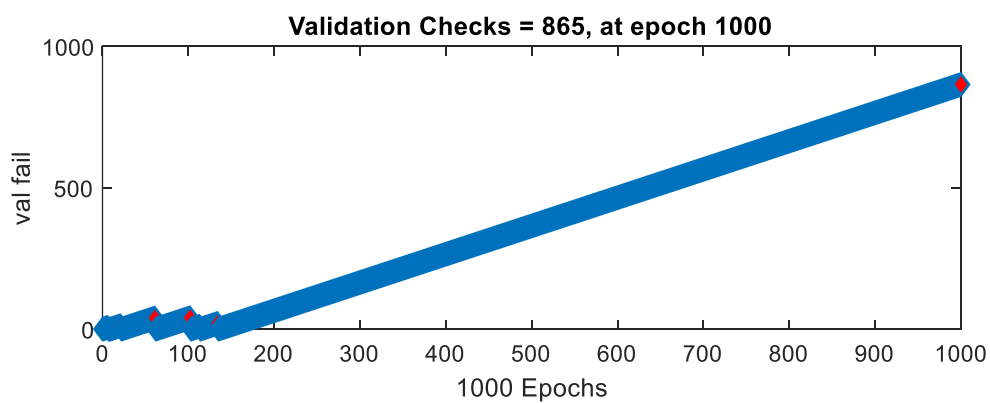
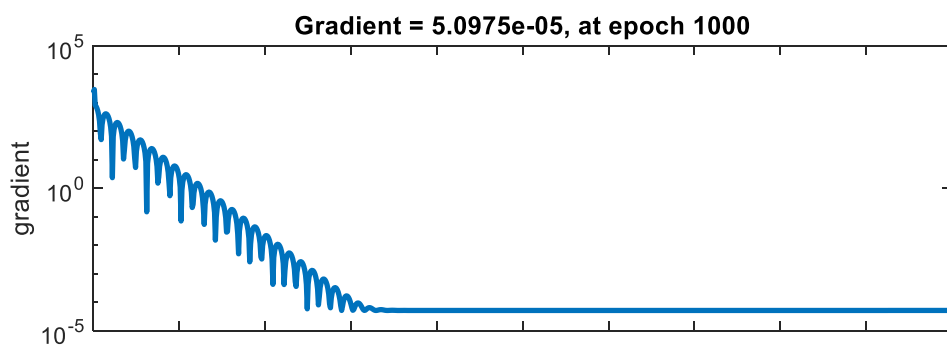
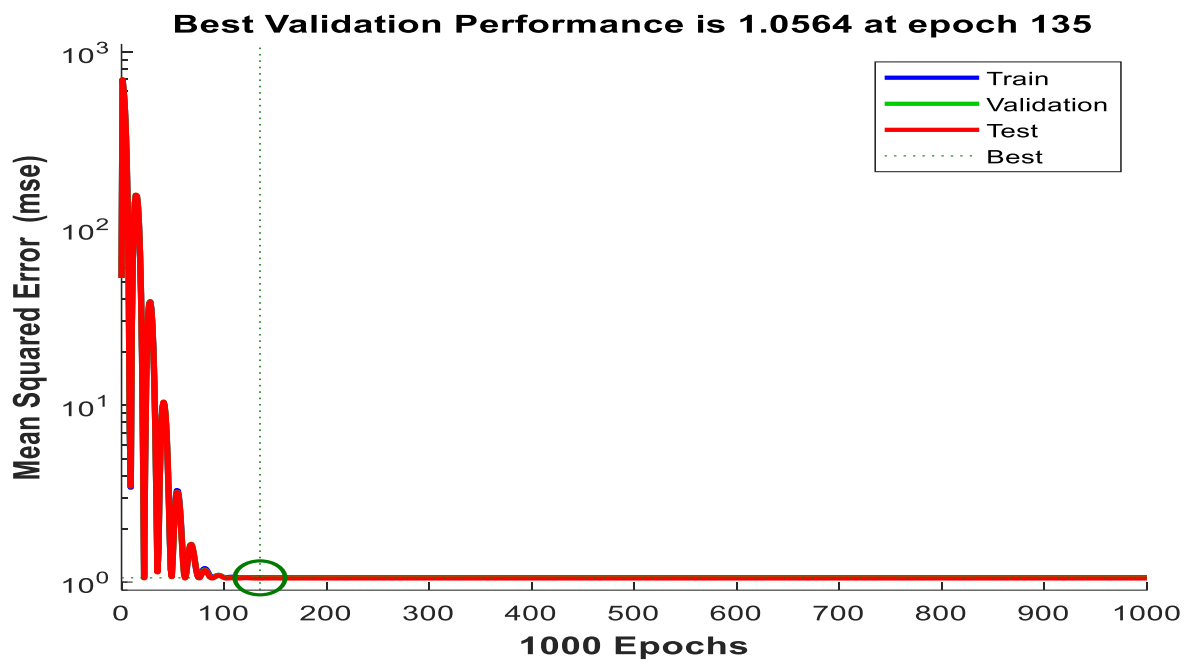


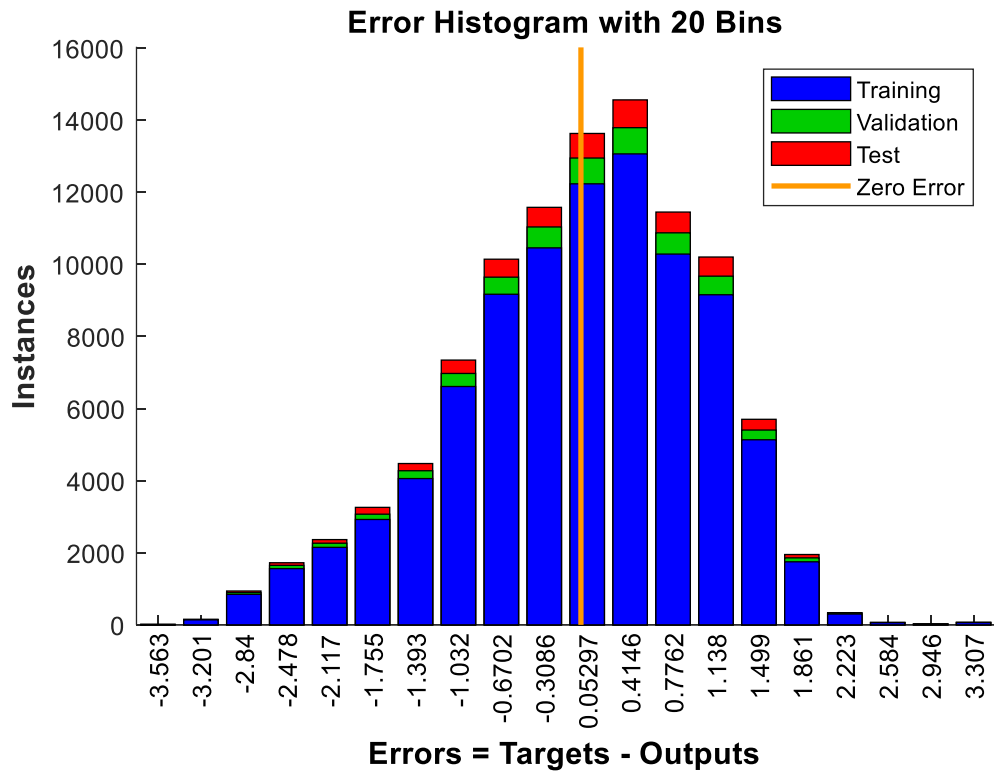


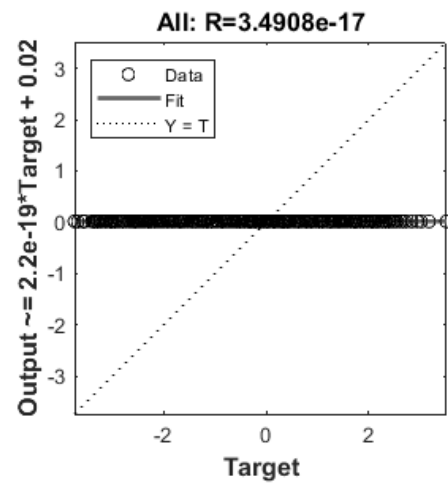
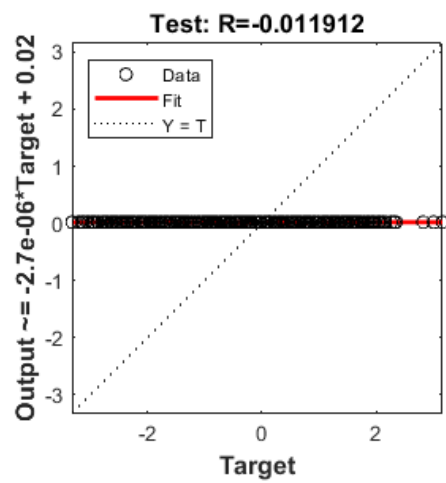
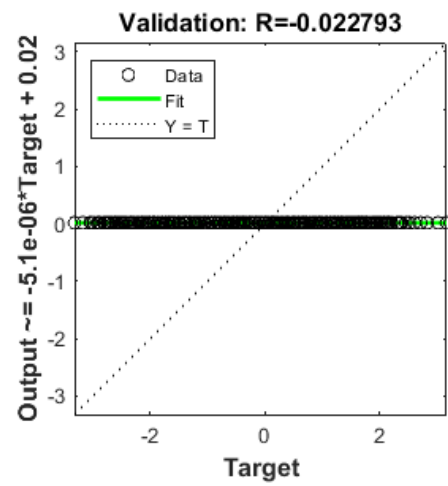
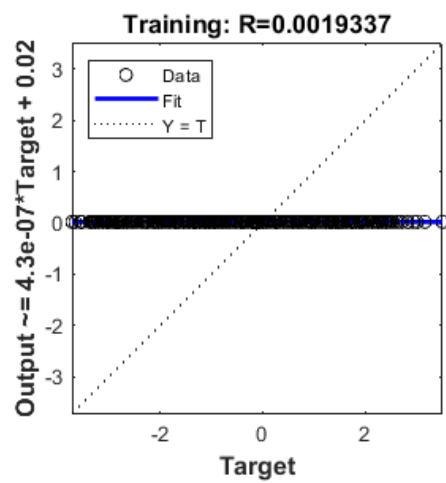
Για H=20 νευρώνες στο στο κρυφό επίπεδο

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.035$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.824$



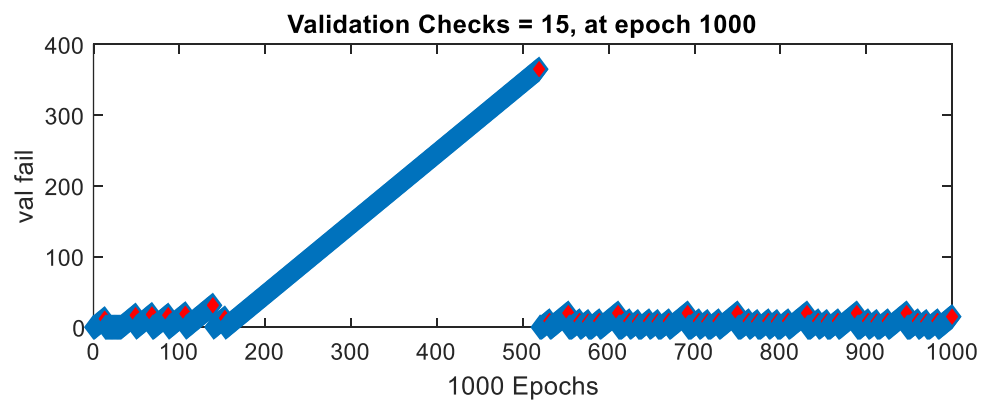
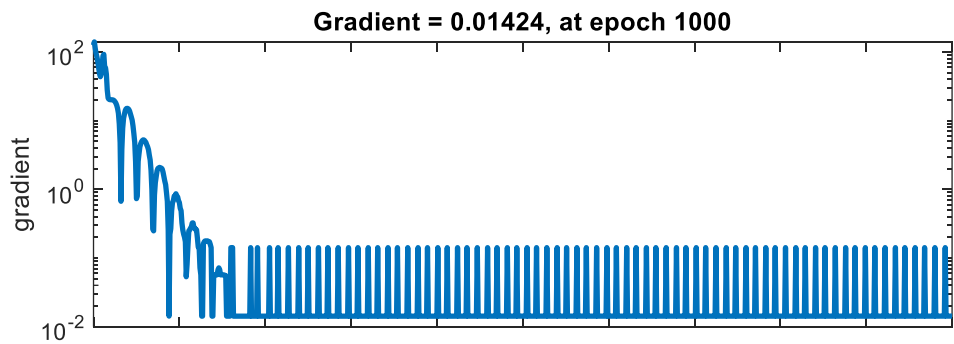
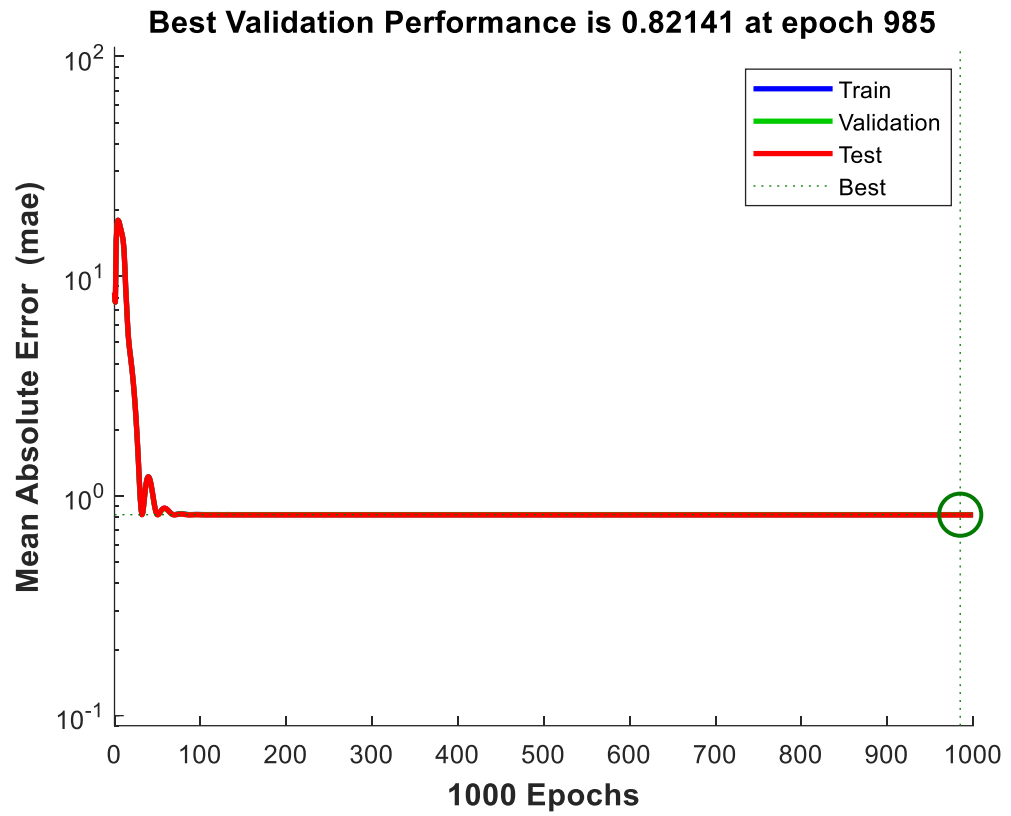


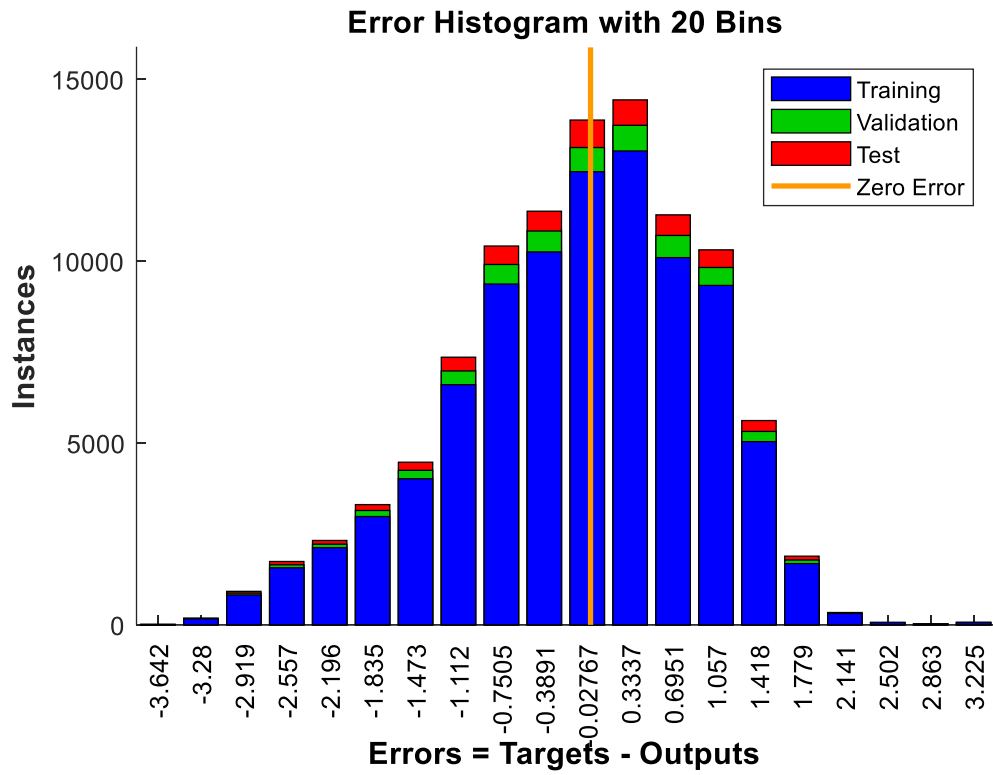


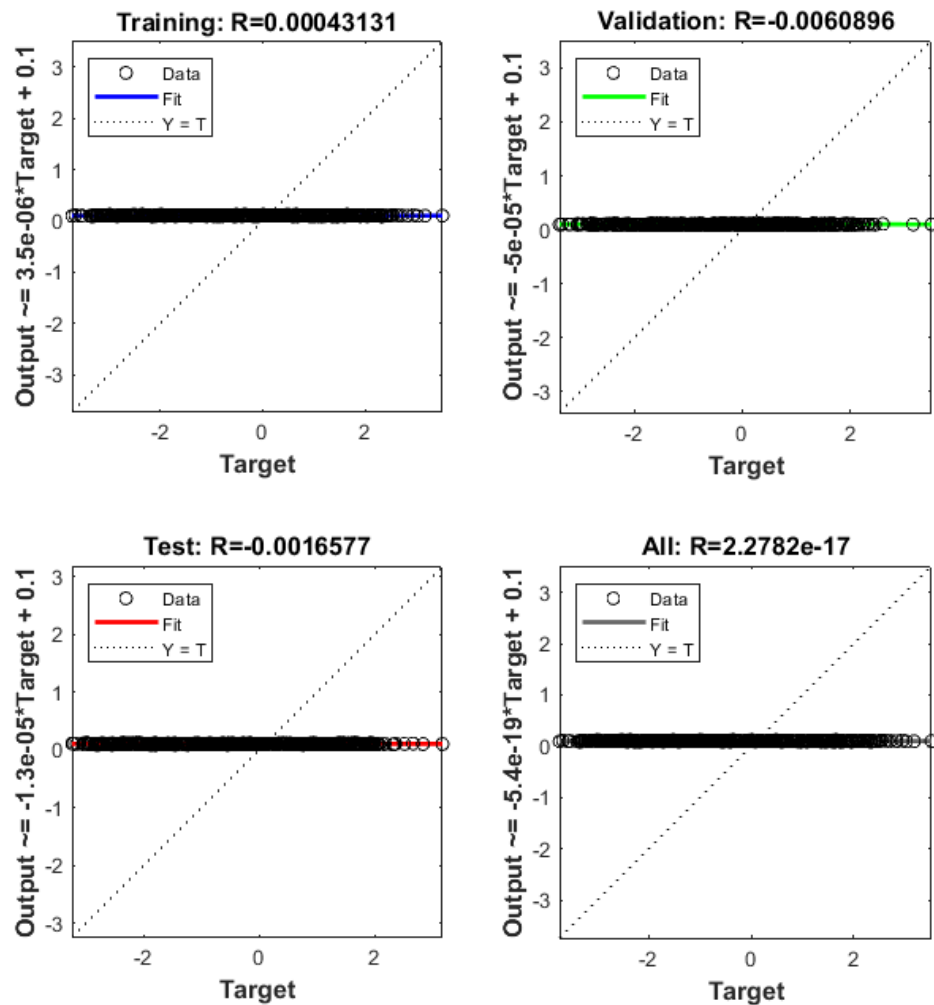
Για H=30 νευρώνες στο στο κρυφό επίπεδο

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.034$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.821$







Επομένως, το νευρωνικό δίκτυο με τους 30 νευρώνες στο κρυφό επίπεδο έχει καλύτερη απόδοση.

ζ) Θα χρησιμοποιήσουμε την τεχνική early stopping. Το early stopping απαιτεί να ρυθμίσουμε το δίκτυο μας έτσι ώστε να είναι υπό περιορισμό, πράγμα που σημαίνει ότι έχει μεγαλύτερη χωρητικότητα από ότι απαιτείται για το πρόβλημα. Το επιλέγουμε γιατί παρακολουθούμε την απόδοση του μοντέλου, δίνει το έναυσμα να σταματήσουμε την προπόνηση και επειδή εμείς επιλέγουμε το μοντέλο που θα χρησιμοποιήσουμε.

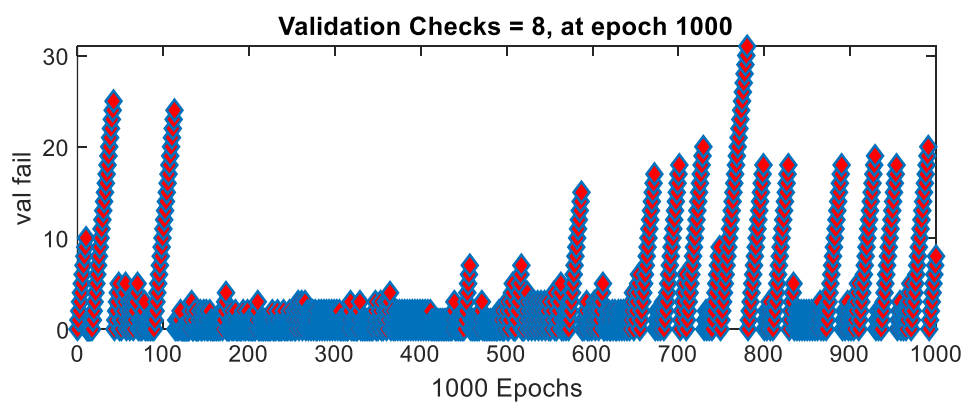
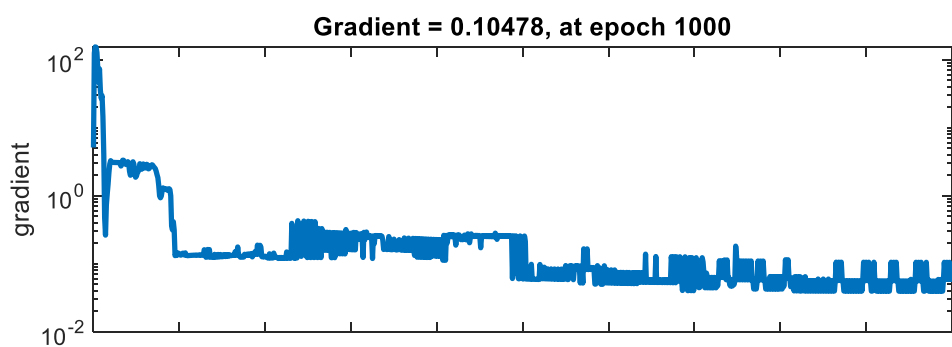
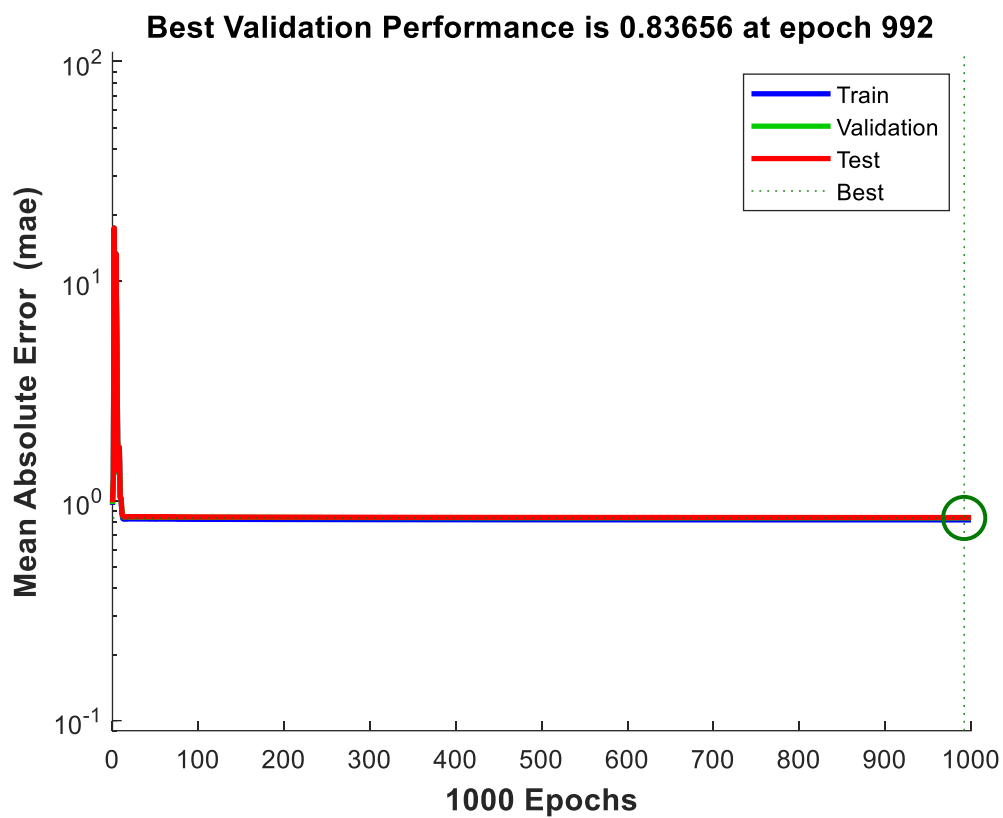
A3)

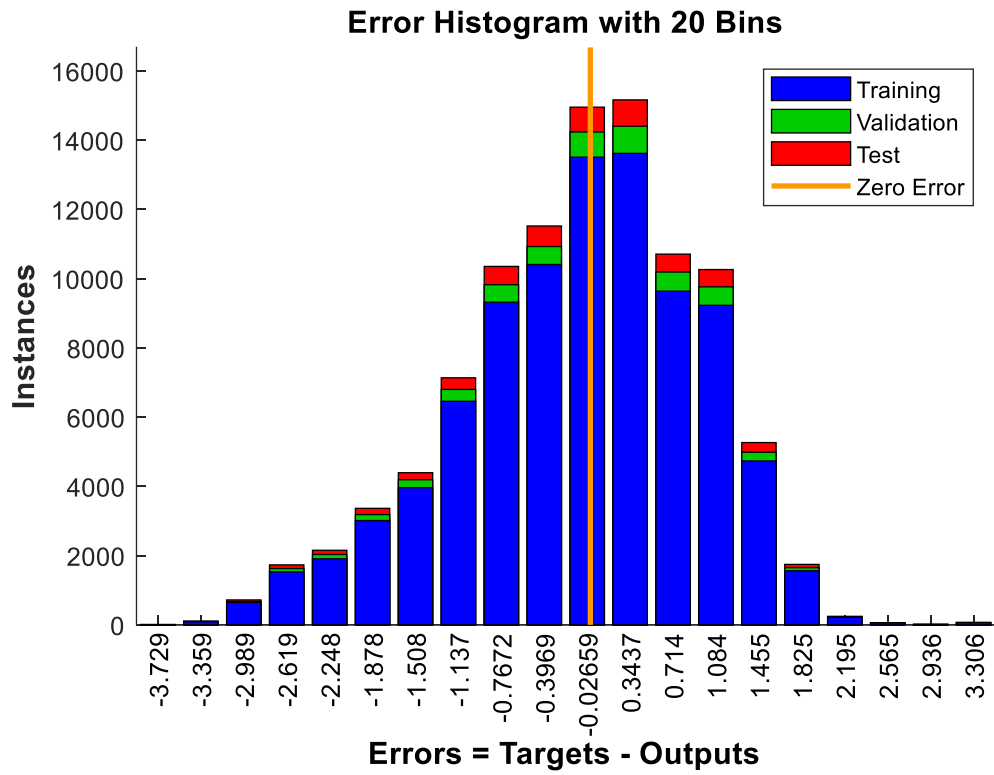
Η τοπολογία με το καλύτερο αποτέλεσμα είναι αυτή με τους 30 νευρώνες στο κρυφό επίπεδο, επομένως θα εκπαιδεύσουμε αυτή.

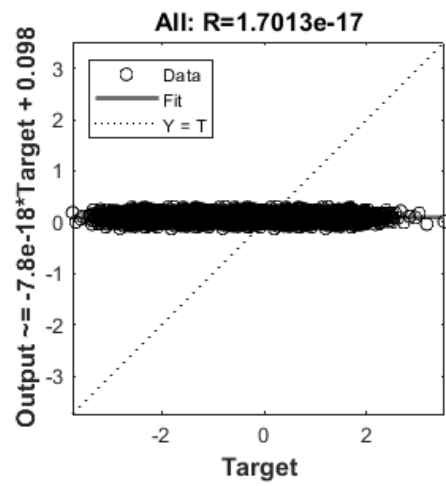
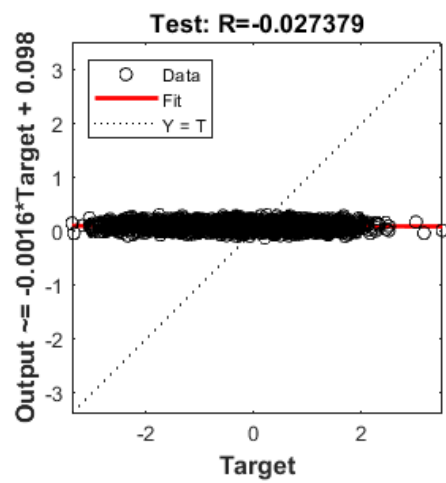
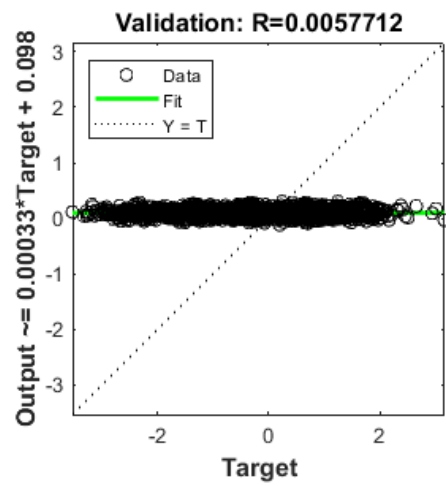
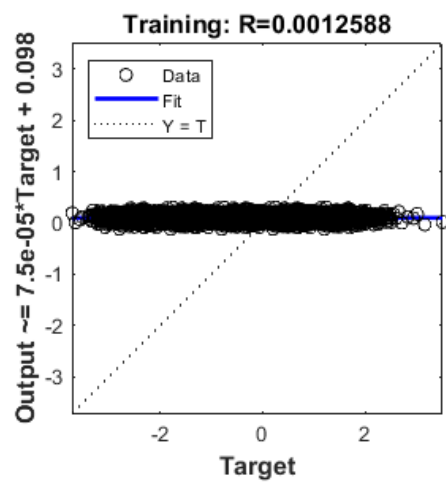
Για $\eta = 0.001$ και $m = 0.2$

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $rmse = 1.0372$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $mae = 0.836$



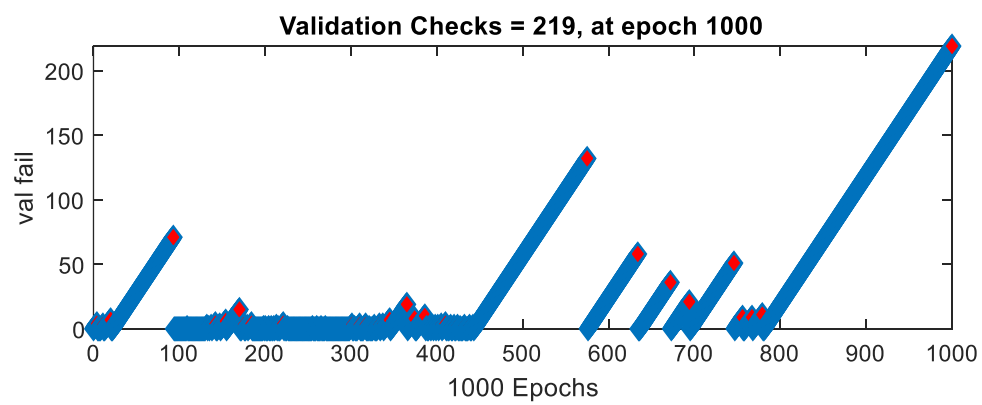
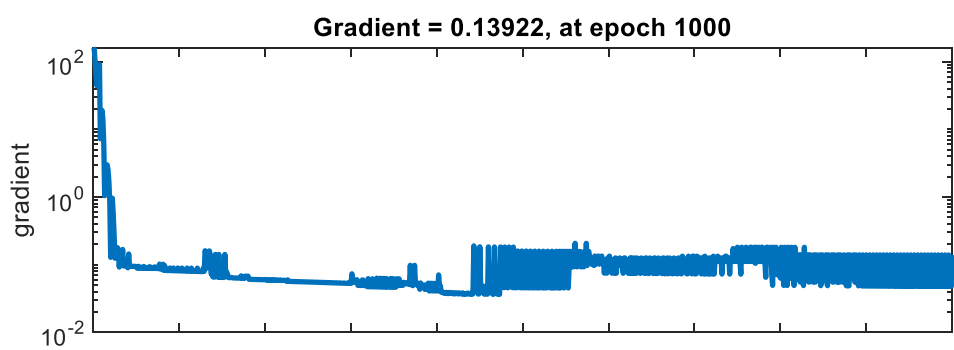
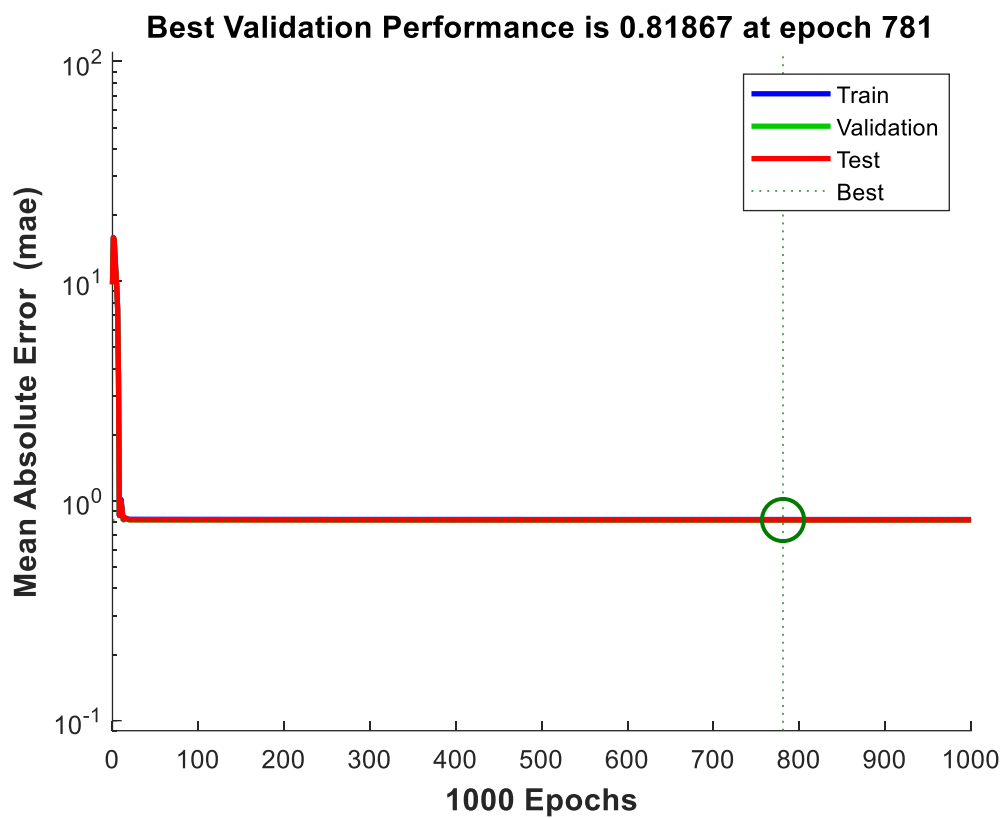


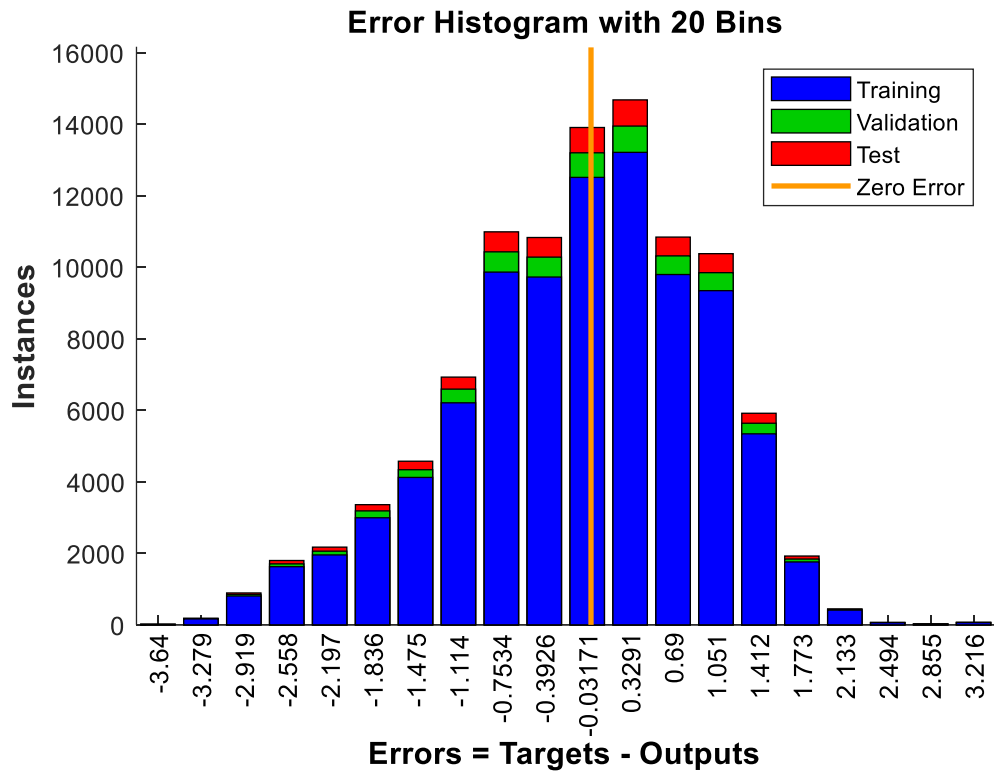


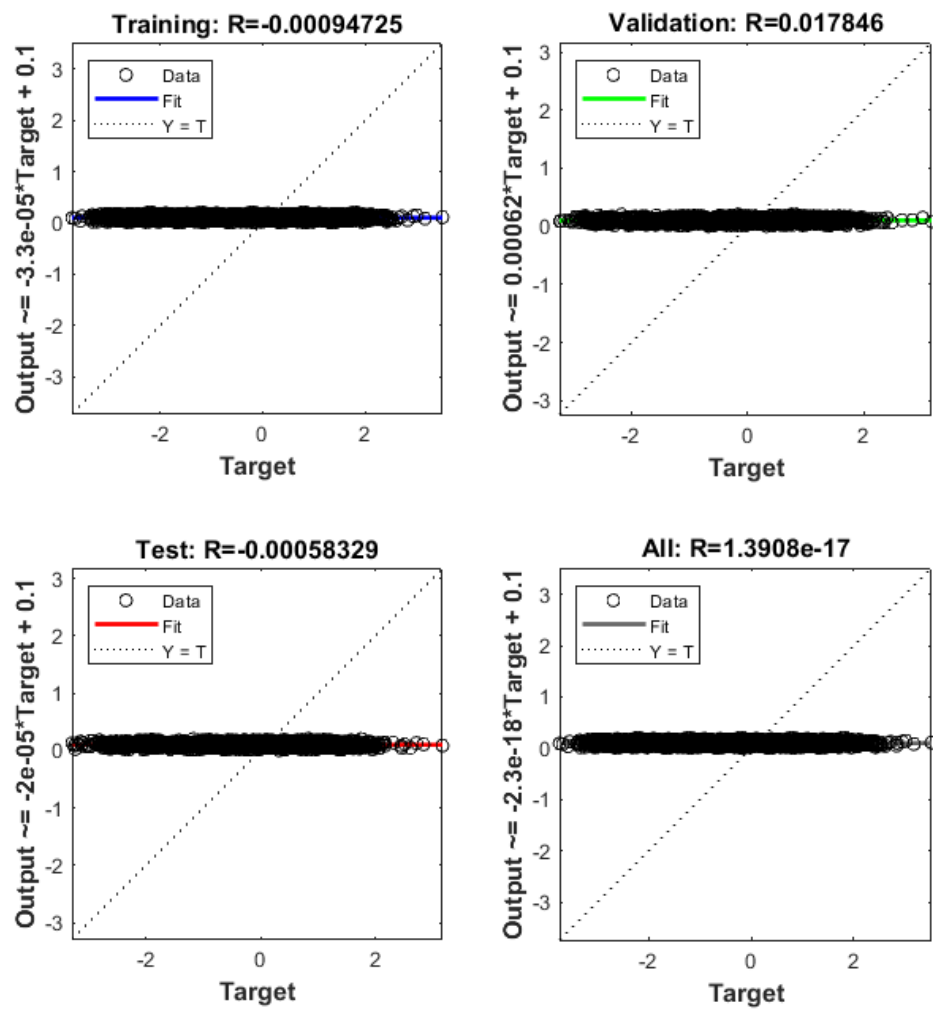
Για $\eta = 0.001$ και $m = 0.6$

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0364$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.818$



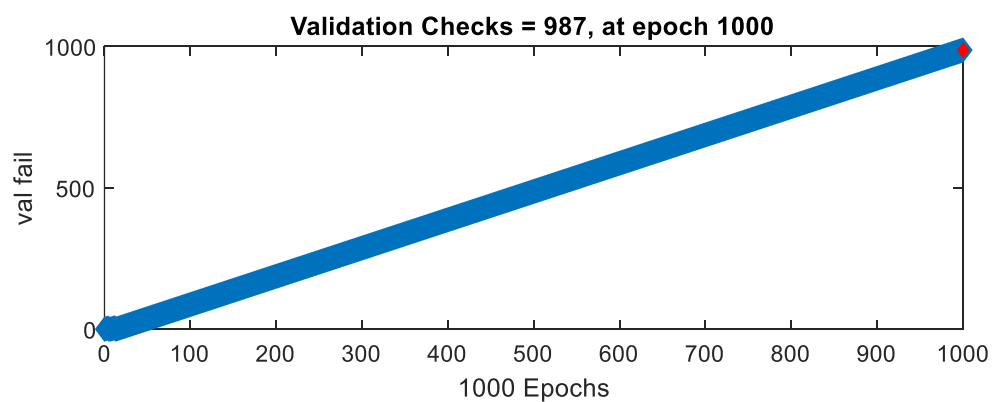
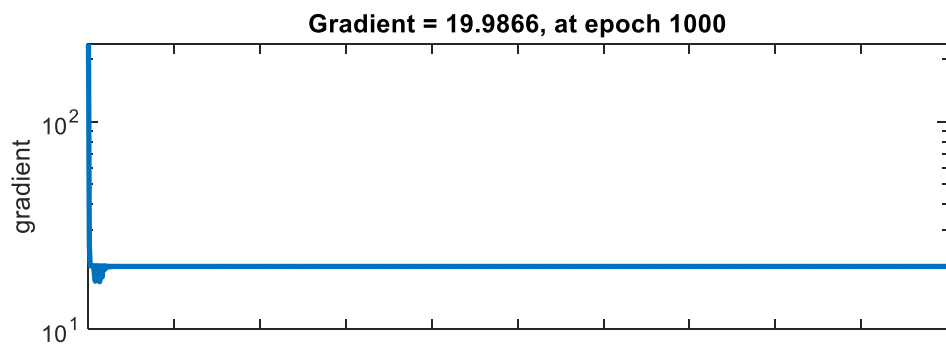
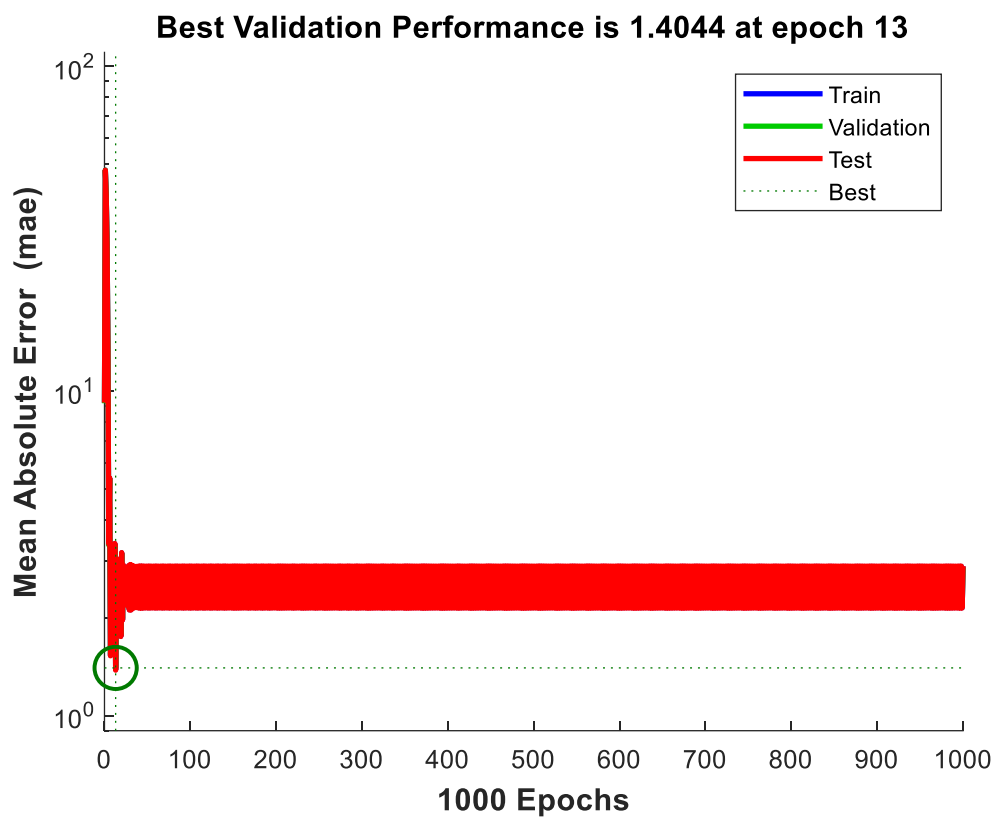


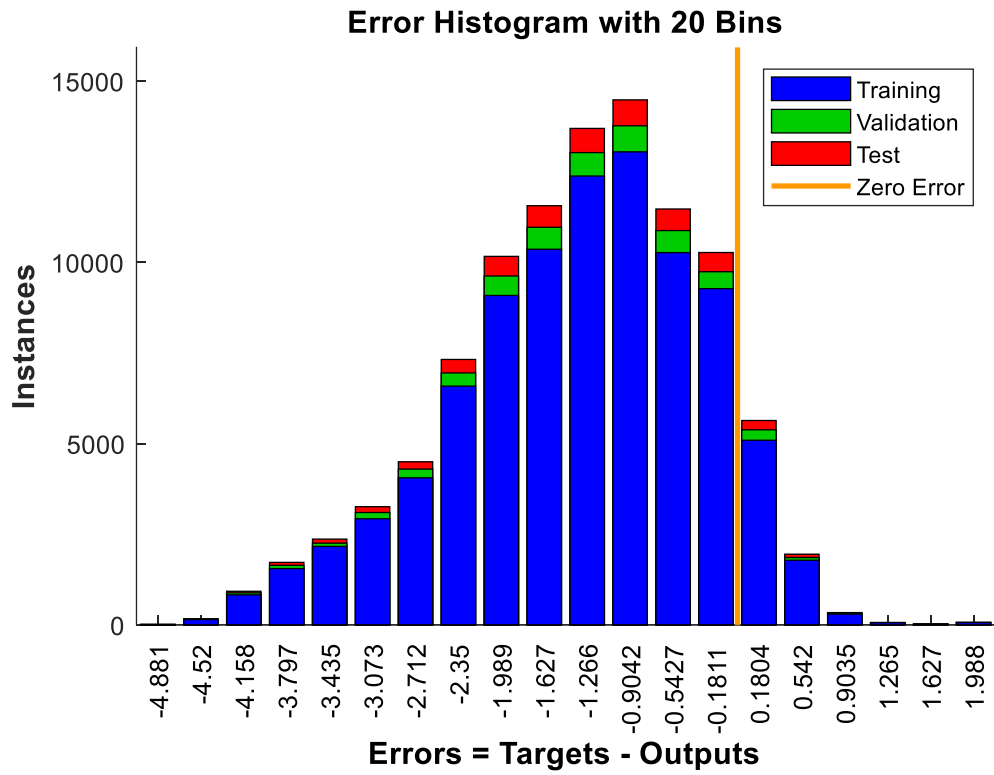


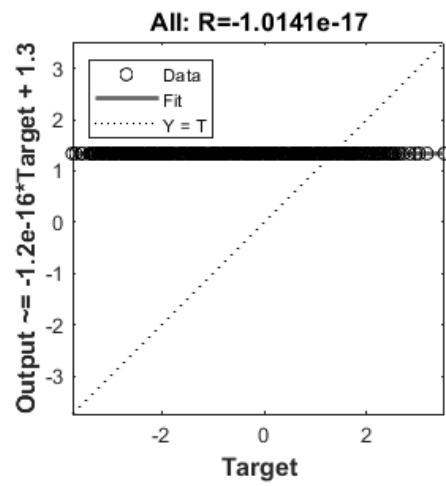
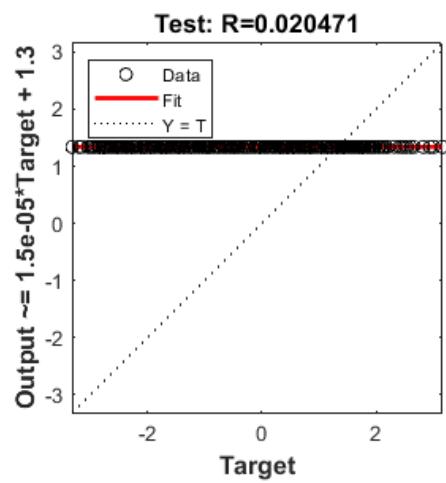
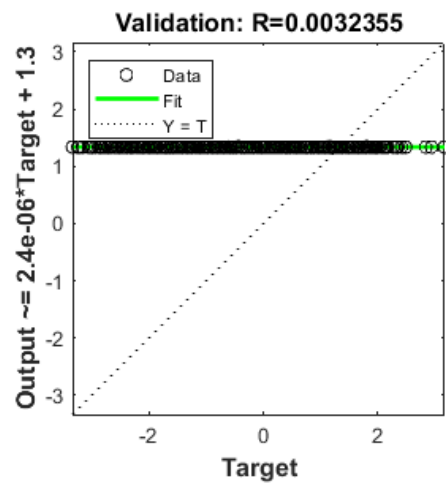
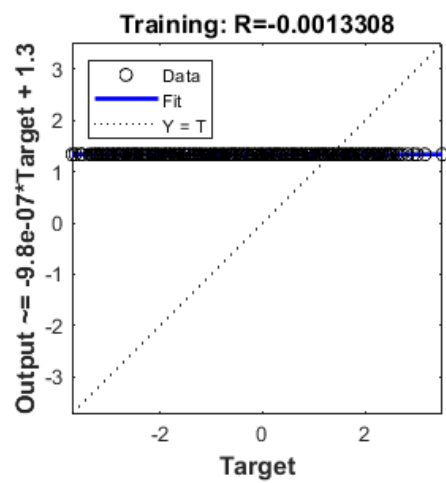
Για $\eta = 0.05$ και $m = 0.6$

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.6897$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 1.404$



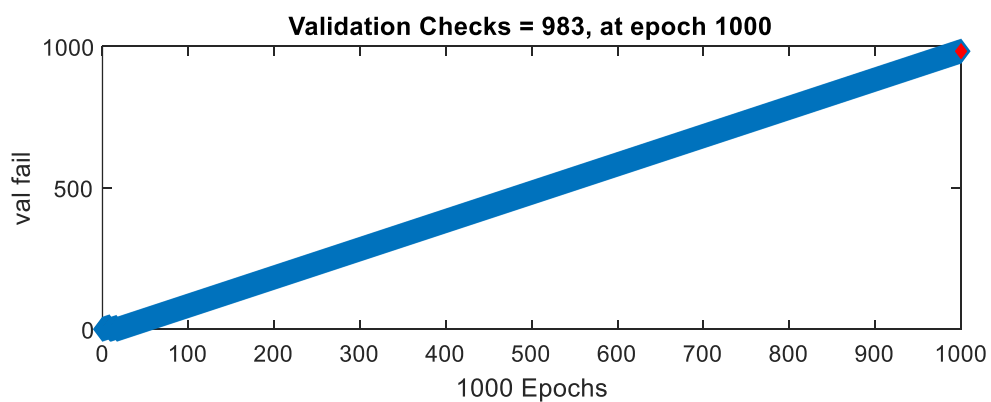
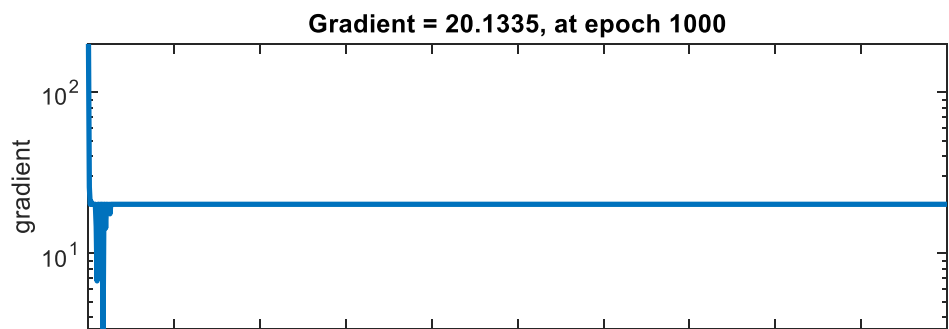
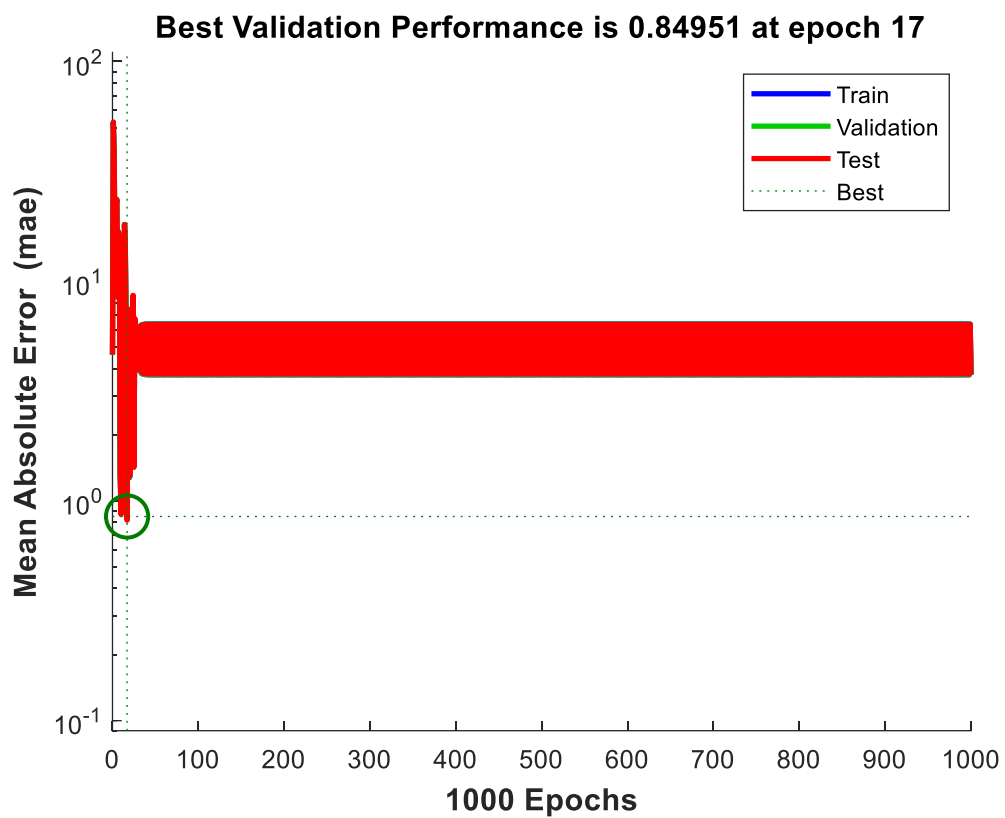


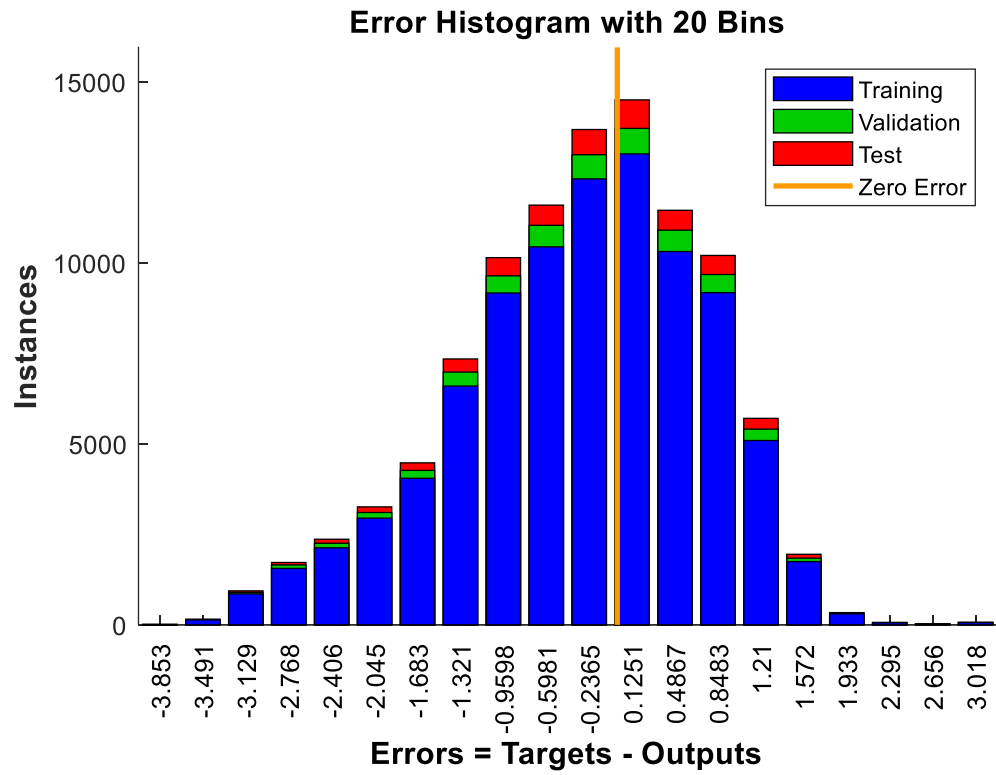


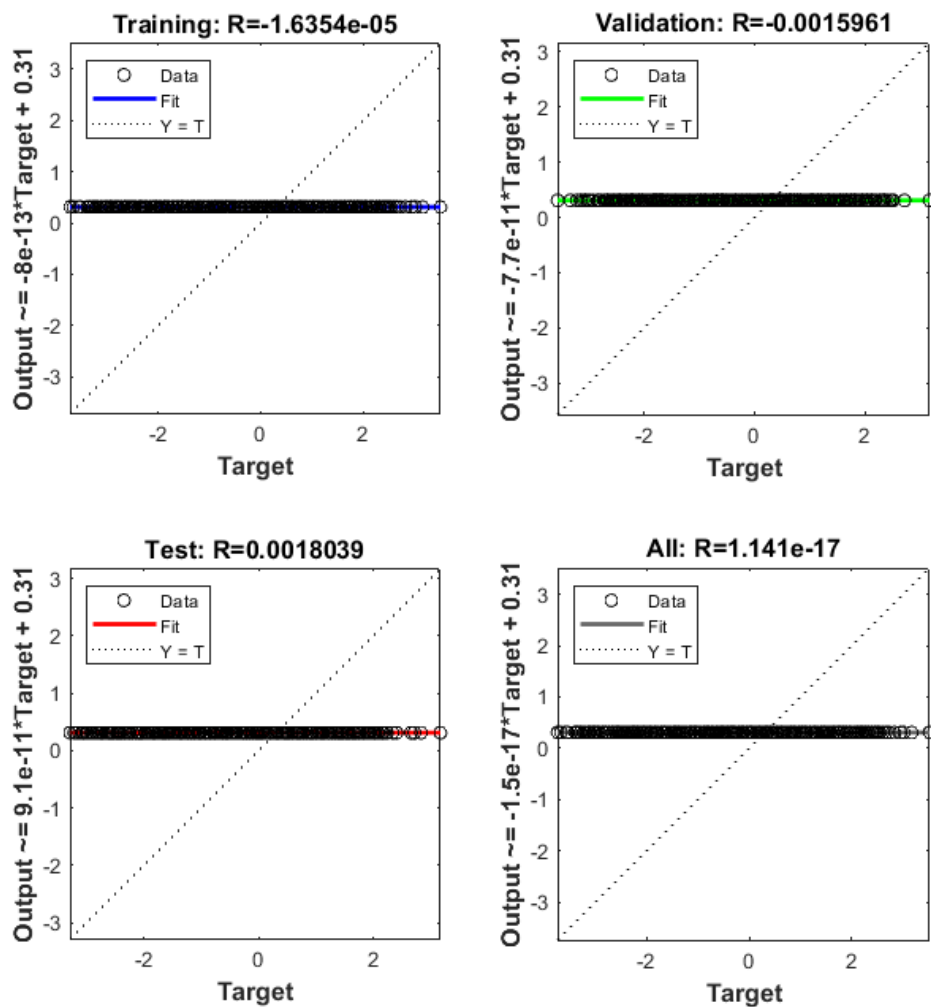
Για $\eta = 0.1$ και $m = 0.6$

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0762$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.839$







Παρατηρούμε ότι η εκπαίδευση με ρυθμό εκπαίδευσης $\eta = 0,001$ και σταθερά ορμής $m = 0,6$ έχει καλύτερη απόδοση.

Να τεκμηριώσετε θεωρητικά γιατί $m < 1$:

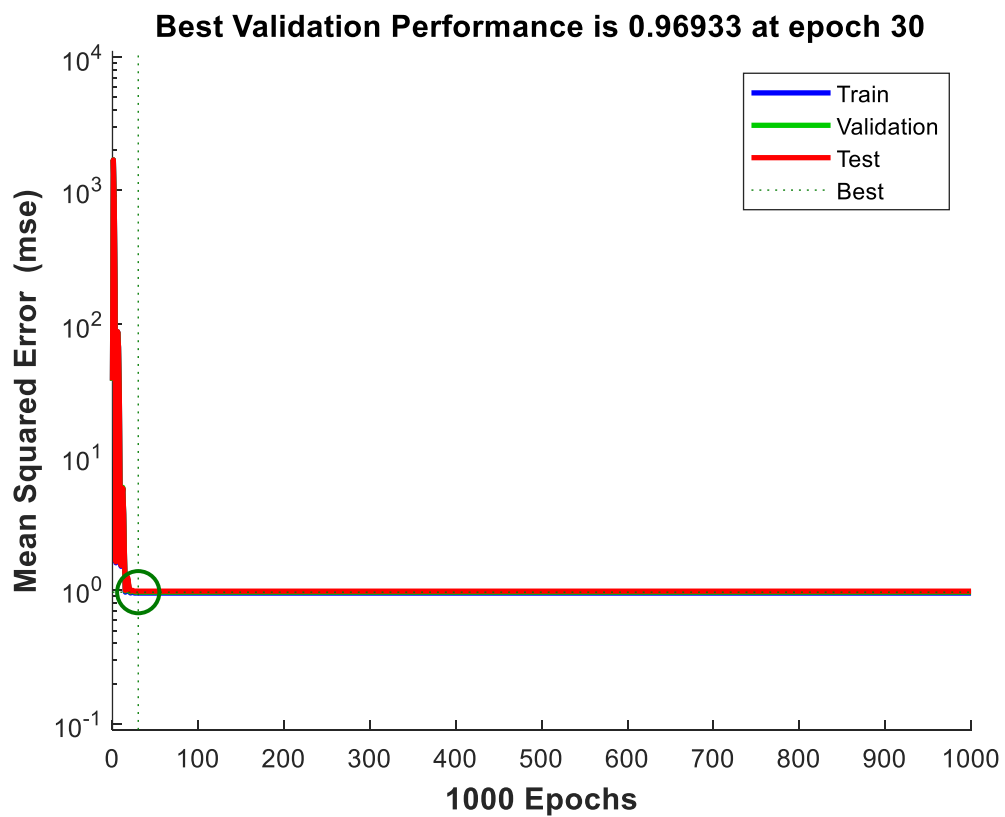
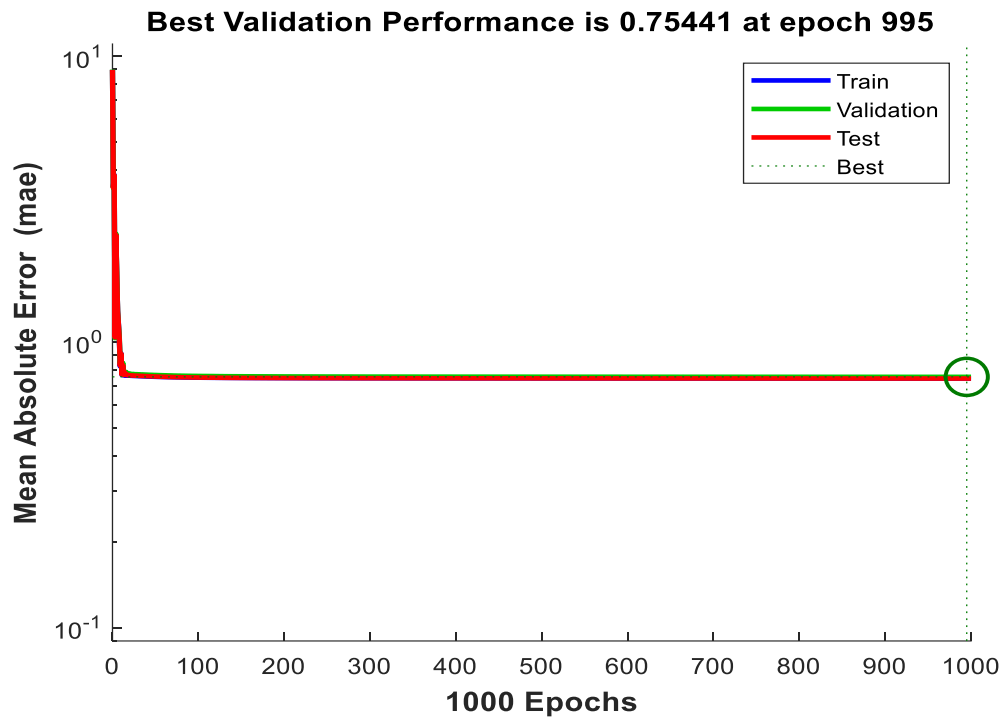
Η ορμή κάνει τα βάρη να μαθαίνουν λιγότερο, δηλαδή ο αντίκτυπος της νέας μάθησης είναι λιγότερος από τον αντίκτυπο των προηγούμενων μαθησεων. Εάν η ορμή είναι πολύ υψηλή το νευρωνικό δίκτυο θα δυσκολευτεί να μαθαίνει νέα πράγματα και τα παλαιότερα πράγματα που έχει μάθει θα αλλάξουν τα βάρη σε εκθετική τιμή. Ακόμη, αν ο gradient descent είναι σαν να κατεβαίνει μια πλαγιά η ορμή είναι η κυριολεκτική ορμή του παράγοντα που διασχίζει το όριο απόφασης. Έτσι, ο συντελεστής της ορμής θα ήταν ανάλογος με τον συντελεστή τριβής, με το 1 να είναι η μέγιστη τριβή και το 0 να μην είναι τριβή. Δεν μπορεί να υπάρχει τριβή πέρα από αυτό το εύρος γιατί: Αν η τριβή = 1 θα ήταν ίδια με το να μην έχουμε τριβή. Αν η τριβή ≤ 0 τότε με διατήρηση του gradient descent δεν θα βρεθεί τοπικό ελάχιστο. Αν η τριβή > 1 τότε το gradient descent θα κινείται προς τα πίσω.

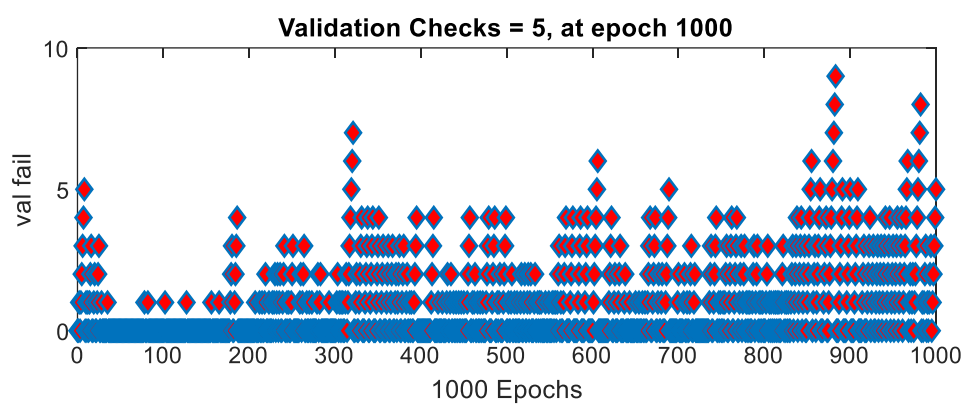
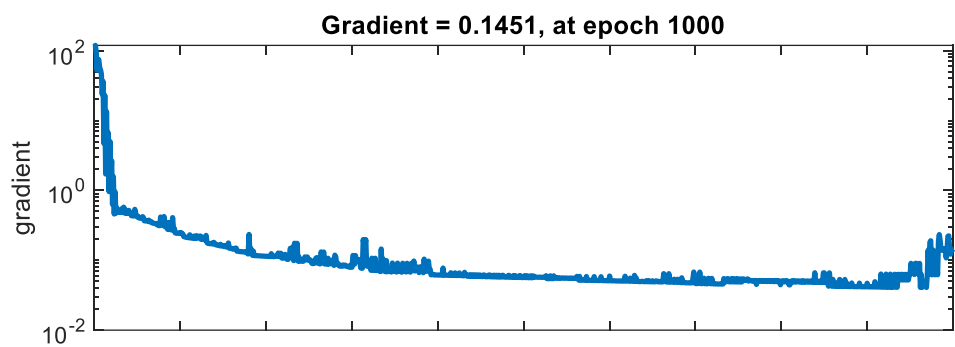
A4)

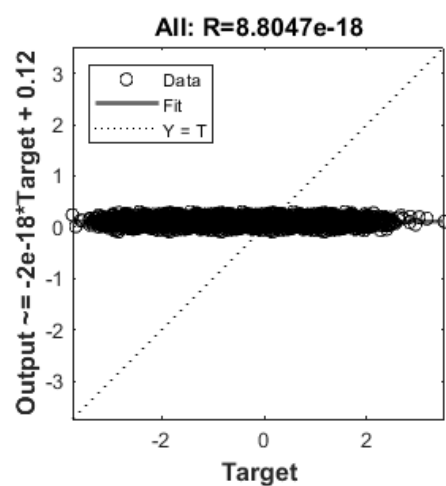
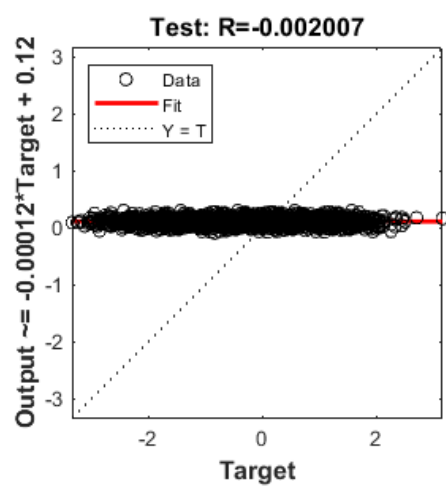
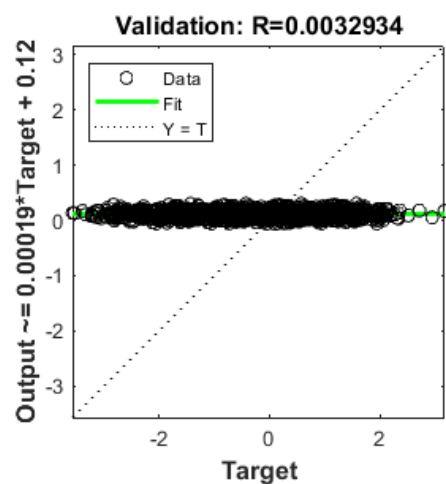
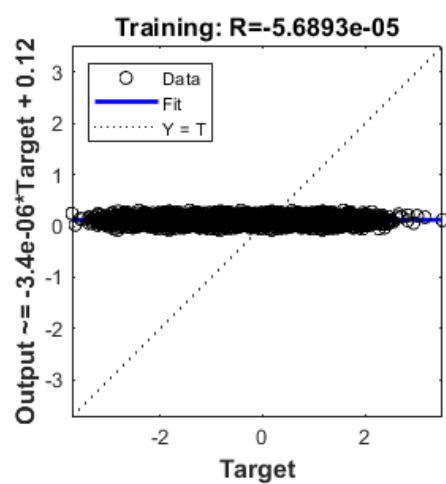
Για $r = 0.1$

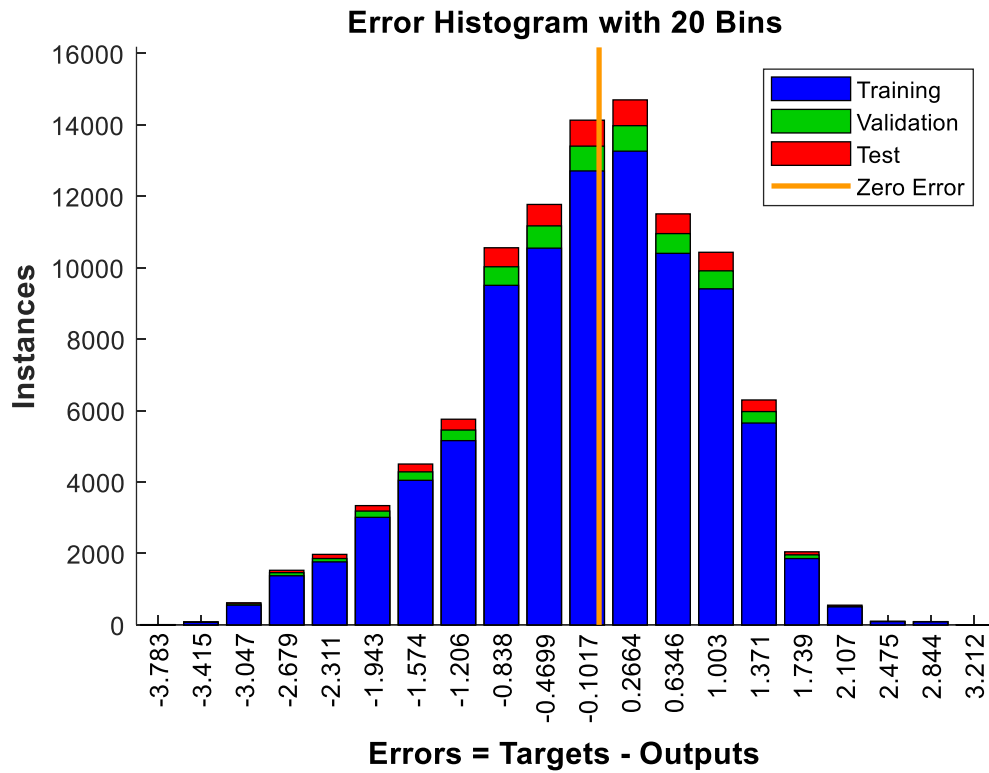
η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 0,984$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.744$





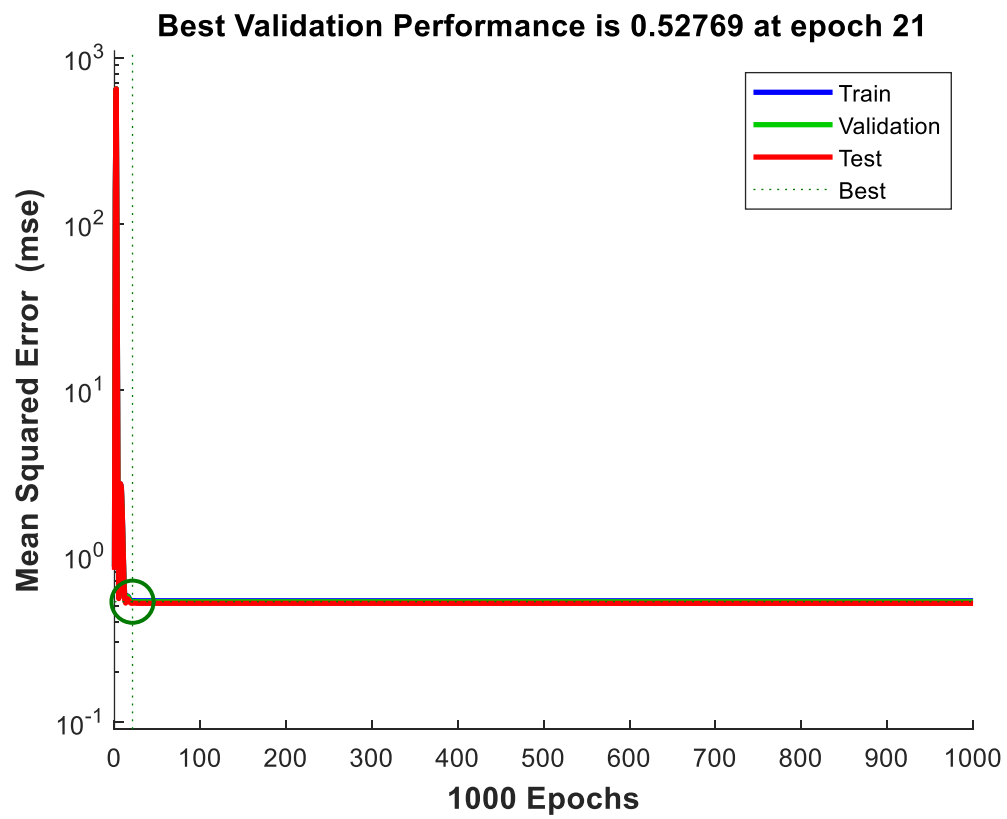
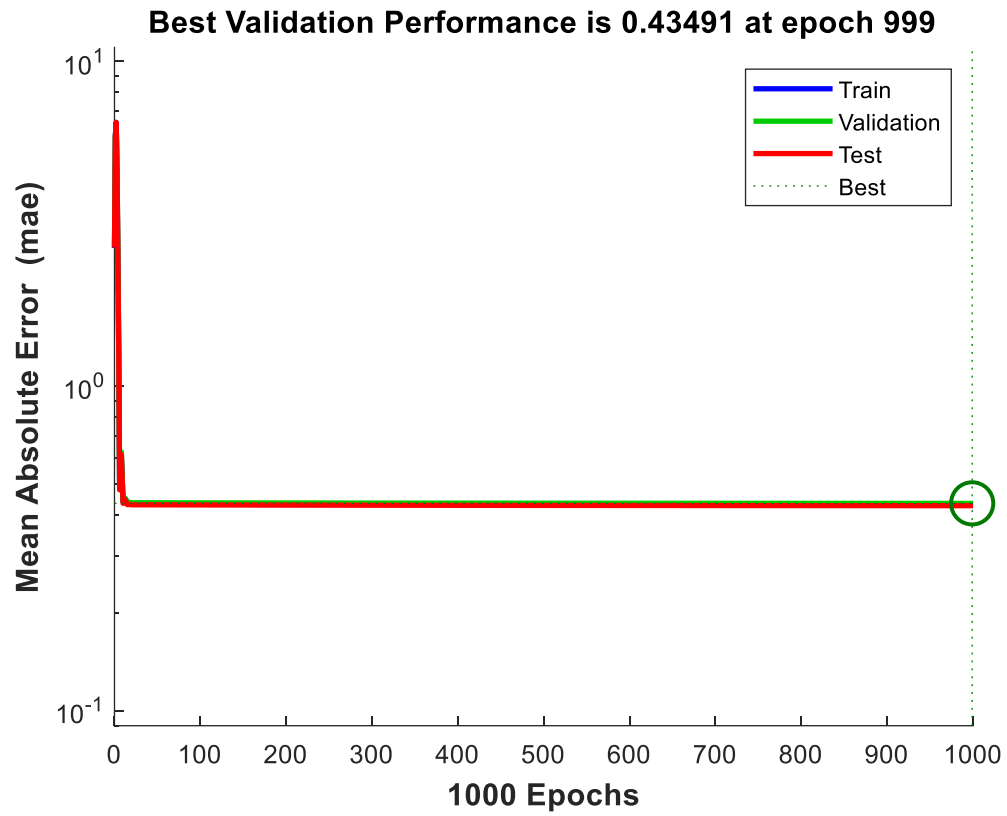


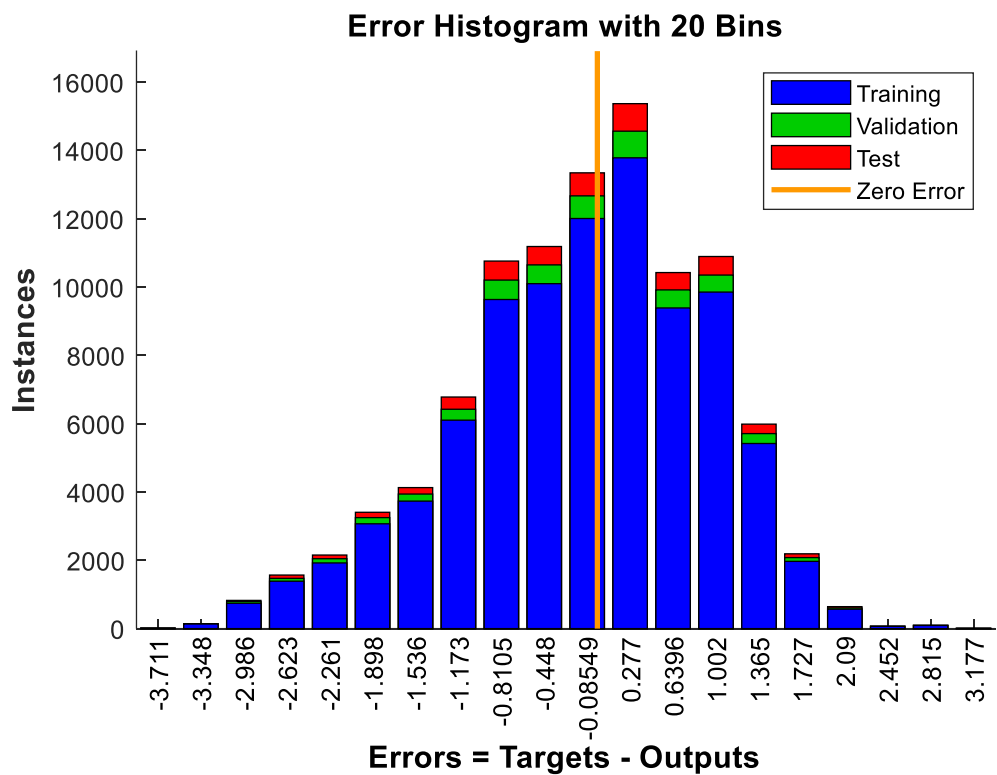
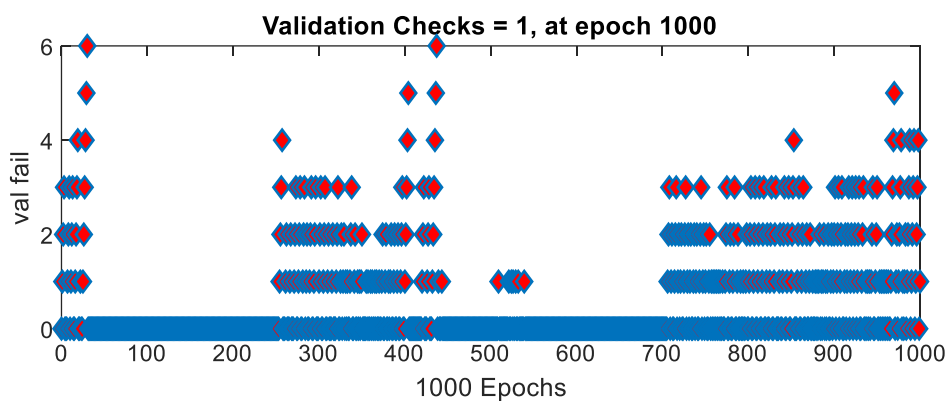
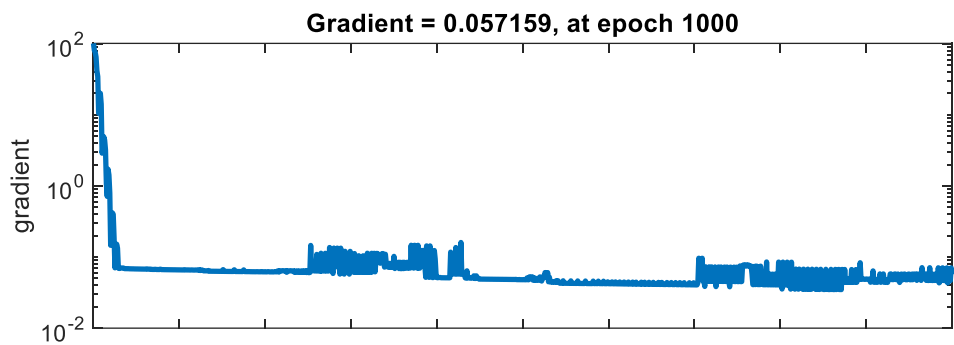


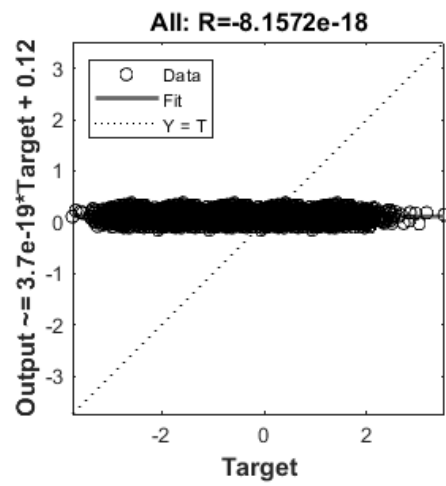
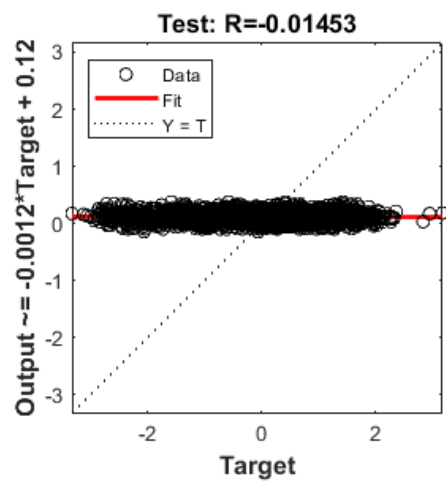
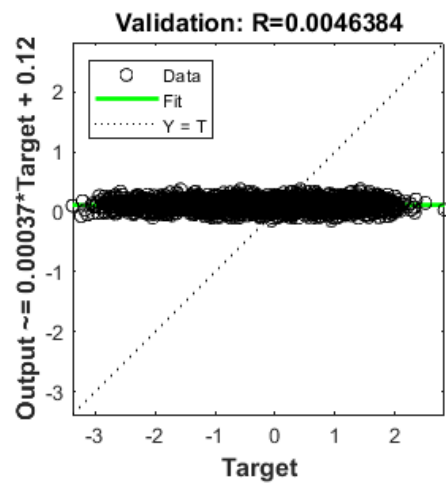
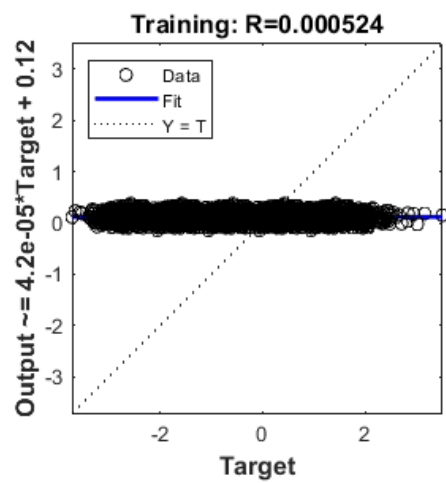
Για $r = 0.5$

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $rmse = 0.725$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $mae = 0.432$



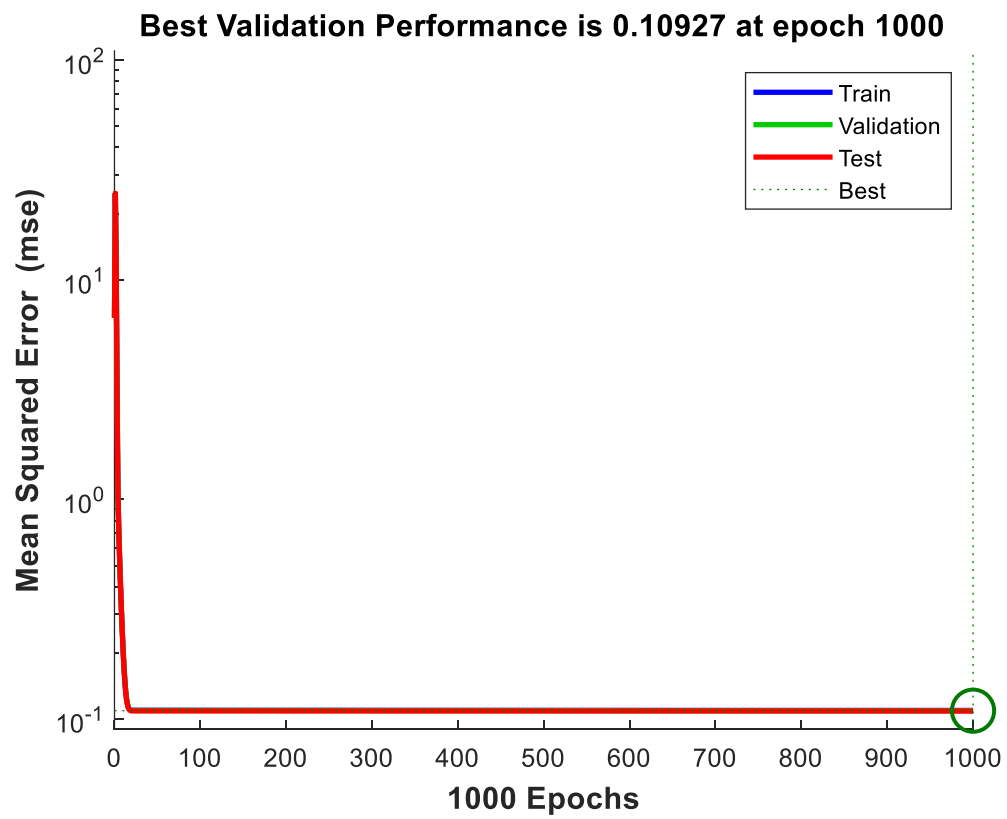
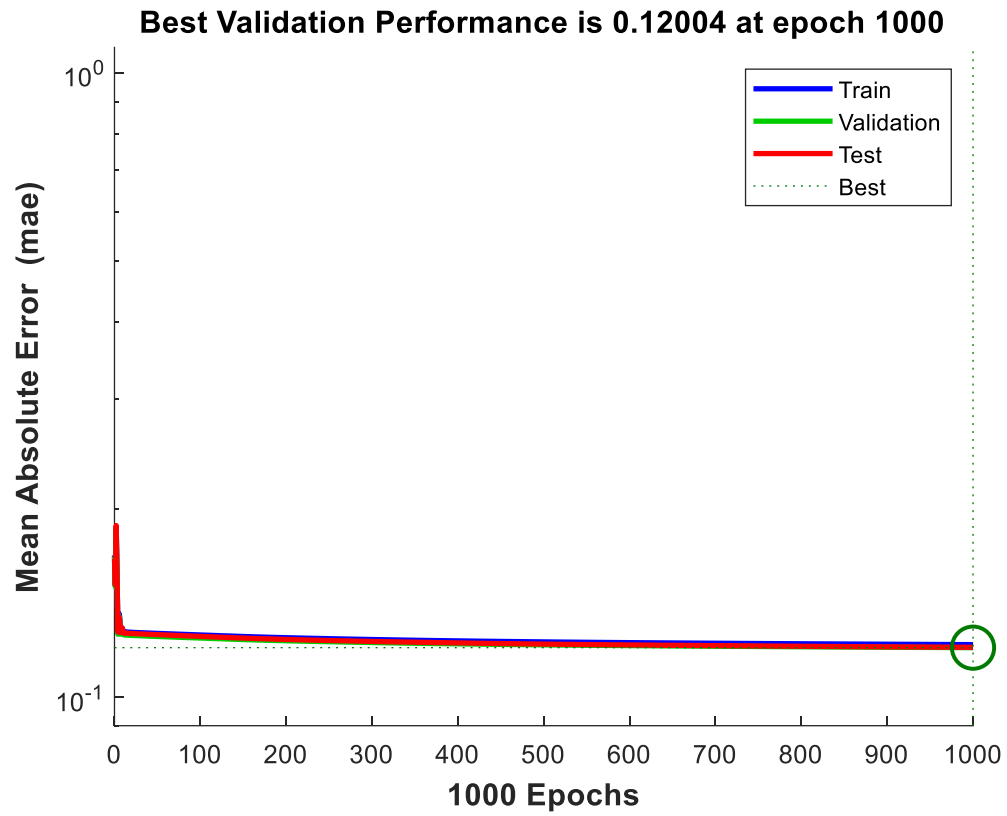


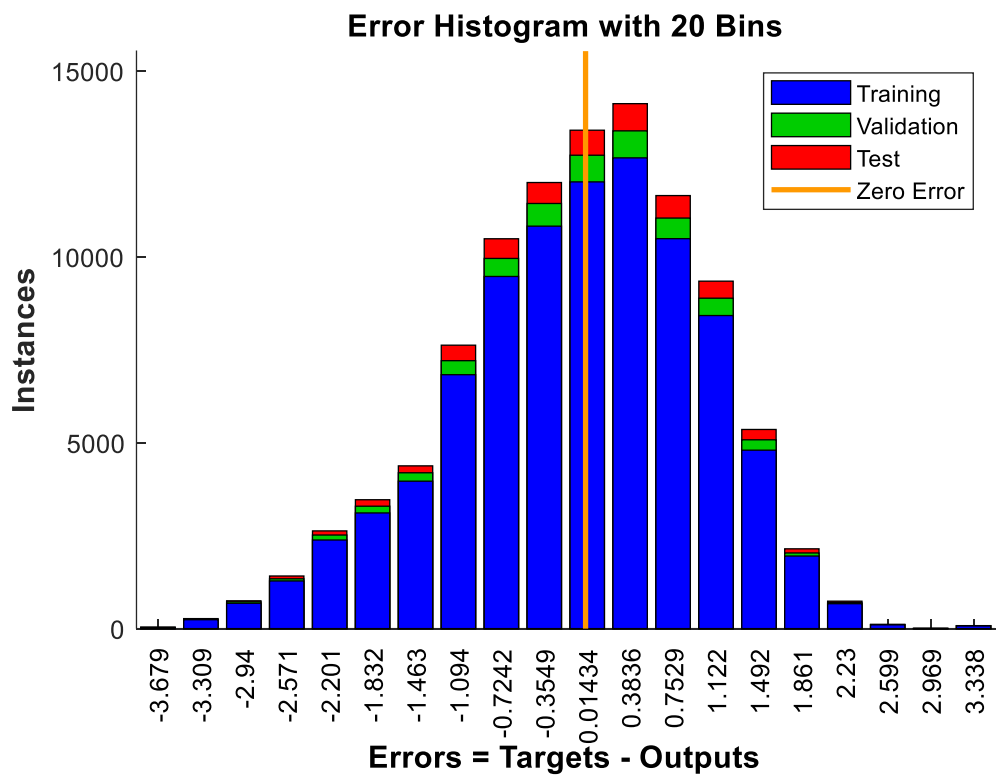
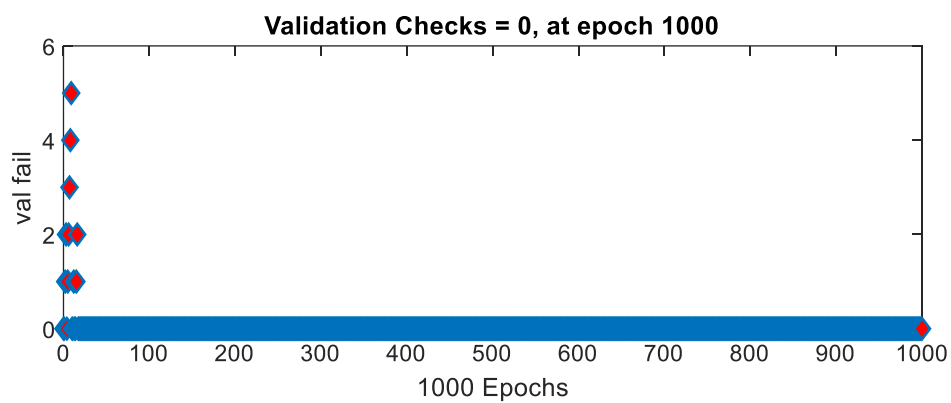
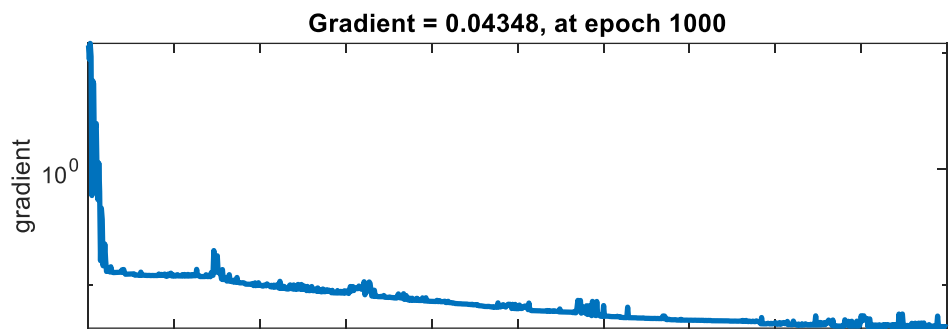


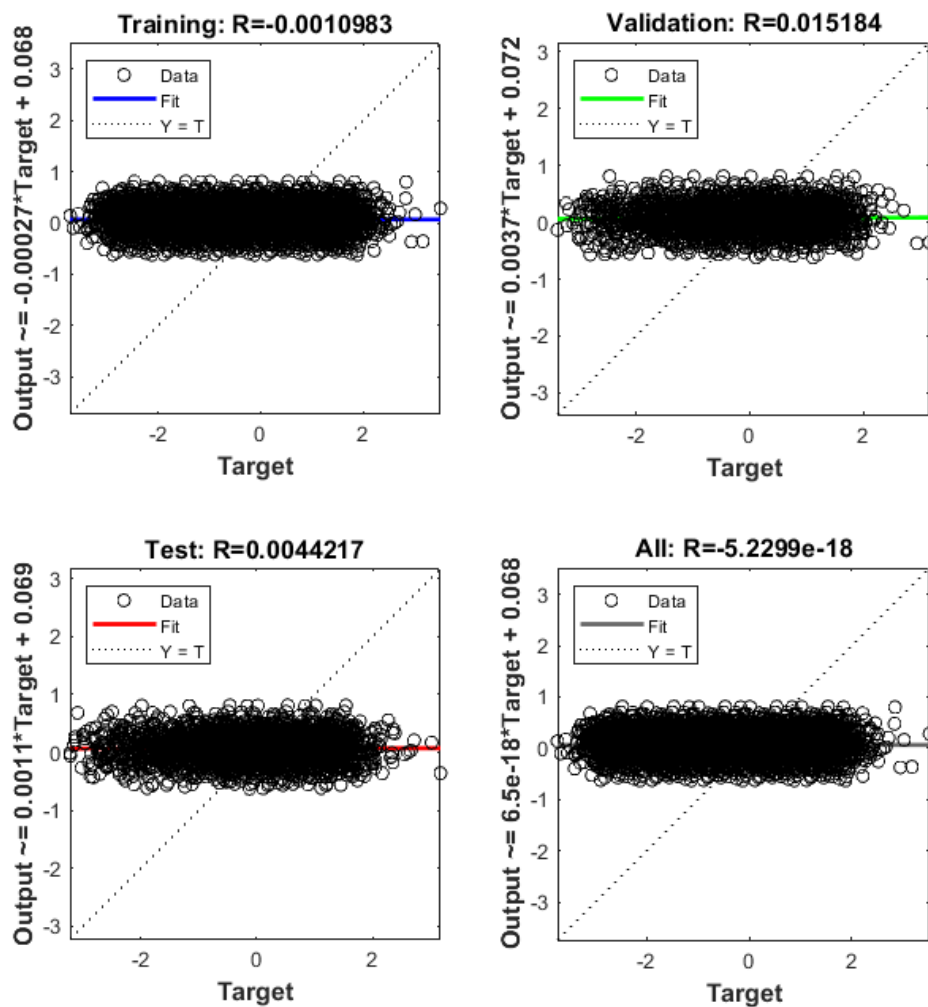
Για $r = 0.9$

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 0.33$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.121$







Η L1 ομαλοποίηση συρρικνώνει τον συντελεστή του λιγότερο σημαντικού χαρακτηριστικού στο μηδέν, αφαιρώντας εντελώς κάποια λειτουργία. Αυτό λειτουργεί καλά σε περίπτωση που έχουμε τεράστιο αριθμό χαρακτηριστικών. Για αυτό είναι προτιμότερη η L1 ομαλοποίηση στην περίπτωση μας.

Παρατηρούμε ότι εκπαιδεύοντας το νευρωνικό δίκτυο με συντελεστή φθοράς $r = 0.9$ πετυχαίνουμε καλύτερη απόδοση. Αυτό γιατί έχουμε μειώσει την υπερεκπαίδευση και την πολυπλοκότητα του μοντέλου. Γενικά με την L1 ομαλοποίηση επιτυγχάνουμε αρκετά καλή γενίκευση για το δίκτυο μας.

A5)

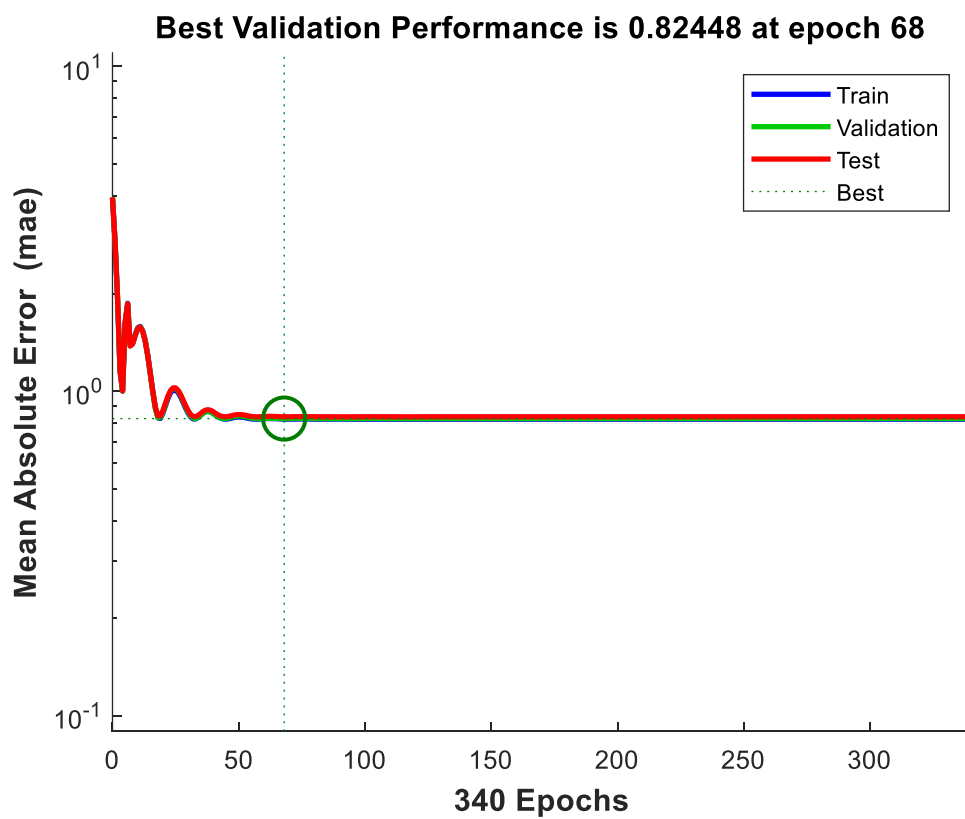
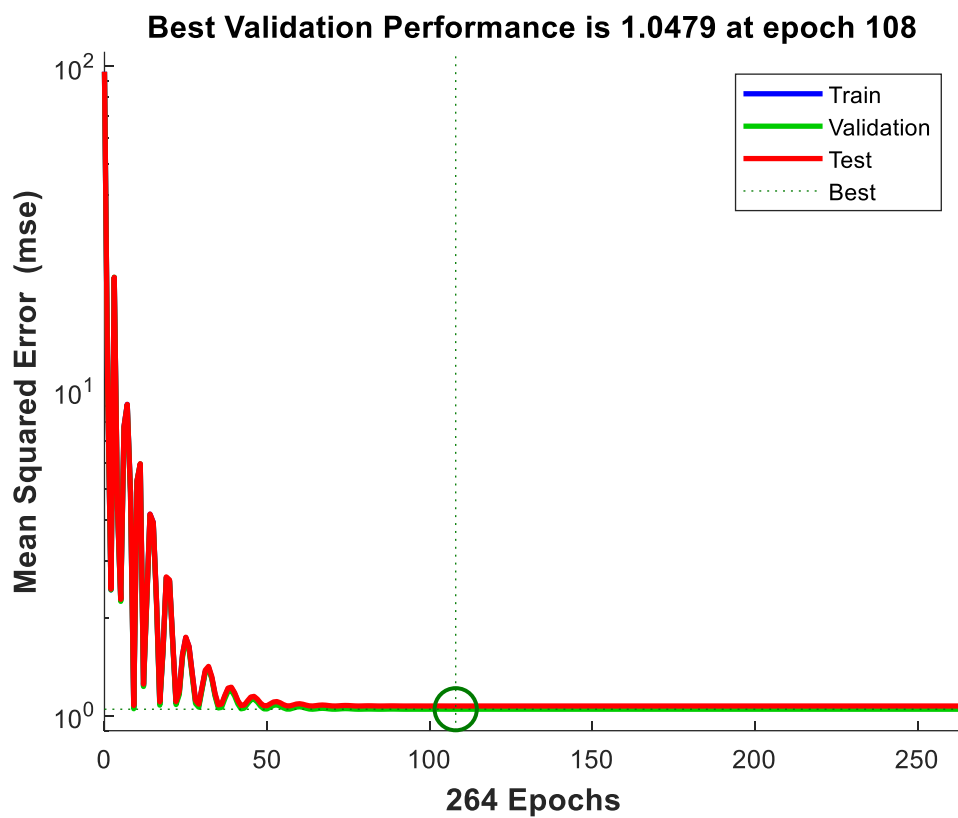
Κόμβοι 1^{ου} κρυφού επιπεδου: 3

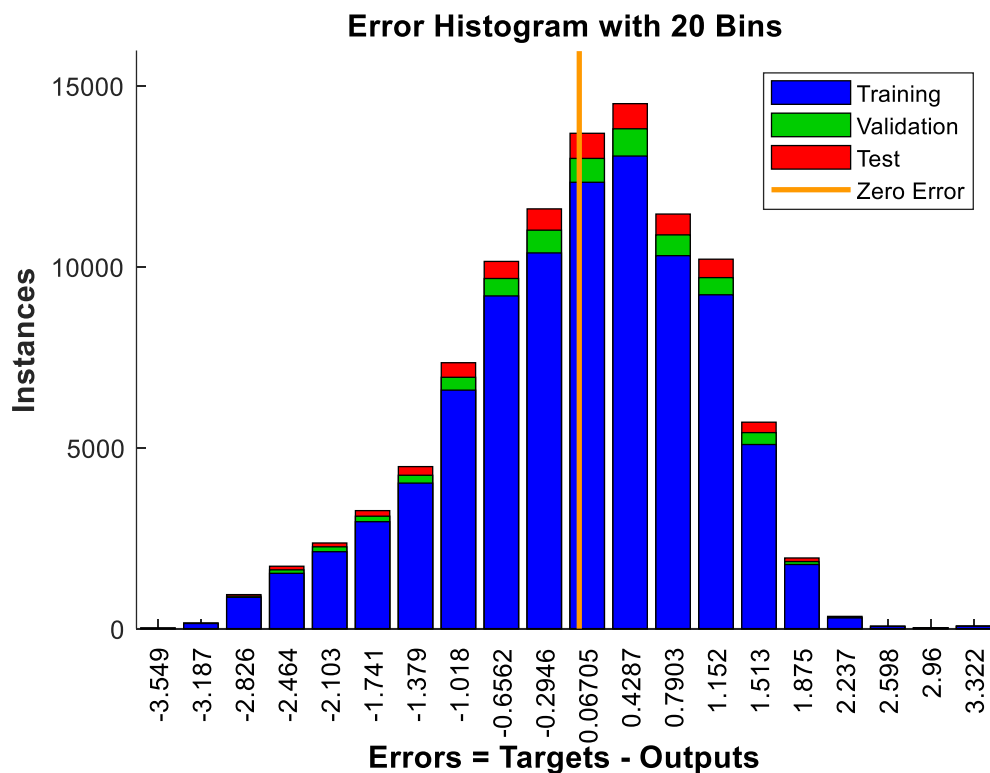
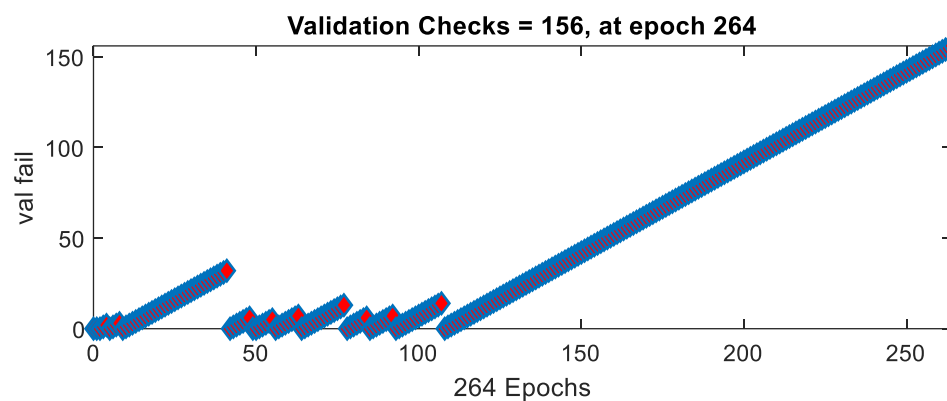
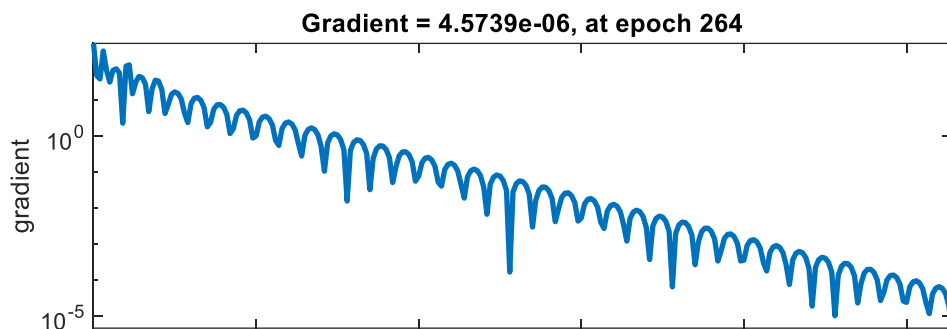
Κόμβοι 2^{ου} κρυφού επιπεδου: 4

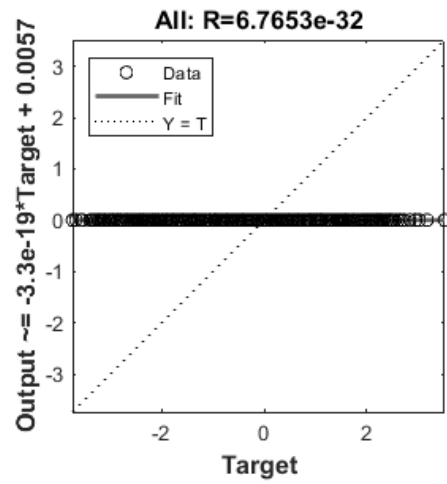
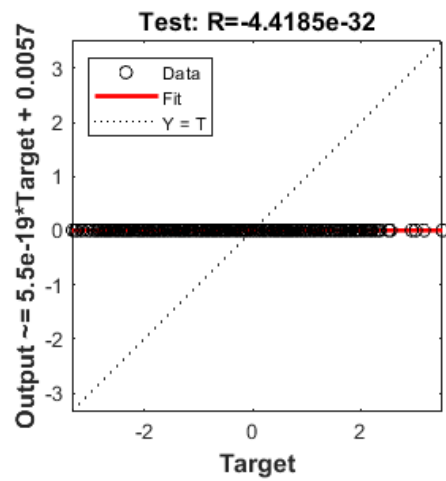
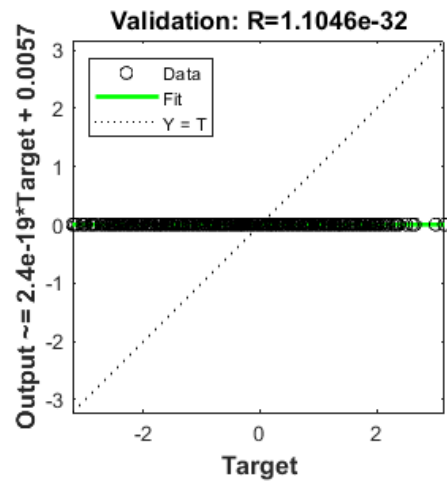
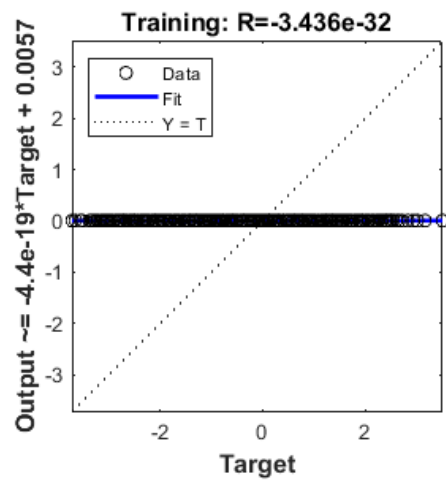
Κόμβοι 3^{ου} κρυφού επιπεδου: 5

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0236$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.8244$







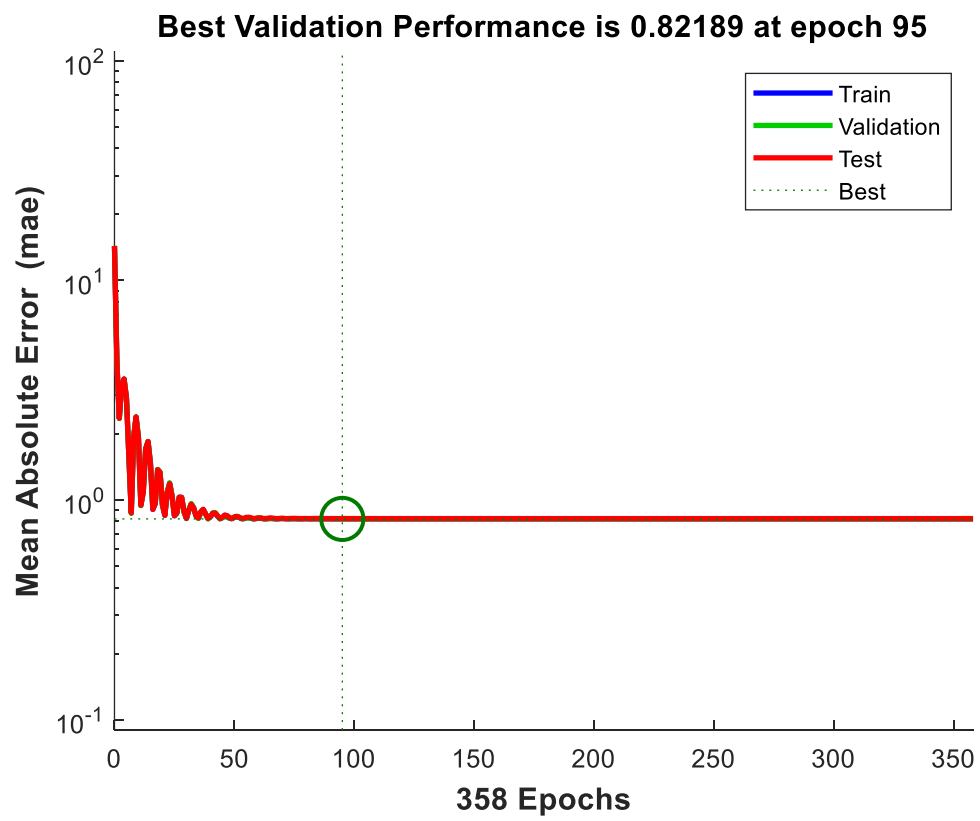
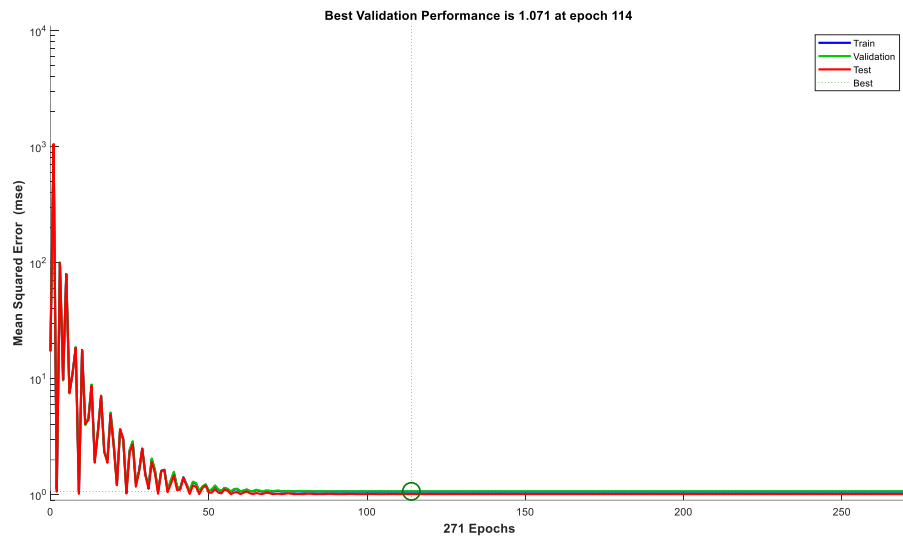
Κόμβοι 1^{ου} κρυφού επιπεδου: 10

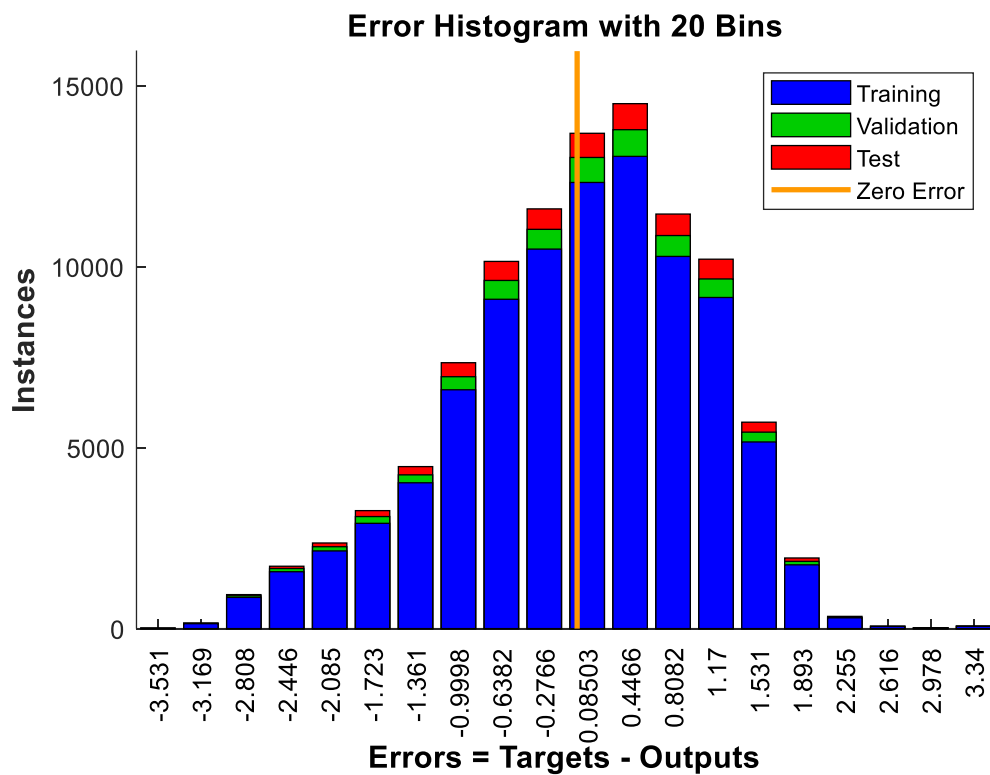
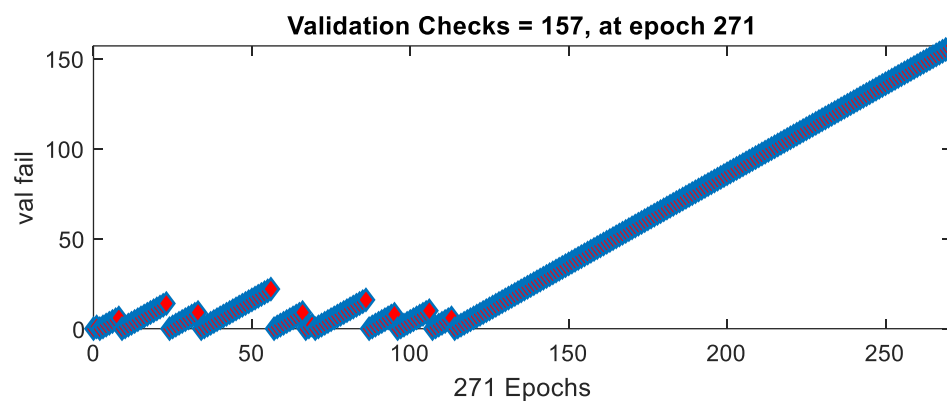
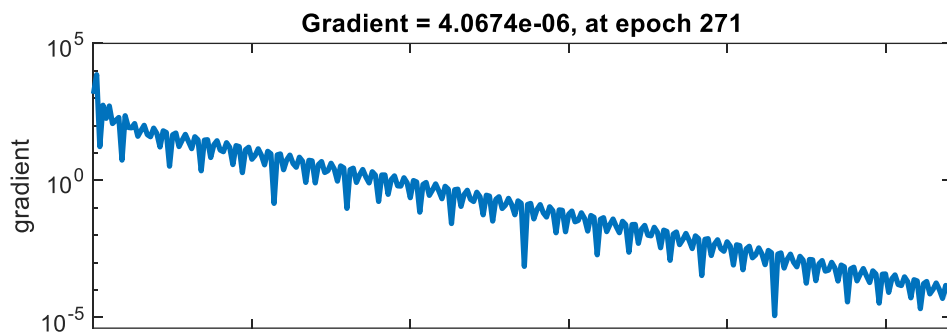
Κόμβοι 2^{ου} κρυφού επιπεδου: 20

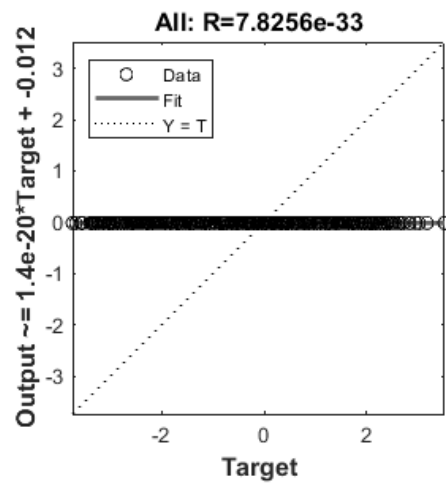
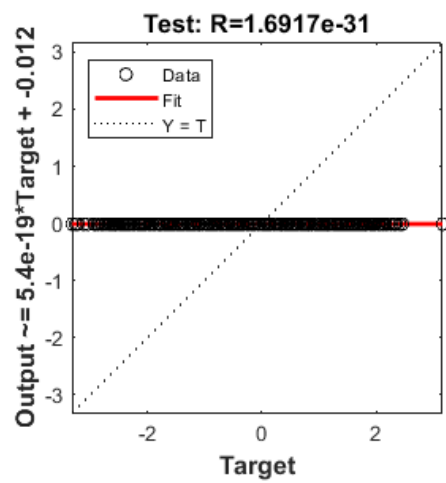
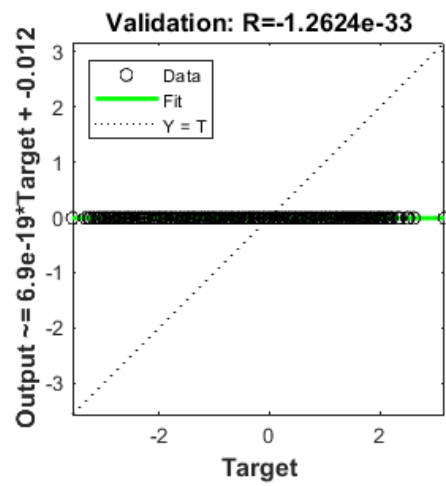
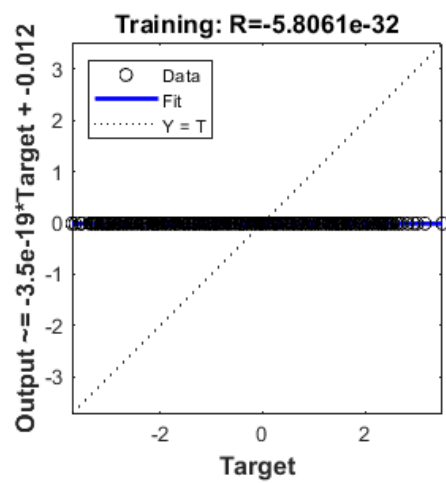
Κόμβοι 3^{ου} κρυφού επιπεδου: 30

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0348$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.8218$







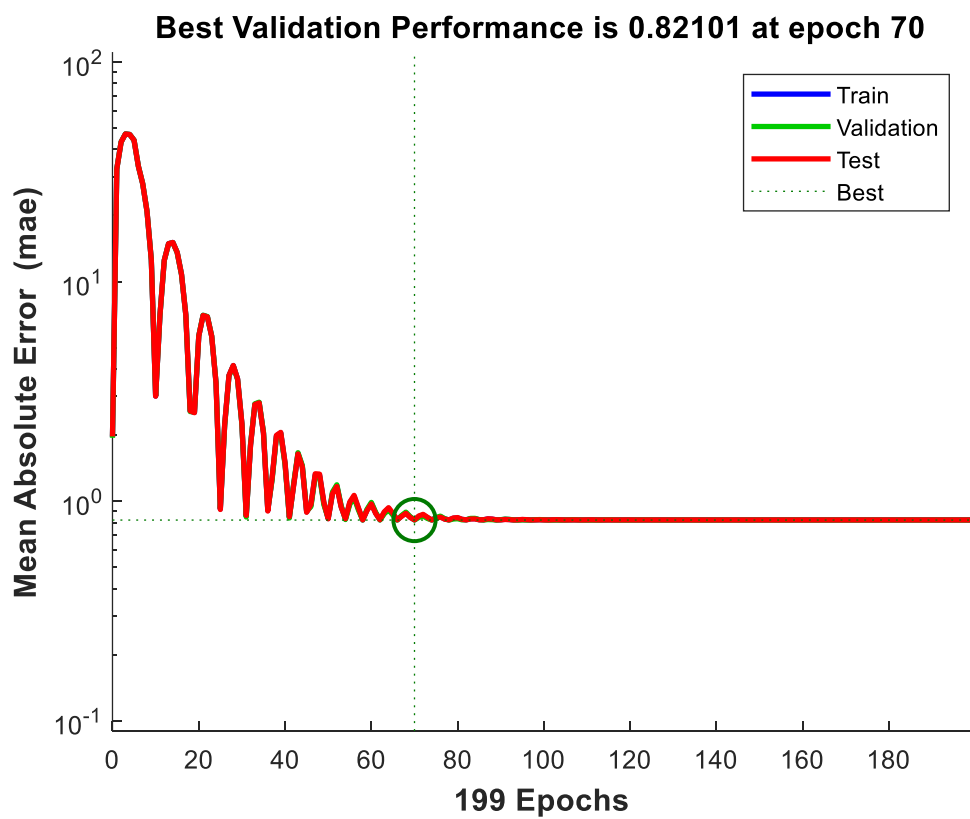
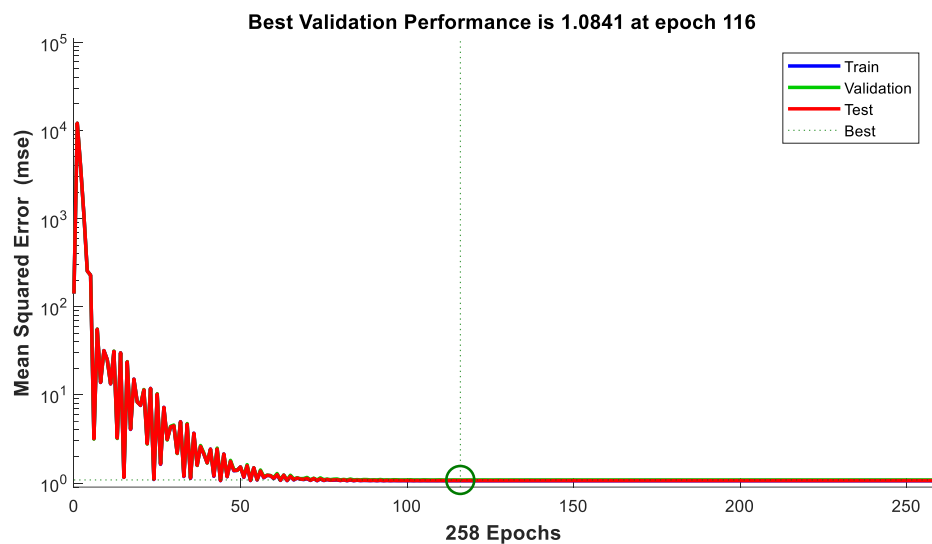
Κόμβοι 1^{ου} κρυφού επιπέδου: 20

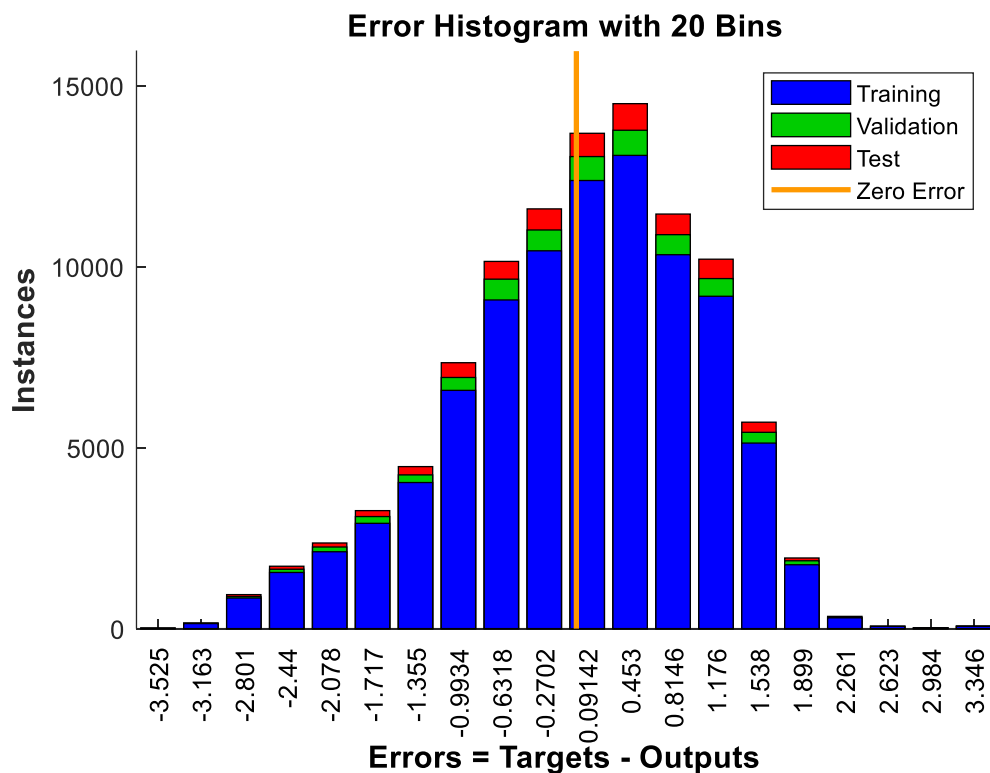
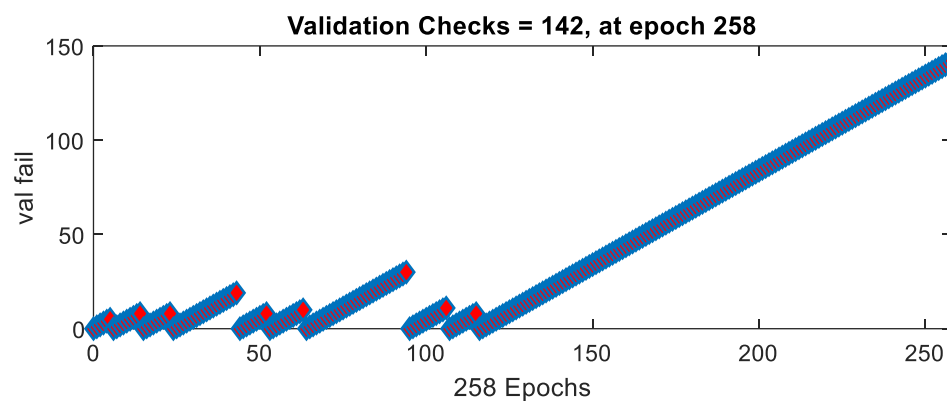
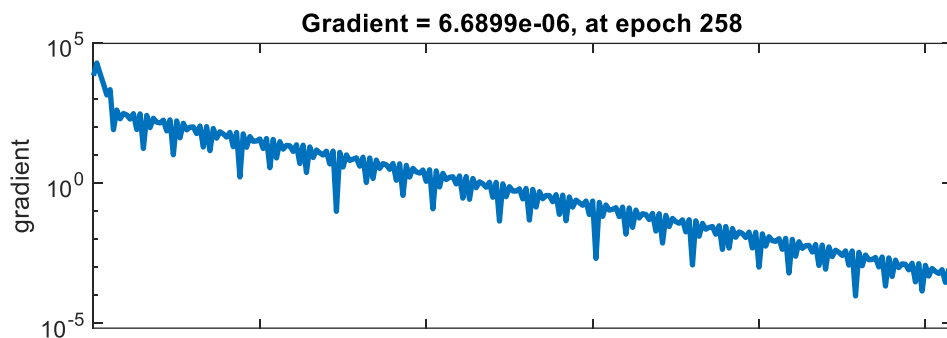
Κόμβοι 2^{ου} κρυφού επιπέδου: 40

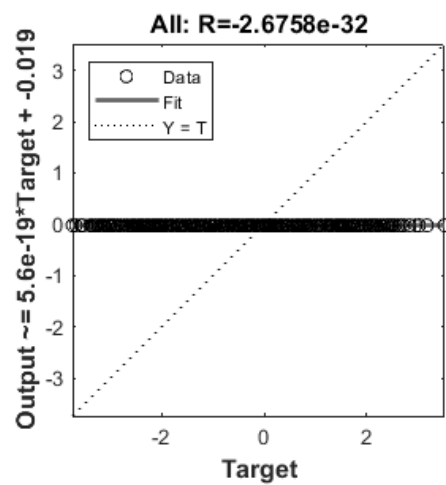
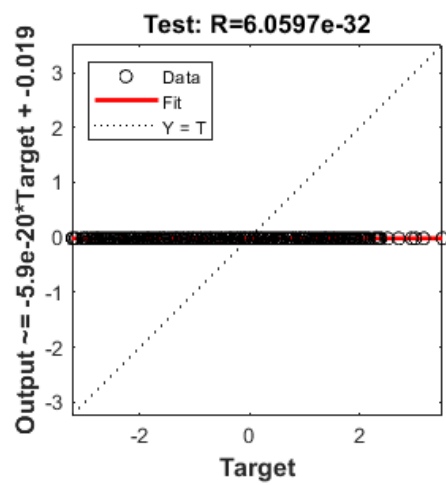
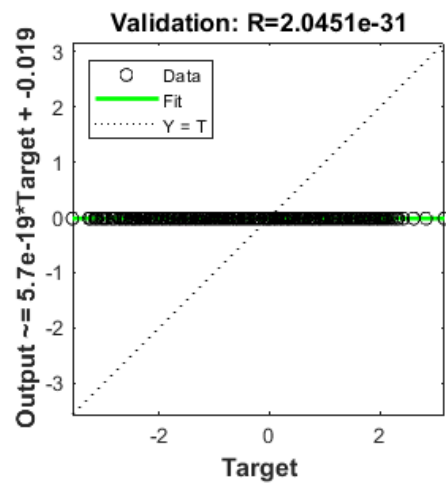
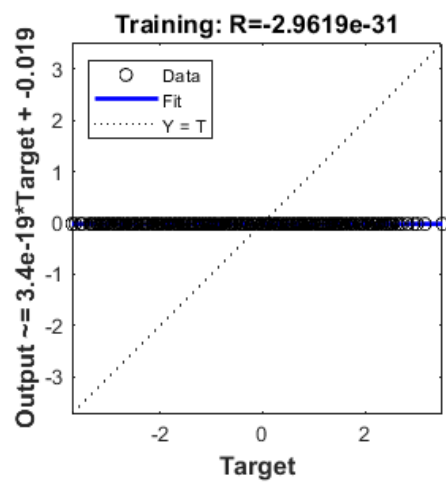
Κόμβοι 3^{ου} κρυφού επιπέδου: 60

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0841$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.821$







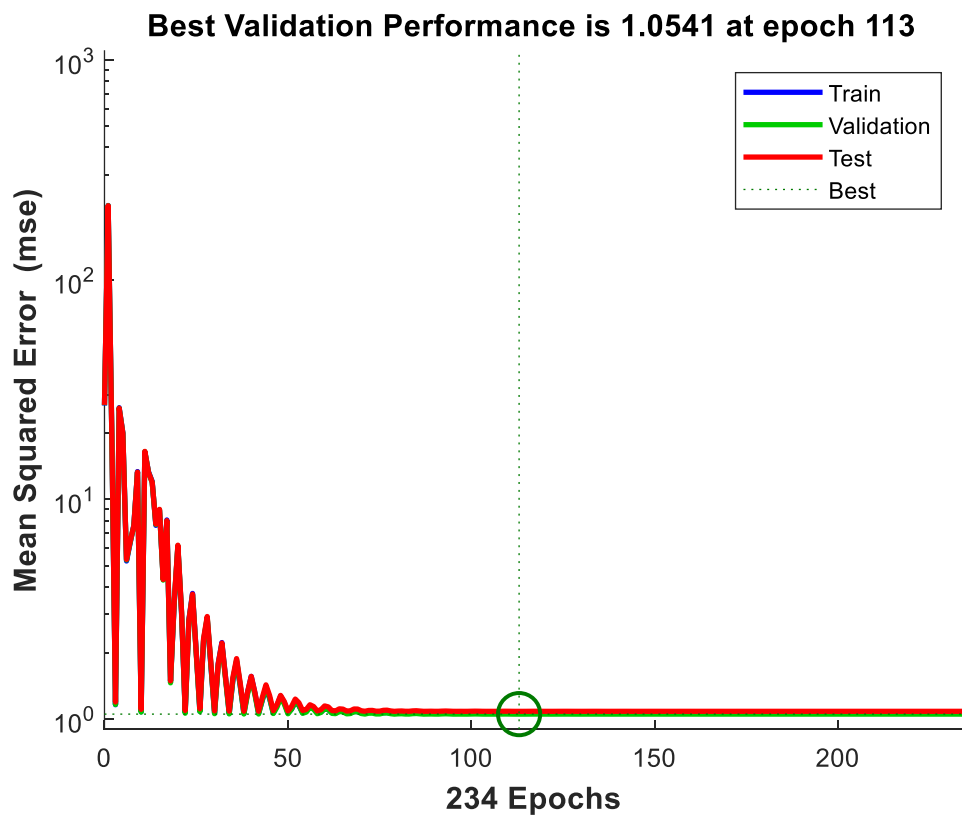
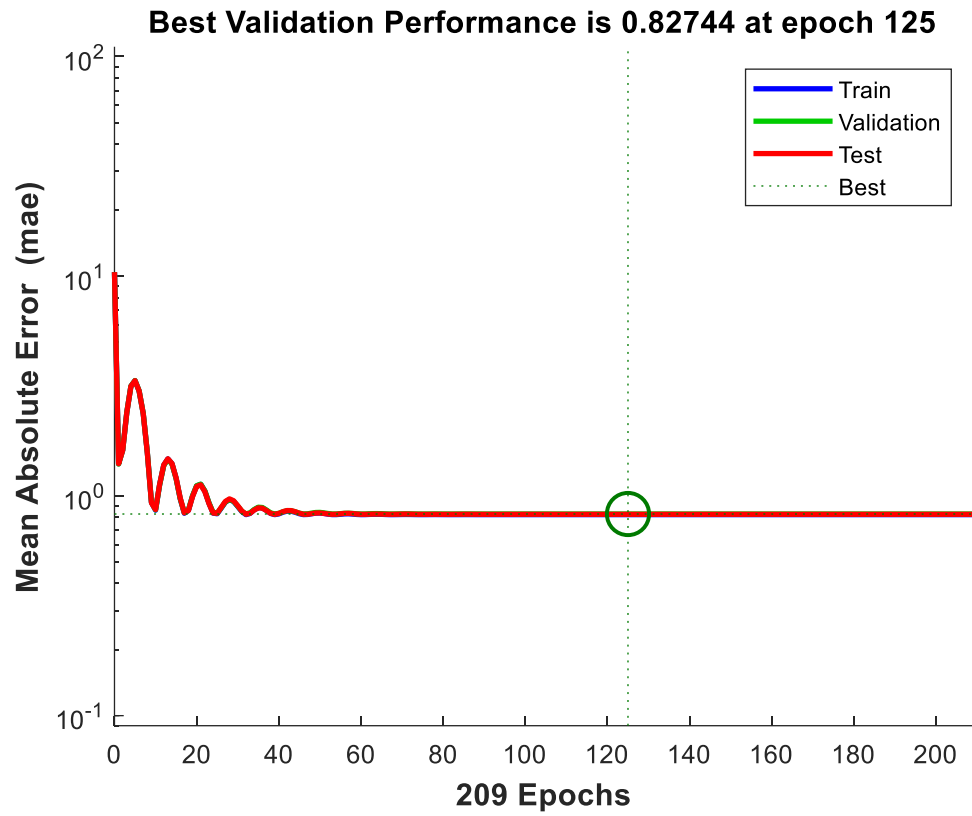
Κόμβοι 1^{ου} κρυφού επιτεδου: 20

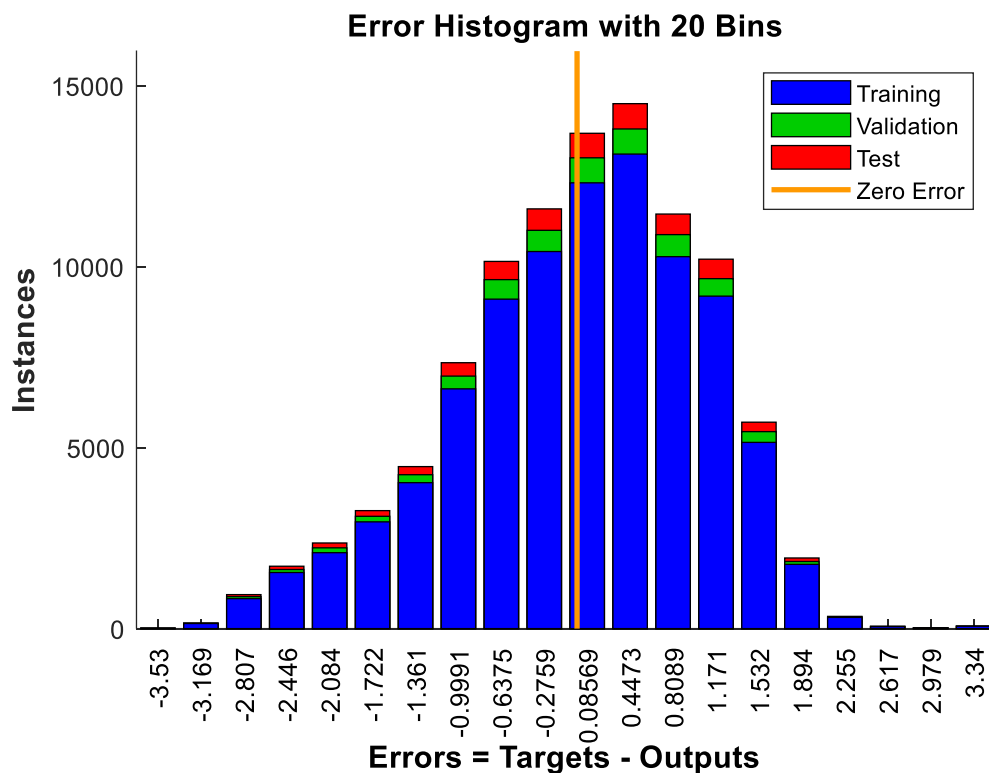
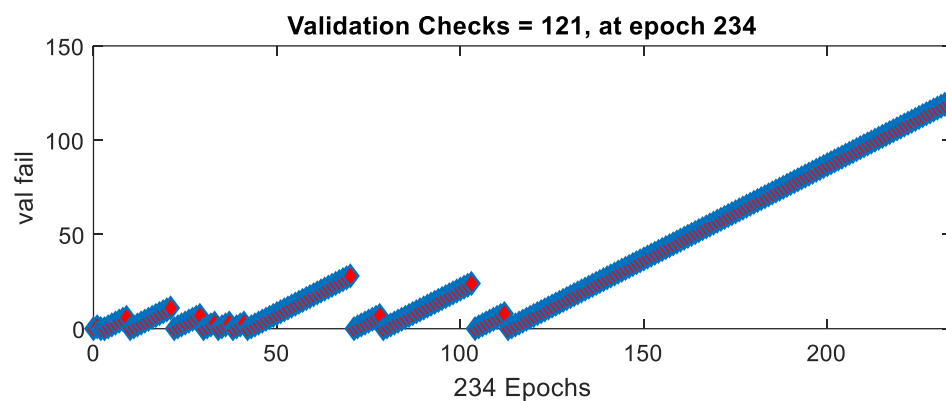
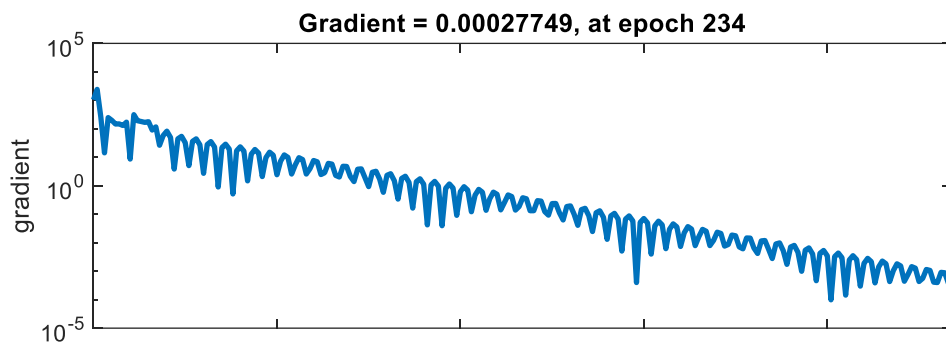
Κόμβοι 2^{ου} κρυφού επιτεδου: 20

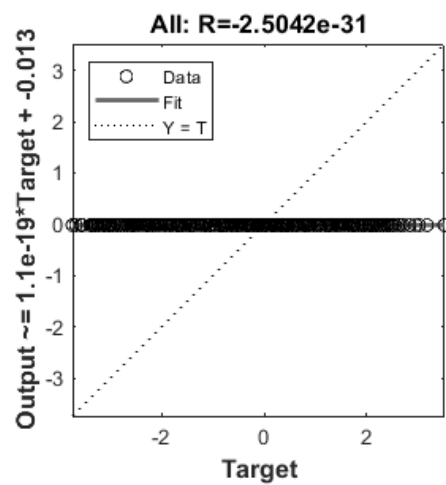
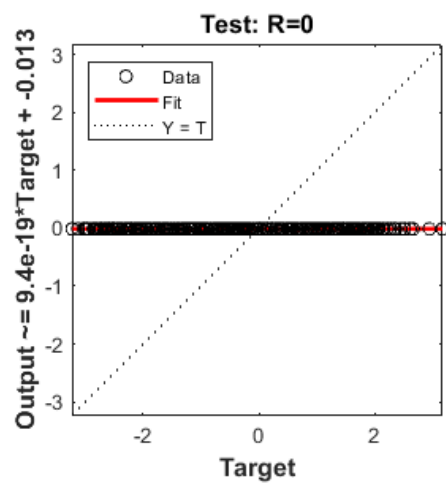
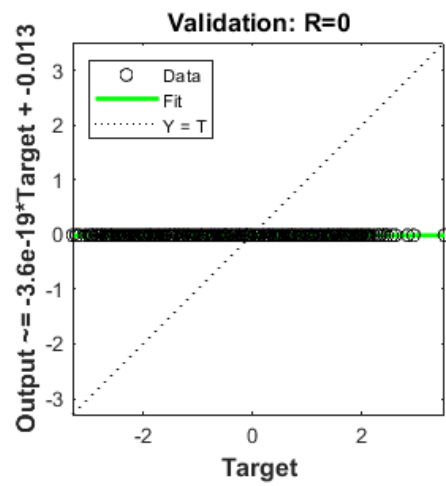
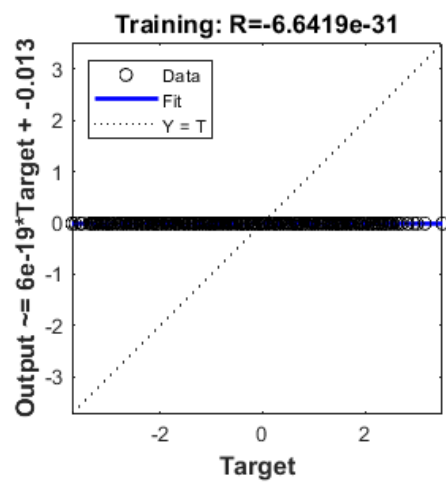
Κόμβοι 3^{ου} κρυφού επιτεδου: 20

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0266$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.8274$







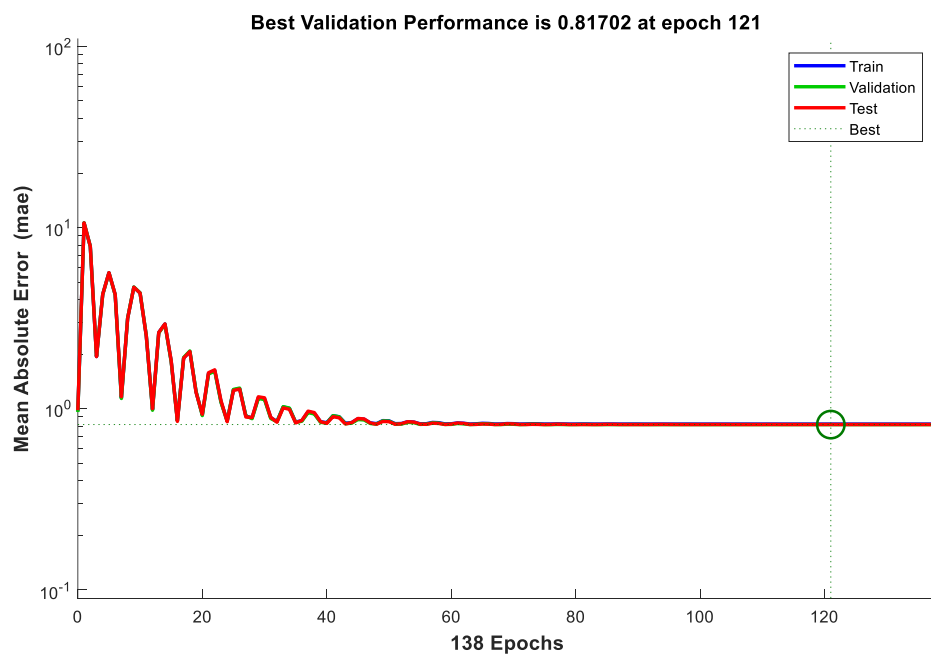
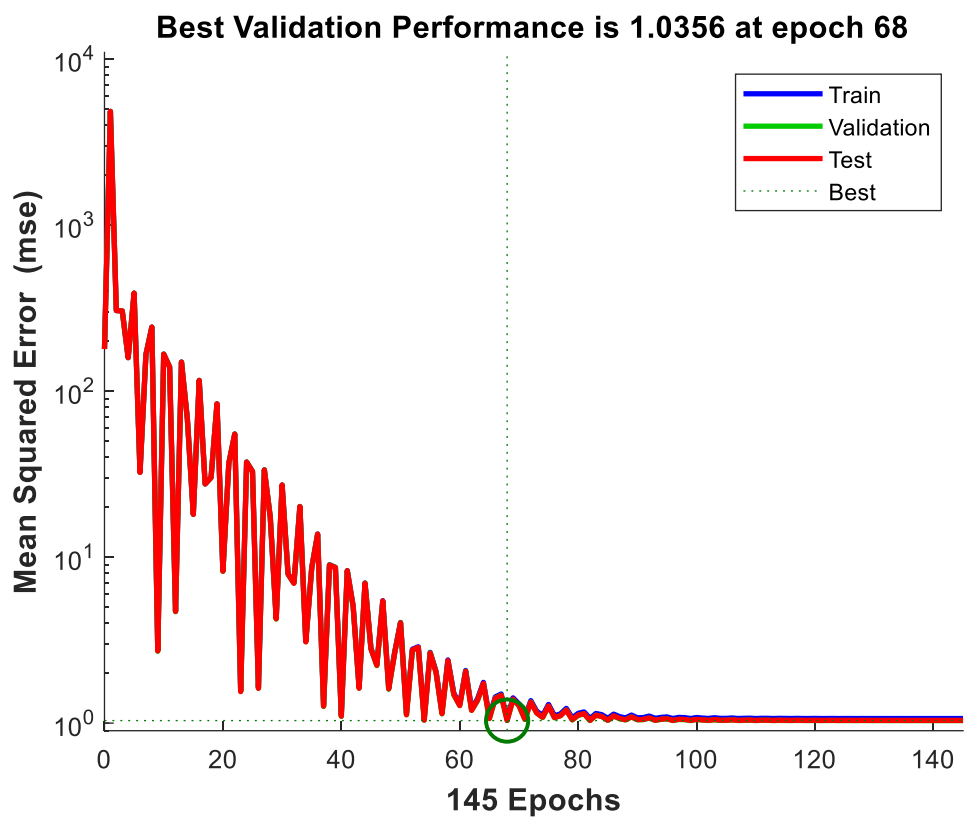
Κόμβοι 1^{ου} κρυφού επιπεδου: 40

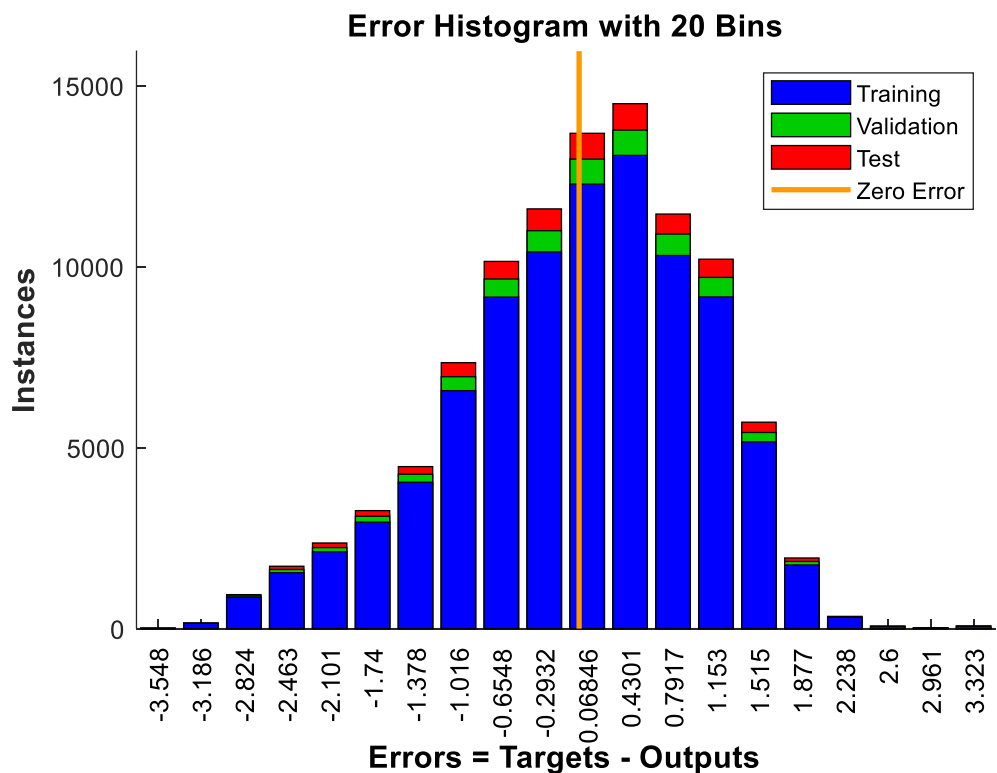
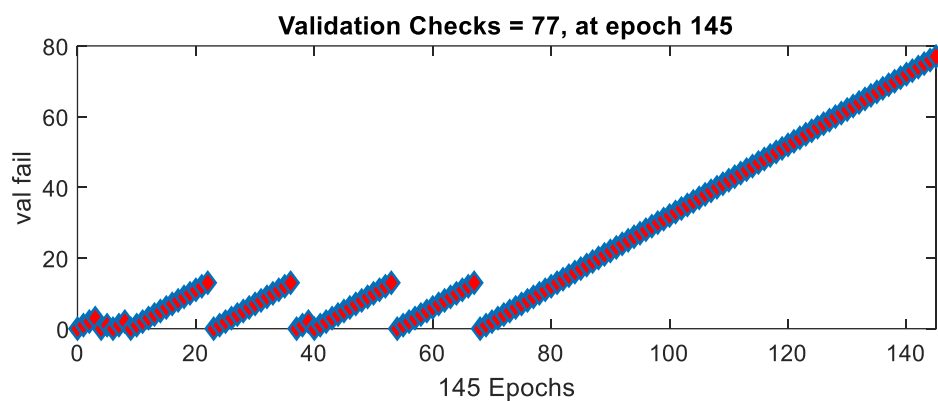
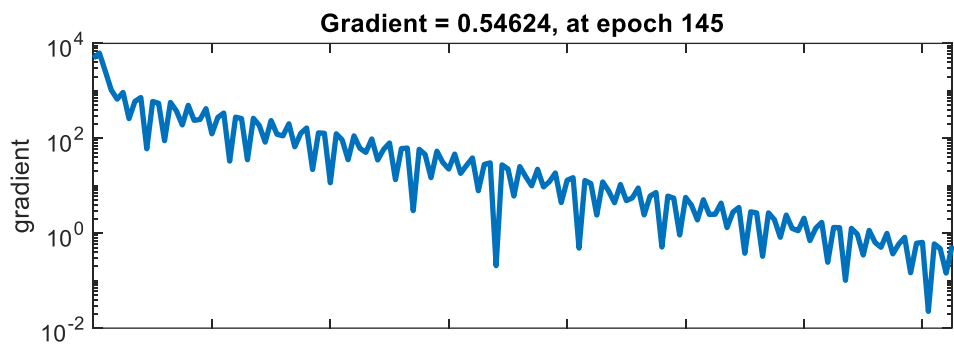
Κόμβοι 2^{ου} κρυφού επιπεδου: 40

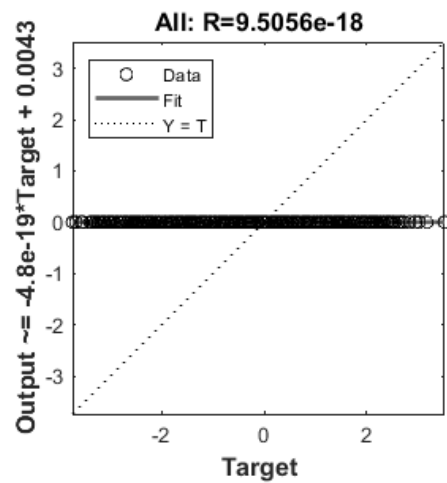
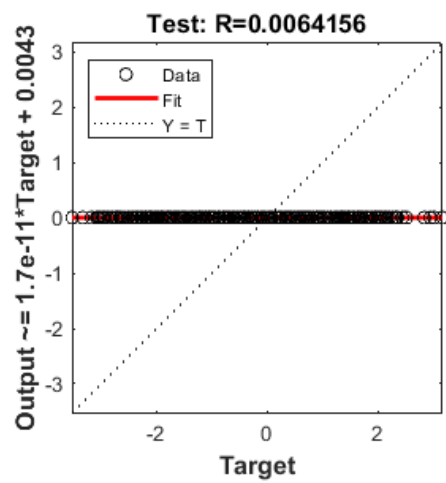
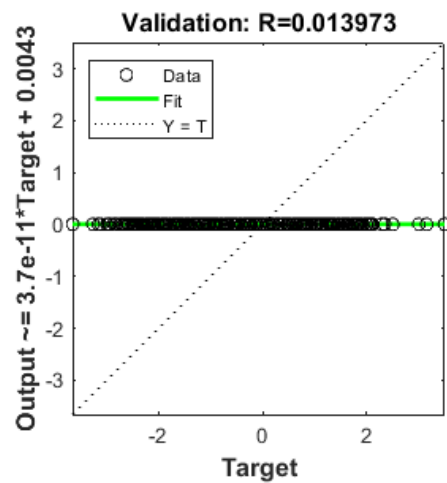
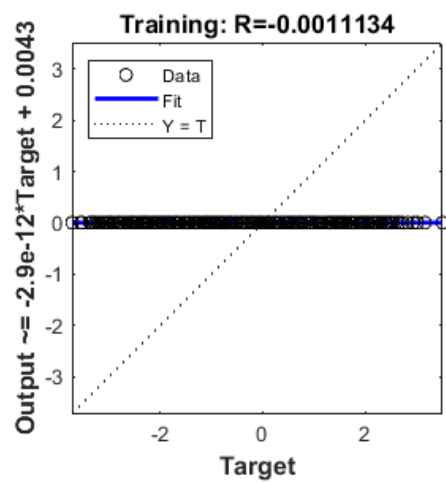
Κόμβοι 3^{ου} κρυφού επιπεδου: 40

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0176$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.817$







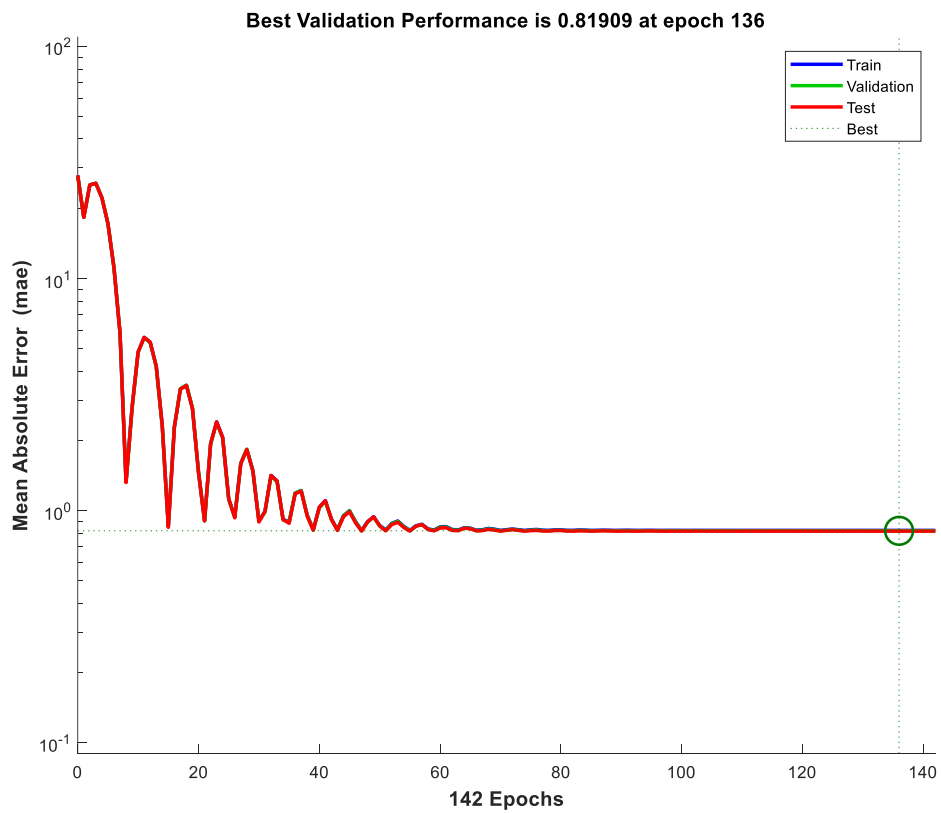
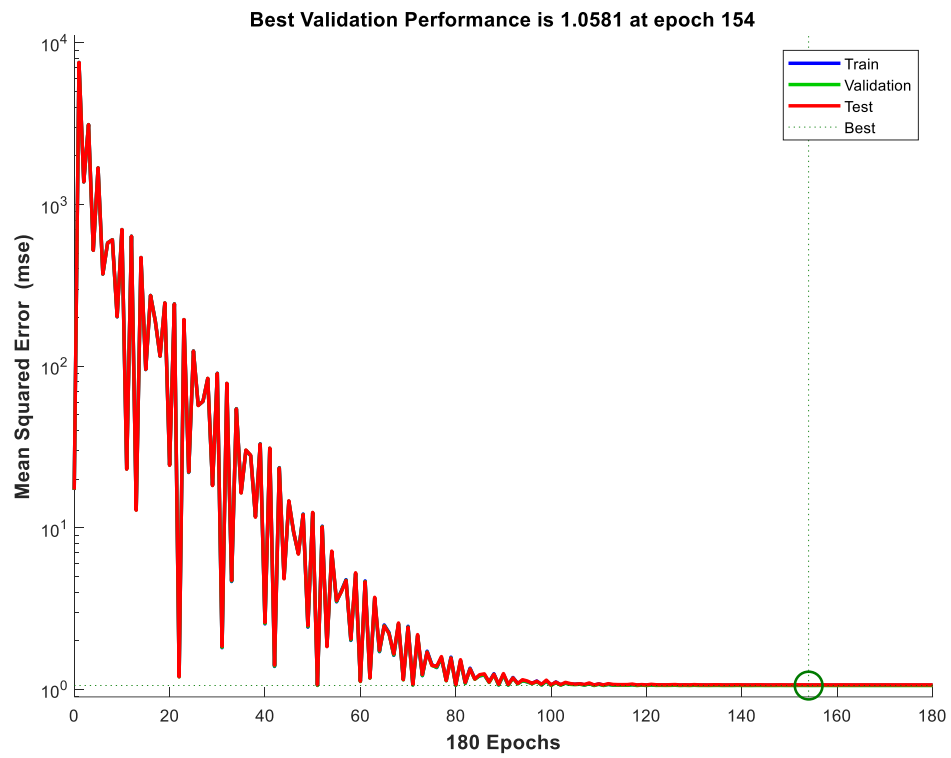
Κόμβοι 1^{ου} κρυφού επιπεδου: 60

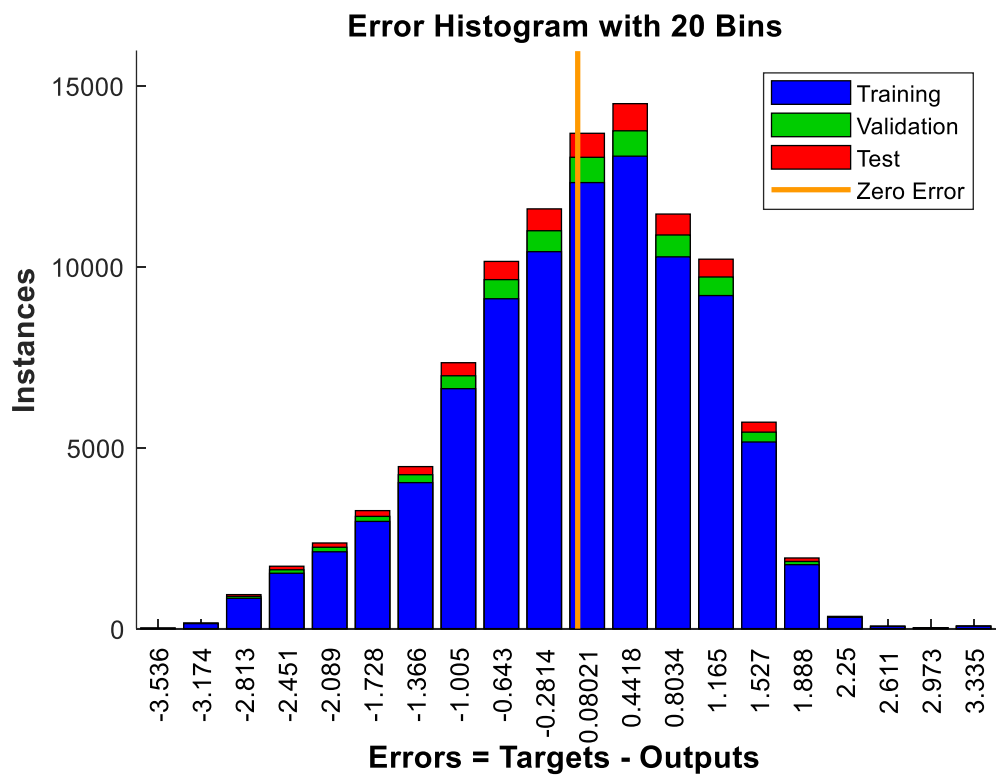
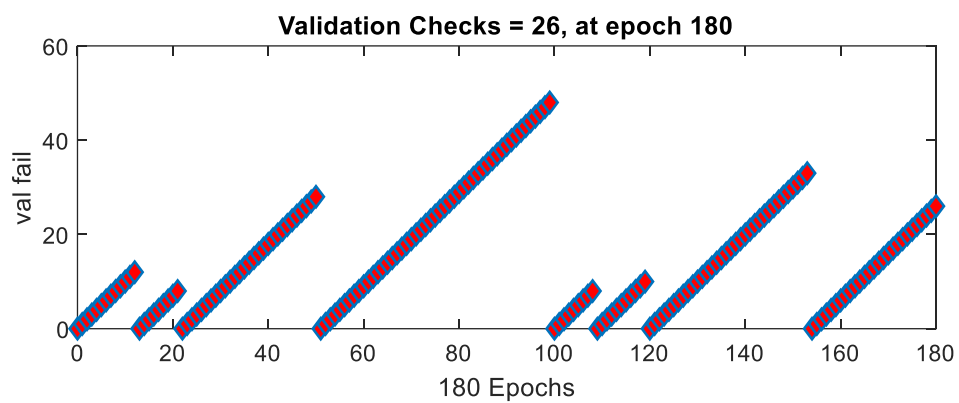
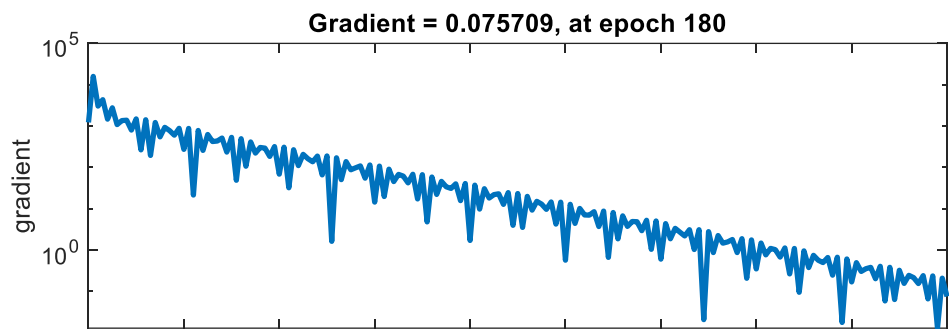
Κόμβοι 2^{ου} κρυφού επιπεδου: 60

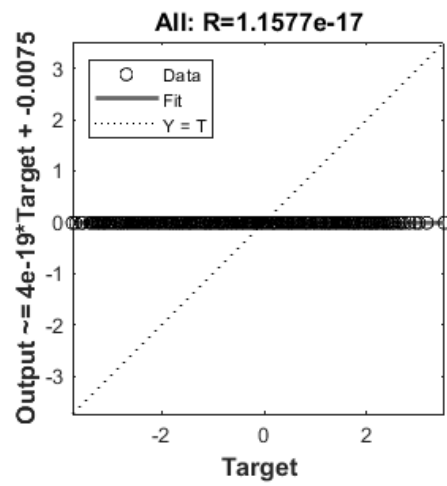
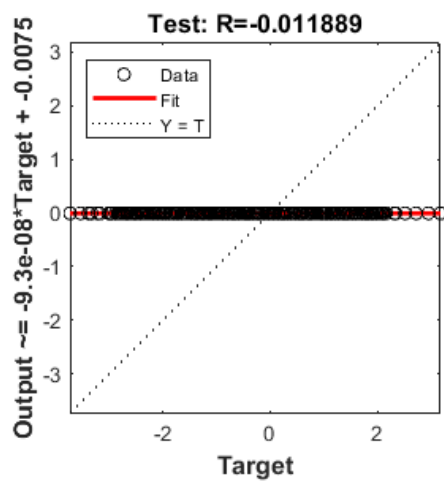
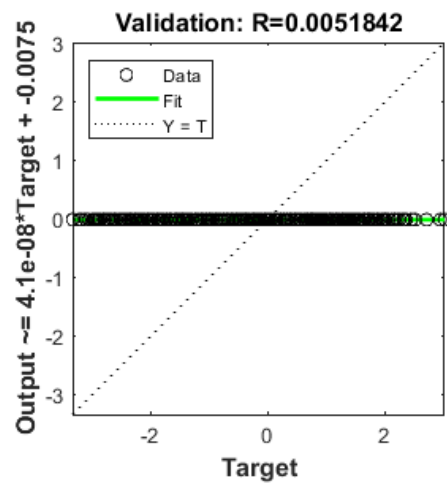
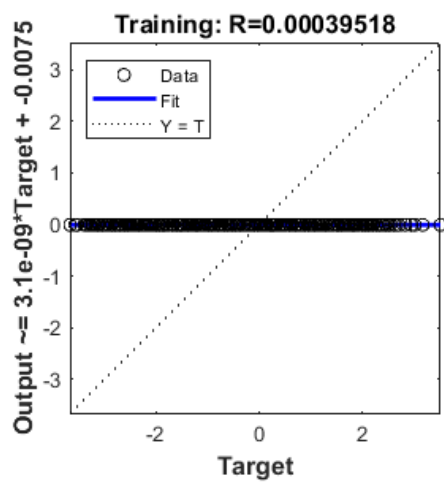
Κόμβοι 3^{ου} κρυφού επιπεδου: 60

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0286$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.819$







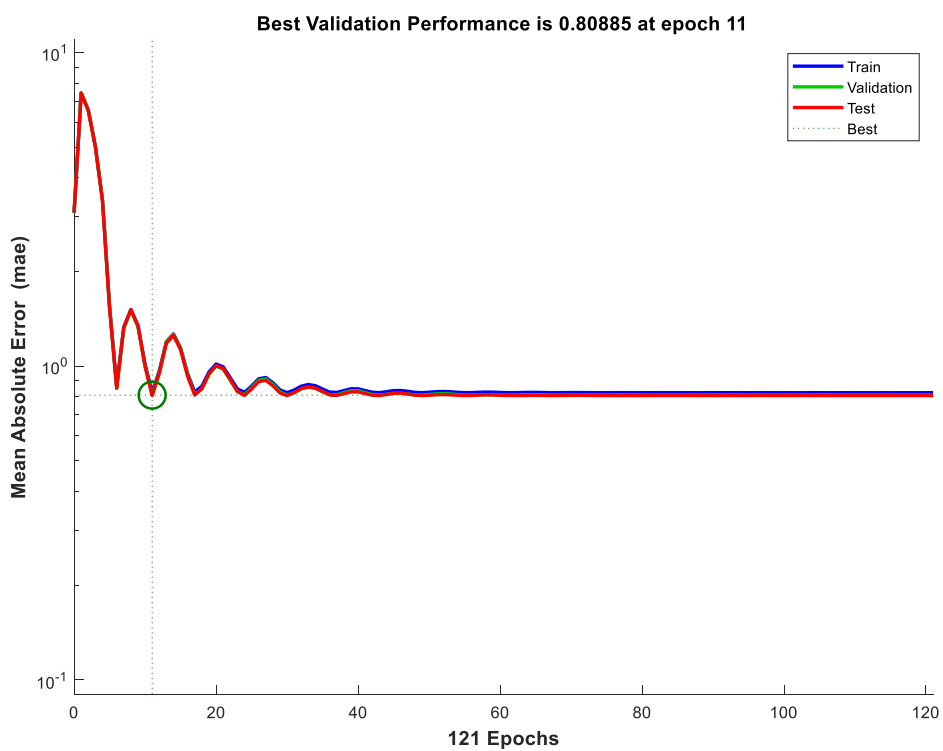
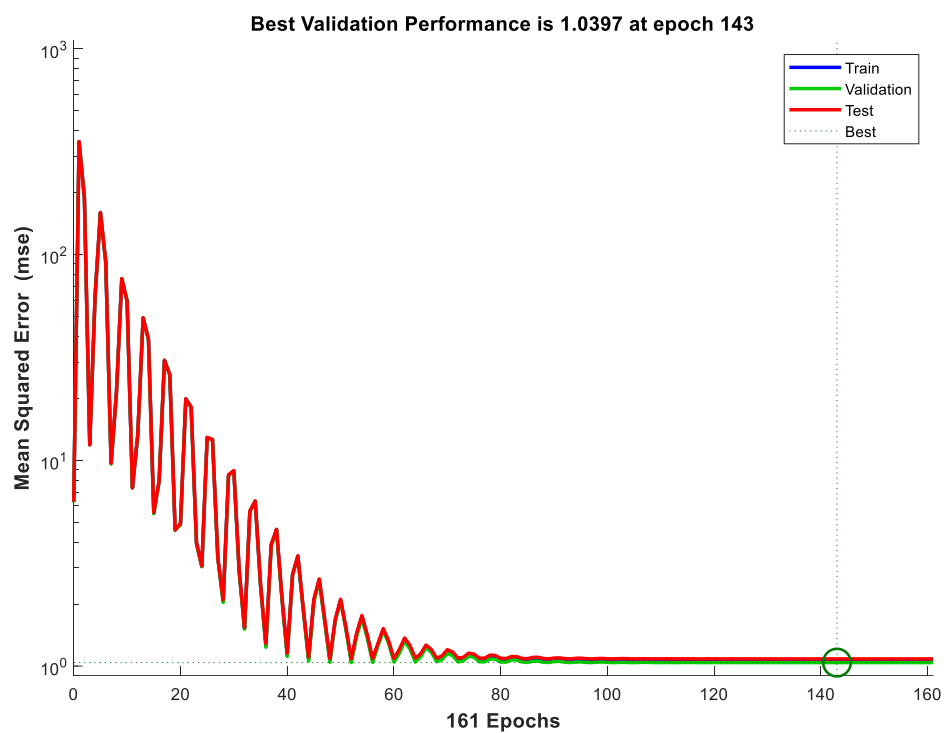
Κόμβοι 1^{ου} κρυφού επιπεδου: 60

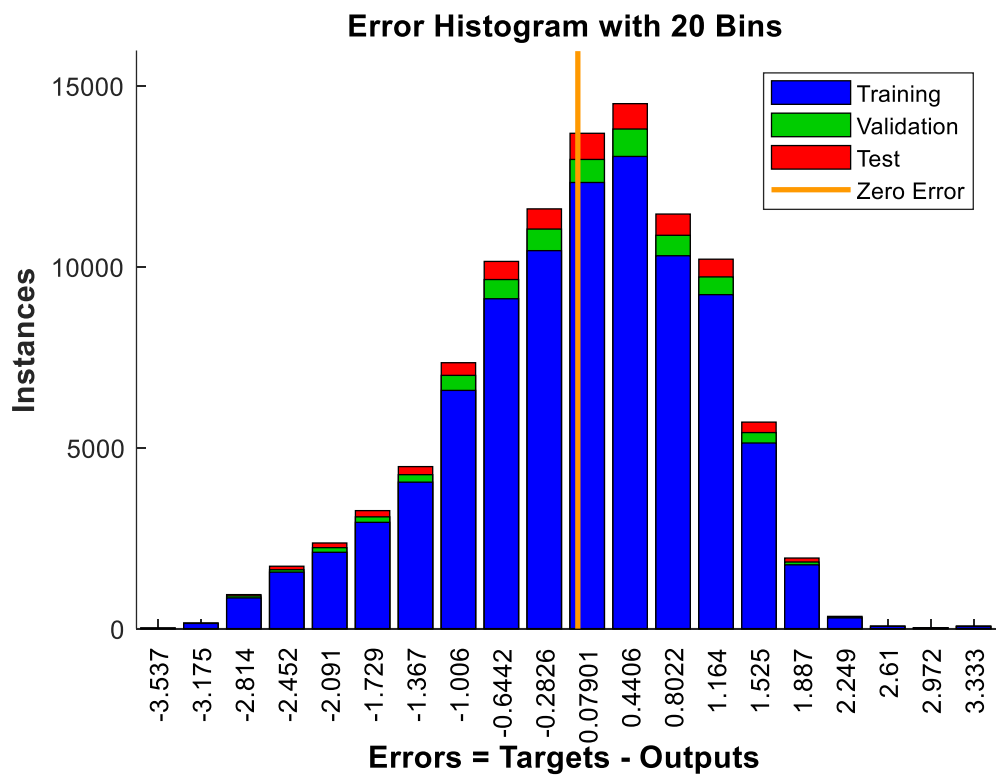
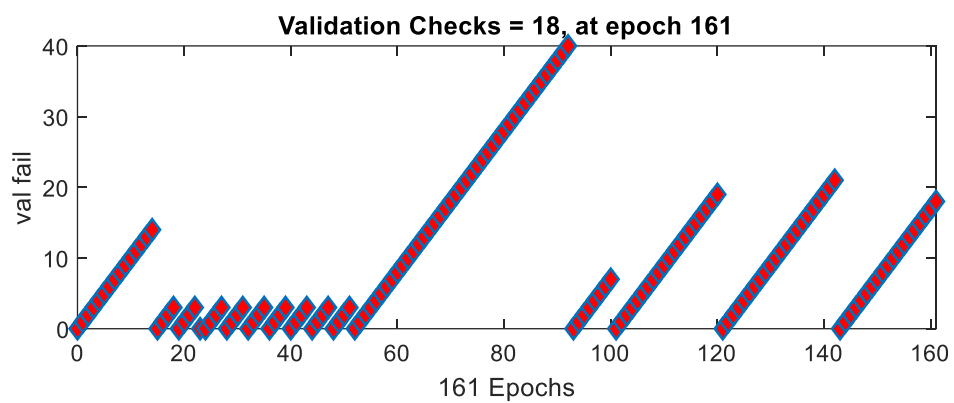
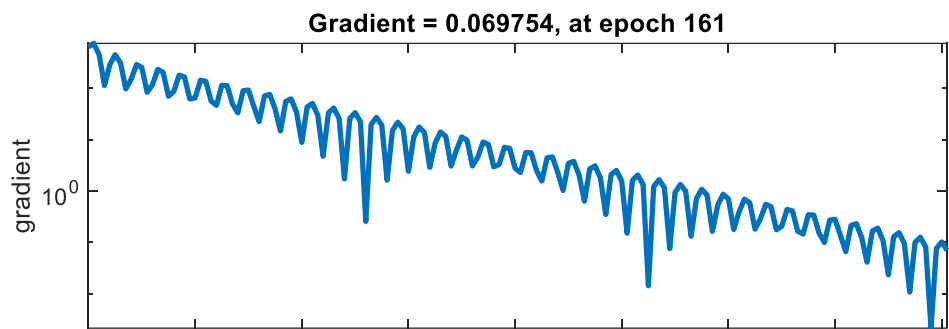
Κόμβοι 2^{ου} κρυφού επιπεδου: 40

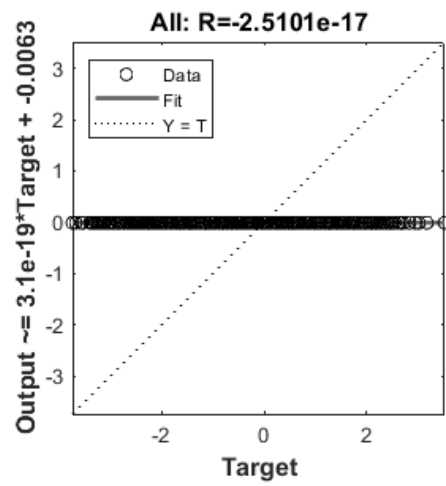
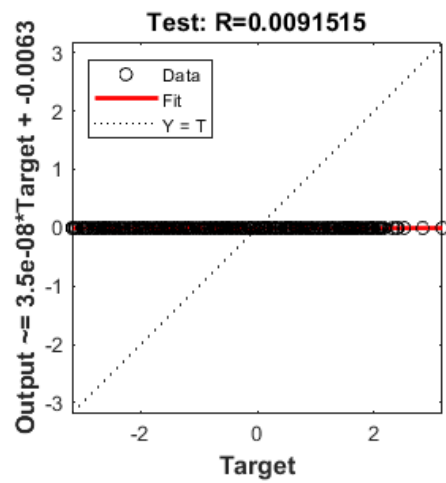
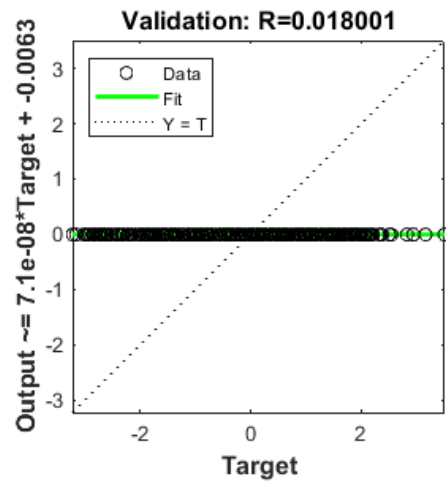
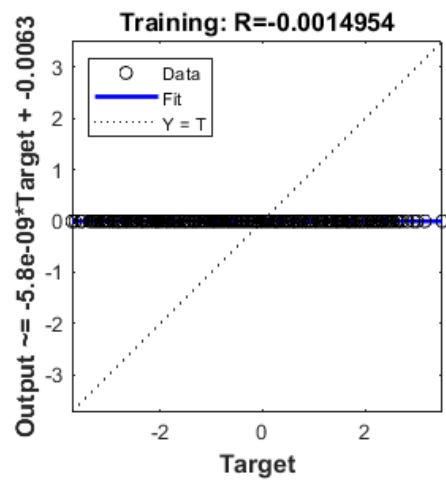
Κόμβοι 3^{ου} κρυφού επιπεδου: 20

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0196$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.808$







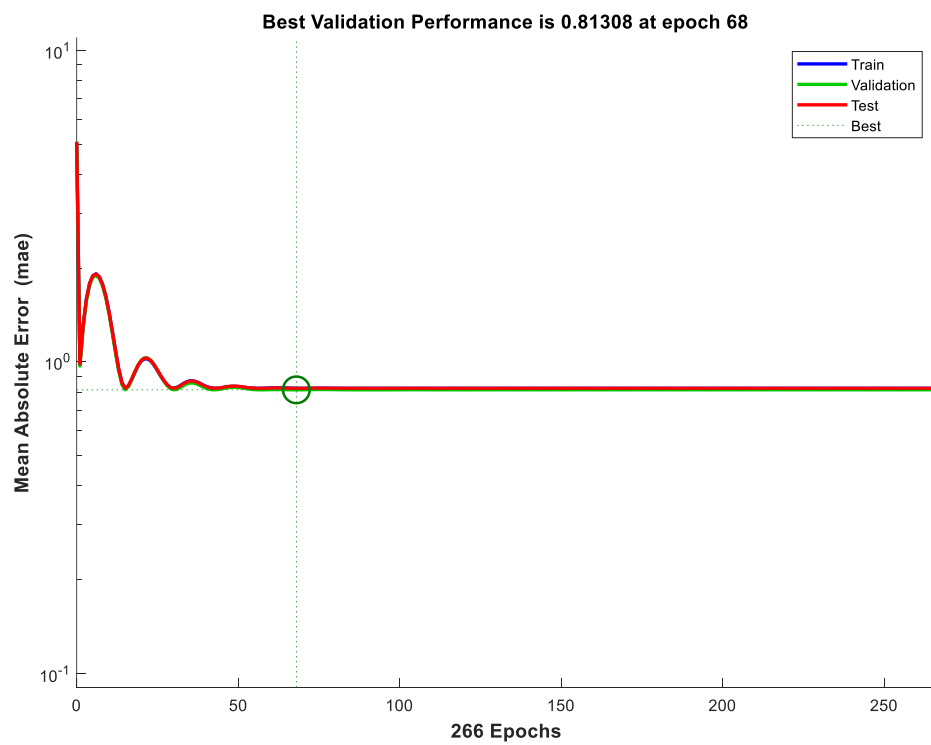
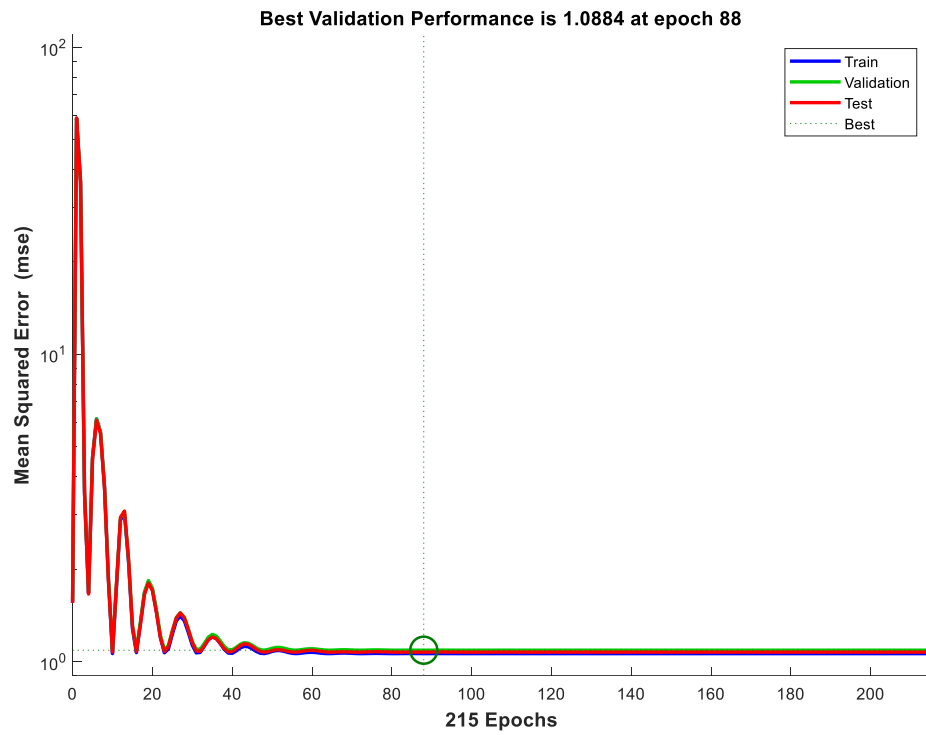
Κόμβοι 1^{ου} κρυφού επιπέδου: 30

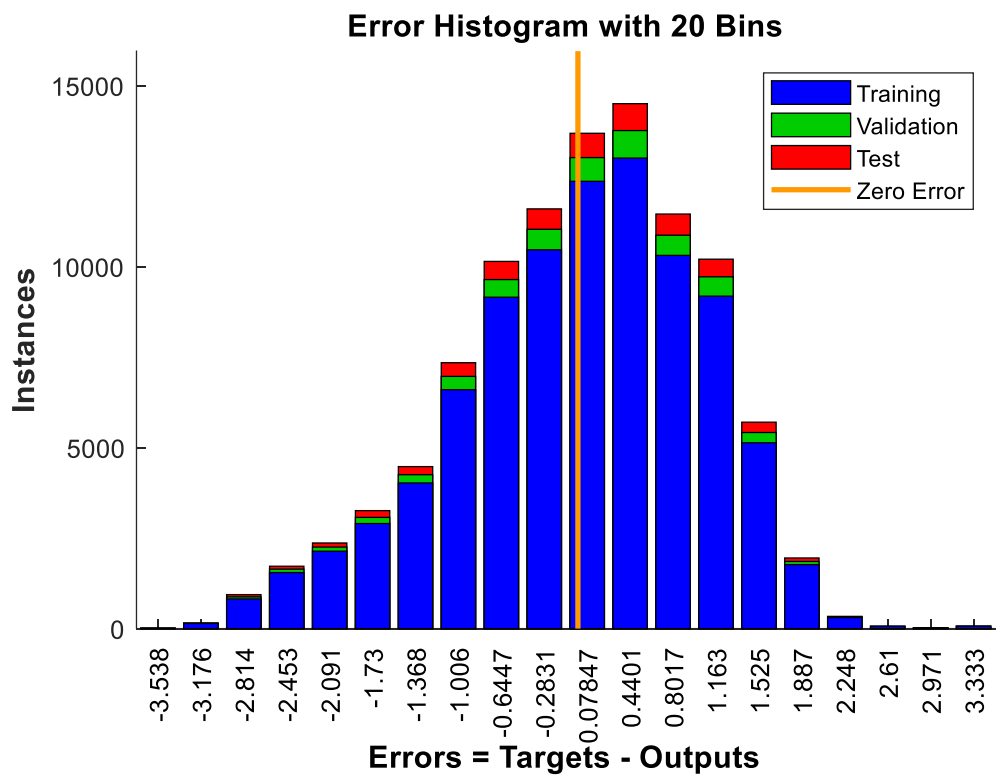
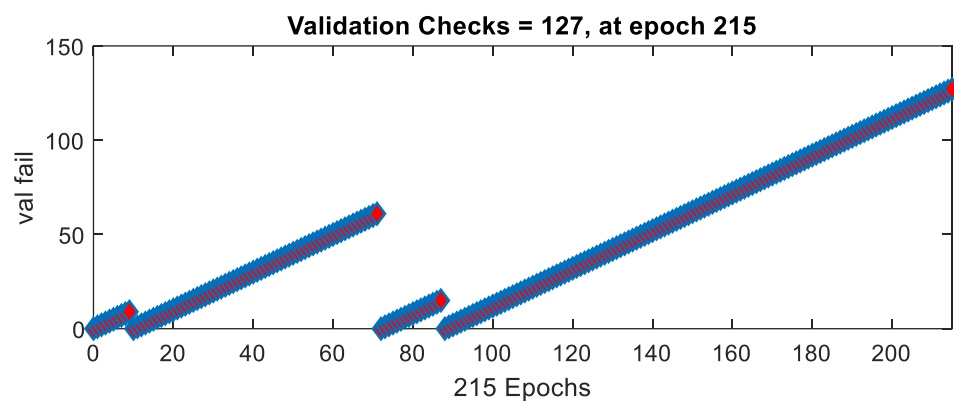
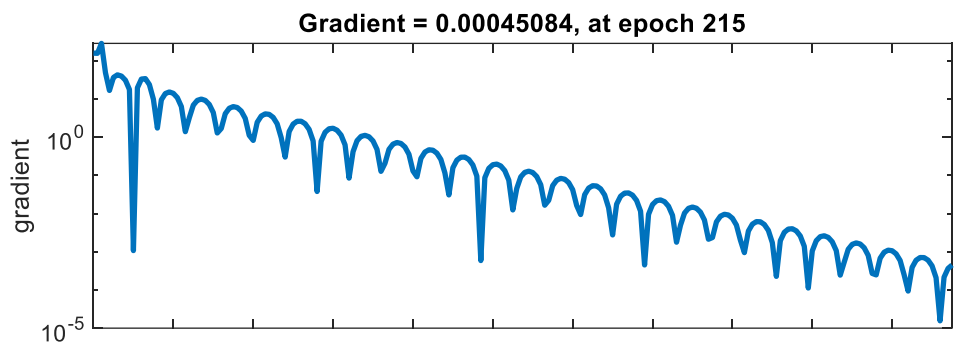
Κόμβοι 2^{ου} κρυφού επιπέδου: 15

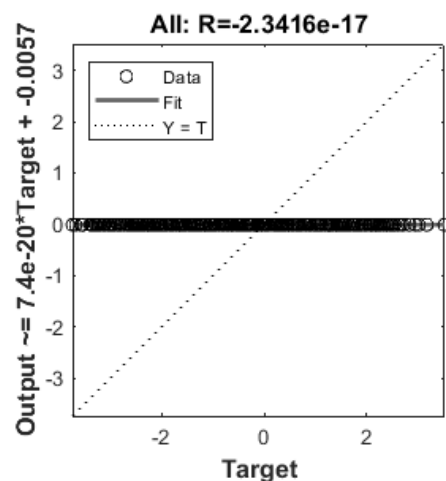
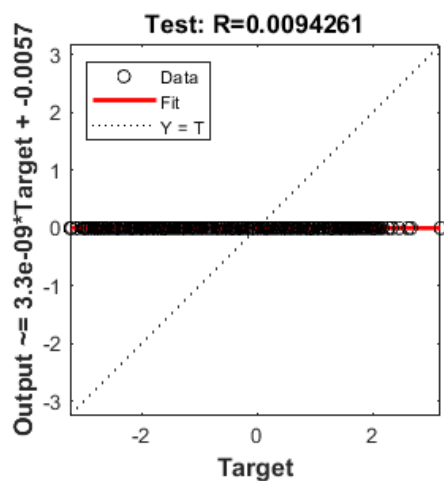
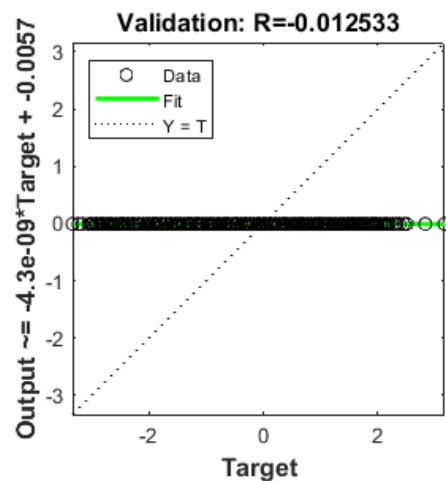
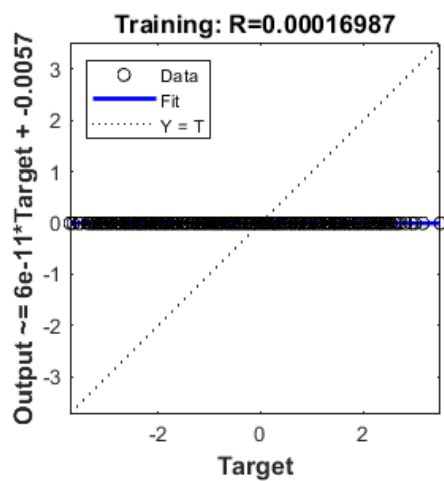
Κόμβοι 3^{ου} κρυφού επιπέδου: 5

η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) είναι $\text{rmse} = 1.0432$ και

το Μέσο Απόλυτο Σφάλμα (MAE) είναι $\text{mae} = 0.813$







Παρατηρούμε ότι την καλύτερη επίδοση σε rmse την έχει το νευρωνικό δίκτυο με 40 νευρώνες στο πρώτο κρυφό επίπεδο, 40 νευρώνες στο δεύτερο κρυφό επίπεδο και 40 νευρώνες στο τρίτο κρυφό επίπεδο.

Παρατηρούμε ότι την καλύτερη επίδοση σε mae την έχει το νευρωνικό δίκτυο με 60 νευρώνες στο πρώτο κρυφό επίπεδο, 40 νευρώνες στο δεύτερο κρυφό επίπεδο και 20 νευρώνες στο τρίτο κρυφό επίπεδο.

Επομένως, η καλύτερη τοπολογία είναι το κάθε κρυφό επίπεδο είναι να έχει τον ίδιο αριθμό νευρώνων με τα άλλα κρυφά επίπεδα αλλά ο αριθμός των νευρώνων σε κάθε κρυφό επίπεδο να μην είναι πολύ μεγάλος για να μην αυξάνεται η πολυπλοκότητα.

Εξίσου καλή είναι η τοπολογία που ο αριθμός των νευρώνων μειώνεται σε κάθε κρυφό επίπεδο.