



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

**Πρόβλεψη Συγκέντρωσης Ατμοσφαιρικών Σωματιδίων με
Χρήση Μεθόδων Μηχανικής Μάθησης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΖΑΪΜΙΩΤΗ ΗΛΙΑΝΑ

Επιβλέπων: Κουσουρής Κωνσταντίνος,

Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2020



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Πρόβλεψη Συγκέντρωσης Ατμοσφαιρικών Σωματιδίων με Χρήση Μεθόδων Μηχανικής Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΖΑΪΜΙΩΤΗ ΗΛΙΑΝΑ

Επιβλέπων: Κουσουρής Κωνσταντίνος,
Επίκουρος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 10η Μαρτίου 2020.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Κουσουρής Κωνσταντίνος
Επίκουρος Καθηγητής
Ε.Μ.Π.

.....
Παπαγιάννης Αλέξανδρος
Καθηγητής
Ε.Μ.Π.

.....
Τσιπολίτης Γεώργιος
Καθηγητής
Ε.Μ.Π.

Αθήνα, Μάρτιος 2020

(Υπογραφή)

.....

Ζαϊμώτη Ηλιάννα

Διπλωματούχος Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών
Ε.Μ.Π.

© 2020 – All rights reserved



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Copyright © - All rights reserved Ηλιάνα Ζαϊμιώτη, 2020.

Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα

Περίληψη

Στόχος της παρούσας διπλωματικής εργασίας είναι η ανάδειξη της Μηχανικής Μάθησης, η οποία αποτελεί ένα διεπιστημονικό πεδίο που περιλαμβάνει τη θεωρία πιθανοτήτων και υπολογιστικής πολυπλοκότητας, τη θεωρία προσεγγίσεων, καθώς επίσης και πολλούς άλλους τομείς της επιστήμης των υπολογιστών. Η Μηχανική Μάθηση είναι ένας από τους πιο ραγδαία αναπτυσσόμενους κλάδους της Πληροφορικής και με την πληθώρα διαθέσιμων όγκων δεδομένων οι δυνατότητες είναι απεριόριστες.

Για τη διπλωματική αυτή εργασία, δεδομένα που αφορούν μετρήσεις συγκέντρωσης ατμοσφαιρικών σωματιδίων αξιοποιούνται για την εφαρμογή μοντέλων Μηχανικής Μάθησης. Πιο συγκεκριμένα, κατασκευάζονται γραμμικά και μη γραμμικά μοντέλα παλινδρόμησης με μεθόδους ενδυνάμωσης για δέντρα απόφασης, καθώς επίσης και νευρωνικά δίκτυα.

Το σύνολο των δεδομένων που χρησιμοποιείται συνδυάζει μετρήσεις για τη συγκέντρωση του αιωρούμενου σωματιδίου $PM_{2.5}$, αλλά και μετεωρολογικές μετρήσεις για ένα χρονικό διάστημα 5 ετών (2010 – 2014) που αφορούν την πόλη του Πεκίνου. Τα δεδομένα αυτά επεξεργάζονται κατάλληλα ώστε να μπορούν να αξιοποιηθούν για τη μελέτη. Για κάθε ένα από τα μοντέλα που δημιουργούνται αναλύεται η διαδικασία επιλογής των παραμέτρων και ακολουθούν τεχνικές αξιολόγησης των προβλεπόμενων αποτελεσμάτων σύμφωνα με τις πραγματικές παρατηρήσεις με σκοπό την εύρεση εκείνου με την καλύτερη απόδοση και τα λιγότερα σφάλματα.

Λέξεις κλειδιά

Μηχανική Μάθηση, Αιωρούμενο Σωματίδιο, Παλινδρόμηση, Πρόβλεψη, Νευρωνικά Δίκτυα, Μέθοδοι Ενδυνάμωσης, Δέντρα απόφασης, Ενδυναμωμένα Δέντρα Απόφασης

Abstract

This diploma thesis aims to highlight the use of Machine Learning, which is an interdisciplinary field that includes probability theory and computational complexity, theory of approaches, as well as many other areas of computer science. Machine Learning is one of the fastest growing branches of IT and with the abundance of available data volumes the possibilities are unlimited.

In this thesis, data related to atmospheric particle concentration measurements are utilized for the application of Machine Learning models. More specifically, linear and non-linear regression models are constructed with amplification methods for decision trees, as well as neural networks.

The dataset used combines measurements for the concentration of the particulate matter PM_{2.5}, but also meteorological measurements for a period of 5 years (2010 - 2014) over Beijing. This data is processed appropriately so that it can be used for the study. For each of the models that are created, the process of selecting the parameters is analyzed and techniques of evaluation of the predicted results compared to the real observations are used in order to find the one with the best performance and the least errors.

Keywords

Machine Learning, Particulate Matter, Regression, Prediction, Neural Networks, Boosting Methods, Decision Trees, Boosted Decision Trees

Ευχαριστίες

Η εκπόνηση της συγκεκριμένης διπλωματικής εργασίας πραγματοποιήθηκε υπό την επίβλεψη του καθηγητή κ. Κουσουρή, τον οποίο θα ήθελα να ευχαριστήσω θερμά τόσο για την ανάθεση της εργασίας αυτής, όσο και για την καθοδήγηση, την υποστήριξη και την εμπιστοσύνη που μου έδειξε. Θα ήθελα επίσης να ευχαριστήσω και τους υπόλοιπους καθηγητές της σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών για τη συμβολή τους στην εξέλιξη των γνώσεων μου στον τομέα των Εφαρμοσμένων Μαθηματικών.

Τέλος, θα ήθελα να εκφράσω την ευγνωμοσύνη μου στην οικογένειά μου για την αμέριστη υποστήριξη που μου προσφέρει και ήταν πάντα δίπλα μου σε όλη τη διάρκεια των σπουδών μου, καθώς επίσης και στους φίλους μου που με στηρίζουν καθημερινά.

Περιεχόμενα

Περίληψη	6
Abstract	7
Ευχαριστίες	8
1 Το Πρόβλημα	12
1.1 Εισαγωγή στο Πρόβλημα.....	13
1.2 Περιγραφή του Προβλήματος.....	16
1.3 Περιγραφή των Δεδομένων	16
1.4 Επεξεργασία των Δεδομένων	18
1.5 Περιγραφική Στατιστική των Δεδομένων.....	28
2 Εισαγωγή στη Μηχανική Μάθηση	31
2.1 Εισαγωγή στη Μηχανική Μάθηση	32
2.2 Εφαρμογές Μηχανικής Μάθησης	32
2.3 Είδη Μηχανικής Μάθησης	33
2.3.1 Εποπτευόμενη Μάθηση (Supervised Learning)	33
2.3.2 Μη Εποπτευόμενη Μάθηση (Unsupervised Learning)	34
2.4 Υπερεκπαίδευση και Ανεπαρκής Εκπαίδευση	34
3 Γραμμική Παλινδρόμηση	37
3.1 Εισαγωγή στη Γραμμική Παλινδρόμηση.....	38
3.1.1 Απλό Γραμμικό Μοντέλο	38
3.1.2 Πολλαπλό Γραμμικό Μοντέλο	40
3.2 Μέθοδος Ελαχίστων Τετραγώνων	41
4 Δέντρα Απόφασης	43
4.1 Εισαγωγή στα Δέντρα Απόφασης.....	44
4.2 Δέντρα Απόφασης για Προβλήματα Παλινδρόμησης	45
4.3 Μέθοδος Bootstrap	47
4.4 Μέθοδος Bagging	48
4.5 Μέθοδοι Boosting - «Ενίσχυσης».....	48
4.5.1 AdaBoost.....	50
4.5.2 Gradient Boosting	53
5 Νευρωνικά Δίκτυα	56
5.1 Εισαγωγή στα Νευρωνικά Δίκτυα	57
5.1.1 Εφαρμογές των Νευρωνικών Δικτύων	58
5.2 Αρχιτεκτονική Νευρωνικού Δικτύου.....	58

5.3 Perceptron	59
5.4 Συνάρτηση Ενεργοποίησης.....	60
5.5 Feedforward Propagation – Εμπροσθοτροφοδοτούμενο Νευρωνικό Δίκτυο...	61
5.6 Backpropagation	64
5.7 Διαμόρφωση Υπερπαραμέτρων Νευρωνικού Δικτύου	65
6 Εφαρμογή Αλγορίθμων	68
6.1 Εισαγωγή των Δεδομένων στο Μοντέλο	69
6.2 Μέτρα Αξιολόγησης του Μοντέλου	70
6.3 Γραμμική Παλινδρόμηση	72
6.4 Δέντρα Απόφασης.....	79
6.4.1 AdaBoost.....	79
6.4.2 Gradient Boosting	85
6.5 Νευρωνικά Δίκτυα	100
6.5.1 Κατασκευή Νευρωνικών Δικτύων με 1 Κρυφό Στρώμα.....	102
6.5.2 Κατασκευή Νευρωνικών Δικτύων με 2 Κρυφά Στρώματα	105
6.5.3 Κατασκευή Νευρωνικών Δικτύων με 3 Κρυφά Στρώματα	110
6.5.4 Σύγκριση Νευρωνικών Δικτύων για το Διαφορετικό Αριθμό των Κρυφών Στρωμάτων.....	116
6.6 Σύγκριση Μοντέλων	121
Βιβλιογραφία	126
Παράρτημα	128

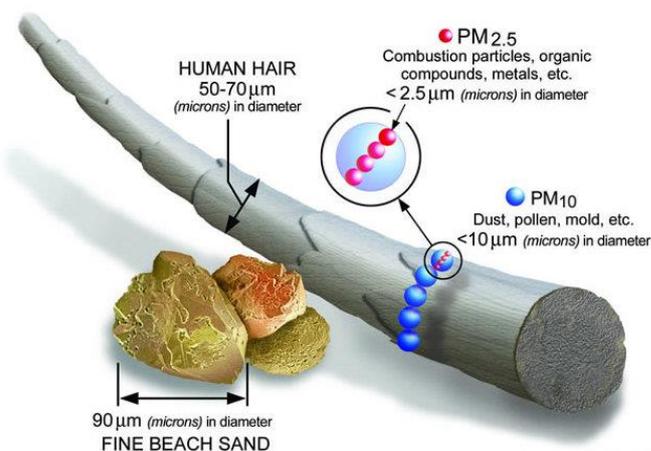
1 Το Πρόβλημα

1.1 Εισαγωγή στο Πρόβλημα

Η ατμοσφαιρική μόλυνση αποτελεί ένα θέμα καθολικού ενδιαφέροντος. Ειδικότερα, τα τελευταία χρόνια με τη ραγδαία αστική ανάπτυξη και τον εκσυγχρονισμό αυτής, η ρύπανση του αέρα επιδεινώνεται και ο αντίκτυπος στην ανθρώπινη υγεία είναι πλέον φλέγον ερευνητικό θέμα. Από τους κύριους ατμοσφαιρικούς ρύπους που απασχολούν τις ανά τον κόσμο υπηρεσίες προστασίας του περιβάλλοντος είναι τα αιωρούμενα σωματίδια – κυρίως τα PM_{10} και $PM_{2,5}$ (PM : Particulate Matter).

Τα τελευταία αποτελούν ένα μείγμα από στερεά σωματίδια και σταγονίδια που βρίσκονται στον αέρα, ενώ μπορεί να παραχθούν είτε από μία συγκεκριμένη πηγή απευθείας, είτε να σχηματίσουν πολύπλοκες χημικές ενώσεις στην ατμόσφαιρα. Τα αιωρούμενα σωματίδια εκκρίνονται κατά την ανάφλεξη στερεών και υγρών καυσίμων, όπως κατά την παραγωγή ηλεκτρικής ενέργειας, οικιακής θέρμανσης και στους κινητήρες οχημάτων. Τα σωματίδια αυτά ποικίλουν σε μέγεθος, δηλαδή διαφέρουν σε διάμετρο και εύρος, με αποτέλεσμα να κατηγοριοποιούνται με βάση την αεροδυναμική τους διάμετρο. Έρευνες έχουν δείξει ότι αιωρούμενα σωματίδια με αεροδυναμική διάμετρο μικρότερη από $10\ \mu\text{m}$ ($1\ \mu\text{m} = 10^{-6}\ \text{m}$) έχουν μεγαλύτερη επίπτωση στην ανθρώπινη υγεία [7].

Μία κατηγορία αιωρούμενων σωματιδίων, τα σωματίδια $PM_{2,5}$, έχουν πολύ μικρή διάμετρο, ωστόσο καλύπτουν μεγάλες εκτάσεις και μπορεί να μεταφέρουν τοξική ύλη, η οποία διασχίζοντας το ρινικό φίλτρο μέχρι το τέλος της αναπνευστικής οδού, μεταδίδεται μέσω της ανταλλαγής αέρα στους πνεύμονες με αποτέλεσμα να φθείρει άλλα σημεία του ανθρώπινου σώματος. Συγκεκριμένα, η ονομασία $PM_{2,5}$ οφείλεται στη διάμετρο των σωματιδίων, η οποία είναι μικρότερη από $2,5\ \mu\text{m}$ και είναι περίπου όσο το 3% της διαμέτρου μίας ανθρώπινης τρίχας μαλλιών. Συνεπώς, μπορούν να γίνουν αντιληπτά μόνο με χρήση ενός ηλεκτρονικού μικροσκοπίου.



Εικόνα 1.1 : Σύγκριση της διαμέτρου ενός αιωρούμενου σωματιδίου $PM_{2,5}$ με αυτή μίας ανθρώπινης τρίχας και ενός

Επιπτώσεις των PM_{2.5} στην υγεία

Λόγω του εξαιρετικά μικρού μεγέθους των αιωρούμενων σωματιδίων PM_{2.5}, αυτά τείνουν να παραμένουν στον αέρα για περισσότερο διάστημα από ότι άλλα πιο βαριά σωματίδια. Το γεγονός αυτό αυξάνει τις πιθανότητες εισπνευσής τους από ανθρώπινους και ζωικούς οργανισμούς, το οποίο μπορεί να οδηγήσει σε δυσμενείς επιπτώσεις για την υγεία τους. Μελέτες έχουν δείξει μια στενή σχέση μεταξύ της έκθεσης σε σωματίδια PM_{2.5} και της εμφάνισης σοβαρών ασθενειών, όπως πνευμονία και χρόνια βρογχίτιδα, καρδιαγγειακές παθήσεις όπως στεφανιαία νόσο και συμφορητική καρδιακή ανεπάρκεια, καθώς και πρόωρο θάνατο. Ειδικότερα, παιδιά, ηλικιωμένοι και όσοι έχουν προδιάθεση για αναπνευστική και καρδιαγγειακή νόσο φαίνεται να είναι πιο ευαίσθητοι στις επιπτώσεις που προκαλούνται από την ατμοσφαιρική ρύπανση στην υγεία [8].

Health category	24-hour PM _{2.5} µg/m ³
Low	0–8.9
Moderate	9.0–25.9
Unhealthy – sensitive	26.0–39.9
Unhealthy – all	40.0–106.9
Very unhealthy – all	107.0–177.9
Hazardous (high)	Greater than 177.9
Hazardous (extreme)	Greater than 250

Εικόνα 1.2 : Κατηγορίες κατάστασης υγείας ανάλογα με τη συγκέντρωση του σωματιδίου PM_{2.5} στον αέρα.

Επιπρόσθετα, με βάση μία σχετικά πρόσφατη μελέτη στην Αμερική, αποδείχθηκε ότι η μακροχρόνια έκθεση στα ατμοσφαιρικά σωματίδια PM_{2.5} συνδέεται με τον αυξημένο κίνδυνο θνησιμότητας των ενηλίκων. Συγκεκριμένα, εκτιμήθηκε ότι περίπου 1 εκατομμύριο πρόωροι θάνατοι ετησίως οφείλονται στην ατμοσφαιρική ρύπανση που προκαλείται από μικρά σωματίδια [9].

Air quality category	24-hr PM _{2.5} μg/m ³	One-hour PM _{2.5} μg/m ³
Very good	0–8.2	0–13.1
Good	8.3–16.4	13.2–26.3
Fair	16.5–25.0	26.4–39.9
Poor	25.1–37.4	40–59.9
Very poor	37.5 or greater	60 or greater

Εικόνα 1.2 : Κατηγορίες ποιότητας αέρα ανάλογα με τη συγκέντρωση του σωματιδίου PM_{2.5} στην ατμόσφαιρα.

Ατμοσφαιρική ρύπανση στο Πεκίνο

Λόγω της ραγδαίας οικονομικής ανάπτυξης, της εκβιομηχάνισης και της αστικοποίησης, η Κίνα πλήττεται από την ατμοσφαιρική ρύπανση, με αποτέλεσμα την αύξηση των κινδύνων για τον εκτεθειμένο πληθυσμό της. Ανάμεσα σε 61 πόλεις με ετήσια συγκέντρωση του σωματιδίου PM_{2.5} πάνω από 80 μg/m³ – αυτή η συγκέντρωση αντιστοιχεί σε πολύ κακή ποιότητα αέρα και ανθυγιεινά επίπεδα για τους οργανισμούς, σύμφωνα με την *Εικόνα 1.2* και *Εικόνα 1.3* – 10 από αυτές βρίσκονται στην Κίνα και συγκεκριμένα, περισσότεροι από 1.3 εκατομμύρια πρόωροι θάνατοι ενηλίκων αποδόθηκαν στην έκθεση αυτών στο αιωρούμενο σωματίδιο PM_{2.5} το 2013 [9].

Το Πεκίνο, με πληθυσμό 21,7 εκατομμύρια ανθρώπους (στοιχεία μετρήσεων το 2015) χαρακτηρίζεται ως μία από τις πιο ρυπαρές πόλεις της χώρας, ενώ συχνά απαντάται και ένα γκρι στρώμα ομίχλης πάνω από αυτή, προσελκύνοντας την παγκόσμια προσοχή εξαιτίας των υψηλών επιπέδων συγκέντρωσης ρύπων.



Εικόνα 1.3: Νέφος πάνω από την πόλη του Πεκίνο λόγω της υψηλής συγκέντρωσης ατμοσφαιρικών ρύπων.

Έρευνες για την ατμοσφαιρική ρύπανση στην οποία συμβάλλει το αιωρούμενο σωματίδιο $PM_{2.5}$ που διεξήχθησαν τον Ιανουάριο του 2013 έδειξαν ότι οι μετεωρολογικές συνθήκες, τα δευτερογενή αερολύματα και τα μέσα μετακίνησης συντελούν στη διαμόρφωση και την ανάπτυξη του σωματιδίου $PM_{2.5}$ στο Πεκίνο. Επιπλέον, από διερευνητικές αναλύσεις βρέθηκε ότι τις υψηλές συγκεντρώσεις του $PM_{2.5}$ επηρεάζουν και οι ανώμαλες συνθήκες ανέμου και υγρασίας.

1.2 Περιγραφή του Προβλήματος

Με τη χρήση δεδομένων που συνδυάζουν μετρήσεις για τη συγκέντρωση του αιωρούμενου σωματιδίου $PM_{2.5}$, αλλά και μετεωρολογικές για ένα χρονικό διάστημα 5 ετών (2010 – 2014) που αφορούν την πόλη του Πεκίνου, εφαρμόζονται διάφορα μοντέλα Μηχανικής Μάθησης για την εκτίμηση των τιμών συγκέντρωσης του σωματιδίου στην ατμόσφαιρα κατά τη διάρκεια αυτών των πέντε ετών. Για κάθε ένα από τα μοντέλα που δημιουργούνται ακολουθούν και τεχνικές αξιολόγησης των προβλεπόμενων αποτελεσμάτων σύμφωνα με τις πραγματικές παρατηρήσεις (βλ. Κεφάλαιο 6).

1.3 Περιγραφή των Δεδομένων

Τα δεδομένα

Για τις ανάγκες της μελέτης που ακολουθεί, χρησιμοποιήθηκαν δεδομένα που παρέχονται ελεύθερα από τη διαδικτυακή αποθήκη δεδομένων προς χρήση σε προβλήματα Μηχανικής Μάθησης, UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>). Αυτά περιέχουν ωριαίες μετρήσεις για τη συγκέντρωση του σωματιδίου $PM_{2.5}$ από την Αμερικάνικη Πρεσβεία στο Πεκίνο για το διάστημα [01.01.2010 – 31.12.2014], ενώ συμπεριλαμβάνονται και μετεωρολογικά δεδομένα από το Διεθνές Αεροδρόμιο του Πεκίνου. Το σύνολο των δεδομένων είναι 43824 και ο αριθμός των μεταβλητών είναι 13.

Οι μεταβλητές

Οι μεταβλητές του προβλήματος με βάση τις οποίες γίνεται η ανάλυση βρίσκονται στον ακόλουθο πίνακα :

Μεταβλητή	Περιγραφή μεταβλητής	Τύπος μεταβλητής	Παρατηρήσεις
No	Αριθμός γραμμής	Αριθμητική (ακέραιος)	
Year	Έτος που πραγματοποιήθηκε η μέτρηση των δεδομένων σε αυτή τη γραμμή	Αριθμητική (ακέραιος)	
Month	Μήνας που πραγματοποιήθηκε η μέτρηση των δεδομένων σε αυτή τη γραμμή	Αριθμητική (ακέραιος)	
Day	Ημέρα του μήνα που πραγματοποιήθηκε η μέτρηση των δεδομένων σε αυτή τη γραμμή	Αριθμητική (ακέραιος)	
Hour	Ώρα της ημέρας που πραγματοποιήθηκε η μέτρηση των δεδομένων σε αυτή τη γραμμή	Αριθμητική (ακέραιος)	
PM _{2.5}	Συγκέντρωση του αιωρούμενου σωματιδίου PM _{2.5} (μg/m ³)	Αριθμητική (πραγματικός)	Μεταβλητή απόκρισης
DEWP	Σημείο δρόσου	Αριθμητική (ακέραιος)	Σημείο δρόσου : η θερμοκρασία με την οποία όταν ψυχθούν οι υδρατμοί δημιουργούν το φαινόμενο της δρόσου, δηλαδή τη δημιουργία λεπτού στρώματος υδροσταγονιδίων πάνω σε επιφάνειες διαφόρων σωμάτων στην ύπαιθρο.
TEMP	Θερμοκρασία	Αριθμητική (πραγματικός)	
PRES	Πίεση (hPa)	Αριθμητική (πραγματικός)	
cbwd	Συνδυασμένη διεύθυνση ανέμου	Κατηγορική	
Iws	Ταχύτητα αέρα (m/s)	Αριθμητική (πραγματικός)	
Is	Συνολικές ώρες χιονόπτωσης	Αριθμητική (ακέραιος)	
Ir	Συνολικές ώρες βροχής	Αριθμητική (ακέραιος)	

Πίνακας 1.1 : Πίνακας περιγραφής των μεταβλητών του προβλήματος.

1.4 Επεξεργασία των Δεδομένων

Αντιμετώπιση δεδομένων με κενές τιμές για μεταβλητές

Ένα συχνό φαινόμενο στα σύνολα δεδομένων που αφορούν πραγματικές μετρήσεις είναι η απουσία τιμών για τις διάφορες μεταβλητές που αυτό περιέχει και η ανομοιογένεια. Για το λόγο αυτό, απαραίτητο βήμα πριν την οποιαδήποτε χρήση ή εισαγωγή τους σε κάποιο μοντέλο καθίσταται απαραίτητη η επεξεργασία τους για ανίχνευση τυχόν κενών ή λάθος τιμών και η κατάλληλη αντιμετώπισή τους. Διαφορετικά, η εκπαίδευση ενός μοντέλου με δεδομένα που έχουν πολλές κενές τιμές μπορεί να επηρεάσει δραστικά την ποιότητα αυτού. Με αυτόν τον τρόπο επομένως, μπορεί να αποφευχθεί η κατάληξη σε ανακριβή και εσφαλμένα συμπεράσματα.

Ορισμένοι από τους τρόπους αντιμετώπισης ελλιπών δεδομένων αναφέρονται ακολούθως :

- Διαγραφή γραμμών που περιέχουν μεταβλητές για τις οποίες απουσιάζουν τιμές.
- Αντικατάσταση των κενών τιμών με άλλες μηδενικές ή τυχαίες θετικές τιμές.
- Αντικατάσταση των κενών τιμών με τη διάμεσο ή τη μέση τιμή των μεταβλητών όπου βρίσκονται.
- Αντικατάσταση των κενών τιμών με άλλες τυχαίες στο διάστημα $[\mu + 1\sigma, \mu - 1\sigma]$, όπου μ η μέση τιμή και σ η τυπική απόκλιση της μεταβλητής για την οποία απουσιάζουν τιμές.

Στην περίπτωση της συγκεκριμένης μελέτης, επιλέχθηκε η πρώτη προτεινόμενη στρατηγική για την αντιμετώπιση των κενών τιμών των μεταβλητών. Καθώς αυτές είναι συνολικά 2067 – όλες βρίσκονται στη στήλη με τις μετρήσεις του αιωρούμενου σωματιδίου $PM_{2.5}$ – δηλαδή αποτελούν λιγότερο από το 5% του συνόλου δεδομένων και επομένως δεν χάνεται μεγάλο κομμάτι πληροφοριών, μπορούν να αφαιρεθούν οι αντίστοιχες γραμμές μετρήσεων.

Επιπλέον, παρουσιάζονται και δύο μηδενικές μετρήσεις του σωματιδίου, οι οποίες αφαιρούνται καθώς δηλώνουν λάθος μέτρηση.

Τέλος, από το σύνολο των δεδομένων αφαιρείται η μεταβλητή '**No**' (βλ. Πίνακα 1.1) καθώς δεν προσφέρει κάποια πληροφορία που μπορεί να χρησιμοποιηθεί σε κάποιο μοντέλο, πέρα από τον αριθμό της γραμμής που βρίσκονται οι μετρήσεις.

Επεξεργασία κατηγορικών μεταβλητών

Πολλές φορές στα σύνολα δεδομένων περιέχονται κατηγορικές μεταβλητές. Αυτές δεν είναι αριθμητικές, αλλά περιγράφουν δεδομένα τα οποία μπορεί να ανήκουν σε κάποια ομάδα – κατηγορία, όπως το φύλο, η εθνικότητα και το επίπεδο εκπαίδευσης. Ωστόσο, τα μοντέλα μηχανικής μάθησης χρησιμοποιούν μόνο αριθμητικά δεδομένα, επομένως είναι απαραίτητη η μετατροπή όποιων κατηγορικών μεταβλητών σε αριθμητικές.

Στο συγκεκριμένο πρόβλημα, όπως φαίνεται και από τον Πίνακα 1.1, η μεταβλητή ‘*cbwd*’ που περιγράφει τη διεύθυνση του ανέμου είναι κατηγορική. Έτσι, δημιουργούνται εικονικές μεταβλητές ή αλλιώς ψευδομεταβλητές, οι οποίες είναι όσες και οι διαφορετικές κατηγορίες της δοσμένης μεταβλητής και λαμβάνουν τις τιμές 0 ή 1, ανάλογα με την κατηγορία στην οποία ανήκει η κάθε μέτρηση, με 0 για τις νέες μεταβλητές – κατηγορίες που δεν ανήκει η μέτρηση και 1 για εκείνη όπου αντιστοιχίζεται.

Αποτελεσματικά, οι τέσσερις νέες μεταβλητές που προκύπτουν περιγράφονται στον ακόλουθο πίνακα :

Μεταβλητή	Περιγραφή μεταβλητής	Τύπος μεταβλητής
wind_direction_NE	Βορειοανατολική κατεύθυνση ανέμου	Δυαδική
wind_direction_NW	Βορειοδυτική κατεύθυνση ανέμου	Δυαδική
wind_direction_SE	Νοτιοανατολική κατεύθυνση ανέμου	Δυαδική
wind_direction_cv	Ήρεμη και μεταβλητή κατεύθυνση ανέμου	Δυαδική

Πίνακας 1.2 : Πίνακας περιγραφής των νέων δυαδικών μεταβλητών.

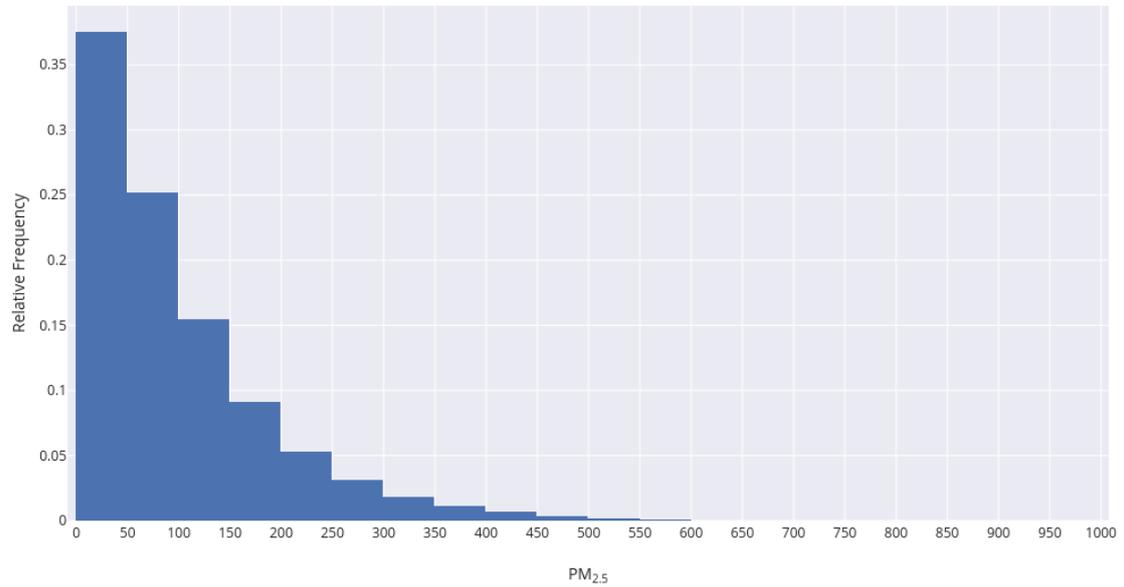
Στον Πίνακα 1.3 παρουσιάζονται συνοπτικά ορισμένα στατιστικά στοιχεία για τις αριθμητικές μεταβλητές του προβλήματος. Συγκεκριμένα, αυτά είναι ο συνολικός αριθμός των μετρήσεων, η μέση τιμή και η τυπική απόκλιση, το ελάχιστο στοιχείο, οι τιμές των τεταρτημόριων και το μέγιστο στοιχείο για την κάθε μεταβλητή.

	year	month	day	hour	pm2.5	DEWP	TEMP	PRES	lws	ls	lr
count	41755.00	41755.00	41755.00	41755.00	41755.00	41755.00	41755.00	41755.00	41755.00	41755.00	41755.00
mean	2012.04	6.51	15.68	11.50	98.62	1.75	12.40	1016.44	23.86	0.06	0.19
std	1.42	3.45	8.79	6.92	92.05	14.43	12.18	10.30	49.61	0.78	1.42
min	2010.00	1.00	1.00	0.00	1.00	-40.00	-19.00	991.00	0.45	0.00	0.00
25%	2011.00	4.00	8.00	5.00	29.00	-10.00	2.00	1008.00	1.79	0.00	0.00
50%	2012.00	7.00	16.00	12.00	72.00	2.00	14.00	1016.00	5.37	0.00	0.00
75%	2013.00	10.00	23.00	18.00	137.00	15.00	23.00	1025.00	21.91	0.00	0.00
max	2014.00	12.00	31.00	23.00	994.00	28.00	42.00	1046.00	565.49	27.00	36.00

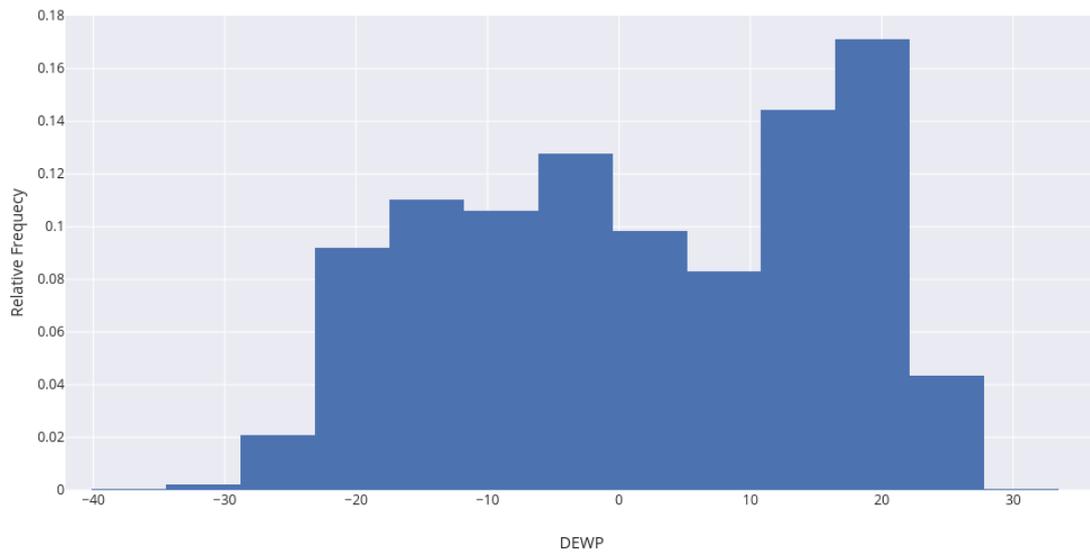
Πίνακας 1.2 : Στατιστικά στοιχεία των μεταβλητών.

Τα Διαγράμματα 1.1 - 1.7 αποτελούν ιστογράμματα σχετικών συχνοτήτων, όπου ο *yy*’ άξονας εκφράζει πυκνότητα και το συνολικό εμβαδόν του ιστογράμματος ολοκληρώνει

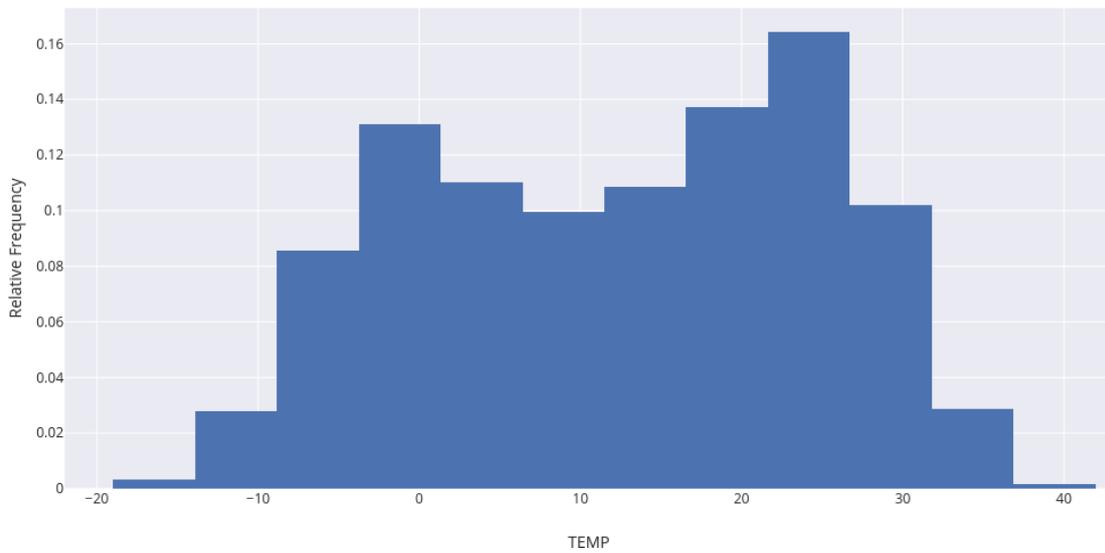
στη μονάδα. Με αυτόν τον τρόπο λαμβάνεται μία εκτίμηση για την κατανομή των μεταβλητών που μελετώνται.



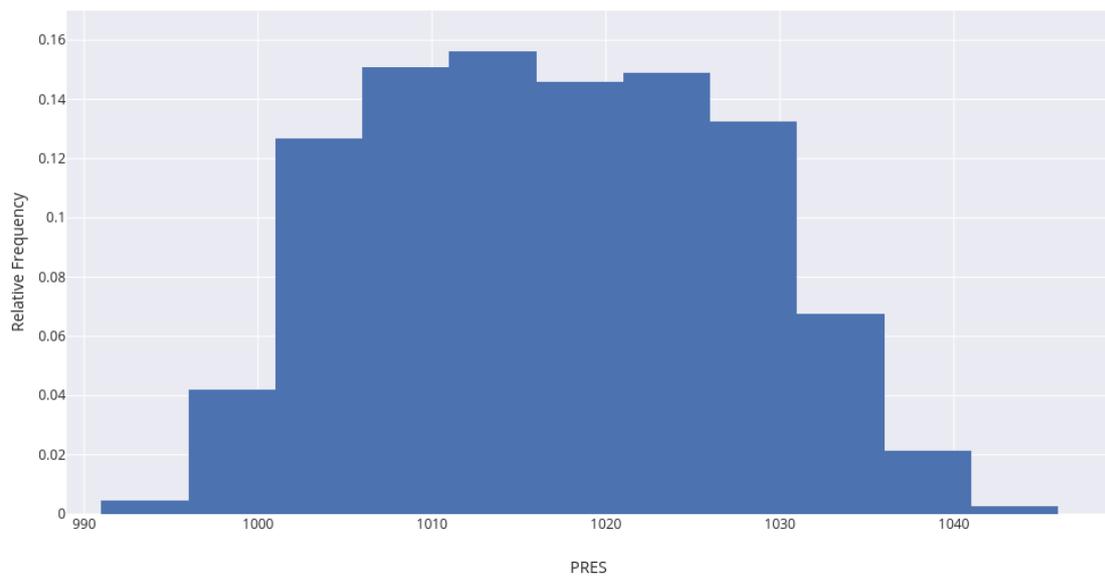
Διάγραμμα 1.1 : Ιστόγραμμα σχετικών συχνοτήτων για τη μεταβλητή $PM_{2.5}$ του προβλήματος.



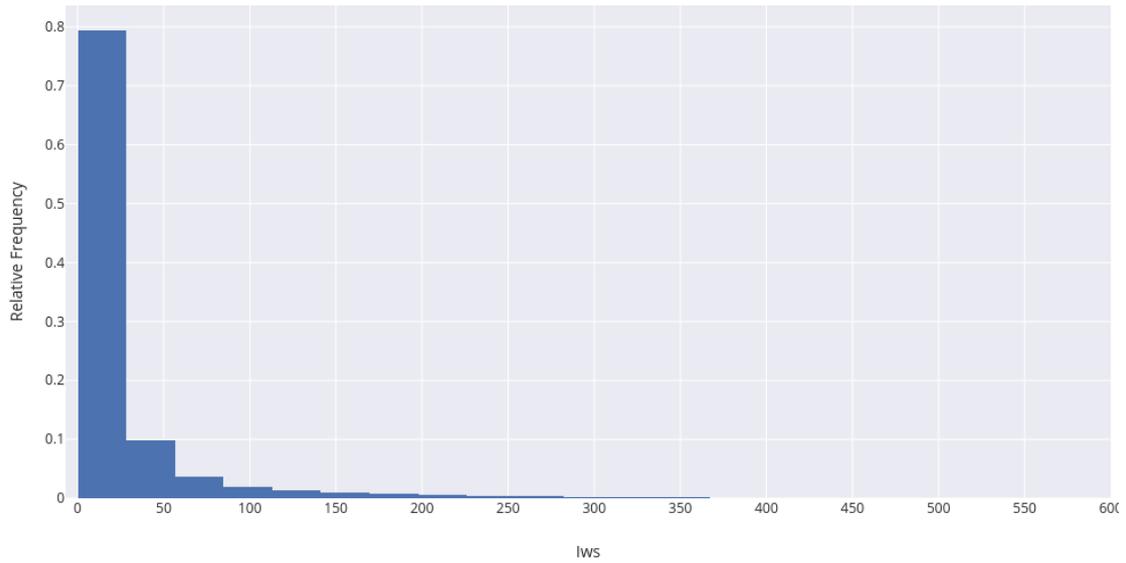
Διάγραμμα 1.2 : Ιστόγραμμα σχετικών συχνοτήτων για τη μεταβλητή $DEWP$ του προβλήματος.



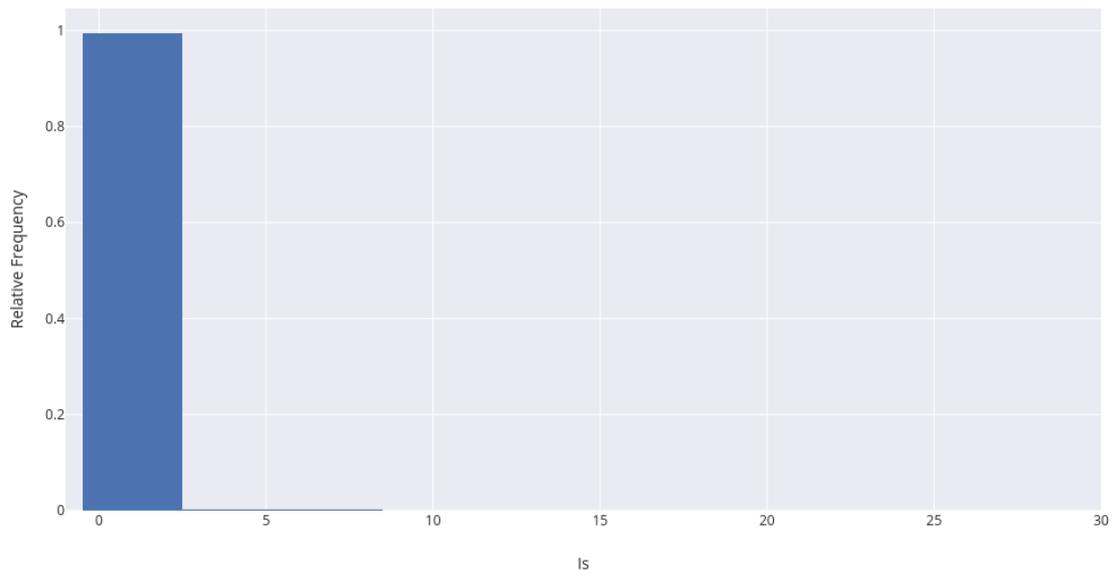
Διάγραμμα 1.3 : Ιστόγραμμα σχετικών συχνοτήτων για τη μεταβλητή TEMP του προβλήματος.



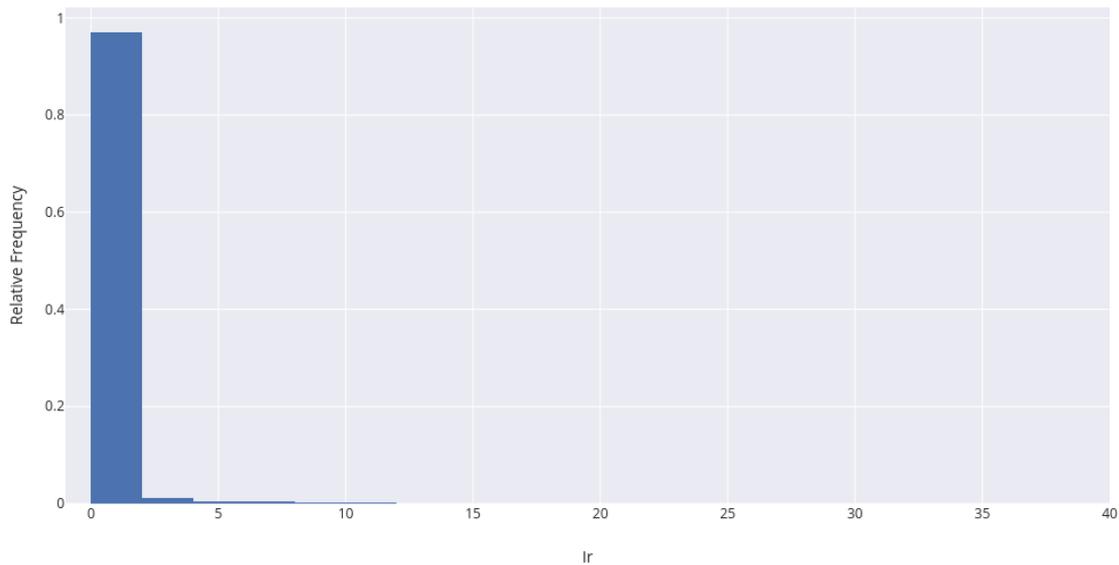
Διάγραμμα 1.4: Ιστόγραμμα σχετικών συχνοτήτων για τη μεταβλητή PRES του προβλήματος.



Διάγραμμα 1.5 : Ιστόγραμμα σχετικών συχνοτήτων για τη μεταβλητή lws του προβλήματος.



Διάγραμμα 1.6 : Ιστόγραμμα σχετικών συχνοτήτων για τη μεταβλητή Is του προβλήματος.



Διάγραμμα 1.7 : Ιστόγραμμα σχετικών συσχετίσεων για τη μεταβλητή I_r του προβλήματος

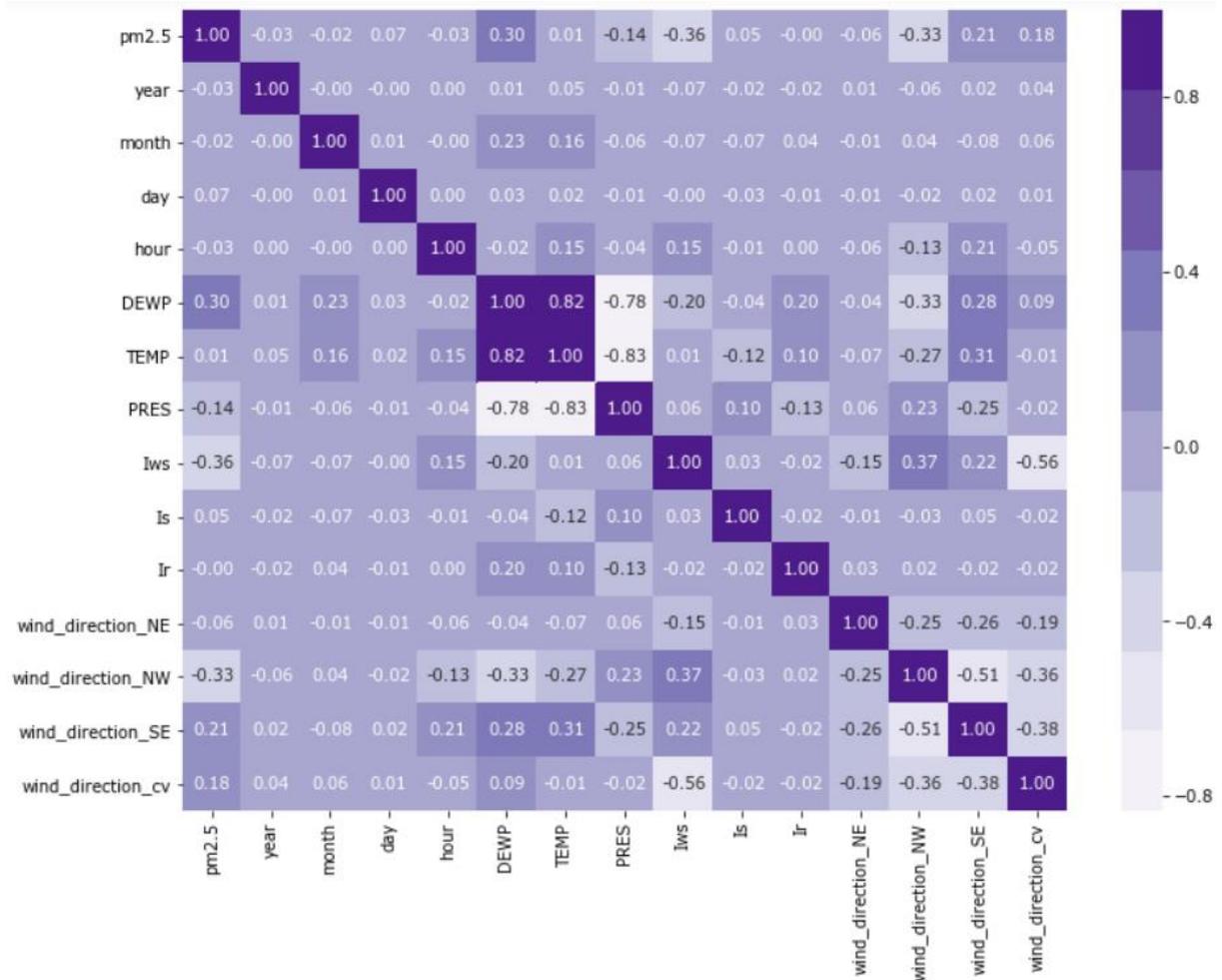
Συσχετίσεις μεταβλητών

Ένα σημαντικό κομμάτι κατά την επεξεργασία των δεδομένων είναι η εύρεση συσχετίσεων μεταξύ των μεταβλητών – ο όρος ‘συσχέτιση’ αναφέρεται στην επιρροή μιας μεταβλητής σε μία άλλη ή αν δεν αλληλοεπηρεάζονται, στην επιρροή που τους ασκεί μία άλλη μεταβλητή. Αυτό γίνεται με τον υπολογισμό του συντελεστή συσχέτισης για κάθε ένα ζευγάρι ποσοτήτων, ο οποίος περιγράφει τόσο το βαθμό, όσο και τη διεύθυνση της μεταξύ τους σχέσης και λαμβάνει τιμές στο διάστημα $[-1, 1]$. Οι ακραίες τιμές ± 1 σημαίνουν ισχυρή θετική ή αρνητική συσχέτιση αντίστοιχα, ενώ όσο πιο κοντά στο 0 βρίσκεται ο συντελεστής εκτιμάται ότι οι ποσότητες είναι ασυσχέτιστες. Ειδικότερα, η θετική συσχέτιση συνεπάγεται την αύξηση της μίας ποσότητας όταν η άλλη επίσης αυξάνεται, αντίθετα η αρνητική συσχέτιση δηλώνει τη μείωση της μίας ποσότητας όταν η άλλη αυξάνεται.

Στόχος του υπολογισμού συσχετίσεων μεταξύ των μεταβλητών είναι η εύρεση και αφαίρεση εκείνων που είναι υψηλά συσχετισμένες μεταξύ τους. Αποτελεσματικά, ο αριθμός των μεταβλητών μειώνεται, καθώς αν υπάρχει ισχυρή συσχέτιση μεταξύ δύο ποσοτήτων, τότε η μία από αυτές δεν χρειάζεται αφού δεν προσφέρει επιπλέον πληροφορία, ενώ παράλληλα ο μαθησιακός αλγόριθμος που εφαρμόζεται θα είναι ταχύτερος.

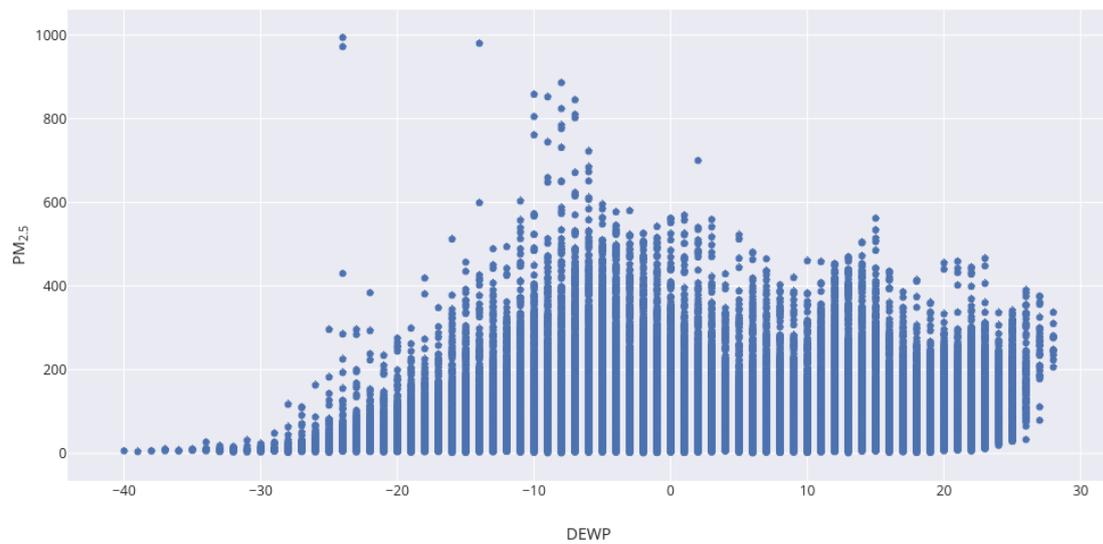
Για την εύρεση των συσχετίσεων μεταξύ των μεταβλητών υπολογίζεται ο συντελεστής συσχέτισης *Spearman*, ο οποίος δεν περιορίζεται μόνο σε γραμμικές συσχετίσεις, αλλά εκτιμά τη μονοτονική σχέση μεταξύ δύο μεταβλητών.

Τα αποτελέσματα παρατίθενται στον Πίνακα 1.4, με το μωβ χρώμα να δηλώνει θετική συσχέτιση, ενώ αντίθετα το λευκό αρνητική. Όσο πιο έντονο είναι το χρώμα στον πίνακα, τόσο πιο ισχυρή είναι και η συσχέτιση.

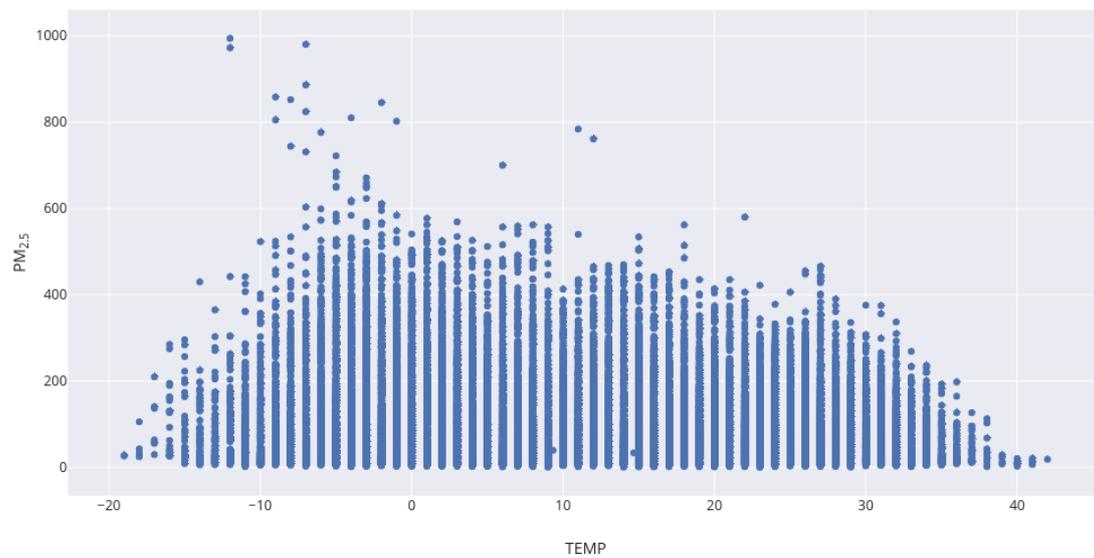


Πίνακας 1.3 : Πίνακας με τους συντελεστές συσχέτισης *Spearman* για τις μεταβλητές του προβλήματος.

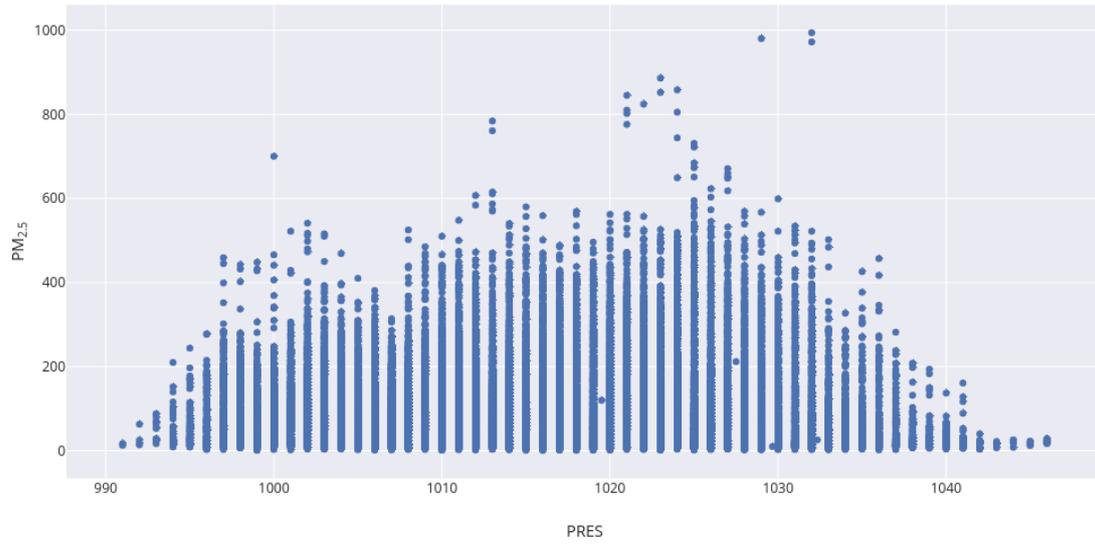
Τα Διαγράμματα 1.8 – 1.13 περιγράφουν τη συσχέτιση μεταξύ της μεταβλητής απόκρισης, $PM_{2.5}$, και των υπόλοιπων αριθμητικών μεταβλητών του προβλήματος.



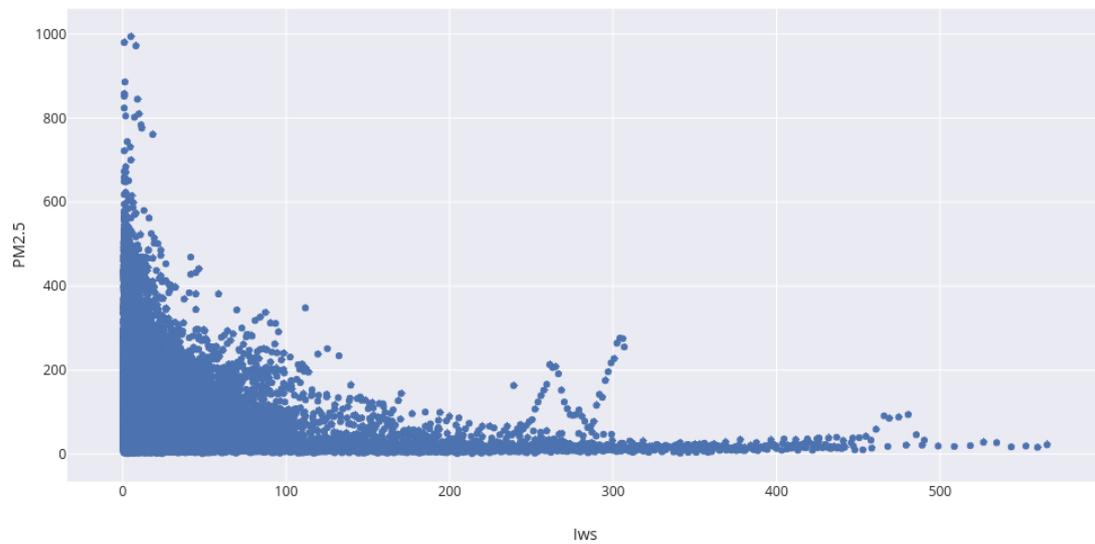
Διάγραμμα 1.8 : Διάγραμμα συσχέτισης μεταξύ των μεταβλητών DEWP και PM_{2.5}.



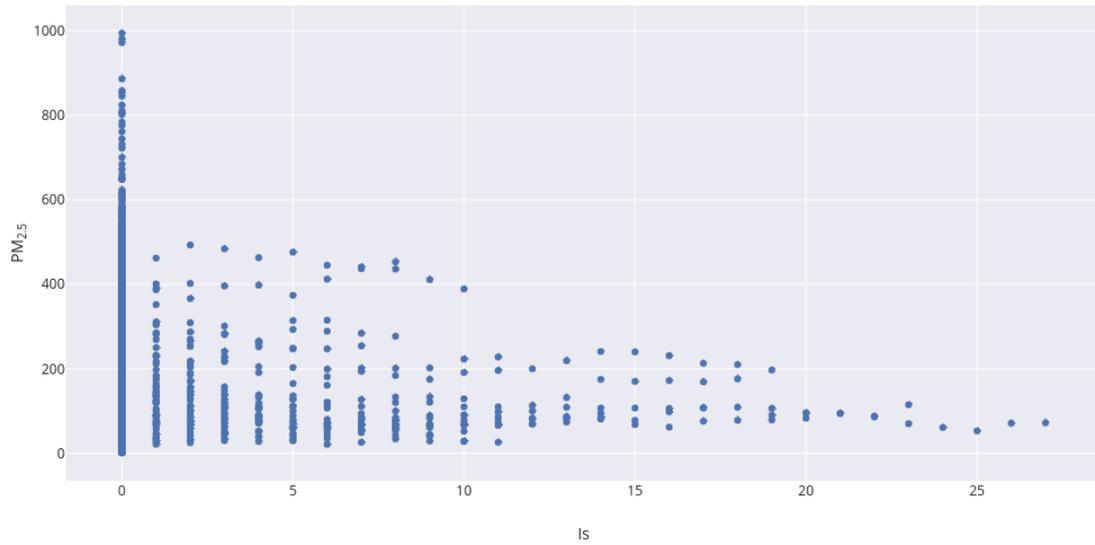
Διάγραμμα 1.9 : Διάγραμμα συσχέτισης μεταξύ των μεταβλητών TEMP και PM_{2.5} .



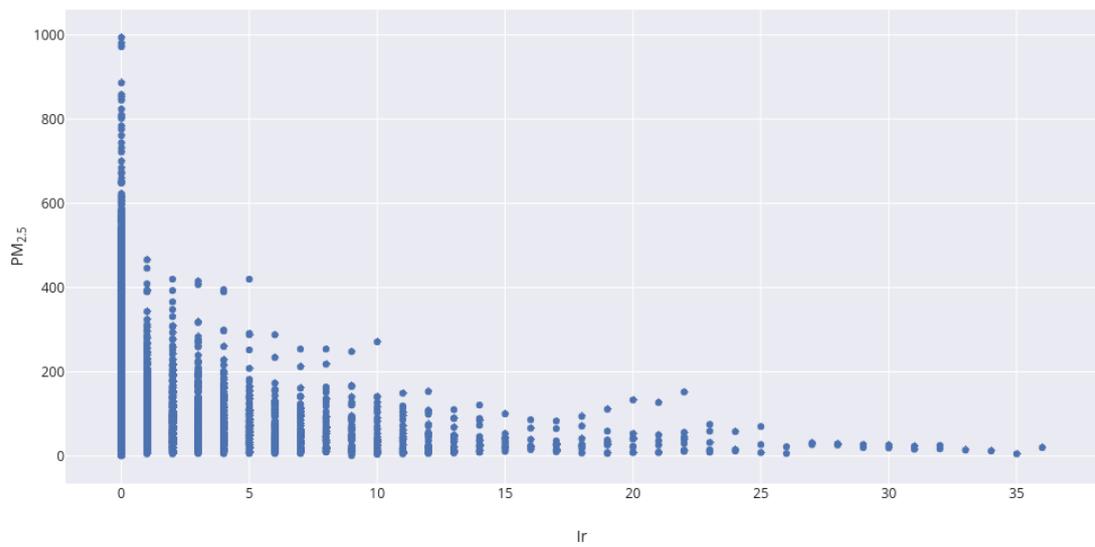
Διάγραμμα 1.10 : Διάγραμμα συσχέτισης μεταξύ των μεταβλητών PRES και $PM_{2.5}$.



Διάγραμμα 1.11 : Διάγραμμα συσχέτισης μεταξύ των μεταβλητών Iws και $PM_{2.5}$.



Διάγραμμα 1.12 : Διάγραμμα συσχέτισης μεταξύ των μεταβλητών Is και $PM_{2.5}$.



Διάγραμμα 1.13 : Διάγραμμα συσχέτισης μεταξύ των μεταβλητών Ir και $PM_{2.5}$.

Από τον Πίνακα 1.4 και τα διαγράμματα 1.8 – 1.13 συσχετίσεων δεν παρατηρείται κάποια ιδιαίτερα ισχυρή συσχέτιση μεταξύ των μεταβλητών. Επομένως, δεν χρειάζεται η αφαίρεση κάποιας ποσότητας του προβλήματος.

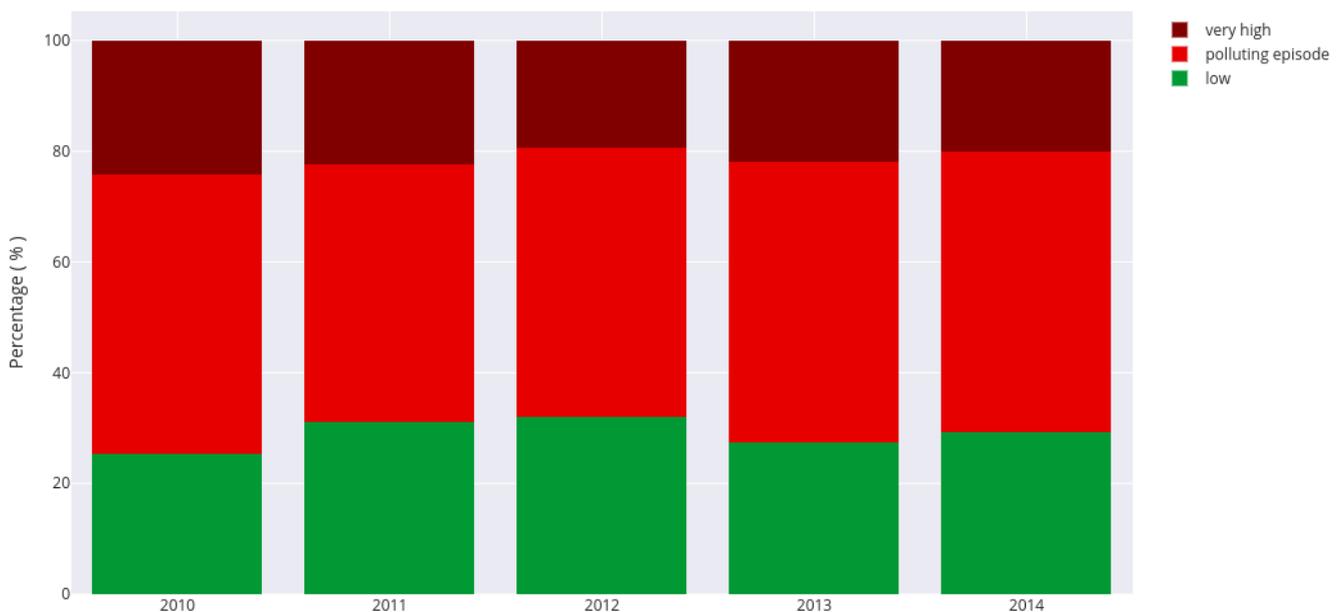
1.5 Περιγραφική Στατιστική των Δεδομένων

Σημείωση: Τα δεδομένα και τα αποτελέσματα που παρατίθενται στη συγκεκριμένη ενότητα ακολουθούν τα πρότυπα της μελέτης που θεωρείται ως αρχική εισήγηση και δεν ανταποκρίνονται στις σημερινές συνθήκες (περισσότερες λεπτομέρειες αναφορικά με τα ισχύοντα νόμους σήμερα μπορούν να αποκτηθούν και να αναζητηθούν από τον Παγκόσμιο Οργανισμό Υγείας (WHO)) [20].

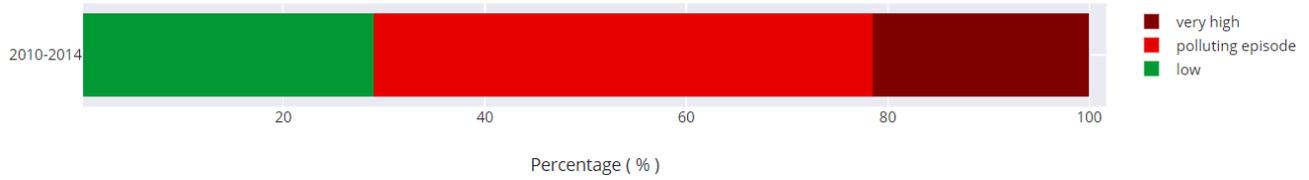
Σύμφωνα με τα πρότυπα των Ηνωμένων Πολιτειών της Αμερικής, η υψηλότερη τιμή μετρήσεων της συγκέντρωσης του αιωρούμενου σωματιδίου $PM_{2.5}$ που χαρακτηρίζει την ποιότητα του αέρα ‘αποδεκτή’ είναι $35 \mu\text{g m}^{-3}$, ενώ η τιμή $150 \mu\text{g m}^{-3}$ θεωρείται ανθυγιεινή, ακόμα και επικίνδυνη. Με βάση τις Εικόνες 1.2 και 1.3, μπορούμε να χωρίσουμε τις μετρήσεις για το αιωρούμενο σωματίδιο $PM_{2.5}$ σε τρία επίπεδα ανάλογα με την ποιότητα του αέρα :

- Χαμηλό επίπεδο (*low*), για μετρήσεις $PM_{2.5} \leq 35 \mu\text{g m}^{-3}$ (καλή ποιότητα αέρα).
- Ρυπογόνο επίπεδο (*polluting episode*), για μετρήσεις $35 < PM_{2.5} \leq 150 \mu\text{g m}^{-3}$.
- Πολύ υψηλό επίπεδο (*very high*), για μετρήσεις $PM_{2.5} > 150 \mu\text{g m}^{-3}$ (πολύ ανθυγιεινή ποιότητα αέρα).

Το Διάγραμμα 1.14 περιγράφει το ετήσιο ποσοστό (%) μετρήσεων συγκέντρωσης του σωματιδίου $PM_{2.5}$ που αντιστοιχίζεται σε κάθε ένα από τα παραπάνω επίπεδα για το διάστημα 2010 – 2014 και το Διάγραμμα 1.15 αντίστοιχα για το ίδιο χρονικό διάστημα συνολικά. Όμοια, το Διάγραμμα 1.16 αναπαριστά το ίδιο ποσοστό ανά εποχή για τα έτη 2010 – 2014.

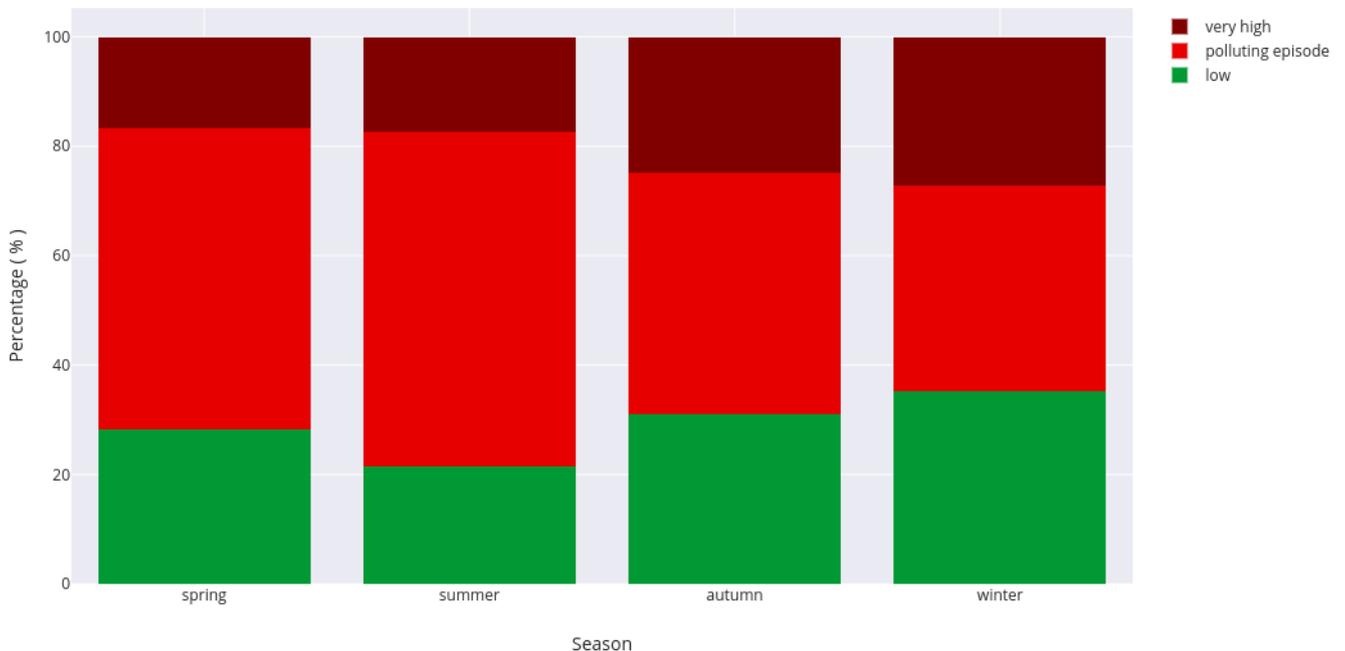


Διάγραμμα 1.14 : Ποσοστό μετρήσεων συγκέντρωσης του σωματιδίου $PM_{2.5}$ ετησίως για κάθε επίπεδο.



Διάγραμμα 1.15 : Ποσοστό μετρήσεων συγκέντρωσης του σωματιδίου PM_{2.5} για το διάστημα 2010 – 2014.

Από το πρώτο διάγραμμα, Διάγραμμα 1.14, παρατηρείται ότι για το διάστημα 2010 – 2014, οι περισσότερες μετρήσεις για τη συγκέντρωση του σωματιδίου PM_{2.5} κυμαίνονται μεταξύ 35 – 150 $\mu\text{g m}^{-3}$ ετησίως, το οποίο σημαίνει κακή ποιότητα αέρα και επομένως αρκετά επιβλαβείς συνθήκες για την υγεία. Αυτό επιβεβαιώνεται και από το Διάγραμμα 1.15, καθώς φαίνεται ότι οι μισές περίπου μετρήσεις για το πενταετές αυτό διάστημα αντιστοιχούν σε υψηλά επίπεδα συγκέντρωσης του σωματιδίου. Όσον αφορά το χαμηλό επίπεδο ενδείξεων του PM_{2.5}, αυτό αναλογεί στο 29 % των συνολικών μετρήσεων, ενώ οι τιμές του σωματιδίου που ξεπερνούν τις 150 μονάδες ισοδυναμεί περίπου με το 22 % των ολικών μετρήσεων (βλ. *Παράρτημα, Πίνακας Α.1*).



Διάγραμμα 1.16 : Ποσοστό μετρήσεων συγκέντρωσης του σωματιδίου PM_{2.5} για κάθε επίπεδο, ανά εποχή στο διάστημα 2010 – 2014.

Στο Διάγραμμα 1.16 σημειώνονται κάποιες διαφορές των μετρήσεων για το αιωρούμενο σωματίδιο PM_{2.5} ανά τις εποχές. Συγκεκριμένα, τόσο οι χαμηλές, όσο και

οι πολύ υψηλές μετρήσεις της συγκέντρωσης του σωματιδίου αυξάνονται κατά το φθινόπωρο και το χειμώνα, ενώ οι μετρήσεις που ανήκουν στο διάστημα [35 , 150] [$\mu\text{g m}^{-3}$] αποτελούν σχεδόν το 61% των συνολικών μετρήσεων κατά τους καλοκαιρινούς μήνες (βλ. Παράρτημα, Πίνακας Α.2).

2 Εισαγωγή στη Μηχανική Μάθηση

2.1 Εισαγωγή στη Μηχανική Μάθηση

Ένα κύριο χαρακτηριστικό που διακρίνει τον άνθρωπο από τους υπολογιστές είναι ο τρόπος αντιμετώπισης προβλημάτων. Ο άνθρωπος τείνει να μαθαίνει από τα λάθη του και προσπαθεί να λύσει το όποιο πρόβλημα διορθώνοντάς τα ή επιχειρώντας μία διαφορετική προσέγγιση. Αντίθετα, τα προγράμματα στους υπολογιστές δεν επεξεργάζονται το αποτέλεσμα των εργασιών τους, επομένως δεν είναι δυνατή η βελτίωσή τους. Ο τομέας της Μηχανικής Μάθησης απευθύνεται σε αυτό το πρόβλημα και συμβάλει στη δημιουργία προγραμμάτων που μπορούν να βελτιώσουν τις επιδόσεις τους συλλέγοντας δεδομένα και ‘εμπειρία’. Αποτελεί ένα διεπιστημονικό πεδίο που περιλαμβάνει τη θεωρία πιθανοτήτων και υπολογιστικής πολυπλοκότητας, τη στατιστική και τη θεωρία προσεγγίσεων, καθώς επίσης και πολλούς άλλους τομείς της επιστήμης των υπολογιστών.

Η Μηχανική Μάθηση είναι ένα πεδίο το οποίο επικεντρώνεται στη δημιουργία αλγορίθμων που κάνουν προβλέψεις με βάση τα δεδομένα που προσφέρονται. Ειδικότερα, οι αλγόριθμοι αυτοί καθιστούν τον υπολογιστή ικανό να εκπαιδευτεί με βάση το δοσμένο σύνολο δεδομένων, το οποίο αποτελεί τα δεδομένα εκπαίδευσης (*Training Data*), ακόμα και να βελτιώσει τις προβλέψεις του χωρίς να έχει ουσιαστικά αυστηρά προγραμματιστεί. Στόχος είναι η εκμάθηση μίας συνάρτησης $f : X \rightarrow Y$ που αντιστοιχίζει τα εισαγόμενα δεδομένα X σε μία μεταβλητή απόκρισης Y (η μεταβλητή που προβλέπει η εφαρμογή της μεθόδου Μηχανικής Μάθησης), ώστε να μπορεί να εκτελεί την ίδια διεργασία αποτελεσματικά και για νέα σύνολα δεδομένων [1].

2.2 Εφαρμογές Μηχανικής Μάθησης

Η Μηχανική Μάθηση αποτελεί σημαντικό εργαλείο στις περισσότερες εταιρίες που χειρίζονται μεγάλο όγκο δεδομένων λόγω της αποτελεσματικότητάς της. Τράπεζες και άλλες επιχειρήσεις της χρηματοπιστωτικής βιομηχανίας χρησιμοποιούν τεχνικές Μηχανικής Μάθησης τόσο για τον εντοπισμό σημαντικών πληροφοριών στα δεδομένα, όπως τον προσδιορισμό επενδυτικών ευκαιριών ή την ‘επίδειξη’ κατάλληλης στιγμής για διαπραγματεύσεις, όσο και για την πρόληψη απάτης. Επίσης, αποτελεί μία αναπτυσσόμενη τάση στον τομέα της υγειονομικής περίθαλψης χάρη στην εμφάνιση φορητών συσκευών και αισθητήρων που χρησιμοποιούν δεδομένα για την εκτίμηση της κατάστασης της υγείας ενός ασθενούς σε πραγματικό χρόνο. Παράλληλα, φαίνεται χρήσιμη και στους ιατρικούς εμπειρογνώμονες κατά τον εντοπισμό στοιχείων που μπορεί να οδηγήσουν σε βελτιωμένες διαγνώσεις και θεραπείες.

Παράλληλα, η εφαρμογή της Μηχανικής Μάθησης φαίνεται αποτελεσματική τόσο στην εύρεση νέων πηγών ενέργειας, όσο και στην ανάλυση ορυκτών του εδάφους, καθώς μπορεί να προβλέψει την αποτυχία του αισθητήρα διυλιστηρίου. Χρήσιμη ωστόσο φαίνεται και για την ανάλυση δεδομένων με στόχο τη διαμόρφωση της βέλτιστης διαδρομής για αποφυγή κυκλοφοριακής συμφόρησης, καθώς σε αυτή την περίπτωση εκτιμώνται οι περιοχές που μπορεί να βρεθεί συμφόρηση βάσει καθημερινών εμπειριών. Τέλος, εμφανίζεται αποτελεσματική στην αναγνώριση ανεπιθύμητης ηλεκτρονικής αλληλογραφίας και κακόβουλων προγραμμάτων, όπως επίσης στην αναγνώριση προσώπου και σε ιστοτόπους ηλεκτρονικών αγορών. Ειδικότερα για την τελευταία κατηγορία, αναλύεται το ιστορικό αγορών των χρηστών για τη δημιουργία εξατομικευμένης εμπειρίας αγορών ώστε να προτείνονται προϊόντα βάσει αυτού [2,3].

2.3 Είδη Μηχανικής Μάθησης

Η Μηχανική Μάθηση εφαρμόζεται σε ένα ευρύ φάσμα προβλημάτων, με αποτέλεσμα να διαχωρίζεται σε διάφορα είδη. Δύο από τις βασικότερες κατηγορίες είναι η *Εποπτευόμενη* (*Supervised*) και η *Μη Εποπτευόμενη Μάθηση* (*Unsupervised Learning*). Η συγκεκριμένη μελέτη αναφέρεται στην *Εποπτευόμενη Μάθηση*.

2.3.1 Εποπτευόμενη Μάθηση (*Supervised Learning*)

Στην *Εποπτευόμενη Μάθηση* (*Supervised Learning*) κάθε δείγμα του συνόλου δεδομένων αποτελεί ένα ζευγάρι διανύσματος των εισαγόμενων μεταβλητών, $x_n \in \mathcal{X}$ και της τιμής (ή διανύσματος) της μεταβλητής απόκρισης, $y_n \in \mathcal{Y}$, που το μοντέλο προσπαθεί να προβλέψει. Χρησιμοποιείται, δηλαδή, ένας αλγόριθμος για να ‘εκπαιδεύσει’ το μοντέλο ώστε να αντιστοιχίσει τις ποσότητες που εισάγονται, στη μεταβλητή απόκριση μέσω μιας συνάρτησης $f : \mathcal{X} \rightarrow \mathcal{Y}$. Στόχος είναι η βέλτιστη προσέγγιση της συνάρτησης αντιστοιχίσης, ώστε όταν εισαχθούν νέα δεδομένα, $x_j \in \mathcal{X}$, να μπορεί να προβλεφθεί η μεταβλητή απόκριση \mathcal{Y} . Η διαδικασία εκπαίδευσης τερματίζει όταν ο αλγόριθμος επιτύχει ένα αποδεκτό επίπεδο απόδοσης [4,5].

Τα προβλήματα Εποπτευόμενης Μηχανικής Μάθησης μπορούν να διαχωριστούν περεταίρω σε *προβλήματα Ταξινόμησης* (*Classification Problem*), όταν η μεταβλητή απόκριση είναι κατηγορική και σε *προβλήματα Παλινδρόμησης* (*Regression Problem*), όταν η μεταβλητή απόκριση λαμβάνει πραγματικές τιμές.

Από τους πιο διαδεδομένους αλγορίθμους που χρησιμοποιούνται σε προβλήματα Εποπτευόμενης Μάθησης είναι οι ακόλουθοι :

- Γραμμική Παλινδρόμηση (Linear Regression).
- Λογιστική Παλινδρόμηση (Logistic Regression).
- Δέντρα Αποφάσεων (Decision Trees), για προβλήματα Ταξινόμησης και Παλινδρόμησης.
- Νευρωνικά Δίκτυα (Neural Networks).

2.3.2 Μη Εποπτευόμενη Μάθηση (Unsupervised Learning)

Στη *Μη Εποπτευόμενη Μάθηση* (*Unsupervised Learning*) υπάρχουν μόνο οι μεταβλητές εισόδου, $\mathbf{x}_n \in \mathcal{X}$, χωρίς την αντίστοιχη μεταβλητή απόκρισης, $y_n \in \mathcal{Y}$. Στόχος είναι η εύρεση μοτίβων, της υποκείμενης δομής ή κατανομής των δεδομένων, ώστε να ληφθούν περισσότερες πληροφορίες για αυτά [4].

Τα προβλήματα Μη Εποπτευόμενης Μάθησης μπορούν να κατηγοριοποιηθούν σε *προβλήματα Ομαδοποίησης* (*Clustering Problem*), όταν τα δεδομένα χωρίζονται σε ομάδες βάσει της ομοιότητας των χαρακτηριστικών γνωρισμάτων τους και σε *προβλήματα Συσχέτισης* (*Association Problem*), όταν είναι επιθυμητή η εύρεση κανόνων που περιγράφουν μεγάλα τμήματα του συνόλου δεδομένων.

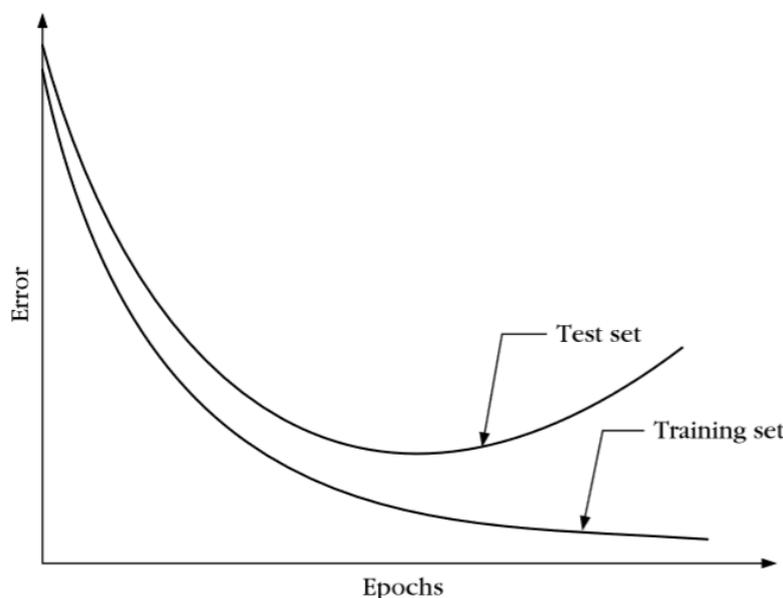
2.4 Υπερεκπαίδευση και Ανεπαρκής Εκπαίδευση

Ο όρος γενικευμένη απόδοση ενός μοντέλου Μηχανικής Μάθησης αναφέρεται στην ικανότητα αυτού να λειτουργεί αποτελεσματικά για σύνολα δεδομένων στα οποία δεν έχει εκπαιδευτεί. Αυτό σημαίνει ότι βάσει των πληροφοριών που έλαβε από τα δεδομένα εκπαίδευσης, θα πρέπει να ταξινομήσει κατάλληλα ή να προβλέψει τη μεταβλητή απόκρισης που αντιστοιχεί στα νέα δεδομένα. Πολλές φορές για να βρεθεί πόσο αποδοτικό είναι ένα μοντέλο, το αρχικό σύνολο δεδομένων χωρίζεται σε δύο επιμέρους τμήματα, τα οποία δεν είναι απαραίτητο να έχουν τον ίδιο αριθμό δειγμάτων. Από αυτά, το ένα προορίζεται για εκπαίδευση και αντιστοιχεί στο δείγμα εκπαίδευσης (*training set*) και το άλλο για έλεγχο, δείγμα αξιολόγησης (*test set*). Με αυτή τη

μέθοδο μπορεί να γίνει μία πρώτη προσέγγιση της απόδοσης του μοντέλου σε νέα δεδομένα.

Το φαινόμενο κατά το οποίο η επίδοση του μοντέλου στο δείγμα εκπαίδευσης είναι πολύ ικανοποιητική, ενώ η απόδοσή του στο δείγμα αξιολόγησης δεν είναι τόσο καλή ονομάζεται *υπερεκπαίδευση*. Αυτό συμβαίνει λόγω της προσαρμογής του μοντέλου στα δεδομένα εκπαίδευσης, δηλαδή εξειδικεύεται σε τυχόν διακυμάνσεις που αυτά εμφανίζουν και οδηγεί σε κακή γενικευμένη απόδοση όταν πρόκειται να εφαρμοστεί σε διανύσματα μεταβλητών από νέα σύνολα δεδομένων [6].

Η υπερεκπαίδευση ενός μοντέλου μπορεί να ανιχνευθεί εκτιμώντας τα σφάλματα που αυτό κάνει κατά την ταξινόμηση ή την πρόβλεψη των μεταβλητών ενός προβλήματος, δημιουργώντας ένα διάγραμμα της απόδοσης του μοντέλου συναρτήσει του χρόνου. Το Διάγραμμα 1.1 περιγράφει τα σφάλματα του μοντέλου συναρτήσει των βημάτων επανάληψης για τα δείγματα εκπαίδευσης και αξιολόγησης.



Διάγραμμα 2.1 : Διάγραμμα της συνάρτησης σφάλματος του μοντέλου σε σχέση με τα βήματα επανάληψης για τα δείγματα εκπαίδευσης και αξιολόγησης.

Η καμπύλη που αντιστοιχεί στο δείγμα εκπαίδευσης (training set) δείχνει το σφάλμα συνεχώς να μειώνεται, ενώ από την αντίστοιχη για το δείγμα εκπαίδευσης φαίνεται πως το σφάλμα μειώνεται μέχρι ένα σημείο και στη συνέχεια αρχίζει να αυξάνεται. Το διάγραμμα αυτό περιγράφει το φαινόμενο της υπερεκπαίδευσης. Το μοντέλο εξειδικεύεται στα δεδομένα εκπαίδευσης και αρχίζει να μην έχει καλή απόδοση όταν εισάγονται τα δεδομένα αξιολόγησης. Με άλλα λόγια, απομνημονεύει τα αποτελέσματα του δείγματος εκπαίδευσης και δεν γενικεύεται κατά την εφαρμογή του στο δείγμα αξιολόγησης.

Το αντίθετο φαινόμενο της υπερεκπαίδευσης αποτελεί η *ανεπαρκής εκπαίδευση*. Αυτό συμβαίνει όταν ένα μοντέλο αποτυγχάνει να αντιληφθεί σημαντικές διαφορές και

μοτίβα των δεδομένων, δηλαδή όταν δεν είναι αρκετά σύνθετο ώστε να συλλάβει τις συσχετίσεις μεταξύ των χαρακτηριστικών μεταβλητών του μοντέλου και της μεταβλητής απόκρισης. Ένα μοντέλο που δεν έχει εκπαιδευτεί επαρκώς καταλήγει σε εσφαλμένα αποτελέσματα για τα νέα δεδομένα στα οποία εφαρμόζεται, ενώ πολλές φορές έχει χαμηλές επιδόσεις ακόμα και στο δείγμα εκπαίδευσης.

3 Γραμμική Παλινδρόμηση

3.1 Εισαγωγή στη Γραμμική Παλινδρόμηση

Στόχος της παλινδρόμησης είναι η πρόβλεψη μίας ή περισσότερων ποσοτικών μεταβλητών απόκρισης y , με βάση ένα διάνυσμα διάστασης p , x , ανεξάρτητων μεταβλητών, ενώ χρησιμοποιείται και για την έρευνα της μεταξύ τους συσχέτισης. Τα πολυώνυμα αποτελούν παράδειγμα μίας ευρείας ομάδας συναρτήσεων που ονομάζονται *γραμμικά μοντέλα παλινδρόμησης*, ενώ την απλούστερη μορφή αυτών αποτελούν γραμμικές συναρτήσεις των μεταβλητών που εισάγονται στο μοντέλο.

Δεδομένου ενός συνόλου δεδομένων εκπαίδευσης που περιλαμβάνει N παρατηρήσεις $x = \{x_n\}, n = 1, \dots, N$, μαζί με τις αντίστοιχες μεταβλητές απόκρισης $y = \{y_n\}$, σκοπός είναι η πρόβλεψη της τιμής της μεταβλητής y για μία νέα τιμή της μεταβλητής x . Ο πιο απλός τρόπος για να επιτευχθεί αυτό είναι η δημιουργία μίας κατάλληλης συνάρτησης $f(x)$, οι τιμές της οποίας για τα νέα στοιχεία της μεταβλητής x αποτελούν τις προβλέψεις για τις αντίστοιχες τιμές της y . Γενικότερα, σύμφωνα με τη θεωρία πιθανοτήτων, στόχος είναι η δημιουργία ενός μοντέλου $p(y/x)$, καθώς εκφράζει την αβεβαιότητα για τον προσδιορισμό της τιμής της μεταβλητής y με βάση την τιμή που λαμβάνει η μεταβλητή x [10].

Η *γραμμική παλινδρόμηση* αποτελεί ένα γραμμικό μοντέλο, δηλαδή ένα μοντέλο που υποθέτει ότι υπάρχει γραμμική συσχέτιση μεταξύ της επεξηγηματικής μεταβλητής x και της μεταβλητής απόκρισης y . Ειδικότερα, η τιμή της μεταβλητής y μπορεί να υπολογιστεί από έναν γραμμικό συνδυασμό της μεταβλητής x .

Όταν η μεταβλητή x αποτελεί ένα μονοδιάστατο διάνυσμα, δηλαδή υπάρχει μόνο μία επεξηγηματική μεταβλητή, τότε το μοντέλο ονομάζεται *απλό γραμμικό μοντέλο παλινδρόμησης*. Αντίστοιχα, όταν υπάρχουν περισσότερες από μία επεξηγηματικές μεταβλητές, $x = \{x_n\}, n = 1, \dots, N$, τότε το μοντέλο ονομάζεται *πολλαπλό μοντέλο γραμμικής παλινδρόμησης* [11].

3.1.1 Απλό Γραμμικό Μοντέλο

Το απλό γραμμικό μοντέλο περιλαμβάνει δύο μεταβλητές, την ανεξάρτητη ή αλλιώς επεξηγηματική μεταβλητή $\mathcal{X} = \{x_n\}, n = 1, \dots, N$ και την εξαρτημένη ή διαφορετικά τη μεταβλητή απόκρισης $\mathcal{Y} = \{y_n\}, n = 1, \dots, N$, οι οποίες συνδέονται μεταξύ τους με τη γραμμική συνάρτηση παλινδρόμησης. Επομένως, η μεταβλητή x επηρεάζει γραμμικά την αναμενόμενη τιμή της μεταβλητής y , δηλαδή :

$$\mathcal{Y} = w_0 + w_1 \mathcal{X} + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \quad (3.1)$$

, όπου w_0, w_1 είναι άγνωστες ποσότητες που καλούνται *συντελεστές του μοντέλου*.

Για συγκεκριμένη τιμή x της μεταβλητής \mathcal{X} και λαμβάνοντας υπόψη ότι η μέση τιμή του τυχαίου σφάλματος ε είναι μηδέν, προκύπτει η ακόλουθη σχέση :

$$\mathbb{E} [\mathcal{Y} | \mathcal{X} = x] = w_0 + w_1 x \quad (3.2)$$

Η παραπάνω σχέση μπορεί να εφαρμοστεί σε οποιοδήποτε ζεύγος του δείγματος $(y_1, x_1), \dots, (y_n, x_n)$. Έτσι για κάθε $n = 1, \dots, N$ ισχύει ότι :

$$y_n = w_0 + w_1 x_n + \varepsilon_n \quad (3.3)$$

, όπου ε_n είναι τα τυχαία σφάλματα που θεωρούνται ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την Κανονική Κατανομή με μηδενική μέση τιμή και τυπική απόκλιση σ . Για συγκεκριμένη τιμή x_n της μεταβλητής \mathcal{X} ισχύει η εξής σχέση :

$$\mathbb{E} [y_n | \mathcal{X}_n = x_n] = w_0 + w_1 x_n \quad (3.4)$$

Το παραπάνω μοντέλο καλείται *απλό γραμμικό μοντέλο παλινδρόμησης*.

Οι ποσότητες w_0 , w_1 και σ^2 είναι άγνωστες παράμετροι του μοντέλου, οι οποίες εκτιμώνται με τη βοήθεια των παρατηρήσεων $(y_1, x_1), \dots, (y_n, x_n)$, των τιμών δηλαδή του δείγματος $(\mathcal{Y}_1, \mathcal{X}_1), \dots, (\mathcal{Y}_n, \mathcal{X}_n)$. Επιπλέον, η ποσότητα w_0 εκφράζει τη μέση τιμή της μεταβλητής \mathcal{Y} , όταν η μεταβλητή \mathcal{X} έχει μηδενική τιμή, ενώ η ποσότητα w_1 εκφράζει την αναμενόμενη μεταβολή της μεταβλητής \mathcal{Y} για μία μονάδα αύξησης της μεταβλητής \mathcal{X} .

Εκτιμώντας τα w_0 και w_1 από τα \hat{w}_0 και \hat{w}_1 αντίστοιχα, με τη βοήθεια των παρατηρήσεων, προκύπτει η εκτιμώμενη ευθεία της απλής γραμμικής παλινδρόμησης:

$$\hat{\mathcal{Y}} = \hat{w}_0 + \hat{w}_1 x \quad (3.5)$$

Η τιμή $\hat{\mathcal{Y}}$ είναι η προβλεπόμενη τιμή της μεταβλητής \mathcal{Y} με βάση το απλό μοντέλο γραμμικής παλινδρόμησης και ισούται με δεσμευμένη μέση τιμή της μεταβλητής \mathcal{Y} , όταν $\mathcal{X} = x$, όπως αυτή εκτιμήθηκε χρησιμοποιώντας το μοντέλο παλινδρόμησης (3.2). Για κάθε παρατήρηση x_n μπορούν να υπολογιστούν οι παρατηρούμενες προβλεπόμενες τιμές $\hat{y}_n = \hat{w}_0 + \hat{w}_1 x_n$.

Τα σημεία (x_n, y_n) , $n = 1, \dots, N$, είναι πιθανόν να διαφέρουν από τα σημεία (x_n, \hat{y}_n) , επομένως υπάρχουν αποκλίσεις μεταξύ των πραγματικών παρατηρήσεων y_n και των προβλεπόμενων τιμών \hat{y}_n . Οι ποσότητες $e_n = y_n - \hat{y}_n$ αποτελούν τις εκτιμήσεις των άγνωστων σφαλμάτων ε_n και ονομάζονται *υπόλοιπα*.

3.1.2 Πολλαπλό Γραμμικό Μοντέλο

Όταν υπάρχει πάνω από μία επεξηγηματική μεταβλητή, οι οποίες με τη βοήθεια του τυχαίου διανύσματος \mathcal{X} συμβολίζονται ως $\mathcal{X} = (X_1, X_2, \dots, X_p)^T$, απαιτείται μία επέκταση του απλού γραμμικού μοντέλου (βλ. Παράγραφο 3.1.1). Δηλαδή, σύμφωνα με αυτό, το τυχαίο διάνυσμα \mathcal{X} επηρεάζει γραμμικά την τιμή της μεταβλητή \mathcal{Y} και επομένως το γενικό γραμμικό μοντέλο δίνεται από τη σχέση :

$$\mathcal{Y} = w_0 + w_1 X_1 + \dots + w_p X_p + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \quad (3.6)$$

Για γνωστό $\mathcal{X} = x = (x_1, x_2, \dots, x_p)^T$ και λαμβάνοντας υπόψη ότι η μέση τιμή του τυχαίου σφάλματος ε είναι μηδέν, προκύπτει η ακόλουθη σχέση :

$$\mathbb{E}[\mathcal{Y} | \mathcal{X} = x] = w_0 + w_1 x_1 + \dots + w_p x_p \quad (3.7)$$

Ισοδύναμα, αν $(Y_1, X_{11}, \dots, X_{1p})^T, \dots, (Y_n, X_{n1}, \dots, X_{np})^T$ είναι ένα τυχαίο δείγμα μεγέθους N , ισχύει η σχέση :

$$Y_n = w_0 + w_1 X_{n1} + \dots + w_p X_{np} + \varepsilon_n, \quad n = 1, \dots, N \quad (3.8)$$

με $\varepsilon_n \sim N(0, \sigma^2)$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ονομάζονται *τυχαία σφάλματα*.

Συμβολίζοντας με $\mathcal{X}_n = (X_{n1}, X_{n2}, \dots, X_{np})^T$ το διάνυσμα των επεξηγηματικών μεταβλητών και κάθε μονάδα n του δείγματος και θεωρώντας ότι έχει γνωστή τιμή $x_n = (x_{n1}, x_{n2}, \dots, x_{np})^T, n = 1, \dots, N$, προκύπτει ότι :

$$\mathbb{E}[Y_n | \mathcal{X}_n = x_n] = w_0 + w_1 x_{n1} + \dots + w_p x_{np}, \quad n = 1, \dots, N \quad (3.9)$$

Το μοντέλο (3.7) καλείται *πολλαπλό γραμμικό μοντέλο παλινδρόμησης*.

Οι ποσότητες w_0, w_1, \dots, w_p καλούνται συντελεστές του μοντέλου και μαζί με τη διασπορά σ^2 του τυχαίου σφάλματος αποτελούν τις άγνωστες παραμέτρους που εκτιμώνται από τις παρατηρήσεις $(y_1, x_{11}, \dots, x_{1p})^T, \dots, (y_n, x_{n1}, \dots, x_{np})^T$, των παρατηρηθεισών δηλαδή τιμών του τυχαίου δείγματος $(Y_1, X_{11}, \dots, X_{1p})^T, \dots, (Y_n, X_{n1}, \dots, X_{np})^T$. Το πρόσημο του συντελεστή $w_j, j = 1, \dots, p$, προσδιορίζει τη σχέση εξάρτησης, θετική ή αρνητική, μεταξύ των μεταβλητών \mathcal{Y} και X_j , όταν όλες οι υπόλοιπες επεξηγηματικές μεταβλητές $X_l, l \neq j$, παραμείνουν σταθερές.

Το πολλαπλό γραμμικό μοντέλο (3.7) μπορεί με τη βοήθεια διανυσμάτων να πάρει την ακόλουθη μορφή :

$$\mathbb{E} [\mathcal{Y} | \mathcal{X} = x] = \tilde{x}w \quad (3.10)$$

με $\tilde{x} = (1, x_1, \dots, x_p)^T$ και $w = (w_0, \dots, w_p)^T$.

Αντίστοιχα, η σχέση (3.8) μπορεί να παρασταθεί με μορφή πινάκων ως :

$$\mathcal{Y} = \tilde{X}w + \varepsilon \quad (3.11)$$

με $\mathcal{Y} = (Y_1, \dots, Y_n)^T$, $w = (w_0, \dots, w_p)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ και

$$\tilde{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \vdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}$$

Ο $n \times (p + 1)$ πίνακας \tilde{X} ονομάζεται *πίνακας σχεδιασμού*.

Οι άγνωστοι συντελεστές του μοντέλου εκτιμώνται από της παρατηρήσεις $(y_1, x_{11}, \dots, x_{1p})^T, \dots, (y_n, x_{n1}, \dots, x_{np})^T$ και έτσι προκύπτει η *εκτιμώμενη ευθεία της πολλαπλής γραμμικής παλινδρόμησης* :

$$\hat{Y} = \hat{w}_0 + \hat{w}_1 x_1 + \cdots + \hat{w}_p x_p$$

όπου \hat{Y} είναι η προβλεπόμενη τιμή της μεταβλητής Y όπως προκύπτει από το μοντέλο παλινδρόμησης (3.7).

Για κάθε $n = 1, \dots, N$ υπολογίζονται οι *παρατηρούμενες* ή *προβλεπόμενες* τιμές $\hat{y}_n = \hat{w}_0 + \hat{w}_1 x_{n1} + \cdots + \hat{w}_p x_{np}$. Οι ποσότητες $e_n = y_n - \hat{y}_n$ ονομάζονται *υπόλοιπα*.

3.2 Μέθοδος Ελαχίστων Τετραγώνων

Οι συντελεστές του γενικού γραμμικού μοντέλου, w_0, w_1, \dots, w_p , εκτιμώνται βάσει των παρατηρήσεων $(y_1, x_{11}, \dots, x_{1p})^T, \dots, (y_n, x_{n1}, \dots, x_{np})^T$ εφαρμόζοντας τη *μέθοδο ελαχίστων τετραγώνων*. Με τη μέθοδο αυτή, επιλέγεται η ευθεία που προσαρμόζεται καλύτερα στα δεδομένα. Ειδικότερα, σκοπός είναι η επιλογή της ευθείας για την οποία τα υπόλοιπα $e_n, n = 1, \dots, N$ να είναι τα μικρότερα δυνατά, δηλαδή ελαχιστοποιείται το συνολικό σφάλμα (σφάλμα : απόκλιση από την επιθυμητή τιμή). Η ποσότητα που εκφράζει το συνολικό σφάλμα ονομάζεται *συνάρτηση κόστους*. Έτσι, επιλέγονται τα w_0, w_1, \dots, w_p που ελαχιστοποιούν το άθροισμα τετραγώνων των υπολοίπων, δηλαδή οι εκτιμητές $\hat{w}_0, \hat{w}_1, \dots, \hat{w}_p$ είναι εκείνες οι τιμές των w_0, w_1, \dots, w_p που ελαχιστοποιούν τη συνάρτηση :

$$\sum_{n=1}^N e_n^2 = \sum_{n=1}^N [y_n - \hat{y}_n]^2 = \sum_{n=1}^N [y_n - (\hat{w}_0 + \hat{w}_1 x_{n1} + \dots + \hat{w}_p x_{np})]^2 \quad (3.11)$$

Η διαφορετικά, τη συνάρτηση :

$$S(w) = (y - \tilde{X}b)^T (y - \tilde{X}b) \quad (3.12)$$

με $y = (y_1, \dots, y_N)^T$.

Παραγωγίζοντας την τελευταία σχέση ως προς b , θέτοντάς τη ίση με το μηδέν και λύνοντας το $(p+1) \times (p+1)$ σύστημα προκύπτει ότι :

$$\hat{w} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y \quad (3.13)$$

Απαραίτητη προϋπόθεση για τα παραπάνω είναι να ισχύει $N \geq (p+1)$ και οι επεξηγηματικές μεταβλητές να μην είναι γραμμικά εξαρτημένες μεταξύ τους, δηλαδή θα πρέπει $\text{rank}(\tilde{X}) = p + 1$. Τότε ο $(p+1) \times (p+1)$ πίνακας $\tilde{X}^T \tilde{X}$ αντιστρέφεται και είναι θετικά ορισμένος.

Σύμφωνα με τα παραπάνω, το εκτιμώμενο μοντέλο υπολογίζεται από τη σχέση :

$$\hat{Y} = X\hat{w} \quad (3.14)$$

Η συνάρτηση κόστους που χρησιμοποιείται στη γραμμική παλινδρόμηση, η οποία ονομάζεται *συνάρτηση μέσου τετραγωνικού σφάλματος*, είναι η ακόλουθη :

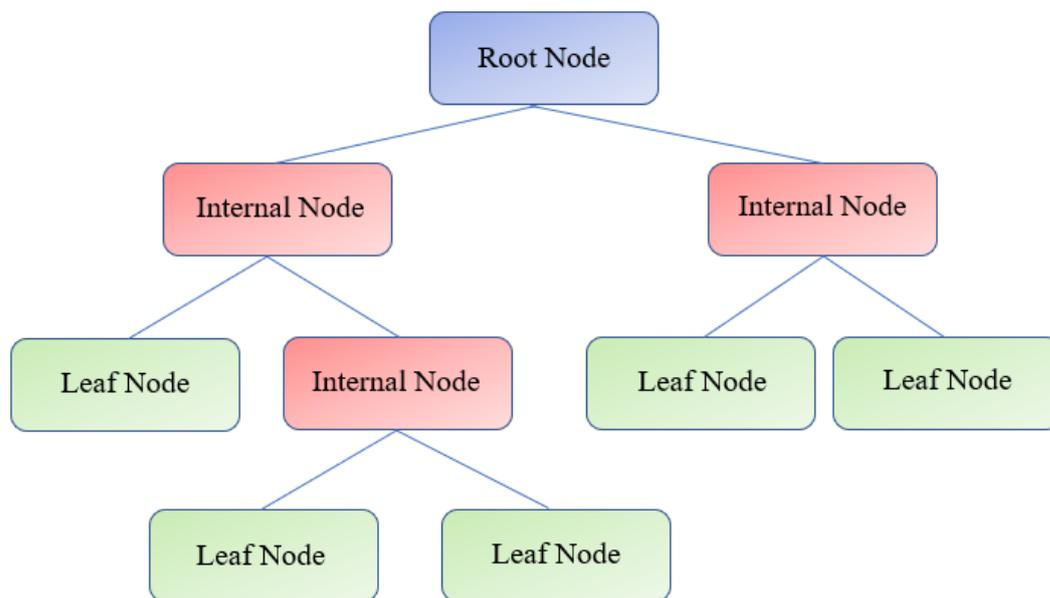
$$J(w) = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2 \quad (3.15)$$

4 Δέντρα Απόφασης

4.1 Εισαγωγή στα Δέντρα Απόφασης

Τα Δέντρα Απόφασης (*Decision Trees*) αποτελούν μία μη παραμετρική μέθοδο της Εποπτευόμενης Μηχανικής Μάθησης και εφαρμόζεται τόσο σε πρόβλημα ταξινόμησης, όσο και σε πρόβλημα παλινδρόμησης, με το οποίο ασχολείται και η συγκεκριμένη μελέτη. Στόχος είναι η δημιουργία ενός μοντέλου που προβλέπει την τιμή της μεταβλητής απόκρισης καθώς εκπαιδεύεται μέσα από μία αλληλουχία δυαδικών αποφάσεων που προκύπτουν από τα χαρακτηριστικά γνωρίσματα των δεδομένων.

Τα Δέντρα Απόφασης αποτελούν ουσιαστικά ένα πολυεπίπεδο σύστημα κόμβων, οι οποίοι απαρτίζονται από τη ρίζα, που είναι και η κορυφή του δέντρου, τους ενδιάμεσους κόμβους απόφασης, οι οποίοι διασπώνται σε δύο κλαδιά με βάση τα γνωρίσματα των μεταβλητών του συνόλου δεδομένων και οδηγούν είτε σε άλλους κόμβους για επιμέρους διαχωρισμούς, είτε στα φύλλα, τα οποία δεν διαχωρίζονται περαιτέρω και αποτελούν το τελευταίο επίπεδο του δέντρου. Ένα παράδειγμα ενός Δέντρου Απόφασης φαίνεται στην Εικόνα 4.1. Συνοπτικά, οι εκτιμήσεις που προκύπτουν από ένα Δέντρο Απόφασης είναι αποτέλεσμα μίας σειράς ερωτήσεων πάνω στις παρατηρήσεις των δεδομένων, που επιδέχονται απαντήσεις της μορφής ναι ή όχι.



Εικόνα 4.4 : Οπτικοποίηση Δέντρου Απόφασης.

Τα Δέντρα Απόφασης διαχωρίζουν τις επεξηγηματικές μεταβλητές σε μία σειρά από κόμβους για να καταλήξουν στα φύλλα που περιέχουν την πιο ομοιογενή συλλογή πιθανών αποτελεσμάτων. Η αντίστοιχη επιλογή μεταβλητών που γίνεται σε προβλήματα παλινδρόμησης είναι ανάλογη των διαχωρισμών αυτών που δημιουργούνται. Για τον κάθε διαχωρισμό χρησιμοποιείται μόνο μία επεξηγηματική μεταβλητή και επιλέγεται εκείνη για την οποία η ανομοιογένεια στον κόμβο είναι η ελάχιστη δυνατή [12].

4.2 Δέντρα Απόφασης για Προβλήματα Παλινδρόμησης

Τα Δέντρα Απόφασης που χρησιμοποιούνται για προβλήματα παλινδρόμησης αποτελούν μία κατηγορία των Δέντρων Απόφασης, όπου κάθε φύλλο αναπαριστά μία αριθμητική τιμή. Κάθε τερματικός κόμβος, φύλλο, του δέντρου αντιπροσωπεύει ένα κελί του διαχωρισμού που γίνεται και σε αυτό είναι συνδεδεμένο ένα μοντέλο που εφαρμόζεται σε αυτό. Μία παρατήρηση του συνόλου δεδομένων ανήκει σε ένα φύλλο, αν κατατάσσεται σε αυτό μέσω το διαχωρισμού που σχηματίζεται. Η κατασκευή των φύλλων προκύπτει ακολουθώντας μία ‘διαδρομή’ από τον αρχικό κόμβο του δέντρου, ρίζα, κάνοντας διαδοχικές ερωτήσεις για τις μεταβλητές. Οι ερωτήσεις αυτές γίνονται στους εσωτερικούς κόμβους του δέντρου, οι οποίοι συνδέονται μεταξύ τους με κλαδιά που φέρουν τις απαντήσεις και κάθε μία από αυτές εξαρτάται από τις απαντήσεις στις προηγούμενες ερωτήσεις.

Το σύνολο των δεδομένων εκπαίδευσης αποτελείται από p επεξηγηματικές μεταβλητές και μία μεταβλητή απόκρισης, για κάθε μία από τις οποίες υπάρχει N αριθμός παρατηρήσεων, δηλαδή $(x_1, y_1), \dots, (x_N, y_N)$ με $x_n = (x_{n1}, \dots, x_{np}) \in \mathcal{X}$ και $n = 1, \dots, N$. Ο αλγόριθμος καλείται να αποφασίσει για την επιλογή των επεξηγηματικών μεταβλητών και των οριακών τιμών με βάση τις οποίες θα γίνουν οι διαχωρισμοί στο δέντρο απόφασης.

Έστω ότι αριθμός των κόμβων του δέντρου είναι M , οι οποίοι συμβολίζονται ως R_m , $m = 1, \dots, M$ και ότι στον κόμβο R_m βρίσκεται ένα υποσύνολο \mathcal{X}_m του \mathcal{X} , δηλαδή το σύνολο των παρατηρήσεων $x_n \in \mathcal{X}$ που ικανοποιούν τις προϋποθέσεις των κόμβων κατά τη διαδρομή από τη ρίζα του δέντρου μέχρι τον κόμβο R_m . Ορίζεται :

$$b_m(x_n) = \begin{cases} 1, & \text{αν } x_n \in R_m \\ 0, & \text{διαφορετικά} \end{cases} \quad (4.1)$$

Στην παλινδρόμηση, ο διαχωρισμός ενός κόμβου σε επιμέρους αξιολογείται με βάση την τιμή του Μέσου Τετραγωνικού Σφάλματος (MSE : Mean Squared Error) για τις

εκτιμώμενες τιμές της μεταβλητής απόκρισης [4]. Αν $\hat{y}_n^{(m)} = f_m(\mathbf{x}_n)$ είναι η εκτιμώμενη τιμή στον κόμβο R_m , τότε το Μέσο Τετραγωνικό Σφάλμα ορίζεται ως :

$$MSE_m = \frac{1}{N_m} \sum_{n=1}^N (y_n - f_m(\mathbf{x}_n))^2 b_m(\mathbf{x}_n) \quad (4.2)$$

με $N_m = |\mathcal{X}_m| = \sum_{n=1}^N b_m(\mathbf{x}_n)$, ο αριθμός των στοιχείων βρίσκονται στον κόμβο R_m .

Σύμφωνα με τα παραπάνω, χρησιμοποιώντας ως συνάρτηση κόστους το μέσο όρο σφάλματος, η καλύτερη τιμή για την $f_m(\mathbf{x}_n)$ είναι η μέση τιμή των y_n που βρίσκονται στον κόμβο R_m [16]. Δηλαδή :

$$f_m(\mathbf{x}_n) = \frac{\sum_{n=1}^N b_m(\mathbf{x}_n) y_n}{\sum_{n=1}^N b_m(\mathbf{x}_n)} \quad (4.3)$$

Σημαντικό σημείο στη μέθοδο αποτελεί η κατάλληλη επιλογή της μεταβλητής και του οριακού σημείου με βάση τα οποία γίνεται ο διαχωρισμός στους κόμβους. Ξεκινώντας διαθέτοντας όλα τα δεδομένα του συνόλου εκπαίδευσης, έστω j η μεταβλητή για την οποία γίνεται ο πρώτος διαχωρισμός και s το αντίστοιχο οριακό σημείο. Οι δύο κόμβοι που προκύπτουν από το διαχωρισμό αυτό ορίζονται ως :

$$R_1(j, s) = \{ X \mid X_j \leq s \} \text{ και } R_2(j, s) = \{ X \mid X_j > s \} \quad (4.4)$$

Η διαχωριστική μεταβλητή j και το οριακό σημείο s βρίσκονται σύμφωνα με την ακόλουθη σχέση :

$$\min_{j,s} \left\{ \min_{\hat{y}_n^{(1)}} \left[\frac{1}{N_1} \sum_{\mathbf{x}_n \in R_1(j,s)} (y_n - \hat{y}_n^{(1)})^2 \right] + \min_{\hat{y}_n^{(2)}} \left[\frac{1}{N_2} \sum_{\mathbf{x}_n \in R_2(j,s)} (y_n - \hat{y}_n^{(2)})^2 \right] \right\} \quad (4.5)$$

Για οποιαδήποτε επιλογή των j και s , η εσωτερική ελαχιστοποίηση γίνεται για :

$$\hat{y}_n^{(1)} = \frac{\sum_{n=1}^N b_1(\mathbf{x}_n) y_n}{\sum_{n=1}^N b_1(\mathbf{x}_n)} \text{ και } \hat{y}_n^{(2)} = \frac{\sum_{n=1}^N b_2(\mathbf{x}_n) y_n}{\sum_{n=1}^N b_2(\mathbf{x}_n)} \quad (4.6)$$

Έχοντας βρει τον καλύτερο διαχωρισμό βάσει της μεταβλητής j και του οριακού σημείου s , τα δεδομένα χωρίζονται σε δύο κόμβους και η ίδια διαδικασία επαναλαμβάνεται για κάθε έναν από αυτούς, αλλά και για όσους επιμέρους κόμβους προκύπτουν.

Συνοψίζοντας τα παραπάνω, για την εύρεση της κατάλληλης μεταβλητής διαχωρισμού και του αντίστοιχου οριακού σημείου, υπολογίζεται για κάθε μία από αυτές η τιμή του

μέσου τετραγωνικού σφάλματος, με τη χρήση των σχέσεων και επιλέγεται εκείνη που δίνει τη μικρότερη τιμή για αυτό.

| Κριτήριο τερματισμού του διαχωρισμού σε επιπλέον κόμβους

Ένα δέντρο απόφασης μεγάλης διάστασης, δηλαδή που αποτελείται από πολλά επίπεδα, μπορεί να προσαρμοστεί πολύ καλά στα δεδομένα εκπαίδευσης και να οδηγηθεί σε υπερεκπαίδευση. Το μέγεθος ενός δέντρου αποτελεί μία παράμετρο που μπορεί να επηρεάσει σημαντικά την πολυπλοκότητα του μοντέλου. Για το λόγο αυτό, το βέλτιστο μέγεθος για ένα δέντρο θα πρέπει να προκύπτει προσαρμοστικά από τα δεδομένα.

Μία επιλογή είναι οι κόμβοι του δέντρου να διαχωρίζονται σε επιμέρους, μόνο όταν το μέσο τετραγωνικό σφάλμα μειώνεται αρκετά ώστε να ξεπερνά ένα δοσμένο όριο. Ωστόσο, αυτή η προσέγγιση δεν είναι ιδιαίτερα αποτελεσματική καθώς ένας διαχωρισμός κόμβου που μπορεί να φαίνεται ασήμαντος, μπορεί στη συνέχεια να οδηγήσει σε έναν άλλο πολύ ισχυρό διαχωρισμό. Μία προτεινόμενη λύση είναι ο ορισμός ενός ελάχιστου αριθμού παρατηρήσεων που περιέχονται στα φύλλα. Έτσι, αν το άθροισμα των παρατηρήσεων για κάποιο φύλλο δέντρου δεν ξεπερνά τον αριθμό αυτό δεν πραγματοποιείται επιπλέον διαχωρισμός του κόμβου [13, 16]. Εναλλακτικά, μπορεί να οριστεί μία μέγιστη τιμή για το βάθος του δέντρου [1].

4.3 Μέθοδος Bootstrap

Η μέθοδος *Bootstrap* αποτελεί μία στατιστική μέθοδο για την εκτίμηση μίας ποσότητας από ένα σύνολο δεδομένων, υπολογίζοντας τη μέση τιμή των εκτιμητών από πολλαπλά μικρότερα δείγματα του αρχικού συνόλου. Συγκεκριμένα, επιλέγονται τυχαίες παρατηρήσεις από τα δεδομένα με επανατοποθέτηση – δηλαδή η κάθε παρατήρηση εισάγεται ξανά στο αρχικό σύνολο και μπορεί να επιλεγθεί πάνω από μία φορά – και δημιουργείται ένας αριθμός ισόποσων δειγμάτων. Για κάθε ένα από αυτά υπολογίζεται η ποσότητα της οποίας ζητείται η εκτίμηση από τα συνολικά δεδομένα και υπολογίζεται η μέση τιμή αυτών [13]. Το τελικό αποτέλεσμα αποτελεί την εκτίμηση της ζητούμενης ποσότητας.

Η μέθοδος αυτή μπορεί να χρησιμοποιηθεί για την εκτίμηση της μέσης τιμής ή της διασποράς ενός συνόλου δεδομένων, καθώς επίσης και για την εκτίμηση της απόδοσης

ενός μοντέλου της Μηχανικής Μάθησης. Αυτό επιτυγχάνεται εκπαιδεύοντας το μοντέλο στο ένα δείγμα και ελέγχοντας την απόδοσή του στα υπόλοιπα.

4.4 Μέθοδος Bagging

Η μέθοδος *Bootstrap Aggregation* ή διαφορετικά εν συντομία μέθοδος *Bagging* είναι μία απλή και αποτελεσματική συνδυαστική μέθοδος. Δηλαδή, συνδυάζει τις προβλέψεις διάφορων αλγορίθμων Μηχανικής Μάθησης ώστε να προβεί σε ακριβέστερες προβλέψεις από ότι σαν ένα μεμονωμένο μοντέλο. Επιπλέον, αποτελεί μία διαδικασία με την οποία επιτυγχάνεται η μείωση της διασποράς (βλ. *Παράρτημα, Ερμηνευτικοί Όροι*) για αλγορίθμους με υψηλή διασπορά. Αλγόριθμοι με υψηλή διασπορά είναι τα Δέντρα Απόφασης, είτε εφαρμόζονται για προβλήματα ταξινόμησης είτε για προβλήματα παλινδρόμησης. Τα Δέντρα Απόφασης είναι ιδιαίτερα 'ευαίσθητα' ως προς τα δεδομένα με βάση τα οποία εκπαιδεύονται, επομένως όταν εισαχθεί ένα νέο σύνολο δεδομένων τα αποτελέσματά τους μπορεί είναι αρκετά διαφορετικά, άρα και οι προβλέψεις.

Με άλλα λόγια, η μέθοδος Bagging είναι η εφαρμογή της διαδικασίας της μεθόδου Bootstrap σε έναν αλγόριθμο υψηλής διασποράς, συνήθως Δέντρα Απόφασης. Δηλαδή το αρχικό σύνολο δεδομένων χωρίζεται σε επιμέρους ισόποσα δείγματα, το μοντέλο εκπαιδεύεται σε κάθε ένα από αυτά και όταν εισάγονται τα νέα δεδομένα υπολογίζεται η μέση τιμή των προβλέψεων του κάθε μοντέλου. Στην περίπτωση που αντιμετωπίζεται ένα πρόβλημα παλινδρόμησης, αντί για τη μέση τιμή μπορεί να χρησιμοποιηθεί και η διάμεσος συνδυάζοντας τις προβλέψεις [4].

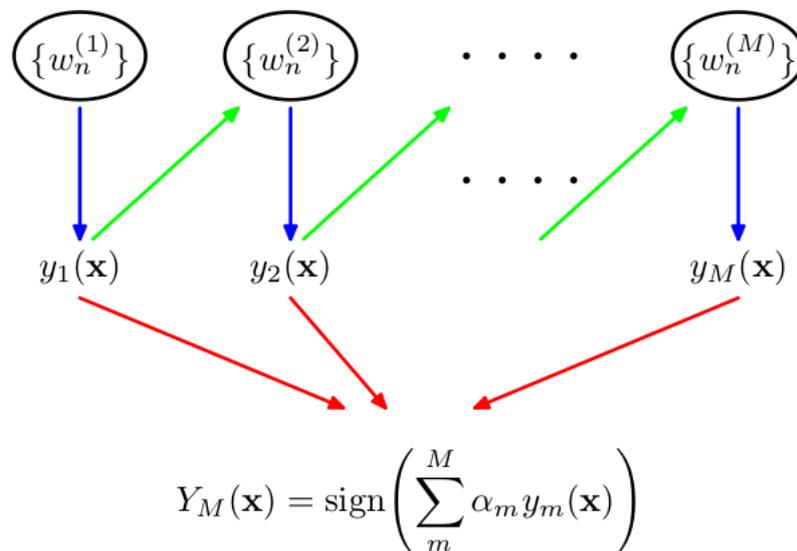
4.5 Μέθοδοι Boosting - «Ενίσχυσης»

Η μέθοδος *Boosting* είναι μία συνδυαστική μέθοδος που δημιουργεί ένα ισχυρό και όσο το δυνατόν αποδοτικό μοντέλο από έναν αριθμό από *weak learners* (βλ. *Παράρτημα, Ερμηνευτικοί Όροι*). Αυτό γίνεται δημιουργώντας ένα μοντέλο από τα δεδομένα εκπαίδευσης και στη συνέχεια ένα δεύτερο, που στοχεύει στη διόρθωση των σφαλμάτων του πρώτου. Αρχικά, η μέθοδος αυτή προοριζόταν για επίλυση προβλημάτων ταξινόμησης, ωστόσο επεκτείνεται και σε προβλήματα παλινδρόμησης [10].

Βασική ιδέα της μεθόδου αυτής είναι η επαναλαμβανόμενη εφαρμογή του μοντέλου στα δεδομένα εκπαίδευσης, προσαρμόζοντας βάρη για τις παρατηρήσεις σε κάθε βήμα.

Με αυτόν τον τρόπο το δέντρο απόφασης τροποποιείται διαδοχικά ώστε να εξαλειφθούν οι τυχαίες διακυμάνσεις του συνόλου εκπαίδευσης, ενώ χρησιμοποιώντας τα ίδια δεδομένα πολλές φορές με διαφορετικό βάρος κάθε φορά και εκπαιδεύοντας ξεχωριστά δέντρα απόφασης μικρού βάθους, προκύπτει τελικά ο βεβαρημένος μέσος όρος όλων των δέντρων. Για κάθε βήμα, οι εκτιμήσεις που απέχουν κατά πολύ από τις πραγματικές παρατηρήσεις αντιστοιχίζονται σε μεγαλύτερη τιμή βάρους στο επόμενο [14].

Η κύρια διαφορά μεταξύ των μεθόδων Bagging και Boosting είναι ότι στην πρώτη κάθε μοντέλο εκπαιδεύεται με ένα τυχαίο δείγμα από το σύνολο των δεδομένων εκπαίδευσης και εφαρμόζεται ανεξάρτητα από τα υπόλοιπα, ενώ η τελική εκτίμηση προκύπτει από το μέσο όρο των επιμέρους εκτιμήσεων. Αντίθετα στη μέθοδο Boosting, τα μοντέλα εκπαιδεύονται διαδοχικά και κάθε ένα από αυτά λαμβάνει υπόψη του μία σταθμισμένη μορφή των συνολικών δεδομένων εκπαίδευσης, όπου η τιμή του βάρους που αποδίδεται στον συντελεστή μίας παρατήρησης εξαρτάται από την απόδοση του προηγούμενου μοντέλου. Πιο συγκεκριμένα, παρατηρήσεις των οποίων οι εκτιμήσεις απέχουν αρκετά από την πραγματική τους τιμή λαμβάνουν μεγαλύτερη τιμή για το βάρος όταν χρησιμοποιούνται στην εκπαίδευση του επόμενου μοντέλου. Όταν πλέον έχουν εκπαιδευτεί όλα τα μοντέλα, οι εκτιμήσεις τους συνδυάζονται κατάλληλα με βάση τα τελικά τους βάρη. Μία απεικόνιση της λειτουργίας της μεθόδου Boosting αποτελεί η Εικόνα 4.2 .



Εικόνα 4.5 : Απεικόνιση της λειτουργίας της μεθόδου Boosting. Οι ποσότητες αποτελούν τα μοντέλα που εκπαιδεύονται με τη σταθμισμένη μορφή των δεδομένων (μπλε βέλη), όπου τα βάρη $w_m^{(n)}$ προκύπτουν από την απόδοση του προηγούμενου μοντέλου, $y_{m-1}(x)$ (πράσινα βέλη). Ο συνδυασμός των μοντέλων μετά το πέρας της εκπαίδευσής τους καταλήγει στο σχηματισμό του τελικού μοντέλου, $Y_m(x)$ (κόκκινα βέλη).

Δύο από τους πιο διαδεδομένους αλγορίθμους που εφαρμόζουν τη μέθοδο Boosting είναι οι :

- **AdaBoost (Adaptive Boosting)** (βλ. Παράγραφο 4.5.1)
- **Gradient Boosting** (βλ. Παράγραφο 4.5.2)

4.5.1 AdaBoost

Η *AdaBoost* (συντομία για Adaptive Boosting) αποτελεί μία μέθοδο Boosting που χρησιμοποιεί ένα σύνολο από weak learners (βλ. *Ερμηνευτικοί Όροι*) ώστε να κατασκευάσει ένα ισχυρό τελικό μοντέλο. Ειδικότερα, χρησιμοποιούνται δέντρα μικρού βάθους, δηλαδή με δύο ή τρία το πολύ επίπεδα, ενώ συνήθως αντί για δέντρα χρησιμοποιούνται ‘κορμοί’, δηλαδή ένας κόμβος και δύο φύλλα. Αυτός ο τύπος δέντρων δεν είναι πολύ ακριβής, καθώς κάνει χρήση μόνο μίας μεταβλητής για να προβεί σε μία εκτίμηση. Επιπλέον, καθώς η κατασκευή των δέντρων γίνεται διαδοχικά, το σφάλμα που υπολογίζεται σε κάθε ένα από αυτά επηρεάζει την κατασκευή του επόμενου. Για το λόγο αυτό, κάποια από τα δέντρα που δημιουργούνται συμβάλουν περισσότερο την τελική εκτίμηση της μεθόδου από ότι άλλα βάσει του σφάλματος που κάνουν.

Για παράδειγμα, την κατασκευή του δεύτερου δέντρου επηρεάζουν τα σφάλματα που υπολογίστηκαν από τη δημιουργία του πρώτου, επομένως αυτό μπορεί να περιέχει διαφορετικές μεταβλητές και οριακό σημείο διαχωρισμού από ότι το πρώτο. Στη συνέχεια, το μοντέλο ανανεώνεται και οδηγείται στην κατασκευή ενός ακόμα δέντρου σύμφωνα με το σφάλμα που σημειώθηκε στο προηγούμενο και η διαδικασία αυτή υλοποιείται διαδοχικά. Αυτό σημαίνει ότι τα ήδη υπάρχοντα δέντρα δεν τροποποιούνται καθώς το μοντέλο διευρύνεται και μόνο η προβλεπόμενη τιμή για τη μεταβλητή απόκρισης κάθε παρατήρησης επανεκτιμάται κάθε φορά, ώστε να αξιολογηθεί η συμβολή του νέου προστιθέμενου δέντρου.

Για προβλήματα παλινδρόμησης, ο αλγόριθμος που υλοποιείται, σύμφωνα με τη βιβλιοθήκη Scikit-learn της γλώσσας προγραμματισμού Python καλείται *AdaBoost.R2* [15].

Για την εφαρμογή του αλγορίθμου ορίζονται αρχικά βάρη, w_n , $n=1, \dots, N$, για τις παρατηρήσεις του συνόλου εκπαίδευσης, $(x_1, y_1), \dots, (x_N, y_N)$ με $x_n = (x_{n1}, \dots, x_{np}) \in \mathcal{X}$ όπου p ο αριθμός των επεξηγηματικών μεταβλητών και y_n οι παρατηρήσεις για τη μεταβλητή απόκρισης, κάθε ένα από τα οποία είναι ίσο με $1/N$,

όπου N ο συνολικός αριθμός των παρατηρήσεων. Η κάθε τιμή βάρους δηλώνει την επιρροή της παρατήρησης και χρησιμοποιείται στον υπολογισμό του σφάλματος της εκτίμησης που προκύπτει από το μοντέλο. Σε κάθε βήμα του αλγορίθμου, για κάθε νέο μοντέλο που εκπαιδεύει η AdaBoost, τα βάρη ανανεώνονται ανάλογα με την απόδοση του προηγούμενου μοντέλου, δηλώνοντας κατά πόσο η κάθε παρατήρηση έχει εκτιμηθεί σωστά και αν απέχει από την πραγματική τιμή να αποδοθεί μεγαλύτερη τιμή βάρους.

Η μέθοδος που χρησιμοποιείται στην AdaBoost.R2 εκφράζει το σφάλμα που αντιστοιχεί στην κάθε εκτίμηση μίας παρατήρησης σε σχέση με τη μεγαλύτερη τιμή σφάλματος, έτσι ώστε η συνάρτηση κόστους να λαμβάνει τιμές στο διάστημα $[0,1]$ [19]. Ο αλγόριθμος ακολουθεί :

AdaBoost.R2 Algorithm

- Ορισμός κατασκευής T συνολικών δέντρων.
- Αρχικοποίηση του μέσου σφάλματος $\bar{L} = 0$.
- Αρχικοποίηση των συντελεστών βάρους w_n , θέτοντας $w_n^{(1)} = \frac{1}{N}$, $n = 1, \dots, N$
- Για $t = 1, \dots, T$ ή $\bar{L} < 0.5$:
 - Κατασκευή μοντέλου παλινδρόμησης $f^{(t)} : x \rightarrow y$ και υπολογισμός της εκτιμώμενης τιμής $f^{(t)}(x_n)$, $n = 1, \dots, N$

- Υπολογισμός του σφάλματος κάθε παρατήρησης :

$$e_n^{(t)} = |f^{(t)}(x_n) - y_n|, n = 1, \dots, N \quad (4.7)$$

- Υπολογισμός της συνάρτησης κόστους $L_n^{(t)}$ για κάθε παρατήρηση :

$$L_n^{(t)} = L[|f^{(t)}(x_n) - y_n|], n = 1, \dots, N \quad (4.8)$$

- Υπολογισμός του μέσου σφάλματος, $\bar{L}^{(t)}$:

$$\bar{L}^{(t)} = \sum_{n=1}^N L_n^{(t)} w_n^{(t)} \quad (4.9)$$

- $\beta^{(t)} = \frac{\bar{L}^{(t)}}{1 - \bar{L}^{(t)}}$ (4.10)

- Ανανέωση της τιμής του βάρους :

$$w_n^{(t+1)} = \frac{w_n^{(t)} \beta^{(t)} [1 - L_n^{(t)}]}{Z^{(t)}}, Z^{(t)} \text{ όρος κανονικοποίησης} \quad (4.11)$$

- Υπολογισμός της σταθμισμένης διαμέσου, $f_{fin}(x)$, για τα $f^{(t)}(x)$, $t = 1, \dots, T$:

$$f_{fin}(x) = \inf \left[y \in Y : \sum_{t: f^{(t)}(x) \leq y} \log \frac{1}{\beta^{(t)}} \geq \frac{1}{2} \sum_t \log \frac{1}{\beta^{(t)}} \right] \quad (4.12)$$

Σύμφωνα με τον παραπάνω αλγόριθμο, τα βάρη που αντιστοιχούν στις παρατηρήσεις κατά την κατασκευή του πρώτου δέντρου είναι όλα ίσα και λαμβάνουν την τιμή $1/N$, όπου N το σύνολο των παρατηρήσεων. Η σχέση (4.7) περιγράφει τον υπολογισμό του σφάλματος μεταξύ της εκτιμώμενης τιμής που προέκυψε από το μοντέλο και της πραγματικής τιμής της παρατήρησης, που χρησιμοποιείται για τον υπολογισμό της συνάρτησης κόστους όπως φαίνεται από τη σχέση (4.8).

Η συνάρτηση κόστους μπορεί να έχει οποιαδήποτε μορφή αρκεί να παίρνει τιμές στο διάστημα $[0, 1]$. Συγκεκριμένα, για :

$$D = \sup |f^{(t)}(x_n) - y_n|, n = 1, \dots, N \quad (4.13)$$

τότε υπάρχουν τρεις υποψήφιες συναρτήσεις κόστους και είναι οι εξής :

- ♦ Γραμμική, με τύπο $L_n^{(t)} = \frac{|f^{(t)}(x_n) - y_n|}{D}$ (4.14)

- ♦ Τετραγωνική, με τύπο $L_n^{(t)} = \frac{|f^{(t)}(x_n) - y_n|^2}{D^2}$ (4.15)

- ♦ Εκθετική, με τύπο $L_n^{(t)} = 1 - \exp\left(\frac{-|f^{(t)}(x_n) - y_n|}{D}\right)$ (4.16)

Η προκαθορισμένη συνάρτηση κόστους της βιβλιοθήκης Scikit-learn της γλώσσας προγραμματισμού Python είναι η γραμμική, η οποία είναι και αυτή που χρησιμοποιείται στη συγκεκριμένη μελέτη.

Στη συνέχεια, η σχέση (4.9) δίνει τον τύπο για τον υπολογισμό του μέσου σφάλματος, ενώ μέσω της σχέσης (4.10) ορίζεται η ποσότητα $\beta^{(t)}$, η οποία αποτελεί ένα μέτρο εμπιστοσύνης για τις εκτιμήσεις του μοντέλου. Όσο πιο χαμηλές τιμές λαμβάνει, το μοντέλο αποδεικνύεται και πιο έμπιστο για τις εκτιμώμενες τιμές που προκύπτουν. Σύμφωνα με τους υπολογισμούς που προηγήθηκαν, οι τιμές για τα βάρη των παρατηρήσεων ανανεώνονται βάσει της σχέσης (4.11), με τον όρο $Z^{(t)}$ να είναι ένας

όρος κανονικοποίησης ώστε το σύνολο των βαρών όλων των παρατηρήσεων να αθροίζει στη μονάδα. Όσο πιο μικρή είναι τιμή του σφάλματος, τόσο μειώνεται το βάρος που αντιστοιχεί στην παρατήρηση. Τέλος, η σχέση (4.12) περιγράφει τη σταθμισμένη διάμεσο. Αφού από κάθε μοντέλο $f^{(t)}$ προκύπτει μία εκτίμηση $f^{(t)}(x_n)$ για την n -οστή παρατήρηση του συνόλου εκπαίδευσης και μία αντίστοιχη τιμή $\beta^{(t)}$, τότε αυτές ταξινομούνται ως $f^{(1)}(x_n) < f^{(2)}(x_n) < \dots < f^{(T)}(x_n)$ και αθροίζονται οι ποσότητες $\log\left(\frac{1}{\beta^{(t)}}\right)$ μέχρι να βρεθεί το μικρότερο t για το οποίο ισχύει η ανισότητα. Η εκτίμηση που προκύπτει από το μοντέλο t θεωρείται ως η συνολική πρόβλεψη, ενώ αν οι ποσότητες $\beta^{(t)}$ ήταν όλες ίσες για $t = 1, \dots, T$, τότε αυτή θα ήταν η διάμεσος [15].

4.5.2 Gradient Boosting

Η *Gradient Boosting* αποτελεί μία μέθοδο Boosting που στοχεύει στη διαδοχική μείωση των σφαλμάτων που προκύπτουν κατά την αλληπάλλληλη προσαρμογή μοντέλων. Δεδομένου ενός συνόλου εκπαίδευσης με N αριθμό παρατηρήσεων $(x_1, y_1), \dots, (x_N, y_N)$ με $x_n = (x_{n1}, \dots, x_{np}) \in \mathcal{X}$ όπου p ο αριθμός των επεξηγηματικών μεταβλητών και $n = 1, \dots, N$ και y_n η μεταβλητή απόκρισης, κατασκευάζεται αρχικά ένας κόμβος – και όχι ένα δέντρο – με μία πρώτη εκτίμηση της μεταβλητής απόκρισης που είναι ίσος με το μέσο όρο όλων των παρατηρήσεων. Στη συνέχεια, δημιουργούνται διαδοχικά weak learners – με τιμή για μέγιστο αριθμό φύλλων στο διάστημα [8, 32] – που με βάση τα σφάλματα που σημειώνονται στα προηγούμενα δέντρα ελαχιστοποιούν τη συνάρτηση κόστους, $L(y, f(x))$, με $f(x)$ να είναι η εκτιμώμενη τιμή της μεταβλητής απόκρισης.

Η μέθοδος Gradient Boosting στηρίζεται στη δημιουργία μοντέλων που εκτιμούν τις τιμές των υπολοίπων των παρατηρήσεων, $y_n - f(x_n)$ και όχι τις τιμές της μεταβλητής απόκρισης όπως η AdaBoost. Συγκεκριμένα, η AdaBoost υλοποιείται χρησιμοποιώντας βάρη για τις παρατηρήσεις που ανανεώνονται επαναληπτικά σε κάθε βήμα του αλγορίθμου, ενώ η Gradient Boosting χρησιμοποιεί ένα μοντέλο παλινδρόμησης που εκπαιδεύεται διαδοχικά στα υπόλοιπα. Με αυτόν τον τρόπο, κάθε νέο μοντέλο – weak learner – δημιουργείται λαμβάνοντας υπόψη τις παρατηρήσεις για τις οποίες οι τιμές των υπολοίπων είναι μεγάλες.

Ο ακόλουθος αλγόριθμος περιγράφει τα βήματα υλοποίησης του αλγορίθμου Gradient Boosting για παλινδρόμηση [16]. Σημειώνεται ότι η συνάρτηση κόστους είναι διαφορίσιμη και πως για προβλήματα παλινδρόμησης αυτή ορίζεται να είναι ίση με το μέσο τετραγωνικό σφάλμα (MSE : Mean Squared Error), $MSE(y, \hat{y}) = \frac{1}{N} \sum_{n=1}^N [y_n - \hat{y}_n]^2$, όπου $\hat{y}_n = f(x_n)$ και N ο συνολικός αριθμός παρατηρήσεων του συνόλου εκπαίδευσης.

Gradient Boosting Algorithm

▪ Αρχικοποίηση $f_0(x) = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, \gamma)$

▪ Για $m = 1, \dots, M$:

• Για $i = 1, \dots, N$:

$$r_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f=f_{m-1}} \quad (4.5)$$

• Δημιουργία δέντρου απόφασης παλινδρόμησης με μεταβλητή απόκρισης r_{im} και φύλλα $R_{jm}, j = 1, \dots, J_m$.

• Για $j = 1, \dots, J_m$:

$$\gamma_{jm} = \arg \min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma) \quad (4.6)$$

• $f_m(x) = f_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$ (4.7)

▪ $\hat{f}(x) = f_M(x)$ (4.8)

Στην πρώτη γραμμή του αλγορίθμου, αρχικοποιείται ένα σταθερό μοντέλο το οποίο δίνει μία πρώτη εκτιμώμενη τιμή της μεταβλητής απόκρισης ελαχιστοποιώντας το άθροισμα και αποτελεί έναν απλό τερματικό κόμβο, δηλαδή ένα φύλλο. Η ποσότητα που περιγράφεται από τη σχέση (4.5) είναι τα γενικευμένα ή ψευδό-υπόλοιπα που υπολογίζονται σε κάθε βήμα, για κάθε δέντρο m , όπου M το τελευταίο δέντρο και για όλες τις παρατηρήσεις. Στη συνέχεια, κατασκευάζεται ένας weak learner που δέχεται ως μεταβλητή απόκρισης την ποσότητα r_{im} με J_m συνολικά φύλλα. Η σχέση (4.6) αναπαριστά την εκτιμώμενη τιμή που βρίσκεται σε κάθε φύλλο του δέντρου που έχει δημιουργηθεί και η σχέση (4.7) αποτελεί την εκτίμηση που υπολογίζεται για κάθε παρατήρηση. Τέλος, η σχέση (4.8) δίνει τις τελικές εκτιμήσεις της μεταβλητής απόκρισης για όλες τις παρατηρήσεις, που προκύπτουν έπειτα από τη δημιουργία M δέντρων.

Με την παραπάνω διαδικασία, σε κάθε κατασκευή ενός νέου δέντρου οι εκτιμήσεις που προκύπτουν για τη μεταβλητή απόκρισης είναι αποτέλεσμα συνδυασμού της αρχικής εκτιμώμενης τιμής, $f_0(x)$ και της τιμής γ που προκύπτει από κάθε δέντρο πολλαπλασιασμένη με μία σταθερά v , η οποία λαμβάνει τιμές στο διάστημα $(0, 1)$. Η σταθερά αυτή ονομάζεται *ρυθμός εκμάθησης* (learning rate) και δηλώνει το βαθμό συνεισφοράς κάθε δέντρου στην εκτίμηση της τιμής της μεταβλητής απόκρισης.

Σύμφωνα με τα παραπάνω, φαίνεται πως οι ποσότητες v και M καθορίζουν τις τελικές εκτιμήσεις που θα προκύψουν από το σύνολο εκπαίδευσης. Ωστόσο, αυτές δεν δρουν

ανεξάρτητα η μία από την άλλη. Συγκεκριμένα, όσο πιο μικρές είναι οι τιμές που λαμβάνει ο ρυθμός εκμάθησης, οδηγούν σε μεγαλύτερες τιμές για την ποσότητα M , δηλαδή κατασκευάζονται περισσότερα δέντρα απόφασης.

Η ονομασία της μεθόδου Gradient Boosting προέκυψε από την Gradient Descent (βλ. Παράρτημα, *Gradient Descent*). Δεδομένης της συνάρτησης κόστους, η μέθοδος Gradient Descent χρησιμοποιείται για την εύρεση των παραμέτρων που την ελαχιστοποιούν. Σκοπός είναι η εύρεση των βέλτιστων παραμέτρων που έχουν ως αποτέλεσμα τη μεγαλύτερη μείωση της τιμής της συνάρτησης κόστους. Με αυτόν τον τρόπο η μέθοδος Gradient Boosting επιχειρεί να ελαχιστοποιήσει το σφάλμα. Καταληκτικά, μειώνοντας διαδοχικά το σφάλμα, το μοντέλο βελτιώνεται σταδιακά μέχρι να βρεθεί η τελική εκτίμηση για τις παρατηρήσεις.

5 Νευρωνικά Δίκτυα

5.1 Εισαγωγή στα Νευρωνικά Δίκτυα

Ένα τεχνητό νευρωνικό δίκτυο αποτελεί ένα υπολογιστικό μοντέλο επηρεασμένο από τη δομή και τη λειτουργία των νευρωνικών δικτύων που βρίσκονται στον ανθρώπινο εγκέφαλο. Ο εγκέφαλος είναι μία μηχανή επεξεργασίας πληροφοριών με απίστευτες ικανότητες που σε πολλούς τομείς ξεπερνά τα διάφορα μηχανικά προϊόντα, όπως για παράδειγμα στην όραση ή την αναγνώριση φωνής. Αυτές οι λειτουργίες αποδεικνύονται χρήσιμες αν εφαρμοστούν σε μηχανές, επομένως αν ο τρόπος εκτέλεσής τους από τον εγκέφαλο γίνει κατανοητός, μπορούν να οριστούν αλγόριθμοι που τους εξηγούν και να εφαρμοστούν σε υπολογιστές.

Ο ανθρώπινος εγκέφαλος είναι πολύ διαφορετικός από έναν ηλεκτρονικό υπολογιστή. Ενώ ο τελευταίος λειτουργεί με έναν επεξεργαστή, ο εγκέφαλος αποτελείται από έναν πολύ μεγάλο αριθμό μονάδων επεξεργασίας, τους νευρώνες, οι οποίοι μέσω διάφορων συνδέσεων, τις συνάψεις, συνδέονται με άλλους νευρώνες και λειτουργούν παράλληλα. Σε έναν υπολογιστή ο επεξεργαστής είναι ενεργός, ενώ η μνήμη του βρίσκεται ξεχωριστά από αυτόν και είναι αδρανής. Αντίθετα, θεωρείται ότι στον ανθρώπινο εγκέφαλο τόσο ο επεξεργαστής, όσο και η μνήμη διανέμονται μαζί στο δίκτυο που σχηματίζεται, με τους νευρώνες να υλοποιούν την επεξεργασία και τη μνήμη να βρίσκεται στις συνάψεις μεταξύ των νευρώνων [4].

Ένα τεχνητό νευρωνικό δίκτυο μπορεί να περιγραφεί ως ένας γράφος, οι κόμβοι του οποίου αντιστοιχούν στους νευρώνες και οι ακμές του στις συνάψεις που συνδέουν τους νευρώνες μεταξύ τους. Κάθε νευρώνας δέχεται ως τιμή το σταθμισμένο άθροισμα του αποτελέσματος των νευρώνων από τις εισερχόμενες ακμές του [5].

Σύμφωνα με τα παραπάνω, ένα τεχνητό νευρωνικό δίκτυο, ή απλά νευρωνικό δίκτυο, αποτελεί μία προσπάθεια προσομοίωσης του δικτύου που σχηματίζουν οι νευρώνες του ανθρώπινου εγκεφάλου, ώστε να μπορέσει ένας υπολογιστής να εκπαιδευτεί και να λάβει αποφάσεις όπως οι άνθρωποι. Έτσι, τα νευρωνικά δίκτυα κατασκευάζονται προγραμματίζοντας υπολογιστές να λειτουργούν σαν να ήταν συνδεδεμένα εγκεφαλικά κύτταρα.

Συνοψίζοντας, ένα νευρωνικό δίκτυο ορίζεται ως ένα υπολογιστικό σύστημα που αποτελείται από ένα πλήθος απλών, αλλά με μεγάλο βαθμό διασύνδεσης, στοιχείων ή κόμβων που ονομάζονται 'νευρώνες' και είναι διατεταγμένοι σε στρώματα που επεξεργάζονται πληροφορίες. Με αυτόν τον τρόπο, μπορούν να αναγνωριστούν πρότυπα που είναι αρκετά σύνθετα ώστε να εξαχθούν αποτελέσματα με το «χέρι» και εκπαιδεύονται με σκοπό να ανιχνεύονται από τη μηχανή.

Ένα νευρωνικό δίκτυο μπορεί να εφαρμοστεί τόσο σε προβλήματα παλινδρόμησης, όσο και σε προβλήματα ταξινόμησης. Για προβλήματα παλινδρόμησης, συνήθως υπάρχει μόνο μία μεταβλητή απόκρισης, ωστόσο τέτοια συστήματα μπορούν να χειριστούν και μεταβλητές απόκρισης μεγαλύτερης διάστασης [16].

5.1.1 Εφαρμογές των Νευρωνικών Δικτύων

Τα νευρωνικά δίκτυα αποτελούν μία ευρέως διαδεδομένη τεχνική επίλυσης προβλημάτων και έχει πολλές εφαρμογές που περιλαμβάνουν την αναγνώριση προτύπων και τη λήψη αποφάσεων βάσει αυτών. Για παράδειγμα, στα αεροπλάνα ένα νευρωνικό δίκτυο μπορεί να χρησιμοποιηθεί ως αυτόματος πιλότος, με τους νευρώνες να λαμβάνουν αναγνωριστικά σήματα από διάφορα όργανα του πιλοτηρίου ώστε να καταλήγουν στην κατάλληλη τροποποίηση των χειριστηρίων του αεροσκάφους για την ασφαλή διατήρηση της πορείας του. Επιπλέον, χρησιμοποιείται σε πολλά εργοστάσια για έλεγχο ποιότητας των προϊόντων, ενώ μπορεί να χρησιμοποιηθεί και ως μέτρο ασφάλειας για την ανίχνευση ύποπτων συναλλαγών σε τράπεζα.

Επιπρόσθετα, τα νευρωνικά δίκτυα μπορούν να φανούν χρήσιμα σε προβλέψεις της χρηματιστηριακής αγοράς, ενώ λειτουργούν και ως συστήματα ανίχνευσης που αναγνωρίζουν αυτόματα τυχόν εχθρικά αεροσκάφη ή πλοία και ακόμη μπορούν να συμβάλουν στην ιατρική κάνοντας διαγνώσεις σύνθετων ασθενειών βάσει των συμπτωμάτων τους. Τέλος, αποτέλεσμα της εφαρμογής τους είναι και η αναγνώριση λέξεων κατά τη σύνταξη προτάσεων σε μία οθόνη κινητού, προβλέποντας τους χαρακτήρες που πρόκειται να χρησιμοποιήσει ο χρήστης, η αναγνώριση φωνής και ανεπιθύμητης ηλεκτρονικής αλληλογραφίας [17].

5.2 Αρχιτεκτονική Νευρωνικού Δικτύου

Ένα νευρωνικό δίκτυο αποτελεί ένα σύστημα αλληλοσυνδεόμενων νευρώνων που είναι ομαδοποιημένοι σε στρώματα. Συγκεκριμένα, το πρώτο και το τελευταίο στρώμα του δικτύου ονομάζεται 'στρώμα εισόδου' (input layer) και 'στρώμα εξόδου' (output layer) αντίστοιχα, ενώ τα υπόλοιπα στρώματα ονομάζονται 'κρυφά στρώματα' (hidden layers). Σε κάθε νευρώνα αντιστοιχίζεται και μία παράμετρος βάρους, ενώ σε κάθε στρώμα, εκτός από το εξωτερικό, προστίθεται και η παράμετρος της μεροληψίας (bias). Το στρώμα εισόδου, εισάγει τα δεδομένα στο σύστημα για περαιτέρω επεξεργασία από τα ακόλουθα στρώματα νευρώνων και έχει ένα νευρώνα για κάθε επεξηγηματική μεταβλητή. Στα ενδιάμεσα στρώματα, κρυφά στρώματα, οι νευρώνες λαμβάνουν κάποια βάρη και εξάγουν ένα μετασχηματισμένο αποτέλεσμα μέσω μιας συνάρτησης ενεργοποίησης ώστε να μπορεί να χρησιμοποιηθεί από το τελευταίο στρώμα, στρώμα εξόδου, το οποίο δίνει τα τελικά αποτελέσματα της διαδικασίας.

Υπάρχουν πολλές διαφορετικές δομές νευρωνικών δικτύων ανάλογα με τον αριθμό των κρυμμένων στρωμάτων και το πλήθος των νευρώνων σε κάθε στρώμα. Ένα νευρωνικό δίκτυο που αποτελείται από ένα στρώμα, δηλαδή αποτελεί ένα νευρώνα, ονομάζεται *Perceptron* και ένας πολυεπίπεδος *Perceptron*, με περισσότερα από ένα στρώματα δηλαδή, είναι ένα νευρωνικό δίκτυο.

5.3 Perceptron

Ο *Perceptron* είναι ένας νευρώνας που αποτελεί το βασικό στοιχείο επεξεργασίας του νευρωνικού δικτύου, ο οποίος δέχεται δυαδικές μεταβλητές x_m , $m = 1, \dots, M$ και παράγει μία μοναδική δυαδική τιμή. Σε κάθε τιμή των μεταβλητών αντιστοιχίζεται ένας συντελεστής βάρους, w_m , $m = 1, \dots, M$, ο οποίος είναι ένας πραγματικός αριθμός και εκφράζει τη σημασία που έχει η τιμή αυτή στον υπολογισμό του αποτελέσματος.

Το τελικό αποτέλεσμα ενός *Perceptron*, 0 ή 1, καθορίζεται από το αν το σταθμισμένο άθροισμα $\sum_{m=1}^M w_m x_m$ είναι μεγαλύτερο ή μικρότερο από μία οριακή τιμή, η οποία είναι ένας πραγματικός αριθμός, όπως τα βάρη και είναι μία παράμετρος του νευρώνα. Ως οριακή τιμή ορίζεται μία παράμετρος μεροληψίας (bias), b , η οποία ισούται με τη μονάδα. Επιπλέον, το άθροισμα $\sum_{m=1}^M w_m x_m$ μπορεί να γραφεί ως γινόμενο δύο διανυσμάτων, δηλαδή $w^T x = \sum_{m=1}^M w_m x_m$, όπου στα διανύσματα w και x περιέχονται οι τιμές για τα βάρη και τις μεταβλητές αντίστοιχα.

Θεωρώντας μία συνάρτηση :

$$f(a) = \begin{cases} 1, & \text{αν } a > 0 \\ 0, & \text{διαφορετικά} \end{cases} \quad (5.1)$$

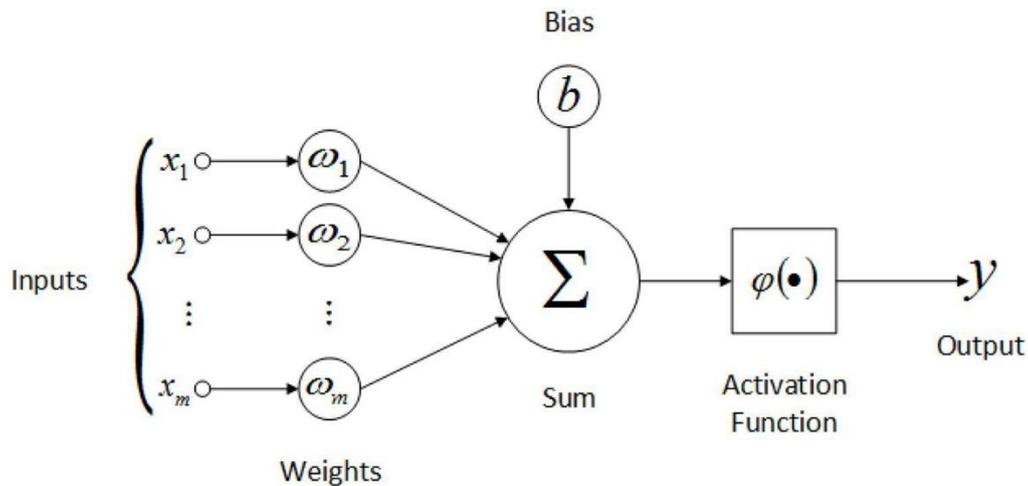
Τότε το αποτέλεσμα που προκύπτει από τον *Perceptron* υπολογίζεται βάσει της ακόλουθης σχέσης :

$$y = \begin{cases} 1, & \text{αν } f(w^T x + b) > 0 \\ 0, & \text{διαφορετικά} \end{cases} \quad (5.2)$$

Εισάγοντας την παράμετρο μεροληψίας στο διάνυσμα του βάρους με $w = [b, w_1, \dots, w_m]^T$ και $x = [1, x_1, \dots, x_m]^T$ η σχέση (5.2) μπορεί να γραφεί ως [4,18] :

$$y = \begin{cases} 1, & \text{αν } f(w^T x) > 0 \\ 0, & \text{διαφορετικά} \end{cases} \quad (5.3)$$

Η Εικόνα 5.1 περιγράφει τη λειτουργία ενός *Perceptron* όπως περιγράφηκε παραπάνω :



Εικόνα 5.6 : Περιγραφή λειτουργίας του Perceptron.

5.4 Συνάρτηση Ενεργοποίησης

Η συνάρτηση ενεργοποίησης εφαρμόζεται σε κάθε κόμβο που ανήκει σε κρυφό και εξωτερικό στρώμα του νευρωνικού δικτύου. Αυτή μπορεί να είναι είτε γραμμική, δηλαδή η ταυτοτική συνάρτηση, είτε μη γραμμική. Ωστόσο, τις περισσότερες φορές δεν είναι γραμμική και αυτός είναι ο λόγος που χρησιμοποιείται σε ένα νευρωνικό δίκτυο. Συγκεκριμένα, στόχος της είναι ο μη γραμμικός μετασχηματισμός των εισερχόμενων τιμών των δεδομένων σε άλλες, οι οποίες εν συνεχεία εισάγονται στα επόμενα στρώματα νευρώνων του δικτύου για να γίνουν νέοι υπολογισμοί μέχρι την τελική εκτίμηση.

Όταν δεν υπάρχει η συνάρτηση ενεργοποίησης, τότε τα βάρη και η παράμετρος μεροληψίας οδηγούν σε ένα γραμμικό μετασχηματισμό. Μία γραμμική εξίσωση είναι εύκολο να λυθεί, ωστόσο είναι περιορισμένων δυνατοτήτων για επίλυση πιο σύνθετων προβλημάτων. Επομένως, δίχως την συνάρτηση ενεργοποίησης ένα νευρωνικό δίκτυο αποτελεί ουσιαστικά ένα γραμμικό μοντέλο παλινδρόμησης. Η συνάρτηση ενεργοποίησης μετασχηματίζει με μη γραμμικό τρόπο τα εισερχόμενα δεδομένα, με αποτέλεσμα να επιτρέπει στο νευρωνικό δίκτυο να εκπαιδευτεί και να εκτελέσει πιο σύνθετες εργασίες, όπως για παράδειγμα γλωσσικές μεταφράσεις και ταξινομήσεις εικόνων που οι γραμμικοί μετασχηματισμοί καθιστούν αδύνατο να υλοποιηθούν.

Υπάρχουν πολλές διαφορετικές συναρτήσεις ενεργοποίησης (βλ. Παράρτημα, Διαφορετικοί Τύποι Συναρτήσεων Ενεργοποίησης), όμως η επιλογή της κατάλληλης εξαρτάται από τη φύση των δεδομένων. Για προβλήματα παλινδρόμησης, η συνάρτηση

ενεργοποίησης που χρησιμοποιείται είναι η γραμμική. Τέλος, σε ένα νευρωνικό δίκτυο μπορούν να υπάρξουν διαφορετικές συναρτήσεις ενεργοποίησης που εφαρμόζονται και κάθε στρώμα.

5.5 Feedforward Propagation – Εμπροσθοτροφοδοτούμενο Νευρωνικό Δίκτυο

Τα γραμμικά μοντέλα για προβλήματα ταξινόμησης και παλινδρόμησης βασίζονται σε γραμμικούς συνδυασμούς από βασικές μη γραμμικές συναρτήσεις $\phi_j(x)$ [10] και έχουν τη μορφή :

$$y(\mathbf{x}, \mathbf{w}) = f \left(\sum_{j=1}^M w_j \phi_j(x) \right) \quad (5.4)$$

όπου η συνάρτηση $f(\cdot)$ είναι μία μη γραμμική συνάρτηση ενεργοποίησης για προβλήματα ταξινόμησης και η ταυτοτική για προβλήματα παλινδρόμησης. Σκοπός είναι η επέκταση αυτού του μοντέλου κατασκευάζοντας συναρτήσεις $\phi_j(x)$ που εξαρτώνται από τις παραμέτρους, οι οποίες προσαρμόζονται μαζί με τους συντελεστές βάρους, w_j κατά την εκπαίδευση. Τα νευρωνικά δίκτυα χρησιμοποιούν συναρτήσεις όπως αυτή που περιγράφεται από τη σχέση (5.4), με κάθε μία από αυτές να είναι μία μη γραμμική συνάρτηση ενός γραμμικού συνδυασμού των εισερχόμενων παρατηρήσεων με τους συντελεστές να είναι οι προσαρμοσμένες παράμετροι.

Σύμφωνα με τα παραπάνω, δημιουργείται ένα βασικό νευρωνικό δίκτυο, στο οποίο οι εισερχόμενες τιμές των παρατηρήσεων επεξεργάζονται μέσα από μία σειρά συναρτησιακών μετασχηματισμών μεταξύ των στρωμάτων, ώστε να υπολογιστεί η τελική εκτίμηση.

Έστω ότι διατίθεται ένα σύνολο δεδομένων που αποτελείται από p επεξηγηματικές μεταβλητές και μία μεταβλητή απόκρισης, για κάθε μία από τις οποίες υπάρχει N αριθμός παρατηρήσεων, δηλαδή $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)$ με $\mathbf{x}_n = (x_{n1}, \dots, x_{np}) \in \mathcal{X}$ και $n = 1, \dots, N$. Στο πρώτο στρώμα του νευρωνικού δικτύου, στρώμα εισόδου, το οποίο αποτελείται από τόσους νευρώνες όσες είναι και οι επεξηγηματικές μεταβλητές, εισάγεται κάθε φορά ένα δείγμα από τα δεδομένα. Κάθε νευρώνας του στρώματος αυτού συνδέεται με άλλους του επόμενου στρώματος, καθώς κατασκευάζεται ένα πλήρως συνδεδεμένο δίκτυο και κάθε μία από αυτές τις συνδέσεις μεταξύ των νευρώνων συνοδεύεται και από μία τιμή βάρους. Επιπλέον, σε κάθε στρώμα, εκτός από το τελευταίο, προστίθεται και μία τιμή μεροληψίας, η οποία είναι συνήθως ίση με τη μονάδα και ενισχύει στη γενίκευση του μοντέλου.

Το βάρος της σύναψης μεταξύ δύο νευρώνων συμβολίζεται ως w_{jk}^l , όπου το l δηλώνει τον αριθμό του στρώματος, το j τον αριθμό του νευρώνα στο στρώμα l και το k τον αριθμό του νευρώνα στο στρώμα $l+1$, ενώ η μεροληψία συμβολίζεται ως w_{0k}^l . Σύμφωνα με τα παραπάνω, η τιμή στον κάθε νευρώνα του πρώτου κρυφού στρώματος υπολογίζεται από την ακόλουθη σχέση :

$$z_k^{(2)} = h(a_k^{(2)}) \quad (5.5)$$

με

$$a_k^{(2)} = \sum_{j=1}^p w_{jk}^{(1)} x_j + w_{0k}^{(1)} \quad (5.6)$$

Η ποσότητα της σχέσης (5.6), $a_k^{(2)}$, με $k = 1, \dots, K$ και K ο συνολικός αριθμός νευρώνων στο πρώτο κρυφό στρώμα και το (2) στον εκθέτη να δηλώνει το δεύτερο στρώμα του νευρωνικού δικτύου, καλείται *ενεργοποιητής* [10] και αποτελεί γραμμικό συνδυασμό των εισερχόμενων επεξηγηματικών μεταβλητών x_1, \dots, x_p . Η τελική τιμή του κάθε νευρώνα, $z_k^{(2)}$, διαμορφώνεται μέσα από τον μετασχηματισμό των ποσοτήτων $a_k^{(2)}$ με μία διαφορίσιμη, μη γραμμική συνάρτηση ενεργοποίησης $h(\cdot)$ που δίνεται από τη σχέση (5.5).

Με όμοιο τρόπο υπολογίζονται και οι τιμές για τους νευρώνες στο δεύτερο κρυφό στρώμα του δικτύου. Δηλαδή :

$$z_m^{(3)} = h(a_m^{(3)}) \quad (5.7)$$

με

$$a_m^{(3)} = \sum_{k=1}^K w_{km}^{(2)} z_k^{(2)} + w_{0m}^{(2)} \quad (5.8)$$

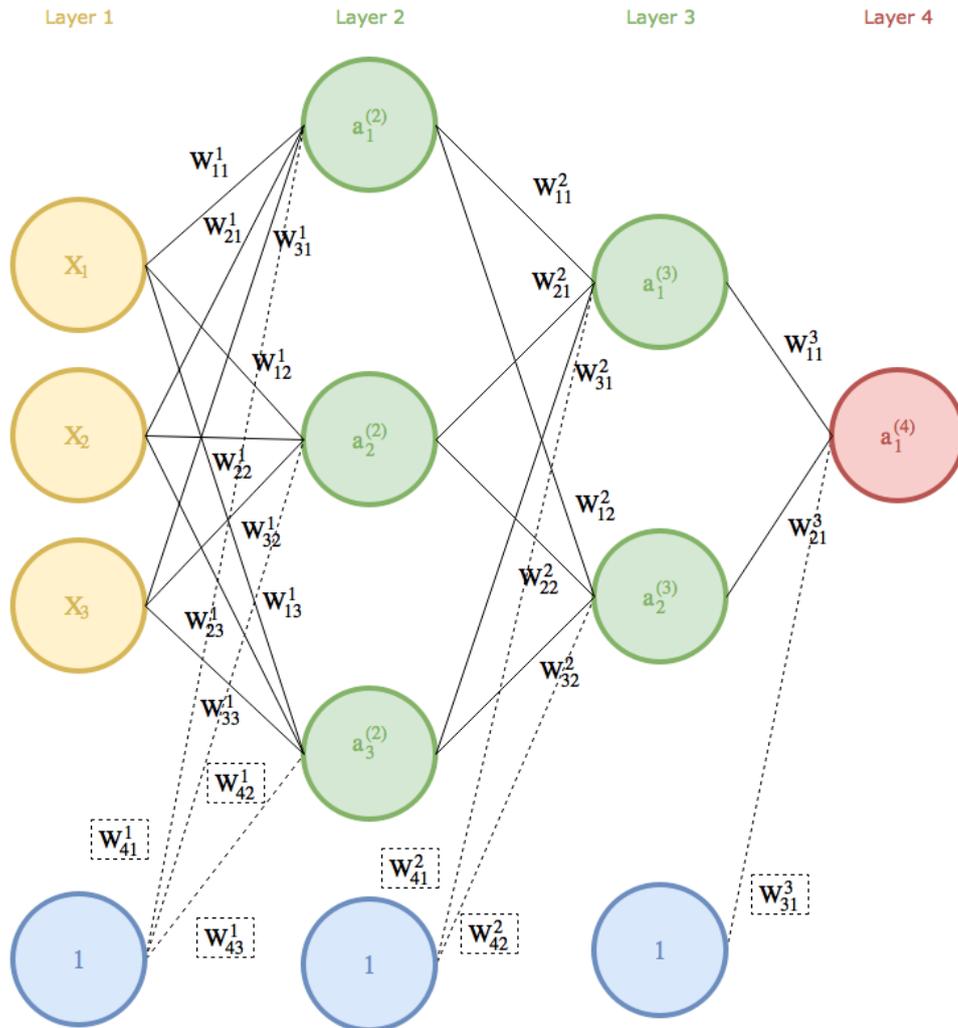
όπου $m = 1, \dots, M$ με M ο συνολικός αριθμός νευρώνων στο δεύτερο κρυφό στρώμα και $w_{0m}^{(2)}$ η παράμετρος μεροληψίας. Με ανάλογη διαδικασία υπολογίζονται και τα αποτελέσματα στους νευρώνες στα υπόλοιπα κρυφά στρώματα του δικτύου.

Στο τελευταίο στρώμα του δικτύου, στρώμα εξόδου, χρησιμοποιείται μία κατάλληλη συνάρτηση ενεργοποίησης η οποία μετασχηματίζει τους ενεργοποιητές που υπολογίζονται και προκύπτει το τελικό αποτέλεσμα για την εκτίμηση της μεταβλητής απόκρισης y .

Η παράμετρος μεροληψίας μπορεί να ενσωματωθεί στο διάνυσμα των βαρών προσθέτοντας μία επιπλέον μεταβλητή x_0 στο διάνυσμα επεξηγηματικών μεταβλητών, για την οποία όλες οι τιμές είναι ίσες με τη μονάδα. Έτσι η σχέση (5.6) μπορεί να γραφεί ως :

$$a_k^{(2)} = \sum_{j=0}^p w_{jk}^{(1)} x_j \quad (5.9)$$

Στην Εικόνα 5.2 αναπαρίσταται η παραπάνω διαδικασία για ένα νευρωνικό δίκτυο με δύο κρυφά στρώματα.



Εικόνα 5.2 : Αναπαράσταση νευρωνικού δικτύου με δύο κρυφά στρώματα (πράσινοι κόμβοι). Το πρώτο στρώμα (κίτρινοι κόμβοι) είναι το στρώμα εισόδου με τις εισαγόμενες μεταβλητές και το τελευταίο στρώμα (κόκκινος κόμβος) αποτελεί το στρώμα εξόδου που αποδίδεται η εκτίμηση της μεταβλητής απόκρισης. Τα βάρη αναπαρίστανται ως οι συνδέσεις μεταξύ των κόμβων και οι μπλε κόμβοι αντιπροσωπεύουν τις παραμέτρους αμεροληψίας.

5.6 Backpropagation

Ο τεχνική *Backpropagation* είναι ένας μηχανισμός ανανέωσης των τιμών για τα βάρη μεταξύ των στρωμάτων του νευρωνικού δικτύου, ώστε οι εκτιμώμενες τιμές που προκύπτουν να μην έχουν μεγάλη απόκλιση από τις πραγματικές τιμές. Σκοπός δηλαδή είναι η μείωση του σφάλματος για τα αποτελέσματα σε κάθε νευρώνα αλλά και στο συνολικό δίκτυο. Αυτό επιτυγχάνεται προσαρμόζοντας τα βάρη έτσι ώστε η συνάρτηση κόστους να ελαχιστοποιείται χρησιμοποιώντας τη μέθοδο Gradient Descent (βλ. Παράρτημα, *Gradient Descent*). Με άλλα λόγια, για να ελαχιστοποιηθεί η συνάρτηση κόστους, πρέπει να υπολογιστεί η επιρροή που δέχεται από τα βάρη. Ο υπολογισμός αυτός γίνεται με κατεύθυνση από το τελευταίο στρώμα, στρώμα εξόδου, προς το πρώτο στρώμα, στρώμα εισόδου.

Μετά τον υπολογισμό της εκτίμησης που προκύπτει από την εισαγωγή και επεξεργασία μίας παρατήρησης από το νευρωνικό δίκτυο, προσδιορίζεται η ποσότητα $\delta_k = \hat{y}_k - y_k$, όπου \hat{y}_k η εκτιμώμενη τιμή του εξωτερικού νευρώνα και περιγράφει το σφάλμα της εκτίμησης. Στη συνέχεια, υπολογίζεται η ποσότητα

$$\delta_j = h'(a_j) \sum_k w_{jk} \delta_k \quad (5.10)$$

η οποία δηλώνει ότι η τιμή του σφάλματος στο συγκεκριμένο νευρώνα υπολογίζεται μέσω των σφαλμάτων νευρώνων που βρίσκονται σε επόμενα στρώματα του δικτύου, όπως φαίνεται και στην Εικόνα 5.3, για κάθε νευρώνα που βρίσκεται σε κρυφό στρώμα. Έπειτα, η ακόλουθη σχέση

$$\frac{\partial E_n}{\partial w_{ij}} = \delta_j z_i \quad (5.11)$$

περιγράφει ότι η ζητούμενη παράγωγος της συνάρτησης κόστους, E_n , για μία παρατήρηση των δεδομένων, προκύπτει από τον πολλαπλασιασμό της τιμής του σφάλματος για το νευρώνα στον οποίο καταλήγει η διεύθυνση του βάρους, με την τιμή του νευρώνα από τον οποίο ξεκινά [10]. Επομένως, χρειάζεται μόνο να υπολογιστεί η τιμή δ_j για κάθε νευρώνα στα κρυφά και το εξωτερικό στρώμα του νευρωνικού δικτύου, ώστε να εφαρμοστεί η τελευταία σχέση.

Η παραπάνω διαδικασία συμβάλει στην ανανέωση των τιμών του βάρους του νευρωνικού δικτύου για την ελαχιστοποίηση του σφάλματος της συνάρτησης κόστους και καλείται *backpropagation*.

Όταν τα δεδομένα χωρίζεται σε μικρότερα σύνολα, *batches*, τότε όλοι αυτοί οι υπολογισμοί αποθηκεύονται και η ανανέωση γίνεται μετά την προσπέλαση όλων των παρατηρήσεων. Δηλαδή η παράγωγος του συνολικού σφάλματος, E , υπολογίζεται επαναλαμβάνοντας τα παραπάνω για κάθε παρατήρηση των δεδομένων και αθροίζοντας για το σύνολο αυτών :

$$\frac{\partial E}{\partial w_{ij}} = \sum_n \frac{\partial E_n}{\partial w_{ij}} \quad (5.12)$$

Ένα πλήρες ‘πέραςμα’ όλων των παρατηρήσεων του συνόλου δεδομένων αποτελεί μία εποχή, *epoch*.

5.7 Διαμόρφωση Υπερπαραμέτρων Νευρωνικού Δικτύου

Για τη δημιουργία ενός νευρωνικού δικτύου είναι απαραίτητος ο ορισμός συγκεκριμένων μεταβλητών. Οι μεταβλητές αυτές ονομάζονται *υπερπαραμέτροι* και καθορίζουν τόσο τη δομή του δικτύου, όσο και την εκπαίδευσή του και ρυθμίζονται πριν την έναρξη της εκπαίδευσης. Η περιγραφή τους ακολουθεί :

- ▶ Αριθμός κρυφών στρωμάτων του δικτύου και πλήθος νευρώνων σε κάθε ένα από αυτά

Κατά την κατασκευή ενός νευρωνικού δικτύου δεν υπάρχει κάποια συγκεκριμένη μεθοδολογία που ακολουθείται για τον καθορισμό του καλύτερου αριθμού κρυφών στρωμάτων και νευρώνων σε αυτά. Η πιο καλή προσέγγιση είναι η εκπαίδευση του μοντέλου με ένα μικρό αρχικό αριθμό νευρώνων για το πρώτο κρυφό στρώμα του δικτύου και η διαδοχική αύξηση των νευρώνων και των στρωμάτων μέχρι να βρεθεί το βέλτιστο αποτέλεσμα, χωρίς να επηρεάζεται η γενικότητα του μοντέλου. Δηλαδή η αρχιτεκτονική του μοντέλου που δίνει το βέλτιστο αποτέλεσμα προκύπτει από το συνεχή πειραματισμό για το συνδυασμό του αριθμού νευρώνων και κρυφών στρωμάτων, υπολογίζοντας κάθε φορά το σφάλμα που προκύπτει.

Γενικότερα, αν ο αριθμός των νευρώνων στα κρυφά στρώματα του δικτύου είναι πολύ μικρός, τότε προκύπτει υψηλό σφάλμα εκπαίδευσης και γενίκευσης λόγω ανεπαρκούς εκπαίδευσης. Αντίστοιχα, ο μεγάλος αριθμός νευρώνων οδηγεί σε ελαχιστοποίηση του σφάλματος εκπαίδευσης, ωστόσο το σφάλμα γενίκευσης του μοντέλου παραμένει υψηλό εξαιτίας της υπερεκπαίδευσης και της μεγάλης διασποράς.

- ▶ Συνάρτηση ενεργοποίησης

Οι διάφορες συναρτήσεις ενεργοποίησης (βλ. Παράρτημα, Διαφορετικοί Τύπου Συναρτήσεων Ενεργοποίησης), όπως περιγράφονται στο κεφάλαιο 5.4, είναι συνήθως μη γραμμικές συναρτήσεις και εισάγονται στα μοντέλα των νευρωνικών δικτύων καθώς τους επιτρέπουν την εκπαίδευση σε σύνθετα προβλήματα.

Όταν ένα νευρωνικό δίκτυο αποτελείται από πολλά διαφορετικά στρώματα, είναι δυνατό σε κάθε ένα από αυτά να εφαρμόζεται διαφορετική συνάρτηση ενεργοποίησης. Συνήθως, σε προβλήματα παλινδρόμησης, όπως είναι και αυτό με το οποίο ασχολείται αυτή η μελέτη, η συνάρτηση ενεργοποίησης του τελευταίου στρώματος του δικτύου είναι η γραμμική.

► Συνάρτηση κόστους

Η συνάρτηση κόστους εκφράζει τη διαφορά μεταξύ των εκτιμώμενων τιμών που προβλέπονται από το μοντέλο και των πραγματικών παρατηρούμενων τιμών. Για προβλήματα παλινδρόμησης, ως συνάρτηση κόστους ορίζεται το μέσο τετραγωνικό σφάλμα, MSE , που περιγράφεται από την ακόλουθη σχέση :

$$MSE = \frac{1}{N} \sum_{n=1}^N (\hat{y}_n - y_n)^2 \quad (5.13)$$

όπου \hat{y}_n η προβλεπόμενη τιμή του μοντέλου για την n-οστή παρατήρηση.

► Μέθοδος βελτιστοποίησης – Optimizer

Οι μέθοδοι βελτιστοποίησης βασίζονται σε αλγόριθμους που στοχεύουν στην ελαχιστοποίηση της συνάρτησης κόστους. Συγκεκριμένα, συνδέουν τη συνάρτηση κόστους και τις παραμέτρους του μοντέλου ανανεώνοντάς το ανάλογα με το αποτέλεσμα της συνάρτησης.

Στη συγκεκριμένη μελέτη επιλέγεται η μέθοδος *Adam*, καθώς είναι γρήγορη και συγκλίνει σύντομα, ενώ διορθώνει και την υψηλή διασπορά – διακύμανση.

► Ρυθμός εκμάθησης – Learning rate

Ο ρυθμός εκμάθησης αποτελεί μία ρυθμιστική παράμετρο που ορίζει πόσο γρήγορα ανανεώνονται τα βάρη με βάση τον επιλεγμένο αλγόριθμο και λαμβάνει τιμές μεταξύ του μηδενός και της μονάδας. Η επιλογή της αριθμητικής τιμής του ρυθμού εκμάθησης παίζει σπουδαίο ρόλο στην εφαρμογή του μοντέλου, καθώς οι μικρές τιμές επιβραδύνουν τη διαδικασία εκπαίδευσης, ενώ με μεγάλες τιμές ο αλγόριθμος υλοποιείται γρήγορα με αποτέλεσμα να μπορεί να προσπεράσει το βέλτιστο σημείο ελαχίστου και να μην συγκλίνει.

► Μέγεθος υποσυνόλου των δεδομένων εκπαίδευσης – Batch size

Το μέγεθος των υποσυνόλων των δεδομένων εκπαίδευσης, *batch size*, αναφέρεται στον αριθμό των δειγμάτων που εισέρχονται στο νευρωνικό δίκτυο πριν την ανανέωση των εσωτερικών του παραμέτρων, δηλαδή των βαρών μεταξύ των νευρώνων.

► Αριθμός εποχών εκπαίδευσης – Epochs

Ο αριθμός των εποχών αποτελεί τον αριθμό που το σύνολο των δεδομένων εκπαίδευσης εισέρχονται στο νευρωνικό δίκτυο. Μία εποχή μπορεί να αποτελείται από ένα ή περισσότερα υποσύνολα των δεδομένων εκπαίδευσης.

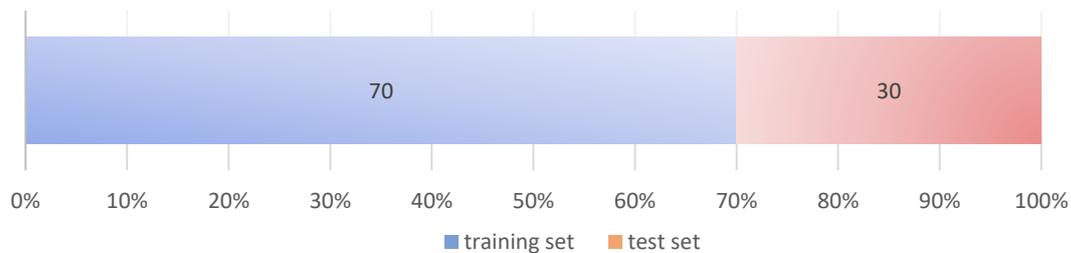
Ο αριθμός των εποχών σε ένα μοντέλο μπορεί να είναι αρκετά μεγάλος ώστε να υλοποιηθεί ο μαθησιακός αλγόριθμος τόσες φορές μέχρι το σφάλμα που προκύπτει να έχει ελαχιστοποιηθεί ικανοποιητικά. Ωστόσο, όσες περισσότερες είναι οι εποχές, το σφάλμα του συνόλου εκπαίδευσης μπορεί να μειώνεται, όμως το σφάλμα στο σύνολο αξιολόγησης αρχίζει να αυξάνεται από ένα σημείο και μετά, δηλαδή η γενίκευση του μοντέλου χάνεται.

6 Εφαρμογή Αλγορίθμων

6.1 Εισαγωγή των Δεδομένων στο Μοντέλο

Το πρόβλημα που μελετάται, όπως περιγράφηκε στο Κεφάλαιο 1, αποτελεί ένα πρόβλημα Εποπτευόμενης Μηχανικής Μάθησης. Είναι απαραίτητο πριν εκπαιδευτεί το μοντέλο με βάση το σύνολο δεδομένων, αυτά να χωριστούν σε δύο σύνολα. Το ένα από αυτά προορίζεται για την εκπαίδευση του μοντέλου και αποτελεί το δείγμα εκπαίδευσης (training set), ενώ το άλλο χρησιμοποιείται για έλεγχο της απόδοσής του και καλείται δείγμα αξιολόγησης (test set). Με αυτόν τον τρόπο, αποφαιίνεται η επίδοση του μοντέλου και αν μπορεί να γενικεύεται κατά την εισαγωγή νέων δεδομένων στα οποία δεν έχει εκπαιδευτεί (βλ. Παράγραφο 2.4). Ο διαχωρισμός των δεδομένων στα δύο αυτά υποσύνολα γίνεται με τυχαίο τρόπο.

Στο συγκεκριμένο πρόβλημα, το σύνολο εκπαίδευσης αποτελεί το 70% των συνολικών δεδομένων, δηλαδή περιέχει 29228 παρατηρήσεις και το σύνολο αξιολόγησης το 30% αυτών, δηλαδή αποτελείται από 12527 παρατηρήσεις.



Διάγραμμα 6.17 : Ποσοστό δεδομένων που αντιστοιχεί στο σύνολο εκπαίδευσης και αξιολόγησης.

Σημείωση: Οι βιβλιοθήκες της προγραμματιστικής γλώσσας *Python* που χρησιμοποιήθηκαν για την κατασκευή των διάφορων μοντέλων παρατίθενται στο Παράρτημα.

6.2 Μέτρα Αξιολόγησης του Μοντέλου

Μετά τη δημιουργία και την προσαρμογή ενός μοντέλου, καθίσταται απαραίτητη η αξιολόγησή του, με σκοπό να διαπιστωθεί αν αυτό έχει ικανοποιητική απόδοση και μπορεί να εφαρμοστεί αποτελεσματικά σε νέα σύνολα δεδομένων στα οποία δεν έχει εκπαιδευτεί. Για το λόγο αυτό, υπάρχουν ορισμένα μέτρα που χρησιμοποιούνται για την εκτίμηση της επίδοσής του, τα οποία περιγράφονται στη συνέχεια.

◇ Γραφική παράσταση των εκτιμώμενων τιμών συναρτήσει των πραγματικών παρατηρηθεισών τιμών

Μετά την εκπαίδευση του μοντέλου, εισάγεται σε αυτό το σύνολο αξιολόγησης και εκτιμώνται οι τιμές για τη μεταβλητή απόκρισης για κάθε μία από τις παρατηρήσεις. Για να διαπιστωθεί αν η απόδοση του μοντέλου είναι ικανοποιητική, δηλαδή αν οι εκτιμώμενες τιμές προσεγγίζουν τις πραγματικές, δημιουργείται ένα διάγραμμα στο οποίο αναπαρίστανται οι εκτιμώμενες τιμές από το μοντέλο συναρτήσει των αντίστοιχων πραγματικών τιμών. Αν τα σημεία αυτά σχηματίζουν τη νοητή ευθεία $y=x$ και την ακολουθούν, τότε φαίνεται πως το μοντέλο είναι αρκετά ικανοποιητικό καθώς οι εκτιμήσεις του πλησιάζουν τις πραγματικές τιμές των παρατηρήσεων. Διαφορετικά, όταν τα σημεία που τοποθετούνται είναι απομακρυσμένα από αυτή τη νοητή γραμμή, συμπεραίνεται ότι οι προβλεπόμενες τιμές αποκλίνουν από την πραγματική τιμή της παρατήρησης στην οποία αντιστοιχούν. Επομένως, το μοντέλο δεν είναι ικανό να εκτιμήσει τις παρατηρήσεις του συνόλου και έχει χαμηλή απόδοση.

◇ Γραφική παράσταση υπολοίπων

Τα υπόλοιπα (βλ. 3.1.1) περιγράφουν τις αποκλίσεις μεταξύ των πραγματικών παρατηρήσεων και των προβλεπόμενων από το μοντέλο τιμών. Οι μεγάλες τιμές των υπολοίπων δηλώνουν μεγάλη διαφορά της εκτιμώμενης από την πραγματική τιμή της παρατήρησης. Αντίθετα, όσο πιο μικρή τιμή λαμβάνουν, σημαίνει πως η εκτίμηση που προκύπτει από το μοντέλο προσεγγίζει την πραγματική τιμή. Για την οπτικοποίηση των αποτελεσμάτων από τον υπολογισμό των υπολοίπων δημιουργείται η γραφική παράσταση αυτών συναρτήσει των πραγματικών τιμών των παρατηρήσεων. Η κατανομή των σημείων όσο το δυνατόν πιο κοντά στην ευθεία $y=0$, δηλώνει ότι τα υπόλοιπα λαμβάνουν μικρές τιμές, επομένως οι αποκλίσεις μεταξύ των εκτιμώμενων και των πραγματικών τιμών των παρατηρήσεων είναι μικρές, άρα είναι παραπλήσιες. Αντίθετα, όσο πιο απομακρυσμένα είναι τα σημεία από αυτή, τα υπόλοιπα αντιστοιχίζονται σε μεγαλύτερες τιμές και αποτελεσματικά φαίνεται ότι το μοντέλο δεν είναι ικανό να εκτιμήσει τις τιμές που αντιστοιχούν στις παρατηρήσεις.

◇ Σχετικό σφάλμα

Το σχετικό σφάλμα δίνει μία ένδειξη για το πόσο καλή είναι η εκτίμηση μίας μέτρησης σε σχέση με το μέγεθος του μετρηθέντος αντικειμένου. Με αυτόν τρόπο μπορεί να προσδιοριστεί το απόλυτο σφάλμα σε σχέση με το πραγματικό μέγεθος της μέτρησης. Το σχετικό σφάλμα υπολογίζεται από τον ακόλουθο τύπο :

$$Relative\ Error = \frac{Absolute\ Error}{Y_{true}} = \frac{\hat{Y} - Y_{true}}{Y_{true}} \quad (6.1)$$

όπου Y_{true} η τιμή της παρατήρησης στο σύνολο δεδομένων και \hat{Y} η εκτιμώμενη τιμή της από το μοντέλο.

Για τις τιμές των σχετικών σφαλμάτων που υπολογίζονται για τις εκτιμήσεις που προκύπτουν από το μοντέλο, κατασκευάζεται η γραφική παράσταση αυτών συναρτήσει των πραγματικών τιμών των παρατηρήσεων. Όταν τα σημεία είναι τοποθετημένα σε μικρή απόσταση κατά μήκος της ευθείας $y=0$, τότε φαίνεται πως το μοντέλο δίνει εκτιμήσεις που πλησιάζουν την πραγματική τιμή των παρατηρήσεων. Διαφορετικά, όσο αυτά έχουν μεγάλη απόσταση από την ευθεία $y=0$, συμπεραίνεται η αδυναμία του μοντέλου να εκτιμήσει ικανοποιητικά τις παρατηρήσεις.

◇ Μέτρα θέσης και διασποράς

Τα αριθμητικά μεγέθη που προσδιορίζουν ένα κεντρικό σημείο γύρω από το οποίο τείνουν να συγκεντρώνονται τα δεδομένα ορίζονται ως μέτρα θέσης. Αποτελούν δηλαδή αριθμητικά μέτρα που περιγράφουν την ‘τάση’ ή την κεντρική τιμή των δεδομένων. Εκτός όμως από την κεντρική τάση, ενδιαφέρον παρουσιάζει και η μεταβλητότητα ή διασπορά των παρατηρήσεων, δηλαδή κατά πόσο αυτές εκτείνονται γύρω από το ‘κέντρο’ τους. Αυτό περιγράφεται από τα μέτρα διασποράς, τα οποία είναι απαραίτητα καθώς η μεταβλητότητα των παρατηρήσεων δεν είναι πάντα φανερή από τα μέτρα θέσης, όπως τη μέση τιμή. Αν τα δεδομένα είναι συγκεντρωμένα γύρω από τη μέση τιμή, δηλαδή η διασπορά είναι μικρή, τότε αυτή αντιπροσωπεύει ικανοποιητικά τα δεδομένα. Σε διαφορετική περίπτωση, τα μέτρα θέσης δε δίνουν καλή περιγραφή των παρατηρήσεων.

Από τα κυριότερα μέτρα θέσης και διασποράς αποτελούν η μέση τιμή και η τυπική απόκλιση αντίστοιχα, τα οποία χρησιμοποιούνται και στη συνέχεια. Ως μέτρο διασποράς επιλέχθηκε η τυπική απόκλιση στη θέση της διακύμανσης, καθώς αποτελεί λύση στη ‘δυσλειτουργία’ που εμφανίζει η τελευταία. Συγκεκριμένα, η τυπική απόκλιση αποτελεί την τετραγωνική ρίζα της διακύμανσης, η οποία δίνει τιμές που υπολογίζονται σε διαφορετικές μονάδες από αυτές του μετρούμενου μεγέθους. Έτσι, η τυπική απόκλιση έχει ίδιες μονάδες μέτρησης με αυτές των παρατηρήσεων και των μέτρων θέσης που έχουν υπολογιστεί.

6.3 Γραμμική Παλινδρόμηση

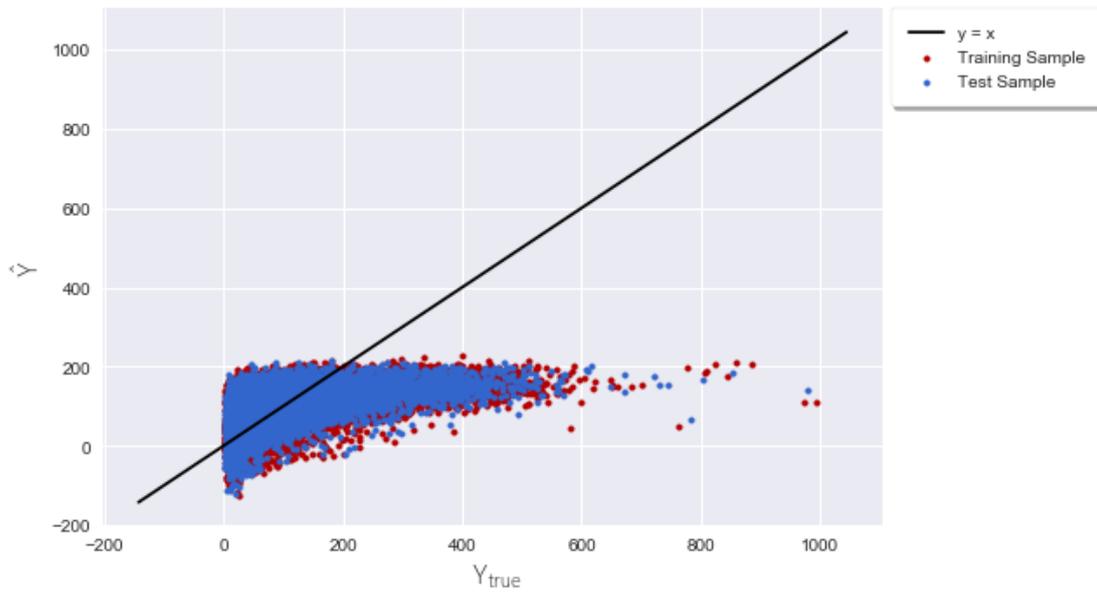
Κατά τη δημιουργία του μοντέλου Γραμμικής Παλινδρόμησης, εισάγεται το σύνολο εκπαίδευσης και γίνεται η προσαρμογή του σε αυτό. Ωστόσο, για να αποφανθεί αν το μοντέλο έχει καλή επίδοση στις εκτιμήσεις όταν εισάγονται νέα δεδομένα, είναι απαραίτητο να γίνουν ορισμένοι έλεγχοι που αφορούν στην απόδοσή του. Για το λόγο αυτό, χρησιμοποιείται το σύνολο αξιολόγησης για το οποίο εκτιμώνται οι τιμές της μεταβλητής απόκρισης για την κάθε παρατήρηση και υπολογίζονται ορισμένα μέτρα απόδοσης. Παράλληλα, κατασκευάζονται και κάποια επιπλέον διαγράμματα, τα οποία περιγράφουν την προσαρμογή του μοντέλου τόσο στο σύνολο εκπαίδευσης, όσο και στο σύνολο αξιολόγησης.

Για ένα μοντέλο παλινδρόμησης που ‘προβλέπει’ συνεχείς ποσοτικές μεταβλητές, η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος, $RMSE$ και ο συντελεστής προσδιορισμού, R^2 – που λαμβάνει τιμές στο διάστημα $[0,1]$ – μπορούν να χρησιμοποιηθούν για να αξιολογήσουν την απόδοση του μοντέλου. Η πρώτη ποσότητα περιγράφει τη διαφορά μεταξύ των εκτιμώμενων και πραγματικών τιμών των παρατηρήσεων, ενώ η δεύτερη εκφράζει το ποσοστό της διασποράς της μεταβλητής απόκρισης που εξηγείται με βάση το μοντέλο παλινδρόμησης. Προσαρμόζοντας το μοντέλο και εκτιμώντας τις τιμές για τη μεταβλητή απόκρισης των παρατηρήσεων του συνόλου αξιολόγησης, οι τιμές για τις δύο παραπάνω ποσότητες ακολουθούν :

Test Set	
RMSE	79.50
R^2	0.28

Πίνακας 6.1 : Πίνακας τιμών της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος, $RMSE$ και του συντελεστή προσδιορισμού, R^2 για το σύνολο αξιολόγησης.

Από τον Πίνακα 6.1 η τιμή για την τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος, $RMSE$ και του συντελεστή προσδιορισμού, R^2 είναι ίση με 79,50 και 0,28 αντίστοιχα. Από τις τιμές αυτές φαίνεται ότι η διαφορά μεταξύ των εκτιμώμενων και των πραγματικών τιμών των παρατηρήσεων είναι αρκετά μεγάλη, ενώ από την τιμή του συντελεστή προσδιορισμού συμπεραίνεται ότι η προσαρμογή του μοντέλου δεν είναι πολύ ικανοποιητική. Τα παραπάνω επιβεβαιώνονται και από το διάγραμμα που ακολουθεί :

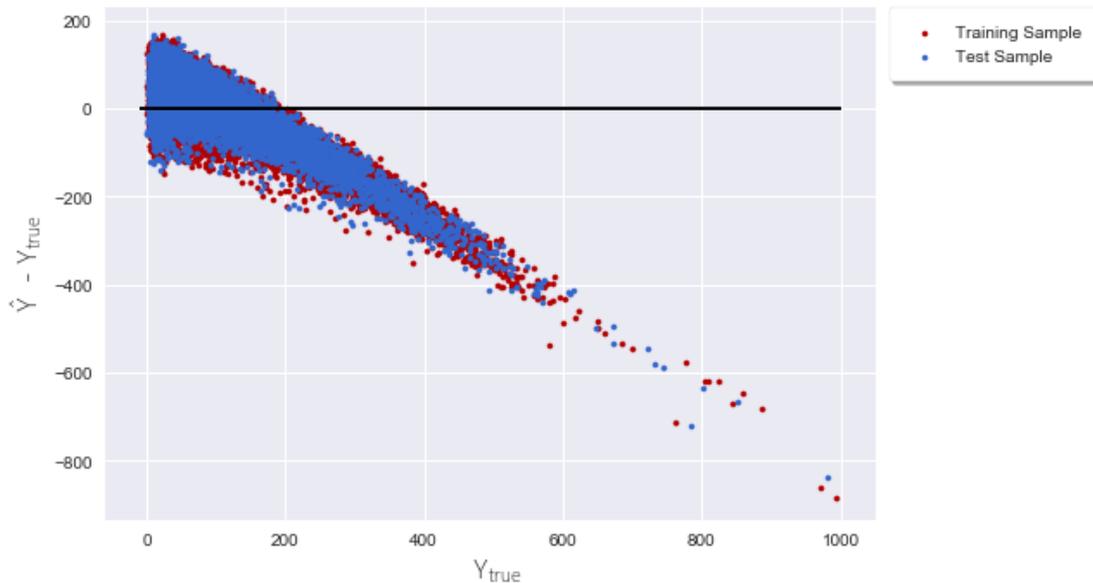


Διάγραμμα 6.2 : Διάγραμμα των εκτιμώμενων τιμών των παρατηρήσεων συναρτήσει των πραγματικών τους τιμών για τα σύνολα εκπαίδευσης και αξιολόγησης.

Στο Διάγραμμα 6.2 αναπαρίστανται οι εκτιμώμενες τιμές των παρατηρήσεων για τα σύνολα εκπαίδευσης και αξιολόγησης συναρτήσει των πραγματικών παρατηρούμενων τιμών. Με την κόκκινη απόχρωση (•) αναπαρίστανται τα σημεία που ανήκουν στο σύνολο εκπαίδευσης, ενώ αυτά με τον μπλε χρωματισμό (•) ανήκουν στο σύνολο αξιολόγησης. Από το διάγραμμα συμπεραίνεται ότι το μοντέλο δεν έχει καθόλου καλή προσαρμογή, ούτε καν για τα δεδομένα στα οποία εκπαιδεύτηκε. Αυτό φαίνεται παρατηρώντας την τοποθέτηση των σημείων, τα οποία από μία τιμή και έπειτα δεν συμβαδίζουν με την ευθεία $y=x$ και όλες οι προβλεπόμενες τιμές ανήκουν στο διάστημα $[0, 200]$ για οποιαδήποτε πραγματική παρατηρούμενη τιμή. Αυτό σημαίνει πως όσο οι εκτιμώμενες από το μοντέλο τιμές απομακρύνονται από την ευθεία $y=x$, τόσο μεγαλύτερη διαφορά θα σημειώνεται μεταξύ αυτών και των πραγματικών τιμών, καθώς αν αντίθετα πλησίαζαν την ευθεία αυτή, σύμφωνα με την εξίσωση που την ορίζει οι τιμές θα ήταν παραπλήσιες.

Συγκεκριμένα, παρατηρείται ότι για πραγματικές τιμές μεγαλύτερες του 200, οι εκτιμώμενες τιμές αντιστοιχίζονται στο διάστημα $[0, 200]$, ενώ για μικρές παρατηρούμενες τιμές, οι εκτιμήσεις που προκύπτουν από το μοντέλο μπορεί να είναι αρκετά μεγαλύτερες. Έτσι, πραγματικές τιμές μικρότερες του 100 αντιστοιχούν σε προβλεπόμενες τιμές μέχρι 200, δηλαδή διπλάσιας εκτίμησης, ενώ υπάρχουν και θετικές πραγματικές τιμές που αντιστοιχίζονται σε αρνητικές προβλέψεις του μοντέλου. Καταληκτικά, το μοντέλο που προκύπτει δεν είναι ικανό να εκτιμήσει με ακρίβεια τις παρατηρήσεις που εισάγονται σε αυτό, με αποτέλεσμα να μην έχει καλή απόδοση.

Τα παραπάνω συμπεράσματα ενισχύει και το Διάγραμμα 6.3, το οποίο περιγράφει τα υπόλοιπα, δηλαδή τη διαφορά μεταξύ της εκτιμώμενης και της πραγματικής τιμής για κάθε παρατήρηση, συναρτήσει της πραγματικής τιμής των στοιχείων.

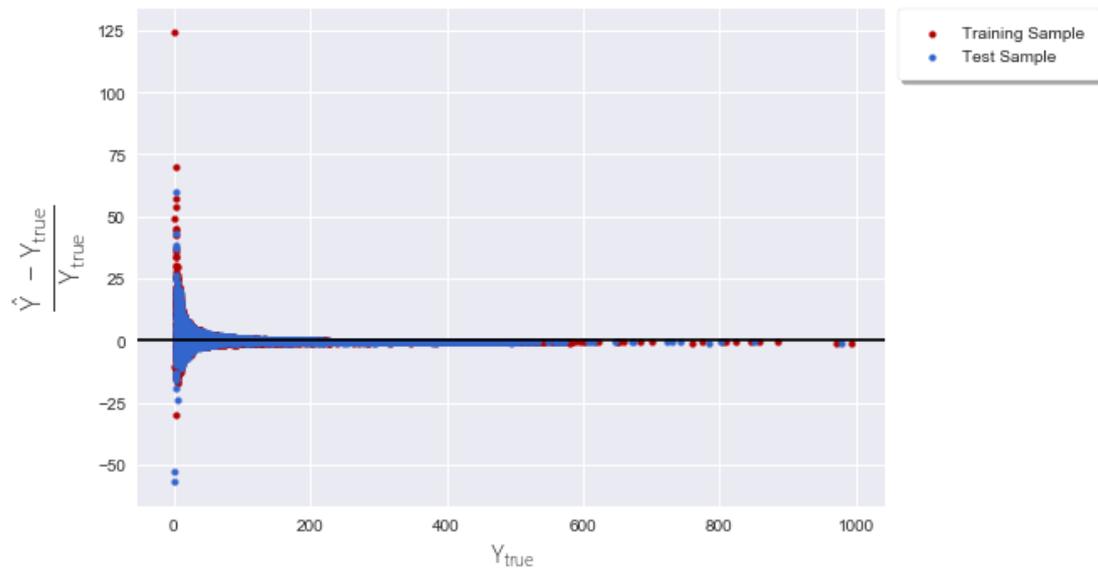


Διάγραμμα 6.3 : Διάγραμμα αναπαράστασης των υπολοίπων συναρτήσει των πραγματικών τιμών των παρατηρήσεων για τα σύνολα εκπαίδευσης και αξιολόγησης.

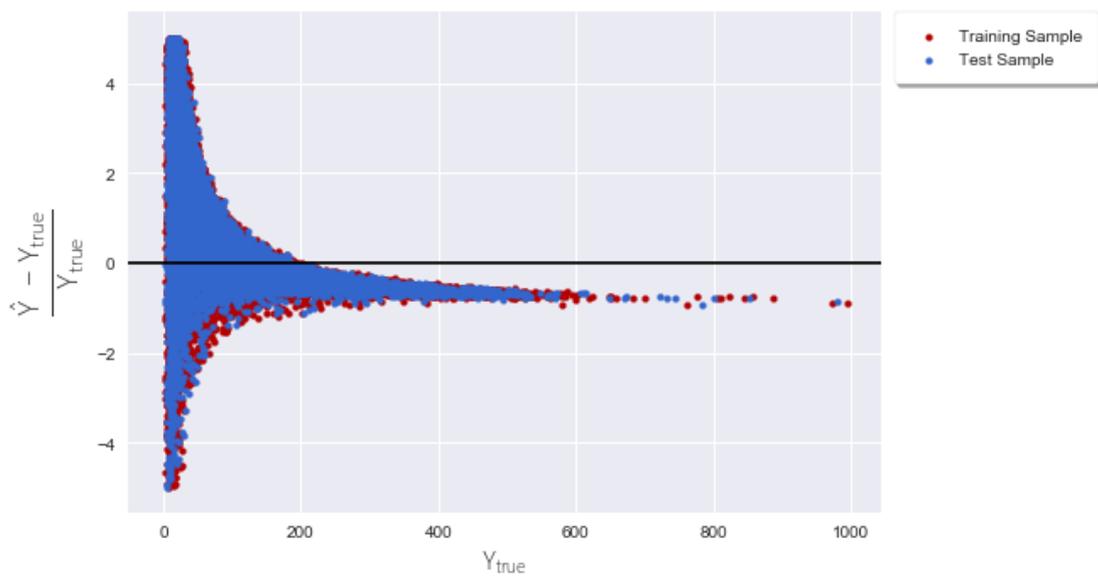
Από το Διάγραμμα 6.3 φαίνεται πως οι τιμές των υπολοίπων για τις παρατηρήσεις είναι αρκετά μεγάλες, καθώς τα περισσότερα σημεία είναι απομακρυσμένα από την ευθεία $y=0$, που σημαίνει ότι όσο πιο μακριά από την ευθεία τοποθετείται ένα σημείο, τόσο η διαφορά $\hat{y}_n - y_n$ μεγαλώνει, όπου \hat{y}_n είναι η εκτίμηση για την παρατήρηση n και y_n η πραγματική τιμή αυτής. Αυτό συνεπάγεται την απόκλιση των εκτιμώμενων τιμών του μοντέλου από τις πραγματικές τιμές των παρατηρήσεων, δηλαδή το μοντέλο δεν είναι ικανό να δώσει εκτιμήσεις με μικρό σφάλμα, το οποίο συμφωνεί με τα συμπεράσματα που προέκυψαν από τη μελέτη του Διαγράμματος 6.2 .

Το Διάγραμμα 6.4 που ακολουθεί αναπαριστά τα σχετικά σφάλματα για τις εκτιμήσεις (βλ. Παράγραφο 6.2) συναρτήσει των πραγματικών τιμών των παρατηρήσεων. Από το διάγραμμα φαίνεται πως για αρκετές από τις εκτιμήσεις του μοντέλου τα σχετικά σφάλματα που υπολογίζονται συγκεντρώνονται σε μικρή απόσταση κατά μήκος του θετικού ημιάξονα Ox . Υπάρχουν όμως και εκτιμήσεις για τις οποίες τα σχετικά σφάλματα λαμβάνουν μεγάλες τιμές και τα αντίστοιχα σημεία είναι απομακρυσμένα από τον ημιάξονα Ox . Ωστόσο επειδή τα σημεία αυτά δεν είναι πολλά – λιγότερο από το 5% των συνολικών δεδομένων – και ορισμένα απέχουν αρκετά από τα υπόλοιπα μπορούν να χαρακτηριστούν ως ακραίες τιμές. Για την περεταίρω μελέτη των σχετικών σφαλμάτων δημιουργείται ένα αντίστοιχο διάγραμμα του Διαγράμματος 6.4, Διάγραμμα 6.5, στο οποίο οι τιμές των σχετικών σφαλμάτων ανήκουν στο διάστημα

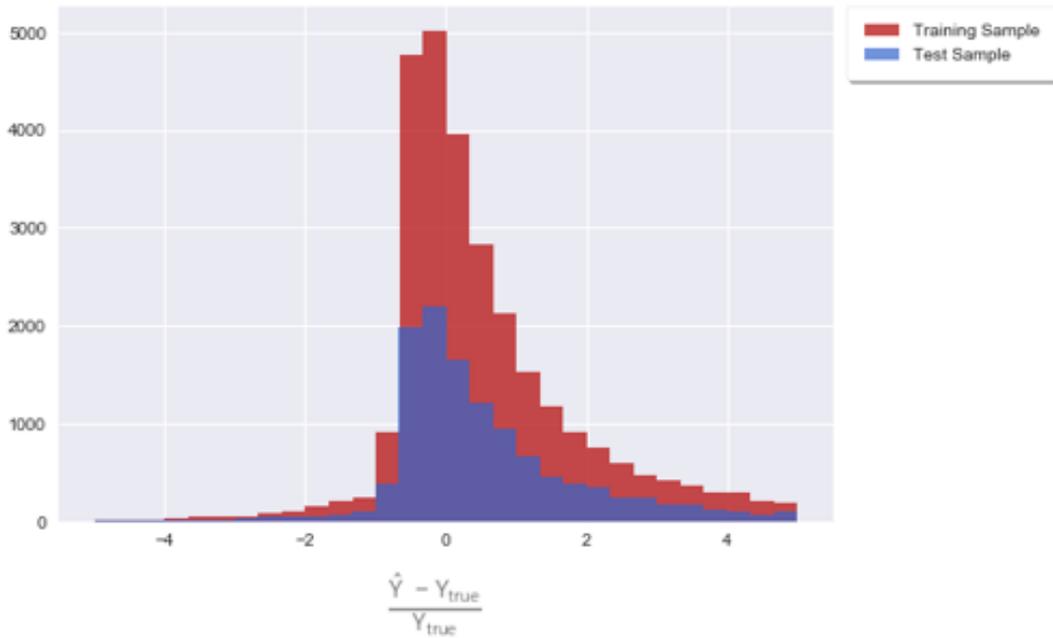
$[-5, 5]$ ώστε να υπάρχει μεγαλύτερη ευκρίνεια στην κατανομή των σημείων. Με αυτό τον τρόπο μπορεί να εκτιμηθεί καλύτερα η απόδοση του μοντέλου, αφού αφαιρούνται οι ακραίες τιμές που μπορεί να είναι παραπλανητικές για την επίδοσή του.



Διάγραμμα 6.4 : Τα σχετικά σφάλματα για τις εκτιμήσεις του μοντέλου συναρτήσει των πραγματικών τιμών των παρατηρήσεων.



Διάγραμμα 6.5 : Τα σχετικά σφάλματα για τις εκτιμήσεις του μοντέλου συναρτήσει των πραγματικών τιμών των παρατηρήσεων στο διάστημα $[-5, 5]$.



Διάγραμμα 6.6 : Ιστόγραμμα συχνοτήτων για τα σχετικά σφάλματα των εκτιμήσεων του μοντέλου, που λαμβάνουν τιμές στο διάστημα $[-5,5]$.

Από το Διάγραμμα 6.6, στο οποίο αναπαρίστανται τα ιστογράμματα συχνοτήτων των αποκομμένων σχετικών σφαλμάτων στο διάστημα $[-5, 5]$, φαίνεται πως τα σύνολα εκπαίδευσης και αξιολόγησης ακολουθούν την ίδια κατανομή, καθώς έχουν την ίδια μορφή, αλλά και σχεδόν ίδια μέση τιμή και τυπική απόκλιση όπως παρατηρείται από τον Πίνακα 6.2.

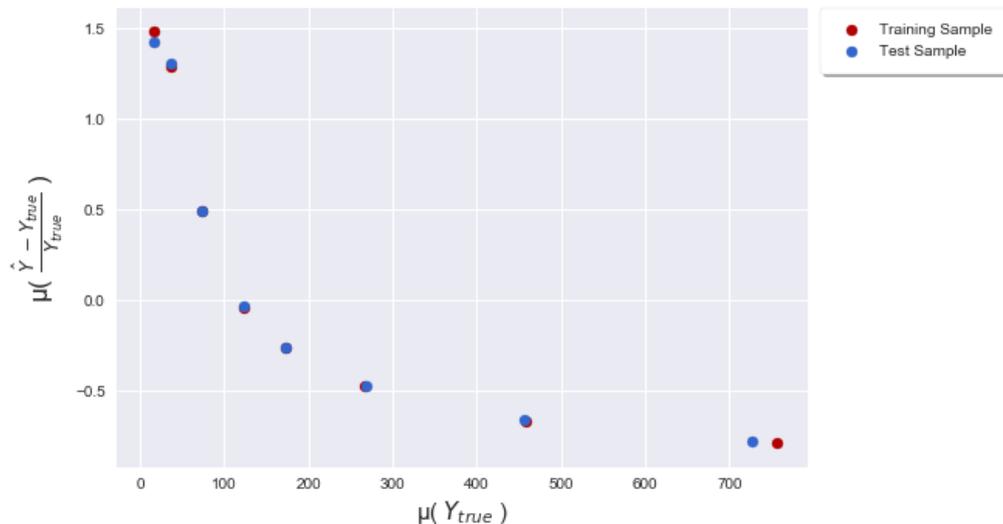
	<i>Training Sample</i>	<i>Test Sample</i>
Mean	0.517	0.516
Standard Deviation	1.307	1.309

Πίνακας 6.4 : Τιμές για τη μέση τιμή και τυπική απόκλιση για τα αποκομμένα σχετικά σφάλματα του συνόλου εκπαίδευσης και αξιολόγησης στο διάστημα $[-5, 5]$.

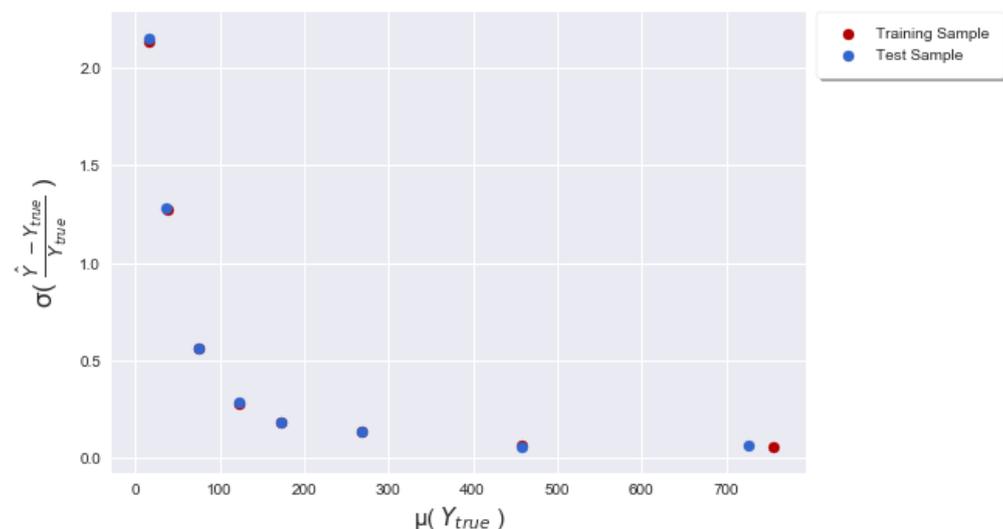
Επιπλέον, από το Διάγραμμα 6.6 υπολογίζεται ότι περίπου το 70% των εκτιμήσεων του συνόλου εκπαίδευσης και αξιολόγησης αντιστοιχούν σε απόλυτο σχετικό σφάλμα μικρότερο της μονάδας. Αντίστοιχα, σχεδόν το 86% των δεδομένων εκπαίδευσης και το 85% των δεδομένων αξιολόγησης παρουσιάζουν απόλυτο σχετικό σφάλμα μικρότερο των 2 μονάδων.

Ακολούθως, οι πραγματικές τιμές των παρατηρήσεων χωρίζονται σε υποσύνολα με τις αντίστοιχες τιμές των σχετικών σφαλμάτων στο διάστημα $[-5, 5]$, για μία καλύτερη και με μεγαλύτερη ευκρίνεια εκτίμηση της απόδοσης του μοντέλου για τις διάφορες παρατηρηθείσες τιμές.

Συγκεκριμένα, το σύνολο των πραγματικών τιμών των παρατηρήσεων χωρίζεται αρχικά σε τέσσερα επιμέρους υποσύνολα με εύρος 200 μονάδες τα πρώτα τρία και 400 το τελευταίο. Ωστόσο, επειδή το πεδίο τιμών $[0, 200]$ περιλαμβάνει ένα μεγάλο αριθμό παρατηρήσεων, αυτό διαμερίζεται σε άλλα πέντε υποσύνολα, εύρους 50 μονάδων το καθένα με εξαίρεση τα πρώτα δύο που έχουν εύρος 25 μονάδες, με σκοπό η συμπεριφορά του μοντέλου να είναι πιο ξεκάθαρη. Στα Διαγράμματα 6.7 και 6.8 που ακολουθούν παρουσιάζονται οι τιμές για τη μέση τιμή και την τυπική απόκλιση των αποκομμένων σχετικών σφαλμάτων στο διάστημα $[-5, 5]$ για τα σύνολα εκπαίδευσης και αξιολόγησης, συναρτήσει της μέσης τιμής των πραγματικών παρατηρήσεων για κάθε ένα από τα υποσύνολα που προέκυψαν.



Διάγραμμα 6.7 : Διάγραμμα μέσης τιμής των αποκομμένων σχετικών σφαλμάτων των εκτιμήσεων στο διάστημα $[-5, 5]$ συναρτήσει της μέσης τιμής των παρατηρηθεισών τιμών για κάθε υποσύνολο αυτών.



Διάγραμμα 6.8 : Διάγραμμα τυπικής απόκλισης των αποκομμένων σχετικών σφαλμάτων των εκτιμήσεων στο διάστημα $[-5, 5]$ συναρτήσει της μέσης τιμής των παρατηρηθεισών τιμών για κάθε υποσύνολο αυτών.

Σημείωση : Στα Διαγράμματα 6.7 και 6.8 όσες τιμές για τη μέση τιμή και την τυπική απόκλιση του συνόλου εκπαίδευσης δεν φαίνεται να παρουσιάζονται στο διάγραμμα, συμπίπτουν ή έχουν πολύ μικρή απόκλιση από τις αντίστοιχες του συνόλου αξιολόγησης, με αποτέλεσμα να καλύπτονται από αυτές.

Η μέση τιμή των σχετικών σφαλμάτων αποτελεί ουσιαστικά το μέσο όρο των σχετικών σφαλμάτων των εκτιμήσεων και περιγράφει τη μέση απόκλιση της κάθε εκτίμησης από την πραγματική τιμή της αντίστοιχης παρατήρησης. Αποτελεσματικά, μπορεί να αποτελέσει ένα σημαντικό μέτρο για την εκτίμηση της απόδοσης και της προσαρμοστικότητας του μοντέλου. Αντίστοιχα, η τυπική απόκλιση που υπολογίζεται για τα σχετικά σφάλματα, δηλώνει πόσο αυτά διαφέρουν από τη μέση τιμή τους.

Από τα Διαγράμματα 6.7 και 6.8, φαίνεται πως για τις μικρότερες τιμές των παρατηρήσεων, οι εκτιμήσεις που προκύπτουν από το μοντέλο έχουν μεγαλύτερη απόκλιση από αυτές, σε σχέση με όσες αντιστοιχούν σε μεγαλύτερες τιμές. Αυτό γίνεται αντιληπτό διότι τόσο η μέση τιμή, όσο και η τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων μειώνονται καθώς αυξάνεται η πραγματική τιμή των παρατηρήσεων, που σημαίνει ότι οι εκτιμώμενες τιμές δεν παρουσιάζουν σημαντικές διαφορές με τις πραγματικές τιμές των παρατηρήσεων. Αυτό συμβαίνει λόγω των υψηλών συσχετίσεων που παρουσιάζει η μεταβλητή απόκρισης με τις επεξηγηματικές μεταβλητές, όταν αυτή λαμβάνει μεγάλες τιμές. Επομένως, για μεγάλες τιμές της μεταβλητής απόκρισης, το μοντέλο φαίνεται να έχει καλύτερη απόδοση, αφού σημειώνονται μικρότερα σφάλματα στις εκτιμήσεις του σε σχέση με αυτές που αντιστοιχούν σε μικρότερες τιμές της.

6.4 Δέντρα Απόφασης

Στη μέθοδο των Δέντρων Απόφασης εφαρμόζονται οι δύο αλγόριθμοι *Boosting*, *AdaBoost* και *Gradient Boosting*, όπως περιγράφηκαν στο Κεφάλαιο 4. Για κάθε έναν από αυτούς τους αλγορίθμους ορίζονται συγκεκριμένες παράμετροι που διαμορφώνουν τα μοντέλα που προκύπτουν.

6.4.1 AdaBoost

Προσδιορισμός Παραμέτρων

Για την κατασκευή ενός μοντέλου βασισμένο στον αλγόριθμο *AdaBoost* είναι απαραίτητο να οριστούν συγκεκριμένες παράμετροι που αφορούν στη μορφή του δέντρου, αλλά και στις διεργασίες που γίνονται στους κόμβους αυτού. Συγκεκριμένα, αυτές αφορούν :

- Τον αριθμό των δέντρων,
- Το ρυθμό εκμάθησης,
- Το μέγιστο βάθος του δέντρου,
- Τη συνάρτηση κόστους,
- Το κριτήριο διαχωρισμού,
- Τον ελάχιστο αριθμό παρατηρήσεων για το διαχωρισμό ενός κόμβου σε επιμέρους ή τη δημιουργία ενός φύλλου.

Όσον αφορά τη συνάρτηση κόστους, αυτή ορίζεται να είναι η γραμμική (βλ. Παράγραφο 4.5.1 , Σχέση 4.14), η οποία είναι και η προκαθορισμένη επιλογή του αλγορίθμου και ως κριτήριο διαχωρισμού επιλέγεται το Μέσο Τετραγωνικό Σφάλμα (βλ. Παράγραφο 4.2 , Σχέση 4.2) το οποίο αξιολογεί το διαχωρισμό του κόμβου που προκύπτει. Επιπλέον, καθίσταται αναγκαίο να προσδιοριστεί ο ελάχιστος αριθμός των παρατηρήσεων που χρειάζεται για το διαχωρισμό ενός εσωτερικού κόμβου σε δύο επιμέρους, καθώς επίσης και ο αντίστοιχος αριθμός για να θεωρηθεί ένας κόμβος φύλλο. Έτσι, για να διαχωριστεί ένας κόμβος απαιτούνται τουλάχιστον δύο παρατηρήσεις, ενώ κάθε φύλλο του δέντρου πρέπει να έχει τουλάχιστον μία παρατήρηση.

Όσον αφορά τις παραμέτρους που αφορούν τον αριθμό και το βάθος των δέντρων, όπως επίσης και το ρυθμό εκμάθησης, χρειάζεται να γίνει περεταίρω διερεύνηση για τον προσδιορισμό τους. Η διερεύνηση αυτή είναι απαραίτητη καθώς για τους

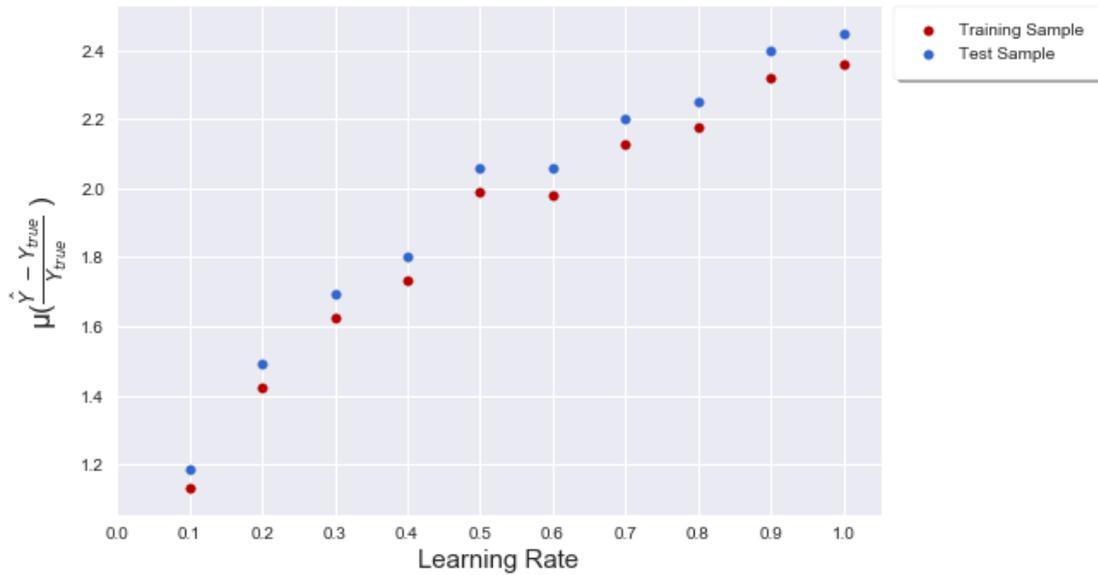
διάφορους συνδυασμούς αυτών προκύπτουν μοντέλα με άλλα αποτελέσματα, που οφείλονται στη διαφορετική προσαρμογή του κάθε μοντέλου ανάλογα με τα ορίσματα που δέχεται.

Διερεύνηση Παραμέτρων

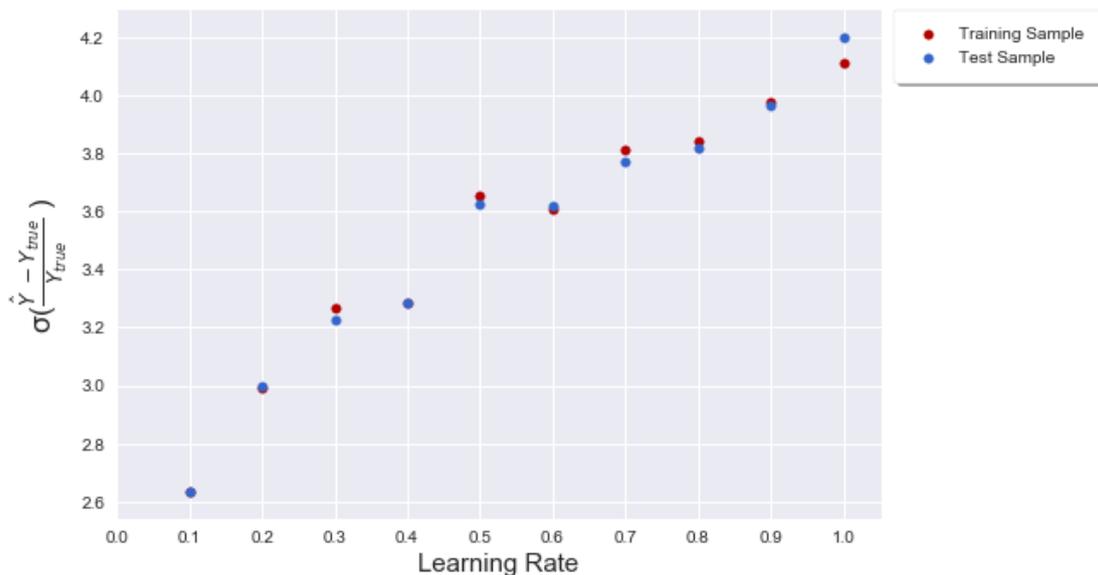
Για τον τελικό προσδιορισμό των παραμέτρων του μοντέλου είναι απαραίτητη η δοκιμή ποικίλων συνδυασμών αυτών, που αφορούν τον αριθμό και το βάθος των δέντρων, όπως και το ρυθμό εκμάθησης. Η διαδικασία αυτή είναι σημαντική για την εύρεση του μοντέλου με τη βέλτιστη απόδοση και την καλύτερη προσαρμογή, καθώς αυτές οι παράμετροι δεν δρουν ανεξάρτητα η μία από την άλλη. Ειδικότερα, όσο μικρότερη είναι η τιμή που λαμβάνει ο ρυθμός εκμάθησης, τόσο μεγαλύτερος είναι ο αριθμός των δέντρων που κατασκευάζονται. Ωστόσο, η κατασκευή ενός μεγάλου αριθμού δέντρων μπορεί να οδηγήσει σε υπερεκπαίδευση του μοντέλου. Το ίδιο μπορεί να συμβεί για μεγάλο βάθος δέντρων, καθώς αυξάνεται η πολυπλοκότητα του μοντέλου. Επομένως, η κατασκευή ενός αποδοτικού μοντέλου χρήζει κατάλληλης επιλογής των παραμέτρων αυτών.

Αρχικά, για σταθερό αριθμό δέντρων ερευνάται η βέλτιστη επιλογή του ρυθμού εκμάθησης και του βάθους των δέντρων. Ο ρυθμός εκμάθησης λαμβάνει τιμές στο διάστημα $(0, 1]$, επομένως δοκιμάζονται όλες οι τιμές από το 0.1 μέχρι τη μονάδα με βήμα 0.1. Αντίστοιχα, για το βάθος του κάθε δέντρου, ελέγχονται οι τιμές 3, 4 και 5 καθώς δέντρα με μεγαλύτερο βάθος αυξάνουν την πολυπλοκότητα του μοντέλου και μπορεί να οδηγήσουν σε υπερεκπαίδευσή του. Όσον αφορά τον αριθμό των δέντρων, κατασκευάζονται 50 δέντρα και στη συνέχεια ο αριθμός αυτός αυξάνεται ανάλογα με τα αποτελέσματα που προκύπτουν, για την κατασκευή ενός βέλτιστου τελικού μοντέλου.

Για κάθε μοντέλο που υλοποιείται, υπολογίζεται το μέσο τετραγωνικό και απόλυτο σφάλμα των εκτιμήσεων, τα σχετικά τους σφάλματα, καθώς επίσης και η μέση τιμή και τυπική απόκλιση αυτών. Στα Διαγράμματα 6.9 και 6.10 αναπαρίστανται οι τιμές για τη μέση τιμή και τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων για μοντέλα με 50 δέντρα βάθους 5. Τα αντίστοιχα διαγράμματα για μοντέλα 50 δέντρων με βάθος 3 και 4, καθώς δεν παρουσιάζουν ιδιαίτερες διαφορές με τα ακόλουθα διαγράμματα, αντιθέτως τόσο η μέση τιμή, όσο και η τυπική απόκλιση των σχετικών σφαλμάτων λαμβάνουν μεγαλύτερες τιμές. Αυτό σημαίνει ότι για δέντρα βάθους 3 και 4 οι εκτιμήσεις που προκύπτουν από τα διάφορα μοντέλα έχουν μεγαλύτερη απόκλιση από την πραγματική τιμή των παρατηρήσεων σε σχέση με τα δέντρα βάθους 5.



Διάγραμμα 6.9 : Διάγραμμα μέσης τιμής των σχετικών σφαλμάτων των εκτιμήσεων για μοντέλα 50 δέντρων βάθους 5 για τις διάφορες τιμές του ρυθμού εκμάθησης.

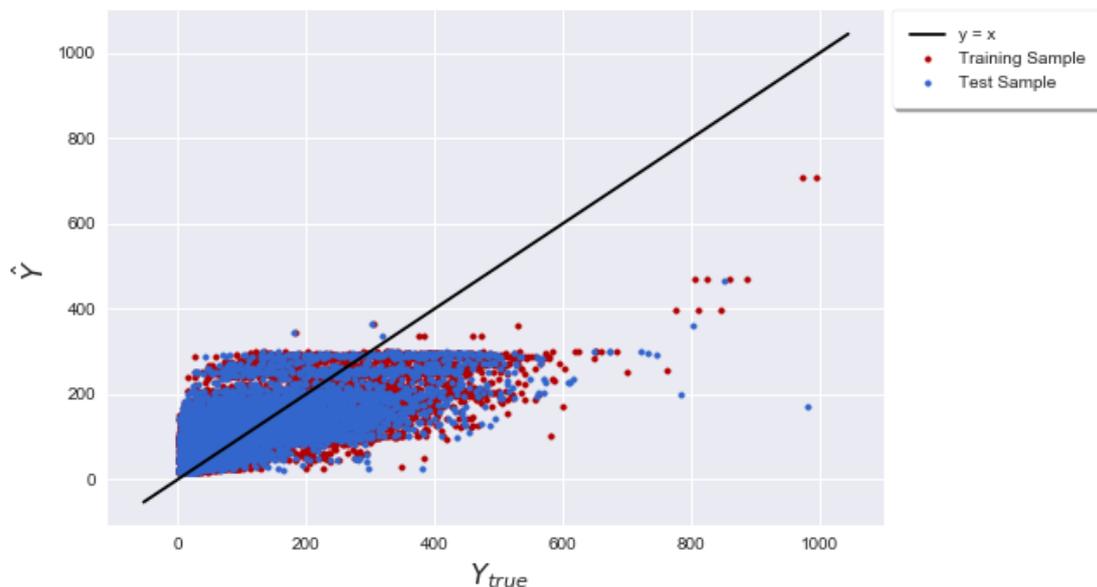


Διάγραμμα 6.10 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων για μοντέλα 50 δέντρων βάθους 5 για τις διάφορες τιμές του ρυθμού εκμάθησης.

Από τα Διαγράμματα 6.9 και 6.10 φαίνεται πως οι τιμές για τη μέση τιμή και την τυπική απόκλιση των σχετικών σφαλμάτων που υπολογίζονται για τις εκτιμήσεις του μοντέλου, δεν εμφανίζουν μεγάλες διαφορές μεταξύ του συνόλου εκπαίδευσης και αξιολόγησης. Επομένως, τα μοντέλα έχουν παρόμοια συμπεριφορά για τα δύο σύνολα. Ωστόσο, παρατηρείται ότι και στα δύο διαγράμματα η μέση τιμή και η τυπική απόκλιση

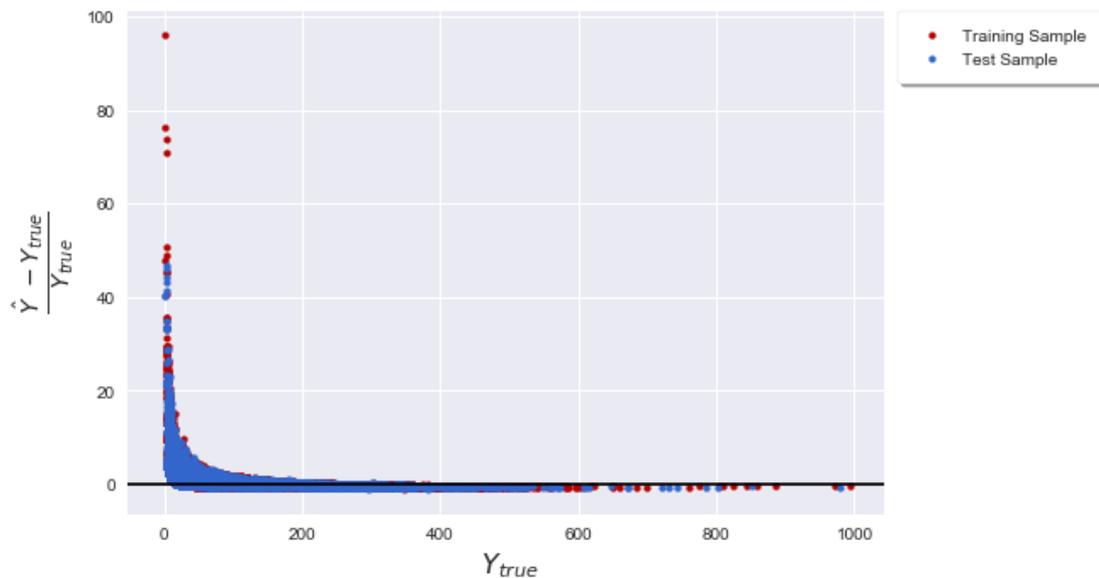
των σχετικών σφαλμάτων λαμβάνουν μεγάλες τιμές, δηλαδή οι εκτιμήσεις του μοντέλου δεν προσεγγίζουν ικανοποιητικά την πραγματική τιμή των παρατηρήσεων. Ειδικότερα από το Διάγραμμα 6.10 συμπεραίνεται ότι για μεγαλύτερες τιμές του ρυθμού εκμάθησης, η απόκλιση των εκτιμώμενων από τις πραγματικές τιμές αυξάνεται.

Τη μικρότερη τυπική απόκλιση σχετικών σφαλμάτων συγκρίνοντας όλα τα μοντέλα, παρουσιάζει αυτό των 50 δέντρων με βάθος 5 και ρυθμό εκμάθησης 0.1. Τα διαγράμματα των εκτιμώμενων τιμών και των σχετικών σφαλμάτων συναρτήσει των πραγματικών τιμών αποτελούν τα Διαγράμματα 6.11 και 6.12 αντίστοιχα.



Διάγραμμα 6.11 : Διάγραμμα των εκτιμώμενων τιμών των παρατηρήσεων συναρτήσει των πραγματικών τους τιμών για τα σύνολα εκπαίδευσης και αξιολόγησης για μοντέλο με 50 δέντρα βάθους 5 και ρυθμό εκμάθησης 0.1.

Από το Διάγραμμα 6.11 φαίνεται πως το μοντέλο δεν είναι ικανό να εκτιμήσει αποτελεσματικά τη μεταβλητή απόκρισης, καθώς τα σημεία που τοποθετούνται δεν ακολουθούν την ευθεία $y=x$ και επομένως οι εκτιμώμενες τιμές δεν προσεγγίζουν τις πραγματικές παρατηρηθείσες τιμές. Το γεγονός αυτό επιβεβαιώνει το Διάγραμμα 6.12, όπου διαπιστώνεται ότι για αρκετές τιμές των παρατηρήσεων τα σχετικά σφάλματα των εκτιμήσεων που τους αντιστοιχούν λαμβάνουν μεγάλες τιμές, το οποίο επισημαίνει την απόκλιση μεταξύ των εκτιμώμενων και των πραγματικών παρατηρούμενων τιμών.



Διάγραμμα 6.12 : Τα σχετικά σφάλματα για τις εκτιμήσεις του μοντέλου συναρτήσεως των πραγματικών τιμών των παρατηρήσεων για τα σύνολα εκπαίδευσης και αξιολόγησης για μοντέλο με 50 δέντρα βάθους 5 και ρυθμό εκμάθησης 0.1.

Σύμφωνα με τα παραπάνω, συμπεραίνεται ότι δεν υπάρχει κάποιος συνδυασμός για το βάθος των δέντρων και το ρυθμό εκμάθησης ώστε ένα μοντέλο με 50 δέντρα να είναι αποδοτικό και να δίνει ικανοποιητικές εκτιμήσεις. Αποτελεσματικά, είναι αναγκαία η περεταίρω διερεύνηση για τον εντοπισμό της κατάλληλης τριπλέτας αριθμού δέντρων, βάθους δέντρου και ρυθμού εκμάθησης, για την κατασκευή ενός βέλτιστου μοντέλου. Για το λόγο αυτό, δημιουργούνται νέα μοντέλα βάθους 3, 4 και 5, με το ρυθμό εκμάθησης να λαμβάνει τις τιμές 0.1, 0.2, 0.4, 0.6, 0.8 και 1.0. Όσον αφορά τον αριθμό των δέντρων που κατασκευάζονται, επιλέγονται όλες οι τιμές από 50 μέχρι 500 με βήμα 50.

Όπως και στα προηγούμενα μοντέλα, υπολογίζεται το μέσο τετραγωνικό και απόλυτο σφάλμα για τις εκτιμήσεις, τα σχετικά τους σφάλματα, καθώς επίσης και η μέση τιμή και τυπική απόκλιση αυτών. Επιπλέον, κατασκευάζονται και όλα τα διαγράμματα των εκτιμώμενων τιμών που προκύπτουν από το κάθε μοντέλο συναρτήσεως των πραγματικών τιμών των παρατηρήσεων.

Για κάθε μία τιμή βάθους δέντρου, δημιουργείται ένας πίνακας που περιέχει τις τιμές που λαμβάνει η μέση τιμή των σχετικών σφαλμάτων για κάθε μοντέλο που προκύπτει από το συνδυασμό των παραπάνω τιμών για τον αριθμό των δέντρων και το ρυθμό εκμάθησης. Αντίστοιχα, ο ίδιος πίνακας κατασκευάζεται για τις τιμές της τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων που προκύπτουν από κάθε διαφορετικό μοντέλο. Στόχος της δημιουργίας αυτών των πινάκων είναι η ανίχνευση των συνδυασμών για τον αριθμό των δέντρων και το ρυθμό εκμάθησης για τους οποίους η μέση τιμή και τυπική απόκλιση των σχετικών σφαλμάτων λαμβάνουν την ελάχιστη τιμή. Με αυτό τον τρόπο, επιχειρείται η εύρεση του μοντέλου για το οποίο οι εκτιμώμενες τιμές προσεγγίζουν όσο το δυνατό καλύτερα τις πραγματικές τιμές των παρατηρήσεων.

Μετά τη μελέτη όλων των πινάκων καθώς και των διαγραμμάτων που περιγράφουν τις εκτιμώμενες τιμές συναρτήσει των πραγματικών τιμών, συμπεραίνεται ότι όλα τα μοντέλα που κατασκευάζονται έχουν παρόμοια συμπεριφορά. Δηλαδή, τόσο η μέση τιμή όσο και η τυπική απόκλιση για τα σχετικά σφάλματα λαμβάνουν κοντινές τιμές στο ίδιο διάστημα και δεν υπάρχει κάποιο μοντέλο για το οποίο οι εκτιμήσεις που προκύπτουν να πλησιάζουν τις τιμές των παρατηρήσεων και άρα τα σφάλματα που σημειώνονται να είναι αρκετά μεγάλα.

Η ελάχιστη τιμή για τη μέση τιμή και την τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων ανιχνεύεται για μοντέλο με δέντρα βάθους 5, όπως φαίνεται στους Πίνακες 6.3 και 6.4. Οι αντίστοιχοι πίνακες για μοντέλα μέγιστου βάθους δέντρων 3 και 4 παρατίθενται στο *Παράρτημα*, καθώς για τους συνδυασμούς που προκύπτουν δεν υπάρχει κάποιο μοντέλο με καλύτερη απόδοση.

		<i>Learning Rate</i>					
		0.1	0.2	0.4	0.6	0.8	1.0
<i>No of Trees</i>	50	1.19	1.49	1.80	2.06	2.25	2.45
	100	1.46	1.89	2.30	2.50	2.64	2.68
	150	1.67	2.15	2.58	2.64	2.77	2.79
	200	1.86	2.27	2.69	2.75	2.94	2.86
	250	2.01	2.38	2.77	2.88	2.97	2.91
	300	2.11	2.47	2.85	2.91	3.01	2.93
	350	2.20	2.56	2.85	2.97	3.01	2.95
	400	2.28	2.63	2.88	3.01	3.03	3.03
	450	2.32	2.67	2.90	3.02	3.03	3.01
	500	2.36	2.69	2.94	3.05	3.04	3.03

Πίνακας 6.5 : Πίνακας με τη μέση τιμή των σχετικών σφαλμάτων για μοντέλα που προκύπτουν από διαφορετικούς συνδυασμούς του συνολικού αριθμού δέντρων (γραμμές) και του ρυθμού εκμάθησης (στήλες) για δέντρα βάθους 5.

		<i>Learning Rate</i>					
		0.1	0.2	0.4	0.6	0.8	1.0
<i>No of Trees</i>	50	2.64	3.00	3.29	3.62	3.82	4.20
	100	2.94	3.43	3.86	4.19	4.29	4.40
	150	3.19	3.72	4.27	4.35	4.47	4.50
	200	3.36	3.81	4.38	4.48	4.71	4.62
	250	3.56	3.96	4.51	4.70	4.77	4.70
	300	3.66	4.08	4.62	4.74	4.81	4.70
	350	3.74	4.21	4.61	4.79	4.83	4.75
	400	3.84	4.33	4.64	4.84	4.89	4.85
	450	3.89	4.38	4.66	4.86	4.88	4.84
	500	3.94	4.39	4.75	4.92	4.90	4.88

Πίνακας 6.6 : Πίνακας με την τυπική απόκλιση των σχετικών σφαλμάτων για μοντέλα που προκύπτουν από διαφορετικούς συνδυασμούς του συνολικού αριθμού δέντρων (γραμμές) και του ρυθμού εκμάθησης (στήλες) για δέντρα βάθους 5.

Σύμφωνα με τους παραπάνω πίνακες φαίνεται πως το μοντέλο για το οποίο προκύπτει η ελάχιστη τιμή για τη μέση τιμή και τυπική απόκλιση των σχετικών σφαλμάτων αποτελείται από 50 δέντρα βάθους 5 με ρυθμό εκμάθησης 0.1, το οποίο περιγράφεται από τα αρχικά Διαγράμματα 6.11 και 6.12. Ωστόσο, όπως έχει ήδη αναφερθεί το μοντέλο αυτό δεν είναι αποδοτικό αφού δεν δίνει ικανοποιητικές εκτιμήσεις που να προσεγγίζουν τις πραγματικές τιμές των παρατηρήσεων και παρουσιάζει μεγάλα σφάλματα. Καταληκτικά, φαίνεται ότι δεν υπάρχει κάποιο καλά αποδοτικό μοντέλο που να κατασκευάζεται με χρήση του αλγορίθμου *AdaBoost* και να προσαρμόζεται στα δεδομένα ώστε να εκτιμά τη μεταβλητή απόκριση με μικρά σφάλματα.

6.4.2 Gradient Boosting

Όμοια με την εφαρμογή του αλγορίθμου *AdaBoost*, έτσι και για τον αλγόριθμο *Gradient Boosting* χρειάζεται να οριστούν συγκεκριμένες παράμετροι για τη δημιουργία μοντέλων που στοχεύουν στην καλύτερη δυνατή εκτίμηση της μεταβλητής απόκρισης του προβλήματος.

Προσδιορισμός Παραμέτρων

Για την κατασκευή ενός μοντέλου βασισμένο στον αλγόριθμο *Gradient Boosting*, οι παράμετροι που ορίζονται αφορούν στη μορφή του δέντρου, αλλά και στις διεργασίες που γίνονται στους κόμβους αυτού. Ειδικότερα, οι παράμετροι που ορίζονται είναι ίδιες με αυτές που περιγράφηκαν στην εφαρμογή του αλγορίθμου *AdaBoost* (βλ. *Παράγραφο 6.4.1*). Ωστόσο, στον συγκεκριμένο αλγόριθμο η συνάρτηση κόστους διαφοροποιείται και συγκεκριμένα υπολογίζεται μέσω της μεθόδου ελαχίστων τετραγώνων.

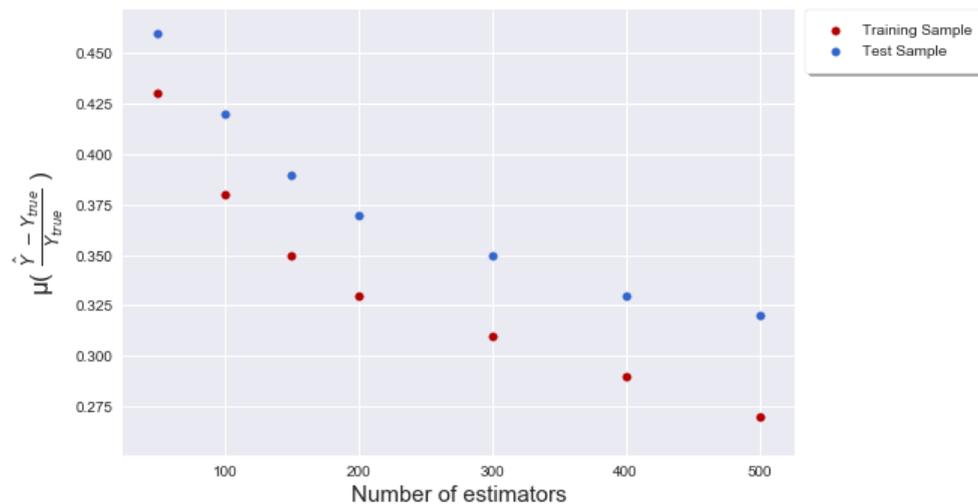
Διερεύνηση Παραμέτρων

Όπως και στην προηγούμενη παράγραφο, κατά την εφαρμογή του αλγορίθμου *Gradient Boosting* είναι απαραίτητο να προσδιοριστεί ο κατάλληλος συνδυασμός του αριθμού των δέντρων που κατασκευάζονται, το βάθος αυτών και ο ρυθμός εκμάθησης για την εύρεση του βέλτιστου μοντέλου. Έτσι, δημιουργούνται αρχικά μοντέλα με σταθερό βάθος και διαφορετικές τιμές για τον αριθμό των δέντρων και το ρυθμό εκμάθησης,

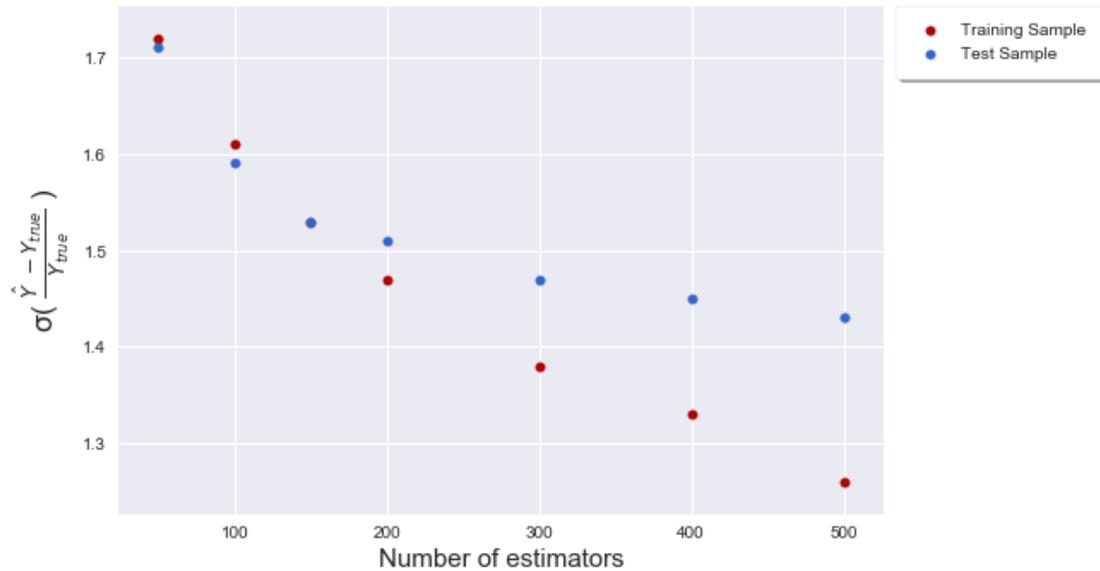
ώστε να βρεθεί ο βέλτιστος συνδυασμός αυτών. Συγκεκριμένα, το βάθος του κάθε δέντρου θέτεται ίσο με 3, ενώ τα συνολικά δέντρα που κατασκευάζονται λαμβάνουν τις τιμές 50, 100, 150, 200, 300, 400 και 500. Αντίστοιχα, ο ρυθμός εκμάθησης λαμβάνει τις τιμές 0.1, 0.2, 0.4, 0.6, 0.8, και 1.0.

Για κάθε ένα από τα 42 μοντέλα βάθους 3 που υλοποιούνται, υπολογίζονται τα σχετικά σφάλματα των εκτιμήσεων, καθώς επίσης και η μέση τιμή και τυπική απόκλιση αυτών (βλ. Παράρτημα). Παράλληλα, για κάθε μοντέλο κατασκευάζεται και το διάγραμμα των εκτιμώμενων τιμών συναρτήσεως των πραγματικών τιμών των παρατηρήσεων. Τα αποτελέσματα αυτά συμβάλουν στην εύρεση του μοντέλου με την καλύτερη προσαρμογή στα δεδομένα, αλλά και με τις μικρότερες τιμές σφαλμάτων. Επομένως, η επιλογή του βέλτιστου μοντέλου κρίνεται από τη μέση τιμή και τυπική απόκλιση των σχετικών σφαλμάτων, ώστε οι αποκλίσεις των εκτιμήσεων από τις παρατηρηθείσες τιμές να μην είναι μεγάλες. Επιπλέον, οι τιμές που λαμβάνουν οι ποσότητες αυτές δεν πρέπει να παρουσιάζουν μεγάλες διαφορές μεταξύ του συνόλου εκπαίδευσης και αξιολόγησης, ώστε το μοντέλο να μην επικεντρώνεται στην εκμάθηση μοτίβων που τυχόν να υπάρχουν στο πρώτο σύνολο και να μην μπορεί να προσαρμοστεί στο δεύτερο.

Συνδυάζοντας τα αποτελέσματα που προκύπτουν, φαίνεται ότι ένα καλό μοντέλο είναι αυτό με μέγιστο αριθμό δέντρων 400 και ρυθμό εκμάθησης 0.6. Ωστόσο, το μοντέλο που παρουσιάζει τη μικρότερη διαφορά για τη μέση τιμή και την τυπική απόκλιση μεταξύ των συνόλων εκπαίδευσης και αξιολόγησης αποτελείται από 150 δέντρα με ρυθμό εκμάθησης 0.6, όπως φαίνεται στα Διαγράμματα 6.13 και 6.14.

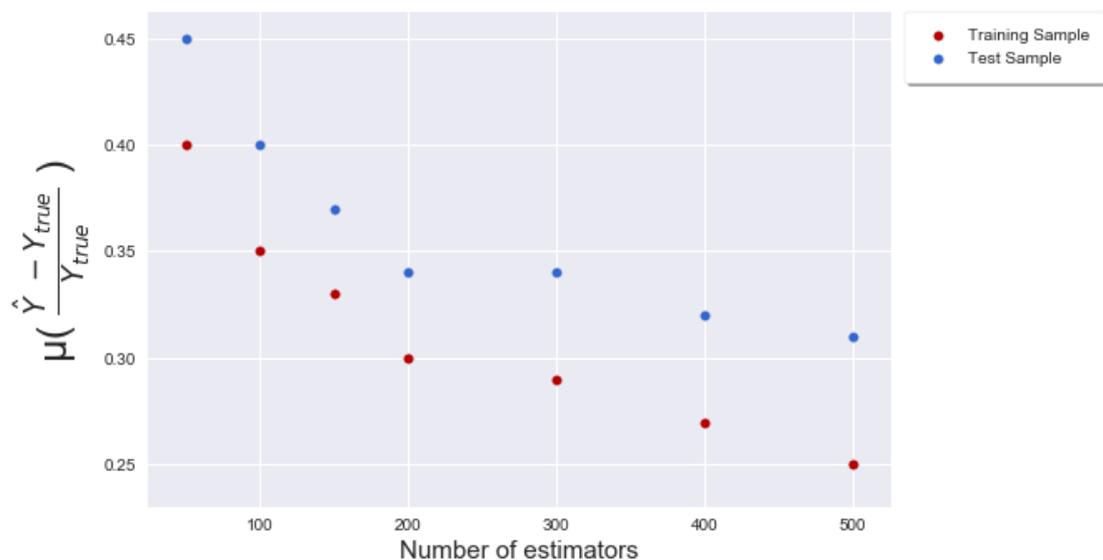


Διάγραμμα 6.13 : Διάγραμμα μέσης τιμής των σχετικών σφαλμάτων των εκτιμήσεων για μοντέλα με δέντρα βάθους 3 και ρυθμό εκμάθησης 0.6 για τις διάφορες τιμές του αριθμού των δέντρων.

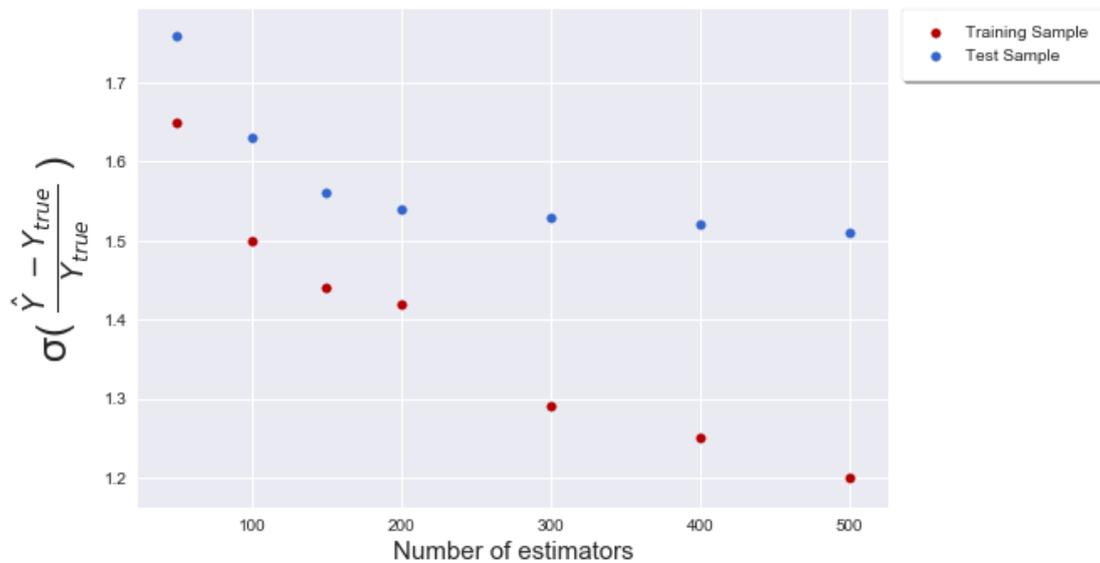


Διάγραμμα 6.14 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων για μοντέλα με δέντρα βάθους 3 και ρυθμό εκμάθησης 0.6 για τις διάφορες τιμές του αριθμού των δέντρων.

Επιπλέον, παρατηρείται ότι καθώς αυξάνεται η τιμή του ρυθμού εκμάθησης, για μοντέλα με ελάχιστο αριθμό δέντρων 300, η μέση τιμή και η τυπική απόκλιση των σχετικών σφαλμάτων αρχίζει να διαφέρει σημαντικά μεταξύ του συνόλου εκπαίδευσης και αξιολόγησης. Συγκεκριμένα, οι τιμές αυτές ολοένα και μειώνονται για το σύνολο εκπαίδευσης, ενώ αντίθετα αυξάνονται για το σύνολο αξιολόγησης, όπως φαίνεται στα Διαγράμματα 6.15 και 6.16 που αντιστοιχούν σε μοντέλα δέντρων βάθους 3 και ρυθμού εκμάθησης 0.8.

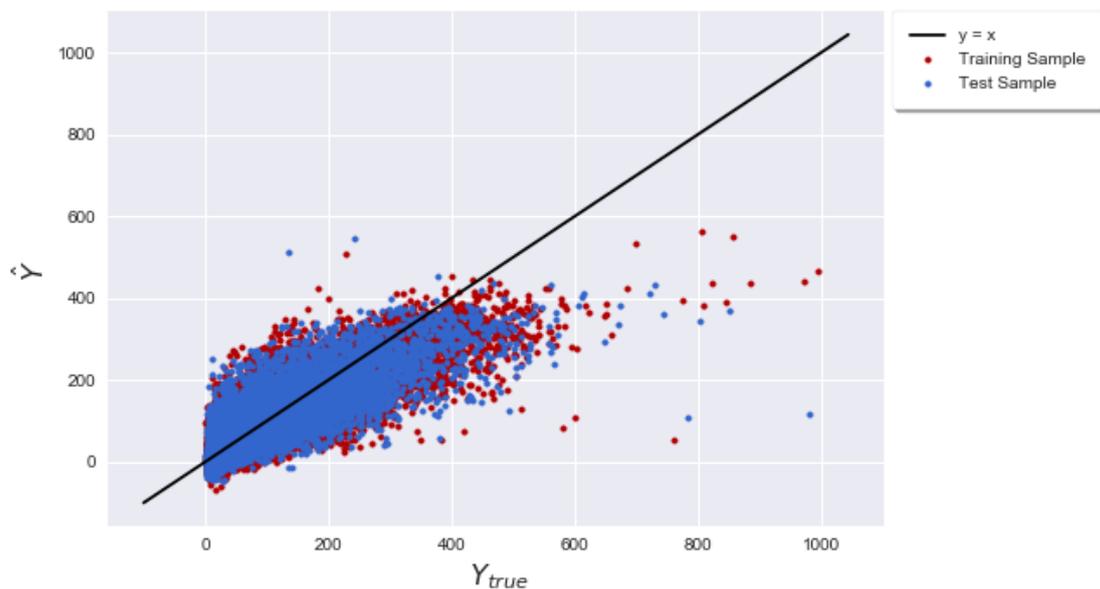


Διάγραμμα 6.15 : Διάγραμμα μέσης τιμής των σχετικών σφαλμάτων των εκτιμήσεων για μοντέλα με δέντρα βάθους 3 και ρυθμό εκμάθησης 0.8 για τις διάφορες τιμές του αριθμού των δέντρων.



Διάγραμμα 6.16 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων για μοντέλα με δέντρα βάθους 3 και ρυθμό εκμάθησης 0.8 για τις διάφορες τιμές του αριθμού των δέντρων.

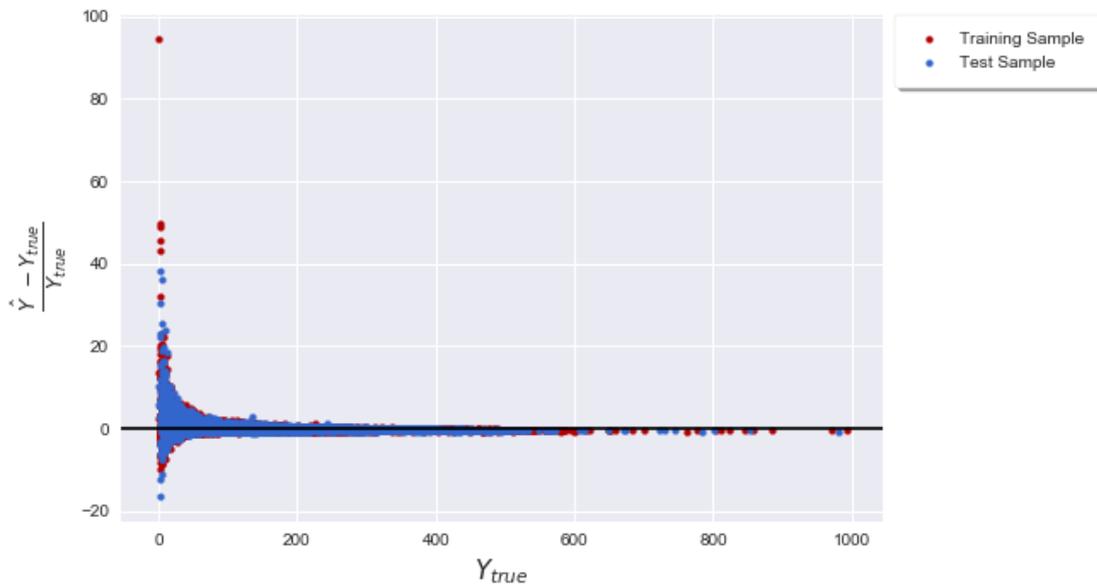
Καταληκτικά, συμπεραίνεται ότι το καλύτερο μοντέλο με δέντρα βάθους 3 αποτελείται από 150 δέντρα και ρυθμό εκμάθησης 0.6. Το Διάγραμμα 6.17 περιγράφει τις εκτιμώμενες τιμές όπως αυτές προκύπτουν από το μοντέλο αυτό συναρτήσει των πραγματικών τιμών των παρατηρήσεων, ενώ στο Διάγραμμα 6.18 αναπαρίστανται τα σχετικά σφάλματα των εκτιμήσεων σε σχέση με τις παρατηρηθείσες τιμές.



Διάγραμμα 6.17 : Διάγραμμα των εκτιμώμενων τιμών των παρατηρήσεων συναρτήσει των πραγματικών τους τιμών για μοντέλο με 150 δέντρα βάθους 3 και ρυθμό εκμάθησης 0.6.

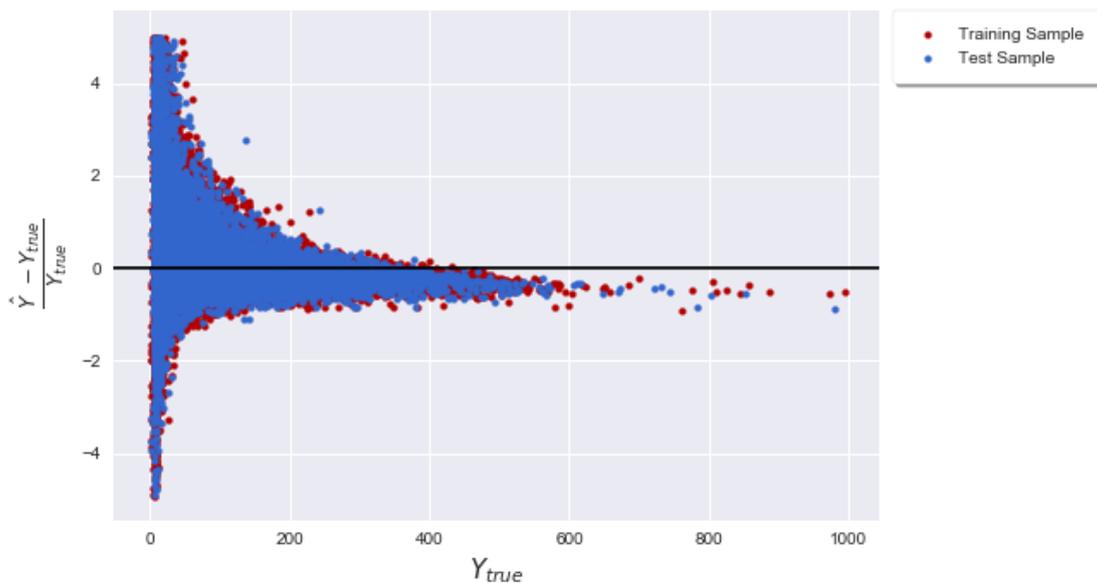
Από το Διάγραμμα 6.17 φαίνεται ότι υπάρχουν σημεία που ακολουθούν την ευθεία $y=x$, ωστόσο υπάρχουν και άλλα για τα οποία παρατηρείται μία απόκλιση μεταξύ της εκτιμώμενης και της πραγματικής τιμής των παρατηρήσεων. Αυτό σημαίνει πως

υπάρχουν τιμές για τις οποίες το μοντέλο δεν μπορεί να δώσει εκτιμήσεις που να προσεγγίζουν την πραγματική τους τιμή.

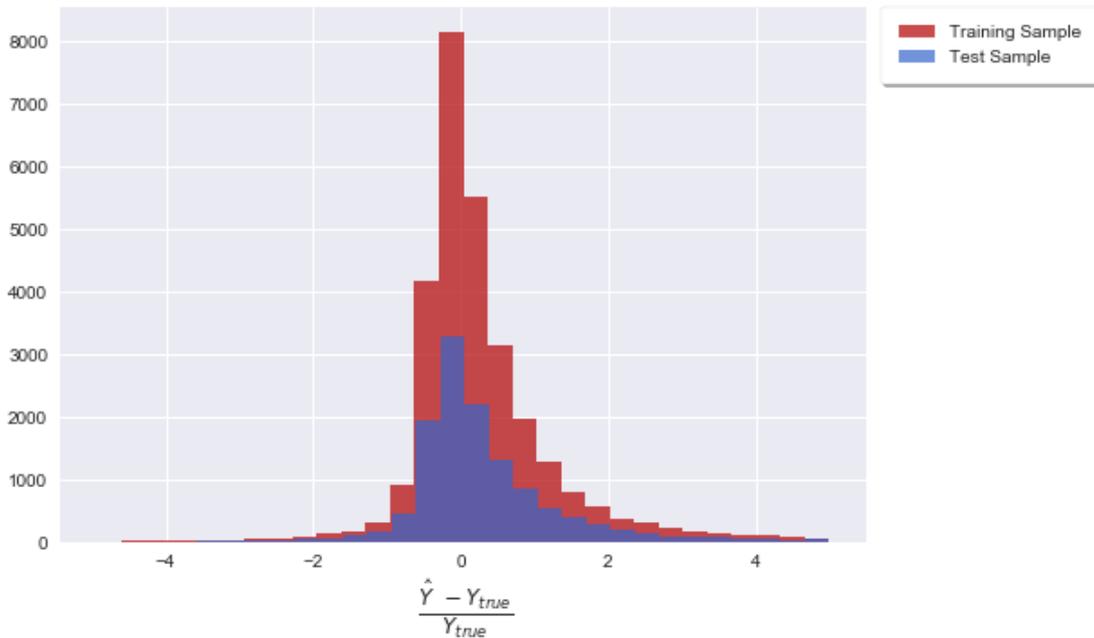


Διάγραμμα 6.18 : Διάγραμμα σχετικών σφαλμάτων για τις εκτιμήσεις του μοντέλου με 150 δέντρα βάθους 3 και ρυθμό εκμάθησης 0.6, συναρτήσει των πραγματικών τιμών των παρατηρήσεων.

Για την καλύτερη κατανόηση της συμπεριφοράς του μοντέλου, δημιουργείται η αντίστοιχη γραφική παράσταση με το Διάγραμμα 6.18, για αποκομμένες τιμές των σχετικών σφαλμάτων στο διάστημα $[-5, 5]$ στο Διάγραμμα 6.19. Με αυτόν τον τρόπο, εστιάζεται η προσαρμογή του μοντέλου αφαιρώντας ακραίες τιμές που υπονομεύουν την απόδοσή του.



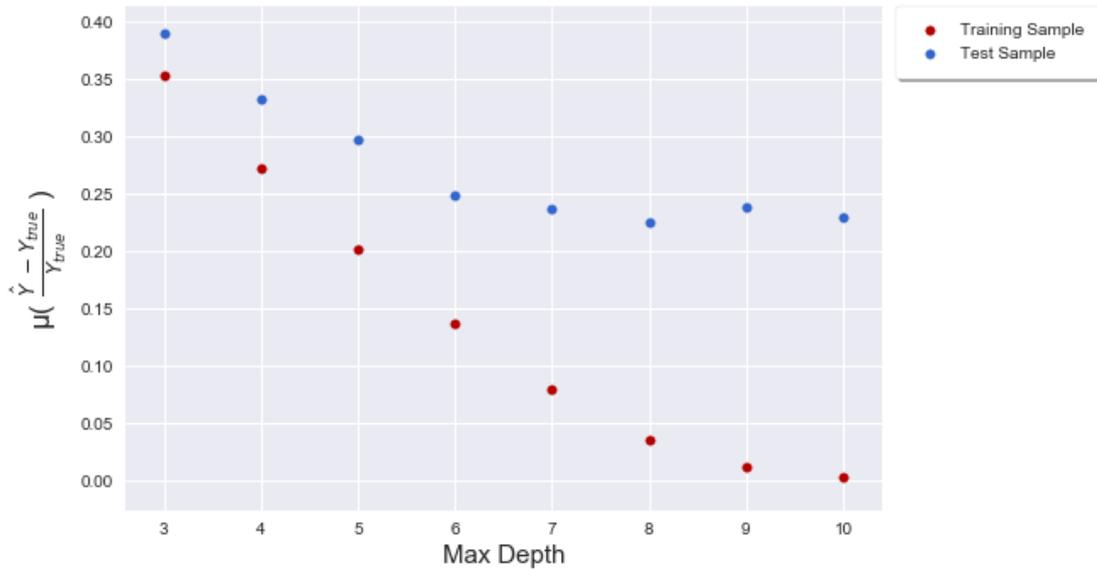
Διάγραμμα 6.19 : Τα σχετικά σφάλματα για τις εκτιμήσεις του μοντέλου 150 δέντρων βάθους 3 με ρυθμό εκμάθησης 0.6, συναρτήσει των πραγματικών τιμών των παρατηρήσεων στο διάστημα $[-5, 5]$.



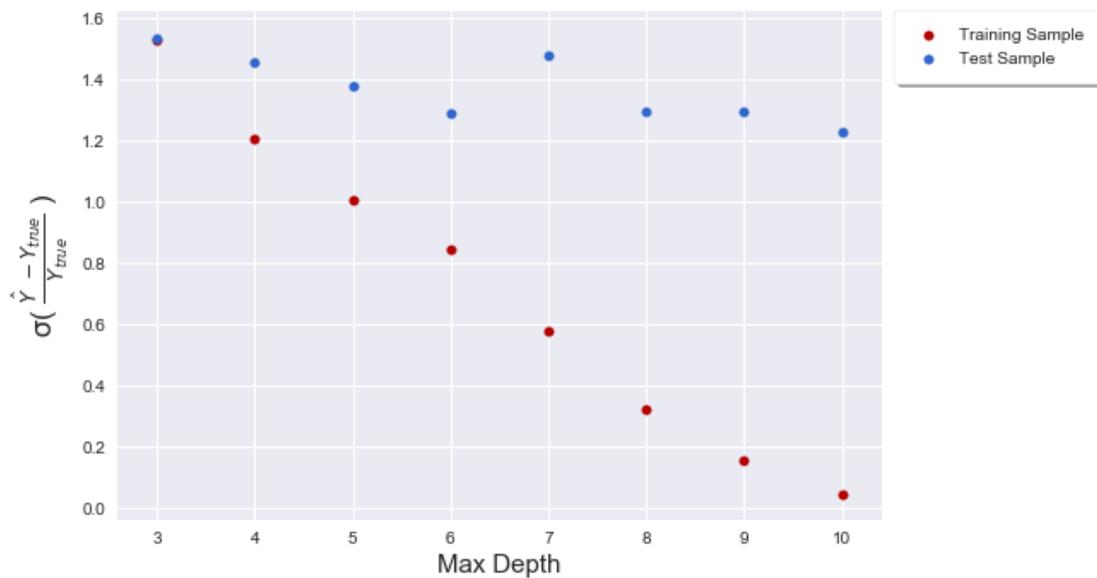
Διάγραμμα 6.20 : Ιστόγραμμα των σχετικών σφαλμάτων για τις εκτιμήσεις του μοντέλου 150 δέντρων βάθους 3 με ρυθμό εκμάθησης 0.6 που λαμβάνουν τιμές στο διάστημα $[-5, 5]$.

Από το Διάγραμμα 6.19 φαίνεται ότι για κάποιες τιμές των παρατηρήσεων, οι εκτιμήσεις που προκύπτουν από το μοντέλο παρουσιάζουν μεγάλο σχετικό σφάλμα. Αυτό παρατηρείται για ορισμένες μικρές παρατηρηθείσες τιμές, που μπορεί να οφείλεται σε υψηλές συσχετίσεις της μεταβλητής απόκρισης με τις επεξηγηματικές μεταβλητές, όταν αυτή λαμβάνει μεγαλύτερες τιμές. Επιπρόσθετα, από το ιστόγραμμα του Διαγράμματος 6.20 υπολογίζεται πως το 81% των εκτιμήσεων του συνόλου εκπαίδευσης και το 83% του συνόλου αξιολόγησης παρουσιάζουν σχετικό σφάλμα μικρότερο της μονάδας, ενώ το 93% και των δύο συνόλων αντιστοιχεί σε σχετικό σφάλμα μικρότερο των 2 μονάδων.

Σύμφωνα με τις παραπάνω παρατηρήσεις, διαπιστώνεται ότι το μοντέλο 150 δέντρων βάθους 3 με ρυθμό εκμάθησης 0.6 επιδέχεται βελτίωσης. Για το λόγο αυτό, καθώς ο συνδυασμός του αριθμού των δέντρων και του ρυθμού εκμάθησης φάνηκε ο πιο ικανοποιητικός για τα διάφορα μοντέλα δέντρων βάθους 3, κρατείται σταθερός και γίνεται έλεγχος για τη βέλτιστη τιμή του μέγιστου βάθους που πρέπει να έχουν τα δέντρα. Έτσι, το βάθος των δέντρων λαμβάνει ακέραιες τιμές στο διάστημα $[3, 10]$ και για κάθε μοντέλο που προκύπτει κατασκευάζονται τα διαγράμματα των εκτιμώμενων τιμών και των σχετικών σφαλμάτων τους συναρτήσει των πραγματικών τιμών των παρατηρήσεων. Επίσης, για κάθε μοντέλο υπολογίζεται η μέση τιμή και τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων και αναπαρίστανται στο Διάγραμμα 6.21 και 6.22 αντίστοιχα.



Διάγραμμα 6.21 : Διάγραμμα μέσης τιμής των σχετικών σφαλμάτων των εκτιμήσεων που προκύπτουν από μοντέλα 150 δέντρων με ρυθμό εκμάθησης 0.6, για κάθε διαφορετική τιμή βάθους.



Διάγραμμα 6.22 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων που προκύπτουν από μοντέλα 150 δέντρων με ρυθμό εκμάθησης 0.6, για κάθε διαφορετική τιμή βάθους.

Από τα Διαγράμματα 6.21 και 6.22 σε συνδυασμό με τα διαγράμματα που περιγράφουν τις εκτιμώμενες τιμές που προκύπτουν από το κάθε μοντέλο συναρτήσεως των πραγματικών παρατηρούμενων τιμών, συμπεραίνεται ότι για μέγιστο βάθος δέντρων μεγαλύτερο του 6 ανιχνεύεται το φαινόμενο της υπερεκπαίδευσης. Συγκεκριμένα, από

τα δύο τελευταία διαγράμματα, όταν το μέγιστο βάθος δέντρου λαμβάνει τιμές μεγαλύτερες του 6, παρατηρείται πως τόσο η μέση τιμή όσο και η τυπική απόκλιση των σχετικών σφαλμάτων μειώνονται απότομα για το σύνολο εκπαίδευσης, ενώ αντίθετα αυξάνονται για το σύνολο αξιολόγησης. Αυτό σημαίνει ότι το μοντέλο εξειδικεύεται σε τυχόν μοτίβα που ανιχνεύονται στα δεδομένα εκπαίδευσης και ελαχιστοποιεί το σφάλμα των εκτιμήσεων για αυτά, ενώ δεν μπορεί να προσαρμοστεί στα δεδομένα αξιολόγησης, με αποτέλεσμα η απόδοσή του για αυτά να μην βελτιώνεται και να μην είναι ικανοποιητική.

Επομένως, είναι αναγκαία μία εκ νέου διερεύνηση για την εύρεση της κατάλληλης τριπλέτας συντελεστών για τον αριθμό και το βάθος των δέντρων του μοντέλου, καθώς και για το ρυθμό εκπαίδευσης. Από τα παραπάνω αποτελέσματα φαίνεται πως το μοντέλο με βάθος 6 είναι αρκετά ικανοποιητικό, επομένως χρειάζεται πειραματισμός για την τελική επιλογή του αριθμού των δέντρων και του ρυθμού εκμάθησης. Για το λόγο αυτό το βάθος των δέντρων σταθεροποιείται και ισούται με 6, ενώ τα συνολικά δέντρα που κατασκευάζονται λαμβάνουν τις τιμές 50, 100, 150, 200, 300, 400 και 500. Αντίστοιχα, ο ρυθμός εκμάθησης λαμβάνει τις τιμές 0.1, 0.2, 0.4, 0.6, 0.8, και 1.0.

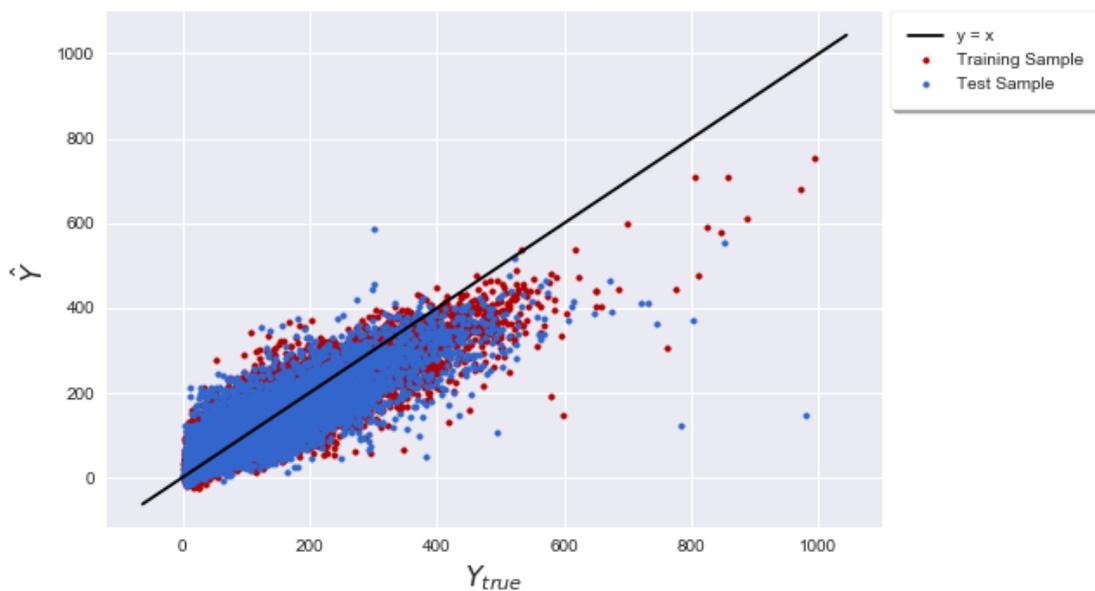
Επαναλαμβάνοντας όμοια στρατηγική με παραπάνω, για κάθε ένα από τα μοντέλα που προκύπτουν υπολογίζονται τα σχετικά σφάλματα των εκτιμήσεων, καθώς επίσης και η μέση τιμή και τυπική απόκλιση αυτών. Παράλληλα, για το κάθε μοντέλο κατασκευάζεται το διάγραμμα των εκτιμώμενων τιμών σε σχέση με τις πραγματικές τιμές των παρατηρήσεων, στο οποίο αποτυπώνεται η απόδοσή του και αποτελεί ένα επιπλέον βοηθητικό μέσο για την επιλογή του βέλτιστου. Σύμφωνα με τα αποτελέσματα που προκύπτουν, φαίνεται πως τα καλύτερα μοντέλα είναι αυτά που κατασκευάζονται για μέγιστο αριθμό δέντρων 200 με ρυθμό εκμάθησης το πολύ 0.4, καθώς για μεγαλύτερες τιμές αριθμού των δέντρων, αλλά και ρυθμού εκμάθησης παρατηρείται το φαινόμενο της υπερεκπαίδευσης. Πιο συγκεκριμένα, για κάθε μία από τις διαφορετικές τιμές ρυθμού εκμάθησης, όσο αυξάνεται ο αριθμός των δέντρων που κατασκευάζονται, σημειώνεται κατακόρυφη πτώση τόσο της μέσης τιμής, όσο και της τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων για το σύνολο εκπαίδευσης, ενώ οι αντίστοιχες τιμές του συνόλου αξιολόγησης δεν παρουσιάζουν την ίδια συμπεριφορά. Συνεπώς, αποδεκτά είναι τα μοντέλα που προκύπτουν από το συνδυασμό αριθμού δέντρων και ρυθμού εκμάθησης για μέγιστη τιμή 200 και 0.4 αντίστοιχα.

Συνδυάζοντας τα αποτελέσματα που προκύπτουν από τα διαγράμματα που κατασκευάζονται για το κάθε μοντέλο, όπως αυτά αναφέρονται παραπάνω, καθώς επίσης και τα διαγράμματα υπολοίπων και σχετικών σφαλμάτων, φαίνεται πως τα πιο ικανοποιητικά μοντέλα είναι τα ακόλουθα :

- [100, 6, 0.1] ,
- [150, 6, 0.1] ,
- [200, 6, 0.1] ,
- [150, 6, 0.2] και
- [50, 6, 0.4] ,

όπου ο παραπάνω συμβολισμός $[x, y, z]$ αναφέρεται στο μοντέλο που κατασκευάζεται με συνολικό αριθμό δέντρων x βάθους y και με ρυθμό εκμάθησης z .

Οι διαφορές των παραπάνω μοντέλων δεν είναι ιδιαίτερα σημαντικές, ωστόσο τα δύο επικρατέστερα είναι τα $[200, 6, 0.1]$ και $[150, 6, 0.2]$. Συγκρίνοντας τα διαγράμματα των εκτιμώμενων τιμών συναρτήσεως των πραγματικών τιμών των παρατηρήσεων, όπως επίσης και αυτά των σχετικών σφαλμάτων των εκτιμήσεων των δύο αυτών μοντέλων, φαίνεται ότι έχουν παρόμοια προσαρμογή στα δεδομένα. Ωστόσο, παρατηρείται ότι στο δεύτερο μοντέλο η διαφορά τόσο της μέσης τιμής, όσο και της τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων μεταξύ των συνόλων εκπαίδευσης και αξιολόγησης είναι αρκετά μεγαλύτερη συγκριτικά με αυτή που παρουσιάζει το πρώτο μοντέλο. Αυτό σημαίνει πως οι εκτιμήσεις που δίνει το δεύτερο μοντέλο για τις παρατηρηθείσες τιμές έχουν μεγαλύτερη απόκλιση από τις πραγματικές τους τιμές, ενώ αντίθετα το πρώτο μοντέλο δίνει εκτιμήσεις μικρότερου σφάλματος. Καταληκτικά, συμπεραίνεται ότι το βέλτιστο μοντέλο που προκύπτει με εφαρμογή του αλγορίθμου Gradient Boosting αποτελείται από 200 δέντρα βάθους 6, με ρυθμό εκμάθησης 0.1.



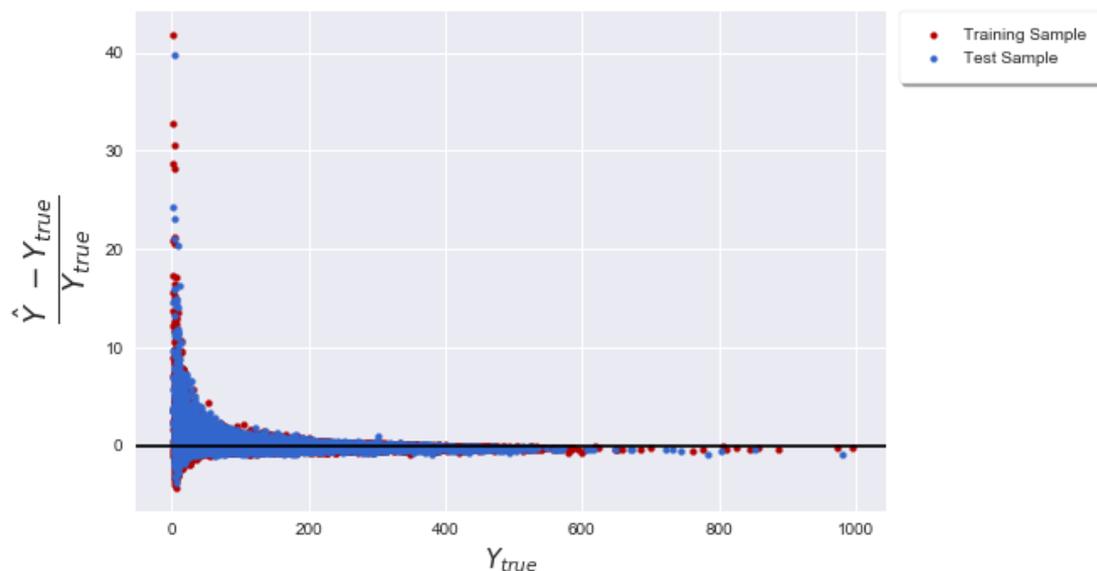
Διάγραμμα 6.183: Διάγραμμα των εκτιμώμενων τιμών των παρατηρήσεων συναρτήσεως των πραγματικών τους τιμών για τα σύνολα εκπαίδευσης και αξιολόγησης για μοντέλο με 200 δέντρα βάθους 6 και ρυθμό εκμάθησης 0.1.

Στο Διάγραμμα 6.23 αναπαρίστανται οι εκτιμώμενες τιμές των παρατηρήσεων για τα σύνολα εκπαίδευσης και αξιολόγησης συναρτήσεως των πραγματικών παρατηρούμενων τιμών. Με την κόκκινη απόχρωση (•) αναπαρίστανται τα σημεία που ανήκουν στο σύνολο εκπαίδευσης, ενώ αυτά με τον μπλε χρωματισμό (•) ανήκουν στο σύνολο αξιολόγησης. Από το διάγραμμα διαπιστώνεται πως το μοντέλο έχει ικανοποιητική προσαρμογή στα δεδομένα, καθώς τα περισσότερα σημεία που τοποθετούνται σε αυτό

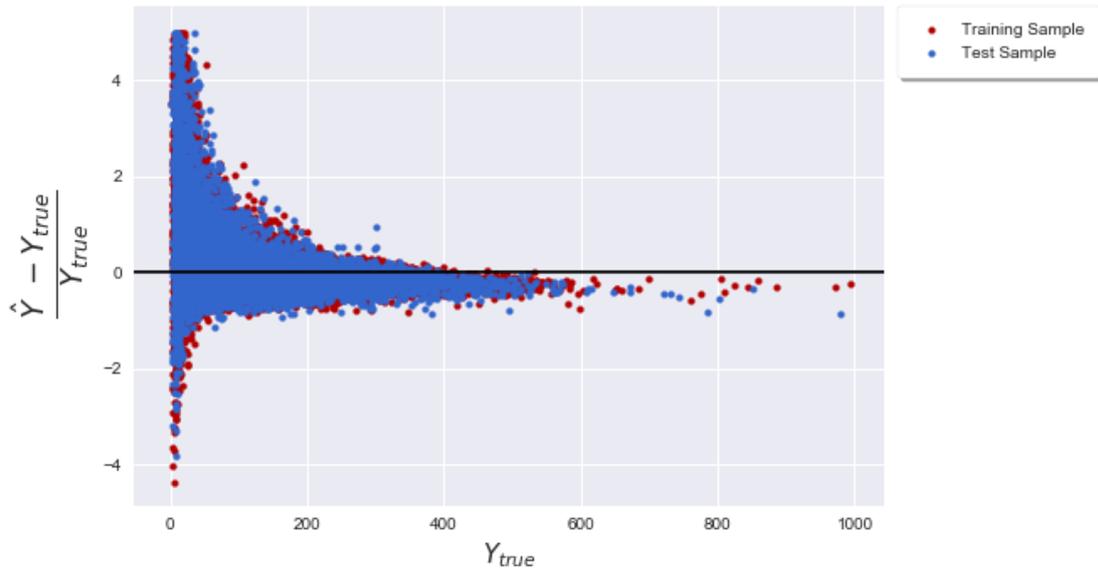
φαίνεται να ακολουθούν νοητά την ευθεία $y=x$ χωρίς να αποκλίνουν με μεγάλες διαφορές από αυτή. Αυτό σημαίνει ότι πολλές εκτιμήσεις που δίνει το μοντέλο δεν απέχουν σημαντικά από τις πραγματικές παρατηρούμενες τιμές. Ωστόσο, υπάρχουν ορισμένες τιμές για τις οποίες το μοντέλο δεν μπορεί να δώσει εκτιμήσεις μικρού σφάλματος. Συγκεκριμένα, φαίνεται να είναι ευαίσθητο στην εκτίμηση αρκετά μικρών παρατηρηθεισών τιμών, καθώς τις αντιστοιχίζει σε μεγαλύτερες τιμές, όπως για παράδειγμα τιμές μικρότερες του 50 που πλησιάζουν το 0, για τις οποίες δίνει πολύ μεγαλύτερες εκτιμήσεις – έως και 100.

Το παραπάνω αποφαίνεται και με τη βοήθεια του Διαγράμματος 6.24, το οποίο περιγράφει τα σχετικά σφάλματα των εκτιμήσεων του μοντέλου συναρτήσει των πραγματικών τιμών των παρατηρήσεων.

Από το διάγραμμα φαίνεται πως για αρκετές από τις εκτιμήσεις του μοντέλου τα σχετικά σφάλματα που υπολογίζονται συγκεντρώνονται σε μικρή απόσταση κατά μήκος του θετικού ημιάξονα Ox . Υπάρχουν όμως και εκτιμήσεις για τις οποίες τα σχετικά σφάλματα λαμβάνουν μεγάλες τιμές και τα αντίστοιχα σημεία είναι απομακρυσμένα από τον ημιάξονα Ox . Ωστόσο επειδή τα σημεία αυτά δεν είναι πολλά – λιγότερο από το 5% των συνολικών δεδομένων – και ορισμένα απέχουν αρκετά από τα υπόλοιπα μπορούν να χαρακτηριστούν ως ακραίες τιμές. Για την περαιτέρω μελέτη των σχετικών σφαλμάτων δημιουργείται ένα αντίστοιχο διάγραμμα του Διαγράμματος 6.24, Διάγραμμα 6.25, στο οποίο οι τιμές των σχετικών σφαλμάτων ανήκουν στο διάστημα $[-5, 5]$ ώστε να υπάρχει μεγαλύτερη ευκρίνεια στην κατανομή των σημείων. Με αυτό τον τρόπο μπορεί να εκτιμηθεί καλύτερα η απόδοση του μοντέλου, αφού αφαιρούνται οι ακραίες τιμές που μπορεί να είναι παραπλανητικές για την επίδοσή του.

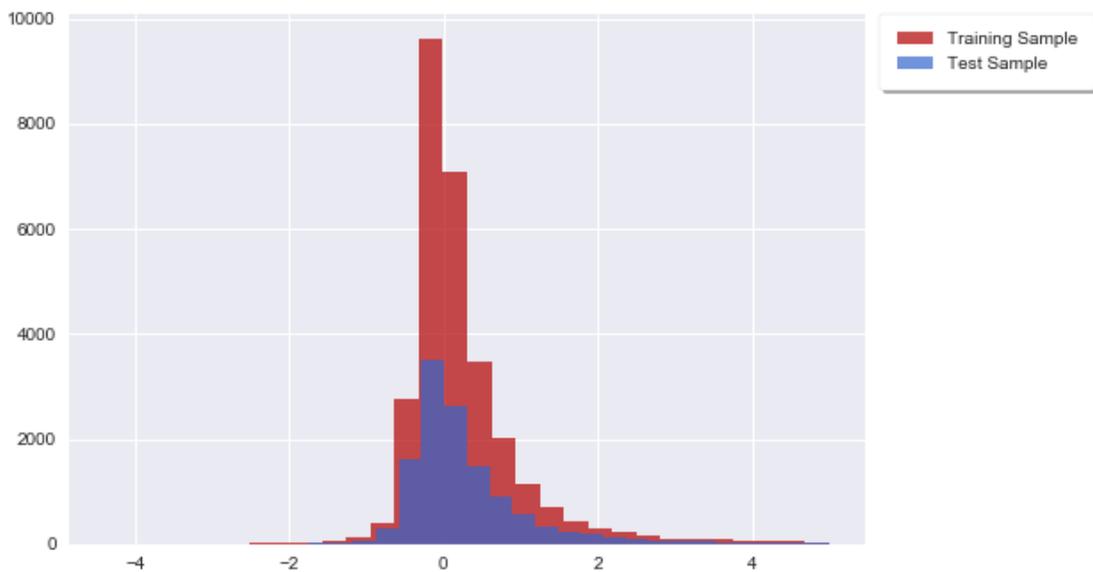


Διάγραμμα 6.24 : Τα σχετικά σφάλματα για τις εκτιμήσεις του μοντέλου 200 δέντρων βάθους 6 με ρυθμό εκμάθησης 0.1, συναρτήσει των πραγματικών τιμών των παρατηρήσεων.



Διάγραμμα 6.25 : Τα σχετικά σφάλματα για τις εκτιμήσεις του μοντέλου 200 δέντρων βάθους 6 με ρυθμό εκμάθησης 0.1, συναρτήσει των πραγματικών τιμών των παρατηρήσεων στο διάστημα $[-5, 5]$.

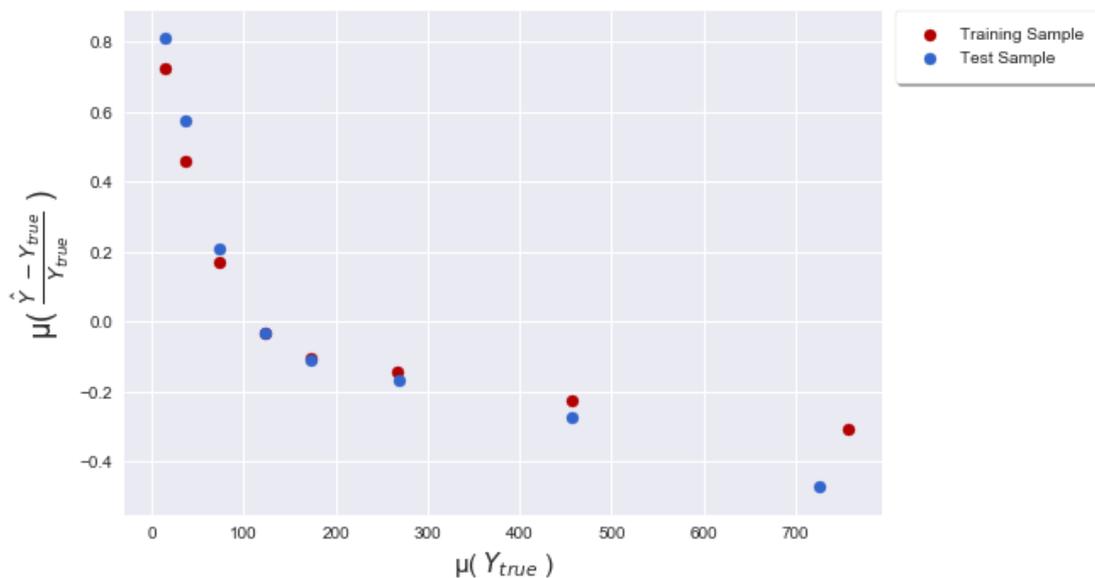
Στο Διάγραμμα 6.26, το οποίο ακολουθεί, αναπαρίστανται τα ιστογράμματα συχνοτήτων των αποκομμένων σχετικών σφαλμάτων στο διάστημα $[-5, 5]$. Από αυτό υπολογίζεται ότι σχεδόν το 89% των εκτιμήσεων του συνόλου εκπαίδευσης και το 87% του συνόλου αξιολόγησης αντιστοιχούν σε σχετικό σφάλμα μικρότερο της μονάδας.



Διάγραμμα 6.26 : Ιστόγραμμα των σχετικών σφαλμάτων για τις εκτιμήσεις του μοντέλου 200 δέντρων βάθους 6 με ρυθμό εκμάθησης 0.1 που λαμβάνουν τιμές στο διάστημα $[-5, 5]$.

Ακολούθως, οι πραγματικές τιμές των παρατηρήσεων χωρίζονται σε υποσύνολα με τις αντίστοιχες τιμές των σχετικών σφαλμάτων στο διάστημα $[-5, 5]$, για μία καλύτερη και

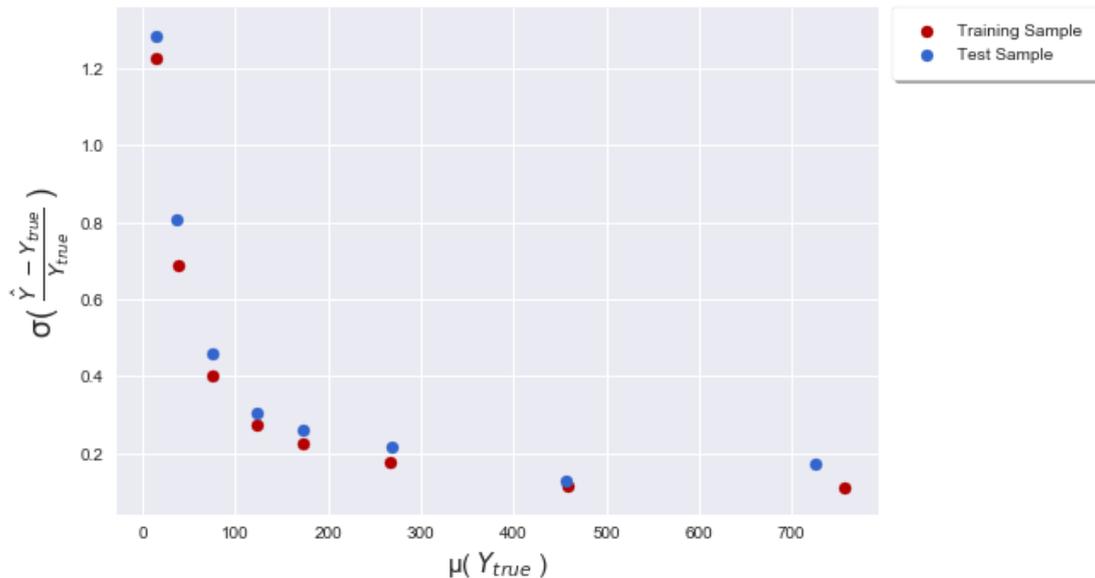
με μεγαλύτερη ευκρίνεια εκτίμηση της απόδοσης του μοντέλου για τις διάφορες παρατηρηθείσες τιμές. Πιο συγκεκριμένα, το σύνολο των πραγματικών τιμών των παρατηρήσεων χωρίζεται αρχικά σε τέσσερα επιμέρους υποσύνολα με εύρος 200 μονάδες τα πρώτα τρία και 400 το τελευταίο. Ωστόσο, επειδή το πεδίο τιμών $[0, 200]$ περιλαμβάνει ένα μεγάλο αριθμό παρατηρήσεων, αυτό διαμερίζεται σε άλλα πέντε υποσύνολα, εύρους 50 μονάδων το καθένα με εξαίρεση τα πρώτα δύο που έχουν εύρος 25 μονάδες, με σκοπό η συμπεριφορά του μοντέλου να είναι πιο ξεκάθαρη. Στα Διαγράμματα 6.27 και 6.28 που ακολουθούν, αναπαρίστανται οι τιμές για τη μέση τιμή και την τυπική απόκλιση των αποκομμένων σχετικών σφαλμάτων στο διάστημα $[-5, 5]$ για τα σύνολα εκπαίδευσης και αξιολόγησης, συναρτήσει της μέσης τιμής των πραγματικών παρατηρήσεων για κάθε ένα από τα υποσύνολα που προέκυψαν, όπως περιγράφηκαν παραπάνω.



Διάγραμμα 6.27 : Διάγραμμα μέσης τιμής των αποκομμένων σχετικών σφαλμάτων των εκτιμήσεων στο διάστημα $[-5, 5]$ συναρτήσει της μέσης τιμής των παρατηρηθεισών τιμών για κάθε υποσύνολο αυτών.

Τόσο από το Διάγραμμα 6.27, όσο και από το Διάγραμμα 6.28 που ακολουθεί, παρατηρείται πως οι τιμές για τη μέση τιμή και την τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων δεν παρουσιάζουν σημαντικές διαφορές μεταξύ του συνόλου εκπαίδευσης και του συνόλου αξιολόγησης. Ειδικότερα, φαίνεται πως όταν οι τιμές των παρατηρήσεων είναι μεγαλύτερες από 100, τότε οι αποκλίσεις είναι ιδιαίτερα μικρές, ενώ παράλληλα το σφάλμα που σημειώνεται είναι και αυτό ελάχιστο. Με άλλα λόγια, τόσο η μέση τιμή, όσο και η τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων μειώνονται καθώς αυξάνεται η πραγματική τιμή των παρατηρήσεων, που σημαίνει ότι οι εκτιμώμενες τιμές δεν παρουσιάζουν σημαντικές διαφορές με τις πραγματικές τιμές των παρατηρήσεων.

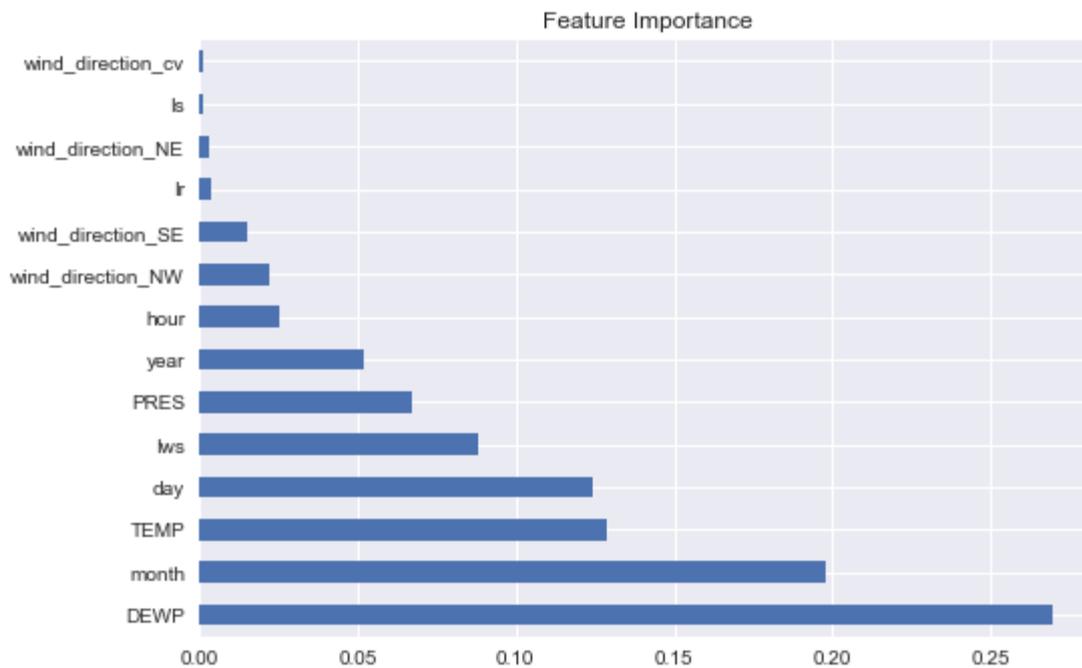
Συνεπώς, φαίνεται πως το μοντέλο ανταποκρίνεται ικανοποιητικά στην εκτίμηση των παρατηρηθεισών τιμών, παρά την ευαισθησία του στην πρόβλεψη των μικρών τιμών του δείγματος – όπως φαίνεται στα Διαγράμματα 6.23 και 6.24. Αυτό συμβαίνει λόγω των υψηλών συσχετίσεων που παρουσιάζει η μεταβλητή απόκρισης με τις εξηγηματικές μεταβλητές, όταν αυτή λαμβάνει μεγάλες τιμές. Επομένως, για μεγάλες τιμές της μεταβλητής απόκρισης, το μοντέλο φαίνεται να έχει καλύτερη απόδοση, αφού σημειώνονται μικρότερα σφάλματα στις εκτιμήσεις του σε σχέση με αυτές που αντιστοιχούν σε μικρότερες τιμές της.



Διάγραμμα 6.28 : Διάγραμμα τυπικής απόκλισης των αποκομμένων σχετικών σφαλμάτων των εκτιμήσεων στο διάστημα $[-5, 5]$ συναρτήσει της μέσης τιμής των παρατηρηθεισών τιμών για κάθε υποσύνολο αυτών.

Το διάγραμμα που ακολουθεί, Διάγραμμα 6.29, αναπαριστά τη συμβολή της κάθε μεταβλητής στο μοντέλο. Συγκεκριμένα, με την εφαρμογή του αλγορίθμου Gradient Boosting, έπειτα από την κατασκευή των δέντρων υπολογίζεται ο βαθμός σημαντικότητας της κάθε μεταβλητής, ο οποίος δηλώνει πόσο χρήσιμη ή πολύτιμη είναι αυτή κατά τη δημιουργία των δέντρων του μοντέλου. Όσο περισσότερο χρησιμοποιείται μία μεταβλητή στη λήψη βασικών αποφάσεων για τα δέντρα απόφασης που κατασκευάζονται, τόσο πιο σημαντική θα είναι.

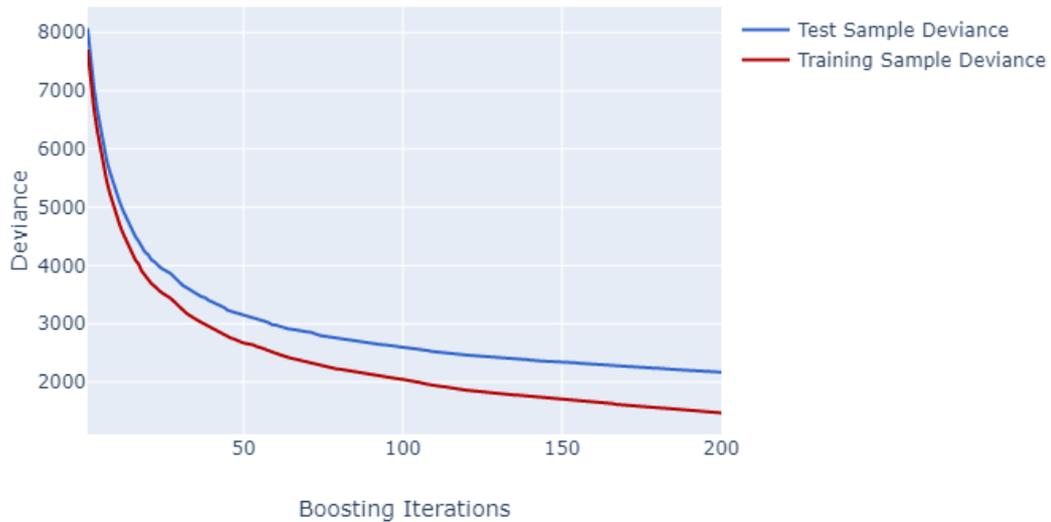
Ο βαθμός σημαντικότητας που αντιστοιχεί στην κάθε μεταβλητή υπολογίζεται έπειτα από την κατασκευή κάθε δέντρου, κατά την επιλογή της μεταβλητής σύμφωνα με την οποία γίνεται ο διαχωρισμός αυτού σε επιπλέον κόμβους, ανάλογα με το πόσο μειώνεται το τετραγωνικό σφάλμα. Δηλαδή, το πόσο σημαντική είναι η κάθε μεταβλητή προκύπτει από το πόσο βελτιώνεται η επίδοση του μοντέλου ελαχιστοποιώντας το σφάλμα, επιλέγοντάς τη ως μεταβλητή διαχωρισμού του κόμβου, σταθμισμένη ανάλογα με τον αριθμό των παρατηρήσεων που βρίσκονταν σε αυτόν πριν διασπαστεί.



Διάγραμμα 6.29 : Διάγραμμα βαθμού σημαντικότητας των μεταβλητών του μοντέλου κατά την εφαρμογή του αλγορίθμου Gradient Boosting σε μοντέλο με συνολικό αριθμό δέντρων 200 με μέγιστο βάθος 6 και ρυθμό εκμάθησης 0.1 .

Όπως διαπιστώνεται από το τελευταίο διάγραμμα, η μεταβλητή που φαίνεται να είναι η πιο σημαντική είναι το σημείο δρόσου (*DEWP*) και στη συνέχεια ακολουθεί ο μήνας κατά τον οποίο έγινε η μέτρηση της παρατήρησης (*month*), καθώς και η θερμοκρασία (*TEMP*). Αυτό σημαίνει πως το σημείο δρόσου παίζει σημαντικό ρόλο ως μεταβλητή διαχωρισμού σε επιπλέον κόμβους, μειώνοντας το τετραγωνικό σφάλμα που υπολογίζεται κάθε φορά.

Το Διάγραμμα 6.30 περιγράφει το σφάλμα που υπολογίζεται σε κάθε βήμα κατά την κατασκευή των 200 δέντρων του μοντέλου. Με την κόκκινη καμπύλη αποτυπώνεται το σφάλμα που υπολογίζεται για το σύνολο εκπαίδευσης, ενώ με την μπλε καμπύλη το αντίστοιχο για το σύνολο αξιολόγησης. Όπως φαίνεται, και οι δύο καμπύλες αρχικά μειώνονται απότομα και στη συνέχεια η μείωση αυτή σταδιακά επιβραδύνεται, ωστόσο συνεχίζουν να ελαχιστοποιούνται όσο αυξάνεται ο αριθμός των δέντρων που κατασκευάζονται.



Διάγραμμα 6.190 : Διάγραμμα συνάρτησης κόστους συναρτήσει του αριθμού των δέντρων που κατασκευάζονται, στο μοντέλο με συνολικό αριθμό δέντρων 200, μέγιστο βάθος 6 και ρυθμό εκμάθησης 0.1, για το σύνολο εκπαίδευσης και αξιολόγησης.

Από το παραπάνω διάγραμμα φαίνεται ότι το σφάλμα που σημειώνεται κατά την κατασκευή του κάθε δέντρου του δείγματος αξιολόγησης έχει μικρή απόκλιση από το αντίστοιχο του δείγματος εκπαίδευσης. Επιπλέον, καθώς ο αριθμός των δέντρων που κατασκευάζονται αυξάνεται, τόσο μειώνεται και το σφάλμα για τα δύο δείγματα, ενώ η διαφορά μεταξύ των δύο καμπυλών δεν αμβλύνει. Συνεπώς, διαπιστώνεται ότι δεν υπάρχει ένδειξη εμφάνισης του φαινομένου της υπερεκπαίδευσης και φαίνεται πως το τελικό αυτό μοντέλο ανταποκρίνεται ικανοποιητικά στα διαθέσιμα δεδομένα.

6.5 Νευρωνικά Δίκτυα

Η κατασκευή ενός Νευρωνικού Δικτύου γίνεται σύμφωνα με όσα περιγράφηκαν στο Κεφάλαιο 5. Ωστόσο, σε σχέση με την εφαρμογή των αλγορίθμων που προηγήθηκαν, για τη δημιουργία ενός νευρωνικού δικτύου είναι απαραίτητη η ειδική μεταχείριση των δεδομένων του δείγματος που διατίθεται, ώστε να μπορούν να εισαχθούν σε αυτό. Επιπλέον, όμοια με τις παραπάνω μεθόδους, ορίζονται κατάλληλες παράμετροι που συμβάλουν στη διαμόρφωση των μοντέλων που προκύπτουν.

Προεπεξεργασία Δεδομένων

Βασικό αρχικό βήμα πριν την κατασκευή του νευρωνικού δικτύου, αποτελεί η τυποποίηση των δεδομένων. Η μέθοδος αυτή έχει ως αποτέλεσμα την αναδιαμόρφωση των τιμών των μεταβλητών, ώστε να έχουν τις ιδιότητες μιας κανονικής κατανομής με τη μέση τιμή και την τυπική απόκλιση να ισούνται με μηδέν και τη μονάδα αντίστοιχα. Ειδικότερα, τα δεδομένα μετασχηματίζονται σύμφωνα με τον τύπο :

$$x' = \frac{x - \mu}{\sigma} \quad (6.1)$$

, όπου x' η νέα μετασχηματισμένη τιμή, x η τιμή που αντιστοιχεί στην κάθε παρατήρηση, μ η μέση τιμή και σ η τυπική απόκλιση της μεταβλητής στην οποία αντιστοιχεί η παρατήρηση x .

Αυτή η διαδικασία κρίνεται απαραίτητη καθώς είναι σημαντικό όλες οι μεταβλητές του προβλήματος να είναι στην ίδια κλίμακα. Στη συγκεκριμένη περίπτωση, όπως φαίνεται και από τον Πίνακα 2.3 στο Κεφάλαιο 2, παρατηρείται ότι υπάρχουν μεταβλητές που διαφέρουν σημαντικά μεταξύ τους. Για παράδειγμα, η μέγιστη τιμή της μεταβλητής που περιγράφει το σημείο δρόσου, *DEWP*, ισούται με 28 βαθμούς, ενώ η αντίστοιχη μέγιστη τιμή για την πίεση, *PRES*, είναι ίση με 1046 hPa. Συνεπώς, η τυποποίηση των δεδομένων είναι αναγκαία με σκοπό να βρίσκονται όλα στην ίδια τάξη κλίμακας.

Με αυτόν τον τρόπο, όλες οι μεταβλητές συνεισφέρουν εξίσου το ίδιο στο μοντέλο που δημιουργείται και αποφεύγεται η επιρροή κάποιας από αυτές δίχως να είναι στατιστικά σημαντική, εξαιτίας των μεγάλων τιμών που λαμβάνει. Επιπλέον, βελτιώνεται η αριθμητική σταθερότητα του μοντέλου, ενώ παράλληλα η διαδικασία εκπαίδευσης γίνεται ταχύτερη.

Προσδιορισμός Παραμέτρων

Για την κατασκευή ενός Νευρωνικού Δικτύου είναι απαραίτητο να οριστούν συγκεκριμένες παράμετροι, ή διαφορετικά *υπερπαράμετροι* όπως παρουσιάζονται στην Παράγραφο 5.7 του Κεφαλαίου 5, που αφορούν στην αρχιτεκτονική του (βλ. *Κεφάλαιο 5.2*), αλλά και στις διεργασίες που γίνονται στους νευρώνες αυτού. Συγκεκριμένα, αυτές αφορούν :

- Τον αριθμό των κρυφών στρωμάτων του δικτύου,
- Το πλήθος των νευρώνων σε κάθε στρώμα του δικτύου,
- Τη συνάρτηση ενεργοποίησης,
- Τη συνάρτηση κόστους,
- Τη μέθοδο βελτιστοποίησης,
- Το μέγεθος του υποσυνόλου των δεδομένων εκπαίδευσης,
- Τον αριθμό των εποχών εκπαίδευσης.

Όσον αφορά τη συνάρτηση κόστους και τη μέθοδο βελτιστοποίησης, επιλέγονται να είναι το Μέσο Τετραγωνικό Σφάλμα και η μέθοδος Adam αντίστοιχα, όπως έχουν οριστεί στην Παράγραφο 5.7 του Κεφαλαίου 5. Στα μοντέλα που κατασκευάζονται, για το μέγεθος του υποσυνόλου των δεδομένων εκπαίδευσης, καθώς και για τον αριθμό των εποχών εκπαίδευσης δοκιμάζονται οι τιμές 50, 100 και 150 για την πρώτη παράμετρο και 500, 1000, 1200 και 1500 για τη δεύτερη αντίστοιχα. Η επιλογή των παραπάνω αριθμών για τις υπαράμετρους αυτές συνδυαστικά μεταξύ τους, πραγματοποιείται για την πιθανή ανίχνευση διαφορετικής συμπεριφοράς του μοντέλου που σχετίζεται με την απόδοσή του. Δηλαδή, αν χρειάζεται μικρότερο ή μεγαλύτερο δείγμα από το σύνολο εκπαίδευσης και λιγότερες ή περισσότερες εποχές ώστε το σφάλμα να είναι το ελάχιστο δυνατό και να μην υπάρχει ένδειξη του φαινομένου της υπερεκπαίδευσης. Επιπλέον, όπως έχει ήδη αναφερθεί και στην Παράγραφο 5.7 του Κεφαλαίου 5, η συνάρτηση ενεργοποίησης που εφαρμόζεται σε κάθε στρώμα του δικτύου μπορεί να διαφέρει από εκείνη κάποιου άλλου, ωστόσο επειδή η συγκεκριμένη μελέτη αφορά πρόβλημα παλινδρόμησης, η συνάρτηση ενεργοποίησης του τελευταίου στρώματος, στρώμα εξόδου, ορίζεται να είναι η γραμμική (βλ. *Παράρτημα, Διαφορετικοί Τύπου Συναρτήσεων Ενεργοποίησης*).

Όσον αφορά τις παραμέτρους που αφορούν τον αριθμό των κρυφών στρωμάτων του δικτύου και το πλήθος των νευρώνων από τους οποίους αποτελούνται, χρειάζεται να γίνει περεταίρω διερεύνηση για τον προσδιορισμό τους. Αντίστοιχα, δοκιμάζονται και διαφορετικές συναρτήσεις ενεργοποίησης μεταξύ των κρυφών στρωμάτων για την εύρεση εκείνης για την οποία το μοντέλο αποδίδει καλύτερα. Η διερεύνηση αυτή είναι απαραίτητη καθώς για τους διάφορους συνδυασμούς των παραμέτρων προκύπτουν μοντέλα με άλλα αποτελέσματα, που οφείλονται στη διαφορετική προσαρμογή του καθενός ανάλογα με τα ορίσματα που δέχεται.

Διερεύνηση Παραμέτρων

Για την εύρεση του βέλτιστου νευρωνικού δικτύου που κατασκευάζεται, είναι απαραίτητη η δοκιμή διάφορων συνδυασμών των παραμέτρων που επιλέγονται για αυτό. Συγκεκριμένα, όσον αφορά τον αριθμό των κρυφών στρωμάτων και το πλήθος των νευρώνων από τους οποίους απαρτίζεται το καθένα, κατασκευάζεται ένας μεγάλος αριθμός δικτύων αρχικά με ένα και στη συνέχεια έως και 3 κρυφά στρώματα, τα οποία σε πρώτο στάδιο αποτελούνται από τόσους νευρώνες όσες και οι μεταβλητές του προβλήματος και έπειτα αυξάνονται πολλαπλασιαστικά μέχρι να βρεθεί το νευρωνικό δίκτυο με την καλύτερη απόδοση. Παράλληλα, πειραματισμός γίνεται και για την επιλογή της κατάλληλης συνάρτησης ενεργοποίησης που εφαρμόζεται μεταξύ των κρυφών στρωμάτων του δικτύου. Συγκεκριμένα, το κάθε νευρωνικό δίκτυο που δημιουργείται με βάση τους συνδυασμούς των παραμέτρων που περιγράφονται παραπάνω, επαναλαμβάνεται για την εφαρμογή των τριών συναρτήσεων ενεργοποίησης που αναφέρονται ακολούθως και περιγράφονται στο Παράρτημα :

- ♦ Relu,
- ♦ Tanh,
- ♦ Sigmoid

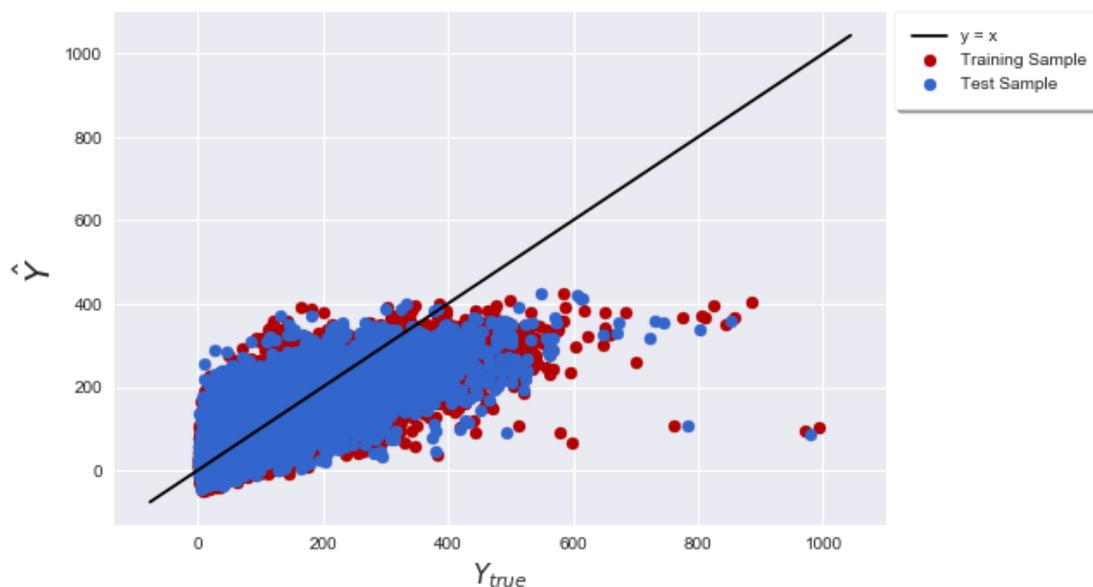
Με βάση τα παραπάνω, εξετάζονται όλα τα μοντέλα που δημιουργούνται και επιλέγεται το βέλτιστο – σύμφωνα με τα διάφορα μέτρα αξιολόγησης, όπως αυτά αναφέρονται στην Παράγραφο 6.2 του Κεφαλαίου 6 – για κάθε νευρωνικό δίκτυο που αποτελείται από συνολικό αριθμό κρυφών στρωμάτων 1, 2 και 3, όπως περιγράφεται στη συνέχεια.

6.5.1 Κατασκευή Νευρωνικών Δικτύων με 1 Κρυφό Στρώμα

Όπως προαναφέρθηκε, αρχικά κατασκευάζονται νευρωνικά δίκτυα για τους διαφορετικούς συνδυασμούς των παραμέτρων όπως περιγράφηκε προηγουμένως και για την κάθε μία συνάρτηση ενεργοποίησης που αναφέρεται παραπάνω αρχικά για 14 νευρώνες στο μοναδικό κρυφό στρώμα του δικτύου – όσες είναι δηλαδή και οι μεταβλητές του προβλήματος – και εν συνεχεία με ολοένα αύξησή τους. Πιο αναλυτικά, δημιουργούνται νευρωνικά δίκτυα με το κρυφό στρώμα να αποτελείται από το διπλάσιο και τριπλάσιο αριθμό νευρώνων από τον αριθμό των μεταβλητών του προβλήματος με διαδοχική αύξηση αυτών μέχρι την εύρεση του δικτύου με τη βέλτιστη απόδοση και τα καλύτερα αποτελέσματα.

Όπως φαίνεται από το Διάγραμμα 6.31, στο οποίο αναπαρίστανται οι εκτιμώμενες τιμές των παρατηρήσεων συναρτήσεων των πραγματικών τους τιμών για νευρωνικό

δίκτυο με ένα κρυφό στρώμα αποτελούμενο από 42 νευρώνες με εφαρμογή της υπερβολικής εφαπτομένης ως συνάρτηση ενεργοποίησης, τα αποτελέσματα δεν είναι ικανοποιητικά. Ειδικότερα, φαίνεται πως δεν γίνεται καλή εκτίμηση των δεδομένων, καθώς τα σημεία που αναπαριστούν τις εκτιμήσεις τοποθετούνται αρκετά απομακρυσμένα από την ευθεία $y=x$, το οποίο δηλώνει πως η εκτιμώμενη τιμή τους απέχει σημαντικά από την πραγματική.



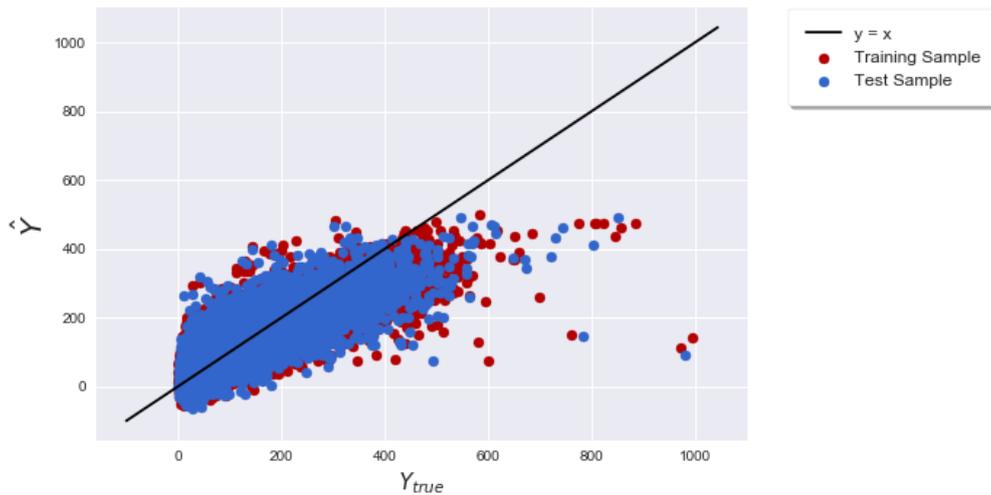
Διάγραμμα 6.31 : Διάγραμμα εκτιμώμενων τιμών των παρατηρήσεων συναρτήσει των πραγματικών τους τιμών για νευρωνικό δίκτυο με ένα κρυφό στρώμα αποτελούμενο από 42 νευρώνες με εφαρμογή της υπερβολικής εφαπτομένης ως συνάρτηση ενεργοποίησης.

Τα αντίστοιχα διαγράμματα για κατασκευή νευρωνικού δικτύου με εφαρμογή των συναρτήσεων Relu και Sigmoid ως συναρτήσεις ενεργοποίησης στο κρυφό στρώμα του δικτύου παρατίθενται στο Παράρτημα. Από το παραπάνω διάγραμμα φαίνεται πως υπάρχουν περιθώρια βελτίωσης της απόδοσης του δικτύου και έτσι το πλήθος των νευρώνων αυξάνεται μέχρι να ανιχνευθεί το κατάλληλο.

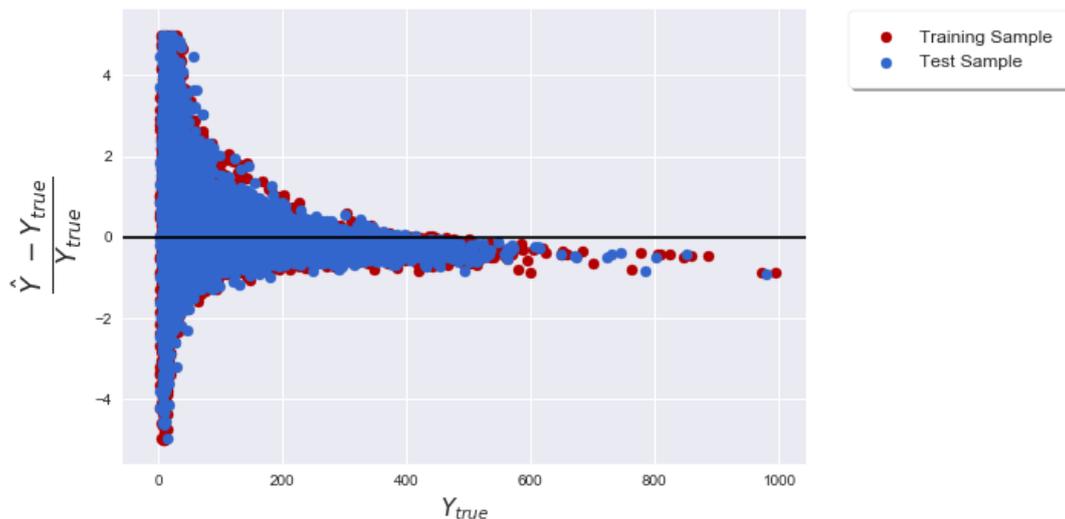
Παρατηρείται ότι καθώς αυξάνεται το πλήθος των νευρώνων που απαρτίζουν το κρυφό στρώμα του δικτύου, τόσο το σφάλμα μειώνεται, το οποίο σημαίνει πως το μοντέλο βελτιώνεται και δίνει καλύτερα αποτελέσματα. Επιπλέον, όπως και στην εφαρμογή των προηγούμενων αλγορίθμων του ίδιου κεφαλαίου, έτσι και στην κατασκευή του νευρωνικού δικτύου υπολογίζονται τα αποκομμένα σχετικά σφάλματα των εκτιμήσεων που προκύπτουν στο διάστημα $[-5, 5]$, με σκοπό να δοθεί μία καλύτερη εικόνα για την απόδοση του μοντέλου, αφαιρώντας τις ακραίες τιμές που υπάρχουν και υπονομεύουν την επίδοσή του. Έτσι, υπολογίζεται η μέση τιμή και η τυπική απόκλιση αυτών τόσο για το σύνολο εκπαίδευσης, όσο και για το σύνολο αξιολόγησης και ανάλογα με τις τιμές που λαμβάνουν επιλέγεται το καταλληλότερο μοντέλο.

Έπειτα από την κατασκευή ποικίλων νευρωνικών δικτύων για τους διάφορους συνδυασμούς που έχουν αναφερθεί, προκύπτει ότι το βέλτιστο δίκτυο που έχει ένα

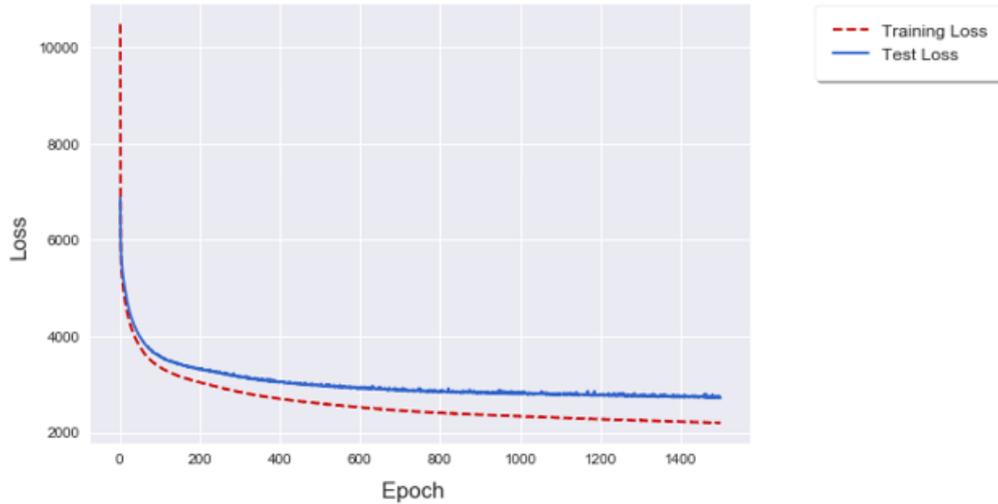
κρυφό στρώμα, αποτελεί αυτό με 300 νευρώνες για αριθμό εποχών 1500 και υποσύνολα δεδομένων εκπαίδευσης μεγέθους 50, ενώ η συνάρτηση ενεργοποίησης που εφαρμόζεται είναι η Relu. Με βάση όλα τα μοντέλα που δημιουργήθηκαν, παρατηρήθηκε πως όταν το κρυφό στρώμα του δικτύου αποτελείται πάνω από 120 νευρώνες, τότε το διάγραμμα των εκτιμώμενων τιμών των παρατηρήσεων συναρτίζει των πραγματικών τους τιμών δεν παρουσιάζει σημαντικές διαφορές αυτών. Ωστόσο, η επιλογή της συγκεκριμένης αρχιτεκτονικής του νευρωνικού δικτύου κρίθηκε από την τιμή της τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων, που ήταν η μικρότερη συγκριτικά με τις υπόλοιπες των άλλων μοντέλων, ενώ η διαφορά της μεταξύ του συνόλου εκπαίδευσης και αξιολόγησης είναι αρκετά μικρή.



Διάγραμμα 6.32 : Διάγραμμα εκτιμώμενων τιμών των παρατηρήσεων συναρτίζει των πραγματικών τους τιμών για νευρωνικό δίκτυο με ένα κρυφό στρώμα αποτελούμενο από 300 νευρώνες με εφαρμογή της συνάρτησης ενεργοποίησης Relu.



Διάγραμμα 6.33 : Τα σχετικά σφάλματα για τις εκτιμήσεις του νευρωνικού δικτύου με ένα κρυφό στρώμα 300 νευρώνων, συναρτίζει των πραγματικών τιμών των παρατηρήσεων στο διάστημα $[-5, 5]$.



Διάγραμμα 6.34 : Διάγραμμα συνάρτησης κόστους συναρτήσει του αριθμού των εποχών για νευρικό δίκτυο αποτελούμενο από ένα κρυφό στρώμα 300 νευρώνων.

Από τα παραπάνω διαγράμματα φαίνεται πως για τη συγκεκριμένη αρχιτεκτονική, η απόδοση του νευρωνικού δικτύου είναι σχετικά ικανοποιητική. Ωστόσο, από το Διάγραμμα 6.5.2. φαίνεται πως υπάρχουν τιμές για τις οποίες δεν μπορεί να δοθεί καλή εκτίμηση με μικρό σφάλμα. Για το λόγο αυτό, και αφού αυξάνοντας το πλήθος των νευρώνων του κρυφού στρώματος δεν εμφανίζεται σημαντική βελτίωση των αποτελεσμάτων, ενώ ελλοχεύει ο κίνδυνος εμφάνισης του φαινομένου υπερεκπαίδευσης αυξάνοντας την πολυπλοκότητα του δικτύου, δοκιμάζεται η κατασκευή ενός επιπλέον κρυφού στρώματος, όπως παρουσιάζεται παρακάτω, με σκοπό να διαπιστωθεί αν υπάρχει βελτίωση της απόκρισης του νευρωνικού δικτύου.

6.5.2 Κατασκευή Νευρωνικών Δικτύων με 2 Κρυφά Στρώματα

Στην ενότητα αυτή κατασκευάζονται νευρωνικά δίκτυα με δύο κρυφά στρώματα ακολουθώντας την ίδια τακτική που περιγράφηκε παραπάνω. Αναλυτικά, δοκιμάζονται πολλοί διαφορετικοί συνδυασμοί για τον αριθμό των νευρώνων του κάθε στρώματος ώστε να βρεθεί ο κατάλληλος για τον οποίο προκύπτουν τα βέλτιστα αποτελέσματα. Πιο συγκεκριμένα, η διερεύνηση για το συνδυασμό του πλήθους των νευρώνων ανά στρώμα στηρίζεται στο προηγούμενο μοτίβο, δηλαδή με το συνολικό αριθμό κόμβων του πρώτου κρυφού στρώματος να ισούται με το πλήθος των μεταβλητών του προβλήματος και με την διαδοχική προσθήκη νέων, ενώ ο αριθμός των αντίστοιχων νευρώνων του δεύτερου κρυφού στρώματος δεν ξεπερνά αυτόν του πρώτου. Για κάθε διαφορετική αρχιτεκτονική δικτύων που προκύπτει υπολογίζονται τα σχετικά σφάλματα των εκτιμήσεων, καθώς επίσης και η μέση τιμή και τυπική απόκλιση αυτών. Παράλληλα, για κάθε ένα νευρωνικό δίκτυο από αυτά κατασκευάζεται και το διάγραμμα των εκτιμώμενων τιμών συναρτήσει των πραγματικών τιμών των

παρατηρήσεων. Τα αποτελέσματα αυτά συμβάλουν στην εύρεση του μοντέλου με την καλύτερη προσαρμογή στα δεδομένα, αλλά και με τις μικρότερες τιμές σφαλμάτων. Επομένως, η επιλογή του βέλτιστου μοντέλου κρίνεται από τη μέση τιμή και τυπική απόκλιση των σχετικών σφαλμάτων, ώστε οι αποκλίσεις των εκτιμήσεων από τις παρατηρηθείσες τιμές να μην είναι μεγάλες. Επιπλέον, οι τιμές που λαμβάνουν οι ποσότητες αυτές δεν πρέπει να παρουσιάζουν μεγάλες διαφορές μεταξύ του συνόλου εκπαίδευσης και αξιολόγησης, ώστε το μοντέλο να μην επικεντρώνεται στην εκμάθηση μοτίβων που τυχόν να υπάρχουν στο πρώτο σύνολο και να μην μπορεί να προσαρμοστεί στο δεύτερο.

Από το σύνολο των νευρωνικών δικτύων που κατασκευάστηκαν, αυτά που ξεχώρισαν για την απόδοσή τους φαίνεται πως προκύπτουν με την εφαρμογή της υπερβολικής εφαπτομένης ως συνάρτηση ενεργοποίησης για το κάθε ένα από αυτά. Συγκεκριμένα, αυτά είναι τα ακόλουθα :

- (1200 , 100) – (28 , 20)
- (1200 , 100) – (42 , 20)
- (1200 , 100) – (56 , 30)

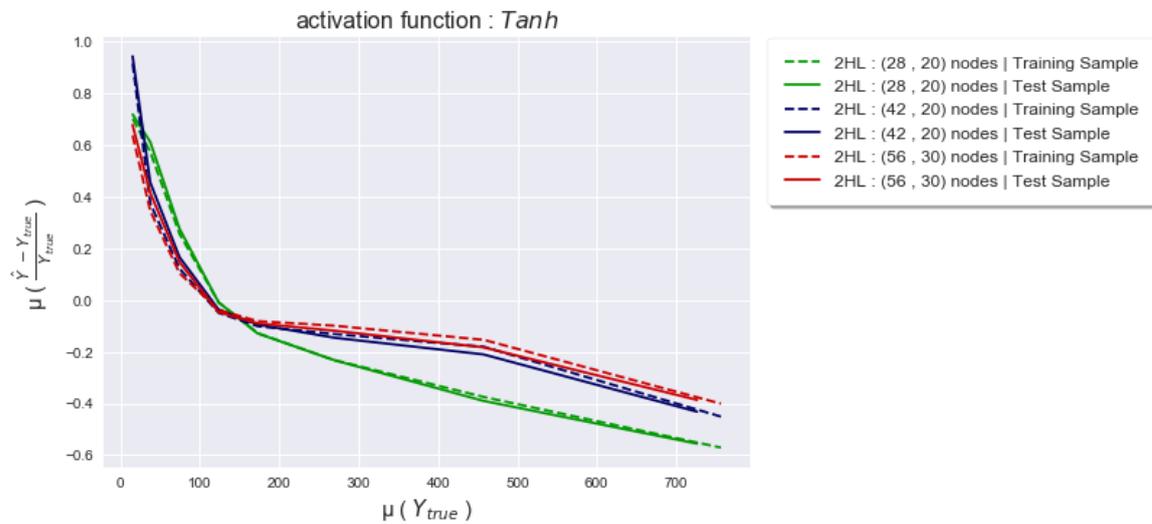
Όπου ο συμβολισμός $(a , \beta) - (\gamma , \delta)$ που περιγράφει τα μοντέλα αναφέρεται στον αριθμό των εποχών, a , το μέγεθος των υποσυνόλων δεδομένων εκπαίδευσης, β , καθώς και στην αρχιτεκτονική τους, δηλαδή με γ νευρώνες στο πρώτο κρυφό στρώμα και δ στο δεύτερο. Περισσότερες πληροφορίες για τα παραπάνω μοντέλα βρίσκονται στον Πίνακα 6.5, καθώς επίσης και στα Διαγράμματα 6.5.6 – 6.5.8.

<i>Model</i>	<i>Mean</i>		<i>Standard Deviation</i>		<i>MSE</i>	
	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>
(28 , 20)	0.35	0.39	1.35	1.37	2182.44	2666.55
(42 , 20)	0.32	0.37	1.16	1.28	1717.68	2301.38
(56 , 30)	0.25	0.30	1.11	1.28	1393.37	2052.01

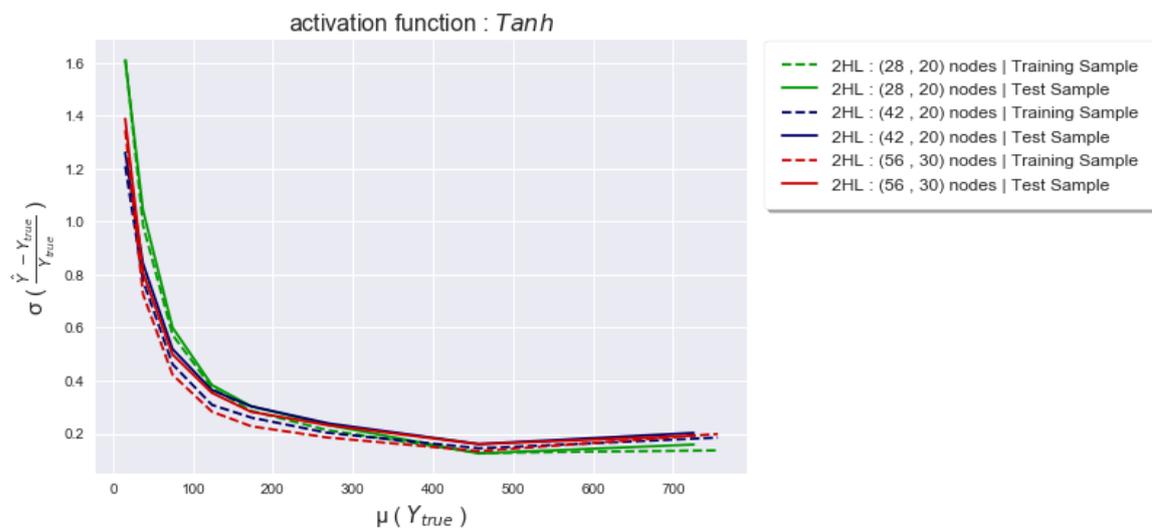
Πίνακας 6.5 : Τιμές για τη μέση τιμή και την τυπική απόκλιση των σχετικών σφαλμάτων του συνόλου εκπαίδευσης και αξιολόγησης, καθώς και το μέσο τετραγωνικό σφάλμα που σημειώνεται για αυτά.

Ο Πίνακας 6.5 περιέχει τις τιμές για τη μέση τιμή και την τυπική απόκλιση των σχετικών σφαλμάτων των συνόλων εκπαίδευσης και αξιολόγησης, όπως επίσης και την τιμή για το μέσο τετραγωνικό σφάλμα που σημειώνεται για τα τρία καλύτερα μοντέλα που ανιχνεύονται για νευρωνικά δίκτυα με δύο κρυφά στρώματα. Από αυτόν, παρατηρείται πως το βέλτιστο μοντέλο αποτελείται από 42 νευρώνες στο πρώτο κρυφό στρώμα και 20 στο δεύτερο καθώς επιτυγχάνεται ταυτόχρονα η ελάχιστη δυνατή τιμή για την τυπική απόκλιση του συνόλου αξιολόγησης σε συνδυασμό με τη μικρότερη απόκλιση από αυτή του συνόλου εκπαίδευσης. Το συμπέρασμα αυτό επιβεβαιώνεται και από τα ακόλουθα διαγράμματα, στα οποία αναπαρίσταται η μέση τιμή και τυπική απόκλιση για τα σχετικά σφάλματα των εκτιμήσεων των δειγμάτων εκπαίδευσης και

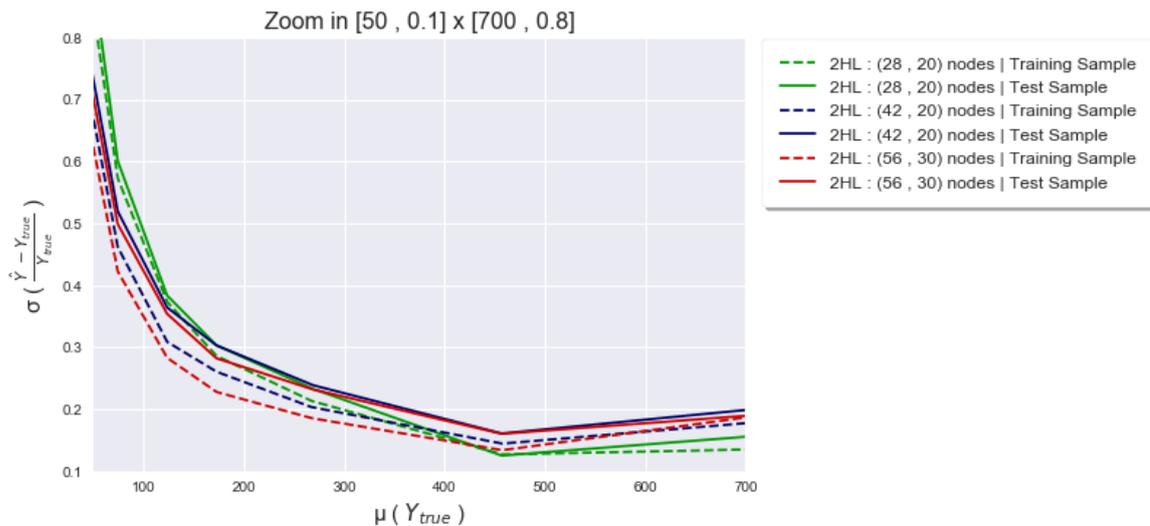
αξιολόγησης στο διάστημα $[-5, 5]$ συναρτήσει της μέσης τιμής των παρατηρήσεων για διάφορα διαστήματα του πεδίου τιμών τους, και για τα τρία αυτά μοντέλα.



Διάγραμμα 6.35 : Διάγραμμα μέσης τιμής των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$, συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα καλύτερα μοντέλα αποτελούμενα από δύο κρυφά στρώματα.



Διάγραμμα 6.36 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$, συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα καλύτερα μοντέλα αποτελούμενα από δύο κρυφά στρώματα.

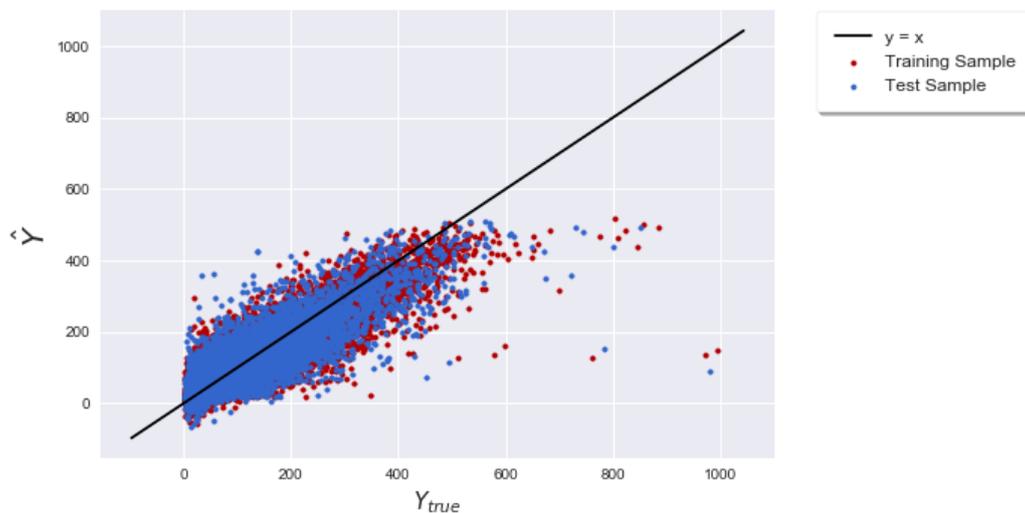


Διάγραμμα 6.37 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$, συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα καλύτερα μοντέλα αποτελούμενα από δύο κρυφά στρώματα, σε μεγέθυνση στο $[50, 0.1] \times [700, 0.8]$.

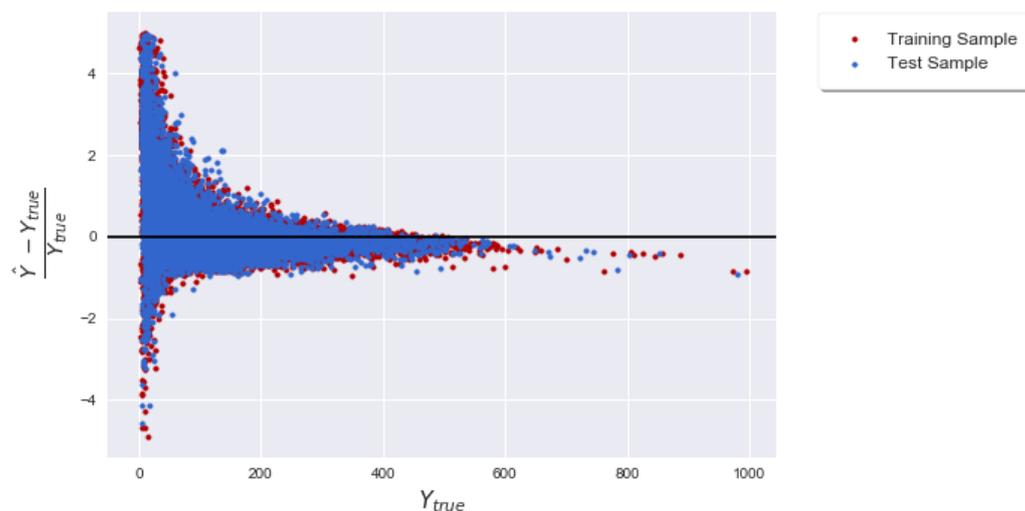
Από τα παραπάνω διαγράμματα φαίνεται πως δεν υπάρχουν ιδιαίτερα σημαντικές αποκλίσεις στις τιμές των συνόλων εκπαίδευσης και αξιολόγησης. Πιο αναλυτικά, από το Διάγραμμα 6.37, το οποίο αποτελεί μεγέθυνση του Διαγράμματος 6.36, παρατηρείται πως για το νευρωνικό δίκτυο που αποτελείται από 56 νευρώνες στο πρώτο κρυφό στρώμα και 30 στο δεύτερο αντίστοιχα, η απόκλιση μεταξύ των γραμμών που αντιστοιχούν στα σύνολα εκπαίδευσης και αξιολόγησης είναι η ελάχιστη. Ωστόσο, από τα συμπεράσματα που προέκυψαν από τον Πίνακα 6.5, καθώς και από τα διαγράμματα των εκτιμώμενων τιμών συναρτήσει των πραγματικών παρατηρούμενων τιμών για τα μοντέλα $(1200, 100) - (42, 20)$ και $(1200, 100) - (56, 30)$ τα οποία έχουν όμοια συμπεριφορά δίχως να σημειώνονται διαφορές μεταξύ αυτών, προτιμάται το πρώτο από αυτά. Η επιλογή αυτή οφείλεται στο ότι για το μοντέλο που αποτελείται από περισσότερους νευρώνες στο πρώτο κρυφό στρώμα δεν παρατηρείται κάποια ιδιαίτερη βελτίωση των αποτελεσμάτων σε σχέση με αυτό με τους λιγότερους νευρώνες, ενώ επιπλέον αυξάνει την πολυπλοκότητα του μοντέλου χωρίς να το καθιστά πιο αποδοτικό.

Αποτελεσματικά, από τα διαφορετικά μοντέλα που προέκυψαν, συμπεραίνεται ότι το καλύτερο είναι αυτό που αποτελείται από 42 νευρώνες στο πρώτο κρυφό στρώμα και 20 αντίστοιχα στο δεύτερο, για εφαρμογή της υπερβολικής επαπτομένης ως συνάρτηση ενεργοποίησης, καθώς παρουσιάζεται η μικρότερη απόκλιση τόσο για τη μέση τιμή, όσο και για την τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων μεταξύ του συνόλου εκπαίδευσης και αξιολόγησης. Το Διάγραμμα 6.38 αναπαριστά τις εκτιμώμενες τιμές συναρτήσει των πραγματικών τιμών των παρατηρήσεων και φαίνεται πως είναι αρκετά ικανοποιητικό αφού και για τα δύο σύνολα έχει την ίδια συμπεριφορά. Ωστόσο, υπάρχουν και κάποιες τιμές για τις οποίες το μοντέλο δεν μπορεί να δώσει ικανοποιητικές εκτιμήσεις, οι οποίες τοποθετούνται απομακρυσμένες από την ευθεία $y=x$, δηλώνοντας ότι για αυτές σημειώνεται μεγαλύτερο σφάλμα

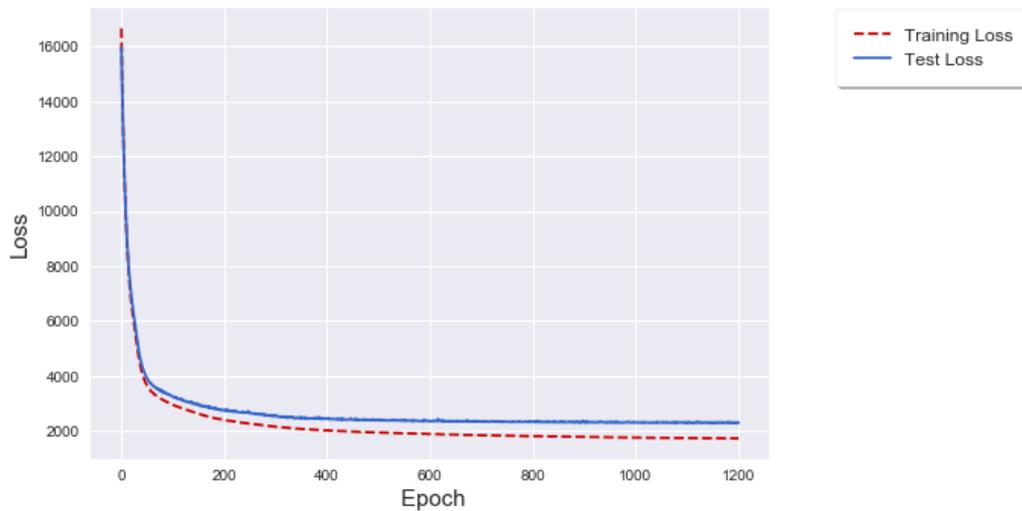
εκτίμησης. Το γεγονός αυτό επιβεβαιώνει το Διάγραμμα 6.39 στο οποίο απεικονίζονται τα σχετικά σφάλματα των εκτιμήσεων που προκύπτουν συναρτήσει των πραγματικών τιμών των παρατηρήσεων στο διάστημα $[-5, 5]$, ώστε να υπάρχει μία πιο ακριβής μελέτη για την απόδοση του δικτύου, έχοντας αφαιρέσει ακραίες τιμές που τυχόν υπονομεύουν την επίδοσή του. Συγκεκριμένα, διαπιστώνεται ότι μεγαλύτερο σχετικό σφάλμα εκτίμησης αντιστοιχεί σε ορισμένες μικρές παρατηρηθείσες τιμές, που μπορεί να οφείλεται σε υψηλές συσχετίσεις της μεταβλητής απόκρισης με τις επεξηγηματικές μεταβλητές, όταν αυτή λαμβάνει μεγαλύτερες τιμές.



Διάγραμμα 6.38 : Διάγραμμα εκτιμώμενων τιμών των παρατηρήσεων συναρτήσει των πραγματικών τους τιμών για νευρωνικό δίκτυο με δύο κρυφά στρώματα αποτελούμενα από 42 και 20 νευρώνες αντίστοιχα, με εφαρμογή της υπερβολικής εφαπτομένης ως συνάρτησης ενεργοποίησης μεταξύ αυτών.



Διάγραμμα 6.39 : Τα σχετικά σφάλματα για τις εκτιμήσεις του νευρωνικού δικτύου με δύο κρυφά στρώματα από 42 και 20 νευρώνες αντίστοιχα, με εφαρμογή της υπερβολικής εφαπτομένης ως συνάρτησης ενεργοποίησης μεταξύ αυτών, συναρτήσει των πραγματικών τιμών των παρατηρήσεων στο διάστημα $[-5, 5]$.



Διάγραμμα 6.40 : Διάγραμμα συνάρτησης κόστους συναρτήσει του αριθμού των εποχών για νευρικό δίκτυο αποτελούμενο από δύο κρυφά στρώματα από 42 και 20 νευρώνες αντίστοιχα, με εφαρμογή της υπερβολικής εφαπτομένης ως συνάρτησης ενεργοποίησης μεταξύ αυτών.

Από το τελευταίο διάγραμμα φαίνεται πως το σφάλμα εκτίμησης των παρατηρήσεων του συνόλου αξιολόγησης έχει μικρή απόκλιση από το αντίστοιχο του συνόλου εκπαίδευσης. Επιπλέον, οι δύο καμπύλες συμβαδίζουν με τη μεταξύ τους απόσταση να παραμένει σχεδόν σταθερή όσο αυξάνεται ο αριθμός των εποχών, που σημαίνει ότι το μοντέλο έχει την ίδια συμπεριφορά για τα δύο σύνολα και δεν φαίνεται να προσαρμόζεται στο δείγμα εκπαίδευσης. Αποτελεσματικά, δεν υπάρχει εμφάνιση του φαινομένου υπερεκπαίδευσης, όπως συμβαίνει σε άλλα δίκτυα όταν αυξάνονται οι νευρώνες των κρυφών στρωμάτων του.

6.5.3 Κατασκευή Νευρωνικών Δικτύων με 3 Κρυφά Στρώματα

Όμοια με την κατασκευή νευρωνικών δικτύων που αποτελούνται από ένα και δύο κρυφά στρώματα, έτσι και για τη δημιουργία δικτύων με τρία κρυφά στρώματα, επαναλαμβάνεται η ίδια στρατηγική. Συγκεκριμένα, κατασκευάζονται νευρωνικά δίκτυα με το πλήθος των νευρώνων του πρώτου στρώματος να είναι πολλαπλάσιο του αριθμού των μεταβλητών του προβλήματος, ενώ ο συνολικός αριθμός των νευρώνων κάθε επόμενου στρώματος είναι μικρότερος από αυτόν του προηγούμενου. Επιπλέον, όπως και στις παραπάνω διαδικασίες, έτσι και τώρα για την κάθε αρχιτεκτονική νευρωνικού δικτύου που προκύπτει, η διαδικασία επαναλαμβάνεται με την εφαρμογή κάθε μίας από τις συναρτήσεις ενεργοποίησης, όπως αναφέρονται στη Διερεύνηση των Παραμέτρων στην αρχή του κεφαλαίου, μεταξύ των κρυφών στρωμάτων.

Από το σύνολο των νευρωνικών δικτύων που κατασκευάζονται, φαίνεται πως με τη συγκεκριμένη αρχιτεκτονική δικτύων λαμβάνονται καλύτερα αποτελέσματα και τα μοντέλα που προκύπτουν είναι πιο αποδοτικά σε σχέση με αυτά που αποτελούνται από λιγότερα κρυφά στρώματα. Αυτό γίνεται αντιληπτό από τις τιμές που λαμβάνει τόσο η μέση τιμή, όσο και η τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων. Συγκεκριμένα για τις τιμές που λαμβάνει η τυπική απόκλιση για τα σχετικά σφάλματα, είναι αρκετά πιο κοντά στη μονάδα συγκριτικά με αυτές που προέκυψαν για τα νευρωνικά δίκτυα με ένα και δύο κρυφά στρώματα, που σημαίνει ότι υπάρχει καλή προσαρμογή στα δεδομένα και προκύπτουν εκτιμήσεις με μικρότερο σφάλμα.

Από το σύνολο των νευρωνικών δικτύων που κατασκευάστηκαν, αυτά που ξεχώρισαν για την απόδοσή τους είναι τα ακόλουθα :

- (1200 , 100) – (28 , 20 , 14), με την εφαρμογή της υπερβολικής εφαπτομένης ως συνάρτηση ενεργοποίησης
- (1200 , 100) – (28 , 25 , 20), με την εφαρμογή της *Relu* ως συνάρτηση ενεργοποίησης
- (1200 , 100) – (50 , 25 , 10), με την εφαρμογή της *Relu* ως συνάρτηση ενεργοποίησης
- (1500 , 100) – (60 , 40 , 10), με την εφαρμογή της *Relu* ως συνάρτηση ενεργοποίησης
- (1200 , 100) – (80 , 30 , 10), με την εφαρμογή της *Relu* ως συνάρτηση ενεργοποίησης

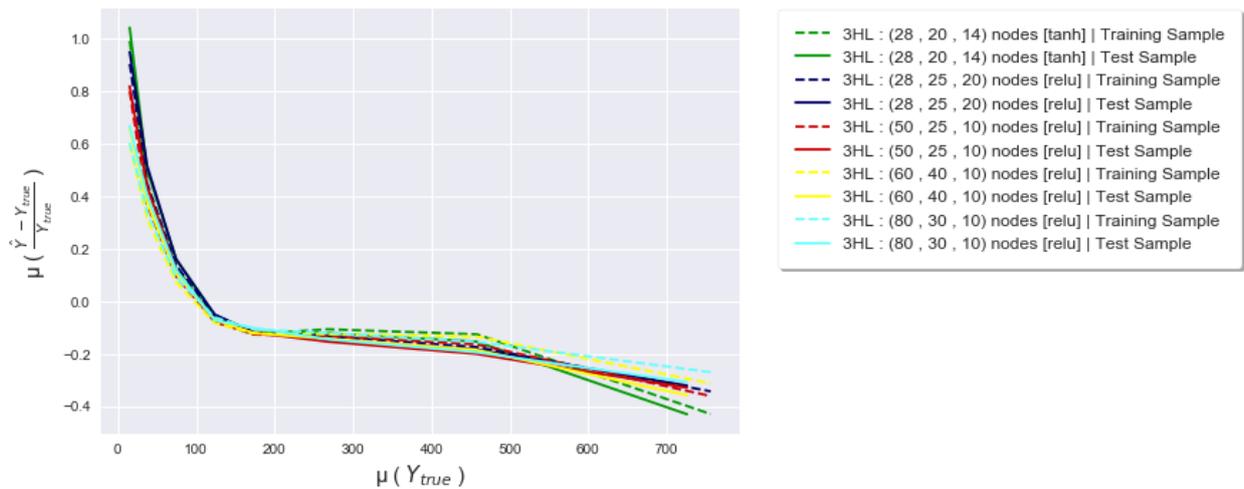
Όπου ο συμβολισμός $(\alpha, \beta) - (\gamma, \delta, \epsilon)$ που περιγράφει τα μοντέλα αναφέρεται στον αριθμό των εποχών, α , το μέγεθος των υποσυνόλων δεδομένων εκπαίδευσης, β , καθώς και στην αρχιτεκτονική τους, δηλαδή με γ νευρώνες στο πρώτο κρυφό στρώμα, δ στο δεύτερο και ϵ στο τρίτο. Περισσότερες πληροφορίες για τα παραπάνω μοντέλα βρίσκονται στον Πίνακα 6.6, καθώς επίσης και στα Διαγράμματα 6.5.11 – 6.5.14.

<i>Model</i>	<i>Mean</i>		<i>Standard Deviation</i>		<i>MSE</i>	
	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>
<i>(28, 20, 14) – tanh</i>	0.34	0.39	1.18	1.26	1592.82	2183.12
<i>(28, 25, 20) – relu</i>	0.33	0.38	1.22	1.27	1792.82	2228.14
<i>(50, 25, 10) – relu</i>	0.26	0.30	1.02	1.14	1624.43	2137.89
<i>(60, 40, 10) – relu</i>	0.20	0.26	1.00	1.13	1365.07	2079.33
<i>(80, 30, 10) – relu</i>	0.21	0.26	1.00	1.11	1392.25	2067.86

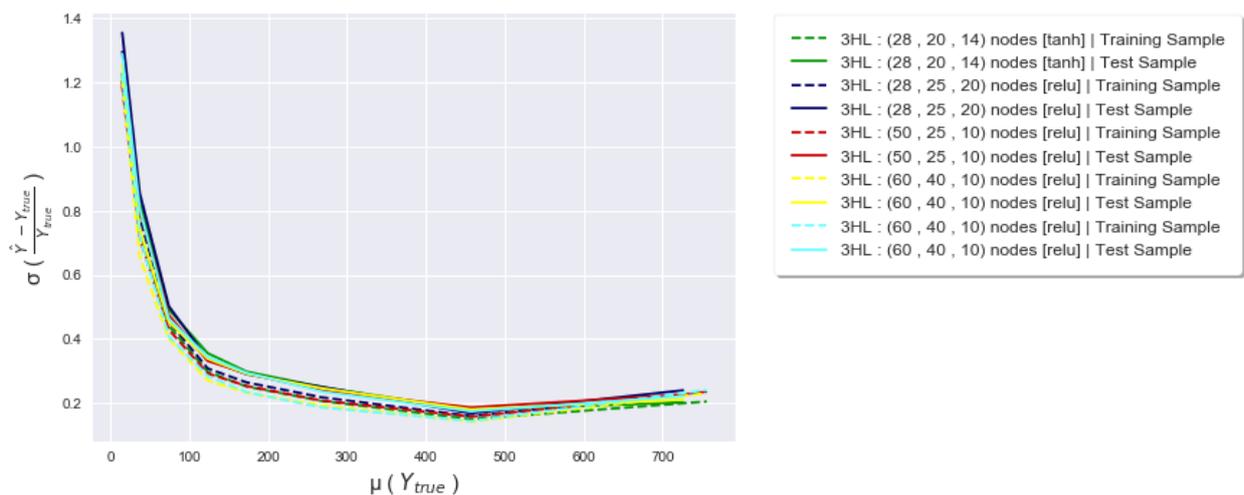
Πίνακας 6.6 : Τιμές για τη μέση τιμή και την τυπική απόκλιση των σχετικών σφαλμάτων του συνόλου εκπαίδευσης και αξιολόγησης, καθώς και το μέσο τετραγωνικό σφάλμα που σημειώνεται για αυτά.

Ο Πίνακας 6.6 περιέχει τις τιμές για τη μέση τιμή και την τυπική απόκλιση των σχετικών σφαλμάτων των συνόλων εκπαίδευσης και αξιολόγησης, όπως επίσης και την

τιμή για το μέσο τετραγωνικό σφάλμα που σημειώνεται για τα καλύτερα μοντέλα που ανιχνεύονται για νευρωνικά δίκτυα με τρία κρυφά στρώματα. Από αυτόν, συμπεραίνεται πως με βάση το κριτήριο της ελάχιστης διαφοράς μεταξύ των τιμών για τη μέση τιμή και τυπική απόκλιση των συνόλων εκπαίδευσης και αξιολόγησης, το βέλτιστο μοντέλο είναι το νευρωνικό δίκτυο που αποτελείται από 28 νευρώνες στο πρώτο κρυφό στρώμα, 25 στο δεύτερο και 20 στο τρίτο αντίστοιχα.

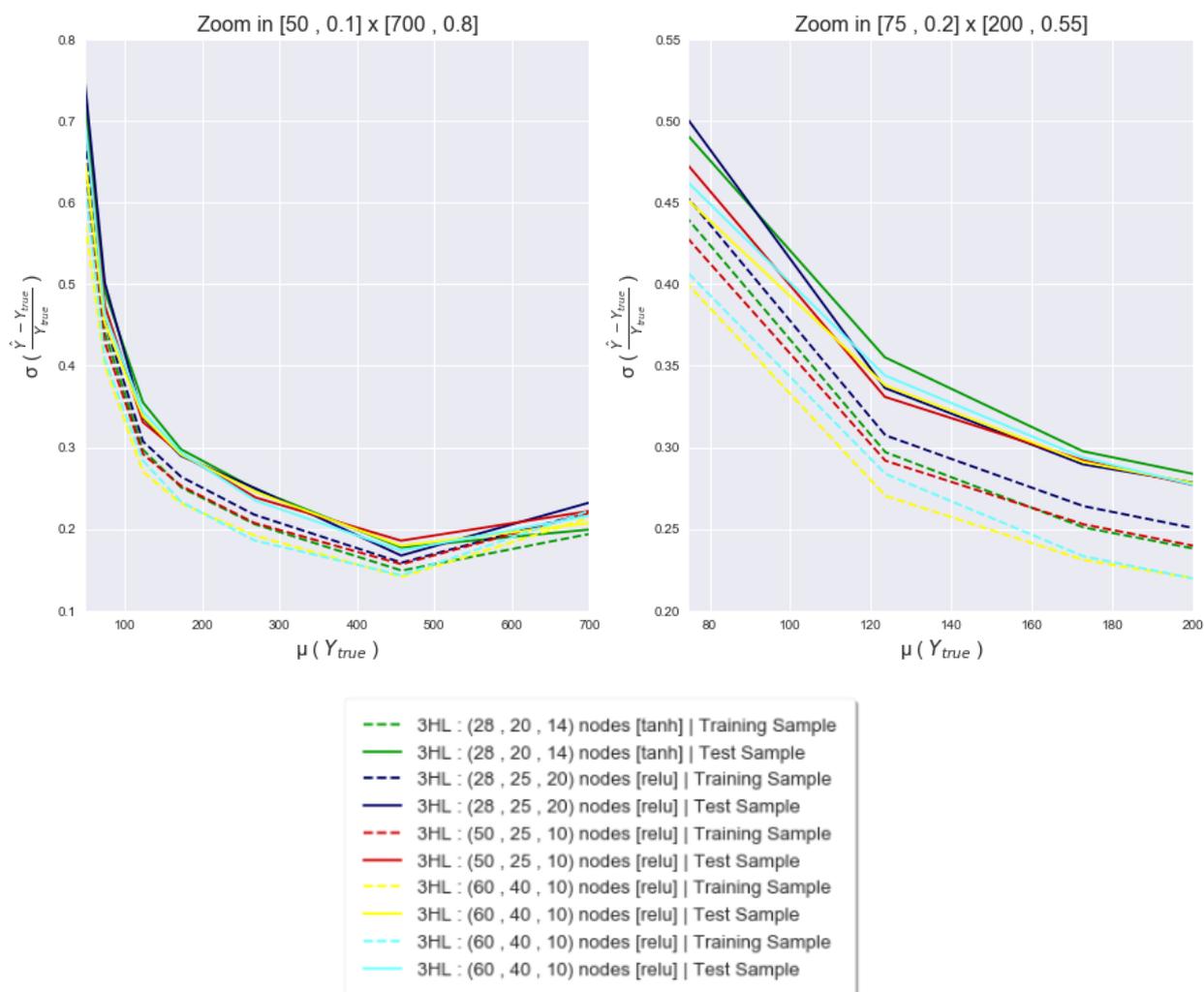


Διάγραμμα 6.43 : Διάγραμμα μέσης τιμής των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$, συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα καλύτερα μοντέλα αποτελούμενα από τρία κρυφά στρώματα.



Διάγραμμα 6.44 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$, συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα καλύτερα μοντέλα αποτελούμενα από τρία κρυφά στρώματα.

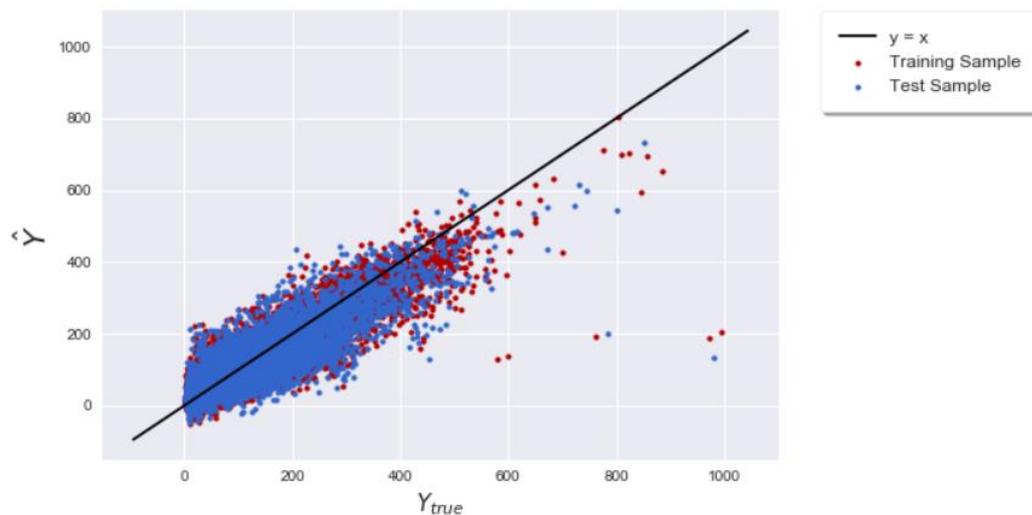
Από τα παραπάνω διαγράμματα φαίνεται πως δεν υπάρχουν ιδιαίτερα σημαντικές αποκλίσεις μεταξύ των τιμών των συνόλων εκπαίδευσης και αξιολόγησης. Ωστόσο, για μεγαλύτερη ευκρίνεια, δημιουργούνται τα Διαγράμματα 6.45 και 6.46, τα οποία αποτελούν μεγεθύνσεις του Διαγράμματος 6.44, ώστε να γίνει μία πιο ακριβής μελέτη ως προς την επιλογή της βέλτιστης αρχιτεκτονικής.



Διαγράμματα 6.45 – 6.46 : Διαγράμματα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$, συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα καλύτερα μοντέλα αποτελούμενα από τρία κρυφά στρώματα, σε μεγέθυνση στο $[50, 0.1] \times [700, 0.8]$ και $[75, 0.2] \times [200, 0.55]$ αντίστοιχα.

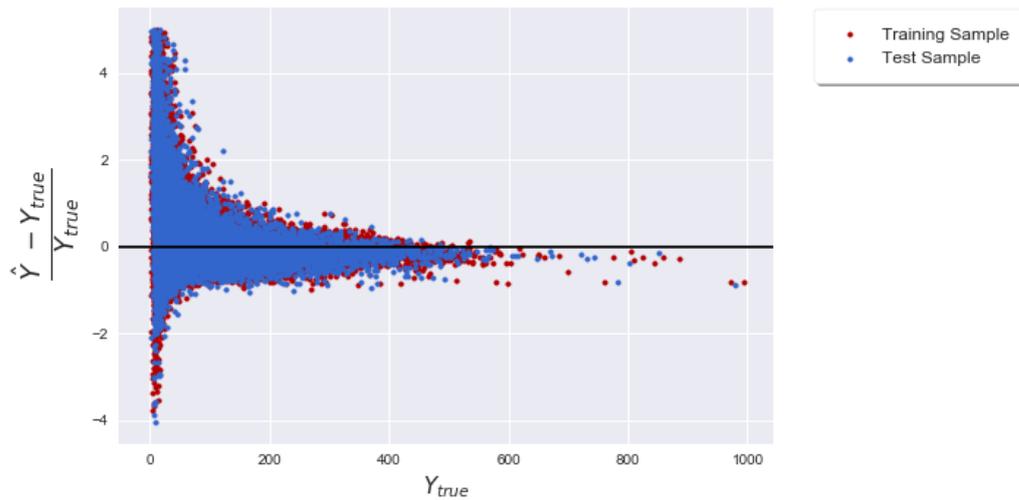
Πιο αναλυτικά, από το Διάγραμμα 6.46, το οποίο αποτελεί μεγέθυνση του Διαγράμματος 6.44, διαπιστώνεται ότι η μικρότερη απόκλιση μεταξύ των τιμών για τα σύνολα εκπαίδευσης και αξιολόγησης σημειώνεται για το νευρωνικό δίκτυο που αποτελείται από 28 νευρώνες στο πρώτο κρυφό στρώμα, 25 στο δεύτερο και 20 στο τρίτο αντίστοιχα.

Αποτελεσματικά, από τα διαφορετικά μοντέλα που προέκυψαν, συμπεραίνεται ότι το καλύτερο είναι αυτό που αποτελείται από 28 νευρώνες στο πρώτο κρυφό στρώμα, 25 στο δεύτερο και 20 στο τρίτο αντίστοιχα, για εφαρμογή της *Relu* ως συνάρτηση ενεργοποίησης, καθώς παρουσιάζεται η μικρότερη απόκλιση τόσο για τη μέση τιμή, όσο και για την τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων μεταξύ του συνόλου εκπαίδευσης και αξιολόγησης. Το Διάγραμμα 6.47 περιγράφει τις εκτιμώμενες τιμές συναρτήσει των πραγματικών τιμών των παρατηρήσεων και φαίνεται πως είναι αρκετά ικανοποιητικό αφού και για τα δύο σύνολα έχει παρόμοια συμπεριφορά. Ωστόσο, υπάρχουν και κάποιες τιμές για τις οποίες το μοντέλο δεν μπορεί να δώσει ικανοποιητικές εκτιμήσεις, οι οποίες τοποθετούνται απομακρυσμένες από την ευθεία $y=x$, δηλώνοντας ότι για αυτές σημειώνεται μεγαλύτερο σφάλμα εκτίμησης. Το γεγονός αυτό επιβεβαιώνει το Διάγραμμα 6.48 στο οποίο απεικονίζονται τα σχετικά σφάλματα των εκτιμήσεων που προκύπτουν συναρτήσει των πραγματικών τιμών των παρατηρήσεων στο διάστημα $[-5, 5]$, ώστε να υπάρχει μία πιο ακριβής μελέτη για την απόδοση του δικτύου, έχοντας αφαιρέσει τις ακραίες τιμές που τυχόν υπονομεύουν την επίδοσή του. Συγκεκριμένα, διαπιστώνεται ότι μεγαλύτερο σχετικό σφάλμα εκτίμησης αντιστοιχεί σε ορισμένες μικρές παρατηρηθείσες τιμές, που μπορεί να οφείλεται σε υψηλές συσχετίσεις της μεταβλητής απόκρισης με τις επεξηγηματικές μεταβλητές, όταν αυτή λαμβάνει μεγαλύτερες τιμές.



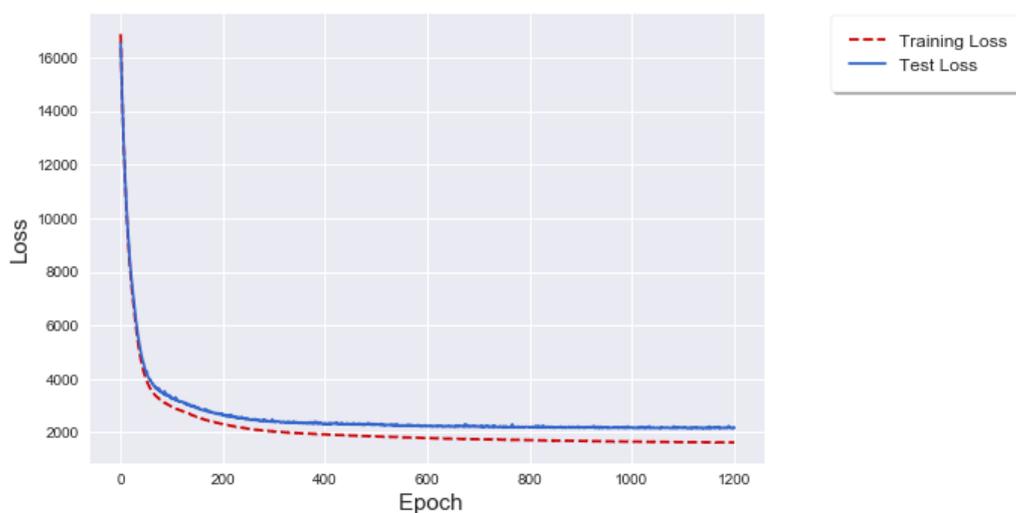
Διάγραμμα 6.47 : Διάγραμμα εκτιμώμενων τιμών των παρατηρήσεων συναρτήσει των πραγματικών τους τιμών για νευρωνικό δίκτυο με τρία κρυφά στρώματα αποτελούμενα από 28, 25 και 20 νευρώνες αντίστοιχα, με εφαρμογή της *Relu* ως συνάρτησης ενεργοποίησης μεταξύ αυτών.

Η καλή απόδοση του νευρωνικού δικτύου επιβεβαιώνεται και από το Διάγραμμα 6.48 στο οποίο απεικονίζονται τα σχετικά σφάλματα των εκτιμήσεων που προκύπτουν συναρτήσει των πραγματικών τιμών των παρατηρήσεων στο διάστημα $[-5, 5]$, καθώς οι περισσότερες τιμές συγκεντρώνονται κοντά στο θετικό ημιάξονα $x'x$ που σημαίνει ότι τα σχετικά σφάλματα λαμβάνουν μικρές τιμές.



Διάγραμμα 6.48 : Τα σχετικά σφάλματα για τις εκτιμήσεις του νευρωνικού δικτύου με τρία κρυφά στρώματα από 28, 25 και 20 νευρώνες αντίστοιχα, με εφαρμογή της Relu ως συνάρτησης ενεργοποίησης μεταξύ αυτών, συναρτήσεως των πραγματικών τιμών των παρατηρήσεων στο διάστημα $[-5, 5]$.

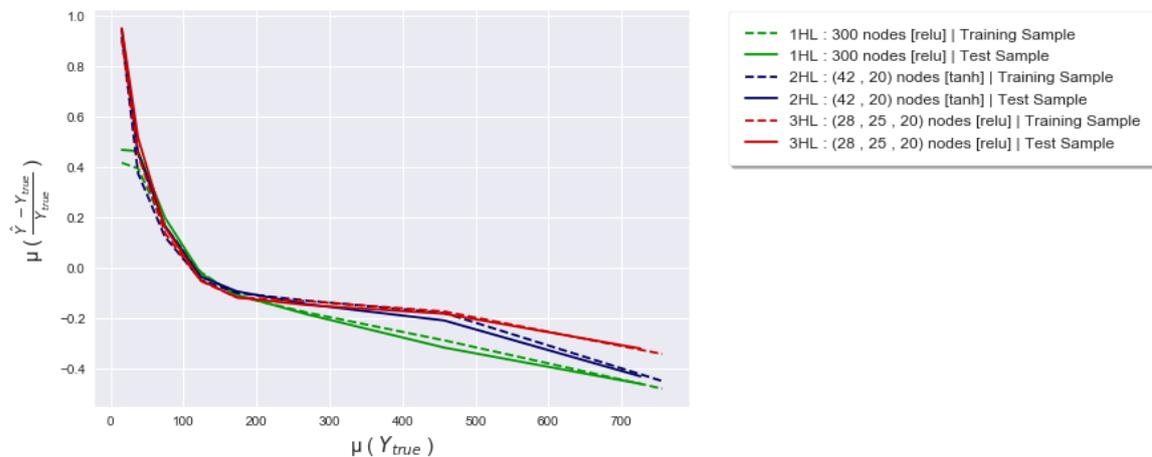
Ωστόσο, υπάρχουν και εκτιμήσεις στις οποίες σημειώνεται ένα μεγαλύτερο σχετικό σφάλμα και τα αντίστοιχα σημεία είναι πιο απομακρυσμένα από τον ημιάξονα. Αυτό το σχετικό σφάλμα εκτίμησης αντιστοιχεί σε ορισμένες μικρές παρατηρηθείσες τιμές, που μπορεί να οφείλεται σε υψηλές συσχετίσεις της μεταβλητής απόκρισης με τις επεξηγηματικές μεταβλητές, όταν αυτή λαμβάνει μεγαλύτερες τιμές. Το Διάγραμμα 6.49 περιγράφει το σφάλμα εκτίμησης των παρατηρήσεων για το σύνολο εκπαίδευσης και αξιολόγησης. Από αυτό φαίνεται πως οι δύο καμπύλες συμβαδίζουν και η μεταξύ τους απόκλιση διατηρείται σταθερή για το συνολικό αριθμό των εποχών που πραγματοποιείται η εκπαίδευση του δικτύου. Επιπλέον, καθώς η καμπύλη της συνάρτησης κόστους που αντιστοιχεί στο σύνολο εκπαίδευσης δεν μειώνεται απότομα από κάποιο σημείο και έπειτα και η αντίστοιχη του συνόλου αξιολόγησης δεν αυξάνεται, δεν υπάρχει ένδειξη εμφάνισης υπερεκπαίδευσης του μοντέλου.



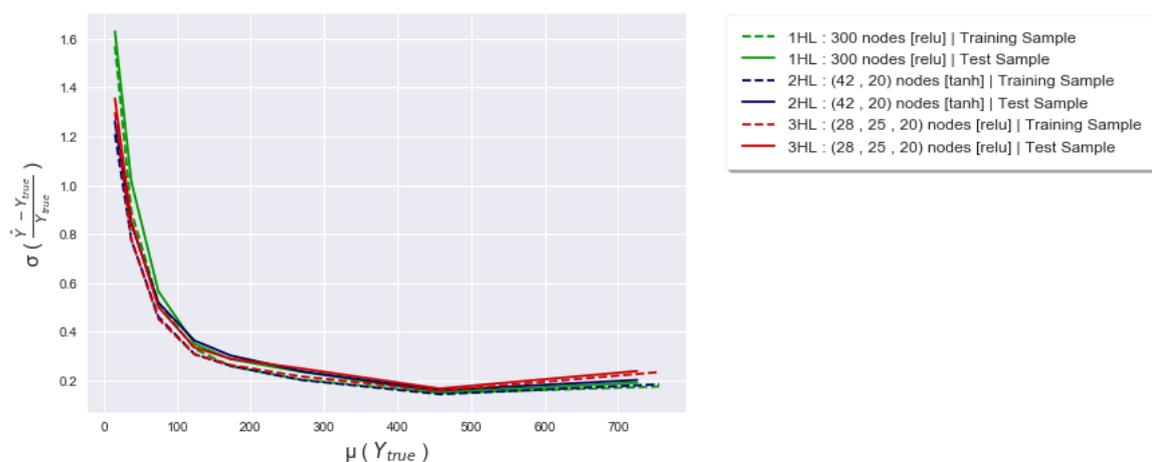
Διάγραμμα 6.49 : Διάγραμμα συνάρτησης κόστους συναρτήσεως του αριθμού των εποχών για νευρικό δίκτυο αποτελούμενο από τρία κρυφά στρώματα από 28, 25 και 20 νευρώνες αντίστοιχα, με εφαρμογή της Relu ως συνάρτησης ενεργοποίησης μεταξύ αυτών.

6.5.4 Σύγκριση Νευρωνικών Δικτύων για το Διαφορετικό Αριθμό των Κρυφών Στρωμάτων

Σύμφωνα με τα παραπάνω, έχει προκύψει ένα βέλτιστο μοντέλο για νευρωνικά δίκτυα με ένα, δύο και τρία κρυφά στρώματα αντίστοιχα. Τα μοντέλα αυτά συγκρίνονται με τη σειρά τους μεταξύ τους, με σκοπό την ανίχνευση του καλύτερου μοντέλου που προκύπτει με χρήση νευρωνικών δικτύων. Για το λόγο αυτό, δημιουργούνται τα Διαγράμματα 6.50 – 6.54, τα οποία βοηθούν στην εύρεση του συγκεκριμένου μοντέλου.

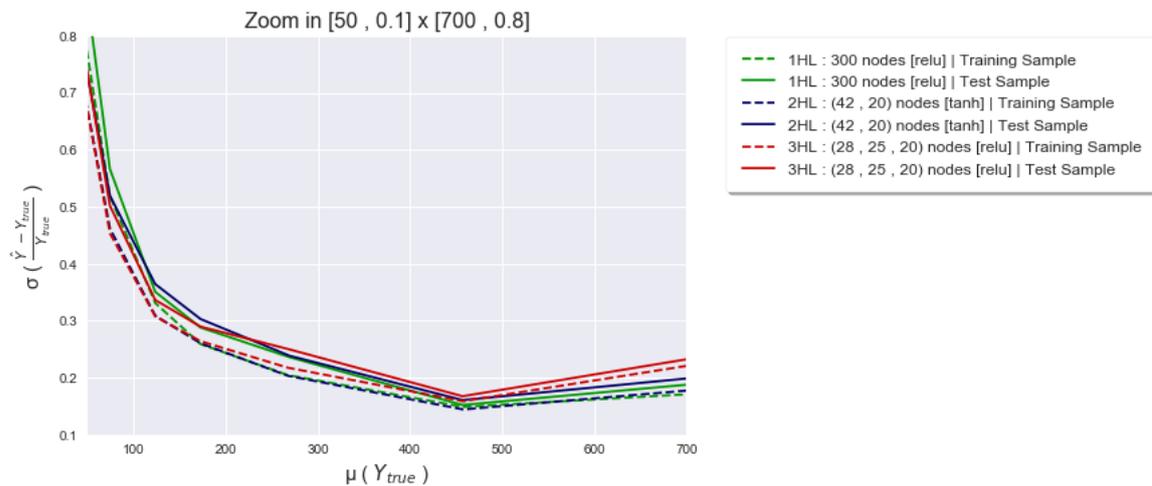


Διάγραμμα 6.50 : Διάγραμμα μέσης τιμής των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα [-5, 5], συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα καλύτερα μοντέλα που προκύπτουν με χρήση νευρωνικών δικτύων.



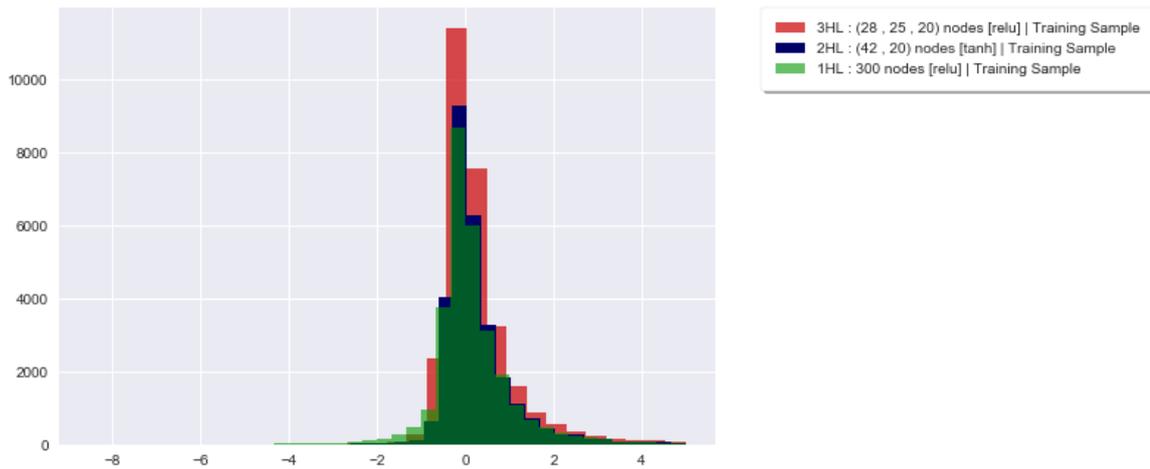
Διάγραμμα 6.51 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα [-5, 5], συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα καλύτερα μοντέλα που προκύπτουν με χρήση νευρωνικών δικτύων.

Για την καλύτερη μελέτη του τελευταίου διαγράμματος, δημιουργείται το Διάγραμμα 6.52, το οποίο αποτελεί μεγέθυνση του πρώτου, ώστε να διαπιστωθεί ποιο είναι το βέλτιστο μοντέλο που προκύπτει από κατασκευή νευρωνικών δικτύων, με βάση το κριτήριο της ελάχιστης απόκλισης των τιμών για τη μέση τιμή και τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων μεταξύ των συνόλων εκπαίδευσης και αξιολόγησης.

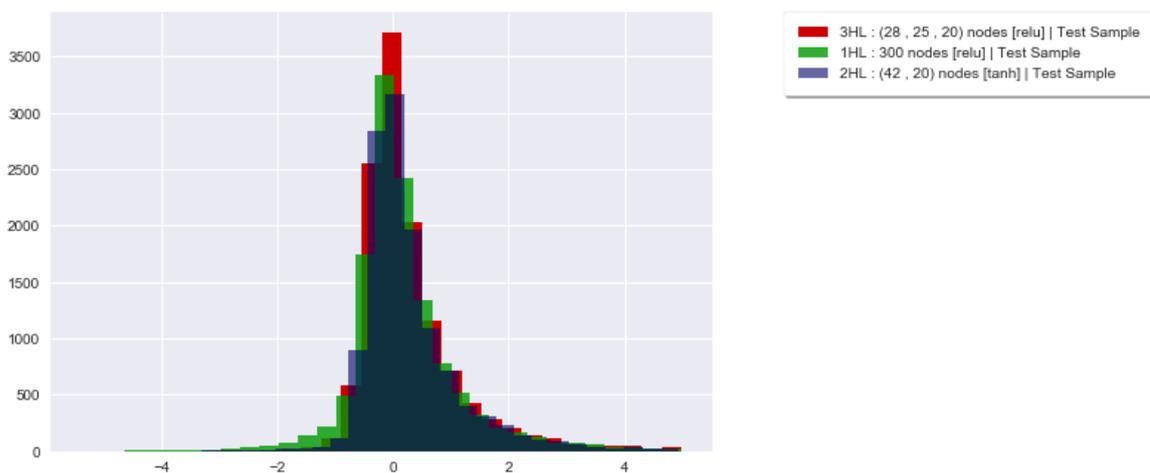


Διάγραμμα 6.52 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα [-5, 5], συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα καλύτερα μοντέλα που προκύπτουν με χρήση νευρωνικών δικτύων, σε μεγέθυνση στο [50, 0.1]x[700, 0.8].

Από το Διάγραμμα 6.52 φαίνεται πως το νευρωνικό δίκτυο που αποτελείται από τρία κρυφά στρώματα είναι αυτό για το οποίο σημειώνεται η μικρότερη απόκλιση μεταξύ των τιμών για τη μέση τιμή και τυπική απόκλιση για το σύνολο εκπαίδευσης και αξιολόγησης. Αποτελεσματικά, φαίνεται να είναι το βέλτιστο μοντέλο που προκύπτει από κατασκευή νευρωνικού δικτύου. Το συμπέρασμα αυτό ενισχύουν και τα ακόλουθα διαγράμματα, τα οποία αποτελούν τα ιστογράμματα των σχετικών σφαλμάτων των εκτιμήσεων στο διάστημα [-5, 5] για τα παραπάνω μοντέλα για το σύνολο εκπαίδευσης και αξιολόγησης αντίστοιχα.



Διάγραμμα 6.53 : Ιστόγραμμα σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$ για τα καλύτερα μοντέλα που προκύπτουν από κατασκευή νευρωνικών δικτύων για το σύνολο εκπαίδευσης.

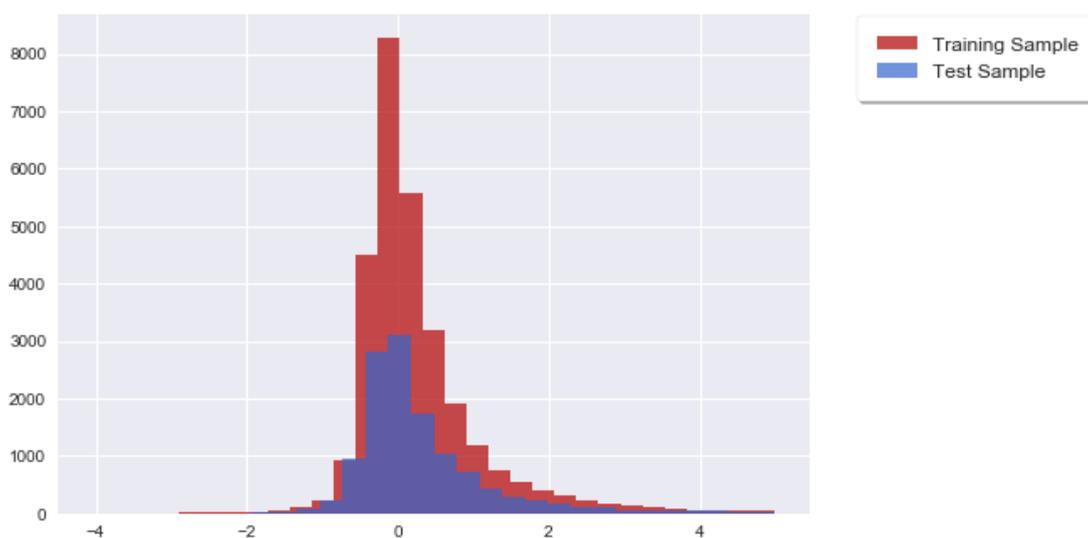


Διάγραμμα 6.54 : Ιστόγραμμα σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$ για τα καλύτερα μοντέλα που προκύπτουν από κατασκευή νευρωνικών δικτύων για το σύνολο αξιολόγησης.

Από τα δύο τελευταία διαγράμματα φαίνεται πως για το νευρωνικό δίκτυο που αποτελείται από τρία κρυφά στρώματα, το μεγαλύτερο ποσοστό των σχετικών σφαλμάτων των εκτιμήσεων συγκεντρώνεται κοντά στο μηδέν, σε σχέση με τα άλλα δύο μοντέλα. Αυτό σημαίνει πως για αυτό, οι περισσότερες εκτιμήσεις που προκύπτουν σημειώνουν μικρό σφάλμα, δηλαδή δεν υπάρχει μεγάλη απόκλιση από την πραγματική τιμή των παρατηρήσεων.

Το διάγραμμα που περιγράφει τις εκτιμώμενες τιμές των παρατηρήσεων συναρτήσει των πραγματικών τους τιμών για το συγκεκριμένο νευρωνικό δίκτυο αποτελεί το Διάγραμμα 6.47 (βλ. Κεφάλαιο 6.5.3), ενώ η παρουσίαση των σχετικών σφαλμάτων των εκτιμήσεων στο διάστημα $[-5, 5]$ συναρτήσει των παρατηρηθεισών τιμών αποτελεί

το Διάγραμμα 6.51 (βλ. Κεφάλαιο 6.5.3). Ωστόσο, μία πιο αναλυτική μελέτη για τη συμπεριφορά του μοντέλου είναι επιτακτική για τη διαπίστωση της καλής του απόδοσης. Το Διάγραμμα 6.55 που ακολουθεί αποτελεί το ιστόγραμμα των σχετικών σφαλμάτων για τις εκτιμήσεις των παρατηρήσεων στο διάστημα $[-5, 5]$ – ώστε να υπάρχει μεγαλύτερη ακρίβεια έχοντας αφαιρέσει ορισμένες ακραίες τιμές που μπορεί να είναι παραπλανητικές για την καλή λειτουργία του μοντέλου.



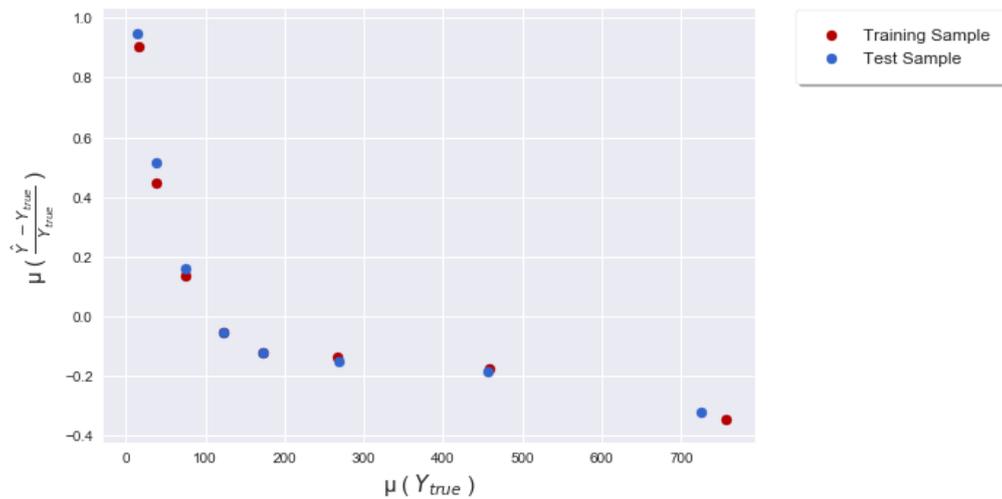
Διάγραμμα 6.55 : Ιστόγραμμα συχνοτήτων για τα σχετικά σφάλματα των εκτιμήσεων του νευρωνικού δικτύου, λαμβάνουν τιμές στο διάστημα $[-5, 5]$.

Από το Διάγραμμα 6.55, στο οποίο αναπαρίστανται τα ιστογράμματα συχνοτήτων των αποκομμένων σχετικών σφαλμάτων στο διάστημα $[-5, 5]$, φαίνεται πως τα σύνολα εκπαίδευσης και αξιολόγησης ακολουθούν την ίδια κατανομή, καθώς έχουν την ίδια μορφή, αλλά και σχεδόν ίδια μέση τιμή και τυπική απόκλιση όπως παρατηρείται από τον Πίνακα 6.6. Επιπλέον, υπολογίζεται ότι περίπου το 86% των εκτιμήσεων του συνόλου εκπαίδευσης και αξιολόγησης αντιστοιχούν σε απόλυτο σχετικό σφάλμα μικρότερο της μονάδας.

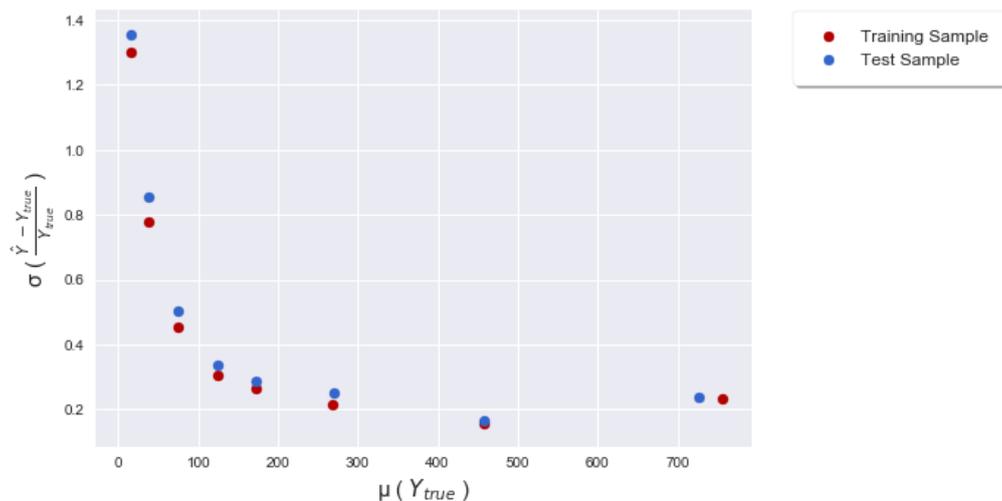
Ακολούθως, οι πραγματικές τιμές των παρατηρήσεων χωρίζονται σε υποσύνολα με τις αντίστοιχες τιμές των σχετικών σφαλμάτων στο διάστημα $[-5, 5]$, για μία καλύτερη και με μεγαλύτερη ευκρίνεια εκτίμηση της απόδοσης του μοντέλου για τις διάφορες παρατηρηθείσες τιμές. Συγκεκριμένα, το σύνολο των πραγματικών τιμών των παρατηρήσεων χωρίζεται αρχικά σε τέσσερα επιμέρους υποσύνολα με εύρος 200 μονάδες τα πρώτα τρία και 400 το τελευταίο. Ωστόσο, επειδή το πεδίο τιμών $[0, 200]$ περιλαμβάνει ένα μεγάλο αριθμό παρατηρήσεων, αυτό διαμερίζεται σε άλλα πέντε υποσύνολα, εύρους 50 μονάδων το καθένα με εξαίρεση τα πρώτα δύο που έχουν εύρος 25 μονάδες, με σκοπό η συμπεριφορά του μοντέλου να είναι πιο ξεκάθαρη. Στα Διαγράμματα 6.56 και 6.57 που ακολουθούν παρουσιάζονται οι τιμές για τη μέση τιμή και την τυπική απόκλιση των αποκομμένων σχετικών σφαλμάτων στο διάστημα $[-5, 5]$

για τα σύνολα εκπαίδευσης και αξιολόγησης, συναρτήσει της μέσης τιμής των πραγματικών παρατηρήσεων για κάθε ένα από τα υποσύνολα που προέκυψαν.

Τα αντίστοιχα ιστογράμματα που δημιουργήθηκαν για κάθε ένα από τα διαστήματα όπως αυτά περιγράφηκαν παραπάνω βρίσκονται στο Παράρτημα.



Διάγραμμα 6.56 : Διάγραμμα μέσης τιμής των αποκομμένων σχετικών σφαλμάτων των εκτιμήσεων στο διάστημα $[-5, 5]$ συναρτήσει της μέσης τιμής των παρατηρηθεισών τιμών για κάθε υποσύνολο αυτών.



Διάγραμμα 6.57 : Διάγραμμα τυπικής απόκλισης των αποκομμένων σχετικών σφαλμάτων των εκτιμήσεων στο διάστημα $[-5, 5]$ συναρτήσει της μέσης τιμής των παρατηρηθεισών τιμών για κάθε υποσύνολο αυτών.

Η μέση τιμή των σχετικών σφαλμάτων αποτελεί το μέσο όρο των σχετικών σφαλμάτων των εκτιμήσεων και περιγράφει τη μέση απόκλιση της κάθε εκτίμησης από την πραγματική τιμή της αντίστοιχης παρατήρησης. Επομένως, μπορεί να αποτελέσει ένα σημαντικό μέτρο για την εκτίμηση της απόδοσης και της προσαρμοστικότητας του μοντέλου. Αντίστοιχα, η τυπική απόκλιση που υπολογίζεται για τα σχετικά σφάλματα, δηλώνει πόσο αυτά διαφέρουν από τη μέση τιμή τους.

Από τα Διαγράμματα 6.56 και 6.57, παρατηρείται πως οι αποκλίσεις μεταξύ των τιμών για το σύνολο εκπαίδευσης και αξιολόγησης είναι πολύ μικρές, το οποίο δηλώνει την καλή προσαρμογή του μοντέλου στα δεδομένα και την ικανοποιητική απόδοσή του. Επιπρόσθετα, φαίνεται ότι για το πρώτο διάστημα, στο οποίο περιέχονται οι παρατηρήσεις που λαμβάνουν τις μικρότερες τιμές, η τιμή της τυπικής απόκλισης των σχετικών σφαλμάτων είναι η μεγαλύτερη για τα δύο σύνολα συγκριτικά με τις τιμές που λαμβάνει στα υπόλοιπα διαστήματα, το οποίο παρατηρείται και στο Διάγραμμα 6.5.25. Αυτό συμβαίνει λόγω των υψηλών συσχετίσεων που παρουσιάζει η μεταβλητή απόκρισης με τις επεξηγηματικές μεταβλητές, όταν αυτή λαμβάνει μεγάλες τιμές. Επομένως, για μεγάλες τιμές της μεταβλητής απόκρισης, το μοντέλο φαίνεται να έχει καλύτερη απόδοση, αφού σημειώνονται μικρότερα σφάλματα στις εκτιμήσεις του σε σχέση με αυτές που αντιστοιχούν σε μικρότερες τιμές της.

6.6 Σύγκριση Μοντέλων

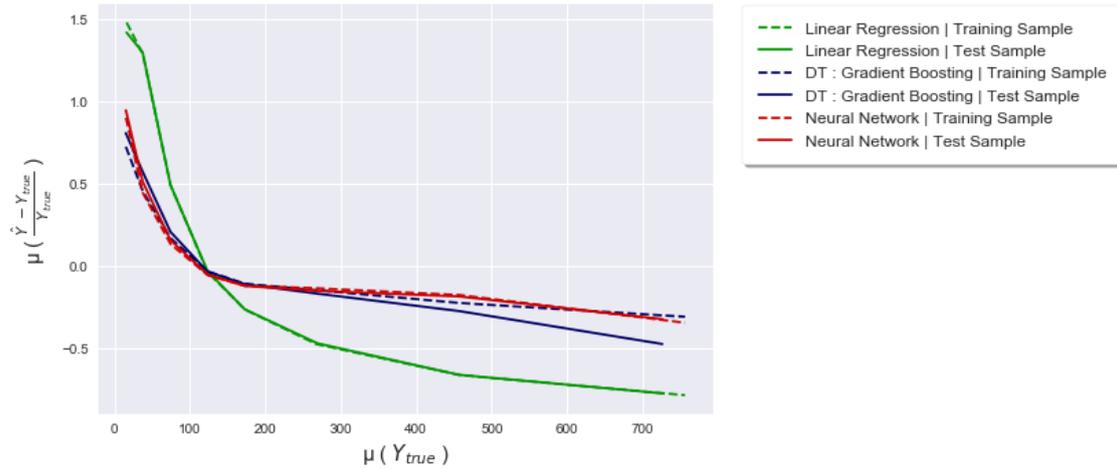
Με βάση όλα τα αποτελέσματα που έχουν προηγηθεί στο κεφάλαιο αυτό, ακολουθεί η σύγκριση των τελικών μοντέλων που επιλέχθηκαν μετά την εφαρμογή της κάθε μεθόδου, με σκοπό την ανίχνευση της πιο αποδοτικής μέσω της οποίας προκύπτουν τα πιο ικανοποιητικά αποτελέσματα για τα δεδομένα που διατίθενται. Σύμφωνα λοιπόν με τα παραπάνω, τα τελικά μοντέλα που έχουν προκύψει προς σύγκριση είναι τα εξής :

- ♦ Το μοντέλο που προέκυψε μέσω Γραμμικής Παλινδρόμησης,
- ♦ Το μοντέλο που προέκυψε με τη δημιουργία Δέντρων Απόφασης, και συγκεκριμένα με τη μέθοδο *Gradient Boosting*, με συνολικό αριθμό δέντρων 200, μέγιστο βάθος 6 και ρυθμό εκμάθησης 0.1,
- ♦ Το μοντέλο που προέκυψε από την κατασκευή Νευρωνικού Δικτύου, ειδικότερα που αποτελείται από τρία κρυφά στρώματα με 28, 25 και 20 νευρώνες σε κάθε ένα από αυτά αντίστοιχα, με εφαρμογή της συνάρτησης *Relu* ως συνάρτηση ενεργοποίησης.

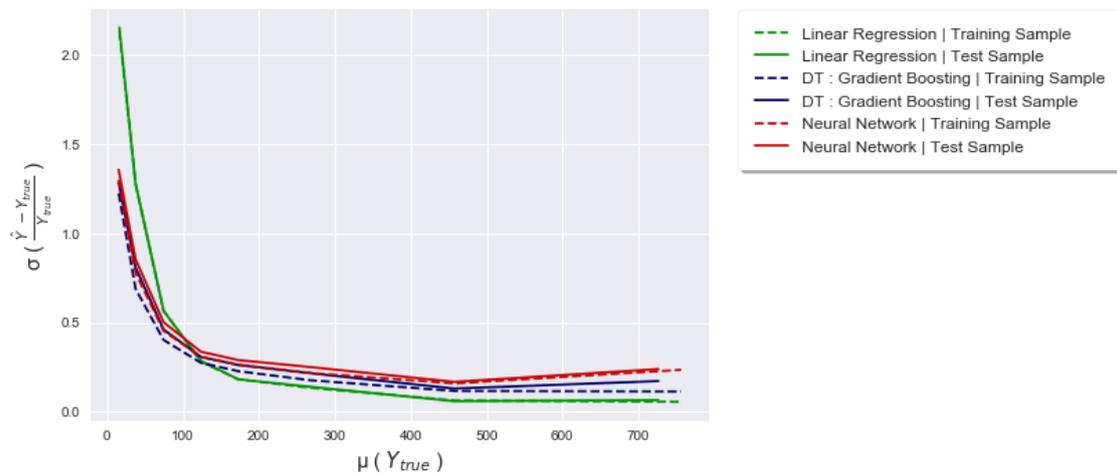
Σημείωση : Από τα παραπάνω μοντέλα απουσιάζει αυτό που προέκυψε με την εφαρμογή της μεθόδου *Adaboost*, καθώς τα αποτελέσματα δεν ήταν καθόλου

ικανοποιητικά και για το λόγο η μελέτη προχώρησε στην κατασκευή δέντρων απόφασης με χρήση της μεθόδου *Gradient Boosting*.

Τα Διαγράμματα 6.58 και 6.59 που ακολουθούν παρουσιάζουν τις τιμές για τη μέση τιμή και τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων, οι οποίες λαμβάνουν τιμές στο διάστημα $[-5, 5]$, συναρτήσει της μέσης τιμής των πραγματικών τιμών των παρατηρήσεων, για τα διάφορα διαστήματα στα οποία αυτές διαμερίζονται – όπως έχει αναφερθεί και παραπάνω.



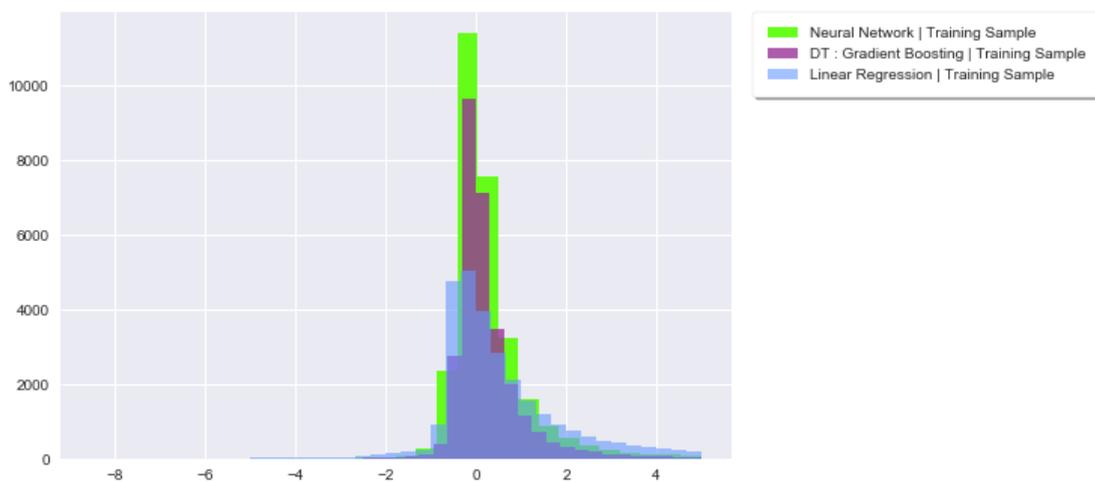
Διάγραμμα 6.58 : Διάγραμμα μέσης τιμής των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$, συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα συνολικά καλύτερα μοντέλα.



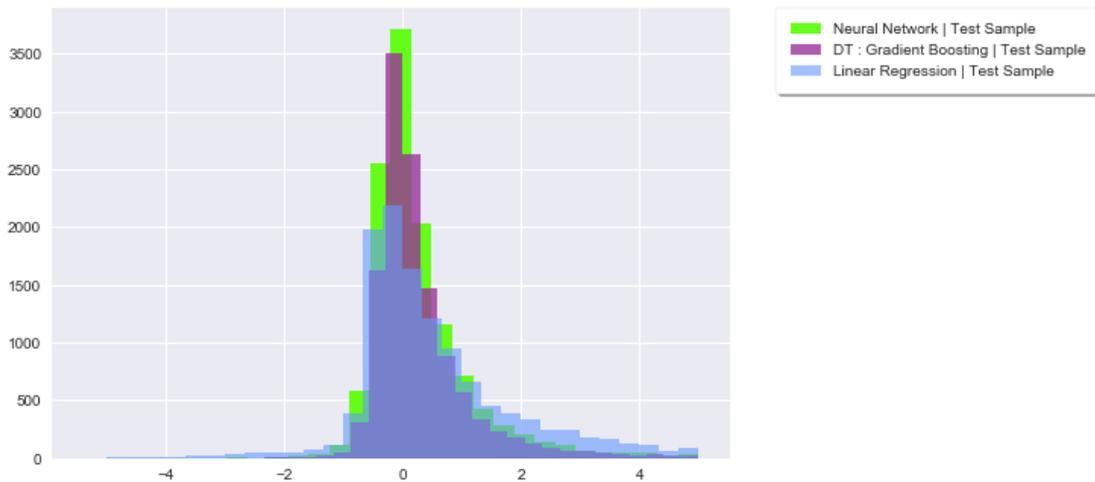
Διάγραμμα 6.59 : Διάγραμμα τυπικής απόκλισης των σχετικών σφαλμάτων των εκτιμήσεων των παρατηρήσεων στο διάστημα $[-5, 5]$, συναρτήσει της μέσης τιμής των παρατηρούμενων τιμών ανά τμήματα του πεδίου τιμών τους, για τα συνολικά καλύτερα μοντέλα.

Από τα δύο τελευταία διαγράμματα παρατηρείται πως οι τιμές μεταξύ των συνόλων εκπαίδευσης και αξιολόγησης παρουσιάζουν μικρές διαφορές, το οποίο πιστοποιεί την επιλογή τους ως τα βέλτιστα μοντέλα της κάθε μεθόδου. Επιπλέον, όμοια με προηγουμένως, φαίνεται πως όταν οι πραγματικές τιμές των παρατηρήσεων ανήκουν στο διάστημα $[0, 25]$, δηλαδή λαμβάνουν τις μικρότερες τιμές, τα μοντέλα δυσκολεύονται να τις εκτιμήσουν με μικρό σφάλμα. Αυτό γίνεται αντιληπτό διότι τόσο η μέση τιμή, όσο και η τυπική απόκλιση των σχετικών σφαλμάτων των εκτιμήσεων μειώνονται καθώς αυξάνεται η πραγματική τιμή των παρατηρήσεων, που σημαίνει ότι οι εκτιμώμενες τιμές δεν παρουσιάζουν σημαντικές διαφορές με τις πραγματικές τιμές των παρατηρήσεων. Το γεγονός αυτό δηλώνει την ύπαρξη υψηλών συσχετίσεων που παρουσιάζει η μεταβλητή απόκρισης με τις επεξηγηματικές μεταβλητές, όταν αυτή λαμβάνει μεγάλες τιμές. Επομένως, για μεγάλες τιμές της μεταβλητής απόκρισης, τα μοντέλα φαίνεται να έχουν καλύτερη απόδοση, αφού σημειώνονται μικρότερα σφάλματα στις εκτιμήσεις του σε σχέση με αυτές που αντιστοιχούν σε μικρότερες τιμές της.

Επιπλέον, από τα Διαγράμματα 6.58 και 6.59 φαίνεται πως το μοντέλο που προκύπτει από Γραμμική Παλινδρόμηση έχει αρκετά καλή απόδοση, πράγμα που δε δικαιολογείται από τα Διαγράμματα 6.60 και 6.61, που ακολουθούν, καθώς και ούτε από τα διαγράμματα της Παραγράφου 6.3. Τα Διαγράμματα 6.60 και 6.61 που παρατίθενται στη συνέχεια αποτελούν τα ιστογράμματα συχνότητας για τα αποκομμένα σχετικά σφάλματα των εκτιμώμενων τιμών στο διάστημα $[-5, 5]$, που προκύπτουν από τα συνολικά καλύτερα μοντέλα για την κάθε μέθοδο που εφαρμόστηκε, για το σύνολο εκπαίδευσης και αξιολόγησης αντίστοιχα.



Διάγραμμα 6.60 : Ιστόγραμμα συχνότητας για τα σχετικά σφάλματα των εκτιμήσεων για τα συνολικά καλύτερα μοντέλα που λαμβάνουν τιμές στο διάστημα $[-5, 5]$ και αφορούν το σύνολο εκπαίδευσης.



Διάγραμμα 6.61 : Ιστόγραμμα συχνοτήτων για τα σχετικά σφάλματα των εκτιμήσεων για τα συνολικά καλύτερα μοντέλα που λαμβάνουν τιμές στο διάστημα $[-5, 5]$ και αφορούν το σύνολο αξιολόγησης.

Από τα Διαγράμματα 6.60 και 6.61 φαίνεται πως και τα τρία μοντέλα έχουν παρόμοια συμπεριφορά για τα σύνολα εκπαίδευσης και αξιολόγησης. Ωστόσο, από αυτά συμπεραίνεται ότι αυτό που προκύπτει από Γραμμική Παλινδρόμηση υστερεί σε σχέση με τα άλλα, καθώς το ποσοστό των σχετικών σφαλμάτων που είναι μικρότερα κατά απόλυτη τιμή από τη μονάδα είναι μικρότερο από εκείνο των άλλων δύο μοντέλων. Ειδικότερα, φαίνεται πως το μοντέλο που προκύπτει από κατασκευή Νευρωνικού Δικτύου είναι αυτό για το οποίο οι περισσότερες τιμές των σχετικών σφαλμάτων των εκτιμήσεων συγκεντρώνονται αρκετά κοντά στο μηδέν. Επομένως, οι εκτιμήσεις που προκύπτουν από την εφαρμογή του είναι αυτές που σημειώνουν τις μικρότερες αποκλίσεις από τις πραγματικές τιμές των παρατηρήσεων.

Το γεγονός αυτό επιβεβαιώνει και ο ακόλουθος πίνακας, ο οποίος περιέχει τις τιμές του μέσου τετραγωνικού σφάλματος για τα σύνολα εκπαίδευσης και αξιολόγησης των τριών παραπάνω μοντέλων. Όπως φαίνεται η μικρότερη διαφορά μεταξύ των τιμών αυτών για τα δύο σύνολα σημειώνεται για το μοντέλο που προκύπτει από την κατασκευή του Νευρωνικού Δικτύου.

	<i>MSE</i>	
	<i>Train</i>	<i>Test</i>
<i>Linear Regression</i>	6064.925	6320.837
<i>Decision Trees : Gradient Boosting</i>	1469.073	2168.020
<i>Neural Network</i>	1892.816	2228.143

Πίνακας 6.7 : Πίνακας τιμών του μέσου τετραγωνικού σφάλματος για τα σύνολα εκπαίδευσης και αξιολόγησης των τελικών καλύτερων μοντέλων.

Καταληκτικά, σύμφωνα με όλες τις παραπάνω συγκρίσεις, καθώς επίσης και από τα διαγράμματα των εκτιμώμενων τιμών που προέκυψαν από το κάθε μοντέλο συναρτήσει των πραγματικών τιμών των παρατηρήσεων, συμπεραίνεται πως το πιο αποδοτικό μοντέλο είναι αυτό που προκύπτει από την κατασκευή του νευρωνικού Δικτύου, όπως αυτό περιγράφεται στην αρχή της Παραγράφου 6.6. Το συμπέρασμα αυτό ενισχύεται παρατηρώντας επίσης και τις γραφικές παραστάσεις που περιγράφουν τη συνάρτηση κόστους των μοντέλων. Από αυτές φαίνεται πως εκείνη που αντιστοιχεί στο Νευρωνικό Δίκτυο έχει την πιο ομαλή συμπεριφορά, που δηλώνει ότι οι καμπύλες του συνόλου εκπαίδευσης και αξιολόγησης συμβαδίζουν παράλληλα, αποκλείοντας την εμφάνιση του φαινομένου υπερεκπαίδευσης, αφού το μοντέλο δε μοιάζει να προσαρμόζονται σε τυχόν μοτίβα του πρώτου και να εξειδικεύεται στην εκτίμηση των τιμών αυτού, επομένως κρίνεται ικανοποιητικό.

Βιβλιογραφία

1. Bekkerman, R., Bilenko, M., & Langford, J. (Eds.). (2011). *Scaling up machine learning: Parallel and distributed approaches*. Cambridge University Press.
2. Medium. (2019). *9 Applications of Machine Learning from Day-to-Day Life*. [online] Available at: <https://medium.com/app-affairs/9-applications-of-machine-learning-from-day-to-day-life-112a47a429d0> [Accessed 4 Sep. 2019].
3. Sas.com. (2019). *Machine Learning: What it is and why it matters*. [online] Available at: https://www.sas.com/en_us/insights/analytics/machine-learning.html [Accessed 4 Sep. 2019].
4. Alpaydin, E. (2009). *Introduction to machine learning*. MIT press.
5. Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.
6. Theodoridis, S., & Koutroumbas, K. (2009). *Pattern recognition*. 2003. Elsevier Inc.
7. Xing, Y. F., Xu, Y. H., Shi, M. H., & Lian, Y. X. (2016). The impact of PM_{2.5} on the human respiratory system. *Journal of thoracic disease*, 8(1), E69.
8. Epa.vic.gov.au. (2019). *PM_{2.5} particles in air | Environment Protection Authority Victoria | EPA Victoria*. [online] Available at: <https://www.epa.vic.gov.au/your-environment/air/air-pollution/pm25-particles-in-air> [Accessed 5 Sep. 2019].
9. Miller, L., & Xu, X. (2018). Ambient PM_{2.5} Human Health Effects—Findings in China and Research Directions. *Atmosphere*, 9(11), 424.
10. Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.
11. Φουσκάκης, Δ. (2013). *Ανάλυση Δεδομένων με Χρήση της R*. Εκδόσεις ΤΣΟΤΡΑΣ.
12. Moisen, G. G. (2008). Classification and regression trees. In: *Jørgensen, Sven Erik; Fath, Brian D.(Editor-in-Chief). Encyclopedia of Ecology, volume 1. Oxford, UK: Elsevier. p. 582-588.*, 582-588.
13. Brownlee, J. (2016). *Master Machine Learning Algorithms: discover how they work and implement them from scratch*. Machine Learning Mastery.
14. Matloff, N. (2017). *Statistical regression and classification: from linear models to machine learning*. Chapman and Hall/CRC.
15. Drucker, H. (1997, July). Improving regressors using boosting techniques. In *ICML* (Vol. 97, pp. 107-115).
16. Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2), 83-85.
17. Explain that Stuff. (2019). *How neural networks work - A simple introduction*. [online] Available at: <https://www.explainthatstuff.com/introduction-to-neural-networks.html> [Accessed 11 Sep. 2019].

18. Nielsen, M. A. (2015). *Neural networks and deep learning* (Vol. 25). San Francisco, CA, USA:: Determination press.
19. Pardoe, D., & Stone, P. (2010, June). Boosting for regression transfer. In *Proceedings of the 27th International Conference on International Conference on Machine Learning* (pp. 863-870). Omnipress.
20. Liang X, Zou T, Guo B, Li S, Zhang H, Zhang S, Huang H, Chen SX. 2015 Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating.

Παράρτημα

Πίνακες ποσοστών συγκέντρωσης του αιωρούμενου σωματιδίου PM_{2.5}

Year	Low	Polluting episode	Very high
2010	25.45 %	50.43 %	24.13 %
2011	31.00 %	46.66 %	22.34 %
2012	31.97 %	48.60 %	19.43 %
2013	27.47 %	50.68 %	21.85 %
2014	29.27 %	50.68 %	20.06 %
2010 – 2014	29.03 %	49.44 %	21.53 %

Διάγραμμα A.1 : Ετήσια ποσοστά συγκέντρωσης του αιωρούμενου σωματιδίου PM_{2.5} ανά κατηγορία ποιότητα αέρα.

Season	Low	Polluting episode	Very high
Spring	28.32 %	54.95 %	16.74 %
Summer	21.57 %	61.15 %	17.28 %
Autumn	30.96 %	44.12 %	24.92 %
Winter	35.26 %	37.47 %	27.27 %

Διάγραμμα A.2 : Ποσοστά συγκέντρωσης του αιωρούμενου σωματιδίου PM_{2.5} ανά εποχή για το διάστημα των ετών 2010-2014 για κάθε μία κατηγορία ποιότητας αέρα.

Gradient Descent

Η *Gradient Descent* αποτελεί έναν αλγόριθμο βελτιστοποίησης που χρησιμοποιείται για την εύρεση των τιμών των παραμέτρων που ελαχιστοποιούν μία συνάρτηση κόστους. Συνήθως χρησιμοποιείται όταν οι παράμετροι δεν μπορούν να υπολογιστούν αναλυτικά, δηλαδή με χρήση γραμμικής άλγεβρας και χρειάζεται να προσδιοριστούν από έναν αλγόριθμο βελτιστοποίησης [13]. Στόχος είναι οι επαναλαμβανόμενες δοκιμές διαφορετικών τιμών για τις παραμέτρους και ο υπολογισμός της συνάρτησης κόστους σε κάθε βήμα, μέχρι η τιμή της να είναι η ελάχιστη δυνατή. Τότε ο συνδυασμός των τιμών των παραμέτρων συνεπάγεται το ελάχιστο κόστος.

Η κλίση μίας διαφορίσιμης συνάρτησης $f : \mathbb{R}^d \rightarrow \mathbb{R}$ στο \mathbf{w} , συμβολίζεται ως $\nabla f(\mathbf{w})$ και είναι ένα διάνυσμα των μερικών παραγώγων της f , συγκεκριμένα $\nabla f(\mathbf{w}) = \left(\frac{\partial f(\mathbf{w})}{\partial w_1}, \dots, \frac{\partial f(\mathbf{w})}{\partial w_d} \right)$. Ο αλγόριθμος της μεθόδου ξεκινά ορίζοντας μία τυχαία αρχική τιμή

για το διάνυσμα \mathbf{w} , για παράδειγμα $\mathbf{w}^{(l)} = \mathbf{0}$. Σε κάθε βήμα της διαδικασίας, το \mathbf{w} ανανεώνεται, ακολουθώντας αντίθετη κατεύθυνση από αυτή της κλίσης, δηλαδή :

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta \nabla f(\mathbf{w}^{(t)})$$

όπου η παράμετρος η καλείται μέγεθος του βήματος ή ρυθμός εκμάθησης και καθορίζει τη μεταβολή του στην κατεύθυνση αυτή σε κάθε βήμα.

Η επιλογή κατάλληλης τιμής για την παράμετρο η αποτελεί σημαντικό παράγοντα του αλγορίθμου, καθώς αν είναι πολύ μικρή τότε η αυτός μπορεί να συγκλίνει πολύ αργά, ενώ αν είναι αρκετά μεγάλη μπορεί να προκαλέσει ταλαντώσεις ή να μην οδηγήσει σε σύγκλιση του αλγορίθμου [4].

Έπειτα από T βήματα, ο αλγόριθμος καταλήγει στο μέσο όρο $\bar{\mathbf{w}} = \frac{1}{T} \sum_{t=1}^T \mathbf{w}^{(t)}$. Το τελικό αποτέλεσμα μπορεί επίσης να είναι και το τελευταίο διάνυσμα που υπολίζεται από την τελευταία σχέση ή το διάνυσμα με την καλύτερη επίδοση, δηλαδή $\arg \min_{t \in [T]} f(\mathbf{w}^{(t)})$. Παίρνοντας τον μέσο όρο φαίνεται πως είναι ιδιαίτερα χρήσιμο όταν η μέθοδος γενικεύεται για μη παραγωγίσιμες συναρτήσεις [5].

Ερμηνευτικοί Όροι

Διασπορά (Variance)

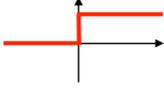
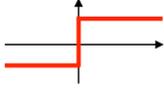
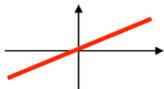
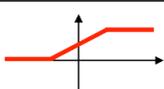
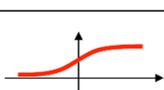
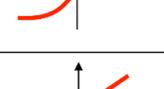
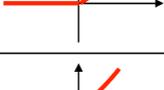
Η διασπορά αποτελεί τη διαφορά μεταξύ των προβλέψεων ενός μοντέλου κατά την εφαρμογή του σε διαφορετικά σύνολα δεδομένων από αυτό στο οποίο εκπαιδεύτηκε.

Μοντέλα με μεγάλη διασπορά προσαρμόζονται στο δείγμα εκπαίδευσης και δεν γενικεύονται όταν εισάγονται σε αυτά νέα δεδομένα. Αποτελεσματικά, έχουν καλή απόδοση στα δεδομένα εκπαίδευσης, αλλά υψηλό ποσοστό σφάλματος στα δεδομένα αξιολόγησης.

Weak learner

Weak learner είναι ένα μοντέλο το οποίο αποδίδει καλύτερα από ότι αν τα αποτελέσματά του προέκυπταν από τύχη προσπαθώντας να ταξινομήσει τα δεδομένα, για οποιαδήποτε κατανομή ακολουθούν τα δεδομένα εκπαίδευσης. Αποδίδοντας καλύτερα από ότι με τυχαίο τρόπο σημαίνει ότι πάντα θα υπάρχει ένα ποσοστό σφάλματος μικρότερο του $\frac{1}{2}$. Έτσι, ο αλγόριθμος πάντα θα 'μαθαίνει' κάτι από τα δεδομένα, αλλά δεν θα είναι πάντα ακριβές με αποτέλεσμα να έχει χαμηλή απόδοση εκτιμώντας τις συσχετίσεις μεταξύ των επεξηγηματικών μεταβλητών και της μεταβλητής απόκρισης. Στις μεθόδους Boosting ένας weak learner είναι συνήθως ένα δέντρο μικρού βάθους, δηλαδή έχει 2 ή 3 επίπεδα.

Διαφορετικοί Τύποι Συναρτήσεων Ενεργοποίησης

Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multi-layer Neural Networks	
Rectifier, ReLU (Rectified Linear Unit)	$\phi(z) = \max(0, z)$	Multi-layer Neural Networks	
Rectifier, softplus	$\phi(z) = \ln(1 + e^z)$	Multi-layer Neural Networks	

Copyright © Sebastian Raschka 2016
(<http://sebastianraschka.com>)

Βιβλιοθήκες της προγραμματιστικής γλώσσας Python

Για την κατασκευή των μοντέλων που περιγράφονται στη συγκεκριμένη διπλωματική εργασία χρησιμοποιήθηκαν οι ακόλουθες βιβλιοθήκες της προγραμματιστικής γλώσσας *Python*:

- NumPy, SciPy & Pandas, για την επεξεργασία των δεδομένων
- Scikit-learn, για την κατασκευή των γραμμικών και μη γραμμικών μοντέλων (Linear Regression, Adaboost Regressor, Gradient Regressor)
- Keras, για την κατασκευή των νευρωνικών δικτύων
- Matplotlib, Seaborn & Plotly, για την κατασκευή των διαγραμμάτων

Εφαρμογή αλγορίθμου AdaBoost

- *Max depth = 3*

		<i>Learning Rate</i>					
		0.1	0.2	0.4	0.6	0.8	1.0
<i>No of Trees</i>	50	1.58	1.91	2.3	2.61	2.59	2.93
	100	1.95	2.40	2.3	2.61	2.59	2.93
	150	2.19	2.46	2.3	2.61	2.59	2.93
	200	2.34	2.46	2.3	2.61	2.59	2.93
	250	2.34	2.46	2.3	2.61	2.59	2.93
	300	2.34	2.46	2.3	2.61	2.59	2.93
	350	2.34	2.46	2.3	2.61	2.59	2.93
	400	2.34	2.46	2.3	2.61	2.59	2.93
	450	2.34	2.46	2.3	2.61	2.59	2.93
	500	2.34	2.46	2.3	2.61	2.59	2.93

Πίνακας με τη μέση τιμή των σχετικών σφαλμάτων για μοντέλα που προκύπτουν από διαφορετικούς συνδυασμούς του συνολικού αριθμού δέντρων (γραμμές) και του ρυθμού εκμάθησης (στήλες) για δέντρα βάθους 3.

		<i>Learning Rate</i>					
		0.1	0.2	0.4	0.6	0.8	1.0
<i>No of Trees</i>	50	3.28	3.80	4.24	4.6	4.55	5.1
	100	3.87	4.36	4.24	4.6	4.55	5.1
	150	4.15	4.45	4.24	4.6	4.55	5.1
	200	4.32	4.45	4.24	4.6	4.55	5.1
	250	4.32	4.45	4.24	4.6	4.55	5.1
	300	4.32	4.45	4.24	4.6	4.55	5.1
	350	4.32	4.45	4.24	4.6	4.55	5.1
	400	4.32	4.45	4.24	4.6	4.55	5.1
	450	4.32	4.45	4.24	4.6	4.55	5.1
	500	4.32	4.45	4.24	4.6	4.55	5.1

Πίνακας με την τυπική απόκλιση των σχετικών σφαλμάτων για μοντέλα που προκύπτουν από διαφορετικούς συνδυασμούς του συνολικού αριθμού δέντρων (γραμμές) και του ρυθμού εκμάθησης (στήλες) για δέντρα βάθους 3.

- *Max depth = 4*

		<i>Learning Rate</i>					
		0.1	0.2	0.4	0.6	0.8	1.0
<i>No of Trees</i>	50	1.39	1.71	2.22	2.56	2.70	2.92
	100	1.70	2.20	2.81	3.02	2.89	3.34
	150	1.98	2.52	3.10	3.02	2.89	3.34
	200	2.21	2.76	3.10	3.02	2.89	3.34
	250	2.38	2.76	3.10	3.02	2.89	3.34
	300	2.54	2.76	3.10	3.02	2.89	3.34
	350	2.66	2.76	3.10	3.02	2.89	3.34
	400	2.76	2.76	3.10	3.02	2.89	3.34
	450	2.89	2.76	3.10	3.02	2.89	3.34
	500	3.03	2.76	3.10	3.02	2.89	3.34

Πίνακας με τη μέση τιμή των σχετικών σφαλμάτων για μοντέλα που προκύπτουν από διαφορετικούς συνδυασμούς του συνολικού αριθμού δέντρων (γραμμές) και του ρυθμού εκμάθησης (στήλες) για δέντρα βάθους 4.

		<i>Learning Rate</i>					
		0.1	0.2	0.4	0.6	0.8	1.0
<i>No of Trees</i>	50	3.00	3.32	3.94	4.31	4.64	4.84
	100	3.30	3.93	4.63	4.91	4.85	5.33
	150	3.63	4.36	5.00	4.91	4.85	5.33
	200	3.95	4.64	5.00	4.91	4.85	5.33
	250	4.16	4.64	5.00	4.91	4.85	5.33
	300	4.34	4.64	5.00	4.91	4.85	5.33
	350	4.47	4.64	5.00	4.91	4.85	5.33
	400	4.58	4.64	5.00	4.91	4.85	5.33
	450	4.74	4.64	5.00	4.91	4.85	5.33
	500	4.92	4.64	5.00	4.91	4.85	5.33

Πίνακας με την τυπική απόκλιση των σχετικών σφαλμάτων για μοντέλα που προκύπτουν από διαφορετικούς συνδυασμούς του συνολικού αριθμού δέντρων (γραμμές) και του ρυθμού εκμάθησης (στήλες) για δέντρα βάθους 4.

Εφαρμογή αλγορίθμου Gradient Boosting

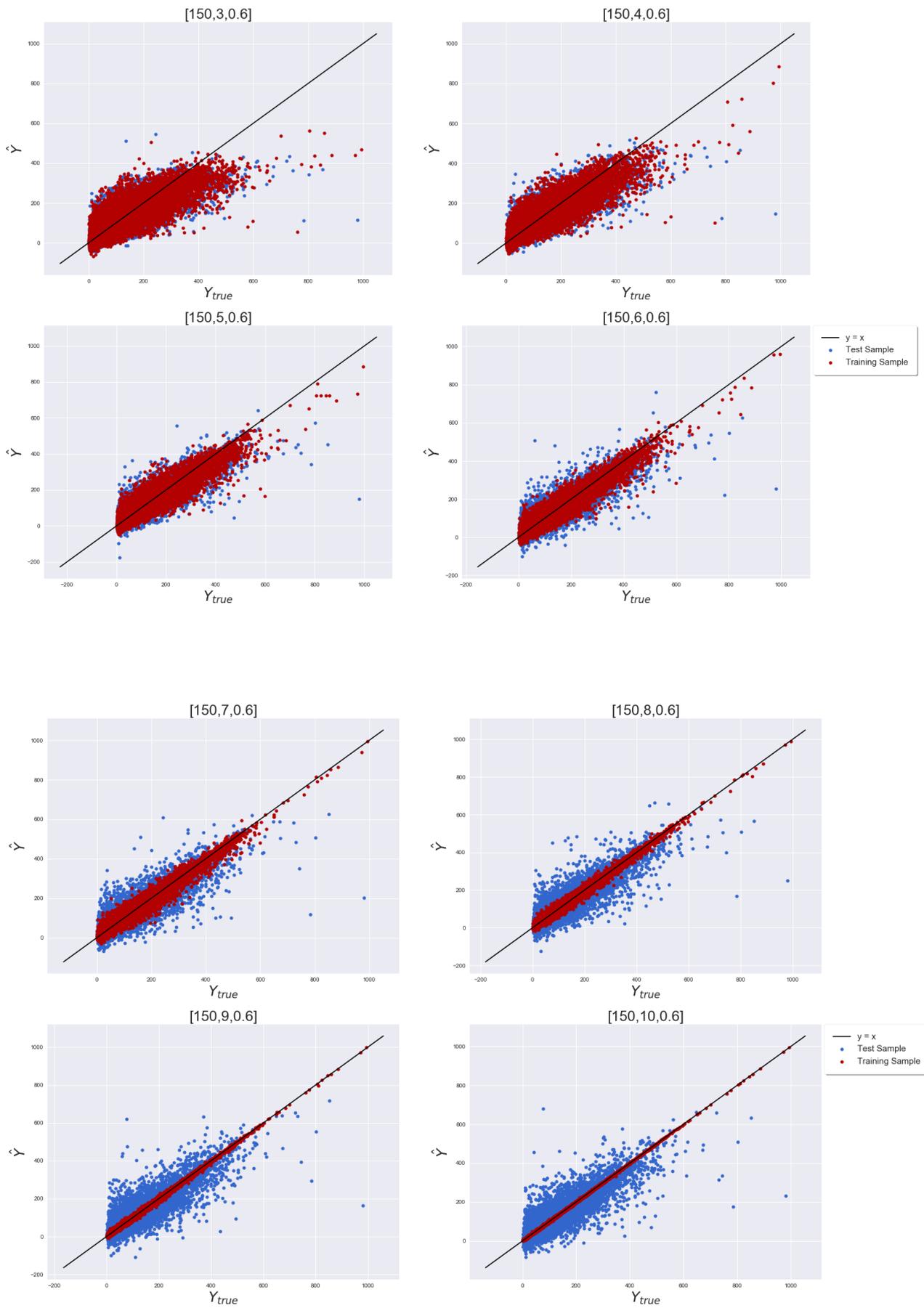
		Learning Rate					
		0.1	0.2	0.4	0.6	0.8	1.0
No of Trees	50	0.89	0.69	0.54	0.46	0.45	0.43
	100	0.7	0.54	0.45	0.42	0.40	0.37
	150	0.62	0.49	0.42	0.39	0.37	0.36
	200	0.56	0.46	0.40	0.37	0.34	0.33
	300	0.51	0.42	0.38	0.35	0.34	0.33
	400	0.47	0.41	0.36	0.33	0.32	0.30
	500	0.45	0.39	0.34	0.32	0.31	0.30

Πίνακας με τη μέση τιμή των σχετικών σφαλμάτων για μοντέλα που προκύπτουν από διαφορετικούς συνδυασμούς του συνολικού αριθμού δέντρων (γραμμές) και του ρυθμού εκμάθησης (στήλες) για δέντρα βάθους 3.

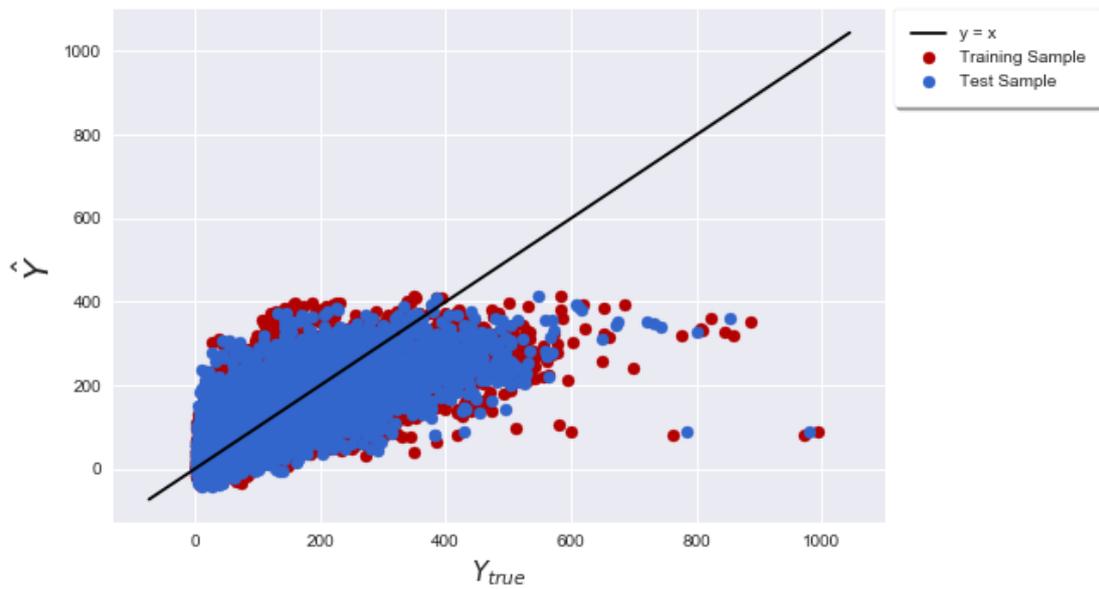
		Learning Rate					
		0.1	0.2	0.4	0.6	0.8	1.0
No of Trees	50	2.11	1.81	1.73	1.71	1.76	1.87
	100	1.82	1.70	1.63	1.59	1.63	1.71
	150	1.73	1.64	1.57	1.53	1.56	1.67
	200	1.69	1.59	1.54	1.51	1.54	1.67
	300	1.65	1.52	1.49	1.47	1.53	1.64
	400	1.61	1.47	1.45	1.45	1.52	1.60
	500	1.59	1.44	1.43	1.43	1.51	1.59

Πίνακας με την τυπική απόκλιση των σχετικών σφαλμάτων για μοντέλα που προκύπτουν από διαφορετικούς συνδυασμούς του συνολικού αριθμού δέντρων (γραμμές) και του ρυθμού εκμάθησης (στήλες) για δέντρα βάθους 3.

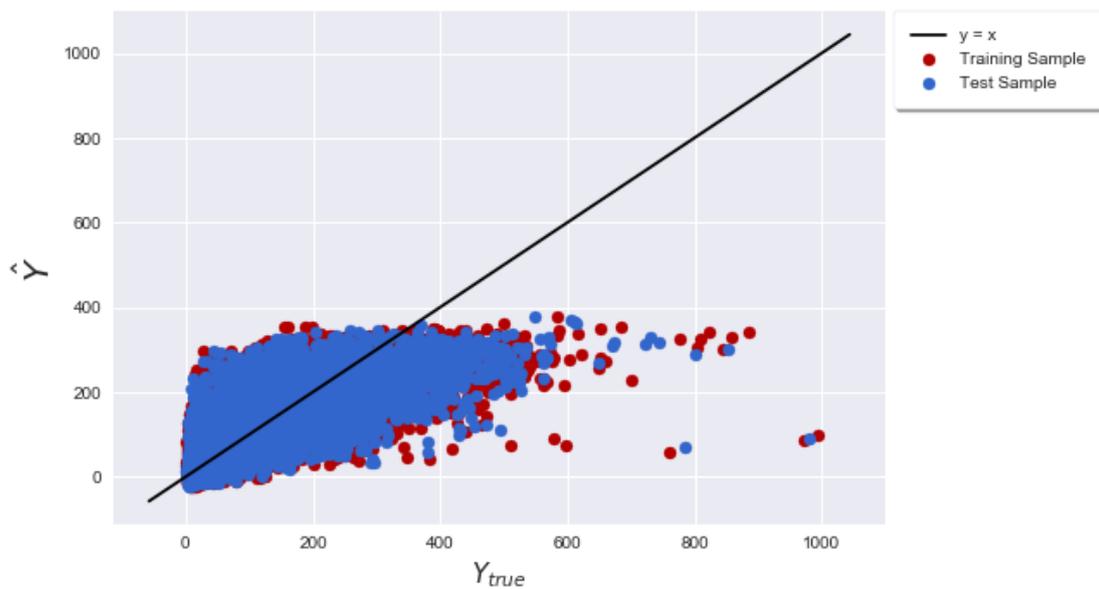
- Τα παρακάτω διαγράμματα αναπαριστούν τις εκτιμώμενες τιμές που προκύπτουν από μοντέλα 150 δέντρων με ρυθμό εκμάθησης 0.6 όταν το βάθος του κάθε δέντρου λαμβάνει τιμές στο διάστημα [3, 10]. Ο συμβολισμός [x, y, z] που τοποθετείται ως τίτλος σε κάθε διάγραμμα αναφέρεται στο μοντέλο που κατασκευάζεται με x συνολικό αριθμό δέντρων βάθους y και με ρυθμό εκμάθησης z.



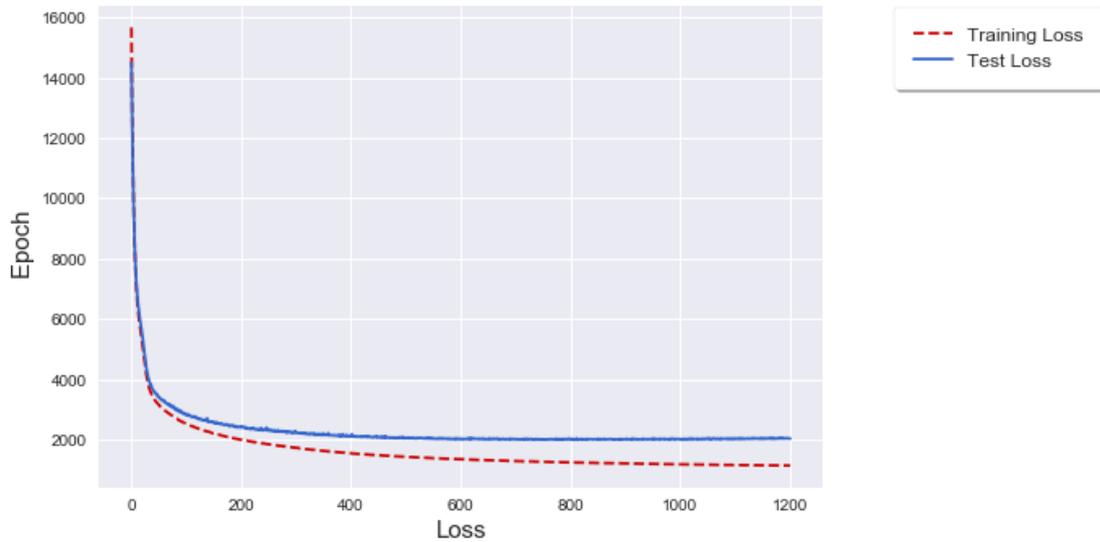
Κατασκευή Νευρωνικού Δικτύου



Διάγραμμα εκτιμώμενων τιμών των παρατηρήσεων συναρτήσει των πραγματικών τους τιμών για νευρωνικό δίκτυο με ένα κρυφό στρώμα αποτελούμενο από 42 νευρώνες με εφαρμογή της *Relu* ως συνάρτηση ενεργοποίησης.



Διάγραμμα εκτιμώμενων τιμών των παρατηρήσεων συναρτήσει των πραγματικών τους τιμών για νευρωνικό δίκτυο με ένα κρυφό στρώμα αποτελούμενο από 42 νευρώνες με εφαρμογή της *Sigmoid* ως συνάρτηση ενεργοποίησης.



Διάγραμμα συνάρτησης κόστους συναρτήσει του αριθμού των εποχών για νευρικό δίκτυο αποτελούμενο από δύο κρυφά στρώματα από 70 και 40 νευρώνες αντίστοιχα, με εφαρμογή της υπερβολικής εφαπτομένης ως συνάρτησης ενεργοποίησης μεταξύ αυτών.

- Τα παρακάτω ιστογράμματα αντιστοιχούν στο τελικό μοντέλο που περιγράφεται στην ενότητα 6.5.

