



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ
& ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Πρόβλεψη ζήτησης προϊόντων σε κεντρικές αποθήκες με χρήση αλγορίθμων μηχανικής μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αναστάσιος Δ. Καλτσούνης

Επιβλέπων : Ασημακόπουλος Βασίλειος
Καθηγητής Ε.Μ.Π.

Υπεύθυνος : Ευάγγελος Σπηλιώτης
Διδάκτωρ Ε.Μ.Π.

Αθήνα, Μάρτιος 2020



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ
& ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Πρόβλεψη ζήτησης προϊόντων σε κεντρικές αποθήκες με χρήση αλγορίθμων μηχανικής μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αναστάσιος Δ. Καλτσούνης

Επιβλέπων : Ασημακόπουλος Βασίλειος
Καθηγητής Ε.Μ.Π.

Υπεύθυνος : Ευάγγελος Σπηλιώτης
Διδάκτωρ Ε.Μ.Π.

Εγκρίθηκε από την τριμελή επιτροπή την <ημερομηνία>

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Βασίλειος Ασημακόπουλος

.....
Ιωάννης Ψαρράς

.....
Δημήτριος Ασκούνης

Αθήνα, Μάρτιος 2020



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ
& ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Copyright © Αναστάσιος Καλτσούνης, 2020.

Με την επιφύλαξη παντός δικαιώματος. All rights reserved

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 τους συγγραφείς.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

(Υπογραφή)

.....

Καλτσούνης Αναστάσιος
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Ηλεκτρονικών
Υπολογιστών

Ευχαριστίες

Η διπλωματική αυτή εργασία εκπονήθηκε στα πλαίσια των ερευνητικών δραστηριοτήτων της Μονάδας Συστημάτων Πρόβλεψης και Στρατηγικής κατά το ακαδημαϊκό έτος 2019 – 2020. Η μονάδα υπάγεται στον Τομέα Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων, της Σχολής Ηλεκτρολόγων Μηχανικών & Μηχανικών Η/Υ, του Εθνικού Μετσόβιου Πολυτεχνείου.

Πριν την παρουσίαση της διπλωματικής μου εργασίας, θα ήθελα να εκφράσω τις ευχαριστίες μου σε όσους συνέβαλλαν στη διεκπεραίωσή της.

Αρχικά, θα ήθελα να ευχαριστήσω θερμότατα τον επιβλέποντα καθηγητή κ. Ασημακόπουλο Βασίλειο για την ευκαιρία την οποία μου έδωσε μέσω της ανάθεσης αυτής της διπλωματικής εργασίας να ασχοληθώ σε βάθος με δύο τομείς ιδιαίτερα σημαντικούς για εμένα. Τόσο ο τομέας των Προβλέψεων όσο και ο τομέας της Μηχανικής Μάθησης και των Νευρωνικών Δικτύων αποτελούν κυρίαρχους άξονες για την μελλοντική επιστημονική και επαγγελματική μου ενασχόληση.

Επίσης, θα ήθελα να ευχαριστήσω τον καθηγητή Ψαρρά Ιωάννη και τον καθηγητή Ασκούνη Δημήτριο για την τιμή που μου έκαναν να συμμετάσχουν στην επιτροπή εξέτασης της εργασίας.

Θα ήθελα να κάνω ιδιαίτερη αναφορά και να δώσω πολλές ευχαριστίες στον Διδάκτορα κ. Σπηλιώτη Ευάγγελο. Η γνώση, η καθοδήγησή του, αλλά και η βοήθεια που μου παρείχε κατά την εκπόνηση της εργασίας, έδρασαν καταλυτικά στην προσπάθειά μου μέχρι και την ολοκλήρωσή της.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου. Τους γονείς μου Δημήτριο και Παναγιώτα και την πρωτίστως ψυχολογική και μετέπειτα οικονομική στήριξη που μου παρείχαν στα χρόνια των σπουδών μου. Τον αδερφό μου Θωμά για την αδιάκοπη συμπαράσταση του όλα αυτά τα χρόνια, αλλά και όλους τους φίλους μου που ήταν συνεχώς δίπλα μου στην πορεία μου ως φοιτητής της σχολής αυτής.

Καλτσούνης Αναστάσιος
Αθήνα, Μάρτιος 2020

Περίληψη

Στόχος της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη και η αξιολόγηση μίας μεθόδου πρόβλεψης που βασίζεται σε αλγόριθμο μηχανικής μάθησης και θα αποσκοπεί στην παραγωγή προβλέψεων σε χρονοσειρές διακοπτόμενης ζήτησης. Πιο συγκεκριμένα, η μέθοδος ασχολείται με χρονοσειρές που υπόκεινται σε ιεραρχικές δομές και καλείται να συνδυάσει τα δεδομένα από τις χρονοσειρές αυτές για την παραγωγή τελικής πρόβλεψης. Για το συγκεκριμένο εγχείρημα χρησιμοποιήθηκε ένα νευρωνικό δίκτυο μαζί με τη μέθοδο της παλινδρόμησης για την επεξεργασία των δεδομένων και την παραγωγή των προβλέψεων.

Εισαγωγικά, γίνεται μία αναφορά στο κομμάτι της ενασχόλησης που έχει διενεργηθεί στο κομμάτι της πρόβλεψης με νευρωνικά δίκτυα. Στη συνέχεια, παρουσιάζονται οι κυριότερες μορφές πρόβλεψης αλλά και οι ευρέως χρησιμοποιούμενες μέθοδοι στατιστικής πρόβλεψης. Γίνεται μία αναλυτική περιγραφή της έννοιας των ιεραρχικών δομών στις χρονοσειρές, αλλά και στις κυριότερες μεθόδους συμψηφισμού που εφαρμόζονται σε αυτές.

Έπειτα, γίνεται μία εισαγωγή στον τομέα της μηχανικής μάθησης και των νευρωνικών δικτύων. Αναφέρονται οι κυριότερες κατηγορίες νευρωνικών δικτύων, τα πλεονεκτήματα αυτών και οι κυριότερες αρχιτεκτονικές δομές στις οποίες αυτά συναντώνται. Ύστερα, τίθεται ένα θεωρητικό υπόβαθρο πάνω στην μηχανική μάθηση και γίνεται ιδιαίτερη αναφορά στο νευρωνικό δίκτυο που χρησιμοποιήθηκε για την ανάπτυξη της μεθόδου.

Τέλος, παρατίθεται αναλυτικά η πρόταση της διπλωματικής εργασίας για την ανάπτυξη της μεθόδου πρόβλεψης. Συγκεκριμένα, γίνεται εκτενής αναφορά στις παραμέτρους που απαιτούνται για την ανάπτυξη και την χρήση της μεθόδου και στις παραλλαγές που μελετήθηκαν στα πλαίσια της εργασίας. Η διπλωματική εργασία ολοκληρώνεται με την αναλυτική παρουσίαση των δεδομένων που χρησιμοποιήθηκαν για την ανάπτυξη και την αξιολόγηση της μεθόδου καθώς και των αποτελεσμάτων αυτής.

Λέξεις Κλειδιά: Τεχνικές Προβλέψεων, Μηχανική Μάθηση, Νευρωνικά Δίκτυα, Ιεραρχικές Δομές, Μέθοδοι Συμψηφισμού, Διακοπτόμενη Ζήτηση.

Abstract

The aim of this thesis is the development and the assessment of a forecasting method which is based on a machine learning algorithm and its purpose will be the production of forecasts for intermittent demand time series. More specifically, the method is dealing with time series that are subjected to hierarchical structures and is expected to combine the data from these time series in order to produce the final forecast. For this particular task a neural network was used alongside with the method of regression in order to process the data and to produce the forecasts.

First of all, a reference is made to the extent of work that has been done on the part of forecasting with neural networks. Moreover, the main forecasting forms and the widely used forecasting methods are presented. A thorough description of the hierarchical structures is made alongside with the main reconciliation methods that are used on these structures.

Afterwards, there is an introduction of the area of machine learning and neural networks. The main categories of neural networks, their advantages and the main architectural structures that they are formed in are being mentioned. Afterwards, a theoretical background on the machine learning area is set and there is a particular mention on the neural network that was used for the method development.

Finally, there is an analytical disclosure of the thesis proposal on the method development. More specifically, there is an extensive report on the variables that are required in order to develop and use the method and on the alternate forms of the method that were studied within the preparation of this thesis. The thesis is concluded with the analytical presentation of the data that was used for the development and the assessment of the method and of the results that it provided.

Keywords: Forecasting Methods, Machine Learning, Neural Networks, Hierarchical Structures, Reconciliation Methods, Intermittent Demand.

Πίνακας περιεχομένων

Ευχαριστίες.....	i
Περίληψη.....	iii
Abstract	v
Κεφάλαιο 1 . Εισαγωγή	1
1.1. Αντικείμενο της Εργασίας	1
1.2. Τεχνικές Προβλέψεων και Μηχανική Μάθηση	2
1.3. Δομή της Εργασίας	3
Κεφάλαιο 2 . Μέθοδοι Πρόβλεψης.....	5
2.1. Εισαγωγή	5
2.2. Αναπαράσταση Χρονοσειρών	5
2.3. Στατιστικές Προβλέψεις	6
2.3.1. Η Μέθοδος Naive	7
2.3.2. Μέθοδοι Μέσου Όρου	7
2.3.2.α. Απλός Μέσος Όρος	7
2.3.2.β. Κινητός Μέσος Όρος.....	7
2.3.3. Μέθοδοι Εκθετικής Εξομάλυνσης.....	8
2.3.3.α. Απλή Εκθετική Εξομάλυνση (Simple Exponential Smoothing - SES)	8
2.3.3.β. Μοντέλο Γραμμικής Τάσης (Holt Exponential Smoothing).....	10
2.3.3.γ. Μοντέλο Μη Γραμμικής Τάσης (Damped Exponential Smoothing).....	11
2.3.4. Μοντέλα Παλινδρόμησης	11
2.3.4.α. Απλή Γραμμική Παλινδρόμηση.....	12
2.3.4.β. Πολλαπλή Γραμμική Παλινδρόμηση	13
2.3.5. Η Μέθοδος Theta	13
2.3.6. Μοντέλα Διακοπτόμενης Ζήτησης.....	14
2.3.6.α. Η Μέθοδος Croston.....	15
2.3.6.β. Η Μέθοδος SBA	16
2.3.6.γ. Η Μέθοδος TSB.....	17
2.3.6.δ. Η Μέθοδος ADIDA.....	17
2.4. Κριτικές Προβλέψεις	19
2.5. Προβλέψεις Προϋπολογισμού	19
2.6. Αξιολόγηση Προβλέψεων	20
Κεφάλαιο 3 . Μέθοδοι Συμψηφισμού	23
3.1. Εισαγωγή	23
3.2. Η Μέθοδος Bottom-Up	26

3.3. Η Μέθοδος Top-Down.....	26
3.4. Optimal Reconciliation	27
Κεφάλαιο 4 . Νευρωνικά Δίκτυα και Μηχανική Μάθηση	31
4.1. Εισαγωγή	31
4.2. Νευρωνικά Δίκτυα.....	32
4.2.1.Πλεονεκτήματα των Νευρωνικών Δικτύων	32
4.2.2. Αρχιτεκτονικές Νευρωνικών Δικτύων	35
4.3. Μηχανική Μάθηση.....	38
4.3.1. Μάθηση με εκπαιδευτή	39
4.3.2. Μάθηση χωρίς εκπαιδευτή.....	40
4.3.2.α. Ενισχυτική Μάθηση.....	40
4.3.2.β. Μη Επιβλεπόμενη Μάθηση	41
4.3.3. Εργασίες Μάθησης.....	42
4.3.3.α. Συσχέτιση Προτύπων	42
4.3.3.β. Αναγνώριση Προτύπων	43
4.3.3.γ. Προσέγγιση Συνάρτησης.....	45
4.3.3.δ. Έλεγχος.....	47
4.4. Perceptron Πολλών Επιπέδων (Multilayer Perceptron)	47
4.4.1. Συνάρτηση Ενεργοποίησης	51
4.4.2. Ρυθμός Μάθησης	51
4.4.3.Κριτήρια Τερματισμού	51
Κεφάλαιο 5 . Προτεινόμενη Μεθοδολογία.....	53
5.1. Εισαγωγή	53
5.2. Benchmarking.....	56
5.3. Προετοιμασία Δεδομένων	58
5.4. Ανάπτυξη της Μεθόδου Πρόβλεψης	62
5.4.1. Παραλλαγές Μεθόδου	64
5.4.1.α. Χρήση Index	64
5.4.1.β. Χρήση ADI και CV ²	65
Κεφάλαιο 6 . Δεδομένα και Αποτελέσματα.....	70
6.1. Εισαγωγή	70
6.2. Πειραματική διαδικασία και αποτελέσματα	74
Κεφάλαιο 7 .Συμπεράσματα και Προεκτάσεις	88
7.1. Αποτίμηση των αποτελεσμάτων - Συμπεράσματα	88
7.2. Επίλογος	89

Βιβλιογραφία	92
Παράρτημα.....	94
Μέρος 1. Πίνακες Σφαλμάτων	94
Μέρος 2. Παρουσίαση Κώδικα	111

Περιεχόμενα Πινάκων

Πίνακας 2.1 Δείκτες Ακρίβειας.....	22
Πίνακας 5.1 Πίνακας Σφαλμάτων.....	58
Πίνακας 6.1 Πίνακας Σφαλμάτων της Μεθόδου SES.....	74
Πίνακας 6.2 Πίνακας Σφαλμάτων της Μεθόδου Croston.....	75
Πίνακας 6.3 Πίνακας Σφαλμάτων της Optimal Reconciliation Μεθόδου εφαρμοσμένης στα αποτελέσματα της Croston	75
Πίνακας 6.4 Πίνακας Σφαλμάτων της Μεθόδου SBA	75
Πίνακας 6.5 Πίνακας Σφαλμάτων της Μεθόδου Optimal Reconciliation εφαρμοσμένης στα αποτελέσματα της SBA.....	75
Πίνακας 6.6 Πίνακας Σφαλμάτων της Μεθόδου TSB.....	75
Πίνακας 6.7 Πίνακας Σφαλμάτων της Μεθόδου Optimal Reconciliation εφαρμοσμένης στα αποτελέσματα της TSB.....	76
Πίνακας 0.1 Πίνακας Σφαλμάτων (MLP) $W=9, NoO = 2$	94
Πίνακας 0.2 Πίνακας Σφαλμάτων (MLP) $W=6, NoO = 5$	94
Πίνακας 0.3 Πίνακας Σφαλμάτων (MLP) $W=15, NoO = 2$	94
Πίνακας 0.4 Πίνακας Σφαλμάτων (Index) $W=12, NoO = 10$	99
Πίνακας 0.5 Πίνακας Σφαλμάτων (Index) $W=6, NoO = 5$	99
Πίνακας 0.6 Πίνακας Σφαλμάτων (Index) $W=9, NoO = 5$	100
Πίνακας 0.7 Πίνακας Σφαλμάτων (ADI-CV ²) $W=6, NoO = 5$	105
Πίνακας 0.8 Πίνακας Σφαλμάτων (ADI-CV ²) $W=12, NoO = 2$	105
Πίνακας 0.9 Πίνακας Σφαλμάτων (ADI-CV ²) $W=15, NoO = 5$	105

Περιεχόμενα Σχημάτων

Σχήμα 3.1 Διάγραμμα Ιεραρχίας Τριών Επιπέδων	26
Σχήμα 4.1 Νευρωνικό Δίκτυο Ενός Επιπέδου.....	36
Σχήμα 4.2 Νευρωνικό Δίκτυο Πολλαπλών Επιπέδων.....	37
Σχήμα 4.3 Αναδρομικό Νευρωνικό Δίκτυο.....	38
Σχήμα 4.4	43
Σχήμα 4.5	45
Σχήμα 4.6 Perceptron Πολλών Επιπέδων.....	49
Σχήμα 5.1 Workflow Chart.....	54
Σχήμα 5.2 Εκπαίδευση του Νευρωνικού Δικτύου Πολλών Επιπέδων	55
Σχήμα 5.3 Πρόβλεψη με Νευρωνικό Δίκτυο Πολλών Επιπέδων	56
Σχήμα 5.4	59
Σχήμα 5.5	60
Σχήμα 5.6	61
Σχήμα 5.7	65
Σχήμα 5.8 Τιμές αποκοπής για την κατηγοριοποίηση των Syntetos, Boylan, Croston.....	67
Σχήμα 5.9	67
Σχήμα 6.1 Διάγραμμα της Ιεραρχίας των Δεδομένων	70
Σχήμα 6.2 Γραφική Παράσταση της Χρονοσειράς Νο 8 (Κατηγορία 1).....	71
Σχήμα 6.3 Γραφική Παράσταση της Χρονοσειράς Νο 87 (Κατηγορία 1).....	71
Σχήμα 6.4 Γραφική Παράσταση της Χρονοσειράς Νο 27 (Κατηγορία 2).....	72

Σχήμα 6.5 Γραφική Παράσταση της Συναθροισμένης Χρονοσειράς για την Κατηγορία 1	72
Σχήμα 6.6 Γραφική Παράσταση της Συναθροισμένης Χρονοσειράς για την Κατηγορία 2	73
Σχήμα 6.7 Γραφική Παράσταση της Συναθροισμένης Χρονοσειράς για τη Σύνολο των Δεδομένων	73
Σχήμα 6.8 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Top Level)	78
Σχήμα 6.9 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Bottom Level)	78
Σχήμα 6.10 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Top Level)	79
Σχήμα 6.11 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Bottom Level)	79
Σχήμα 6.12 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Top Level)	81
Σχήμα 6.13 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Bottom Level)	81
Σχήμα 6.14 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Top Level)	82
Σχήμα 6.15 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Bottom Level)	82
Σχήμα 6.16 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Top Level)	83
Σχήμα 6.17 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Bottom Level)	83
Σχήμα 6.18 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Top Level)	84
Σχήμα 6.19 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Bottom Level)	84
Σχήμα 6.20 Σφάλματα της μεθόδου κατά μέσο όρο σε σύγκριση με τα σφάλματα της TSB	85
Σχήμα 6.21 Σφάλματα της μεθόδου ($W=9, NoO=2$) σε σύγκριση με τα σφάλματα της TSB	85
Σχήμα 6.22 Σφάλματα της μεθόδου (Index) κατά μέσο όρο σε σύγκριση με τα σφάλματα της TSB	86
Σχήμα 6.23 Σφάλματα της μεθόδου (Index) ($W=12, NoO=10$) σε σύγκριση με τα σφάλματα της TSB	86
Σχήμα 6.24 Σφάλματα της μεθόδου (ADI-CV ²) κατά μέσο όρο σε σύγκριση με τα σφάλματα της TSB	87
Σχήμα 6.25 Σφάλματα της μεθόδου (ASI-CV ²) ($W=6, NoO=5$) σε σύγκριση με τα σφάλματα της TSB	87
Σχήμα 0.1 MSE/M ² Errors – Level 0 (MLP)	95
Σχήμα 0.2 MSE/M ² Errors – Level 1 (MLP)	95
Σχήμα 0.3 MSE/M ² Errors – Level 2 (MLP)	96
Σχήμα 0.4 MAE/M Errors – Level 0 (MLP)	96
Σχήμα 0.5 MAE/M Errors – Level 1 (MLP)	97
Σχήμα 0.6 MAE/M Errors – Level 2 (MLP)	97
Σχήμα 0.7 ME/M Errors – Level 0 (MLP)	98
Σχήμα 0.8 ME/M Errors – Level 1 (MLP)	98
Σχήμα 0.9 ME/M Errors – Level 2 (MLP)	99
Σχήμα 0.10 MSE/M ² Errors – Level 0 (Index)	100
Σχήμα 0.11 MSE/M ² Errors - Level 1 (Index)	101
Σχήμα 0.12 MSE/M ² Errors - Level 2 (Index)	101
Σχήμα 0.13 MAE/M Errors - Level 0 (Index)	102
Σχήμα 0.14 MAE/M Errors - Level 1 (Index)	102
Σχήμα 0.15 MAE/M Errors - Level 2 (Index)	103
Σχήμα 0.16 ME/M Errors - Level 0 (Index)	103
Σχήμα 0.17 ME/M Errors - Level 1 (Index)	104
Σχήμα 0.18 ME/M Errors - Level 2 (Index)	104
Σχήμα 0.19 MSE/M ² Errors - Level 0 (ADI-CV ²)	106
Σχήμα 0.20 MSE/M ² Errors - Level 1 (ADI-CV ²)	106
Σχήμα 0.21 MSE/M ² Errors - Level 2 (ADI-CV ²)	107
Σχήμα 0.22 MAE/M Errors - Level 0 (ADI-CV ²)	107
Σχήμα 0.23 MAE/M Errors - Level 1 (ADI-CV ²)	108
Σχήμα 0.24 MAE/M Errors - Level 2 (ADI-CV ²)	108
Σχήμα 0.25 ME/M Errors - Level 0 (ADI-CV ²)	109
Σχήμα 0.26 ME/M Errors - Level 1 (ADI-CV ²)	109
Σχήμα 0.27 ME/M Errors - Level 2 (ADI-CV ²)	110

Κεφάλαιο 1 . Εισαγωγή

1.1. Αντικείμενο της Εργασίας

Η ανάπτυξη στον τομέα των Τεχνικών Προβλέψεων τα τελευταία χρόνια είναι διαρκώς αυξανόμενη, κυρίως λόγω της ανάγκης παραγωγής αποδοτικών και λειτουργικών στατιστικών μοντέλων πρόβλεψης. Οι εμπειρικές προβλέψεις του παρελθόντος πλέον δρουν επικουρικά δίπλα σε στατιστικές μεθόδους που αναλύουν και επεξεργάζονται τα δεδομένα που διαθέτει κάποιος στη διάθεσή του. Το επόμενο βήμα στην παραγωγή των προβλέψεων αποτελούν οι νέες τεχνολογίες που αναπτύσσονται παράλληλα με την πρόοδο των επιστημών μέσω των οποίων δίνεται η δυνατότητα για περαιτέρω βελτίωση της αποδοτικότητας της παραγωγικής διαδικασίας της πρόβλεψης.

Μία τέτοια τεχνολογία αποτελεί η μηχανική μάθηση, η οποία χρησιμοποιείται ευρέως σε πάρα πολλούς τομείς και από ένα πολύ μεγάλο μέρος της επιστημονικής κοινότητας. Παρά την είσοδό της στην πλειοψηφία των επιστημών, η μηχανική μάθηση δεν έχει χρησιμοποιηθεί ευρέως στον τομέα των Προβλέψεων.

Η παρούσα διπλωματική εργασία αποτελεί μία προσέγγιση στο συνδυασμό της μηχανικής μάθησης με την παραγωγή προβλέψεων. Χρησιμοποιώντας ως εργαλείο τα νευρωνικά δίκτυα και τις δυνατότητές τους επιδιώκουμε την εφαρμογή αλγορίθμων μηχανικής μάθησης με σκοπό την αποδοτική παραγωγή προβλέψεων. Πιο συγκεκριμένα, αντικείμενο της εργασίας αποτελεί η ανάπτυξη μίας μεθοδολογίας που θα βασίζεται στην Παλινδρόμηση με νευρωνικό δίκτυο για την παραγωγή πρόβλεψης.

Τα δεδομένα τα οποία εξετάζουμε και χρησιμοποιούμε στην εργασία αυτή αποτελούν μέρος ιεραρχικής δομής. *Οι ιεραρχίες*, όπως αποκαλούνται, αποτελούν ένα μεγάλο κεφάλαιο των τεχνικών προβλέψεων και μία ξεχωριστή κατηγορία δεδομένων στην οποία αναπτύχθηκαν τα τελευταία χρόνια και ειδικές τεχνικές ή αλλιώς *μέθοδοι συμψηφισμού (reconciliation methods)* για την επεξεργασία των δεδομένων τους. Παρατηρείται ωστόσο ένα «κενό» στις τεχνικές αυτές, καθώς ο τρόπος με τον οποίον συνδυάζεται η πληροφορία είναι αποσπασματικός (χρησιμοποιείται πληροφορία μόνο από κάποια επίπεδα της ιεραρχίας), και καθαρά γραμμικός (συναθροίζουμε ή διαιρούμε ή επεξεργαζόμαστε γενικότερα με γραμμικές συναρτήσεις).

Στόχος, λοιπόν, της εργασίας είναι η μελέτη και η επεξεργασία τέτοιων δεδομένων, και συγκεκριμένα χρονοσειρών διακοπτόμενης ζήτησης που δομούν μία ιεραρχία με απώτερο σκοπό τον αποδοτικό συνδυασμό τους για την παραγωγή προβλέψεων. Σε αντίθεση με τις μεθόδους reconciliation, επιδιώκουμε την ανάπτυξη ενός μη γραμμικού συνδυασμού της πληροφορίας που έχουμε σε όλα τα επίπεδα μιας ιεραρχίας. Η μηχανική μάθηση και τα νευρωνικά δίκτυα αποτελούν χρήσιμα και αναντικατάστατα εργαλεία στην προσπάθεια αυτή.

1.2. Τεχνικές Προβλέψεων και Μηχανική Μάθηση

Όπως ήδη αναφέραμε, η μηχανική μάθηση δεν έχει ακόμα εισαχθεί σε μεγάλο βαθμό στον τομέα των προβλέψεων. Αυτό βέβαια δε σημαίνει ότι δεν έχουν γίνει κάποια πρώτα βήματα στη μελέτη της αποδοτικότητας μεθόδων πρόβλεψης που βασίζονται σε αλγορίθμους μηχανικής μάθησης.

Ξεκινώντας το 1989, οι Hornik, Stinchcombe και White, μελετώντας Νευρωνικά Δίκτυα πρόσθιας τροφοδότησης, δείχνουν ότι τα νευρωνικά δίκτυα αποτελούν γενικούς εκτιμητές οι οποίοι είναι επιρρεπείς να παρουσιάσουν το φαινόμενο που στην στατιστική αποκαλούμε "overfitting". Το φαινόμενο αυτό παρουσιάζεται όταν τα στοιχεία που δίνονται για ένα μοντέλο είναι περισσότερες από αυτές που μπορούν να «δικαιολογηθούν» από τα δοσμένα δεδομένα. Όταν δηλαδή έχουμε πολλές πληροφορίες αλλά όχι αρκετά δεδομένα που να τις επιβεβαιώνουν. Στο φαινόμενο αυτό έχουμε εξαιρετική λειτουργία του μοντέλου στα δεδομένα μας, αλλά δεν συμπεριφέρεται το ίδιο καλά όταν καλούμαστε να προβλέψουμε ή προσθέτουμε δεδομένα στο μοντέλο. Στα νευρωνικά δίκτυα αυτό μπορεί να αποφευχθεί με κατάλληλη προετοιμασία των δεδομένων στο κομμάτι της εκπαίδευσης του δικτύου.

Το 1998 οι Markham και Rakes μελέτησαν την ανάγκη των νευρωνικών δικτύων για δεδομένα για την παραγωγή προβλέψεων με μικρά σφάλματα και κατέληξαν ότι προκειμένου να παράγουν προβλέψεις που θα είναι πιο ακριβείς από τις προβλέψεις που παράγονται από συμβατικές μεθόδους πρόβλεψης απαιτούν ιδιαίτερα μεγάλο αριθμό δεδομένων.

Την ίδια χρονιά οι Zhang, G., Patuwo, B. E. και Hu, M. Y. κατέγραψαν μία λίστα με Νευρωνικά Δίκτυα που είχαν ήδη χρησιμοποιηθεί στον τομέα των προβλέψεων επιτυχώς. Τα Νευρωνικά Δίκτυα που κατέληξαν ότι χρησιμοποιούνται περισσότερο ήταν τα πολυεπίπεδα δίκτυα πρόσθιας τροφοδότησης (feedforward Multilayer Perceptron (MLP)). Στην εργασία τους συμπεριλαμβάνεται και μία λεπτομερής περιγραφή των μοντέλων αυτών και του τρόπου χρήσης τους στον τομέα των προβλέψεων. Ακόμα, οι Adya και Collory (1998) βρήκαν ότι το 73% των papers που περιλάμβαναν Νευρωνικά Δίκτυα ανέφεραν ότι τα Δίκτυα απέδιδαν καλύτερα αποτελέσματα από τα benchmarks που έθεταν με τις υπάρχουσες τεχνικές.

Το 2008 οι Gutierrez, R. S., Solis, A. O. και Mukhopadhyay, S. πρότειναν μία μέθοδο που χρησιμοποιεί Νευρωνικό Δίκτυο για να προβλέψει χρονοσειρές που χαρακτηρίζονται ως "lumpy". Χρησιμοποίησαν κλασικά Perceptron Πολλών Επιπέδων (MLP) με συγκεκριμένη αρχιτεκτονική και αριθμό εισόδων. Προτείνουν τη χρήση τριών κρυμμένων κόμβων (hidden nodes) σε ένα κρυφό επίπεδο (hidden layer) με δύο εισόδους. Η τελευταία μη μηδενική παρατήρηση αποτελεί την πρώτη είσοδο, ενώ η δεύτερη είναι ο αριθμός των περιόδων που παρεμβάλλονται μεταξύ των δύο τελευταίων μη μηδενικών τιμών της χρονοσειράς. Η έξοδος της μεθόδου αποτελεί την μελλοντική ζήτηση, εκφρασμένη με τρόπο ανάλογο αυτού της μεθόδου SES. Η μέθοδος αυτή μάλιστα, στα πειράματα τα οποία πραγματοποιήθηκαν, φάνηκε να αποδίδει καλύτερα από τη μέθοδο του Croston και την παραλλαγή αυτής από τους Syntetos και Boylan (2005).

Τέλος, το 2013 ο Kourentzes, N. προτείνει μία μέθοδο πρόβλεψης διακοπτόμενων χρονοσειρών με χρήση νευρωνικών δικτύων. Στο δημοσιευμένο paper του “Intermittent demand forecasts with neural networks” επικεντρώνεται σε χρονοσειρές διακοπτόμενης ζήτησης με περιορισμένο αριθμό δεδομένων και τροποποιεί την μέθοδο Croston με χρήση Νευρωνικών Δικτύων. Συγκεκριμένα, αφαιρεί το κομμάτι της SES από την κλασική μέθοδο Croston και εισάγει την πρόβλεψη των demands και intervals από νευρωνικό δίκτυο. Το νευρωνικό δίκτυο λαμβάνει ως εισόδους τα γνωστά demands και intervals και παράγει ως έξοδο μία πρόβλεψη για καθένα από αυτά ξεχωριστά.

Όπως γίνεται αντιληπτό, λοιπόν, η υπάρχουσα ενασχόληση με την παραγωγή προβλέψεων χρησιμοποιώντας νευρωνικά δίκτυα επικεντρώνεται είτε στην μετατροπή υπάρχουσών μεθόδων σε «υβριδικές», είτε στην παραγωγή πρόβλεψης χρησιμοποιώντας αποσπασματικές πληροφορίες από τις χρονοσειρές και όχι τις χρονοσειρές αυτές καθ' αυτές. Η καινοτομία της μεθόδου που προτείνεται από τη διπλωματική αυτή είναι ότι χρησιμοποιείται το σετ δεδομένων στην ολότητά του. Οι χρονοσειρές εισάγονται ως έχουν για την εκπαίδευση του νευρωνικού δικτύου (στην κατάλληλη μορφή βέβαια) και οποιαδήποτε αποκοπή πληροφορίας γίνεται για τη βελτιστοποίηση των αποτελεσμάτων.

1.3. Δομή της Εργασίας

Στο δεύτερο κεφάλαιο της παρούσας εργασίας γίνεται μία εκτενής αναφορά στις κατηγορίες προβλέψεων. Δίνονται οι ορισμοί τους και μετέπειτα γίνεται παράθεση και ανάλυση των κυριότερων στατιστικών μεθόδων που χρησιμοποιούνται, τόσο σε χρονοσειρές συνεχούς ζήτησης όσο και σε χρονοσειρές διακοπτόμενης ζήτησης. Επίσης, γίνεται μία σύντομη αλλά πλήρης αναφορά στις κύριες μεθόδους αξιολόγησης των προβλέψεων.

Τις τεχνικές προβλέψεων διαδέχονται οι τεχνικές συμφηφισμού ή αλλιώς reconciliation methods. Οι τεχνικές αυτές αναφέρονται στον συνδυασμό δεδομένων από χρονοσειρές μιας ιεραρχικής δομής για την βελτίωση των προβλέψεων που παρέχουν για αυτές οι στατιστικές μέθοδοι πρόβλεψης. Γίνεται αναφορά στη γενικότερη δομή των τεχνικών αυτών και επισκόπηση των κυριότερων μεθόδων – Bottom-Up και Top-Down – καθώς και στην τεχνική Optimal Reconciliation που εισηγήθηκαν το 2008 οι Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos και Han Lin Shang.

Στο τέταρτο κεφάλαιο αναφερόμαστε στις έννοιες της μηχανικής μάθησης και των νευρωνικών δικτύων. Δίνονται ορισμοί αυτών και τα κυριότερα χαρακτηριστικά τους. Αναφέρονται επιγραμματικά οι κατηγορίες νευρωνικών δικτύων και τα είδη μηχανικής μάθησης και γίνεται ιδιαίτερη αναφορά στο νευρωνικό δίκτυο που χρησιμοποιήθηκε για την υλοποίηση της μεθόδου.

Στο πέμπτο κεφάλαιο επεξηγείται αναλυτικά η πρόταση της διπλωματικής εργασίας. Αρχικά, αναφερόμαστε στο benchmarking, δηλαδή στην δημιουργία επιπέδων σύγκρισης τα οποία θα χρησιμοποιήσουμε έπειτα για την αξιολόγηση της μεθόδου. Δίνονται αναλυτικά οι παράμετροι που χρησιμοποιήθηκαν, ο αλγόριθμος

Κεφάλαιο 1. Εισαγωγή

που ακολουθήθηκε και ο τρόπος με τον οποίο αντιμετωπίζονται αρχικά προβλήματα της μεθόδου.

Στο κεφάλαιο 6 παραθέτουμε ένα Case Study της μεθόδου, δηλαδή μία πρακτική εφαρμογή της μεθόδου που αναπτύξαμε στο κεφάλαιο 5 σε πραγματικά δεδομένα με σκοπό την αξιολόγησή της. Γίνεται αναλυτική περιγραφή των δεδομένων που υπήρχαν στη διάθεσή μας, η δομή αυτών και ο τρόπος χρήσης και επεξεργασίας τους προκειμένου να καταλήξουμε στη μέθοδο πρόβλεψης.

Η διπλωματική εργασία ολοκληρώνεται στο κεφάλαιο 7 με σχολιασμό και αποτίμηση των αποτελεσμάτων με βάση την αποδοτικότητα της μεθόδου. Επίσης, σημειώνονται ορισμένα συμπεράσματα, αλλά και κάποιες σκέψεις για προέκταση της μεθοδολογίας προς μελλοντική μελέτη.

Κεφάλαιο 2 . Μέθοδοι Πρόβλεψης

2.1. Εισαγωγή

Το πρόβλημα της παραγωγής προβλέψεων απασχολεί εδώ και πολλά χρόνια την κοινωνία τόσο σε καθημερινό – επιχειρηματικό επίπεδο, όσο και σε επιστημονικό. Με την πάροδο των ετών έχουν αναπτυχθεί αρκετές θεωρίες ως προς την μεθοδολογία παραγωγής προβλέψεων αλλά και μέτρα σύγκρισης των τεχνικών προβλέψεων.

Ένας πρώτος τρόπος κατηγοριοποίησης των τεχνικών προβλέψεων είναι ο διαχωρισμός τους σε:

- Στατιστικές Προβλέψεις, στις οποίες εντάσσονται οι προβλέψεις που προκύπτουν μέσω στατιστικών μεθόδων ανάλυσης χρονοσειρών.
- Κριτικές προβλέψεις, στις οποίες εντάσσονται οι προβλέψεις που προκύπτουν μέσω εμπειρικής προσέγγισης του εκάστοτε ατόμου ή του συνόλου ατόμων (επιτροπή) που καλούνται να παράγουν την πρόβλεψη.
- Προβλέψεις Προϋπολογισμού, στις οποίες εντάσσονται προβλέψεις – στόχοι μιας εταιρίας η επιχείρησης και αποτελούν περισσότερο ιδανική προσέγγιση, παρά ρεαλιστική.

Οποιοσδήποτε συνδυασμός αυτών μπορεί να χαρακτηριστεί και «Τελική Πρόβλεψη».

2.2. Αναπαράσταση Χρονοσειρών

Έχοντας κατανοήσει την έννοια της χρονοσειράς, γίνεται εύκολα αντιληπτό ότι η αύξηση του μεγέθους των ιστορικών χρονολογικών παρατηρήσεων της χρονοσειράς, ή αλλιώς των ιστορικών δεδομένων της, συντελεί και σε αυξημένη δυσκολία κατανόησης και ανάλυσης της. Το φαινόμενο αυτό γίνεται εύκολα αντιληπτό παίρνοντας σαν παράδειγμα τις ιστορικές ανά λεπτό τιμές μίας ή περισσότερων μετοχών στο Χρηματιστήριο.

Είναι, λοιπόν, προφανές ότι η δισδιάστατη γραφική αναπαράστασή τους θα συνέβαλλε σημαντικά στο να είναι ευκολότερη η μελέτη των δεδομένων. Οι κυριότεροι τύποι γραφημάτων για την αναπαράσταση χρονοσειρών είναι οι ακόλουθοι:

- Διαγράμματα Χρόνου (Time Plots): Τα διαγράμματα χρόνου είναι το πλέον χρησιμοποιούμενο διάγραμμα, εφόσον απεικονίζει τις τιμές μίας μεταβλητής στην πάροδο του χρόνου. Λόγω της ευκολίας στην κατανόησης και αναπαράστασης των δεδομένων, τα διαγράμματα χρόνου είναι ιδιαίτερα διαδεδομένα στο χώρο των επιχειρήσεων (π.χ. για την αναπαράσταση των πωλήσεων ή των εξαγωγών/εισαγωγών) όπως επίσης και των Χρηματιστηρίων (π.χ. αναπαράσταση των ιστορικών τιμών μίας μετοχής) και της στατιστικής (π.χ. αναπαράσταση κοινωνικών, πολιτικών και

δημογραφικών δεδομένων). Μέσω της χρήσης τους είναι παράλληλα εφικτή η παρατήρηση βασικών χαρακτηριστικών των χρονοσειρών, όπως είναι για παράδειγμα η τάση (time trend) και η εποχικότητα (seasonality). Η αναπαράσταση αυτή επιλέχτηκε για την αναπαράσταση των δεδομένων στην εργασία.

- **Εποχιακά Διαγράμματα (Seasonal Plots):** Όπως προδίδει και το όνομα της κατηγορίας αυτής, τα εποχιακά διαγράμματα είναι ιδιαίτερα χρήσιμα για χρονοσειρές αυτές που εμφανίζουν ιδιαίτερη εποχικότητα. Για παράδειγμα, ένα χρονοδιάγραμμα θα μπορούσε να εμφανίζει τις πωλήσεις μια εταιρείας εποχιακών ειδών κατά τη διάρκεια ενός χρόνου.
- **Διαγράμματα Διασποράς (Scatter Plots):** Τα διαγράμματα διασποράς ενδείκνυνται στις περιπτώσεις που απαιτείται η σύγκριση μεταξύ δύο διαφορετικών μεγεθών ή χαρακτηριστικών που αφορούν τα δεδομένα αυτά. Για παράδειγμα θα μπορούσαν να χρησιμοποιηθούν για τη σύγκριση μεταξύ διαφορετικών προϊόντων ή υπηρεσιών μιας εταιρείας. Μέσω της οπτικοποίησης των ιστορικών δεδομένα διευκολύνεται ιδιαίτερα η εύρεση ακραίων και ιδιαίτερων τιμών καθώς και η αναγνώριση και ανάλυση των βασικών ποιοτικών χαρακτηριστικών μιας χρονοσειράς, τα οποία περιγράφονται στην υποενότητα που ακολουθεί.

2.3. Στατιστικές Προβλέψεις

Οι στατιστικές προβλέψεις αναφέρονται στην εφαρμογή στατιστικών μοντέλων χρονοσειρών ή αιτιοκρατικών μοντέλων επί μίας σειράς δεδομένων με σκοπό την αυτοματοποιημένη και συστηματική παραγωγή προβλέψεων. Οι στατιστικές προβλέψεις είναι άμεσα εφαρμόσιμες και αποδεκτά ακριβείς, αν συνδυαστούν με κατάλληλα διαστήματα εμπιστοσύνης. Επίσης, μπορούν να εφαρμοστούν μαζικά σε μεγάλο πλήθος χρονοσειρών (batch forecasting), απαιτώντας ελάχιστο χρόνο και λίγους υπολογιστικούς πόρους.

Πέρα από πολλά και σημαντικά πλεονεκτήματα των στατιστικών προβλέψεων, υπάρχουν ορισμένα προβλήματα στην εφαρμογή αυτών. Συγκεκριμένα, οι στατιστικές μέθοδοι προβλέψεων προϋποθέτουν ότι το πρότυπο (συμπεριφορά) της εκάστοτε χρονοσειράς θα συνεχιστεί στο μέλλον, γεγονός που δεν συμβαίνει πάντα. Επίσης, οι στατιστικές μέθοδοι δεν λαμβάνουν υπόψη ειδικά γεγονότα και ενέργειες που ενδέχεται να πραγματοποιηθούν στο άμεσο μέλλον. Τέλος, η συλλογή των δεδομένων μπορεί πολλές φορές να είναι μία διαδικασία δύσκολη, χρονοβόρα και ζημιογόνα, τη στιγμή που αρκετές στατιστικές μέθοδοι, προκειμένου να παράγουν προβλέψεις απαιτούν αρκετές παρατηρήσεις ιστορικών δεδομένων. Οι κυριότερες στατιστικές μέθοδοι πρόβλεψης παρουσιάζονται περιληπτικά στη συνέχεια:

2.3.1. Η Μέθοδος Naive

Η Μέθοδος Naive αποτελεί την πιο απλή στατιστική μέθοδο. Χρησιμοποιείται κυρίως ως benchmark για τις πιο πολύπλοκες μεθόδους και λιγότερο ως μέθοδος παραγωγής τελικής πρόβλεψης, καθώς οι προβλέψεις που παράγονται από αυτή δεν είναι ακριβείς. Η πρόβλεψη που προκύπτει από τη μέθοδο Naive για μία χρονική στιγμή t ισούται με την πραγματική παρατήρηση της προηγούμενης χρονικής περιόδου $(t-1)$. Δηλαδή:

$$F_t = Y_{t-1}$$

2.3.2. Μέθοδοι Μέσου Όρου

Μεθοδολογίες που εμπεριέχουν υπολογισμούς του μέσου όρου τμημάτων χρονοσειρών μπορούν να χρησιμοποιηθούν όχι μόνο στην εξομάλυνση τους αλλά και στην παραγωγή προβλέψεων.

2.3.2.α. Απλός Μέσος Όρος

Η πρόβλεψη χρονοσειρών με τη μέθοδο του απλού μέσου όρου βασίζεται σε ένα σύνολο των παρατηρήσεων του δείγματος, το μέγεθος του οποίου καθορίζεται στην αρχή της πρόβλεψης. Ο μέσος όρος των παρατηρήσεων θα χρησιμοποιηθεί ως πρόβλεψη για την επόμενη χρονική στιγμή $t + 1$. Η μέθοδος του απλού μέσου όρου περιγράφεται με την ακόλουθη εξίσωση:

$$F_{t+1} = \frac{1}{t} \sum_{i=1}^t Y_i$$

Λαμβάνοντας υπόψη το γεγονός ότι βασίζεται σε μεγάλο όγκο δεδομένων με ίσα βάρη (equal weighted), η μέθοδος του απλού μέσου όρου ενδείκνυται σε περιπτώσεις όπου οι χρονοσειρές έχουν σταθεροποιηθεί στην πάροδο του χρόνου και δεν χαρακτηρίζονται από εποχικότητα.

2.3.2.β. Κινητός Μέσος Όρος

Για τον υπολογισμό του κινητού μέσου όρου των δεδομένων μιας χρονοσειράς καθορίζεται αρχικά ο αριθμός των παρατηρήσεων που θα συμπεριληφθεί στον υπολογισμό αυτό. Η ονομασία κινητός μέσος όρος χρησιμοποιείται για να περιγράψει τη διαδικασία κατά την οποία καθώς μια νέα παρατήρηση γίνεται διαθέσιμη, ένας νέος μέσος όρος μπορεί να υπολογιστεί, στον οποίο παραλείπεται η πιο παλιά παρατήρηση προκειμένου να συμπεριληφθεί η πιο

πρόσφατη. Ο κινητός αυτός μέσος όρος αποτελεί την πρόβλεψη για την επόμενη περίοδο. Ο αριθμός των δεδομένων που χρησιμοποιούνται στον υπολογισμό του μέσου όρου παραμένει σταθερός και εμπεριέχει πάντα τις πιο πρόσφατες παρατηρήσεις της χρονοσειράς.

Έστω ότι η χρονοσειρά αποτελείται από n δεδομένα και ότι θα χρησιμοποιηθούν k παρατηρήσεις για τον υπολογισμό κάθε μέσου όρου, ο οποίος ονομάζεται κινητός μέσος όρος μήκους k ή ΚΜΟ(k) ή ΜΑ(k). Ο υπολογισμός της πρόβλεψης η οποία προκύπτει από την εφαρμογή ενός κινητού μέσου όρου μήκους k δίνεται από τον τύπο:

$$F_{t+1} = \frac{1}{k} \cdot \sum_{i=t-k+1}^t Y_i$$

2.3.3. Μέθοδοι Εκθετικής Εξομάλυνσης

Η Εκθετική Εξομάλυνση είναι μία μέθοδος πρόβλεψης η οποία προεκτείνει στοιχεία του προτύπου των ιστορικών δεδομένων, όπως τάσεις και εποχιακούς κύκλους, στο μέλλον. Οι προβλέψεις υπολογίζονται μετά από εξομάλυνση των δεδομένων, προκειμένου να απομονωθούν τα πραγματικά πρότυπα από τις καθαρά τυχαίες διακυμάνσεις. Οι μέθοδοι εκθετικής εξομάλυνσης έχουν ως βασική αρχή το γεγονός ότι, όσο πιο πρόσφατα είναι τα δεδομένα, τόσο περισσότερη πληροφορία εμπεριέχουν. Για το λόγο αυτό μεγαλύτερη βαρύτητα αποδίδεται στα πρόσφατα δεδομένα και φθίνει εκθετικά καθώς αναφερόμαστε σε δεδομένα που αντιστοιχούν σε παλαιότερες χρονικές περιόδους. Οι μέθοδοι εκθετικής εξομάλυνσης χρησιμοποιούνται ιδιαίτερα σε περιπτώσεις βραχυπρόθεσμου σχεδιασμού.

Τα μοντέλα εκθετικής εξομάλυνση κατηγοριοποιούνται ανάλογα με τη γενική μορφή της γραφικής παράστασης των ιστορικών δεδομένων με οριζόντιο άξονα τον χρόνο.

2.3.3.α. Απλή Εκθετική Εξομάλυνση (Simple Exponential Smoothing - SES)

Το μοντέλο σταθερού επιπέδου, που αναφέρεται επίσης και ως απλή εκθετική εξομάλυνση, περιγράφεται από τις εξής εξισώσεις:

$$\begin{aligned} e_t &= Y_t - F_t \\ S_t &= S_{t-1} + a \cdot e_t \\ F_{t+1} &= S_t \end{aligned}$$

Όπου e δηλώνει το σφάλμα, S δηλώνει το επίπεδο, F την πρόβλεψη και t τη χρονική περίοδο. Η παράμετρος a αποτελεί το συντελεστή εξομάλυνσης της μεθόδου και μπορεί να λάβει τιμές στο διάστημα $[0,1]$.

Κεφάλαιο 2. Μέθοδοι Πρόβλεψης

Είναι αναγκαίο, λόγω της αναδρομικότητας του μοντέλου, να οριστεί ένα αρχικό επίπεδο (S_0) προκειμένου να ξεκινήσει η διαδικασία υπολογισμού του μοντέλου πρόβλεψης αρχικά και των ζητούμενων σημειακών προβλέψεων δεδομένου χρονικού ορίζοντα. Ως αρχικό επίπεδο συνήθως χρησιμοποιείται:

- Ο Μέσος όρος όλων των παρατηρήσεων
- Ο Μέσος όρος των n πρώτων παρατηρήσεων
- Η πρώτη παρατήρηση
- Το σταθερό επίπεδο από το μοντέλο της απλής γραμμικής παλινδρόμησης

Η επιλογή μιας από τις εναλλακτικές έχει να κάνει με τα ποιοτικά χαρακτηριστικά της εκάστοτε χρονοσειράς. Αν το αρχικό επίπεδο δεν είναι αντιπροσωπευτικό των δεδομένων, θα έχει αρνητική επίδραση στις τιμές των προβλέψεων. Επιπλέον, το αρχικό επίπεδο επηρεάζει την επιλογή του συντελεστή εξομάλυνσης. Το αρχικό επίπεδο το οποίο θα επιλεγεί αποτελεί και την αρχική πρόβλεψη (F_1).

Ο βέλτιστος συντελεστής εξομάλυνσης καθορίζεται από δύο παράγοντες οι οποίοι αλληλεξαρτώνται. Ο ένας παράγοντας είναι το ποσοστό θορύβου στη χρονοσειρά. Όσο περισσότερος θόρυβος υπάρχει στα δεδομένα της χρονοσειράς, τόσο μικρότερη πρέπει να είναι η τιμή του συντελεστή εξομάλυνσης για να αποφύγουμε την υπερβολική αντίδραση στο θόρυβο. Ο άλλος παράγοντας είναι η σταθερότητα του μέσου όρου της χρονοσειράς. Αν ο μέσος όρος μεταβάλλεται, ο συντελεστής εξομάλυνσης θα πρέπει να είναι μεγάλος ώστε οι προβλέψεις να παρακολουθούν τις μεταβολές που παρουσιάζουν τα δεδομένα. Αντίθετα, αν ο μέσος όρος είναι σχετικά σταθερός, η τιμή του συντελεστή εξομάλυνσης θα είναι μικρή. Όταν η παράμετρος εξομάλυνσης λάβει τιμή κοντά στη μονάδα, τότε το σφάλμα της τελευταίας περιόδου διαδραματίζει σημαντικό ρόλο στον υπολογισμό της επόμενης σημειακής πρόβλεψης. Σε αντίθετη περίπτωση, η νέα πρόβλεψη θα μεταβληθεί σε πολύ μικρό βαθμό. Στην ακραία περίπτωση όπου $\alpha = 1$, το μοντέλο ταυτίζεται με την απλοϊκή μέθοδο Naive, ενώ στην περίπτωση που $\alpha = 0$ τότε κάθε πρόβλεψη είναι ίση με το αρχικό επίπεδο.

2.3.3.β. Μοντέλο Γραμμικής Τάσης (Holt Exponential Smoothing)

Το μοντέλο εξομάλυνσης για γραμμική τάση είναι μια επέκταση της απλής εκθετικής εξομάλυνσης η οποία μπορεί επιπρόσθετα να διαχειριστεί τη συνιστώσα της τάσης η οποία συχνά παρατηρείται στα επιχειρησιακά δεδομένα. Η επέκταση αυτή προτάθηκε από τον Holt το 1957 και για το λόγο αυτό η μέθοδος διατηρεί μέχρι σήμερα το όνομά του. Το μοντέλο εξομάλυνσης γραμμικής τάσης περιγράφεται από τις ακόλουθες εξισώσεις:

$$\begin{aligned} e_t &= Y_t - F_t \\ S_t &= S_{t-1} + T_{t-1} + a \cdot e_t \\ T_t &= T_{t-1} + a \cdot b \cdot e_t \\ F_{t+m} &= S_t + m \cdot T_t \end{aligned}$$

Όπου e δηλώνει το σφάλμα, S το επίπεδο, T την τάση, F την πρόβλεψη, t την χρονική περίοδο και m το χρονικό ορίζοντα της πρόβλεψης. Η παράμετρος a αποτελεί το συντελεστή εξομάλυνσης για το επίπεδο, ενώ η παράμετρος b είναι ο συντελεστής εξομάλυνσης της τάσης. Και οι δύο τιμές λαμβάνουν τιμές στο διάστημα $[0,1]$.

Όπως και στην απλή εκθετική εξομάλυνση είναι αναγκαίο να οριστεί ένα αρχικό επίπεδο (S_0) και μία αρχική τάση (T_0), προκειμένου να ξεκινήσει η διαδικασία υπολογισμού του μοντέλου πρόβλεψης αρχικά και των ζητούμενων σημειακών προβλέψεων δεδομένου χρονικού ορίζοντα, εν συνεχεία. Το αρχικό επίπεδο υπολογίζεται όπως και στην απλή εκθετική εξομάλυνση. Ως αρχική τάση, συνήθως χρησιμοποιείται:

- Διαφορά δεύτερης και πρώτης παρατήρησης ($Y_2 - Y_1$)
- Διαφορά n -στης και πρώτης παρατήρησης διαιρεμένης με $n-1$ (π.χ. $(Y_4 - Y_1)/3$)
- Η σταθερά της κλίσης από το μοντέλο της απλής γραμμικής παλινδρόμησης

Απαιτείται προσεκτική επιλογή τόσο του αρχικού επιπέδου όσο και της αρχικής τάσης, καθώς οι τιμές αυτές επηρεάζουν το μοντέλο πρόβλεψης και κατ'επέκταση τις σημειακές προβλέψεις. Η τελική επιλογή συνήθως γίνεται σύμφωνα με τα ποιοτικά χαρακτηριστικά των διαθέσιμων ιστορικών δεδομένων που θέλουμε να εφαρμόσουμε στο εκάστοτε μοντέλο. Στην περίπτωση μαζικής πρόβλεψης σε μεγάλο πλήθος χρονοσειρών (batch forecasting) συνίσταται η εφαρμογή γραμμικής παλινδρόμησης με ανεξάρτητη μεταβλητή το χρόνο.

2.3.3.γ. Μοντέλο Μη Γραμμικής Τάσης (Damped Exponential Smoothing)

Το μοντέλο μη γραμμικής τάσης εισάχθηκε από τους Gardner και McKenzie το 1985, αφότου είχε παρατηρηθεί ότι το μοντέλο γραμμικής τάσης πολλές φορές παρουσίαζε θετική προκατάληψη, ειδικά όταν εφαρμοζόταν με στόχο μεσοπρόθεσμες ή μακροπρόθεσμες προβλέψεις. Το μοντέλο περιγράφεται από τις εξής σχέσεις:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + T_{t-1} + a \cdot e_t$$

$$T_t = T_{t-1} + a \cdot b \cdot e_t$$

$$F_{t+m} = S_t + \sum_{i=1}^m \varphi^i \cdot T_t$$

Παρατηρούμε ότι η αλλαγή σε σχέση με το μοντέλο γραμμικής τάσης (Holt) εντοπίζεται στη τελευταία εξίσωση που, αντί να πολλαπλασιάζεται η τελευταία τάση που εντοπίσαμε με τη χρονική διαφορά σε περιόδους της τιμής προς πρόβλεψη από την τελευταία τιμή των δεδομένων μας, έχουμε την τελευταία τάση να πολλαπλασιάζεται με το άθροισμα των στοιχείων γεωμετρικής προόδου με λόγο ίσο με φ και μήκος m . Η παράμετρος φ , που ονομάζεται παράμετρος διόρθωσης τάσης, καθορίζει το μοντέλο ως εξής:

- Για $\varphi = 0$, το μοντέλο ταυτίζεται με αυτό της απλής εκθετικής εξομάλυνσης
- Για $0 < \varphi < 1$, το μοντέλο χαρακτηρίζεται από φθίνουσα τάση
- Για $\varphi = 1$, το μοντέλο ταυτίζεται με το μοντέλο γραμμικής τάσης
- Για $\varphi > 1$, το μοντέλο περιγράφει εκθετική εξομάλυνση με εκθετική τάση

Για τον προσδιορισμό του αρχικού επιπέδου και της αρχικής τάσης, χρησιμοποιούμε τις μεθόδους που περιεγράφηκαν στις προηγούμενες ενότητες.

2.3.4. Μοντέλα Παλινδρόμησης

Η ανάλυση της παλινδρόμησης αποσκοπεί στην εύρεση συσχετίσεων μεταξύ μιας εξαρτημένης μεταβλητής και μίας ή περισσότερων ανεξάρτητων μεταβλητών. Έχει ευρεία χρήση στη διαδικασία των προβλέψεων, τόσο ως μοντέλο πρόβλεψης αλλά και ως βοήθημα σε άλλες μεθόδους, όπως θα δούμε παρακάτω. Κυρίως, όμως μας επιτρέπει να βγάλουμε συμπεράσματα για τη συσχέτιση της ανεξάρτητης μεταβλητής και των εξαρτημένων μεταβλητών.

2.3.4.α. Απλή Γραμμική Παλινδρόμηση

Η μέθοδος, που φέρει και το όνομα μέθοδος των ελαχίστων τετραγώνων, περιγράφει μία ευθεία με την ελάχιστη απόσταση ανά σημείο από τα πραγματικά δεδομένα. Για να τη λάβουμε πρέπει να ελαχιστοποιηθεί το άθροισμα σφαλμάτων στη δεύτερη εξίσωση που ακολουθεί. Επίσης, φαίνεται αναλυτικά πως προκύπτουν και οι συντελεστές:

$$\hat{Y}_i = \alpha + \beta \cdot X_i$$

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

$$\beta = \frac{\frac{\sum X_i Y_i}{n} - \bar{X}\bar{Y}}{\frac{\sum X_i^2}{n} - \bar{X}^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$a = \bar{Y} - \beta\bar{X}$$

Για να χρησιμοποιήσουμε τη μέθοδο της Απλής Γραμμικής Παλινδρόμησης, πρέπει να υπάρχει εξάρτηση της μεταβλητής πρόβλεψης από τη τιμή ή τη μεταβολή κάποιας άλλης. Για να ελέγξουμε αν αυτό συμβαίνει, χρησιμοποιούμε τον συντελεστή γραμμικής συσχέτισης r που αποτελεί έναν δείκτη του βαθμού που συσχετίζονται δύο μεταβλητές μεταξύ τους και για δύο μεταβλητές X, Y προκύπτει ως εξής:

$$Cov_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

$$Cov_{XX} = \frac{\sum (X_i - \bar{X})^2}{n} = Var_X = S_X^2$$

$$Cov_{YY} = \frac{\sum (Y_i - \bar{Y})^2}{n} = Var_Y = S_Y^2$$

$$r_{XY} = \frac{Cov_{XY}}{\sqrt{Cov_{YY}Cov_{XX}}} = \frac{Cov_{XY}}{S_Y S_X}$$

Προκύπτει ότι ο συντελεστής r_{XY} λαμβάνει τιμές στο διάστημα $[-1, 1]$. Ο συντελεστής γραμμικής συσχέτισης ερμηνεύεται με δύο τρόπους:

- Μας δείχνει την κατεύθυνση της σχέσης μεταξύ των δύο μεταβλητών, δηλαδή αν όταν οι τιμές της μίας αυξάνονται, οι τιμές της άλλης μειώνονται οι αυξάνονται. Επίσης, μπορεί να μας δείξει ότι η μεταβολές της μίας είναι ανεξάρτητες από τις άλλης.

Κεφάλαιο 2. Μέθοδοι Πρόβλεψης

- Μας δείχνει τον βαθμό συσχέτισης και συνεπώς τη δυνατότητα της γραμμής παλινδρόμησης να εκφράσει τη σχέση μεταξύ των μεταβλητών. Όσο το r_{xy} πλησιάζει κατά απόλυτη τιμή τη μονάδα τόσο μικρότερη είναι η απόκλιση των πραγματικών τιμών της εξαρτημένης μεταβλητής από αυτές του μοντέλου.

Η μέθοδος των ελάχιστων τετραγώνων χρησιμοποιείται ως μοντέλο πρόβλεψης χρονοσειρών, απλά θέτοντας ως ανεξάρτητη μεταβλητή το χρόνο. Η γραμμή προκύπτει όπως περιγράψαμε προηγουμένως από τα ιστορικά δεδομένα, και προεκτείνοντας τη στο χρόνο λαμβάνουμε τις τιμές της πρόβλεψης για τις χρονικές στιγμές που μας ενδιαφέρουν.

2.3.4.β. Πολλαπλή Γραμμική Παλινδρόμηση

Υπάρχουν περιπτώσεις που έχουμε πληροφορία για περισσότερες ανεξάρτητες μεταβλητές, τότε χρησιμοποιούμε τη μέθοδο της πολλαπλής παλινδρόμησης, η παίρνει την εξής μορφή:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_k \cdot X_k + e$$

Αντίστοιχα με πριν, πρέπει να ελαχιστοποιηθεί το τετραγωνικό σφάλμα. Για να βρούμε τις τιμές των συντελεστών που μας οδηγούν στο επιθυμητό αποτέλεσμα, προσδιορίζουμε τις τιμές των μερικών παραγώγων της συνάρτησης του σφάλματος για κάθε συντελεστή. Κατόπιν λύνουμε το γραμμικό σύστημα που προκύπτει αν θέσουμε τη κάθε μερική παράγωγο ίση με το μηδέν.

2.3.5. Η Μέθοδος Theta

Η μέθοδος Theta είναι μια μονοδιάστατη μέθοδος πρόβλεψης που εισάχθηκε από τον Ασημακόπουλο το 2000. Κατά το μοντέλο, η χρονοσειρά αναλύεται σε δύο ή περισσότερες χρονοσειρές ή αλλιώς γραμμές Theta. Κάθε μία από τις προκύπτουσες σειρές προβλέπεται ξεχωριστά, είτε με το ίδιο είτε με διαφορετικό μοντέλο πρόβλεψης, και η τελική πρόβλεψη είναι ο συνδυασμός των επιμέρους αποτελεσμάτων.

Στην απλούστερη περίπτωση, το κλασσικό μοντέλο Θ , έχουμε δύο γραμμές Theta. Η πρώτη είναι μία ευθεία γραμμή που προκύπτει από την ευθεία γραμμικής παλινδρόμησης που μοντελοποιεί τα δεδομένα, η Theta Line (0). Η δεύτερη, Theta Line (2) προκύπτει ως εξής:

$$ThetaLine(2) = 2 \cdot Y - LRL$$

Όπου Y η αρχική χρονοσειρά και LRL η ευθεία που προκύπτει από τη μέθοδο ελάχιστων τετραγώνων. Η μέθοδος Theta βασίζεται στην μεταβολή των τοπικών

καμπυλοτήτων μιας χρονοσειράς. Η μεταβολή αυτή λαμβάνει χώρα με τη χρήση της παραμέτρου θ που εφαρμόζεται στις δεύτερες διαφορές τις χρονοσειράς ως εξής:

$$Y_t^\theta = \theta \cdot Y''_t$$

$$Y''_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

Διακρίνουμε διαφορετικές συμπεριφορές των γραμμών θ ανάλογα με τις τιμές τις παραμέτρου:

- Αν $\theta = 0$ η σειρά ταυτίζεται με την ευθεία γραμμικής παλινδρόμησης
- Αν $\theta = -1$ η σειρά που προκύπτει είναι η συμμετρική της αρχικής ως προς την ευθεία γραμμικής παλινδρόμησης
- Αν $\theta > 1$ έχουμε πιο έντονες καμπυλότητες, ανάλογες του βαθμού της παραμέτρου

Το τελικό μοντέλο πρόβλεψης προκύπτει ως γραμμικός συνδυασμός των γραμμών θ , έτσι ώστε τα βάρη της κάθε σειράς να αθροίζουν στο 1. Στην κλασσική μέθοδο, όπως είδαμε στην εξίσωση υπολογισμού της Theta Line (2), έχουμε πάρει τα βάρη ίσα με 0.5 και για τις δύο γραμμές.

2.3.6. Μοντέλα Διακοπτόμενης Ζήτησης

Η διακοπτόμενη ζήτηση έχει την χαρακτηριστική ιδιότητα να εμφανίζεται σποραδικά, δηλαδή να μεσολαβούν περίοδοι όπου αυτή είναι μηδενική. Αυτό είναι και το χαρακτηριστικό της που την κάνει ιδιαίτερη ως περίπτωση και χρήζει διαφορετικής αντιμετώπισης. Όταν εμφανιστεί ζήτηση, τότε το μέγεθος της ποικίλει και σε πολλές περιπτώσεις σε σημαντικό βαθμό. Η διακοπτόμενη ζήτηση συναντάται σε μονάδες διαχείρισης αποθεμάτων (stock keeping units) και ανταλλακτικών (spare parts). Η διακοπτόμενη φύση των δεδομένων αυτών δημιουργεί σημαντικά προβλήματα τόσο στον έλεγχο των αποθεμάτων όσο και στη διαδικασία της πρόβλεψης. Η δυσκολία της πρόβλεψης δεν έγκειται μόνο στην ασυνέχεια των παρατηρήσεων αλλά και στη μεγάλη διακύμανση μεταξύ δύο μη μηδενικών παρατηρήσεων.

Στην πράξη, οι μέθοδοι της εκθετικής εξομάλυνσης χρησιμοποιούνται συχνά όταν απαιτείται χειρισμός και πρόβλεψη δεδομένων διακοπτόμενης φύσης. Καθώς όμως οι μέθοδοι αυτοί αποδίδουν μεγαλύτερο βάρος στα πιο πρόσφατα δεδομένα, καταλήγουν σε ένα μοντέλο πρόβλεψης όπου οι εκτιμήσεις είναι μέγιστες έπειτα από μία εμφάνιση ζήτησης και ελάχιστες ακριβώς πριν από περίοδο μη μηδενικής ζήτησης. Η αδυναμία λοιπόν, των μοντέλων εξομάλυνσης οδήγησε τους ερευνητές στην αναζήτηση νέων μεθόδων και τεχνικών για την αποτελεσματική πρόβλεψη χρονοσειρών με έντονη παρουσία μηδενικών τιμών.

Η πρόβλεψη χρονοσειρών διακοπτόμενης ζήτησης αποτελεί ξεχωριστό κομμάτι στις μεθόδους πρόβλεψης, λόγω των ιδιομορφιών που παρουσιάζουν οι χρονοσειρές αυτές. Πιο συγκεκριμένα, η διακοπτόμενη ζήτηση (intermittent demand) εμφανίζει τα εξής χαρακτηριστικά σε μία χρονοσειρά:

- Η χρονοσειρά εμπεριέχει πολλαπλές χρονικές στιγμές στις οποίες το μέγεθος που απεικονίζεται λαμβάνει την τιμή μηδέν.
- Η χρονοσειρά δεν παρουσιάζει εποχιακή συμπεριφορά.

2.3.6.α. Η Μέθοδος Croston

Η ακανόνιστη και κυματοειδής φύση των δεδομένων διακοπτόμενης ζήτησης είναι η κύρια αιτία δυσκολίας εξαγωγής των προβλέψεων και ο λόγος ενασχόλησης της κοινότητας των προβλέψεων με αυτά. Όπως έχει ήδη αναφερθεί, η μέθοδος που εν γένει χρησιμοποιείται για την πρόβλεψη για τέτοια δεδομένα ήταν η μέθοδος εκθετικής εξομάλυνσης. Το 1972, ο Croston πρότεινε μία εναλλακτική μέθοδο, η οποία λαμβάνει υπόψη τόσο το μέγεθος της ζήτησης όσο και το χρόνο μεταξύ των ζητήσεων. Στην πράξη ο Croston εξήγαγε προβλέψεις εφαρμόζοντας ανεξάρτητα απλή εκθετική εξομάλυνση τόσο στις μη μηδενικές τιμές των χρονοσειρών όσο και στα χρονικά διαστήματα μεταξύ των μη μηδενικών τιμών των χρονοσειρών.

Πιο αναλυτικά, η μέθοδος που εισήγαγε ο Croston, διαχωρίζει αρχικά την χρονοσειρά των πραγματικών παρατηρήσεων σε δύο επιμέρους, όπου η μία αποτελείται από τα χρονικά διαστήματα μεταξύ των μη μηδενικών ζητήσεων (intervals) και η άλλη από το πλήθος των ανεξάρτητων μη μηδενικών ζητήσεων (demands). Οι δύο χρονοσειρές προεκτείνονται ανεξάρτητα με χρήση της μεθόδου εκθετικής εξομάλυνσης σταθερού επιπέδου. Ως παράμετρος εξομάλυνσης στην εφαρμογή της μεθόδου είναι συνήθης στη βιβλιογραφία η χρήση της τιμής $\alpha=0.05$. Η προέκταση των χρονοσειρών ακολουθείται από τον υπολογισμό της πρόβλεψης Croston μέσω εύρεσης του ηθλικού των δύο ανεξάρτητων προβλέψεων ως εξής:

$$F_{Croston} = \frac{F_{demands}}{F_{intervals}}$$

Γενικά, η διακοπτόμενη ζήτηση μπορεί να οδηγήσει σε αύξηση του επιπέδου προμήθειας αποθεμάτων επιφέροντας προκατάληψη στις εκτιμήσεις της μέσης ζήτησης. Ο Croston όμως υποστηρίζει πως η προσέγγιση των ξεχωριστών εκτιμήσεων για ζήτηση και χρόνο μεταξύ ζητήσεων, οδηγεί σε μείωση της προκατάληψης. Επίσης, η πρόσθετη συνιστώσα της συχνότητας μεταξύ των ζητήσεων επιτρέπει στον διαχειριστή των αποθεμάτων να καθορίσει τις παραγγελίες και το κόστος με μεγαλύτερη ακρίβεια, αποφεύγοντας της περίσσεια αποθεμάτων. Η μέθοδος έχει το πλεονέκτημα ότι μπορεί να χρησιμοποιηθεί με τον ίδιο ακριβώς τρόπο και για δεδομένα συνεχούς ζήτησης, αφού πλέον συμπίπτει με απλή εφαρμογή της εκθετικής εξομάλυνσης σταθερού επιπέδου. Στην περίπτωση όμως δεδομένων διακοπτόμενης ζήτησης ουσιαστικά αποτελεί μια παραλλαγή της απλής εκθετικής εξομάλυνσης, με τις προβλέψεις να παραμένουν σταθερές και αμετάβλητες μεταξύ των μη μηδενικών ζητήσεων και να ανανεώνονται μόνο μετά την παρατήρηση νέας μη μηδενικής τιμής.

Η μέθοδος Croston εφαρμόζεται κατά κόρον με τον τρόπο που έχει ήδη περιγραφεί. Παρόλα αυτά έχουν προταθεί κατά καιρούς διάφορες παραλλαγές

αυτής, όσον αφορά τη μέθοδο που θα χρησιμοποιηθεί για την πρόβλεψη. Έχει προταθεί για παράδειγμα η χρήση της μεθόδου Theta έναντι της εκθετικής εξομάλυνσης σταθερού επιπέδου (Nikolopoulos et al., 2007). Επίσης, έχει ελεγχθεί σε μεγάλο πλήθος χρονοσειρών η χρήση βελτιστοποιημένης παραμέτρου εκθετικής εξομάλυνσης έναντι σταθερής, ίσης με $\alpha=0.05$, δίνοντας ικανοποιητικά αποτελέσματα (Petroopoulos et al., 2008).

2.3.6.β. Η Μέθοδος SBA

Η μέθοδος Croston έχει αποδειχθεί μεγίστης αξίας για τους κατασκευαστές που αντιμετωπίζουν δεδομένα διακοπτόμενης φύσης. Παρά όμως τη σπουδαία θεωρητική ανωτερότητα αυτής της μεθόδου πρόβλεψης, οι εμπειρικές ενδείξεις δεν συμφωνούν με την εφαρμογή της, καθώς τα κέρδη από την αντίστοιχη εφαρμογή είναι χειρότερα του αναμενόμενου, συγκρινόμενη με απλούστερες τεχνικές πρόβλεψης. Μετά από αυτήν τη διαπίστωση, οι Syntetos και Boylan (2001), προσπαθώντας να εντοπίσουν το αίτιο αυτής της απροσδόκητης συμπεριφοράς, διαπίστωσαν ότι η μέθοδος Croston είναι μια θετικά προκατειλημμένη μέθοδος (positive biased), δηλαδή παρουσιάζει μια αισιόδοξη τάση στα αποτελέσματα των προβλέψεων της. Ένα άλλο σημαντικό επίσης επίτευγμα της έρευνας τους ήταν η σύνδεση του επιπέδου της αισιόδοξης τάσης της μεθόδου με την τιμή της παραμέτρου εξομάλυνσης α που χρησιμοποιείται για την προέκταση των δύο αποσυντιθέμενων χρονοσειρών. Η μέγιστη προκατάληψη μάλιστα παρατηρήθηκε όταν η παράμετρος εξομάλυνσης α λάβει τη μέγιστη τιμή της δηλαδή $\alpha=1$. Γενικά παρατηρήθηκε μία αναλογία της θετικής προκατάληψης της μεθόδου με τις μεγάλες τιμές της παραμέτρου α , και η μέθοδος Croston να ενδείκνυται μόνο όταν το α είναι μικρότερο του 0.15.

Η μέθοδος λοιπόν που αναπτύχθηκε, πήρε το όνομά της από τους προαναφερόμενους, ονομάζεται SBA (Syntetos and Boylan Approximation) και αποτελεί μία τροποποίηση της Croston, στην οποία η πρόβλεψη υπολογίζεται από τον ακόλουθο μαθηματικό τύπο:

$$F_{SBA} = \left(1 - \frac{\alpha}{2}\right) \cdot \frac{F_{demands}}{F_{intervals}} = \left(1 - \frac{\alpha}{2}\right) \cdot F_{Croston}$$

Μεταγενέστερες εμπειρικές μελέτες και μελέτες προσομοίωσης (Syntetos & Boylan, 2005; Teunter & Sani, 2009) έδειξαν πως η αρχική μέθοδος Croston παρουσιάζει μικρότερη θετική προκατάληψη εάν υπάρχουν λίγες μη μηδενικές τιμές στη ζήτηση ενώ η τροποποιημένη μέθοδος των Syntetos και Boylan έχει μικρότερη προκατάληψη εάν παρατηρούνται περισσότερες μηδενικές τιμές στη ζήτηση.

2.3.6.γ. Η Μέθοδος TSB

Μία ακόμα μέθοδος πρόβλεψης διακοπτόμενης ζήτησης αναπτύχθηκε το 2011 από τους Teunter, Syntetos και Babai λαμβάνοντας το όνομά της από τα αρχικά τους. Η μέθοδος TSB αντικαθιστά το demand interval της μεθόδου Croston με demand probability, δηλαδή την πιθανότητα ύπαρξης ζήτησης σε κάθε χρονική περίοδο. Η ανάγκη για τη διαφοροποίηση αυτή προκύπτει καθώς στη μέθοδο Croston η ζήτηση ανανεώνεται μόνο όταν προκύπτει τιμή διάφορη του μηδενός, ενώ σε καταστάσεις και χρονοσειρές σε πρακτικό επίπεδο παρατηρείται συχνά η ύπαρξη πολλών συνεχόμενων μηδενικών τιμών, γεγονός που σηματοδοτεί ότι η τελική πρόβλεψη δεν δύναται να υπολογιστεί λόγω της αδυναμίας υπολογισμού σωστού επιπέδου καθώς έχει περάσει μεγάλο διάστημα από την τελευταία μη μηδενική παρατήρηση.

Η μέθοδος TSB περιγράφεται με τις εξής εξισώσεις:

Αν $D_t = 0$, τότε:

$$\begin{aligned} z'_t &= z'_{t-1} \\ D'_t &= D'_{t-1} + \beta(0 - D'_{t-1}) \end{aligned}$$

Διαφορετικά:

$$\begin{aligned} z'_t &= z'_{t-1} + \alpha(z_t - z'_{t-1}) \\ D'_t &= D'_{t-1} + \beta(1 - D'_{t-1}) \\ Y'_t &= D'_t z'_t \end{aligned}$$

Όπου, το D_t αποτελεί δείκτη ύπαρξης τιμής για τη χρονική περίοδο t , το Y'_t αναπαριστά τη μέση ζήτηση της περιόδου t , το z_t την αληθινή ζήτηση τη χρονική περίοδο t , το z'_t δηλώνει το χρόνο μεταξύ δύο μη μηδενικών τιμών της χρονοσειράς, το D'_t αποτελεί τη διαφοροποίηση που αναφέραμε παραπάνω καθώς εκφράζει την πιθανότητα να υπάρχει ζήτηση στο τέλος της χρονικής περιόδου t και, τέλος, τα α και β αποτελούν σταθερές εξομάλυνσης και λαμβάνουν τιμές στο διάστημα $[0,1]$.

2.3.6.δ. Η Μέθοδος ADIDA

Μία άλλη διαδικασία αντιμετώπισης των χρονοσειρών διακοπτόμενης ζήτησης που μπορεί να ακολουθηθεί προκειμένου να αποφεύγεται η ασυνέχεια των δεδομένων όσον αφορά τις μηδενικές τιμές, είναι η μη επικαλυπτόμενη συνάθροιση (aggregation) των δεδομένων σε περιόδους μικρότερης συχνότητας. Αν για παράδειγμα υπάρχουν μηνιαία δεδομένα, και σε κάποιους μήνες η ζήτηση ήταν μηδενική τότε θα μπορούσε να εφαρμοστεί η συνάθροιση σε τριμηνιαίο επίπεδο, θέτοντας το επίπεδο συνάθροισης ίσο με τρεις περιόδους. Με μία τέτοια συνάρτηση, ενδέχεται να μειωθεί, αν όχι επαλειφθεί, η ασυνέχεια λόγω ύπαρξης μηδενικών παρατηρήσεων. Επίσης, η διακύμανση της προκύπτουσας νέας χρονοσειράς αναμένεται να είναι μικρότερη, έχοντας χρησιμοποιήσει ουσιαστικά μη επικαλυπτόμενο κινητό μέσο όρο για εξομάλυνση. Ο κατάλληλος καθορισμός του επιπέδου συνάθροισης (aggregation level) θα οδηγήσει σε χρονοσειρά συνεχούς ζήτησης, χωρίς μηδενικές τιμές, στην οποία θα μπορεί πλέον να εφαρμοστεί οποιαδήποτε τεχνική πρόβλεψης δεδομένων κατάλληλη για συνεχή ζήτηση για την

Κεφάλαιο 2. Μέθοδοι Πρόβλεψης

παραγωγή προβλέψεων στο επίπεδο συνάθροισης. Εντούτοις, η χρησιμότητα των εξαγομένων συναθροισμένων προβλέψεων μπορεί να αμφισβητηθεί, καθώς τα δεδομένα παρακολουθούνται και ελέγχονται σε υψηλότερη χρονική συχνότητα, με σκοπό την εξυπηρέτηση διαφόρων λειτουργικών αναγκών. Κρίνεται λοιπόν αναγκαία η διάσπαση (disaggregation) των υπολογισμένων προβλέψεων ώστε να υπολογιστούν οι τελικές προβλέψεις χρονικής συχνότητας ίσης με του αρχικού επιπέδου.

Όντας αυτή η θεμελιώδης σκέψη, παρήχθη μία νέα μέθοδος η οποία ονομάζεται ADIDA (Nikolopoulos et al., 2010) και η οποία προτείνει τη συνάθροιση δεδομένων σε ένα υψηλότερο χρονικά επίπεδο όπου έχει αποφευχθεί η ασυνέχεια των δεδομένων. Σε αυτήν μπορούν πλέον να εφαρμοστούν μέθοδοι πρόβλεψης για δεδομένα συνεχούς ζήτησης πέραν των μεθόδων που εφαρμόζονται ως επί τω πλείστω σε δεδομένα διακοπτόμενης ζήτησης. Ύστερα από το στάδιο παραγωγής των προβλέψεων, απαιτείται ο διαχωρισμός των προβλέψεων στα αρχικά επίπεδα των χρονοσειρών, χρησιμοποιώντας εμπειρικές τεχνικές.

Δύο βασικές έννοιες χρησιμοποιεί η μέθοδος ADIDA οι οποίες επηρεάζουν ιδιαίτερα τα αποτελέσματα της: το επίπεδο συνάθροισης και η μέθοδος διαχωρισμού. Και οι δύο αυτές παράμετροι έχουν μεγάλη επίδραση στα αποτελέσματα της συγκεκριμένης μεθόδου και αυτός είναι και ο λόγος που αποτελούν πεδίο έρευνας και μελέτης.

Η κατάλληλη επιλογή του επιπέδου συνάθροισης είναι πολύ σημαντική διαδικασία στην ορθή επιλογή του μοντέλου ADIDA. Η δυσκολία στην επιλογή του επιπέδου συνάθροισης έγκειται στο ότι από την μία πλευρά το επίπεδο συνάθροισης πρέπει να είναι ικανοποιητικά μεγάλο έτσι ώστε να εξαλείφεται πλήρως το φαινόμενο της διακοπτόμενης ζήτησης ενώ από την άλλη, δεν πρέπει να οδηγεί σε μία χρονοσειρά εξαιρετικά λίγων παρατηρήσεων όπου πλέον μεγάλο πλήθος πληροφορίας έχει πια χαθεί. Σε ένα σύστημα διαχείρισης αποθεμάτων θα είχε απόλυτο νόημα ο ορισμός του επιπέδου συνάθροισης ως τον αριθμό των περιόδων που χρειάζεται κάθε SKU (Stock Keeping Unit) από τη στιγμή που θα γίνει η παραγγελία νέας παρτίδας μέχρι τη στιγμή που θα γίνουν αυτά διαθέσιμα (lead time) συν μία περίοδο. Ο Νικολόπουλος έδειξε ότι ο ορισμός αυτός επιφέρει στατιστικά σημαντική μείωση των δεικτών σφάλματος που με τη σειρά τους θα οδηγούσαν σε σημαντική μείωση του κόστους διαχείρισης των αποθεμάτων.

Σχετικά τώρα με τις μεθόδους διαχωρισμού της παραχθείσας πρόβλεψης στο συναθροισμένο επίπεδο μπορεί να γίνει με χρήση διαφόρων εμπειρικών μεθόδων που έχουν προταθεί από τους συγγραφείς της μεθόδου ADIDA. Οι μέθοδοι αυτοί είναι:

- Equal Weights. Απλός ισοβαρής διαχωρισμός και ενδείκνυται για χρονοσειρές με μεγάλη τυχαιότητα και χωρίς εποχιακή συμπεριφορά.
- Previous Weights. Εφαρμογή βαρών που έχουν οι m προηγούμενες παρατηρήσεις, όπου m είναι ισούται με το επίπεδο συνάθροισης.
- Average Weights. Υπολογισμός των μέσων βαρών που υπολογίζονται αν χωρίσουμε τις παρατηρήσεις σε k ομάδες m παρατηρήσεων η κάθε μία, όπου $(k \times m)$ είναι το σύνολο των διαθέσιμων παρατηρήσεων και m είναι το

επίπεδο της συνάθροισης, ενδείκνυται σε περιπτώσεις όπου η συνιστώσα της εποχιακότητας είναι έντονη.

Συνοψίζοντας, η μέθοδος ADIDA μοιάζει να είναι μια πολλά υποσχόμενη προσέγγιση για χρονοσειρές διακοπτόμενης ζήτησης. Στα πλεονεκτήματα της μεθόδου προστίθεται και η δυνατότητα που παρέχεται από τα περισσότερα πακέτα λογισμικού προβλέψεων όσον αφορά τη συνάθροιση των δεδομένων σε υψηλότερα χρονικά επίπεδα. Τέλος, η μέθοδος ADIDA, μπορεί να οδηγήσει σε σημαντική βελτίωση της ακρίβειας πρόβλεψης σε σχέση με μεμονωμένες μεθόδους, γεγονός που την καθιστά μηχανισμό "αυτοβελτίωσης".

2.4. Κριτικές Προβλέψεις

Στη χρήση στατιστικών μεθόδων πρόβλεψης υιοθετούμε μία πολύ βασική παραδοχή, ότι το πρότυπο πρόβλεψης που ανιχνεύεται από τη μέθοδο θα συνεχίσει να ισχύει και στο μέλλον. Αυτό φυσικά δεν συμβαίνει σε όλα τα προβλήματα που καλούνται οι στατιστικές μέθοδοι να δώσουν απάντηση, καθώς συχνά συμβαίνουν μεταβολές που επιφέρουν αλλαγή στο πρότυπο που παρουσίαζε μέχρι πρότινος μία χρονοσειρά. Όσο γρηγορότερα και όσο πιο αποτελεσματικά εντοπίζονται αυτές οι αλλαγές, τόσο καλύτερο είναι το τελικό αποτέλεσμα. Τέτοιες μεταβολές συχνά, προκειμένου να αναγνωριστούν, κρίνεται απαραίτητη η ανθρώπινη κριτική ικανότητα η οποία συχνά αποτελεί το μοναδικό τρόπο να εκτιμηθεί η επίδραση των αλλαγών αυτών στις προβλέψεις. Συνοψίζοντας, οι κριτικές μέθοδοι, οι οποίες χρησιμοποιούνται πολύ συχνά σε επιχειρήσεις και οργανισμούς, εμπεριέχουν δεδομένα που προκύπτουν από την ανθρώπινη διαίσθηση, κρίση και συσσωρευμένη γνώση. Η πρόβλεψη μέσω αυτών μπορεί να βασίζεται είτε στις γνώσεις και την κρίση ενός ατόμου (ατομικές μέθοδοι) είτε στο συνδυασμό απόψεων των μελών κάποιας επιτροπής (μέθοδοι επιτροπής).

2.5. Προβλέψεις Προϋπολογισμού

Οι προβλέψεις προϋπολογισμού διαφέρουν από τις προηγούμενες δύο κατηγορίες που περιγράψαμε. Αναφέρονται σε προβλέψεις που λαμβάνουν χώρα συνήθως στην αρχή της χρονιάς από τον διευθυντή ανάπτυξης μίας εταιρίας. Με δεδομένο έναν καθορισμένο ετήσιο ρυθμό ανάπτυξης (growth rate) καθορίζεται ο επιθυμητός ρυθμός ανάπτυξης και οι προβλέψεις προσαρμόζονται αναλόγως. Σε αντίθεση με τις στατιστικές προβλέψεις, για την παραγωγή μιας πρόβλεψης προϋπολογισμού δε βασιζόμαστε τόσο σε δεδομένα και αλγορίθμους αλλά στους στόχους που θέλουμε να θέσουμε για την εταιρία ή επιχείρηση για το εγγύς μέλλον. Πρόκειται για ένα «wish-future-status» παρά για μία πραγματική πορεία χρονοσειράς και ακριβή πρόβλεψη. Περιέχει αρκετή αισιοδοξία και μεροληψία (biased forecast) και συνήθως μεγάλα σφάλματα. Αποτελεί ένα μέσο καθορισμού στόχων της εταιρίας-επιχείρησης ώστε να υπάρχει και η κατάλληλη πολιτική ανάπτυξης προς την κατεύθυνση αυτή.

2.6. Αξιολόγηση Προβλέψεων

Αφότου έχουμε ολοκληρώσει τη διαδικασία πρόβλεψης πρέπει να αξιολογήσουμε κατά πόσο το μοντέλο μας παρήγαγε ακριβές προβλέψεις και στη περίπτωση που εφαρμόσαμε περισσότερα μοντέλα, πιο από αυτά είναι το πιο ακριβές. Για να το πετύχουμε αυτό χρησιμοποιούμε ένα σύνολο στατιστικών δεικτών αξιολόγησης της ακρίβειας. Αυτοί οι δείκτες, ή αλλιώς σφάλματα, μπορούν να χρησιμοποιηθούν για να μετρήσουν την απόδοση του υπό εφαρμογή μοντέλου στο σύνολο των δεδομένων του ιστορικού της χρονοσειράς, που είναι δηλαδή γνωστά στο μοντέλο μας (in-sample error, είναι στις τιμές του ορίζοντα πρόβλεψης (out-of-sample error)). Αφότου ο σκοπός μας είναι να έχουμε ακριβή αποτύπωση της μελλοντικής εξέλιξης της χρονοσειράς, ενδιαφέρον για τη μέτρηση της απόδοσης του μοντέλου παρουσιάζει η δεύτερη κατηγορία.

Γενικά ορίζουμε ως σφάλμα της πρόβλεψης μιας περιόδου:

$$e_i = Y_i - F_i$$

Το απλό ποσοστιαίο σφάλμα ως:

$$p_i = \frac{100 \cdot e_i}{Y_i} (\%)$$

Επίσης, έχουμε το απλό σχετικό σφάλμα, που είναι ο λόγος του σφάλματος της υπό εξέταση μεθόδου σε σχέση με κάποια άλλη μέθοδο. Συνήθως η μέθοδος σύγκρισης είναι η Naive, που περιεγράφηκε πρωτύτερα:

$$r_i = \frac{e_i}{e_i^*}$$

Το απλό κανονικοποιημένο σφάλμα βρίσκεται από τη παρακάτω σχέση:

$$q_i = \frac{e_i}{\frac{1}{n-1} \cdot \sum_{i=2}^n |Y_i - Y_{i-1}|}$$

Στον Πίνακα 2.1 βλέπουμε τους πιο κοινούς δείκτες ακρίβειας. Το mean υποδηλώνει τον αριθμητικό μέσο όρο, το median τη διάμεσο και το gmean το γεωμετρικό μέσο. Επίσης το 1 λαμβάνει τη τιμή 1 στη περίπτωση που αυτό που εσωκλείει είναι αληθές, αλλιώς παίρνει την τιμή 0.

Κεφάλαιο 2. Μέθοδοι Πρόβλεψης

Στηριζόμενοι στη δουλειά των Hyndman και Koehler μπορούμε να χωρίσουμε τους δείκτες αυτούς στις εξής κατηγορίες:

Δείκτες που εξαρτώνται από τη κλίμακα των δεδομένων

Η κατηγορία αυτή αποτελείται από τους δείκτες MSE, RMSE, MAE, MdAE που συναρτώνται της απόλυτης τιμής των δεδομένων. Βοηθούν ιδιαίτερα όταν καλούμαστε να συγκρίνουμε την απόδοση διαφόρων μοντέλων πρόβλεψης επί της ίδιας χρονοσειράς. Σε περίπτωση που τους εφαρμόζουμε, όμως, σε ένα σύνολο χρονοσειρών με διαφορετική κλίμακα μπορούν να παράγουν αποπροσανατολιστικά αποτελέσματα. Το προτέρημα του δείκτη RMSE έναντι του απλού MSE είναι ότι μας δίνει μετρήσεις στην ίδια κλίμακα με αυτή της χρονοσειράς στην οποία εφαρμόζεται. Επίσης, οι δύο παραπάνω δείκτες είναι ιδιαίτερα ευαίσθητοι στις ακραίες τιμές μιας χρονοσειράς, σε σύγκριση με του MAE, MdAE λόγω του τετραγωνισμού του σφάλματος. Αυτό τους καθιστά ακατάλληλους για χρήση στην αξιολόγηση της ακρίβειας πρόβλεψης.

Δείκτες που βασίζονται σε ποσοστιαία σφάλματα

Σε αυτή την ομάδα, από τους δείκτες που είδαμε, ανήκουν οι MAPE, MdAPE, sMAPE και sMdAPE. Οι δείκτες MAPE και MdAPE υστερούν στο γεγονός ότι δεν δύνανται να λάβουν τιμή όταν οι πραγματικές παρατηρήσεις είναι μηδενικές και επιπρόσθετα παρουσιάζουν έντονη ασυμμετρία όταν οι παρατηρήσεις είναι κοντά στο μηδέν. Έτσι, ο δείκτης MAPE παρουσιάζει χαρακτηριστικά μεγάλες τιμές σε σχέση με τον δείκτη MdAPE. Μπορούμε, βέβαια, με τη χρήση λογαριθμικών μετασχηματισμών να προσδώσουμε μια σταθερότητα στους δείκτες. Επίσης, οι δύο αυτοί δείκτες δίνουν μεγαλύτερη βαρύτητα στα θετικά έναντι των αρνητικών σφαλμάτων. Αντίθετα, οι δείκτες sMAPE και sMdAPE δεν παρουσιάζουν στον ίδιο βαθμό το πρόβλημα των μηδενικών τιμών. Αν και το όνομα τους υποδηλώνει συμμετρία, έχουν το μειονέκτημα ότι οι αισιόδοξες και οι απαισιόδοξες προβλέψεις δεν υπολογίζονται με το ίδιο βάρος.

Δείκτες σχετικών σφαλμάτων

Από τους δείκτες που έχουμε εξετάσει οι MRAE, MdRAE και GMRAE ανήκουν σε αυτή την κατηγορία. Το πρόβλημα με αυτή των ομάδα δεικτών είναι ότι στις περιπτώσεις που το σφάλμα αναφοράς λαμβάνει αρκετά μικρές τιμές ο λόγος r_i τείνει να έχει άπειρη διακύμανση.

Σχετικοί δείκτες

Το πλεονέκτημα των σχετικών δεικτών RelMAE, RelRMSE και PB είναι ότι αποφεύγουν το πρόβλημα που είδαμε στη προηγούμενη κατηγορία δεικτών με τις άπειρες τιμές. Επίσης, μας επιτρέπουν να αποφασίσουμε εύκολα πιο από τις μεθόδους που συγκρίνουμε είναι πιο ακριβής, ανάλογα με τον αν ο δείκτης έχει τιμή

Κεφάλαιο 2. Μέθοδοι Πρόβλεψης

μεγαλύτερη της μονάδας. Βέβαια, καθότι απαιτούν αρκετές προβλέψεις δε μπορούμε να τους εφαρμόσουμε όταν έχουν πρόβλεψη με ορίζοντα ίσο με ένα.

Κανονικοποιημένοι δείκτες

Οι κανονικοποιημένοι δείκτες MAsE, MdAsE προσφέρουν τόσο ευκολία στην ερμηνεία αντίστοιχη των σχετικών δεικτών, ενώ συγχρόνως μπορούν να εφαρμοστούν για μοναδική περίοδο πρόβλεψης. Επιπρόσθετα, είναι απαλλαγμένες από το επίπεδο της κάθε χρονοσειράς και εφαρμόσιμες σε μαζική πρόβλεψη χρονοσειρών, αποφεύγοντας τις απροσδιοριστίες των ποσοστιαίων σφαλμάτων.

Στον πίνακα 2.1 παρατίθενται οι κυριότεροι δείκτες ακρίβειας με το πλήρες όνομα, την συντομογραφία και τον τύπο τους.

Συντομογραφία	Πλήρες Όνομα	Τύπος
<i>ME</i>	<i>Mean Error</i>	$mean(e_i)$
<i>MSE</i>	<i>Mean Squared Error</i>	$mean(e_i^2)$
<i>RMSE</i>	<i>Rooted Mean Squared Error</i>	$\sqrt{mean(e_i^2)}$
<i>MAE</i>	<i>Mean Absolute Error</i>	$mean(e_i)$
<i>MdAE</i>	<i>Median Absolute Error</i>	$median(e_i)$
<i>MAPE</i>	<i>Mean Absolute Percentage Error</i>	$mean(p_i)$
<i>MdAPE</i>	<i>Median Absolute Percentage Error</i>	$median(p_i)$
<i>sMAPE</i>	<i>Symmetric Mean Absolute Percentage Error</i>	$mean\left(\frac{200 \cdot Y_i - F_i }{Y_i + F_i}\right)$
<i>sMdAPE</i>	<i>Symmetric Median Absolute Percentage Error</i>	$median\left(\frac{200 \cdot Y_i - F_i }{Y_i + F_i}\right)$
<i>MRAE</i>	<i>Mean Relative Absolute Error</i>	$mean(r_i)$
<i>MdRAE</i>	<i>Median Relative Absolute Error</i>	$median(r_i)$
<i>GMRAE</i>	<i>Geometric Mean Relative Absolute Error</i>	$gmean(r_i)$
<i>RelMAE</i>	<i>Relative Mean Absolute Error</i>	MAE/MAE_b
<i>RelMSE</i>	<i>Relative Mean Squared Error</i>	$RMSE/RMSE_b$
<i>LMR</i>	<i>Log Mean Squared Error Ratio</i>	$\log(RelMSE)$
<i>PB</i>	<i>Percentage Better</i>	$100 \cdot mean(1\{ r_i < 1\})$
<i>PB(MAE)</i>	<i>Percentage Better (MAE)</i>	$100 \cdot mean(1\{MAE < MAE_b\})$
<i>PB(MSE)</i>	<i>Percentage Better (MSE)</i>	$100 \cdot mean(1\{MSE < MSE_b\})$
<i>MAsE</i>	<i>Mean Absolute Scaled Error</i>	$mean(q_i)$
<i>MdAsE</i>	<i>Median Absolute Scaled Error</i>	$median(q_i)$

Πίνακας 2.1 Δείκτες Ακρίβειας

Κεφάλαιο 3 . Μέθοδοι Συμψηφισμού

3.1. Εισαγωγή

Στις επιχειρήσεις και στα οικονομικά υπάρχουν συχνά εφαρμογές που απαιτούν προβλέψεις από πολλές συσχετιζόμενες χρονοσειρές που οργανώνονται σε ιεραρχικές δομές με βάση κάποια μεγέθη, όπως 'είδος προϊόντος' ή 'τόπος'. Το γεγονός αυτό οδήγησε στην ανάγκη για ύπαρξη μεθόδων οι οποίες «συμψηφίζουν» πληροφορίες από διαφορετικές χρονοσειρές της δομής προκειμένου να βελτιωθεί η ακρίβεια των προβλέψεων στο σύνολο της δομής αυτής. Η διαδικασία αυτή ονομάζεται συμψηφισμός (reconciliation όπως αναφέρεται στη διεθνή βιβλιογραφία) και περιλαμβάνει διαφορετικές μεθόδους που χρησιμοποιούνται για την περαιτέρω αξιοποίηση πληροφοριών των χρονοσειρών που δεν χρησιμοποιούν οι τεχνικές προβλέψεων που αναφέρθηκαν στο Κεφάλαιο 2.

Οι μέθοδοι αυτές εφαρμόζονται σε «ιεραρχικές» δομές πολλών επιπέδων ως εξής. Το επίπεδο 0 υποδεικνύει το συνολικό άθροισμα όλων των χρονοσειρών της δομής, το επίπεδο 1 ακολουθεί ύστερα από την πρώτη διάσπαση των χρονοσειρών σε ευρύτερες ομάδες και φτάνουμε στο επίπεδο K που περιλαμβάνει όλες τις χρονοσειρές διαχωρισμένες μεταξύ τους. Μία απεικόνιση φαίνεται στο σχήμα Fig. 1, στο οποίο φαίνεται πως δομείται μια ιεραρχία τριών επιπέδων κατ' αυτόν τον τρόπο. Ένας άλλος τρόπος για να αναφερθούμε στα διάφορα επίπεδα είναι, όπως φαίνεται και στο εν λόγω σχήμα, να ονομάσουμε το επίπεδο 0 «Total» ή «Σύνολο» και το κάθε επιμέρους υποσύνολο να εκφράζεται με ένα κεφαλαίο γράμμα της αλφαβήτου (για το επίπεδο 1), με δύο γράμματα στο επόμενο επίπεδο, κ.ο.κ..

Έστω τώρα ότι λαμβάνουμε μετρήσεις για τις χρονοσειρές στις χρονικές περιόδους $t = 1, 2, 3, \dots, n$ και ενδιαφερόμαστε να προβλέψουμε τις τιμές των χρονοσειρών αυτών στις χρονικές περιόδους $t = n+1, n+2, \dots, n+h$. Είναι βολικό να χρησιμοποιούμε τον χαρακτήρα X προκειμένου να δηλώσουμε ότι αναφερόμαστε σε μία συγκεκριμένη χρονοσειρά μέσα στην ιεραρχία. Οι παρατηρήσεις για μία X χρονοσειρά δηλώνονται ως $Y_{X,t}$. Έτσι, αν αναφερόμαστε στη δομή του σχήματος Fig. 1 και μιλάμε για την πρόβλεψη τη χρονική περίοδο t για την χρονοσειρά AF, ο συμβολισμός είναι $Y_{AF,t}$. Επίσης χρησιμοποιούμε το Y_t για να δηλώσουμε τη συνάθροιση όλων των χρονοσειρών τη χρονική περίοδο t.

Με τις παραδοχές αυτές προκύπτουν οι εξής σχέσεις:

$$Y_t = \sum_i Y_{i,t}$$

$$Y_{i,t} = \sum_j Y_{ij,t}$$

$$Y_{ij,t} = \sum_k Y_{ijk,t}$$

Κεφάλαιο 3. Μέθοδοι Συμψηφισμού

κ.ο.κ.. Έτσι συμπεραίνουμε ότι οι προβλέψεις των υψηλότερων επιπέδων μπορούν να υπολογιστούν με το άθροισμα των κατώτερων σειρών.

Ας υποθέσουμε τώρα ότι έχουμε m_i χρονοσειρές στο επίπεδο i , όπου $i = 0, 1, 2, \dots, K$. Άρα θα ισχύει ότι $m_i > m_{i-1}$ και ο συνολικός αριθμός χρονοσειρών στη δομή είναι $m = m_0 + m_1 + \dots + m_K$. Στο παράδειγμα του σχήματος Fig. 1 έχουμε $m_i = 3^i$ και $m = 40$.

Η μαθηματική ανάλυση γίνεται με τη χρήση διανυσμάτων και πινάκων. Έστω ότι $Y_{i,t}$ είναι το διάνυσμα των παρατηρήσεων του επιπέδου i τη χρονική περίοδο t , δηλαδή $Y_t = [Y_t, Y_{1,t}, \dots, Y_{K,t}]'$. Η σχέση αυτή μπορεί να γραφτεί και ως $Y_t = SY_{K,t}$, όπου S ένας αθροιστικός πίνακας διαστάσεων $m \times m_K$ που χρησιμοποιείται για τον συμψηφισμό των χαμηλότερων επιπέδων. Στο παράδειγμα του σχήματος Fig. 1 έχουμε $Y_t = [Y_t, Y_{A,t}, Y_{B,t}, Y_{C,t}, Y_{AA,t}, Y_{AB,t}, \dots, Y_{CC,t}, Y_{AAA,t}, Y_{AAB,t}, \dots, Y_{CCC,t}]'$ και ο πίνακας S έχει διαστάσεις 40×27 και δομείται ως εξής:

$$S = \begin{bmatrix} 11111111111111111111111111111111 \\ 1111111110000000000000000000 \\ 0000000001111111110000000000 \\ 0000000000000000000111111111 \\ 1110000000000000000000000000 \\ 0000111000000000000000000000 \\ \vdots \\ 00000000000000000000000000111 \\ 1000000000000000000000000000 \\ 0100000000000000000000000000 \\ \vdots \\ 00000000000000000000000000001 \end{bmatrix}$$

Ο πίνακας S είναι τάξης m_K . Είναι σαφές ότι ο πίνακας S μπορεί να συμπληρωθεί από τα επίπεδα της εκάστοτε ιεραρχίας, καθώς η πρώτη σειρά αποτελείται από ένα μοναδιαίο διάνυσμα μήκους m_K και το κάτω κομμάτι αποτελείται από έναν μοναδιαίο πίνακα διαστάσεων $m_K \times m_K$. Το μεσαίο κομμάτι πρόκειται για «μοναδιαίο διαγώνιο μπλοκ».

Αφού δώσαμε μία πρώτη μαθηματική προσέγγιση, καλό είναι να δούμε πως εφαρμόζονται οι μέθοδοι reconciliation στις εκάστοτε δομές. Έστω ότι υπολογίζουμε πρώτα τις προβλέψεις για κάθε χρονοσειρά σε κάθε επίπεδο δίνοντας m βασικές προβλέψεις για κάθε χρονική περίοδο $n+1, n+2, \dots, n+h$, βασιζόμενοι στα γνωστά δεδομένα μέχρι και την χρονική περίοδο n . Συμβολίζουμε τις προβλέψεις αυτές ως $\hat{Y}_{X,n}(h)$, όπου X η σειρά στην οποία αναφέρεται η πρόβλεψη. Έτσι, έχουμε ότι το $\hat{Y}_n(h)$ δηλώνει την πρόβλεψη του συνόλου τη χρονική περίοδο h , το $\hat{Y}_{A,n}(h)$ την αντίστοιχη πρόβλεψη της σειράς A , το $\hat{Y}_{AC,n}(h)$ της AC , κ.ο.κ.. Τώρα το $\hat{Y}_n(h)$ αποτελεί το διάνυσμα που εμπεριέχει όλες αυτές τις προβλέψεις με την ίδια σειρά που αναφέρεται στις χρονοσειρές και το \hat{Y}_t .

Όλες οι υπάρχουσες μέθοδοι ιεραρχικής πρόβλεψης μπορούν να γραφτούν στη μορφή:

$$\tilde{Y}_n(h) = SP\hat{Y}_n(h)$$

με τον πίνακα P να έχει διαστάσεις $m_K \times m$ και να λαμβάνει τις κατάλληλες τιμές, ανάλογα με τη μέθοδο στην οποία αναφερόμαστε. Άρα, από τη σχέση αυτή προκύπτει ότι οι προβλέψεις των μεθόδων στις οποίες αναφερόμαστε είναι γραμμικός συνδυασμός των βασικών προβλέψεων για κάθε χρονοσειρά. Η δράση του πίνακα P είναι να συλλέγει και να συνδυάζει συσχετιζόμενα στοιχεία από τις βασικές προβλέψεις $\hat{Y}_n(h)$, τα οποία μετά αθροίζει ο πίνακας S κατάλληλα για να προκύψουν οι τελικές προβλέψεις $\tilde{Y}_n(h)$.

Εκτός από τις ευρέως διαδεδομένες Bottom-Up και Top-Down μεθόδους που θα αναλυθούν στη συνέχεια, αλλάζοντας τον πίνακα P κατάλληλα, λαμβάνουμε παραλλαγές όπως προβλέψεις middle-out. Αυτό υποδηλώνει πως άλλες νέες ιεραρχικές μέθοδοι πρόβλεψης μπορούν να οριστούν διαλέγοντας κατάλληλους πίνακες P , με δεδομένου όμως ότι θέτουμε κάποιους περιορισμούς για τον πίνακα αυτό προκειμένου οι προβλέψεις των μεθόδων να είναι σε σωστά πλαίσια.

Αν υποθέσουμε ότι οι βασικές προβλέψεις δεν είναι προκατειλημμένες (unbiased), δηλαδή ότι $E[\hat{Y}_n(h)] = E[Y_n(h)]$ και ότι θέλουμε και οι προβλέψεις του ιεραρχικού μοντέλου να είναι επίσης μη προκατειλημμένες, τότε πρέπει να ισχύει ότι:

$$E[\tilde{Y}_n(h)] = E[Y_n(h)] = SE[Y_{K,n}(h)]$$

Αν υποθέσουμε ότι $\beta_n(h) = E[Y_{K,n+h} | Y_1, \dots, Y_n]$ είναι οι μέσοι όροι των μελλοντικών τιμών του κατώτερου επιπέδου K , τότε:

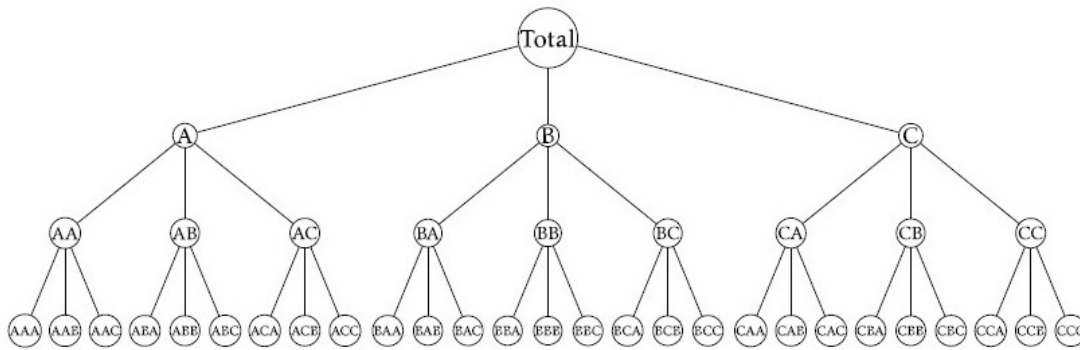
$$E[\tilde{Y}_n(h)] = SPE[\hat{Y}_n(h)] = SPS\beta_n(h).$$

Άρα η αντικειμενικότητα των προβλέψεων θα ισχύει αν και μόνο αν:

$$SPS = S \quad (1)$$

Αν υποθέσουμε τώρα ότι η διακύμανση των βασικών προβλέψεων $\hat{Y}_n(h)$ δίνεται από το Σ_h , τότε η διακύμανση των προβλέψεων του εκάστοτε μοντέλου θα δίνεται από τη σχέση:

$$Var[\tilde{Y}_n(h)] = SP\Sigma_h P'S'$$



Σχήμα 3.1 Διάγραμμα Ιεραρχίας Τριών Επιπέδων

3.2. Η Μέθοδος Bottom-Up

Η πρώτη μέθοδος reconciliation στην οποία θα αναφερθούμε είναι η μέθοδος Bottom-Up. Με τη μέθοδο αυτή, οι χρονοσειρές κατώτερων βαθμίδων δίνουν προβλέψεις τις οποίες στη συνέχεια αθροίζουμε προκειμένου να προκύψουν προβλέψεις για τις ανώτερες. Δηλαδή, στη δομή στο Σχήμα 3.1, για παράδειγμα, αφού προβλέψουμε με κάποια μέθοδο πρόβλεψης της επιλογής μας τις χρονοσειρές του επιπέδου 3, αθροίζουμε τις τιμές που θα προκύψουν προκειμένου να λάβουμε τιμές για το επίπεδο 2, ύστερα για το επίπεδο 1 και τέλος για το επίπεδο 0, δηλαδή για το άθροισμα όλων των χρονοσειρών.

Σύμφωνα με τη μαθηματική ανάλυση που έγινε παραπάνω, για την υλοποίηση της μεθόδου Bottom-Up, ο πίνακας P δομείται ως εξής:

$$P = [0_{m_K \times (m-m_K)} \mid I_{m_K}] \quad (2)$$

όπου, $0_{l \times k}$ είναι μηδενικός πίνακας διαστάσεων $l \times k$ και ο I_k μοναδιαίος πίνακας διαστάσεων $k \times k$. Στην περίπτωση αυτή ο P συλλέγει μόνο τις προβλέψεις του κατώτερου επιπέδου από το $\hat{Y}_n(h)$, τις οποίες εν συνεχεία αθροίζει ο S προκειμένου να λάβουμε τις τελικές Bottom-Up προβλέψεις.

Τέλος, για την μέθοδο αυτή ισχύει ότι $SPS = S$, δηλαδή τα αποτελέσματά της δεν τείνουν να έχουν θετική ή αρνητική προκατάληψη.

3.3. Η Μέθοδος Top-Down

Με αντίστοιχο τρόπο δουλεύει και η μέθοδος Top-Down, μόνο που στην περίπτωση αυτή λαμβάνουμε αρχικά πρόβλεψη για την ανώτερη κλίμακα, δηλαδή το επίπεδο 0 και με χρήση κατάλληλων δεικτών-βαρών διαμερίζουμε την πρόβλεψη αυτή στις χρονοσειρές του επόμενου επιπέδου, δηλαδή του επιπέδου 1, ύστερα του επιπέδου 2 κ.ο.κ.. Στο παράδειγμα του Σχήματος 3.1 θα προβλέπαμε την αθροισμένη

χρονοσειρά του επιπέδου 0 και με κατάλληλους δείκτες θα λαμβάναμε προβλέψεις για το επίπεδο 1, ύστερα για το επίπεδο 2 και, τέλος, για το επίπεδο 3.

Σύμφωνα με την μαθηματική ανάλυση που έγινε παραπάνω, για την υλοποίηση της μεθόδου Top-Down, ο πίνακας P δομείται ως εξής:

$$P = [p \mid 0_{m_K \times (m-1)}] \quad (3)$$

όπου $p = [p_1, p_2, \dots, p_{m_K}]'$ το διάνυσμα που περιέχει τους δείκτες για κάθε χρονοσειρά με $\sum_{i=1}^{m_K} p_i = 1$. Στην περίπτωση αυτή ο πίνακας P συλλέγει την πληροφορία της πρόβλεψης της συναθροισμένης χρονοσειράς και τη διασπείνει προκειμένου να διανεμηθεί στα επόμενα επίπεδα. Διαφορετικές μέθοδοι Top-Down χρησιμοποιούν και διαφορετικούς πίνακες P .

Τέλος, για την μέθοδο αυτή δεν ισχύει ότι $SPS = S$, γεγονός που υποδεικνύει ότι η μέθοδος Top-Down δεν δίνει ποτέ μη προκατειλημμένα αποτελέσματα, ακόμα και αν οι βασικές προβλέψεις είναι αντικειμενικές.

3.4. Optimal Reconciliation

Όπως αναφέρθηκε, με κατάλληλη επιλογή του πίνακα P , μπορούν να προκύψουν πολλές παραλλαγές των μεθόδων reconciliation. Το 2008 οι Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos και Han Lin Shang πρότειναν μία νέα μέθοδο βέλτιστου συμψηφισμού χρησιμοποιώντας παλινδρόμηση. Η πρότασή τους βασίστηκε στο γεγονός ότι μπορούμε να γράψουμε τις βασικές προβλέψεις ως εξής:

$$\hat{Y}_n(h) = S\beta_n(h) + \varepsilon_h \quad (4)$$

όπου $\beta_n(h) = E[Y_{K,n+h} \mid Y_1, \dots, Y_n]$ είναι οι άγνωστοι μέσοι όροι του κατώτερου επιπέδου K και το ε_h έχει μηδενικό μέσο όρο και πίνακα συνδιακύμανσης $\text{Var}(\varepsilon_h) = \Sigma_h$. Αυτό σημαίνει πως μπορούμε να υπολογίσουμε το $\beta_n(h)$ αναλύοντας την αρχική συνάρτηση ως εξίσωση παλινδρόμησης και με τον τρόπο αυτό να λάβουμε προβλέψεις για όλα τα επίπεδα της ιεραρχίας. Αν γνωρίζαμε το Σ_h , τότε θα μπορούσαμε να χρησιμοποιήσουμε τη μέθοδο των γενικευμένων ελάχιστων τετραγώνων για να λάβουμε την αντικειμενική προσέγγιση του $\beta_n(h)$ με την ελάχιστη διακύμανση ως εξής:

$$\hat{\beta}_n(h) = (S' \Sigma_h^{-1} S)^{-1} S' \Sigma_h^{-1} \hat{Y}_n(h) \quad (5)$$

όπου Σ_h^{-1} είναι ο Moore-Penrose γενικευμένος αντίστροφος του Σ_h . Χρησιμοποιούμε τον γενικευμένο αντίστροφο γιατί ο Σ_h είναι συχνά σχεδόν μοναδιαίος λόγω της συναθροίσεως που περιλαμβάνεται στο Y_n . Αυτό μας οδηγεί στις εξής γενικευμένες ανανεωμένες προβλέψεις:

$$\tilde{Y}_n(h) = S\hat{\beta}_n(h) = SP\hat{Y}_n(h) \quad (6)$$

όπου $P = (S' \Sigma_h^T S)^{-1} S' \Sigma_h^T$. Εμφανώς η συνθήκη αντικειμενικότητας ισχύει και η διακύμανση των προβλέψεων δίνεται από τον τύπο:

$$\text{Var}[\tilde{Y}_n(\mathbf{h})] = S(S' \Sigma_h^T S)^{-1} S' \quad (7)$$

Η δυσκολία στη μέθοδο αυτή έγκειται στο γεγονός ότι απαιτεί τη γνώση του Σ_h , ή τουλάχιστον μία καλή εκτίμησή του. Σε μία μεγάλη ιεραρχία με χιλιάδες σειρές αυτό μπορεί να μην είναι δυνατόν.

Ωστόσο μπορούμε να απλοποιήσουμε σε μεγάλο βαθμό τους υπολογισμούς υποθέτοντας ότι το σφάλμα ε_h μπορεί να εκφραστεί ως $\varepsilon_h \approx S \varepsilon_{K,h}$, όπου $\varepsilon_{K,h}$ είναι το σφάλμα πρόβλεψης στο κατώτερο επίπεδο. Αν ισχύει αυτό, υποθέτουμε ότι τα σφάλματα ικανοποιούν την ίδια συνθήκη συνάθροισης με τα αρχικά δεδομένα. Αυτή η υπόθεση επαληθεύεται όταν οι προβλέψεις προσεγγιστικά ικανοποιούν αυτή την συνθήκη συνάθροισης, γεγονός που θα πρέπει να ισχύει για κάθε σετ λογικών προβλέψεων. Σε ορισμένες περιπτώσεις είναι πιθανό τα σφάλματα να ικανοποιούν την έκφραση αυτή ακριβώς και όχι προσεγγιστικά. Για παράδειγμα, αν μια μέθοδος πρόβλεψης που περιλαμβάνει γραμμική παλινδρόμηση (π.χ. ένα μοντέλο ARIMA) με συγκεκριμένες παραμέτρους έχει χρησιμοποιηθεί για όλες τις χρονοσειρές, τότε τα σφάλματα θα αθροίζονται με ακριβώς τον ίδιο τρόπο. Άρα η υπόθεση αυτή είναι μία λογική προσέγγιση των πρακτικών εφαρμογών.

Η προσέγγιση αυτή οδηγεί στο αποτέλεσμα $\Sigma_h \approx S \Omega_h S'$, όπου $\Omega_h = \text{Var}(\varepsilon_{K,h})$. Με τις παραδοχές αυτές δομείται το εξής θεώρημα:

Θεώρημα 1. Ας υποθέσουμε ότι $Y_n(\mathbf{h}) = S\beta_n + \varepsilon$ με $\text{Var}(\varepsilon) = \Sigma_h = S\Omega_h S'$ και ότι ο S είναι πίνακας συνάθροισης. Τότε η προσέγγιση με τη μέθοδο των γενικευμένων ελάχιστων τετραγώνων για το β που λαμβάνεται χρησιμοποιώντας τον Moore-Penrose γενικευμένο αντίστροφο είναι ανεξάρτητη του Ω_h :

$$\hat{\beta}_n = (S' \Sigma_h^T S)^{-1} S' \Sigma_h^T Y = (SS')^{-1} S' Y \quad (8)$$

με πίνακα διακύμανσης $\text{Var}(\hat{\beta}) = \Omega_h$. Επιπλέον, αυτή είναι η ελάχιστη προσέγγιση της γραμμικής αντικειμενικής διακύμανσης.

Απόδειξη. Γράφουμε το $\Sigma_h = BC$, όπου $B = S\Omega_h$ και $C = S'$. Τότε, από το Fact 6.4.8 του Bernstein (2005, p.235), ο Moore-Penrose γενικευμένος αντίστροφος του Σ_h είναι:

$$\Sigma_h^T = C(CC')^{-1}(B'B)^{-1}B' = S(S'S)^{-1}(\Omega_h' S' S \Omega_h)^{-1} \Omega_h' S' \quad (9)$$

Τότε, $(S' \Sigma_h^T S)^{-1} S' \Sigma_h^T = (S'S)^{-1} S'$. Η διακύμανση προκύπτει αντικαθιστώντας την παραπάνω σχέση στο $(S' \Sigma_h^T S)^{-1}$.

Οι Tian και Wiens (2006, Theorem 3) δείχνουν ότι ο εκτιμητής GLS θα είναι η ελάχιστη μη προκατειλημμένη προσέγγιση της διακύμανσης αν και μόνον αν:

$$SS^T \Sigma_h^T \Sigma_h (1 - SS^T) = \mathbf{0}$$

όπου $S^T = (S'S)^{-1}S'$. Με χρήση της σχέσης (9) είναι εύκολο να δείξουμε ότι η συνθήκη ισχύει.

Το αποτέλεσμα αυτό δείχνει ότι μπορούμε να χρησιμοποιήσουμε OLS αντί του GLS όταν υπολογίζουμε τις ανανεωμένες προβλέψεις, χωρίς την ανάγκη προσέγγισης ενός ξεχωριστού πίνακα συνδιακύμανσης. Αυτό διευκολύνει τους υπολογισμούς ειδικά για έναν μεγάλο πίνακα συνδιακύμανσης. Έτσι χρησιμοποιούμε:

$$\tilde{Y}_n(\mathbf{h}) = S(S'S)^{-1}S'\hat{Y}_n(\mathbf{h}) \quad (10)$$

και $P = (S'S)^{-1}S'$. Η ο πίνακας διακύμανσης συνδιακύμανσης των ανανεωμένων προβλέψεων είναι $Var[\tilde{Y}_n(\mathbf{h})] = \Sigma_h$. Για το λόγο αυτό τα διαστήματα προβλέψεων πάλι απαιτούν προσέγγιση του Σ_h .

Η εξίσωση (10) δείχνει ότι υπό την προϋπόθεση ότι $\Sigma_h = S'\Omega_h S$, ο βέλτιστος συνδυασμός των βασικών προβλέψεων είναι ανεξάρτητος των δεδομένων. Για μια απλή ιεραρχική δομή με μόνο ένα επίπεδο διαίρεσης ($K = 1$) και με m_1 κόμβους (χρονοσειρές) στο επίπεδο 1, τα βάρη δίνονται από τον τύπο:

$$S(S'S)^{-1}S' = (m_1 + 1)^{-1} \begin{pmatrix} m_1 & 1 & 1 & \dots & \dots & 1 \\ 1 & m_1 & -1 & -1 & \dots & -1 \\ 1 & -1 & m_1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & -1 & \dots & -1 & \ddots & -1 \\ 1 & -1 & -1 & \dots & -1 & m_1 \end{pmatrix}$$

Με αντίστοιχο τρόπο μπορούμε να υπολογίσουμε βάρη για οποιαδήποτε ιεραρχική δομή, ανεξάρτητα από τα δεδομένα που έχουμε.

Η μέθοδος των Hyndman, Ahmed, Athanasopoulos και Shang «μοιράζεται» κάποιες ομοιότητες με την προσέγγιση που παρουσίασαν οι Stone et al.(1942) και ανέλυσαν περαιτέρω οι Solomou και Weale (1991, 1993, 1996) και Weale (1985, 1988). Στις αναλύσεις αυτές το πρόβλημα ιεραρχίας κατασκευάζεται χρησιμοποιώντας ταυτόχρονες γραμμικές εξισώσεις στα πλαίσια της εξισορρόπησης της εθνικού ταμείου. Το εθνικό ταμείο διασπάται σε παραγωγή, έσοδα και έξοδα και συναλλαγές κεφαλαίων. Η παραγωγή διασπάται περαιτέρω σε παραγωγή εντός της Βρετανίας και σε παραγωγή του υπόλοιπου κόσμου, τα έσοδα και έξοδα και οι συναλλαγές κεφαλαίων κατηγοριοποιούνται περαιτέρω σε άτομα, εταιρίες, δημόσιους οργανισμούς, κυβέρνηση και λοιπός κόσμος. Η λύση που προτείνεται στα paper αυτά είναι ισοδύναμη με την GLS εκτίμηση που αναφέρεται στη μέθοδο που αναλύθηκε αλλά χρησιμοποιεί δείκτες και συμβολισμούς που είναι πιο κατάλληλοι για την ανάλυση λογιστικής εξισορρόπησης παρά για ιεραρχικές προβλέψεις και χωρίς τις απλοποιήσεις που απαιτεί η εξίσωση (10) για να χρησιμοποιηθεί. Επίσης δεν λαμβάνουν υπόψιν την εφαρμογή της λύσης όταν οι διαστάσεις του πίνακα S

Κεφάλαιο 3. Μέθοδοι Συμψηφισμού

είναι μεγάλες, γεγονός που είναι ευρέως διαδεδομένο σε προβλήματα ιεραρχικής πρόβλεψης.

Κεφάλαιο 4 . Νευρωνικά Δίκτυα και Μηχανική Μάθηση

4.1. Εισαγωγή

Το έργο στο επιστημονικό πεδίο των τεχνητών νευρωνικών δικτύων (χάρην συντομίας, αποκαλούνται συνήθως «νευρωνικά δίκτυα») βασίστηκε, από τις απαρχές του, στο γεγονός ότι ο ανθρώπινος εγκέφαλος εκτελεί τους υπολογισμούς του με εντελώς διαφορετικό τρόπο από το συμβατικό ηλεκτρονικό υπολογιστή. Ο εγκέφαλος είναι ένας εξαιρετικά πολύπλοκος, μη γραμμικός, παράλληλος υπολογιστής (σύστημα επεξεργασίας πληροφοριών). Έχει τη δυνατότητα να οργανώνει τα δομικά του στοιχεία, γνωστά ως νευρώνες, με τρόπο ώστε να εκτελούν συγκεκριμένους υπολογισμούς (π.χ. αναγνώριση προτύπων, αντίληψη και έλεγχος της κίνησης) με ταχύτητα πολλαπλάσια από αυτή του ταχύτερου ψηφιακού υπολογιστή που υπάρχει σήμερα. Ένα χαρακτηριστικό παράδειγμα είναι η ανθρώπινη όραση, η οποία είναι μία διαδικασία επεξεργασίας πληροφοριών. Είναι ευθύνη του οπτικού συστήματος να μας παρέχει μία αναπαράσταση του περιβάλλοντός μας και, ακόμη πιο σημαντικό να μας προμηθεύει με τις πληροφορίες που χρειαζόμαστε για να επικοινωνήσουμε με αυτό. Συγκεκριμένα ο εγκέφαλος εκτελεί διαρκώς και ασταμάτητα εργασίες αναγνώρισης που βασίζονται στην αντίληψη (π.χ. αναγνώριση ενός οικείου προσώπου που βρίσκεται σε μία άγνωστη σκηνή) σε χρόνο περίπου 100-200 ms, την ίδια στιγμή που εργασίες πολύ μικρότερου βαθμού πολυπλοκότητας απαιτούν πολύ μεγαλύτερους χρόνους για να εκτελεστούν από έναν ισχυρό υπολογιστή.

Τα πρώτα νευρωνικά μοντέλα είναι αποτέλεσμα της θεμελιώδους έρευνας των W.S. McCulloch και W.A. Pitts αναφορικά με τη θεωρία της νευρωνικής υπολογιστικής κατά τα τέλη της δεκαετίας του 1940, η οποία είχε ξεκινήσει να αναπτύσσεται στα τέλη του 19ου και τις αρχές του 20ου αιώνα μετά τις εξελίξεις στην πληροφορική και την νευροεπιστήμη (de Vos, 2003) . Στη συνέχεια, αναπτύχθηκε η θεωρία για τις τεχνικές των τεχνητών νευρωνικών δικτύων και η πρόοδος που σημειώθηκε οδήγησε στην κατασκευή των πρώτων νευρωνικών υπολογιστών (υπολογιστές των οποίων οι επεξεργαστές λειτουργούν με πρότυπο τα νευρωνικά δίκτυα) τη δεκαετία του 1960. Όμως, διάφοροι περιορισμοί που εντοπίστηκαν, όπως η έλλειψη των υπολογιστικών πόρων και οι ανεπιτυχείς, μέχρι τότε, προσπάθειες για την ανάπτυξη τεχνικών που θα μπορούσαν να επιλύσουν προβλήματα μεγάλης κλίμακας, οδήγησε σε περιορισμένο αριθμό ερευνών στο συγκεκριμένο πεδίο μέχρι τα μέσα της δεκαετίας του 1980, οπότε και αυξήθηκε ξανά το ενδιαφέρον για τη νευρωνική υπολογιστική, κυρίως από τον J.J. Hopfield.

Κατά τα επόμενα χρόνια, πολλοί από τους προηγούμενους περιορισμούς των τεχνητών νευρωνικών δικτύων είχαν ξεπεραστεί και έτσι η εφαρμογή τους ήταν επιτυχής σε πολλά προβλήματα. Από τις τεχνικές που ήρθαν ξανά στο προσκήνιο, η πιο πρωτοποριακή ήταν αυτή της προς τα πίσω διάδοσης (backpropagation), γεγονός που σε συνδυασμό με την αύξηση των υπολογιστικών πόρων οδήγησε στην εκρηκτική ανάπτυξη των τεχνητών νευρωνικών δικτύων. Τα τεχνητά νευρωνικά

δίκτυα χρησιμοποιούνται κυρίως για τη μοντελοποίηση των πολύπλοκων σχέσεων σε συνθήκες όπου η γνώση για το εξεταζόμενο σύστημα δεν είναι επαρκής ώστε να τις περιγράψει μέσω συμβατικών αναλυτικών μοντέλων.

4.2. Νευρωνικά Δίκτυα

Νευρωνικό δίκτυο ονομάζεται ένα κύκλωμα διασυνδεδεμένων νευρώνων. Στην περίπτωση βιολογικών νευρώνων, πρόκειται για ένα τμήμα νευρικού ιστού. Στην περίπτωση τεχνητών νευρώνων, πρόκειται για μία αλγοριθμική δομή η οποία εμπίπτει στον τομέα της υπολογιστικής νοημοσύνης, όταν στόχος του νευρωνικού δικτύου είναι η επίλυση κάποιου υπολογιστικού προβλήματος, ή της υπολογιστικής νευροεπιστήμης, όταν στόχος είναι η υπολογιστική προσομοίωση της λειτουργίας των βιολογικών νευρωνικών δικτύων με βάση κάποιο μαθηματικό μοντέλο τους. Ένα νευρωνικό δίκτυο είναι ένας τεράστιος παράλληλος επεξεργαστής με κατανομημένη αρχιτεκτονική, ο οποίος αποτελείται από απλές μονάδες επεξεργασίας και έχει από τη φύση του τη δυνατότητα να αποθηκεύει εμπειρική γνώση και να την καθιστά διαθέσιμη για χρήση. Η αλγοριθμική δομή την οποία περιγράφουμε μοιάζει με τον ανθρώπινο εγκέφαλο (νευρικό ιστό) σε δύο σημεία:

1. Το δίκτυο προσλαμβάνει τη γνώση από το περιβάλλον του, μέσω μιας διαδικασίας μάθησης.
2. Η ισχύς των συνδέσεων μεταξύ των νευρώνων, που αποκαλείται συναπτικό βάρος, χρησιμοποιείται για την αποθήκευση της γνώσης που αποκτιέται.

Η διαδικασία μέσω της οποίας επιτυγχάνεται η μάθηση αποκαλείται αλγόριθμος μάθησης και η λειτουργία του είναι να τροποποιεί τα συναπτικά βάρη του δικτύου με τον κατάλληλο τρόπο για την επίτευξη του επιθυμητού στόχου. Η τροποποίηση των συναπτικών βαρών αποτελεί την «παραδοσιακή» μέθοδο σχεδιασμού νευρωνικών δικτύων. Αυτή η προσέγγιση βρίσκεται πλησιέστερα στη θεωρία γραμμικών προσαρμοστικών φίλτρων, η οποία είναι από καιρό καθιερωμένη και εφαρμόζεται επιτυχώς σε πολλά διαφορετικά πεδία (Widrow & Stearns, 1985. Haykin, 2002). Ωστόσο, ένα νευρωνικό δίκτυο έχει επίσης τη δυνατότητα να τροποποιεί την τοπολογία του, και αυτό επειδή κάποιοι νευρώνες του ανθρώπινου εγκεφάλου μπορεί να «πεθάνουν», ενώ μπορούν επίσης να αναπτύσσονται νέες συναπτικές συνδέσεις.

4.2.1. Πλεονεκτήματα των Νευρωνικών Δικτύων

Είναι προφανές ότι ένα νευρωνικό δίκτυο οφείλει την υπολογιστική ισχύ του κατά πρώτον στην παράλληλη, κατανομημένη δομή του και κατά δεύτερον στην ικανότητά του να μαθαίνει και, ως εκ τούτου, να γενικεύει. Ο όρος γενίκευση αναφέρεται στην παραγωγή, από το νευρωνικό δίκτυο, λογικών εξόδων για εισόδους τις οποίες δεν έχει συναντήσει κατά τη διάρκεια της εκπαίδευσής του. Αυτές οι δύο δυνατότητες δίνουν στα νευρωνικά δίκτυα τη δυνατότητα να βρίσκουν καλές

προσεγγιστικές λύσεις σε πολύπλοκα (μεγάλης κλίμακας) προβλήματα, τα οποία είναι μη επιδεκτικά σε λύσεις. Στην πράξη, ωστόσο, τα νευρωνικά δίκτυα δεν μπορούν να παρέχουν τη λύση δουλεύοντας ατομικά. Απεναντίας, χρειάζεται να ενταχθούν σε μία ευρύτερη αλλά συνεπή προσέγγιση ανάπτυξης ενός συστήματος. Συγκεκριμένα, ένα πολύπλοκο πρόβλημα αποσυντίθεται σε έναν αριθμό σχετικά απλών εργασιών και τα νευρωνικά δίκτυα αναλαμβάνουν ένα υποσύνολο των εργασιών που ταιριάζουν με τις εγγενείς δυνατότητές τους.

Οι ιδιότητες και οι δυνατότητες των νευρωνικών δικτύων είναι οι εξής:

1. Μη Γραμμικότητα.

Ένας τεχνητός νευρώνας μπορεί να είναι είτε γραμμικός, είτε μη γραμμικός. Ένα νευρωνικό δίκτυο αποτελούμενο από διασυνδεδεμένους μη γραμμικούς νευρώνες είναι, από τη φύση του, μη γραμμικό. Επιπλέον, αυτή η μη γραμμικότητα είναι «ειδικού τύπου», υπο την έννοια ότι είναι κατανεμημένη σε όλη την έκταση του δικτύου. Η μη γραμμικότητα είναι μία εξαιρετικά σημαντική ιδιότητα, κυρίως αν ο υποκειμενικός φυσικός μηχανισμός που είναι υπεύθυνος για την παραγωγή του σήματος εξόδου (π.χ. ομιλία) είναι εκ φύσεως μη γραμμικός.

2. Αντιστοίχιση Εισόδου εξόδου.

Στο παράδειγμα της επιβλεπόμενης μάθησης έχουμε τροποποίηση των συναπτικών βαρών ενός νευρωνικού δικτύου εφαρμόζοντας ένα σύνολο χαρακτηρισμένων παραδειγμάτων εκπαίδευσης, ή παραδειγμάτων εργασιών. Κάθε παράδειγμα αποτελείται από ένα μοναδικό σήμα εισόδου και μία αντίστοιχη επιθυμητή απόκριση (στόχος). Στο δίκτυο παρουσιάζεται ένα παράδειγμα επιλεγμένο τυχαία από το σύνολο και συναπτικά βάρη (ελεύθερες παράμετροι) του δικτύου τροποποιούνται ώστε να ελαχιστοποιηθεί η διαφορά μεταξύ της επιθυμητής απόκρισης και της πραγματικής απόκρισης του δικτύου που παράγεται από το σήμα εισόδου, σύμφωνα με ένα κατάλληλο στατιστικό κριτήριο. Η εκπαίδευση του δικτύου επαναλαμβάνεται για πολλά παραδείγματα του συνόλου εκπαίδευσης, έως ότου το δίκτυο φτάσει σε μία ευσταθή κατάσταση, όπου δεν υπάρχουν περεταίρω σημαντικές μεταβολές στα συναπτικά βάρη. Τα προηγούμενα εφαρμοσθέντα παραδείγματα εκπαίδευσης θα μπορούσαν να εφαρμοστούν εκ νέου κατά τη διάρκεια της εκπαίδευσης, αλλά με διαφορετική σειρά. Έτσι, το δίκτυο μαθαίνει από τα παραδείγματα, κατασκευάζοντας μια αντιστοίχιση εισόδου-εξόδου για το δοθέν πρόβλημα.

3. Προσαρμοστικότητα

Τα νευρωνικά δίκτυα έχουν την εγγενή δυνατότητα να προσαρμόζουν τα συναπτικά βάρη τους ανάλογα με τις μεταβολές που γίνονται στο περιβάλλον τους. Συγκεκριμένα, ένα νευρωνικό δίκτυο εκπαιδευμένο να λειτουργεί σε συγκεκριμένο περιβάλλον, μπορεί εύκολα να εκπαιδευτεί εκ νέου ώστε να χειρίζεται ήσσονος σημασίας μεταβολές στις συνθήκες του περιβάλλοντος λειτουργίας του. Επιπλέον, όταν λειτουργεί σε ένα μη στατικό περιβάλλον (δηλαδή ένα περιβάλλον του οποίου τα στατιστικά στοιχεία μεταβάλλονται

με το χρόνο), ένα νευρωνικό δίκτυο μπορεί να σχεδιαστεί ώστε να μεταβάλλει τα συναπτικά βάρη του σε πραγματικό χρόνο. Η φυσική αρχιτεκτονική ενός νευρωνικού δικτύου για ταξινόμηση προτύπων, επεξεργασία σήματος και εφαρμογές ελέγχου, σε συνδυασμό με την προσαρμοστική δυνατότητα του δικτύου, το καθιστά χρήσιμο εργαλείο για την προσαρμοστική ταξινόμηση προτύπων, την προσαρμοστική επεξεργασία σήματος και τον προσαρμοστικό έλεγχο συστημάτων. Σαν γενικό κανόνα θα μπορούσαμε να πούμε ότι όσο πιο προσαρμοστικό κάνουμε ένα σύστημα, διασφαλίζοντας ταυτόχρονα ότι παραμένει διαρκώς σταθερό, τόσο πιο εύρωστο θα είναι και τόσο καλύτερα θα αποδίδει όταν κληθεί να λειτουργήσει σε ένα μη σταθερό περιβάλλον.

4. Ενδεικτική Απόκριση

Στο πλαίσιο της ταξινόμησης προτύπων, ένα νευρωνικό δίκτυο μπορεί να σχεδιαστεί ώστε να παρέχει πληροφορία όχι μόνο για το ποιο συγκεκριμένο πρότυπο θα επιλεγεί, αλλά επίσης σχετικά με το βαθμό εμπιστοσύνης στη ληφθείσα απόφαση. Αυτή η δεύτερη πληροφορία μπορεί να χρησιμοποιηθεί για την απόρριψη των διφορούμενων μοτίβων, αν προκύψουν, και κατ' επέκταση τη βελτίωση της απόδοσης του δικτύου.

5. Πληροφορία Σχετική με το Περιεχόμενο

Η γνώση αντιπροσωπεύεται από την ίδια τη δομή και την κατάσταση ενεργοποίησης ενός νευρωνικού δικτύου. Κάθε νευρώνας στο δίκτυο ενδεχομένως να επηρεάζεται από τη συνολική δραστηριότητα όλων των άλλων νευρώνων του δικτύου. Αυτό σημαίνει ότι ένα νευρωνικό δίκτυο χειρίζεται με φυσικό τρόπο την σχετική με το περιεχόμενο πληροφορία (contextual information).

6. Ανοχή σε Βλάβες

Ένα νευρωνικό δίκτυο, υλοποιημένο σε μορφή hardware, έχει την εγγενή δυνατότητα να είναι ανεκτικό σε βλάβες, ή εύρωστο, υπό την έννοια ότι η απόδοσή του μειώνεται βαθμιαία και ομαλά υπό αντίξοες συνθήκες λειτουργίας. Για παράδειγμα, αν ένας νευρώνας ή οι συνδέσεις του καταστραφούν, η ποιότητα της ανάκλησης ενός αποθηκευμένου εκεί προτύπου μειώνεται. Ωστόσο, λόγω της κατανεμημένης φύσης της πληροφορίας που αποθηκεύεται στο δίκτυο, μία τέτοια βλάβη θα πρέπει να λάβει μεγάλη έκταση πριν αρχίσει να μειώνεται σοβαρά η συνολική απόκριση του δικτύου. Έτσι, σε επίπεδο αρχών, ένα νευρωνικό δίκτυο επιδεικνύει ομαλή μείωση στην απόδοση και όχι καταστροφική αποτυχία. Υπάρχουν ορισμένα εμπειρικά στοιχεία για την εκτίμηση της ευρωστίας, αλλά συνήθως δεν είναι ελέγξιμη. Για να διασφαλιστεί ότι ένα νευρωνικό δίκτυο είναι πράγματι ανεκτικό σε βλάβες, μπορεί να χρειαστεί να ληφθούν διορθωτικά μέτρα στη σχεδίαση του αλγορίθμου που χρησιμοποιείται για την εκπαίδευση του δικτύου (Kerlirzin & Vallet, 1993).

7. Δυνατότητα υλοποίησης σε VLSI

Η μαζικά παράλληλη φύση ενός νευρωνικού δικτύου το καθιστά ενδεχομένως γρήγορο για τον υπολογισμό συγκεκριμένων εργασιών. Αυτό το χαρακτηριστικό καθιστά επίσης ένα νευρωνικό δίκτυο ιδιαίτερα κατάλληλο για υλοποίηση με χρήση τεχνολογίας πολύ μεγάλης κλίμακας ολοκλήρωσης (VLSI). Ένα συγκεκριμένο πλεονέκτημα της τεχνολογίας VLSI είναι ότι παρέχει ένα μέσο «σύλληψης» πραγματικά πολύπλοκης συμπεριφοράς με εξαιρετικά ιεραρχικό τρόπο.

8. Ομοιομορφία Ανάλυσης και Σχεδίασης

Ουσιαστικά τα νευρωνικά δίκτυα απολαμβάνουν καθολικής αποδοχής ως επεξεργαστές πληροφοριών, υπό την έννοια ότι χρησιμοποιείται η ίδια σημειογραφία σε όλα τα πεδία εφαρμογής των νευρωνικών δικτύων. Αυτό το χαρακτηριστικό εκδηλώνεται με διάφορους τρόπους:

- i. Οι νευρώνες, σε οποιαδήποτε μορφή, αντιπροσωπεύουν ένα συστατικό κοινό σε όλα τα νευρωνικά δίκτυα.
- ii. Αυτός ο «κοινός» χαρακτήρας καθιστά εφικτή τη χρήση των ίδιων θεωριών και αλγορίθμων μάθησης σε διαφορετικές εφαρμογές των νευρωνικών δικτύων.
- iii. «Σπονδυλωτά» δίκτυα μπορούν να κατασκευάζονται με απρόσκοπτη ενοποίηση επιμέρους λειτουργικών μονάδων (modules).

9. Αναλογία με τη Νευροφυσιολογία του Εγκεφάλου

Η σχεδίαση ενός νευρωνικού δικτύου δανείζεται στοιχεία από τη λειτουργία του ανθρώπινου εγκεφάλου, ο οποίος είναι η ζωντανή απόδειξη ότι η εύρωστη, παράλληλη επεξεργασία δεν είναι μόνο φυσικά εφικτή, αλλά επίσης γρήγορη και ισχυρή. Οι νευροβιολόγοι αντιμετωπίζουν τα (τεχνητά) νευρωνικά δίκτυα ως ένα ερευνητικό εργαλείο για την ερμηνεία νευροβιολογικών φαινομένων. Από την άλλη, οι μηχανικοί αναζητούν στον τομέα της νευροβιολογίας νέες ιδέες για την επίλυση προβλημάτων που είναι πολύ πιο πολύπλοκα από αυτά που βασίζονται στις συμβατικές τεχνικές σχεδίασης.

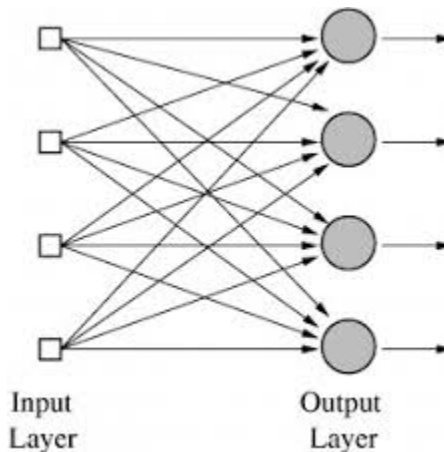
4.2.2. Αρχιτεκτονικές Νευρωνικών Δικτύων

Ο τρόπος με τον οποίο είναι δομημένοι οι νευρώνες ενός νευρωνικού δικτύου σχετίζεται στενά με τον αλγόριθμο μάθησης που χρησιμοποιείται για την εκπαίδευση του δικτύου. Άρα, μπορούμε να αντιμετωπίσουμε τους αλγορίθμους μάθησης (κανόνες) που χρησιμοποιούνται στη σχεδίαση νευρωνικών δικτύων ως δομημένους. Αναφορά στην κατηγοριοποίηση των αλγορίθμων μάθησης θα γίνει στη συνέχεια.

Μπορούμε να αναγνωρίσουμε τρεις θεμελιωδώς διαφορετικές κατηγορίες αρχιτεκτονικών νευρωνικών δικτύων:

1. Ενός Επιπέδου Δίκτυα Πρόσθιας Τροφοδότησης

Σε ένα νευρωνικό δίκτυο οι νευρώνες οργανώνονται σε μορφή επιπέδων. Στην απλούστερη μορφή ενός δικτύου, έχουμε ένα επίπεδο εισόδου (αποτελούμενο από πηγαίους κόμβους) το οποίο συνδέεται απευθείας με ένα επίπεδο νευρώνων εξόδου (υπολογιστικοί κόμβοι), αλλά όχι αντίστροφα. Με άλλα λόγια, αυτό το δίκτυο είναι τύπου πρόσθιας τροφοδότησης (feedforward). Ένα τέτοιο δίκτυο απεικονίζεται στη συνέχεια. Το δίκτυο αυτό αποκαλείται δίκτυο ενός επιπέδου, με το χαρακτηρισμό «ενός επιπέδου» να αναφέρεται στο επίπεδο εξόδου των υπολογιστικών κόμβων. Το επίπεδο εισόδου (των πηγαίων κόμβων) δεν προσμετράται επειδή δεν εκτελείται κανένας υπολογισμός σε αυτό.

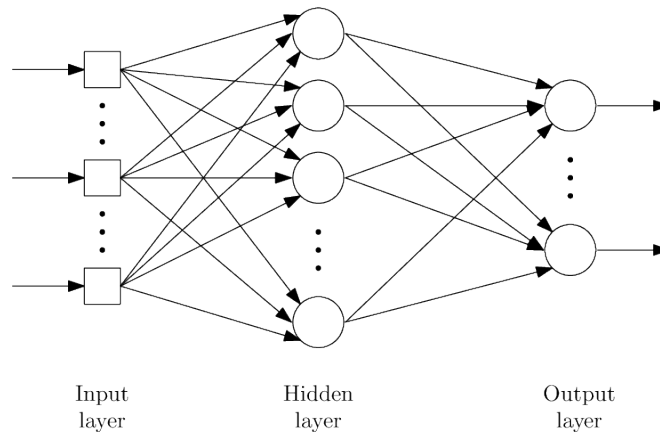


Σχήμα 4.1 Νευρωνικό Δίκτυο Ενός Επιπέδου

2. Πολυεπίπεδα Δίκτυα Πρόσθιας Τροφοδότησης

Η δεύτερη κατηγορία νευρωνικών δικτύων πρόσθιας τροφοδότησης χαρακτηρίζεται από την παρουσία ενός ή περισσότερων κρυφών επιπέδων, των οποίων οι υπολογιστικοί κόμβοι αποκαλούνται *κρυφοί νευρώνες* ή *κρυφές μονάδες*. Ο όρος «κρυφός» αναφέρεται στο γεγονός ότι αυτό το μέρος του νευρωνικού δικτύου δεν είναι άμεσα ορατό ούτε από την είσοδο, ούτε από την έξοδο του δικτύου. Η λειτουργία των κρυφών νευρώνων είναι να παρεμβαίνουν μεταξύ της εξωτερικά προερχόμενης εισόδου και της εξόδου του δικτύου με κάποιο χρήσιμο τρόπο. Προσθέτοντας ένα ή περισσότερα κρυφά επίπεδα, το δίκτυο έχει τη δυνατότητα να εξαγάγει στατιστικά υψηλότερης τάξης από την είσοδό του. Υπό μία χαλαρή έννοια το δίκτυο αποκτά μία γενική προοπτική, παρά τον τοπικό χαρακτήρα των συνδέσεων του, λόγω του επιπλέον συνόλου συναπτικών συνδέσεων και της επιπλέον διάστασης νευρωνικών αλληλεπιδράσεων (Churchland & Sejnowski, 1992). Οι πηγαίοι κόμβοι στο επίπεδο εισόδου του δικτύου παρέχουν τα αντίστοιχα στοιχεία του προτύπου ενεργοποίησης (διάνυσμα εισόδου), τα οποία

συνιστούν τα σήματα εισόδου που εφαρμόζονται στους νευρώνες (υπολογιστικούς κόμβους) του δεύτερου επιπέδου (δηλαδή του πρώτου κρυφού επιπέδου). Τα σήματα εξόδου του δεύτερου επιπέδου χρησιμοποιούνται ως εισόδοι στο τρίτο επίπεδο, και αυτό συνεχίζεται για όλα τα υπόλοιπα επίπεδα του δικτύου. Τυπικά, οι νευρώνες σε κάθε επίπεδο του δικτύου έχουν ως εισόδους τα σήματα εξόδου μόνο του προηγούμενου επιπέδου. Το σύνολο των σημάτων εξόδου των νευρώνων στο επίπεδο εξόδου (το τελικό επίπεδο) του δικτύου συνιστά τη συνολική απόκριση του δικτύου στο πρότυπο ενεργοποίησης που παρέχεται από τους πηγαίους κόμβους στο επίπεδο εισόδου (πρώτο επίπεδο του δικτύου). Χάρην συντομίας, ένα δίκτυο πρόσθιας τροφοδότησης με m πηγαίους κόμβους, h_1 νευρώνες στο πρώτο κρυφό επίπεδο, h_2 νευρώνες στο δεύτερο επίπεδο και q νευρώνες στο επίπεδο εξόδου, αναφέρεται ως δίκτυο $m-h_1-h_2-q$.

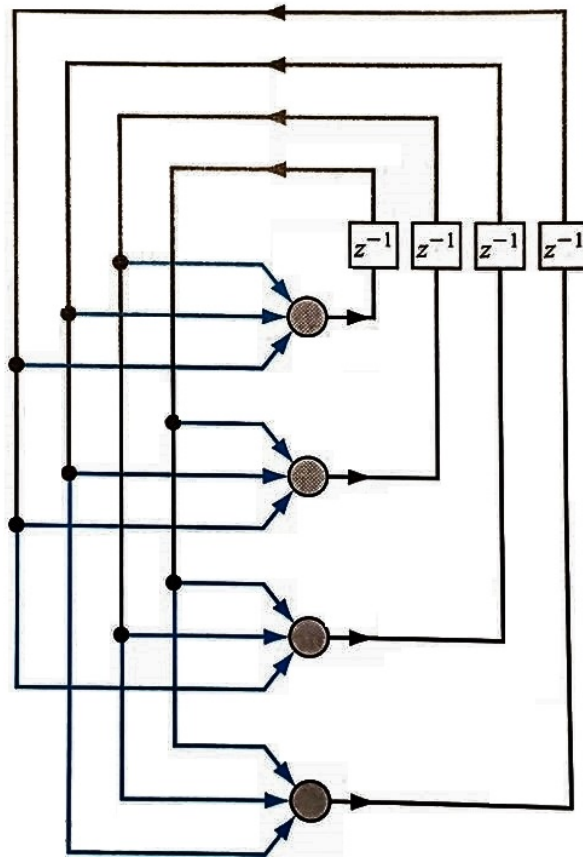


Σχήμα 4.2 Νευρωνικό Δίκτυο Πολλαπλών Επιπέδων

Στο Σχήμα 4.2 απεικονίζεται ένα νευρωνικό δίκτυο πολλαπλών επιπέδων. Το νευρωνικό δίκτυο αυτό λέγεται ότι είναι *πλήρως συνδεδεμένο* υπό την έννοια ότι κάθε κόμβος σε κάθε επίπεδο του δικτύου συνδέεται με κάθε άλλο κόμβο του επόμενου (προς τα εμπρός) επιπέδου. Αν, όμως, λείπουν ορισμένες συνδέσεις επικοινωνίας (συναπτικές συνδέσεις) από το δίκτυο, τότε λέμε ότι είναι *μερικώς συνδεδεμένο*.

3. Αναδρομικά Δίκτυα

Ένα αναδρομικό νευρωνικό δίκτυο (recurrent neural network) διαφέρει από ένα νευρωνικό δίκτυο πρόσθιας τροφοδότησης στο ότι έχει τουλάχιστον ένα *βρόχο ανάδρασης*. Για παράδειγμα, ένα αναδρομικό δίκτυο μπορεί να αποτελείται από ένα μεμονωμένο επίπεδο νευρώνων, με κάθε νευρώνα να τροφοδοτεί το σήμα εξόδου του πίσω, στις εισόδους όλων των άλλων νευρώνων, όπως απεικονίζει ο αρχιτεκτονικός γράφος στο Σχήμα 4.3. Στη δομή που απεικονίζεται σε αυτό το σχήμα, δεν υπάρχουν βρόχοι αυτο-ανάδρασης στο δίκτυο (η αυτό ανάδραση αναφέρεται σε μία κατάσταση όπου η έξοδος ενός νευρώνα επανατροφοδοτείται στην είσοδο του ίδιου νευρώνα). Το αναδρομικό δίκτυο στο Σχήμα 4.3 δεν έχει κρυφούς νευρώνες.



Σχήμα 4.3 Αναδρομικό Νευρωνικό Δίκτυο

4.3. Μηχανική Μάθηση

Ακριβώς όπως υπάρχουν διαφορετικοί τρόποι με τους οποίους μαθαίνουν οι άνθρωποι από το περιβάλλον τους, το ίδιο ισχύει και για τα νευρωνικά δίκτυα. Υπό μία ευρεία έννοια, μπορούμε να κατηγοριοποιήσουμε τις διαδικασίες μάθησης μέσω των οποίων λειτουργούν τα νευρωνικά δίκτυα ως εξής: μάθηση με εκπαιδευτή (supervised learning) και μάθηση χωρίς εκπαιδευτή. Με το ίδιο κριτήριο η δεύτερη μορφή μάθησης μπορεί να κατηγοριοποιηθεί περαιτέρω σε μη επιβλεπόμενη μάθηση (unsupervised learning) και ενισχυτική μάθηση. Αυτές οι διαφορετικές μορφές μάθησης, όπως λαμβάνουν χώρα στα νευρωνικά δίκτυα, προσομοιώνουν την ανθρώπινη μάθηση.

4.3.1. Μάθηση με εκπαιδευτή

Η μάθηση με εκπαιδευτή αναφέρεται επίσης ως *επιβλεπόμενη μάθηση*. Με εννοιολογικούς όρους μπορούμε να θεωρήσουμε ότι ο εκπαιδευτής έχει γνώση του περιβάλλοντος, και αυτή η γνώση αντιπροσωπεύεται από ένα σύνολο παραδειγμάτων εισόδου-εξόδου. Ωστόσο, το περιβάλλον είναι άγνωστο στο νευρωνικό δίκτυο. Ας υποθέσουμε τώρα ότι ο εκπαιδευτής και το νευρωνικό δίκτυο εκτίθενται σε ένα διάνυσμα εκπαίδευσης (δηλ. ένα παράδειγμα) αντλούμενο από το ίδιο περιβάλλον. Χάρη στην εγγενή του γνώση, ο εκπαιδευτής είναι σε θέση να παρέχει στο νευρωνικό δίκτυο μία επιθυμητή απόκριση για το συγκεκριμένο διάνυσμα εκπαίδευσης.

Η επιθυμητή απόκριση αντιπροσωπεύει τη «βέλτιστη» ενέργεια που πρέπει να εκτελείται από το νευρωνικό δίκτυο. Οι παράμετροι του δικτύου προσαρμόζονται υπό τη συνδυασμένη επιρροή του διανύσματος εκπαίδευσης και του σήματος σφάλματος. Το *σήμα σφάλματος* ορίζεται ως η διαφορά μεταξύ της επιθυμητής απόκρισης και της πραγματικής απόκρισης του δικτύου. Αυτή η προσαρμογή εκτελείται με επαναληπτικό τρόπο, βήμα προς βήμα, με στόχο να φέρει τελικά το νευρωνικό δίκτυο σε μία κατάσταση που θα προσομοιώνει τη συμπεριφορά του εκπαιδευτή. Η προσομοίωση αυτή κρίνεται ως «βέλτιστη» με κάποια στατιστική έννοια. Κατ' αυτόν τον τρόπο, η γνώση του περιβάλλοντος που είναι διαθέσιμη στον εκπαιδευτή μεταφέρεται στο νευρωνικό δίκτυο μέσω εκπαίδευσης και αποθηκεύεται με τη μορφή σταθερών συναπτικών βαρών, τα οποία αντιπροσωπεύουν *μακροπρόθεσμη μνήμη*. Όταν επιτευχθεί αυτή η συνθήκη, μπορούμε να απαλλαγούμε από τον εκπαιδευτή και να αφήσουμε το νευρωνικό δίκτυο να αντιμετωπίσει το περιβάλλον εντελώς μόνο του.

Η μορφή της επιβλεπόμενης μάθησης που μόλις περιεγράφηκε αποτελεί τη βάση της *μάθησης μέσω διόρθωσης σφάλματος*. Η διαδικασία επιβλεπόμενης ένα σύστημα ανάδρασης κλειστού βρόχου, αλλά το άγνωστο περιβάλλον βρίσκεται εκτός του βρόχου αυτού. Ως μέτρο απόδοσης για το σύστημα μπορούμε να χρησιμοποιήσουμε το μέσο τετραγωνικό σφάλμα ή το άθροισμα των τετραγώνων των σφαλμάτων επί του δείγματος εκπαίδευσης, ορισμένο ως συνάρτηση των ελεύθερων παραμέτρων (δηλ. των συναπτικών βαρών) του συστήματος. Αυτή η λειτουργία μπορεί να σχηματιστεί ως μια πολυδιάστατη *επιφάνεια σφάλματος-απόδοσης*, ή απλώς *επιφάνεια σφάλματος*, με τις ελεύθερες παραμέτρους σαν συντεταγμένες. Η πραγματική επιφάνεια σφάλματος υπολογίζεται ως μέσος όρος επί όλων των πιθανών παραδειγμάτων εισόδου-εξόδου. Οποιαδήποτε συγκεκριμένη λειτουργία του συστήματος υπό την επίβλεψη του εκπαιδευτή αναπαρίσταται ως ένα σημείο στην επιφάνεια σφάλματος. Για να μπορεί το σύστημα να βελτιώνει την απόδοσή του με την πάροδο του χρόνου και, κατά συνέπεια, να μαθαίνει από τον εκπαιδευτή, το σημείο λειτουργίας του πρέπει να μετακινείται διαδοχικά προς τα κάτω προς ένα ελάχιστο σημείο της επιφάνειας σφάλματος. Το ελάχιστο σημείο μπορεί να είναι ένα τοπικό ελάχιστο ή ένα γενικό ελάχιστο. Ένα σύστημα επιβλεπόμενης μάθησης έχει τη δυνατότητα να το κάνει αυτό με τη χρήσιμη πληροφορία που διαθέτει σχετικά με την κλίση (gradient) της επιφάνειας σφάλματος που αντιστοιχεί στην τρέχουσα συμπεριφορά του συστήματος. Η κλίση

της επιφάνειας σφάλματος σε οποιοδήποτε σημείο είναι ένα διάνυσμα το οποίο δείχνει προς την κατεύθυνση της *πλέον απότομης κατάβασης* (steepest descent). Στην πραγματικότητα, στην περίπτωση της επιβλεπόμενης μάθησης μέσω παραδειγμάτων, το σύστημα μπορεί να χρησιμοποιεί μία στιγμιαία εκτίμηση του διανύσματος κλίσης, με τους δείκτες των παραδειγμάτων να είναι χρονικοί. Η χρήση μίας τέτοιας εκτίμησης έχει ως αποτέλεσμα μία μετακίνηση του σημείου λειτουργίας πάνω στην επιφάνεια σφάλματος, η οποία τυπικά έχει τη μορφή «τυχαίου περιπάτου». Παρά ταύτα με δεδομένα έναν αλγόριθμο ειδικά σχεδιασμένο ώστε να ελαχιστοποιεί τη συνάρτηση κόστους, ένα επαρκές σύνολο παραδειγμάτων εισόδου-εξόδου και επαρκή χρόνο για την εκπαίδευση, ένα σύστημα επιβλεπόμενης μάθησης είναι συνήθως σε θέση να προσεγγίσει μία άγνωστη αντιστοίχιση εισόδου-εξόδου εύλογα καλά.

4.3.2. Μάθηση χωρίς εκπαιδευτή

Στην επιβλεπόμενη μάθηση, η διαδικασία μάθησης λαμβάνει χώρα υπό την καθοδήγηση ενός εκπαιδευτή. Ωστόσο, στο παράδειγμα που είναι γνωστό ως μάθηση χωρίς εκπαιδευτή, όπως υποδηλώνει το όνομά του, δεν υπάρχει εκπαιδευτής που να επιβλέπει τη διαδικασία μάθησης. Δηλαδή, δεν υπάρχουν χαρακτηρισμένα παραδείγματα της λειτουργίας που πρέπει να μάθει το δίκτυο. Στα πλαίσια αυτού του δεύτερου παραδείγματος, μπορούμε να ορίσουμε δύο υποκατηγορίες μάθησης:

4.3.2.α. Ενισχυτική Μάθηση

Στην ενισχυτική μάθηση, η εκμάθηση μίας αντιστοίχισης εισόδου-εξόδου εκτελείται μέσω συνεχούς αλληλεπίδρασης με το περιβάλλον, με στόχο την ελαχιστοποίηση ενός βαθμωτού δείκτη απόδοσης. Το σύστημα σχεδιάζεται ώστε να μαθαίνει βάσει καθυστερούμενης ενίσχυσης, πράγμα το οποίο σημαίνει ότι το σύστημα παρατηρεί μία χρονική ακολουθία ερεθισμάτων που λαμβάνει από το περιβάλλον τα οποία καταλήγουν στην παραγωγή του ευρετικού σήματος ενίσχυσης.

Ο στόχος της ενισχυτικής μάθησης είναι να ελαχιστοποιεί μία συνάρτηση τρέχοντος κόστους, η οποία ορίζεται ως πρόβλεψη του αθροιστικού κόστους ενεργειών που εκτελούνται σε μία αλληλουχία βημάτων αντί απλώς του άμεσου κόστους μιας ενέργειας. Ενδεχομένως ορισμένες από τις ενέργειες που έχουν εκτελεστεί σε αυτήν την αλληλουχία χρονικών βημάτων να είναι οι καλύτερες ορίζουσες της συνολικής συμπεριφοράς του συστήματος. Η λειτουργία του συστήματος μάθησης είναι να ανακαλύψει αυτές τις ενέργειες και να τις τροφοδοτήσει πίσω, στο περιβάλλον.

Η ενισχυτική μάθηση με καθυστέρηση είναι δύσκολο να εκτελεστεί για δύο βασικούς λόγους:

- Δεν υπάρχει εκπαιδευτής για να παρέχει μία επιθυμητή απόκριση σε κάθε βήμα της διαδικασίας μάθησης.
- Η καθυστέρηση με την οποία παράγεται το κύριο σήμα ενίσχυσης υποδηλώνει ότι η μανθάνουσα μηχανή πρέπει να λύσει ένα χρονικό πρόβλημα ανάθεσης εμπιστοσύνης. Με αυτό εννοούμε ότι η μηχανή πρέπει να είναι σε θέση να καθορίζει το βαθμό επιτυχίας ατομικά για κάθε ενέργεια της χρονικής αλληλουχίας βημάτων που οδήγησαν στο τελικό αποτέλεσμα, ενώ ο κύριος μηχανισμός ενίσχυσης μπορεί να αποτιμά μόνο το τελικό αποτέλεσμα.

Σε πείσμα αυτών των δυσκολιών, η ενισχυτική μάθηση με καθυστέρηση είναι μία ελκυστική μέθοδος. Παρέχει στο σύστημα μάθησης μία βάση για να επικοινωνεί με το περιβάλλον του, αναπτύσσοντας έτσι την δυνατότητα να μάθει να εκτελεί μια προκαθορισμένη εργασία βασιζόμενο αποκλειστικά στα αποτελέσματα της εμπειρίας του από την αλληλεπίδραση με το περιβάλλον.

4.3.2.β. Μη Επιβλεπόμενη Μάθηση

Στη μη επιβλεπόμενη, ή αυτο-οργανούμενη, μάθηση, δεν υπάρχει εξωτερικός εκπαιδευτής ή κριτής που να επιβλέπει την διαδικασία μάθησης. Αντ' αυτού, υπάρχει ένα ανεξάρτητο από την εργασία μέτρο της ποιότητας της αναπαράστασης που καλείται να μάθει το δίκτυο και οι ελεύθερες παράμετροι του δικτύου βελτιστοποιούνται σε σχέση με αυτό το μέτρο. Για ένα συγκεκριμένο ανεξάρτητο από την εργασία μέτρο, αφού το δίκτυο «συντονιστεί» στις στατιστικές κανονικότητες των δεδομένων εισόδου, αναπτύσσει την δυνατότητα να σχηματίζει εσωτερικές αναπαραστάσεις για την κωδικοποίηση χαρακτηριστικών της εισόδου και, μέσω αυτών, να δημιουργεί νέες κλάσεις αυτόματα (Becker, 1991).

Για την εκτέλεση μη επιβλεπόμενης μάθησης, μπορούμε να χρησιμοποιήσουμε έναν κανόνα ανταγωνιστική μάθησης. Για παράδειγμα, μπορούμε να χρησιμοποιήσουμε ένα νευρωνικό δίκτυο το οποίο αποτελείται από δύο επίπεδα – ένα επίπεδο εισόδου και ένα ανταγωνιστικό επίπεδο. Το επίπεδο εισόδου λαμβάνει τα διαθέσιμα δεδομένα. Το ανταγωνιστικό επίπεδο αποτελείται από νευρώνες οι οποίοι ανταγωνίζονται ο ένας τον άλλο (σύμφωνα με έναν κανόνα μάθησης) για την ευκαιρία να ανταποκριθούν σε χαρακτηριστικά που περιέχονται στα δεδομένα εισόδου. Στην απλούστερη δυνατή μορφή του, το δίκτυο λειτουργεί σύμφωνα με μία στρατηγική «ο νικητής παίρνει τα πάντα» (“winner takes it all”). Βάσει μιας τέτοιας στρατηγικής, ο νευρώνας με τη μεγαλύτερη συνολική είσοδο «νικά» στον ανταγωνισμό και ενεργοποιείται, ενώ όλοι οι άλλοι νευρώνες του δικτύου απενεργοποιούνται.

4.3.3. Εργασίες Μάθησης

Στην προηγούμενη ενότητα εξετάσαμε διάφορα παραδείγματα μάθησης. Σε αυτήν την ενότητα θα περιγράψουμε ορισμένες βασικές εργασίες μάθησης. Η επιλογή ενός συγκεκριμένου κανόνα μάθησης, προφανώς, επηρεάζεται από τις εργασίες μάθησης, η ποικιλόμορφη φύση των οποίων καταδεικνύει την καθολικότητα εφαρμογής των νευρωνικών δικτύων.

4.3.3.α. Συσχέτιση Προτύπων

Μία συσχετιστική μνήμη είναι μία μορφή κατανεμημένης μνήμης παρόμοια με αυτή του ανθρώπινου εγκεφάλου, η οποία μαθαίνει μέσω συσχέτισης (association). Η συσχέτιση είναι γνωστό ότι αποτελεί κυρίαρχο χαρακτηριστικό της ανθρώπινης μνήμης από την εποχή του Αριστοτέλη, και όλα τα γνωστικά μοντέλα τη χρησιμοποιούν, με κάποια μορφή, ως βασική λειτουργία.

Η «Συσχέτιση» μπορεί να λάβει μία από τις εξής δύο μορφές: *αυτο-συσχέτιση* και *ετερο-συσχέτιση*. Με την αυτο-συσχέτιση (autoassociation), ένα νευρωνικό δίκτυο καλείται να αποθηκεύσει ένα σύνολο προτύπων (διάνυσματα), τα οποία παρουσιάζονται κατ' επανάληψη στο δίκτυο. Στη συνέχεια παρουσιάζεται στο δίκτυο μία ημιτελής περιγραφή ή παραμορφωμένη (με θόρυβο) έκδοση ενός εκ των αρχικών προτύπων που έχει αποθηκεύσει και το ζητούμενο είναι να ανακτήσει (να ανακαλέσει) αυτό το συγκεκριμένο πρότυπο από τη μνήμη του. Η ετερο-συσχέτιση (heteroassociation) διαφέρει από την αυτο-συσχέτιση στο ότι ένα τυχαίο σύνολο προτύπων εισόδου συνδυάζεται με ένα άλλο τυχαίο σύνολο προτύπων εξόδου. Η αυτο-συσχέτιση προϋποθέτει τη χρήση μη επιβλεπόμενης μάθησης, ενώ στην ετερο-συσχέτιση η μάθηση είναι επιβλεπόμενη.

Έστω ότι το x_k είναι ένα πρότυπο-κλειδί (διάνυσμα) που εφαρμόζεται σε μία συσχετιστική μνήμη και το y_k είναι ένα απομνημονευμένο πρότυπο (διάνυσμα). Η συσχέτιση προτύπων που εκτελείται από το δίκτυο περιγράφεται από την σχέση:

$$x_k \rightarrow y_k, \quad k = 1, 2, \dots, q \quad (1)$$

όπου q είναι ο αριθμός των προτύπων που έχουν αποθηκευτεί στο δίκτυο. Το πρότυπο-κλειδί x_k δρα ως ερέθισμα, το οποίο όχι μόνο καθορίζει τη θέση αποθήκευσης του απομνημονευμένου προτύπου y_k , αλλά επίσης «κρατά το κλειδί» για την ανάκτησή του.

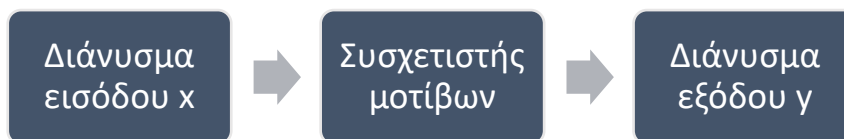
Σε μία αυτό-συσχετιστική μνήμη, $y_k = x_k$, οπότε οι χώροι (δεδομένων) εισόδου και εξόδου του δικτύου έχουν ίδια διαστατικότητα. Σε μία ετερο-συσχετιστική μνήμη, $y_k \neq x_k$ άρα, η διαστατικότητα του χώρου εξόδου σε αυτή τη δεύτερη περίπτωση μπορεί να είναι ίδια με αυτή του χώρου εισόδου, αλλά μπορεί και όχι.

Υπάρχουν δύο φάσεις που εμπλέκονται στη λειτουργία μιας συσχετιστικής μνήμης:

- Η φάση αποθήκευσης, η οποία αναφέρεται στην εκπαίδευση του δικτύου σύμφωνα με την Εξ. (1)
- Η φάση ανάκλησης, η οποία απαιτεί την ανάκτηση ενός απομνημονευμένου προτύπου ως απόκριση στην παρουσίαση μιας «θορυβώδους» ή παραμορφωμένης έκδοσης ενός προτύπου-κλειδιού στο δίκτυο.

Έστω ότι το ερέθισμα (είσοδος) x αναπαριστά μία θορυβώδη ή παραμορφωμένη έκδοση ενός πρότυπου-κλειδιού x_j . Αυτό το ερέθισμα παράγει μία απόκριση (έξοδο) y , όπως υποδεικνύεται στο Σχήμα 4.4. Για τέλεια ανάκληση, θα πρέπει να διαπιστώσουμε ότι $y = y_j$, όπου y_j είναι το απομνημονευμένο πρότυπο που έχει συσχετιστεί με το πρότυπο-κλειδί x_j . Όταν $y \neq y_j$ για $x = x_j$, η συσχετιστική μνήμη λέγεται ότι επιδεικνύει *σφάλμα ανάκλησης*.

Το πλήθος προτύπων q που αποθηκεύονται σε μια συσχετιστική μνήμη παρέχει ένα άμεσο μέτρο της χωρητικότητας αποθήκευσης του δικτύου. Κατά τη σχεδίαση μιας συσχετιστικής μνήμης, η πρόκληση είναι να κάνουμε τη χωρητικότητα αποθήκευσης q (εκφρασμένη ως ποσοστό επί του συνολικού αριθμού N των νευρώνων που χρησιμοποιήθηκαν για την κατασκευή του δικτύου) όσο το δυνατόν πιο μεγάλη, επιδιώκοντας ταυτόχρονα τη σωστή ανάκληση μεγάλου μέρους των απομνημονευμένων προτύπων.



Σχήμα 4.4

4.3.3.β. Αναγνώριση Προτύπων

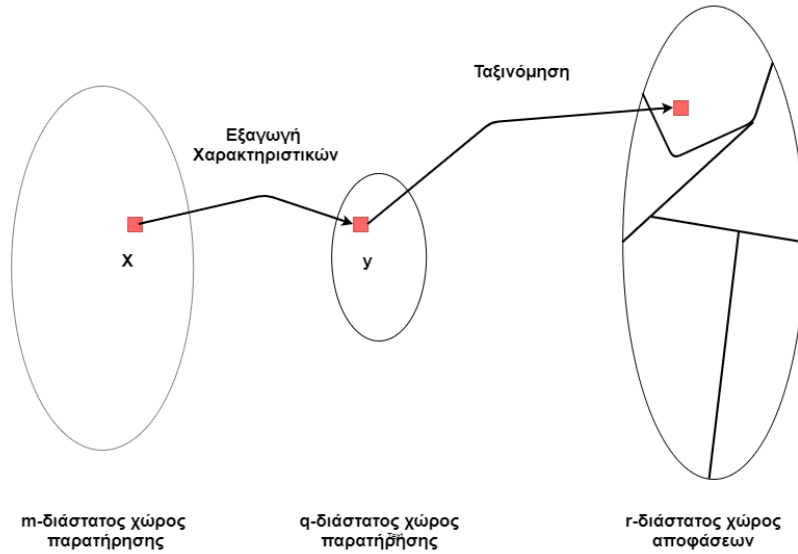
Οι άνθρωποι τα καταφέρνουν πολύ καλά στην αναγνώριση προτύπων. Λαμβάνουμε δεδομένα από τον κόσμο γύρω μας μέσω των αισθήσεών μας και είμαστε σε θέση να αναγνωρίσουμε την προέλευσή τους. Συχνά, αυτό μπορούμε να το κάνουμε σχεδόν άμεσα και με πρακτικά μηδενική προσπάθεια. Για παράδειγμα, μπορούμε να αναγνωρίσουμε το πρόσωπο ενός ατόμου που μας είναι οικείο, ακόμα και αν το άτομο αυτό έχει γεράσει από την τελευταία φορά που το συναντήσαμε. Μπορούμε, επίσης, να αναγνωρίσουμε ένα οικείο άτομο ακούγοντας τη φωνή του στο τηλέφωνο ακόμα και αν η σύνδεση δεν είναι καλή και μπορούμε να ξεχωρίσουμε ένα καλό αβγό από ένα χαλασμένο απλώς και μόνο από τη μυρωδιά του. Οι άνθρωποι

εκτελούν την αναγνώριση προτύπων αφού περάσουν από μία διαδικασία μάθησης και το ίδιο ισχύει και για τα νευρωνικά δίκτυα.

Η αναγνώριση προτύπων ορίζεται φορμαλιστικά ως η διαδικασία δια της οποίας ένα προσλαμβανόμενο σήμα/πρότυπο αντιστοιχίζεται σε μία κλάση που ανήκει σε ένα προκαθορισμένο σύνολο κλάσεων. Ένα νευρωνικό δίκτυο εκτελεί αναγνώριση προτύπων αφού πρώτα περάσει από μία «σύνοδο εκπαίδευσης» κατά τη διάρκεια της οποίας παρουσιάζεται κατ' επανάληψη στο δίκτυο ένα σύνολο προτύπων εισόδου μαζί με την κατηγορία (κλάση) στην οποία ανήκει το καθένα. Στη συνέχεια, παρουσιάζεται στο δίκτυο ένα νέο πρότυπο, το οποίο δεν έχει «δει» ποτέ πριν, αλλά το οποίο ανήκει στον ίδιο πληθυσμό προτύπων που χρησιμοποιήθηκαν για την εκπαίδευση του δικτύου. Το δίκτυο είναι σε θέση να αναγνωρίσει την κλάση στην οποία ανήκει το συγκεκριμένο πρότυπο βάσει της πληροφορίας που έχει εξαγάγει από τα δεδομένα εκπαίδευσης. Η αναγνώριση προτύπων που εκτελεί ένα νευρωνικό δίκτυο είναι στατική φύσης, με τα πρότυπα να αναπαρίστανται από σημεία σε έναν πολυδιάστατο χώρο αποφάσεων. Ο χώρος αποφάσεων διαιρείται σε περιοχές, κάθε μία εκ των οποίων συσχετίζεται με μία κλάση. Τα όρια αποφάσεων καθορίζονται από τη διαδικασία εκπαίδευσης. Η δημιουργία αυτών των ορίων καθίσταται στατιστική λόγω της εγγενούς μεταβλητότητας που υπάρχει εντός και μεταξύ των κλάσεων.

Με γενικούς όρους, οι μηχανές αναγνώρισης προτύπων που χρησιμοποιούν νευρωνικά δίκτυα μπορεί να έχουν μία από τις ακόλουθες δύο μορφές:

- Η μηχανή χωρίζεται σε δύο μέρη, ένα μη επιβλεπόμενο δίκτυο για την εξαγωγή χαρακτηριστικών και ένα επιβλεπόμενο δίκτυο για την ταξινόμηση. Μία τέτοια μέθοδος ακολουθεί την «παραδοσιακή» προσέγγιση της στατιστικής αναγνώρισης προτύπων (Fukunaga, 1990 Duda κ.α., 2001 Theodoridis και Koutroumbas, 2003). Με εννοιολογικούς όρους, ένα πρότυπο αντιπροσωπεύεται από ένα σύνολο m παρατηρήσιμων χαρακτηριστικών (observables), τα οποία μπορούν να θεωρηθούν ένα σημείο x σε έναν m -διάστατο χώρο παρατηρήσεων (δεδομένων). Η εξαγωγή χαρακτηριστικών περιγράφεται από έναν μετασχηματισμό ο οποίος αντιστοιχίζει το σημείο x σε ένα ενδιάμεσο σημείο y ενός q -διάστατου χώρου χαρακτηριστικών, όπου $q < m$, όπως υποδεικνύεται στο Σχήμα 4.5. Αυτός ο μετασχηματισμός μπορεί να εκληφθεί ως μείωση της διαστατικότητας (δηλ. συμπίεση δεδομένων), η χρήση της οποίας αιτιολογείται από το γεγονός ότι απλοποιεί την εργασία ταξινόμησης. Η ταξινόμηση περιγράφεται ως ένας μετασχηματισμός ο οποίος αντιστοιχίζει το ενδιάμεσο σημείο y σε μία από τις κλάσεις ενός r -διάστατου χώρου αποφάσεων, όπου r είναι ο αριθμός των διακριτών κλάσεων.
- Η μηχανή σχεδιάζεται ως ένα δίκτυο πρόσθιας τροφοδότησης χρησιμοποιώντας έναν αλγόριθμο επιβλεπόμενης μάθησης. Σε αυτή τη δεύτερη προσέγγιση, η εργασία της εξαγωγής χαρακτηριστικών εκτελείται από τις υπολογιστικές μονάδες στο κρυφό επίπεδο ή στα κρυφά επίπεδα, αν έχει περισσότερα από ένα, του δικτύου.



Σχήμα 4.5

4.3.3.γ. Προσέγγιση Συνάρτησης

Μία ακόμα διαδικασία μάθησης αποτελεί η αποκαλούμενη προσέγγιση συνάρτησης. Θεωρούμε μία μη γραμμική αντιστοίχιση εισόδου-εξόδου η οποία περιγράφεται από τη σχέση:

$$\mathbf{d} = \mathbf{f}(\mathbf{x}) \quad (1)$$

όπου το διάνυσμα \mathbf{x} είναι η είσοδος και το διάνυσμα \mathbf{d} είναι η έξοδος. Η διανυσματική συνάρτηση $\mathbf{f}(\cdot)$ υποτίθεται ότι είναι άγνωστη. Σαν αντιστάθμισμα για την έλλειψη γνώσης σχετικά με τη συνάρτηση $\mathbf{f}(\cdot)$, μας δίνεται το ακόλουθο σύνολο χαρακτηρισμένων παραδειγμάτων:

$$\mathcal{F} = \{(\mathbf{x}_i, \mathbf{d}_i)\}_{i=1}^N \quad (2)$$

Το ζητούμενο είναι να σχεδιάσουμε ένα νευρωνικό δίκτυο το οποίο θα προσεγγίζει την άγνωστη συνάρτηση $\mathbf{f}(\cdot)$, με τρόπο ώστε η συνάρτηση $\mathbf{F}(\cdot)$ που περιγράφει την αντιστοίχιση εισόδου-εξόδου που υλοποιείται πραγματικά από το δίκτυο, να είναι επαρκώς κοντά στην $\mathbf{f}(\cdot)$ υπό την Ευκλείδεια έννοια για όλες τις εισόδους, όπως υποδεικνύει η σχέση:

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{f}(\mathbf{x})\| < \varepsilon, \forall \mathbf{x} \quad (3)$$

όπου ε είναι ένας μικρός θετικός αριθμός. Υπό τον όρο ότι το μέγεθος N του δείγματος εκπαίδευσης \mathcal{F} είναι επαρκώς μεγάλο και το δίκτυο είναι εφοδιασμένο με επαρκή

αριθμό ελεύθερων παραμέτρων, το σφάλμα προσέγγισης ϵ μπορεί να γίνει επαρκώς μικρό για την εν λόγω εργασία.

Ο τρόπος προσέγγισης του προβλήματος που περιγράφεται εδώ είναι ιδανικός για χρήση επιβλεπόμενης μάθησης, με το x_i να παίζει το ρόλο του διανύσματος εισόδου και το d_i να παίζει το ρόλο της επιθυμητής απόκρισης.

Η δυνατότητα ενός νευρωνικού δικτύου να προσεγγίζει μία άγνωστη αντιστοίχιση εισόδου-εξόδου μπορεί να αξιοποιηθεί με δύο σημαντικούς τρόπους:

i. Αναγνώριση Συστημάτων

Έστω ότι η αρχική σχέση για την έξοδο d περιγράφει τη σχέση εισόδου-εξόδου ενός άγνωστου, άνευ μνήμης συστήματος πολλαπλών εισόδων-πολλαπλών εξόδων (MIMO). Με τον όρο «άνευ μνήμης» εννοούμε ένα σύστημα το οποίο είναι αναλλοίωτο στο χρόνο. Μπορούμε τότε να χρησιμοποιήσουμε το σύνολο χαρακτηρισμένων παραδειγμάτων της εξίσωσης (2) για την εκπαίδευση ενός νευρωνικού δικτύου ως μοντέλο του συστήματος. Έστω ότι το διάνυσμα y_i συμβολίζει την πραγματική έξοδο του νευρωνικού δικτύου που παράγεται σε απόκριση προς ένα διάνυσμα εισόδου x_i . Η διαφορά μεταξύ του d_i (το οποίο σχετίζεται με το x_i) και της εξόδου του δικτύου y_i παρέχει το διάνυσμα σήματος σφάλματος e_i . Αυτό το σήμα σφάλματος, με τη σειρά του, χρησιμοποιείται για την προσαρμογή των ελεύθερων παραμέτρων του δικτύου, με στόχο την ελαχιστοποίηση του τετραγώνου της διαφοράς μεταξύ των εξόδων του αγνώστου συστήματος και του νευρωνικού δικτύου (υπό μία στατιστική έννοια) και υπολογίζεται για ολόκληρο το δείγμα εκπαίδευσης \mathcal{F} .

ii. Αντίστροφη μοντελοποίηση

Υποθέτουμε ότι στη συνέχεια μας δίνεται ένα γνωστό άνευ μνήμης MIMO, του οποίου η σχέση εισόδου-εξόδου περιγράφεται από την εξίσωση (1). Το ζητούμενο σε αυτή την περίπτωση είναι να κατασκευαστεί ένα αντίστροφο μοντέλο το οποίο θα παράγει το διάνυσμα x σε απόκριση προς το διάνυσμα d . Άρα, το αντίστροφο σύστημα μπορεί να περιγραφεί ως:

$$x = f^{-1}(d) \quad (4)$$

όπου η διανυσματική συνάρτηση $f^{-1}(\cdot)$ συμβολίζει αναστροφή της συνάρτησης $f(\cdot)$. Σημειώνεται, ωστόσο, ότι η $f^{-1}(\cdot)$ δεν είναι αντίστροφη (reciprocal) συνάρτηση της $f(\cdot)$, αλλά η χρήση του εκθέτη -1 σηματοδοτεί απλώς μία αναστροφή. Σε πολλές περιπτώσεις που συναντώνται στην πράξη, η διανυσματική συνάρτηση $f(\cdot)$ είναι υπερβολικά πολύπλοκη, δυσχεραίνοντας σε απαγορευτικό βαθμό τη διατύπωση της $f^{-1}(\cdot)$. Δοθέντος του συνόλου χαρακτηρισμένων παραδειγμάτων της εξίσωσης (2), μπορούμε να υλοποιήσουμε μία προσέγγιση της $f^{-1}(\cdot)$ μέσω του νευρωνικού δικτύου. Στην περίπτωση που περιγράφεται εδώ, οι ρόλοι των x_i και d_i εναλλάσσονται: Το διάνυσμα d_i χρησιμοποιείται ως είσοδος και το x_i αντιμετωπίζεται ως επιθυμητή απόκριση. Έστω ότι το διάνυσμα σήματος σφάλματος e_i

συμβολίζει τη διαφορά μεταξύ του x_i και της πραγματικής εξόδου y_i του νευρωνικού δικτύου, η οποία παράγεται σε απόκριση προς το d_i . Όπως και στην περίπτωση της αναγνώρισης συστημάτων, αυτό το διάνυσμα σήματος σφάλματος χρησιμοποιείται για την προσαρμογή των ελεύθερων παραμέτρων του νευρωνικού δικτύου, με στόχο την ελαχιστοποίηση του τετραγώνου της διαφοράς μεταξύ των εξόδων του άγνωστου αντίστροφου συστήματος και του νευρωνικού δικτύου (υπό μία στατιστική έννοια) και υπολογίζεται για όλο το σύνολο των δεδομένων εκπαίδευσης \mathcal{F} . Συνήθως, η αντίστροφη μοντελοποίηση είναι δυσκολότερη εργασία μάθησης συγκριτικά με την αναγνώριση συστημάτων, καθώς μπορεί να μην υπάρχει μία μοναδική για αυτή.

4.3.3.δ. Έλεγχος

Ο έλεγχος μιας μονάδας είναι άλλη μία διαδικασία μάθησης ιδιαίτερα κατάλληλη για υλοποίηση μέσω νευρωνικών δικτύων. Με τον όρο μονάδα εννοούμε μια διεργασία ή ένα ζωτικής σημασίας υποσύστημα ενός μεγαλύτερου συστήματος το οποίο πρέπει να διατηρείται σε μια ελεγχόμενη κατάσταση. Η στενή σχέση της μάθησης με τον εγκέφαλο δεν θα πρέπει να μας εκπλήσσει, καθώς, στο κάτω κάτω, ο ανθρώπινος εγκέφαλος δεν είναι παρά ένας υπολογιστής (επεξεργαστής πληροφοριών), οι έξοδοι του οποίου (ως ένα πλήρες σύστημα) είναι οι ενέργειές μας. Υπ' αυτό το πρίσμα, ο εγκέφαλος είναι η ζωντανή απόδειξη του ότι είναι δυνατόν να κατασκευαστεί ένας «γενικευμένος» μηχανισμός ελέγχου ο οποίος θα αξιοποιεί πλήρως τις δυνατότητες παράλληλης επεξεργασίας συστημάτων hardware με κατανομημένη αρχιτεκτονική, ικανός να ελέγχει πολλές χιλιάδες επενεργητές (actuators, μυϊκές ίνες) ταυτόχρονα, να χειρίζεται επιτυχώς τη μη γραμμικότητα και το θόρυβο και να εκτελεί βελτιστοποιήσεις με μεγάλο ορίζοντα σχεδιασμού (Werbos, 1992).

4.4. *Perceptron Πολλών Επιπέδων (Multilayer Perceptron)*

Μία από τις κατηγορίες νευρωνικών δικτύων αποτελούν τα «Perceptron» Πολλών Επιπέδων. Τα βασικά χαρακτηριστικά των νευρωνικών αυτών δικτύων είναι:

- Το μοντέλο κάθε νευρώνα στο δίκτυο περιλαμβάνει μία μη γραμμική συνάρτηση ενεργοποίησης, η οποία είναι διαφορίσιμη.
- Το δίκτυο περιέχει ένα ή περισσότερα επίπεδα τα οποία παραμένουν κρυφά (hidden layers) για τους κόμβους των επιπέδων εισόδου και εξόδου.
- Το δίκτυο επιδεικνύει μεγάλη διασυνδεσιμότητα, ο βαθμός της οποίας καθορίζεται από τα συναπτικά βάρη του δικτύου.

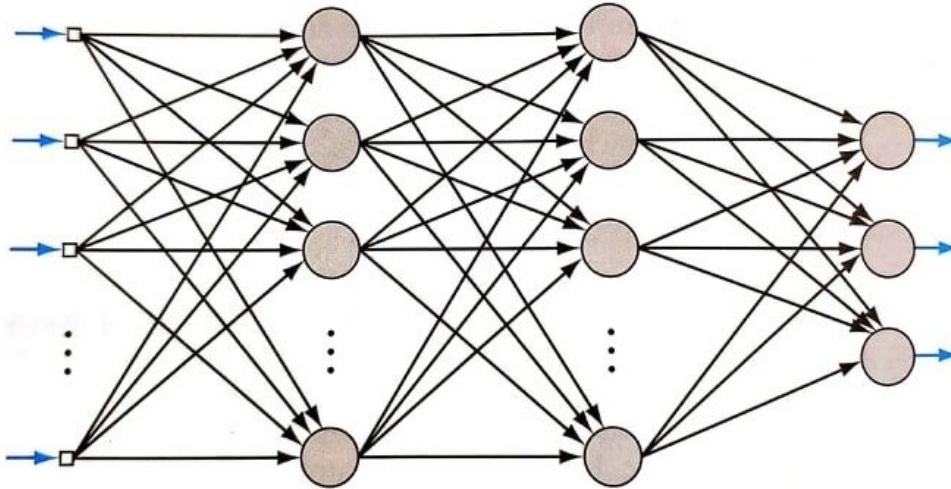
Αυτά τα ίδια χαρακτηριστικά, ωστόσο, είναι επίσης υπεύθυνα για την ανεπάρκεια της γνώσης μας αναφορικά με τη συμπεριφορά του δικτύου. Πρώτον, η παρουσία μίας κατανεμημένης μορφής μη γραμμικότητας και η υψηλή διασυνδεσιμότητα του δικτύου καθιστούν τη θεωρητική ανάλυση ενός perceptron πολλών επιπέδων δύσκολη υπόθεση. Δεύτερον, η χρήση κρυφών νευρώνων καθιστά δυσκολότερη την κατανόηση και σχηματοποίηση της διαδικασίας μάθησης. Εμμέσως, η διαδικασία μάθησης οφείλει να αποφασίσει ποια χαρακτηριστικά του προτύπου εισόδου θα πρέπει να αναπαρίσταται από τους κρυφούς νευρώνες. Έτσι, η διαδικασία μάθησης γίνεται πιο δύσκολη, επειδή η αναζήτηση πρέπει να διεξαχθεί σε έναν πολύ μεγαλύτερο χώρο πιθανών λειτουργιών και πρέπει να κάνει μία επιλογή μεταξύ των εναλλακτικών αναπαραστάσεων του προτύπου εισόδου.

Μια δημοφιλή μέθοδος για την εκπαίδευση των perceptron πολλών επιπέδων είναι ο αλγόριθμος BK (Back Propagation). Η εκπαίδευση λαμβάνει χώρα σε δύο φάσεις:

1. Στη φάση που εξελίσσεται προς τα εμπρός, τα συναπτικά βάρη του δικτύου είναι σταθερά και το σήμα εισόδου διαδίδεται διαμέσου του δικτύου, επίπεδο προς επίπεδο, μέχρι να φτάσει στην έξοδο. Συνεπώς, σε αυτή τη φάση, οι αλλαγές περιορίζονται στα δυναμικά ενεργοποίησης και στις εξόδους των νευρώνων του δικτύου.
2. Στη φάση που εξελίσσεται προς τα πίσω, παράγεται ένα σήμα σφάλματος μέσω της σύγκρισης της εξόδου του δικτύου με μία επιθυμητή απόκριση. Το προκύπτον σήμα σφάλματος διαδίδεται διαμέσου του δικτύου, ξανά επίπεδο προς επίπεδο, αλλά αυτή τη φορά η διάδοση γίνεται με κατεύθυνση προς τα πίσω. Σε αυτή τη δεύτερη φάση, γίνονται διαδοχικές προσαρμογές στα συναπτικά βάρη του δικτύου. Ο υπολογισμός των προσαρμογών για το επίπεδο εξόδου είναι απλή υπόθεση, ενώ για τα κρυφά επίπεδα γίνεται πολύ πιο δύσκολος.

Η ανάπτυξη του αλγορίθμου BK στα μέσα της δεκαετίας του '80 αποτέλεσε ένα ορόσημο στην εξέλιξη του κλάδου των νευρωνικών δικτύων διότι παρείχε μία υπολογιστικώς αποτελεσματική μέθοδο για την εκπαίδευση των perceptron πολλών επιπέδων, αίροντας τον έως τότε πεσιμισμό σχετικά με τις δυνατότητες μάθησής τους που οφείλονταν εν πολλοίς στο βιβλίο των Minsky και Papert (1969).

Στο Σχήμα 4.4.α απεικονίζεται ένας αρχιτεκτονικός γράφος ενός perceptron πολλών επιπέδων με δύο κρυφά επίπεδα και ένα επίπεδο εξόδου. Για λόγους απλότητας και πληρότητας το δίκτυο που παρουσιάζεται είναι πλήρως συνδεδεμένο (fully connected). Αυτό σημαίνει ότι ένας νευρώνας σε οποιοδήποτε επίπεδο του δικτύου συνδέεται με όλους τους νευρώνες (κόμβους) του προηγούμενου επιπέδου. Η ροή σήματος διαμέσου του δικτύου προχωρά με κατεύθυνση προς τα εμπρός, από τα αριστερά προς τα δεξιά και επίπεδο προς επίπεδο.



Σχήμα 4.6 Perceptron Πολλών Επιπέδων

Σε ένα τέτοιο δίκτυο μπορούμε να αναγνωρίσουμε δύο είδη σημάτων:

1. *Λειτουργικά Σήματα.*

Ένα λειτουργικό σήμα (function signal) είναι ένα σήμα εισόδου (ερέθισμα) το οποίο φτάνει στην είσοδο του δικτύου, διαδίδεται προς τα εμπρός (νευρώνα προς νευρώνα) σε όλο το δίκτυο και εμφανίζεται τελικά στην έξοδο του δικτύου, σαν σήμα εξόδου. Αποκαλούμε ένα τέτοιο σήμα «λειτουργικό» για δύο λόγους: Πρώτον, υποτίθεται ότι εκτελεί μια χρήσιμη λειτουργία στην έξοδο του δικτύου. Δεύτερον, σε κάθε νευρώνα του δικτύου διαμέσου του οποίου διέρχεται ένα λειτουργικό σήμα, το σήμα αυτό υπολογίζεται ως συνάρτηση των εισόδων και των σχετιζόμενων βαρών που εφαρμόζονται σε αυτό το νευρώνα. Το λειτουργικό σήμα αποκαλείται επίσης σήμα εισόδου.

2. *Σήματα Σφάλματος.*

Ένα σήμα σφάλματος (error signal) προέρχεται από έναν νευρώνα εξόδου του δικτύου και διαδίδεται προς τα πίσω (επίπεδο προς επίπεδο) διαμέσου του δικτύου. Το αποκαλούμε «σήμα σφάλματος» επειδή ο υπολογισμός του κάθε νευρώνα του δικτύου απαιτεί τη χρήση μιας εξαρτώμενης από το σφάλμα συνάρτησης, με τη μία ή την άλλη μορφή.

Οι νευρώνες εξόδου απαρτίζουν το επίπεδο εξόδου του δικτύου. Οι υπόλοιποι νευρώνες απαρτίζουν τα κρυφά επίπεδα του δικτύου. Έτσι, οι κρυφές μονάδες δεν αποτελούν μέρος ούτε της εξόδου, ούτε της εισόδου του δικτύου – σε αυτό οφείλεται και ο χαρακτηρισμός τους ως «κρυφές». Το πρώτο κρυφό επίπεδο τροφοδοτείται από το επίπεδο εισόδου του δικτύου, το οποίο αποτελείται από αισθητηριακές μονάδες (πηγαίοι κόμβοι). Οι έξοδοι που παράγονται από το πρώτο κρυφό επίπεδο

Κεφάλαιο 4. Νευρωνικά Δίκτυα και Μηχανική Μάθηση

εφαρμόζονται στο επόμενο κρυφό επίπεδο. Η διαδικασία συνεχίζεται κατά τον ίδιο τρόπο για όλο το υπόλοιπο του δικτύου.

Κάθε κρυφός νευρώνας ή νευρώνας εξόδου ενός perceptron πολλών επιπέδων σχεδιάζεται ώστε να εκτελεί δύο υπολογισμούς:

1. Τον υπολογισμό του λειτουργικού σήματος που εμφανίζεται στην έξοδο κάθε νευρώνα, ο οποίος εκφράζεται ως μία συνεχής μη γραμμική συνάρτηση του σήματος εισόδου και των συναπτικών βαρών που σχετίζονται με αυτό το νευρώνα.
2. Τον υπολογισμό μιας εκτίμησης του διανύσματος κλίσης (δηλαδή, των κλίσεων μιας επιφάνειας σφάλματος σε σχέση με τα βάρη που είναι συνδεδεμένα στις εισόδους ενός νευρώνα), ο οποίος χρειάζεται για τη φάση που εξελίσσεται προς τα πίσω στο δίκτυο.

Καθώς ο στόχος αυτής της διπλωματικής εργασίας δεν είναι η εμβάθυνση στη θεωρία της λειτουργίας των MLP νευρωνικών δικτύων, αλλά γίνεται χρήση ενός για την υλοποίηση της τεχνικής πρόβλεψης, θα γίνει μία σύντομη αναφορά σε χαρακτηριστικά και στοιχεία τους που συνδέονται με την υλοποίηση αυτή. Τα χαρακτηριστικά αυτά προκύπτουν από τη χρήση του αλγορίθμου BK στα Multilayer Perceptrons.

Ως πρόλογο στα χαρακτηριστικά αυτά παραθέτουμε σε βασικά σημεία τον αλγόριθμο BK:

1. Η διόρθωση $\Delta_{w_{ji}}(\mathbf{n})$ που αναφέραμε και ανωτέρω ότι εφαρμόζεται στο συναπτικό βάρος που συνδέει τον νευρώνα i με τον νευρώνα j ορίζεται από τον κανόνα:

$$\left\{ \begin{array}{l} \text{Διόρθωση} \\ \text{Βάρους} \\ \Delta w_{ij}(\mathbf{n}) \end{array} \right\} = \left\{ \begin{array}{l} \text{Παράμετρος} \\ \text{ρυθμού} \\ \text{μάθησης} \\ \eta \end{array} \right\} \times \left\{ \begin{array}{l} \text{Τοπική} \\ \text{κλίση} \\ \Delta_j(\mathbf{n}) \end{array} \right\} \times \left\{ \begin{array}{l} \text{Σήμα εισόδου} \\ \text{του νευρώνα } j \\ y_i(\mathbf{n}) \end{array} \right\}$$

2. Η τοπική κλίση $\Delta_j(\mathbf{n})$ εξαρτάται από το εάν ο νευρώνας j είναι ένας κόμβος εξόδου ή κρυφός κόμβος:
 - Αν ο νευρώνας j είναι ένας κόμβος εξόδου, το $\Delta_j(\mathbf{n})$ ισούται με το γινόμενο της παραγώγου $\phi'_j(u_j(\mathbf{n}))$ και του σήματος σφάλματος $e_j(\mathbf{n})$, αμφότερα εκ των οποίων συσχετίζονται με το νευρώνα j .
 - Αν ο νευρώνας j είναι ένας κρυφός κόμβος, το $\Delta_j(\mathbf{n})$ ισούται με το γινόμενο της σχετιζόμενης παραγώγου $\phi'_j(u_j(\mathbf{n}))$ και του σταθμισμένου αθροίσματος των δ που υπολογίζονται για τους νευρώνες του επόμενου κρυφού επιπέδου ή επιπέδου εξόδου και οι οποίοι συνδέονται με τον νευρώνα j .

4.4.1. Συνάρτηση Ενεργοποίησης

Η συνάρτηση ενεργοποίησης αποτελεί απαραίτητο κομμάτι για την σωστή «αρχικοποίηση» ενός νευρωνικού δικτύου πολλών επιπέδων. Ο υπολογισμός του δ για κάθε νευρώνα του perceptron πολλών επιπέδων απαιτεί γνώση της παραγώγου της συνάρτησης ενεργοποίησης $\varphi(\cdot)$ που σχετίζεται με αυτό το νευρώνα. Για να υπάρχει αυτή η παράγωγος, πρέπει η συνάρτηση $\varphi(\cdot)$ να είναι συνεχής. Με απλά λόγια, η διαφορισιμότητα είναι η μόνη απαίτηση που πρέπει να ικανοποιεί μία συνάρτηση ενεργοποίησης. Ένα παράδειγμα συνεχώς διαφορίσιμης, μη γραμμικής συνάρτησης ενεργοποίησης που χρησιμοποιείται ευρέως στα perceptron πολλών επιπέδων είναι μία συνάρτηση με σιγμοειδή μη γραμμικότητα.

4.4.2. Ρυθμός Μάθησης

Ο αλγόριθμος ΒΚ παρέχει μία «προσέγγιση» στην τροχιά που διαγράφεται στο χώρο βαρών, όπως υπολογίζεται από τη μέθοδο της πλέον απότομης κατάβασης (steepest descent). Όσο μικρότερη κάνουμε την παράμετρο ρυθμού μάθησης η , τόσο μικρότερες θα είναι οι αλλαγές στα συναπτικά βάρη του δικτύου από τη μία επανάληψη στην επόμενη – και τόσο περισσότερο ομαλή θα είναι η τροχιά στο χώρο βαρών. Ωστόσο, αυτή βελτίωση επιτυγχάνεται με αντίτιμο τον αργότερο ρυθμό μάθησης. Αν, από την άλλη, κάνουμε την παράμετρο ρυθμού μάθησης η υπερβολικά μεγάλη για να επιταχύνουμε το ρυθμό μάθησης, οι προκύπτουσες (μεγάλες) αλλαγές στα συναπτικά βάρη λαμβάνουν τέτοια μορφή που το δίκτυο μπορεί να γίνει ασταθές (δηλ. να ταλαντώνει).

4.4.3. Κριτήρια Τερματισμού

Γενικά, ο αλγόριθμος ΒΚ δεν μπορεί να αποδειχτεί ότι συγκλίνει και δεν υπάρχουν καλά ορισμένα κριτήρια για τον τερματισμό λειτουργίας του. Αντ' αυτών, υπάρχουν κάποια λογικά κριτήρια, το καθένα με τη δική του πρακτική αξία, τα οποία θα μπορούσαν να χρησιμοποιηθούν για τον τερματισμό της διαδικασίας προσαρμογής των βαρών. Για να καθορίσουμε ένα τέτοιο κριτήριο, είναι λογικό να προσφύγουμε στις μοναδικές ιδιότητες που έχει ένα τοπικό ή γενικό ελάχιστο της επιφάνειας σφάλματος. Έστω ότι το διάνυσμα βαρών w^* συμβολίζει ένα ελάχιστο, είτε τοπικό είτε γενικό. Μία αναγκαία συνθήκη για να είναι το w^* ένα ελάχιστο είναι ότι το διάνυσμα κλίσης $g(w)$ (δηλ. η μερική παράγωγος πρώτης τάξης) της επιφάνειας σφάλματος σε σχέση με το διάνυσμα βαρών w πρέπει να είναι μηδέν στο $w = w^*$. Κατά συνέπεια, μπορούμε να διατυπώσουμε ένα εύλογο κριτήριο σύγκλισης για τη διαδικασία μάθησης με οπισθοδιάδοση σφάλματος, ως εξής (Kramer & Sangiovanni-Vincentelli, 1989):

Ο αλγόριθμος ΒΚ θεωρείται ότι έχει συγκλίνει όταν η Ευκλείδεια νόρμα του διανύσματος κλίσης φτάσει σε ένα επαρκώς μικρό κατώφλι κλίσης.

Το μειονέκτημα αυτού του κριτηρίου σύγκλισης είναι ότι, για επιτυχημένες δοκιμές, οι χρόνοι μάθησης μπορεί να είναι μεγάλοι. Επίσης, απαιτεί τον υπολογισμό του διανύσματος κλίσης $g(w)$.

Μία άλλη μοναδική ιδιότητα ενός ελαχίστου την οποία μπορούμε να χρησιμοποιήσουμε είναι το γεγονός ότι η συνάρτηση κόστους $E_{av}(w)$ είναι στατική στο σημείο $w=w^*$. Μπορούμε, συνεπώς, να προτείνουμε ένα διαφορετικό κριτήριο σύγκλισης:

Ο αλγόριθμος BK θεωρείται ότι έχει συγκλίνει όταν ο απόλυτος ρυθμός μεταβολής του μέσου τετραγωνικού σφάλματος ανά εποχή είναι επαρκώς μικρός.

Ο ρυθμός μεταβολής στο μέσο τετραγωνικό σφάλμα θεωρείται τυπικά ότι είναι επαρκώς μικρός αν βρίσκεται στο εύρος τιμών 0.1 έως 1 επί τοις εκατό ανά εποχή. Ορισμένες φορές, χρησιμοποιούνται πολύ μικρότερες τιμές, έως και 0.01 επί τοις εκατό ανά εποχή. Δυστυχώς, αυτό το κριτήριο μπορεί να έχει ως αποτέλεσμα πρόωμο τερματισμό της διαδικασίας μάθησης.

Υπάρχει ένα ακόμη, χρήσιμο και υποστηριζόμενο από τη θεωρία, κριτήριο για τη σύγκλιση: Μετά από κάθε επανάληψη της διαδικασίας μάθησης, το δίκτυο ελέγχεται για την απόδοσή του αναφορικά με τη γενίκευση που μπορεί να επιτύχει. Η διαδικασία μάθησης τερματίζεται όταν η απόδοσή του στη γενίκευση κρίνεται επαρκής, ή όταν είναι εμφανές ότι η απόδοσή του στη γενίκευση έχει φτάσει σε μέγιστο.

Κεφάλαιο 5 . Προτεινόμενη Μεθοδολογία

5.1. Εισαγωγή

Στόχος της παρούσας διπλωματικής όπως αναφέρεται και στην εισαγωγή είναι η ανάπτυξη μίας μεθοδολογίας με την οποία θα καθίσταται δυνατή η παραγωγή προβλέψεων για όλα τα επίπεδα μιας ιεραρχικής δομής αξιοποιώντας όλα τα δεδομένα ανεξαρτήτως επιπέδου. Ειδικότερα, η πρόταση αυτή αφορά χρονοσειρές διακοπτόμενης ζήτησης που συντελούν μία ιεραρχική δομή. Ας σημειωθεί ότι αυτός ο τρόπος πρόβλεψης δεν δύναται να υλοποιηθεί με τις μεθόδους reconciliation που υπάρχουν, καθώς σύμφωνα με το μοντέλο που αναπτύχθηκε και αναλύθηκε παραπάνω, για την παραγωγή προβλέψεων με μεθόδους reconciliation είναι προαπαιτούμενη η ύπαρξη προβλέψεων για κάποιο από τα επίπεδα της δομής.

Με τις μεθόδους reconciliation που περιγράφονται στα κεφάλαια 3.2 και 3.3 γίνεται μία πρώτη αξιοποίηση της επιπλέον πληροφορίας που παρέχεται σε μία ιεραρχική δομή, καθώς χρησιμοποιούν τμήμα της υπάρχουσας γνώσης – είτε μέσω της αξιοποίησης των προβλέψεων των χαμηλότερων επιπέδων για την προσέγγιση των ανώτερων (Bottom-Up), είτε χρησιμοποιώντας τις προβλέψεις των υψηλότερων επιπέδων και τον διαμερισμό τους για την προσέγγιση των χαμηλότερων (Top-Down). Επίσης, η μέθοδος Optimal reconciliation που περιγράφεται στο κεφάλαιο 3.4 απ' ενός μεν χρησιμοποιεί προβλέψεις από όλη την ιεραρχία απ' ετέρου όμως δεν είναι υλοποιήσιμη χωρίς επικουρική μέθοδο πρόβλεψης για την παραγωγή των βασικών προβλέψεων σε κάθε επίπεδο. Εκτός αυτού, ο συνδυασμός των προβλέψεων γίνεται αποκλειστικά με γραμμικό τρόπο, γεγονός που περιορίζει την ακρίβεια στις προσεγγίσεις της μεθόδου.

Έτσι, η ύπαρξη μιας μεθόδου που αξιοποιεί όλες τις χρονοσειρές της δομής για την παραγωγή προβλέψεων τόσο στα χαμηλότερα επίπεδα, όσο και στα συναθροισμένα παρέχει τη μέγιστη δυνατή εκμετάλλευση πληροφορίας. Το σκεπτικό πίσω από την μέθοδο που δύναται να πραγματοποιηθεί για την εκμετάλλευση της πληροφορίας είναι να συνδυάζονται τα κατάλληλα δεδομένα από όλη την ιεραρχία εξ αρχής για την παραγωγή των βασικών προβλέψεων και να «εμπεριέχουν» ήδη γνώση από όλα τα επίπεδα χωρίς την εκ νέου χρήση κάποιας μεθόδου reconciliation.

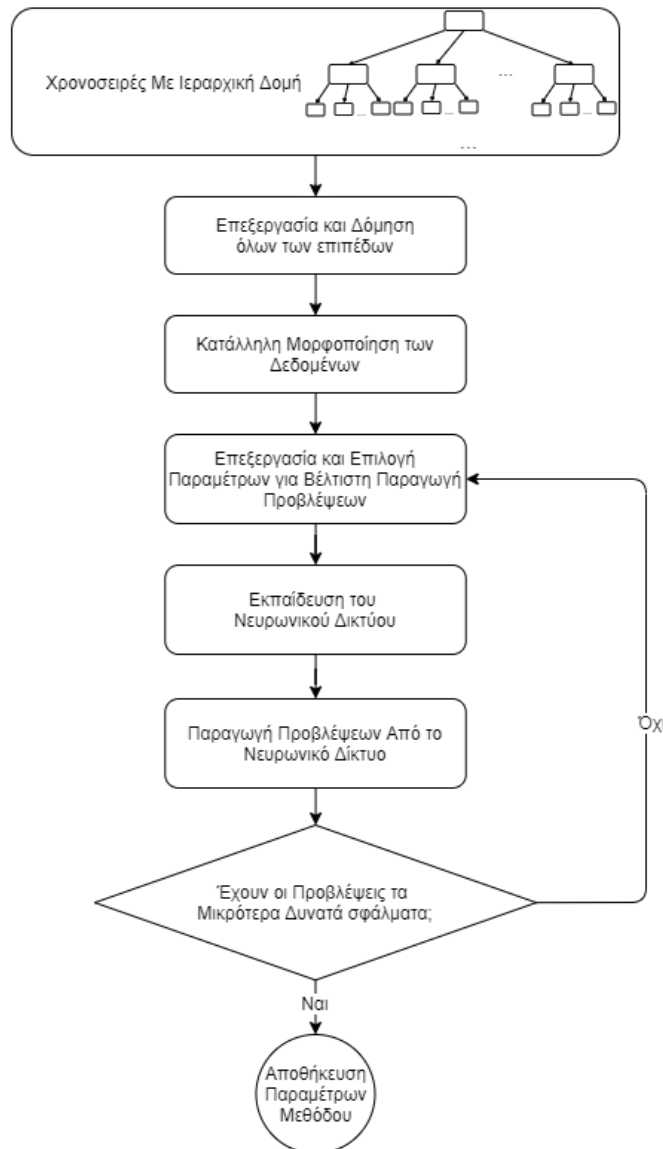
Η μέθοδος που επιλέχθηκε για την υλοποίηση μιας τέτοια μεθοδολογίας ήταν η χρήση machine learning και συγκεκριμένα η μέθοδος της παλινδρόμησης (regression). Ως παλινδρόμηση ορίζεται η μοντελοποίηση της σχέσης μεταξύ μιας βαθμωτής εξαρτημένης μεταβλητής που θα τη συμβολίζουμε με Y και μίας ή περισσοτέρων επεξηγηματικών (ή αλλιώς ανεξάρτητων) μεταβλητών που θα τις συμβολίζουμε με X . Στην περίπτωση που έχουμε μία επεξηγηματική μεταβλητή μιλάμε για απλή γραμμική παλινδρόμηση, ενώ όταν έχουμε δύο και άνω πρόκειται για πολλαπλή γραμμική παλινδρόμηση.

Πιο συγκεκριμένα, με τη βοήθεια της βιβλιοθήκης scikit-learn της γλώσσας προγραμματισμού Python V.3.7.1 αναπτύχθηκε ένα νευρωνικό δίκτυο το οποίο

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία

λαμβάνει ως εισόδους τις χρονοσειρές μιας ιεραρχικής δομής ή τμήματα αυτών μέχρι τη χρονική περίοδο t και παράγει ως έξοδο πρόβλεψη για τη χρονική περίοδο $t+1$.

Ξεκινώντας από ένα τυχαίο dataset με μοναδικό χαρακτηριστικό την ύπαρξη ιεραρχίας μεταξύ των χρονοσειρών του (στην περίπτωση μας συντελεί έξτρα χαρακτηριστικό η διακοπτόμενη φύση αυτών) καλούμαστε να αναπτύξουμε μία δομή που επεξεργάζεται κατάλληλα τα δεδομένα αυτά και τα εισάγει σε ένα νευρωνικό δίκτυο το οποίο με σωστή επεξεργασία παράγει τις προβλέψεις που επιθυμούμε. Σχηματικά, η διαδικασία η οποία ακολουθήθηκε μπορεί να απεικονιστεί ως εξής:

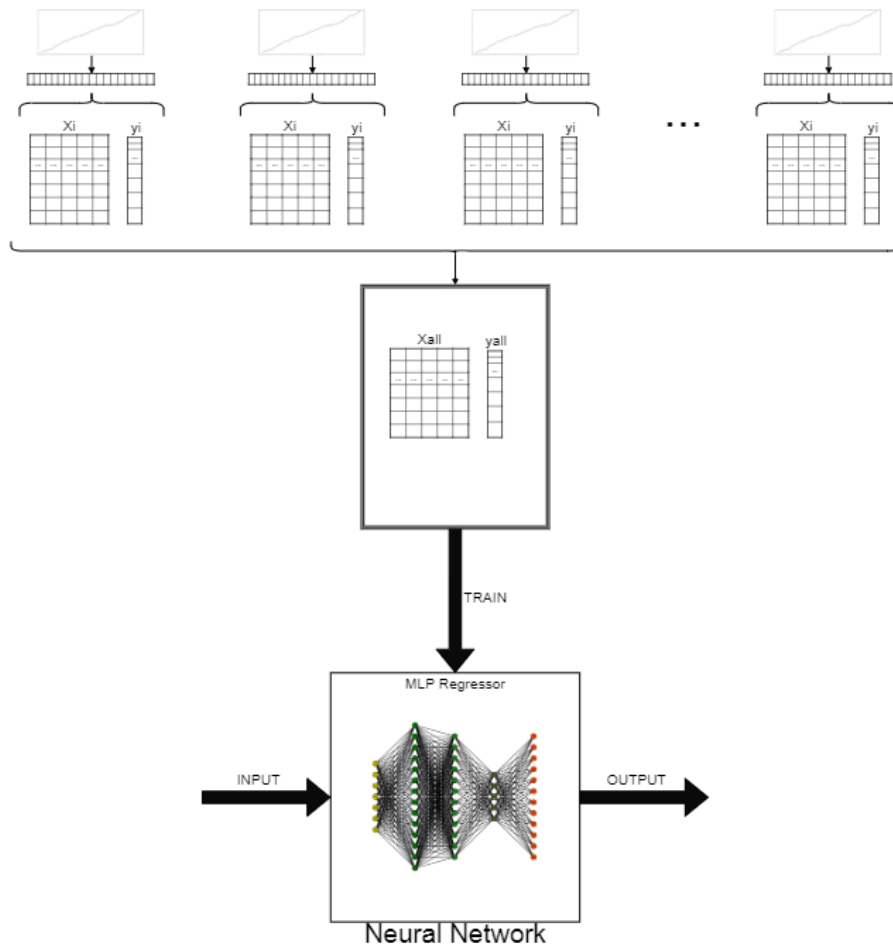


Σχήμα 5.1 Workflow Chart

Η μεθοδολογία, όπως γίνεται σαφές, περιλαμβάνει 2 στάδια: την εκπαίδευση του νευρωνικού δικτύου και την παραγωγή των προβλέψεων από αυτό.

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία

Στο πρώτο στάδιο η κάθε χρονοσειρά του dataset αναλύεται και δομεί κατάλληλα τους «πίνακες εκπαίδευσης» του νευρωνικού δικτύου (X_i), μαζί με το διάνυσμα προτύπων (y_i) τα οποία, εν συνεχεία, συνδυάζονται με όλες τις διαθέσιμες χρονοσειρές προκειμένου να αποτελέσουν ενιαίους πίνακες που τοποθετούνται ως training set του δικτύου.

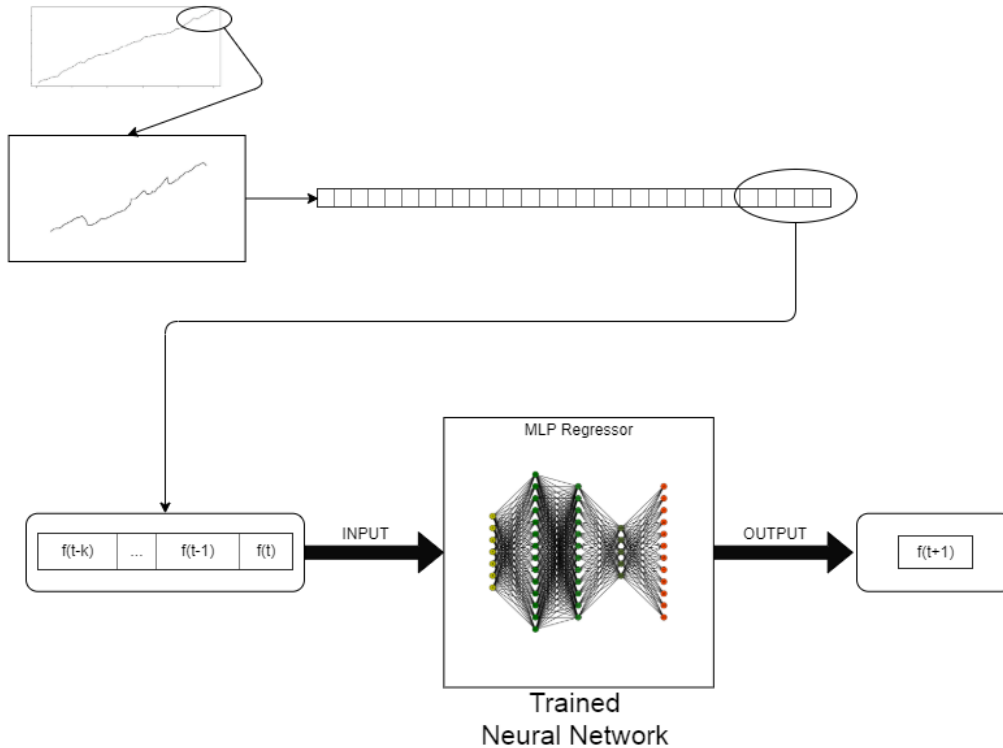


Σχήμα 5.2 Εκπαίδευση του Νευρωνικού Δικτύου Πολλών Επιπέδων

Η εκπαίδευση του δικτύου συντελείται μετά από κατάλληλη επεξεργασία των δεδομένων και παροχή τους στο νευρωνικό δίκτυο.

Στο δεύτερο στάδιο, αφού έχουμε «εκπαιδέσει» το νευρωνικό δίκτυο, αναλύουμε πάλι, αυτή τη φορά τη χρονοσειρά για την οποία επιθυμούμε να λάβουμε πρόβλεψη, και τοποθετείται τμήμα αυτής, κατάλληλα επεξεργασμένο, ως είσοδος στο νευρωνικό δίκτυο.

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία



Σχήμα 5.3 Πρόβλεψη με Νευρωνικό Δίκτυο Πολλών Επιπέδων

Η παραγωγή προβλέψεων καθίσταται δυνατή μορφοποιώντας την χρονοσειρά προς πρόβλεψη κατάλληλα για την ένταξή της ως είσοδο (input) στο δίκτυο. Ως έξοδο (output) λαμβάνουμε την πρόβλεψη του νευρωνικού δικτύου για την χρονική περίοδο $t+1$.

5.2. Benchmarking

Προτού όμως επεξηγηθεί εκτενώς και πλήρως η ανάπτυξη της μεθόδου θα γίνει αναφορά στις μεθόδους και στα μεγέθη που χρησιμοποιήθηκαν ως σημεία αναφοράς (benchmarking) για τον έλεγχο της απόδοσης της.

Οι στατιστικές μέθοδοι πρόβλεψης που εφαρμόστηκαν στο δείγμα των δεδομένων που υπήρχαν στη διάθεσή μας ήταν οι:

1. Απλή Εκθετική Εξομάλυνση (SES), με συντελεστή εξομάλυνσης $\alpha = 0,1$.
2. Croston, στην οποία χρησιμοποιήσαμε απλή εκθετική εξομάλυνση με συντελεστή εξομάλυνσης $\alpha = 0,1$ για την παραγωγή προβλέψεων στα demands και intervals.
3. SBA, με $\alpha = 0,05$ και εφαρμογή του τύπου στην πρόβλεψη της μεθόδου Croston.
4. TSB, με παραμέτρους $\alpha = 0,3$ και $\beta = 0,3$.

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία

Εκτός των στατιστικών μεθόδων, για την καλύτερη σύγκριση με τις υπάρχουσες μεθόδους πρόβλεψης χρησιμοποιήσαμε ως benchmark και δύο μεθόδους reconciliation προκειμένου να απεικονίσουμε και τη δράση των στατιστικών μεθόδων σε συνδυασμό με αυτές σε όλα τα επίπεδα ιεραρχίας. Οι μέθοδοι reconciliation που χρησιμοποιήθηκαν είναι:

1. Bottom-Up, η οποία εφαρμόστηκε σε όλες τις μεθόδους.
2. Optimal Reconciliation (Hyndman), η οποία εφαρμόστηκε στις προβλέψεις των μεθόδων Croston και SBA.

Για τη μέτρηση σφαλμάτων δημιουργήσαμε έναν πίνακα 3x4. Σε κάθε στήλη του πίνακα απεικονίζονται τα σφάλματα των χρονοσειρών ανάλογα με το επίπεδο οργάνωσης στο οποίο βρίσκονται, δηλαδή η πρώτη στήλη απεικονίζει τα σφάλματα του επιπέδου 0 (συναθροισμένη χρονοσειρά), η δεύτερη στήλη τα σφάλματα του επιπέδου 1 και η τρίτη στήλη τα σφάλματα του επιπέδου 2 στο οποίο βρίσκονται όλες οι χρονοσειρές ξεχωριστά. Η τέταρτη στήλη χρησιμοποιήθηκε για την απεικόνιση του μέσου όρου των σφαλμάτων κάθε κατηγορίας μεταξύ των τριών επιπέδων.

Σε κάθε γραμμή του πίνακα απεικονίζονται τα διαφορετικά μέτρα σφάλματος που επιλέξαμε για την ποσοτικοποίηση της αποδοτικότητας της κάθε μεθόδου. Τα σφάλματα που επιλέχθηκαν είναι:

1. Mean Squared Error / Mean² (MSE/M²)
Αποτελεί μία παραλλαγή του MSE. Για τον υπολογισμό του απαιτείται η εύρεση του μέσου τετραγωνικού σφάλματος των προβλέψεων μιας μεθόδου και η διαίρεση αυτού με το τετράγωνο της μέσης τιμής των προβλέψεων αυτών. Δηλαδή:

$$MSE/M^2 = \frac{\sum_{i=1}^n (x_i^{forecasted} - x_i^{actual})^2}{n \cdot Mean^2}$$

2. Mean Absolute Error / Mean (MAE/M)
Αποτελεί μία παραλλαγή του MAE. Για τον υπολογισμό του απαιτείται η εύρεση του μέσου απόλυτου σφάλματος των προβλέψεων μιας μεθόδου και η διαίρεση αυτού με τη μέση τιμή των προβλέψεων αυτών. Δηλαδή:

$$MAE/M = \frac{\sum_{i=1}^n |x_i^{forecasted} - x_i^{actual}|}{n \cdot Mean}$$

3. Mean Error / Mean (ME/M)
Αποτελεί μία παραλλαγή του ME. Για τον υπολογισμό του απαιτείται η εύρεση του μέσου σφάλματος των προβλέψεων μιας μεθόδου και η διαίρεση αυτού με τη μέση τιμή των προβλέψεων αυτών. Δηλαδή:

$$ME/M = \frac{\sum_{i=1}^n (x_i^{forecasted} - x_i^{actual})}{n} = \frac{n}{Mean}$$

Ο λόγος για τον οποίον χρησιμοποιούμε κανονικοποιημένες τιμές των σφαλμάτων και όχι τις αρχικές τιμές αυτών είναι η διαφορά επιπέδου στις τιμές που λαμβάνει η κάθε χρονοσειρά. Υπάρχουν χρονοσειρές που η ζήτησή τους αποτυπώνεται σε εκατοντάδες ή χιλιάδες και άλλες που η ζήτησής τους μπορεί να φτάνει κάποιες δεκάδες. Αν, λοιπόν, δεν κανονικοποιήσουμε τα σφάλματα, το αποτέλεσμα που λαμβάνουμε θα κυμαίνεται σε εκατοντάδες ή χιλιάδες, γεγονός που δεν αντιπροσωπεύει τις χρονοσειρές με μικρότερες τιμές ζήτησης. Έτσι, κανονικοποιούμε τα σφάλματα προκειμένου να έχουμε μία μορφή «στατιστικού σφάλματος επί τοις 100».

Ο πίνακας σφαλμάτων που χρησιμοποιήσαμε, λοιπόν, μπορεί να απεικονιστεί ως εξής:

	Level 0	Level 1	...	Level k	Average
(MSE/M ²)	(1,1)	(1,2)	...	(1,k)	$\frac{(1,1) + (1,2) + \dots + (1,k)}{k}$
(MAE/M)	(2,1)	(2,2)	...	(2,k)	$\frac{(2,1) + (2,2) + \dots + (2,k)}{k}$
(ME/M)	(3,1)	(3,2)	...	(3,k)	$\frac{(3,1) + (3,2) + \dots + (3,k)}{k}$

Πίνακας 5.1 Πίνακας Σφαλμάτων

Η υλοποίηση των ανωτέρω έγινε με τη χρήση ρουτίνων. Ο κώδικας για την κάθε μέθοδο αλλά και για τον πίνακα και τα σφάλματα παρατίθεται στο Παράρτημα.

5.3. Προετοιμασία Δεδομένων

Δεδομένου ότι για τη μέθοδο πρόβλεψης χρησιμοποιούμε νευρωνικό δίκτυο θα πρέπει να προετοιμάσουμε τα δεδομένα που έχουμε στη διάθεσή μας κατάλληλα προκειμένου να αποτελέσουν την είσοδο του δικτύου μας. Από τη στιγμή που ο χρήστης της μεθόδου παρέχει στο δίκτυο αντικείμενα από το περιβάλλον στο οποίο γίνεται το πείραμα (δεδομένα) μαζί με τις επιθυμητές εξόδους πρόκειται για επιβλεπόμενη μάθηση.

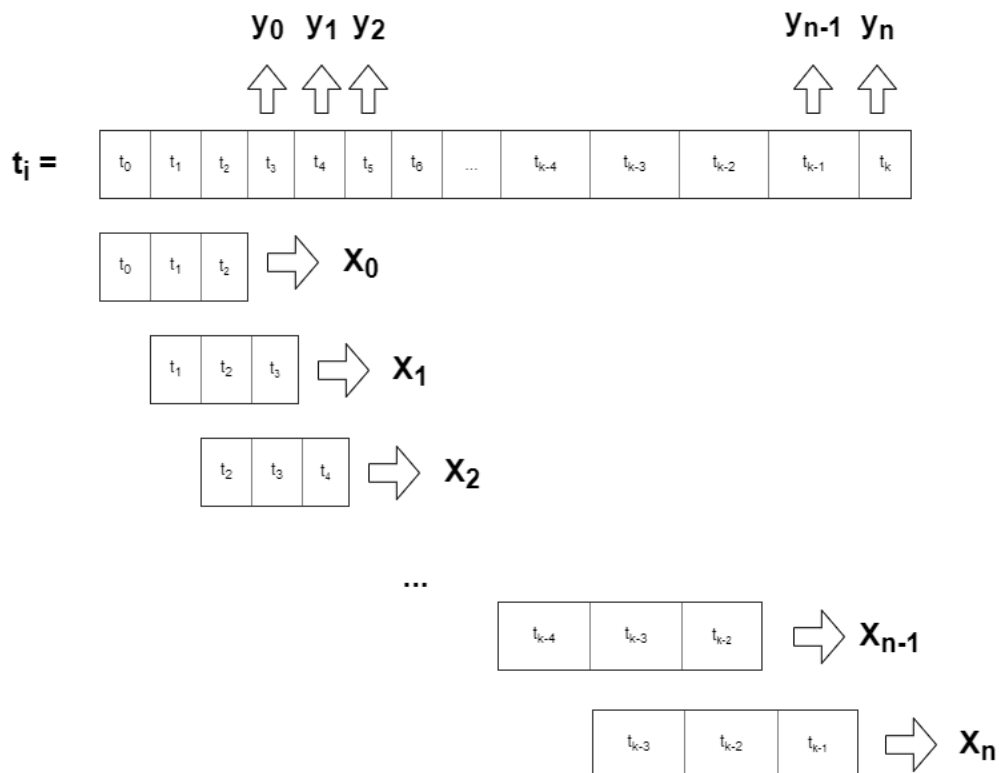
Καταρχάς, ας τονίσουμε ότι το νευρωνικό δίκτυο λαμβάνει ως εισόδους δύο πίνακες X, y. Ο πίνακας X αποτελείται από τα διανύσματα εκπαίδευσης και το διάνυσμα y περιλαμβάνει τις επιθυμητές εξόδους των διανυσμάτων αυτών με την ίδια σειρά που τοποθετούνται στον πίνακα X. Δηλαδή, αν το κάθε διάνυσμα εκπαίδευσης αποτελείται από τρία στοιχεία και το πρότυπο εξόδου από ένα, τότε το

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία

διάνυσμα x_i διάστασης 1×3 αποτελεί την i -οστή γραμμή του πίνακα X και το πρότυπο εξόδου του διανύσματος αυτού y_i τοποθετείται στην i -οστή γραμμή του πίνακα y .

Στην περίπτωσή μας, λοιπόν, τα δεδομένα που έχουμε αποτελούνται από χρονοσειρές – για την ακρίβεια από χρονοσειρές διακοπτόμενης ζήτησης. Η κάθε χρονοσειρά αποτελείται από πολλά στοιχεία τα οποία πρέπει να λάβουν την κατάλληλη στοίχιση και μορφή για να αποτελέσουν «λογικά» δεδομένα εισόδου και όχι απλά αριθμούς.

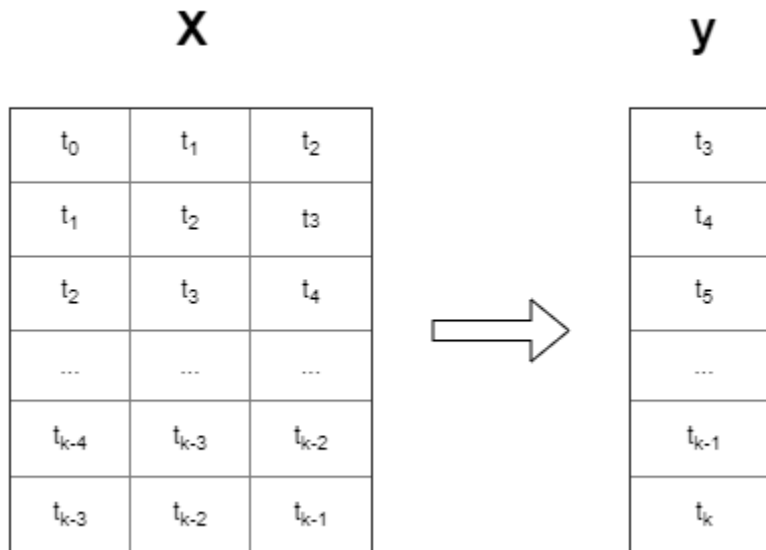
Η λογική για τον καταμερισμό και τη στοίχιση των στοιχείων, όπως συντόμως αναφέρθηκε προηγουμένως ως παράδειγμα είναι ο διαμερισμός τους σε διανύσματα συγκεκριμένου μήκους. Με τη μέθοδο αυτή, επιλέγουμε το μήκος που επιθυμούμε για κάθε διάνυσμα εκπαίδευσης και δημιουργούμε ένα «παράθυρο» με το μήκος αυτό. Έπειτα, θεωρούμε ότι μπορούμε να «δούμε» μόνο μέσω αυτού του παραθύρου τα στοιχεία της χρονοσειράς και ξεκινάμε από το πρώτο πηγαίνοντας προς το τελευταίο προχωρώντας κάθε φορά κατά ένα στοιχείο. Στο Σχήμα 5.4 απεικονίζεται ένα παράδειγμα της μεθόδου.



Σχήμα 5.4

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία

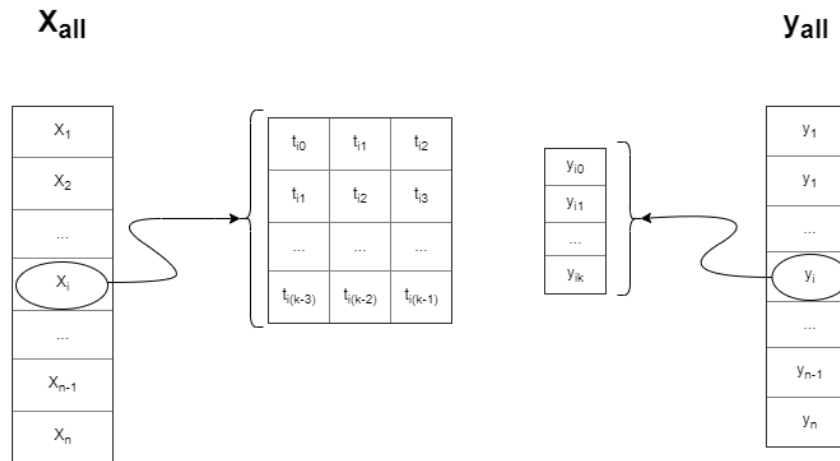
Στο παράδειγμα του σχήματος επιλέγουμε παράθυρο μήκους 3. Τα τρία πρώτα στοιχεία λοιπόν αποτελούν το πρώτο διάνυσμα εκπαίδευσης με το τέταρτο να αποτελεί το πρότυπο εξόδου. Δηλαδή, κατά μία έννοια, υποδεικνύουμε στο νευρωνικό δίκτυο ότι στο περιβάλλον που δουλεύουμε η τριάδα X_0 μας δίνει ως αποτέλεσμα το στοιχείο y_0 . Ύστερα μετακινούμαστε κατά ένα στοιχείο και η καινούρια τριάδα αποτελεί το δεύτερο διάνυσμα εκπαίδευσης και το πέμπτο το πρότυπο εξόδου. Συνεχίζουμε με τον ίδιο τρόπο έως ότου έχουμε σαν πρότυπο εξόδου την τελευταία διαθέσιμη τιμή της χρονοσειράς. Με αυτό τον τρόπο δομούνται οι δύο πίνακες που παρέχουμε σαν είσοδο στο νευρωνικό δίκτυο που έχουν τη μορφή που φαίνεται στο Σχήμα 5.5.



Σχήμα 5.5

Υπενθυμίζεται σε αυτό το σημείο ότι σκοπός της διπλωματικής εργασίας είναι ο συνδυασμός όλων των δεδομένων της ιεραρχικής δομής που έχουμε στη διάθεσή μας. Άρα η είσοδος του νευρωνικού δεν μπορεί να είναι μόνο οι πίνακες X και y μιας μεμονωμένης χρονοσειράς.

Η πρώτη, βασική και κυρίαρχη ιδέα για την εκμετάλλευση όλων των χρονοσειρών είναι η δημιουργία ενός πίνακα X και ενός πίνακα y που εμπεριέχει όλες τις χρονοσειρές. Οι δύο αυτοί πίνακες παράγονται με παράθεση των επιμέρους πινάκων των χρονοσειρών κατά στήλες. Η δομή τους παρουσιάζεται στο Σχήμα 5.6.



Σχήμα 5.6

Τελευταίο αλλά ιδιαίτερα σημαντικό βήμα στην επεξεργασία των δεδομένων ου έχουμε είναι η κανονικοποίησή τους. Όπως γίνεται εύκολα κατανοητό όταν έχουμε δεδομένα σε ιεραρχική δομή, οι χρονοσειρές του επιπέδου 0 (πλήρως συναθροισμένη χρονοσειρά), του επιπέδου 1 κ.ο.κ. θα λαμβάνουν τιμές πολύ μεγαλύτερες από αυτές των κατώτερων επιπέδων. Προκειμένου, λοιπόν, να εισάγουμε «αμεροληψία» στο σύστημα και να μην έχουμε προβλέψεις εκτός πραγματικότητας αν το νευρωνικό «μπλέκει» τα δεδομένα από τις συναθροισμένες χρονοσειρές με τις προβλέψεις των κατώτερων επιπέδων κανονικοποιούμε τις χρονοσειρές μας –των συναθροισμένων και μη επιπέδων ώστε να λαμβάνουν τιμές στο διάστημα $[0,1]$.

Η κάθε χρονοσειρά κανονικοποιείται ανεξάρτητα από τις υπόλοιπες ακολουθώντας τον εξής τύπο:

$$t'_i = \frac{t_i - t_{min}}{t_{max} - t_{min}}, i = 0, 1, \dots, k \quad (5.1)$$

όπου t_i τιμή της χρονοσειράς που τίθεται προς κανονικοποίηση, t_{min} η ελάχιστη τιμή που λαμβάνει η εκάστοτε χρονοσειρά, t_{max} η μέγιστη τιμή που λαμβάνει η εκάστοτε χρονοσειρά και t'_i η νέα κανονικοποιημένη τιμή.

Έτσι, εν τέλει, οι πίνακες X_{all} και y_{all} περιλαμβάνουν τιμές που κυμαίνονται μεταξύ 0 και 1. Μετά την προετοιμασία αυτή είμαστε έτοιμοι να προχωρήσουμε στην παραγωγή προβλέψεων.

5.4. Ανάπτυξη της Μεθόδου Πρόβλεψης

Μετά τον έλεγχο και την προετοιμασία των δεδομένων προχωράμε στην παραγωγή των προβλέψεων. Στο κομμάτι αυτό μας ενδιαφέρει η ακριβής μορφή και επεξεργασία των δεδομένων μας προκειμένου να έχουμε όσο το δυνατόν χαμηλότερα σφάλματα πρόβλεψης.

Δύο κύριες αποφάσεις που καλείται ο χρήστης να λάβει αρχικά είναι:

1. Ποιο είναι το βέλτιστο μήκος παραθύρου (window) για την παραγωγή των διανυσμάτων εκπαίδευσης.
2. Αν θα κρατήσει όλα τα διανύσματα εκπαίδευσης μαζί με τα πρότυπα εξόδου και, αν όχι, πόσα από αυτά θα κρατήσει.

Στο σημείο αυτό αξίζει να σημειωθεί ότι πολλές μέθοδοι πρόβλεψης τείνουν να δίνουν μεγαλύτερη έμφαση στα πρόσφατα δεδομένα χρησιμοποιώντας βάρη – δείκτες σημαντικότητας, προκειμένου να αποτυπώσουν με μαθηματικό τρόπο την προτίμηση αυτή. Κατ' αναλογία, στη μέθοδο που αναπτύσσεται με τη χρήση νευρωνικού δικτύου ακολουθούμε αυτή την τακτική «κόβοντας» τις πρώτες γραμμές, και κατ' επέκταση τα αρχικά δεδομένα, από κάθε πίνακα X_i .

Με τον τρόπο αυτό «εκπαιδύουμε» το νευρωνικό δίκτυο χρησιμοποιώντας ως πρότυπο τη συμπεριφορά των χρονοσειρών – τόσο του χαμηλότερου επιπέδου, όσο και των ανώτερων – στις τελευταίες χρονικές περιόδους. Χρησιμοποιούμε τον όρο «συμπεριφορά» καθώς όντας κανονικοποιημένες, οι χρονοσειρές δεν αποτυπώνουν συγκεκριμένα μεγέθη στη ζήτηση. Αντιθέτως, υποδεικνύουν πότε εμφανίζεται ζήτηση και, όταν εμφανίζεται, τι ποσοστό της μέγιστης τιμής που έχει παρατηρηθεί εμφανίζεται. Έτσι, συνδυάζουμε τη συμπεριφορά όλων των χρονοσειρών σαν να ήταν μία.

Η πρόβλεψη γίνεται αφού ολοκληρωθεί η εκπαίδευση του multilayer perceptron. Όπως αναφέραμε και στην εισαγωγή του κεφαλαίου, η εκπαίδευση γίνεται με τα δεδομένα που έχουμε έως τη χρονική στιγμή t και η πρόβλεψη γίνεται για χρονικό ορίζοντα πρόβλεψης 1, δηλαδή για τη χρονική περίοδο $t+1$. Για να γίνει αυτό παρέχουμε στο νευρωνικό δίκτυο ένα διάνυσμα με μέγεθος ίδιο με αυτό των διανυσμάτων εκπαίδευσης με τα οποία δομήσαμε τον πίνακα X . Αν, για παράδειγμα, είχαμε επιλέξει για τα διανύσματα εκπαίδευσης να έχουν μήκος μ , αυτό σημαίνει ότι και το διάνυσμα που παρέχουμε στο νευρωνικό για την πρόβλεψη θα έχει μήκος μ . Το διάνυσμα αυτό περιλαμβάνει τα μ τελευταία στοιχεία της χρονοσειράς (για την οποία θέλουμε να προβλέψουμε την τιμή στη χρονική περίοδο $t+1$), κανονικοποιημένα με χρήση της εξίσωσης (5.1).

Το νευρωνικό δίκτυο μας παρέχει με τον τρόπο αυτό μία κανονικοποιημένη τιμή η οποία αντιστοιχεί σύμφωνα με τα βάρη του στην τιμή που θα έπεται της ομάδας δεδομένων που του δώσαμε. Στη συνέχεια, με αντίστροφη διαδικασία από αυτή της εξίσωσης (5.1) αποκανονικοποιούμε την τιμή αυτή η οποία αποτελεί την στατιστική πρόβλεψη του μοντέλου στην αρχική του μορφή. Η σχέση αυτή περιγράφεται από την εξίσωση (5.2).

$$t_{min} \quad (5.2) \quad t_{k+1} = (t_{max} - t_{min}) \cdot t'_{k+1} +$$

Επαναλαμβάνουμε αυτή τη διαδικασία για όλες τις χρονοσειρές και έχουμε έτσι πρόβλεψη για όλες. Ωστόσο, από τη στιγμή που το νευρωνικό δίκτυο κάθε φορά που το εκπαιδεύουμε ξεκινά με διαφορετικές αρχικές μεταβλητές, κάθε φορά που χρησιμοποιούμε αυτή τη διαδικασία δεν παρέχει τις ίδιες προβλέψεις για κάθε χρονοσειρά. Το γεγονός αυτό μας ωθεί σε εναλλακτική παραγωγή της τελικής πρόβλεψης.

Για να είναι πιο αξιόπιστη η τελική πρόβλεψη και για να μην παρατηρούμε μεγάλες διακυμάνσεις σε αυτή κάθε φορά που χρησιμοποιούμε αυτή τη διαδικασία με τα ίδια δεδομένα, εκπαιδεύουμε πολλαπλές φορές το νευρωνικό δίκτυο, ή εκπαιδεύουμε πανομοιότυπα νευρωνικά δίκτυα και, ύστερα, επιλέγουμε μέσω των παραγόμενων προβλέψεων την τελική πρόβλεψη της μεθόδου χρησιμοποιώντας κάποια από τις εξής μεθόδους:

- Μέση τιμή των τιμών που έχουμε στη διάθεσή μας για κάθε μία πρόβλεψη.
- Median των τιμών που έχουμε στη διάθεσή μας για κάθε μία πρόβλεψη.
- Winsorizing, δηλαδή μέση τιμή των τιμών γύρω από την median τιμή.

Στα πλαίσια της διπλωματικής εργασίας, αρχικά έγινε δοκιμή της μεθόδου με χρήση μέσης τιμής όλων των τιμών που έχουμε στη διάθεσή μας από το νευρωνικό δίκτυο για μία πρόβλεψη χρονοσειράς. Ωστόσο, με τη μέθοδο αυτή υπάρχει κίνδυνος μεγάλης απόκλισης αν ένα μικρό ποσοστό των προβλέψεων έχει μεγάλη διαφορά από τις υπόλοιπες προβλέψεις.

Για το λόγο αυτό, μετέπειτα χρησιμοποιήθηκε η δεύτερη μέθοδος, δηλαδή η εύρεση της median τιμής όλων των προβλέψεων μιας χρονοσειράς. Με τη μέθοδο αυτή, επιλέγουμε ως τελική πρόβλεψη την τιμή που βρίσκεται στο μέσο του ταξινομημένου διανύσματος προβλέψεων για μία χρονοσειρά.

Τέλος, χρησιμοποιήθηκε και μία ακόμα μέθοδος, η μέθοδος Winsorizing, με την οποία προσπαθούμε να περιορίσουμε την επίδραση των ακραίων τιμών στη μέθοδο της μέσης τιμής. Χρησιμοποιούμε συνδυαστικά τις δύο προηγούμενες μεθόδους (μέση τιμή - median) βρίσκοντας την median τιμή και λαμβάνοντας τη μέση τιμή των προβλέψεων που παρουσιάζονται γύρω από αυτή. Δηλαδή, αντί να χρησιμοποιούμε όλες τις προβλέψεις, χρησιμοποιούμε μόνο αυτές που βρίσκονται κοντά στην median. Στο πειραματικό κομμάτι χρησιμοποιήσαμε ένα διάστημα της μορφής $[median - p \cdot n, median + p \cdot n]$, όπου n ο συνολικός αριθμός προβλέψεων που παρήγαγε το νευρωνικό δίκτυο και p το ποσοστό το οποίο καθορίζει πόσο μεγάλο θέλουμε να είναι το διάστημα που θα χρησιμοποιήσουμε (για το πείραμα οι τιμές που χρησιμοποιήθηκαν ήταν $p = 0.1$).

Εκτός από τη μέθοδο αυτή, σε κάθε δοκιμή παράγουμε και προβλέψεις με τη μέθοδο Bottom-Up, προκειμένου να έχουμε προβλέψεις και με αυτή τη μέθοδο reconciliation, τις οποίες χρησιμοποιούμε για να υλοποιήσουμε τη μέθοδο Optimal Reconciliation του Hyndman.

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία

Με τον τρόπο αυτό σε κάθε πείραμα έχουμε 2 πίνακες σφαλμάτων· έναν για τη μέθοδο του νευρωνικού δικτύου -το οποίο παρέχει προβλέψεις για όλα τα επίπεδα της ιεραρχικής δομής- και έναν για τη «βελτιστοποιημένη» μέθοδο του Hyndman πάνω στις προβλέψεις του νευρωνικού δικτύου, αφού έχουμε συναθροίσει τις προβλέψεις του για τα κατώτερα επίπεδα για να λάβουμε προβλέψεις για να ανώτερα.

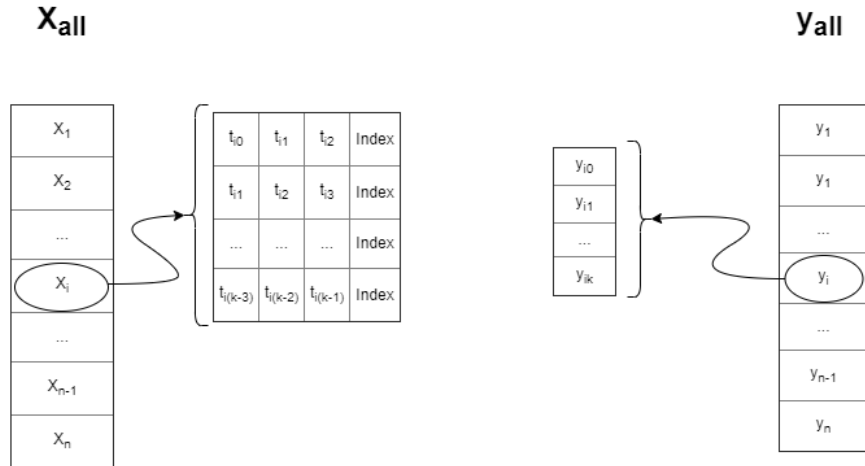
5.4.1. Παραλλαγές Μεθόδου

Η μεθοδολογία που προτάθηκε παραπάνω παρέχει προβλέψεις που παρουσιάζουν αποκλίσεις μεγαλύτερες από αυτές των συμβατικών μεθόδων πρόβλεψης διακοπτόμενων χρονοσειρών. Για το λόγο αυτό έγινε μία προσπάθεια, με αλλαγή στα δεδομένα που παρέχουμε στο νευρωνικό για την εκπαίδευσή του προκειμένου να ελαχιστοποιηθούν τα σφάλματα των προβλέψεών μας.

Οι παραλλαγές που προτάθηκαν αφορούν στον «προσδιορισμό» της κάθε χρονοσειράς από κάποιο δείκτη. Με αυτό εννοούμε ότι, παρόλο που εκπαιδεύουμε το νευρωνικό δίκτυο με όλες τις χρονοσειρές όλων των επιπέδων, εισάγουμε στα διανύσματα εκπαίδευσης κάποιες μεταβλητές προκειμένου κάθε χρονοσειρά να έχει τη δική της «ταυτότητα» εντός του πίνακα X, προκειμένου να υπάρχει μία «καθοδήγηση» του νευρωνικού προς αυτήν όταν καλείται να παράγει πρόβλεψη για εκείνη.

5.4.1.α. Χρήση Index

Για τον «προσδιορισμό» που αναφέραμε παραπάνω μία πρώτη ιδέα είναι η προσαύξηση στα διανύσματα εκπαίδευσης ενός δείκτη ξεχωριστού για κάθε χρονοσειρά. Με τον τρόπο αυτό προσθέτουμε μία στήλη στον πίνακα X στην οποία θα συμπεριλαμβάνονται οι δείκτες αυτοί. Σχηματικά, η δοκιμή αυτή απεικονίζεται στο Σχήμα 5.7.



Σχήμα 5.7

Ο δείκτης αυτός μπορεί να λάβει οποιαδήποτε τιμή επιθυμούμε, με την προϋπόθεση ότι είναι διαφορετικός για κάθε χρονοσειρά. Στο πειραματικό κομμάτι της διπλωματικής χρησιμοποιήθηκαν τιμές αναλόγως με τη θέση της χρονοσειράς στην ιεραρχία.

5.4.1.β. Χρήση ADI και CV²

Μία δεύτερη και αρκετά πιο αποδοτική προσέγγιση για τον «προσδιορισμό» των χρονοσειρών στον πίνακα X είναι τα μεγέθη ADI και CV². Με ανάλογο τρόπο με την τεχνική που αναφέρθηκε προηγουμένως, προσθέτουμε δύο στήλες αυτή τη φορά στον πίνακα X, μία για την συμπερίληψη του δείκτη ADI της συγκεκριμένης χρονοσειράς και μία για τον δείκτη CV². Σχηματικά, η δοκιμή αυτή απεικονίζεται στο Σχήμα 5.9.

Ο δείκτης ADI (Average inter-Demand Interval) πρόκειται για μία μαθηματική απεικόνιση του πόσο «αραιή» είναι μία χρονοσειρά, δηλαδή πόσο συχνά εμφανίζεται ζήτηση. Ο Δείκτης αυτός μοιάζει με τον δείκτη IDI (Intermittent Demand Intervals) της μεθόδου Croston, με τη διαφορά ότι στην μέθοδο Croston φτιάχνουμε ένα διάνυσμα με τις τιμές που λαμβάνει το IDI, ενώ ο δείκτης ADI πρόκειται για την μέση τιμή αυτών. Δηλαδή:

$$ADI = \frac{\sum_n(IDI)}{n}$$

Ο δείκτης CV² (Coefficient of variation) είναι κλασικός δείκτης στη στατιστική, γνωστός και ως σχετική τυπική απόκλιση (relative standard deviation (RSD)). Πρόκειται για έναν δείκτη που απεικονίζει πόσο μεγάλες διακυμάνσεις παρουσιάζει μία χρονοσειρά: μικραίνει όσο πιο ομοιόμορφη και σταθερού επιπέδου είναι μία χρονοσειρά και μεγαλώνει αν εμφανίζονται μεγάλες διακυμάνσεις. Ορίζεται ως το

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία

τετράγωνο της τυπικής απόκλισης διαιρεμένο με το τετράγωνο της μέσης τιμής του δείγματος που έχουμε. Δηλαδή:

$$CV^2 = \frac{s^2}{\bar{x}^2} = \frac{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2}{\bar{x}^2}$$

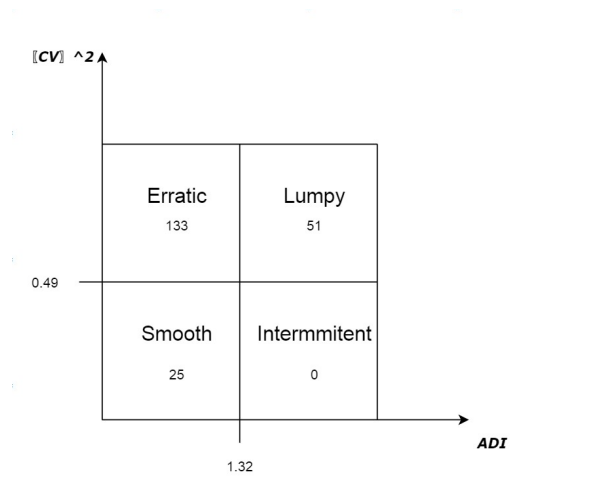
όπου x_i οι τιμές που λαμβάνει η χρονοσειρά, N το σύνολο αυτών, και \bar{x} η μέση τιμή τους.

Οι Syntetos, Boylan και Croston προτείνουν στη δημοσίευσή τους το 2005 ("On the categorization of demand patterns") ένα τρόπο κατηγοριοποίησης των διακοπτόμενων και μη χρονοσειρών με βάση το πόσο μεγάλες διακυμάνσεις εμφανίζουν και πόσο συχνά εμφανίζονται μηδενικές τιμές. Χωρίζουν τις χρονοσειρές σε 4 κατηγορίες που διαχωρίζονται με βάση πειραματικές τιμές από τα δεδομένα που είχαν στη διάθεσή τους. Συγκεκριμένα, χρησιμοποιούν ως μέτρα σύγκρισης το CV^2 που ορίσαμε παραπάνω και το p με το οποίο συμβολίζουν το ADI που περιγράψαμε παραπάνω.

Έτσι, προκύπτουν οι εξής κατηγορίες:

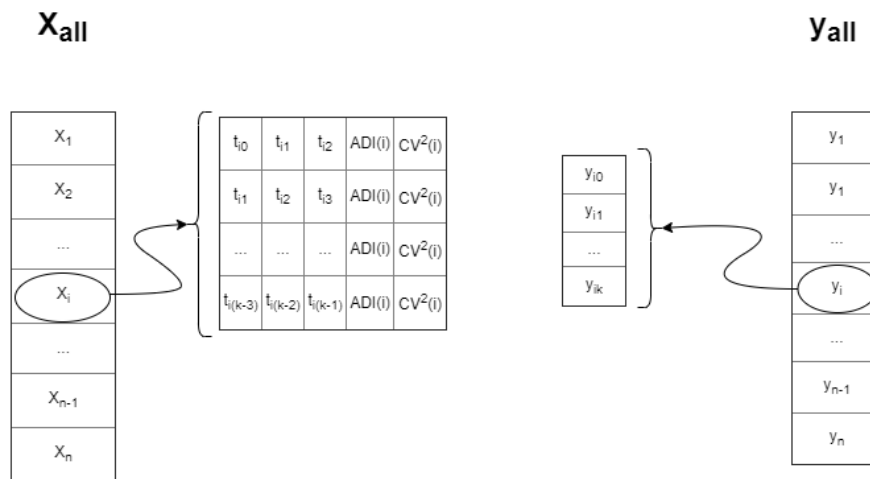
1. Περιοχή 1. Erratic
Στην περιοχή αυτή οι χρονοσειρές εμφανίζουν μεγάλες διακυμάνσεις στη ζήτηση αλλά μικρή συχνότητα εμφάνισης μηδενικών τιμών.
2. Περιοχή 2. Lumpy
Στην περιοχή αυτή οι χρονοσειρές εμφανίζουν μεγάλες διακυμάνσεις στη ζήτηση και μεγάλη συχνότητα εμφάνισης μηδενικών τιμών, γεγονός που καθιστά ιδιαίτερα δύσκολη την παραγωγή προβλέψεων για τις χρονοσειρές της κατηγορίας αυτής.
3. Περιοχή 3. Smooth
Στην περιοχή αυτή οι χρονοσειρές εμφανίζουν μικρές διακυμάνσεις στη ζήτηση και μικρή συχνότητα εμφάνισης μηδενικών τιμών.
4. Περιοχή 4. Intermittent
Στην περιοχή αυτή οι χρονοσειρές εμφανίζουν μικρές διακυμάνσεις στη ζήτηση αλλά μεγάλη συχνότητα εμφάνισης μηδενικών τιμών.

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία



Σχήμα 5.8 Τιμές αποκοπής για την κατηγοριοποίηση των Syntetos, Boylan, Croston

Στο γράφημα το Σχήματος 5.8 αναγράφονται οι περιοχές και οι τιμές που χρησιμοποίησαν οι Syntetos, Boylan και Croston για το διαχωρισμό. Επίσης, σε κάθε περιοχή αναγράφεται και η μέθοδος που φαίνεται να δουλεύει καλύτερα για κάθε κατηγορία χρονοσειρών.



Σχήμα 5.9

Ως τελικό σχόλιο για την ανάπτυξη της μεθόδου αναφέρεται ότι για καλύτερη απόδοση του MLP Regressor επιλέγουμε τις εξής παραμέτρους:

- Ορίζουμε τη συνάρτηση ενεργοποίησης να είναι σιγμοειδής, στην περίπτωσή μας 'logistic'.
- Ορίζουμε το μέγεθος των κρυφών επιπέδων να είναι το διπλάσιο από το μήκος των διανυσμάτων εκπαίδευσης, δηλαδή αν το μήκος είναι $\mu = 6$, τότε το μέγεθος των κρυφών επιπέδων θα είναι 12.

Κεφάλαιο 5. Προτεινόμενη Μεθοδολογία

- Ορίζουμε τη συνάρτηση «επίλυσης», δηλαδή τη συνάρτηση που βελτιστοποιεί τα βάρη του νευρωνικού, αντί για «adam» που δουλεύει καλά σε μεγάλα dataset σε «lbfgs» που δουλεύει καλύτερα σε πιο μικρά και συγκλίνει νωρίτερα.
- Ορίζουμε το ρυθμό μάθησης ως «adaptive». Με τον τρόπο αυτό ο ρυθμός μάθησης παραμένει σταθερός στην αρχική τιμή του όσο μειώνεται το σφάλμα μάθησης. Κάθε φορά που δύο διαδοχικά σφάλματα δεν μειώνονται αρκετά ο ρυθμός μάθησης διαιρείται με 5.

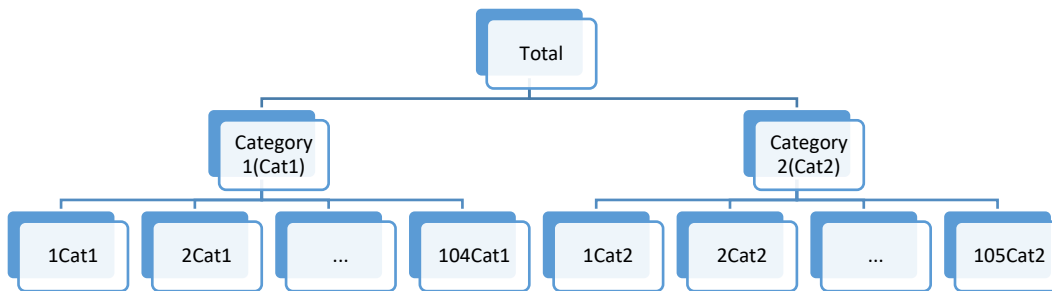
Η υλοποίηση των ανωτέρω έγινε σε μεγάλο βαθμό με τη χρήση ρουτινών. Τα τμήματα κώδικα που θεωρείται αναγκαίο να συμπεριληφθούν παρατίθενται στο Παράρτημα.

Κεφάλαιο 6 . Δεδομένα και Αποτελέσματα

6.1. Εισαγωγή

Για το πειραματικό κομμάτι της εργασίας χρησιμοποιήθηκε ένα αρχείο δεδομένων από αποθήκη κεντρικού σούπερ μάρκετ της Αθήνας. Συγκεκριμένα, τα δεδομένα (dataset) περιλάμβαναν 209 χρονοσειρές με προϊόντα από δύο διαφορετικές κατηγορίες, 104 χρονοσειρές από την κατηγορία 1 και 105 χρονοσειρές από την κατηγορία 2. Η κάθε χρονοσειρά απεικόνιζε τις ημερήσιες ποσότητες ζήτησης του εκάστοτε προϊόντος σε 300+ χρονικές στιγμές. Σημειώνεται εδώ ότι οι χρονοσειρές ήταν διακοπτόμενης ζήτησης, οπότε πολλές τιμές στα δεδομένα μας ήταν μηδενικές, ενώ σε ορισμένες χρονοσειρές υπήρχε και η ένδειξη «NA» σε ορισμένες χρονικές περιόδους, γεγονός που υποδεικνυε ότι δεν υπήρχε δεδομένο ζήτησης του εκάστοτε προϊόντος τη χρονική περίοδο αυτή.

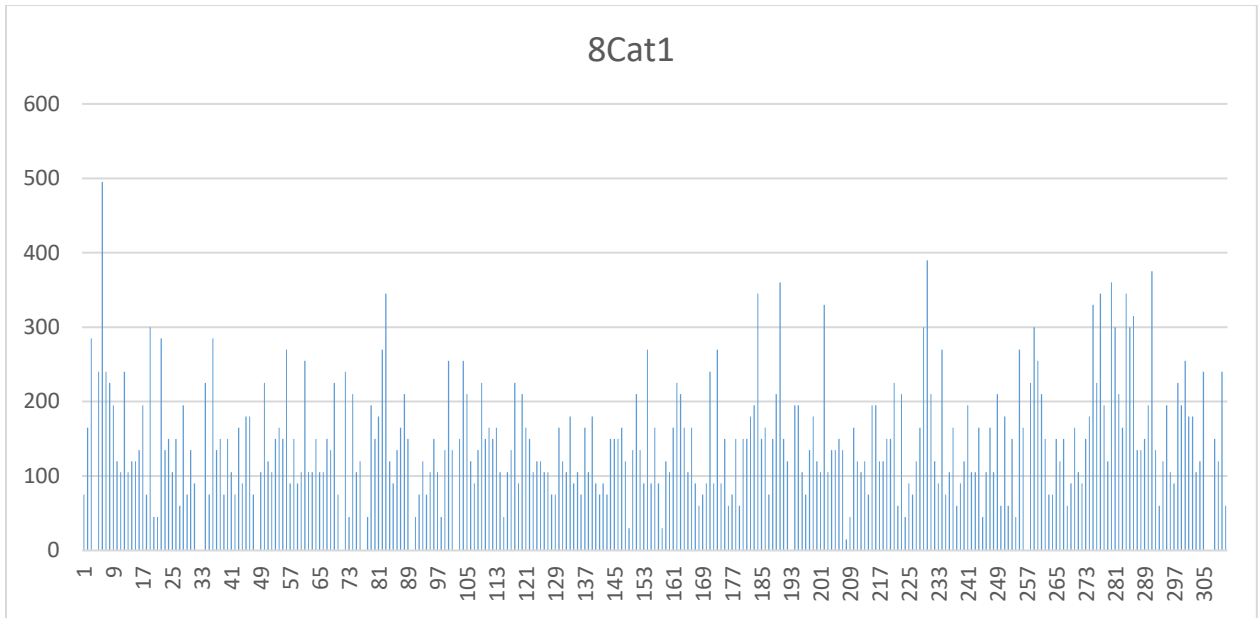
Έτσι, σύμφωνα με την ανάλυση που έγινε στο κεφάλαιο 3, η δομή που παράγεται από το συγκεκριμένο dataset μπορεί να απεικονιστεί ως εξής:



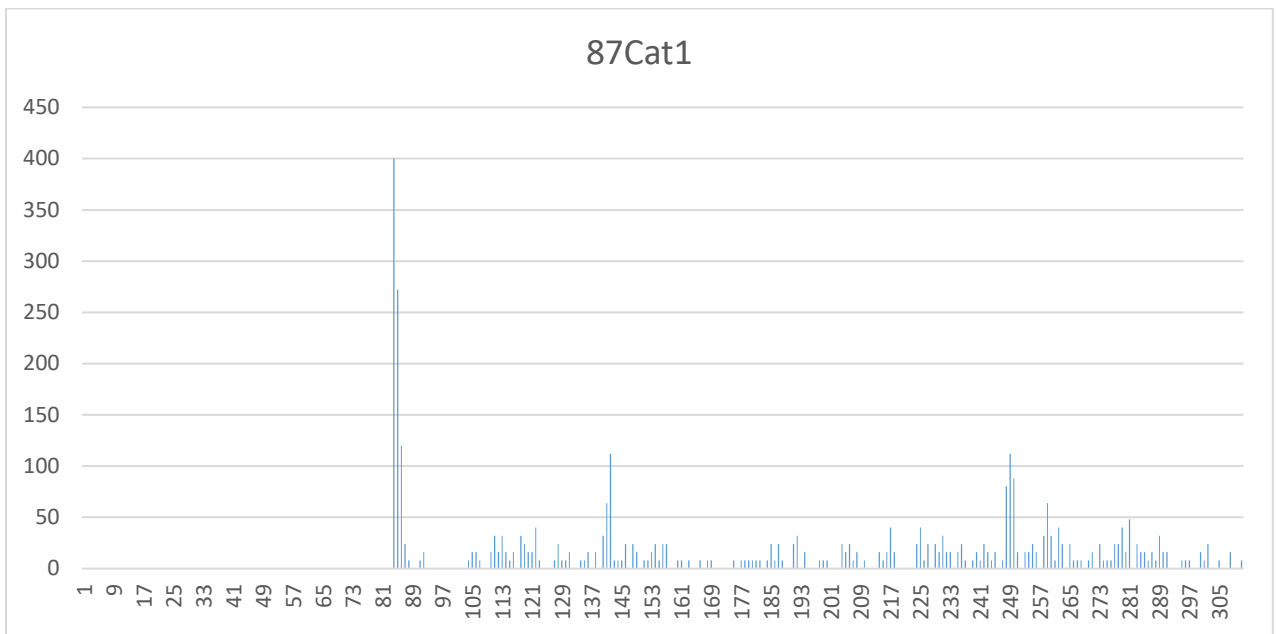
Σχήμα 6.1 Διάγραμμα της Ιεραρχίας των Δεδομένων

Το «Total» αποτελεί το άθροισμα όλων των χρονοσειρών (Επίπεδο 0), ακολούθως τα «Category 1» και «Category 2» αποτελούν το επίπεδο 1 και οι χρονοσειρές που απεικονίζουν τη ζήτηση του κάθε προϊόντος το επίπεδο 2. Ενδεικτικά παρατίθενται οι γραφικές παραστάσεις ορισμένων χρονοσειρών:

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

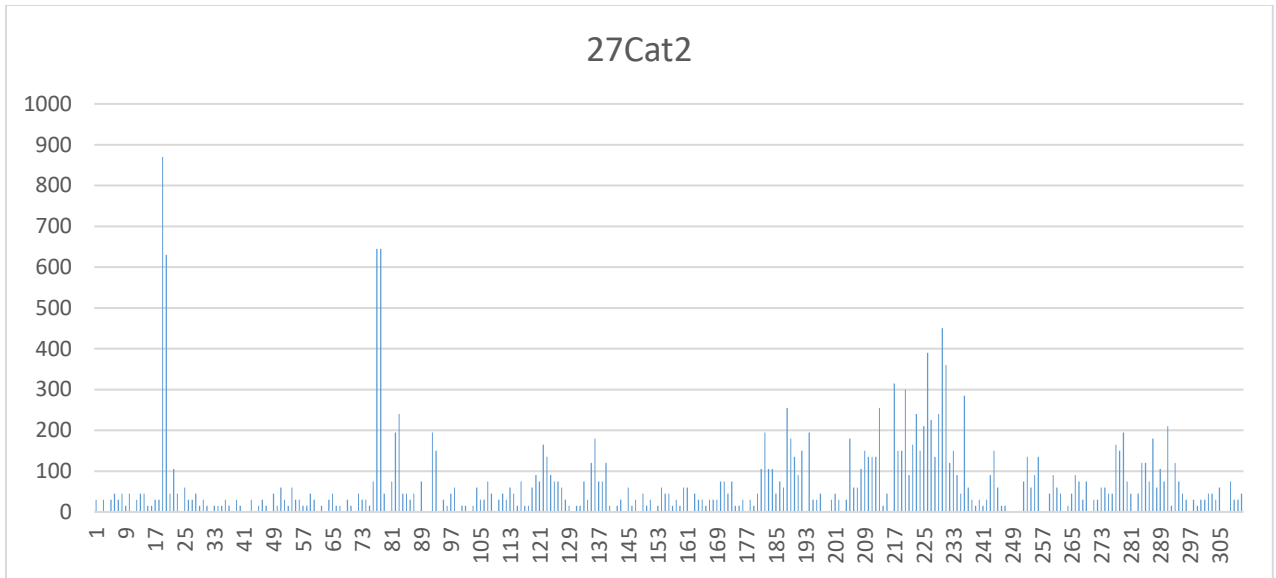


Σχήμα 6.2 Γραφική Παράσταση της Χρονοσειράς Νο 8 (Κατηγορία 1)



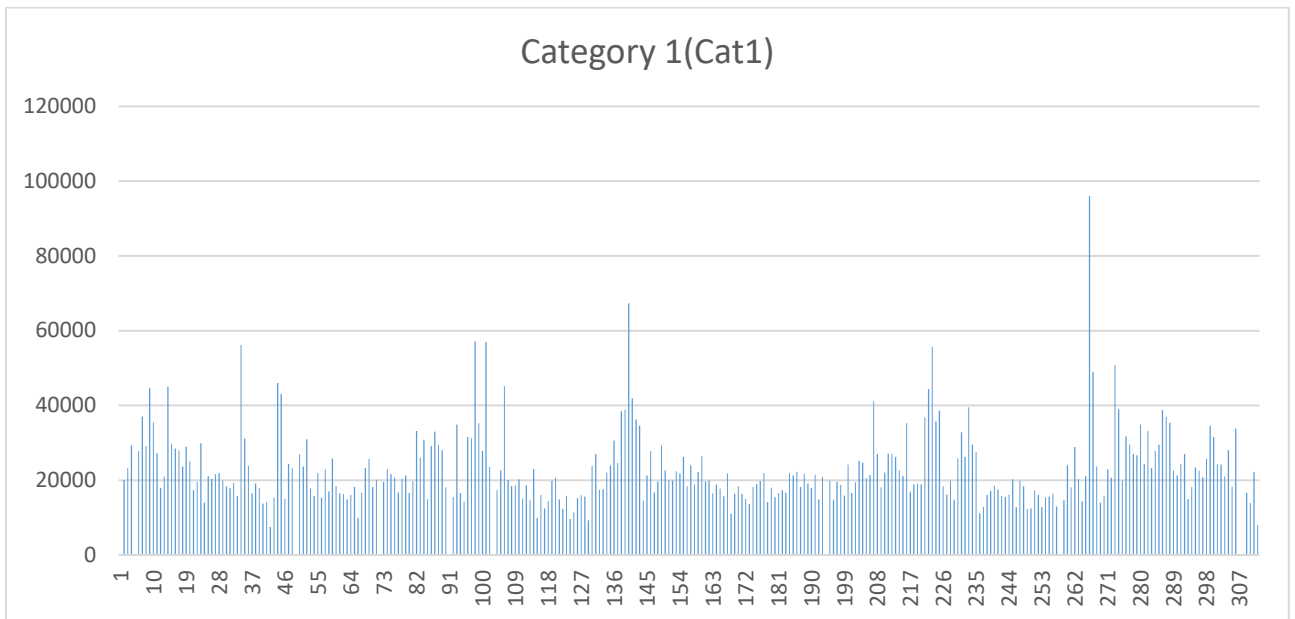
Σχήμα 6.3 Γραφική Παράσταση της Χρονοσειράς Νο 87 (Κατηγορία 1)

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα



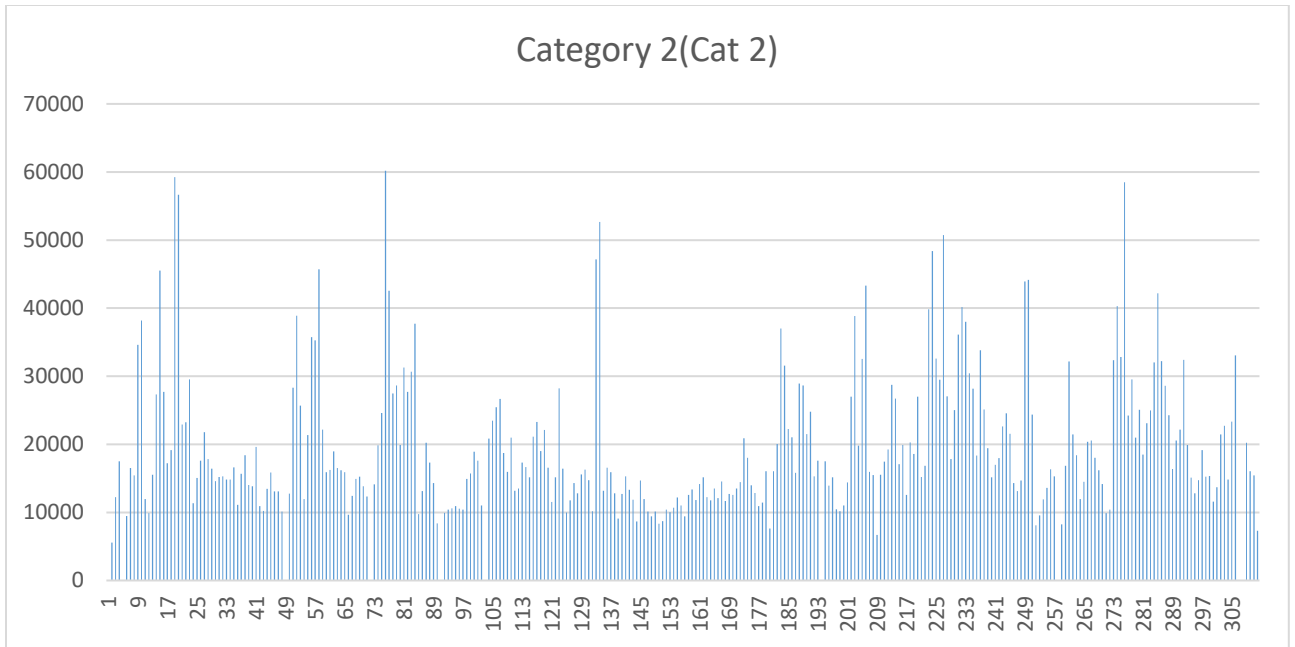
Σχήμα 6.4 Γραφική Παράσταση της Χρονοσειράς Νο 27 (Κατηγορία 2)

Αλλά και αυτές των συναθροισμένων χρονοσειρών:

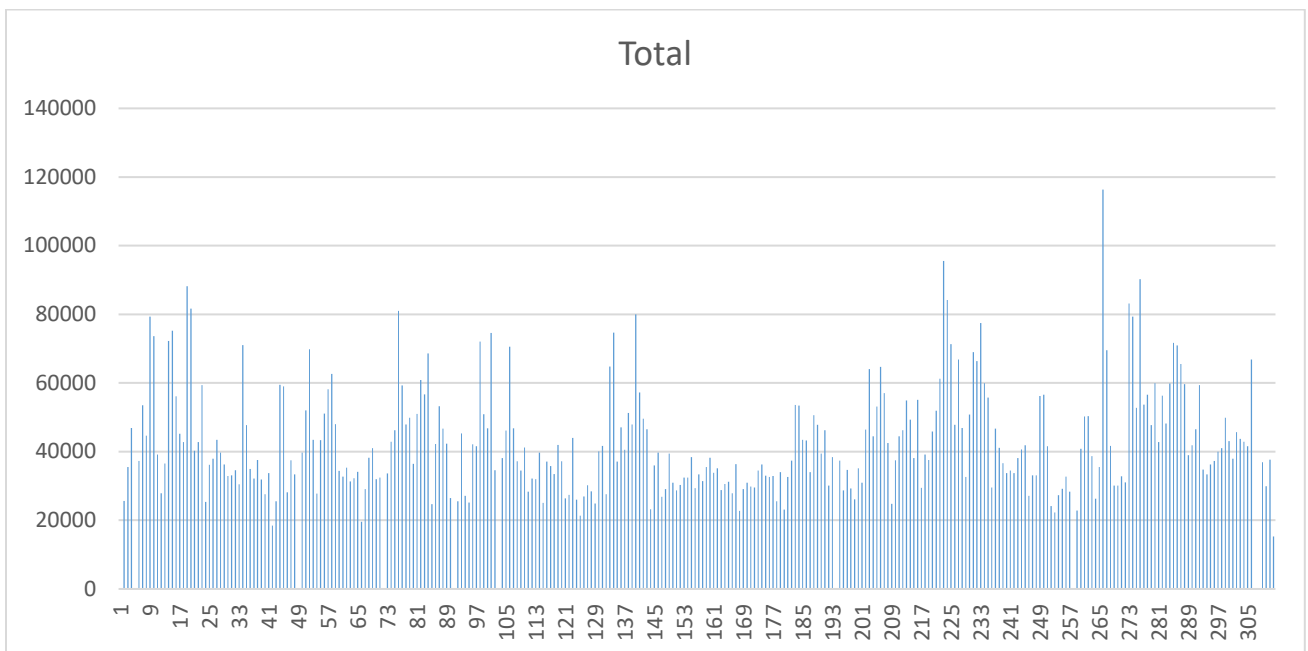


Σχήμα 6.5 Γραφική Παράσταση της Συναθροισμένης Χρονοσειράς για την Κατηγορία 1

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα



Σχήμα 6.6 Γραφική Παράσταση της Συναθροισμένης Χρονοσειράς για την Κατηγορία 2



Σχήμα 6.7 Γραφική Παράσταση της Συναθροισμένης Χρονοσειράς για τη Σύνολο των Δεδομένων

Σημειώνεται στο σημείο αυτό ότι σύμφωνα με την κατηγοριοποίηση που περιγράφηκε στο κεφάλαιο 5.4.1.β οι χρονοσειρές του dataset κατατάσσονται κατά κύριο λόγο στην κατηγορία των erratic. Συγκεκριμένα, εμφανίζουν κατά μέσο όρο $ADI = 1.24$ και $CV^2 = 1.603$ με 133 από αυτές να αντιστοιχίζονται στην κατηγορία erratic, 51 στην κατηγορία lumpry και 25 από αυτές στην κατηγορία smooth. Αντίθετα με ότι θα περίμενε κάποιος, δεν εμφανίζει καμία τιμές που θα την

περιλάμβαναν στην κατηγορία των intermittent, αλλά αυτό μπορεί να οφείλεται και στις τιμές που λαμβάνουμε ως κατώφλια για την κάθε κατηγορία (στην περίπτωση αυτή χρησιμοποιήθηκαν οι τιμές που προτείνονται και στο κεφάλαιο 5.4.1.β).

Όπως αναφέρθηκε, λοιπόν, και στο κεφάλαιο 5, η ιδέα-πρόταση που αναπτύσσεται ως θέμα στη διπλωματική αυτή είναι η δημιουργία ενός νευρωνικού δικτύου, συγκεκριμένα ενός MLP (Multi-layer Perceptron) Regressor, το οποίο λαμβάνοντας κατάλληλα ως εισόδους τις χρονοσειρές όλων των επιπέδων θα αξιοποιεί όλη την πληροφορία αυτών προκειμένου να παράγει ακριβείς προβλέψεις. Για την υλοποίηση αυτή χρησιμοποιήθηκε ως εργαλείο η γλώσσα προγραμματισμού Python V.3.7.1.

6.2. Πειραματική διαδικασία και αποτελέσματα

Η πρώτη φάση του πειράματος πάνω στα δεδομένα περιλάμβανε την «θέσπιση» κάποιων στάθμων σύγκρισης (benchmark) προκειμένου να έχουμε μία γενικότερη εικόνα της επίδρασης των υπάρχουσών τεχνικών πρόβλεψης στο συγκεκριμένο dataset – κυρίως των τεχνικών για διακοπτόμενες χρονοσειρές, αλλά και της μεθόδου SES. Συγκεκριμένα, χρησιμοποιήσαμε ως σημείο αναφοράς τα σφάλματα των μεθόδων:

- i. SES
- ii. Croston
- iii. Croston βελτιστοποιημένη με τη μέθοδο του Hyndman
- iv. SBA
- v. SBA βελτιστοποιημένη με τη μέθοδο του Hyndman
- vi. TSB
- vii. TSB βελτιστοποιημένη με τη μέθοδο του Hyndman

Για την αξιολόγηση των μεθόδων αυτών χρησιμοποιήσαμε τον πίνακα που περιγράφεται στο κεφάλαιο 5.2. Οι πίνακες που προέκυψαν για κάθε μέθοδο ήταν οι εξής:

- i. SES

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,6975	0,3516	3,2416	1,4302
(MAE/M)	0,8351	0,5891	0,6205	0,6816
(ME/M)	0,8351	0,5891	0,5891	0,6711

Πίνακας 6.1 Πίνακας Σφαλμάτων της Μεθόδου SES

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

ii. Croston

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,7169	0,3585	3,2888	1,4547
(MAE/M)	0,8467	0,5949	0,6239	0,6885
(ME/M)	0,8467	0,5949	0,5949	0,6788

Πίνακας 6.2 Πίνακας Σφαλμάτων της Μεθόδου Croston

iii. Optimized Croston

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,7127	0,3573	3,3053	1,4585
(MAE/M)	0,8442	0,5936	0,6237	0,6872
(ME/M)	0,8442	0,5936	0,5936	0,6772

Πίνακας 6.3 Πίνακας Σφαλμάτων της Optimal Reconciliation Μεθόδου εφαρμοσμένης στα αποτελέσματα της Croston

iv. SBA

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,6466	0,3335	3,1934	1,3912
(MAE/M)	0,8041	0,5735	0,6096	0,6624
(ME/M)	0,8041	0,5735	0,5735	0,6504

Πίνακας 6.4 Πίνακας Σφαλμάτων της Μεθόδου SBA

v. Optimized SBA

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,6426	0,3323	3,2095	1,3948
(MAE/M)	0,8016	0,5723	0,6095	0,6611
(ME/M)	0,8016	0,5723	0,5723	0,6487

Πίνακας 6.5 Πίνακας Σφαλμάτων της Μεθόδου Optimal Reconciliation εφαρμοσμένης στα αποτελέσματα της SBA

vi. TSB

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,5430	0,2911	3,1288	1,3210
(MAE/M)	0,7369	0,5385	0,5865	0,6206
(ME/M)	0,7369	0,5385	0,5385	0,6046

Πίνακας 6.6 Πίνακας Σφαλμάτων της Μεθόδου TSB

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

vii. Optimized TSB

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,5429	0,2911	3,1292	1,3211
(MAE/M)	0,7368	0,5385	0,5864	0,6206
(ME/M)	0,7368	0,5385	0,5385	0,6046

Πίνακας 6.7 Πίνακας Σφαλμάτων της Μεθόδου *Optimal Reconciliation* εφαρμοσμένης στα αποτελέσματα της TSB

Σημειώνεται εδώ ότι λόγω της φύσης των σφαλμάτων που χρησιμοποιούμε, οι γραμμές 1 και 2 των πινάκων αποτελούν μία ένδειξης απόκλισης των προβλέψεων από τις πραγματικές τιμές, με τη γραμμή 1 να λαμβάνει μεγαλύτερα νούμερα λόγω της ύπαρξης τετραγώνου στον ορισμό του σφάλματος, ενώ η γραμμή 3 απεικονίζει την γενικότερη προκατάληψη της μεθόδου' καθώς δεν χρησιμοποιούμε απόλυτη τιμή ή δύναμη, το πρόσημο του σφάλματος παραμένει. Έτσι, παρατηρούμε ότι όλες οι μέθοδοι που χρησιμοποιούμε στο συγκεκριμένο dataset εμφανίζουν θετική προκατάληψη στο σύνολο των προβλέψεων, αφού το πρόσημο είναι θετικό. Αναφερόμαστε στο σύνολο των προβλέψεων καθώς μπορεί κάποιες προβλέψεις (ελάχιστες) να ήταν αρνητικά προκατειλημμένες, αλλά στο σύνολο ο δείκτης σφάλματος είναι θετικός.

Επίσης, παρατηρούμε ότι η μέθοδος με τα μικρότερα σφάλματα είναι η TSB, οπότε θα αποπειραθούμε η μέθοδος πρόβλεψης να εμφανίζει σφάλματα μικρότερα αυτής της μεθόδου.

Αφού αποκτήσαμε μία πλήρη εικόνα για την αποτελεσματικότητα των υπαρχουσών τεχνικών προβλέψεων στην παραγωγή προβλέψεων για χρονικό ορίζοντα 1, μπορούμε να προχωρήσουμε στην υλοποίηση της μεθόδου.

Η αρχική απόπειρα αφορούσε στην εκπαίδευση ενός νευρωνικού δικτύου που λαμβάνει αποκλειστικά τις χρονοσειρές στη δομή που περιγράψαμε στο κεφάλαιο 5.4. Αρχικά, το μήκος των διανυσμάτων εκπαίδευσης ορίστηκε στα 3 στοιχεία ανά διάνυσμα. Έγιναν δοκιμές για το βέλτιστο αριθμό γραμμών που θα πρέπει να κρατήσουμε από κάθε πίνακα X_i και y_i για την δόμηση του τελικού πίνακα εκπαίδευσης X_{all} και διανύσματος προτύπων y_{all} . Οι δοκιμές αυτές περιλάμβαναν παρατήρηση των σφαλμάτων αφού κρατήσουμε όλες τις γραμμές από τους πίνακες, αφού κρατήσουμε 50 γραμμές από τον καθένα, αφού κρατήσουμε 15, αφού κρατήσουμε 10, αφού κρατήσουμε 5, αφού κρατήσουμε 2 και τέλος αφού κρατήσουμε 1 γραμμή από τον κάθε πίνακα.

Τα αποτελέσματα της δοκιμής αυτής προέκυψαν αρκετά μεγάλα σε σχέση με αυτά των συμβατικών μεθόδων πρόβλεψης που αναφέραμε προηγουμένως. Λογικό επακόλουθο αυτού αποτελεί η τροποποίηση της τεχνικής με στόχο των περιορισμών των σφαλμάτων σε μικρότερα νούμερα.

Για την επίτευξη του στόχου αυτού, προχωρήσαμε σε μία δεύτερη απόπειρα. Δοκιμάσαμε την αύξηση του μήκους των διανυσμάτων εκπαίδευσης, δηλαδή των παραθύρων (windows), για να δούμε πως επιδρά στα σφάλματα της μεθόδου.

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

Ακολουθήθηκε η ίδια διαδικασία, κρατώντας αρχικά όλες τις γραμμές από τους πίνακες εκπαίδευσης, ύστερα τις 50 τελευταίες γραμμές από τον καθένα και ούτω καθ' εξής μέχρι να φτάσουμε πάλι στη 1 γραμμή ανά πίνακα. Με αύξηση του μήκους των διανυσμάτων σε 6 στοιχεία ανά διάνυσμα παρατηρούμε γενικότερα μικρότερα σφάλματα από την προηγούμενη απόπειρα στην οποία το μήκος των διανυσμάτων ήταν 3 στοιχεία ανά διάνυσμα. Η διαδικασία αυτή επαναλήφθηκε με το μήκος διανυσμάτων εκπαίδευσης να παίρνει τις τιμές 9, 12 και 15. Ορισμένες εκδοχές της μεθόδου φαίνεται να παρουσιάζουν σαφώς καλύτερα σφάλματα από την αρχική απόπειρα αλλά καμία δεν παρουσιάζει σφάλματα που να την καθιστούν πιο αποδοτική στο σύνολό της από τις υπάρχουσες μεθόδους πρόβλεψης.

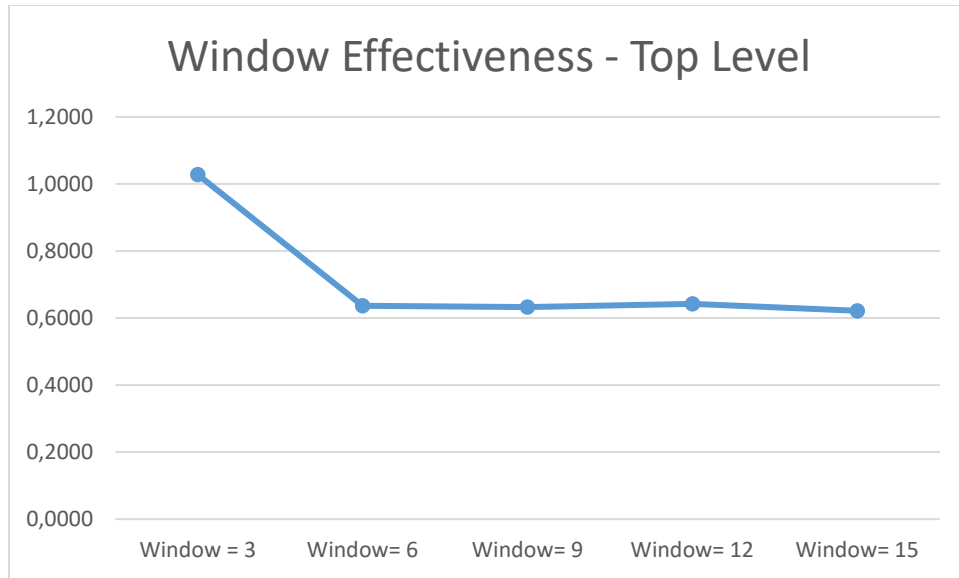
Επίσης, να σημειωθεί στο σημείο αυτό ότι κάθε φορά που εκπαιδεύουμε εκ νέου το νευρωνικό, λόγω των διαφορετικών αρχικών τιμών στα βάρη του δικτύου, τα αποτελέσματα παρουσιάζουν αποκλίσεις.

Για το λόγο αυτό περνάμε στη μεθοδολογία που αναπτύχθηκε στο κεφάλαιο 5.4. κατά την οποία εκπαιδεύουμε το νευρωνικό δίκτυο πολλαπλές φορές, ή καλύτερα εκπαιδεύουμε πανομοιότυπα νευρωνικά δίκτυα με τα δεδομένα μας πολλές φορές, και παράγουμε την τελική πρόβλεψη για κάθε χρονοσειρά χρησιμοποιώντας (i) τον μέσο όρο των παραγόμενων προβλέψεων, (ii) την median τιμή τους ή (iii) winsorizing όπως αναφέρεται στο ίδιο κεφάλαιο. Με την αλλαγή αυτή παρατηρούμε πολύ μεγάλη σταθεροποίηση των σφαλμάτων, γεγονός που τα καθιστά αξιόπιστα.

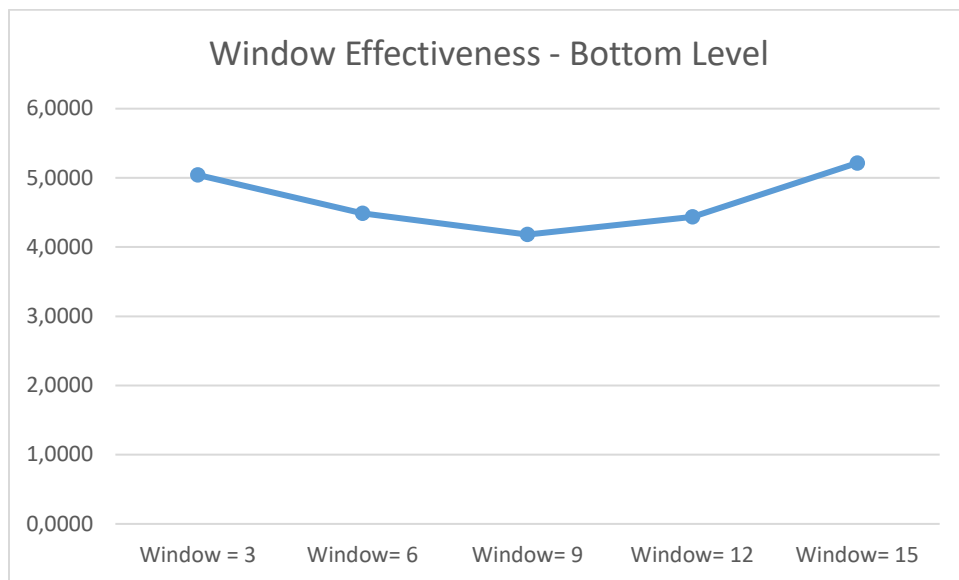
Επαναλαμβάνουμε, τη δεύτερη απόπειρα, λοιπόν, αλλά ,και πάλι, τα σφάλματα της μεθόδου δεν είναι αρκετά μικρά στο σύνολό τους ώστε να αποδεικνύεται πιο αποδοτική από τις μεθόδους πρόβλεψης, καθώς τα σφάλματά τους, τα οποία θέσαμε ως benchmarks, παραμένουν αρκετά μικρότερα.

Στα διαγράμματα που ακολουθούν φαίνεται η επίδραση του μήκους των παραθύρων που επιλέγουμε και των παρατηρήσεων που κρατάμε για κάθε πίνακα στα σφάλματα της μεθόδου. Συγκεκριμένα χρησιμοποιούμε το κανονικοποιημένο MSE, όπως προτάθηκε στο κεφάλαιο 5.2.

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

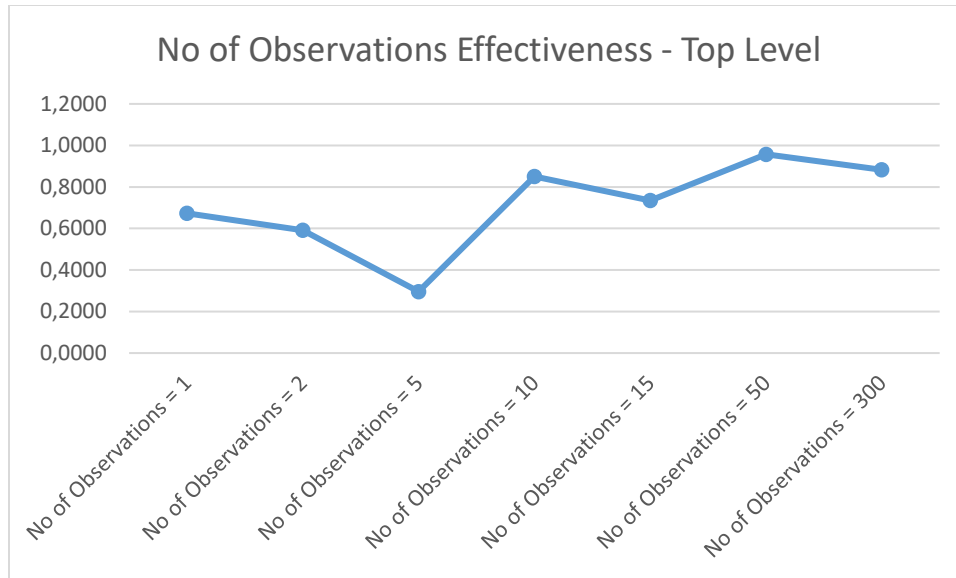


Σχήμα 6.8 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Top Level)

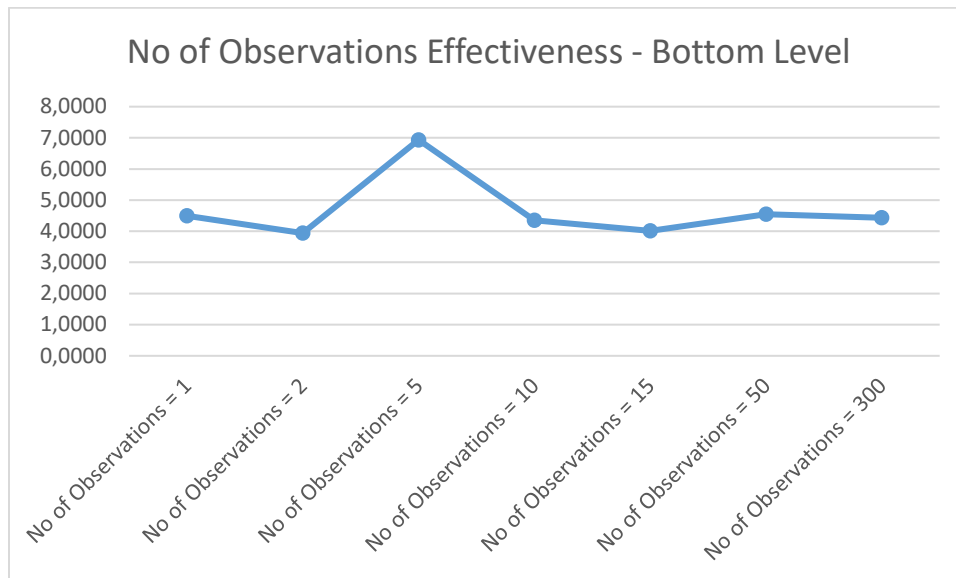


Σχήμα 6.9 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Bottom Level)

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα



Σχήμα 6.10 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Top Level)



Σχήμα 6.11 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Bottom Level)

Το γεγονός ότι τα σφάλματά τα οποία θέσαμε ως benchmarks παραμένουν αρκετά μικρότερα μας οδηγεί στο συμπέρασμα ότι η απλή παράθεση των δεδομένων που έχουμε και ο απλός συνδυασμός τους παράγουν αρκετά καλές προβλέψεις, αλλά όχι αρκετά καλές ώστε να είναι η τεχνική αυτή πιο αποδοτική από τις υπόλοιπες. Έτσι, περνάμε στην εκ νέου εκπαίδευση του νευρωνικού δικτύου με τρόπο διαφορετικό, επιδιώκοντας τον «προσδιορισμό» κάθε χρονοσειράς εντός του συνόλου με τους δείκτες που αναφέραμε στο κεφάλαιο 5.4.1.

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

Η πρώτη παραλλαγή που εφαρμόστηκε στα δεδομένα είναι αυτή που αναφέρεται στο κεφάλαιο 5.4..1.α και περιλαμβάνει τη χρήση ενός δείκτη - Index - στην κατασκευή του πίνακα εκπαίδευσης X_{all} . Συγκεκριμένα, χρησιμοποιούμε τις χρονοσειρές με βάση την ιεραρχία στην οποία υπόκεινται. Δημιουργούμε μία «ιεραρχική» λίστα με πρώτο στοιχείο να είναι το σύνολο των συναθροισμένων τιμών όλων των χρονοσειρών, δεύτερο στοιχείο το διάνυσμα των συναθροισμένων τιμών των προϊόντων της κατηγορίας 1, τρίτο στοιχείο το διάνυσμα των συναθροισμένων τιμών των προϊόντων της κατηγορίας 2 και μετά από αυτά παραθέτουμε όλες τις χρονοσειρές που περιέχονται στην ιεραρχία με αυτές της κατηγορίας 1 να προηγούνται και να παρατίθενται με βάση το δείκτη που έχουν σε αυτή και τις χρονοσειρές της κατηγορίας 2 να ακολουθούν. Έτσι, κάθε χρονοσειρά έχει ένα δείκτη στη λίστα αυτή ανάλογα με τη θέση στην οποία βρίσκεται η χρονοσειρά με τις συναθροισμένες τιμές όλων των άλλων βρίσκεται στην πρώτη θέση και για τον λόγο αυτό λαμβάνει τον δείκτη 1, η χρονοσειρά των συναθροισμένων προϊόντων της κατηγορίας 1 λαμβάνει τον δείκτη 2 κ.ο.κ..

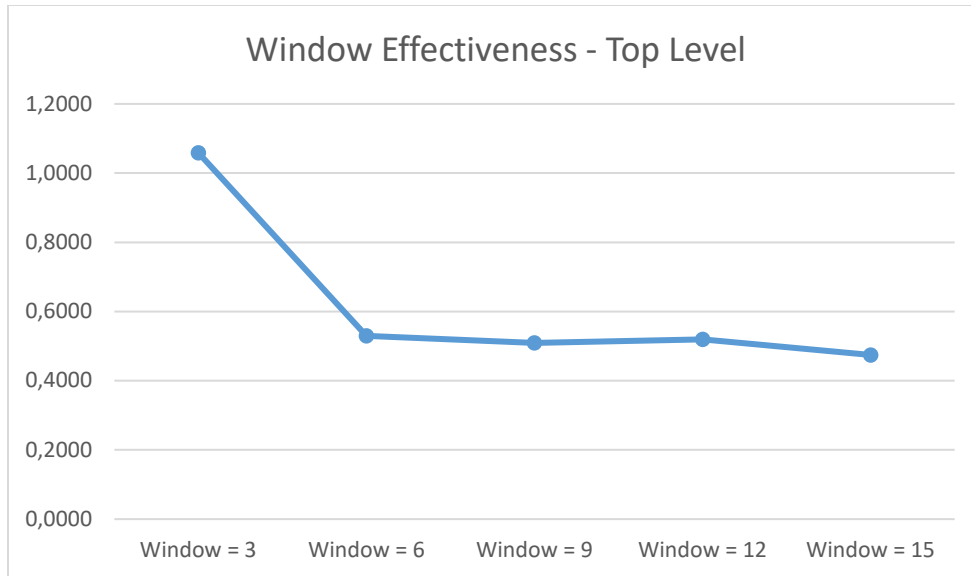
Κατά την κατασκευή των πινάκων εκπαίδευσης X_i παραθέτουμε μία παραπάνω στήλη στον καθένα που περιλαμβάνει τον δείκτη της χρονοσειράς την οποία αναφέρεται ο οποίος προκύπτει όπως αναφέραμε παραπάνω. Επιπλέον, όταν καλούμε το νευρωνικό δίκτυο να προβλέψει με την μέθοδο αυτή, χρησιμοποιούμε ένα διάνυσμα που περιλαμβάνει τις k τελευταίες τιμές της χρονοσειράς για την οποία επιθυμούμε να λάβουμε πρόβλεψη, όπου k το μήκος του διανύσματος εκπαίδευσης που έχουμε επιλέξει, στο οποίο προσθέτουμε εκ νέου στο τέλος τον δείκτη της χρονοσειράς όπως αυτός προέκυψε παραπάνω.

Η μέθοδος αυτή εφαρμόστηκε όπως και προηγουμένως, χρησιμοποιώντας αρχικά μήκος διανύσματος εκπαίδευσης 3, ύστερα 6, 9, 12 και τέλος 15. Επίσης, έγιναν εκ νέου δοκιμές χρησιμοποιώντας αρχικά όλες τις γραμμές των πινάκων X_i και y_i για την κατασκευή του συνολικού πίνακα εκπαίδευσης και του διανύσματος προτύπων, ύστερα 50 από αυτές, μετά 15, 10, 5, 2 και τέλος μόλις 1 από τον κάθε ξεχωριστό πίνακα.

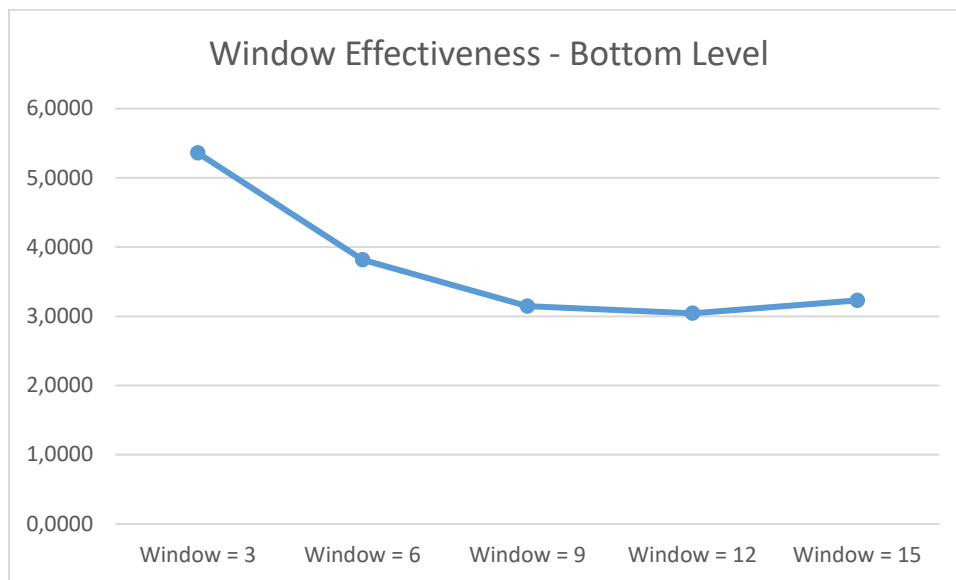
Η παραλλαγή αυτή της μεθόδου μας παρέχει κατά πολύ βελτιωμένες προβλέψεις και, κατ' επέκταση, σφάλματα συγκριτικά με αυτές των μεθόδων που χρησιμοποιούμε ως σημείο αναφοράς. Την καλύτερη επίδοση όσον αφορά στα σφάλματα φαίνεται να εμφανίζεται όταν διατηρούμε το μήκος των διανυσμάτων σε μήκος 12 στοιχεία ανά διάνυσμα και χρησιμοποιούμε 10 γραμμές από κάθε πίνακα εκπαίδευσης.

Στα διαγράμματα που ακολουθούν φαίνεται η επίδραση του μήκους των παραθύρων που επιλέγουμε και των παρατηρήσεων που κρατάμε για κάθε πίνακα στα σφάλματα της μεθόδου. Συγκεκριμένα χρησιμοποιούμε το κανονικοποιημένο MSE, όπως προτάθηκε στο κεφάλαιο 5.2.

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

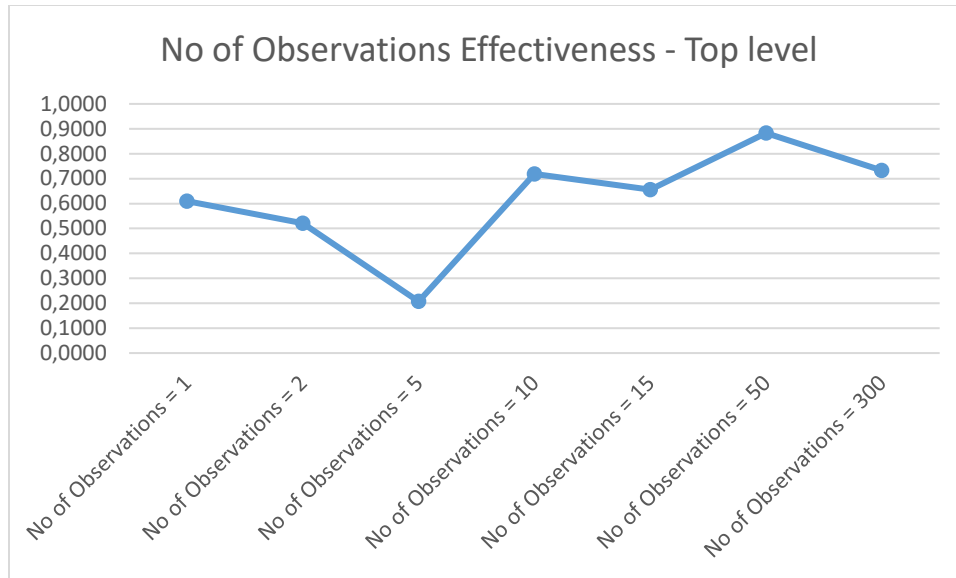


Σχήμα 6.12 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Top Level)

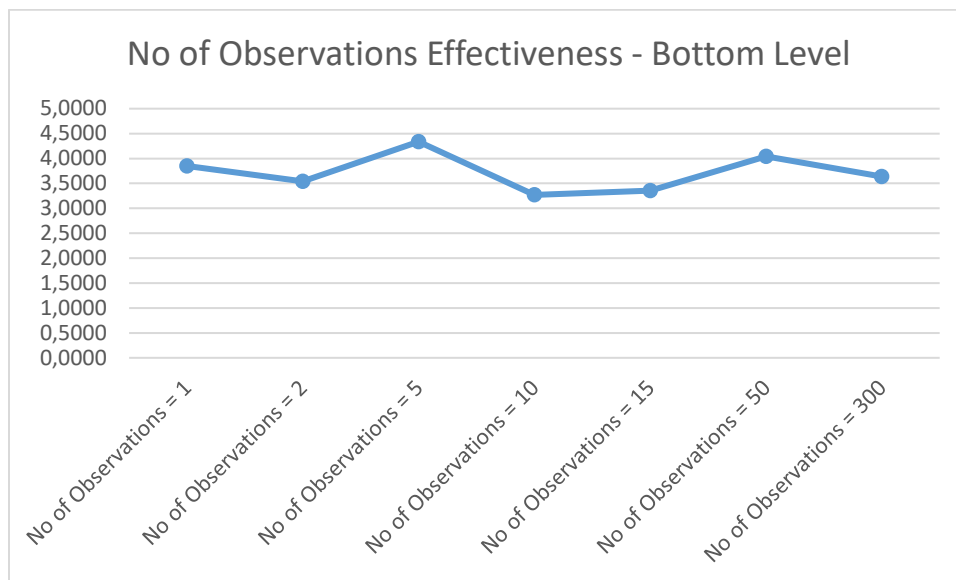


Σχήμα 6.13 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Bottom Level)

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα



Σχήμα 6.14 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Top Level)



Σχήμα 6.15 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Bottom Level)

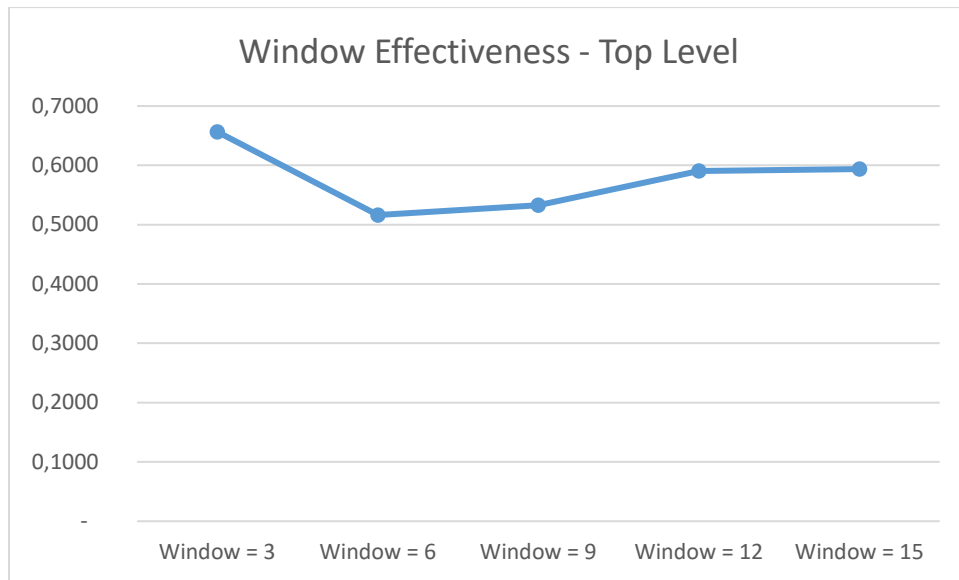
Η δεύτερη παραλλαγή που εφαρμόστηκε στα δεδομένα είναι αυτή που επίσης αναφέρεται στο κεφάλαιο 5.4. και συγκεκριμένα στο κεφάλαιο 5.4.1.β και περιλαμβάνει τη χρήση δύο μεγεθών-δεικτών – ADI και CV^2 – στην κατασκευή του πίνακα εκπαίδευσης. Η μέθοδος αυτή αποσκοπεί στον προσδιορισμό των χρονοσειρών εντός του πίνακα εκπαίδευσης, έτσι ώστε το νευρωνικό δίκτυο να «συσχετίζει» χρονοσειρές με αντίστοιχη δομή και χαρακτηριστικά. Δηλαδή, οι χρονοσειρές που έχουν παρόμοιες τιμές στα μεγέθη αυτά ξεχωρίζουν, ή τουλάχιστον το νευρωνικό δίκτυο έχει έναν παραπάνω δείκτη για να τις συσχετίσει.

Η μέθοδος αυτή βελτιώνει τις προβλέψεις του νευρωνικού σε σύγκριση με τα αρχικά αποτελέσματα. Σε γενικό επίπεδο παρουσιάζει ιδιαίτερα καλές προβλέψεις

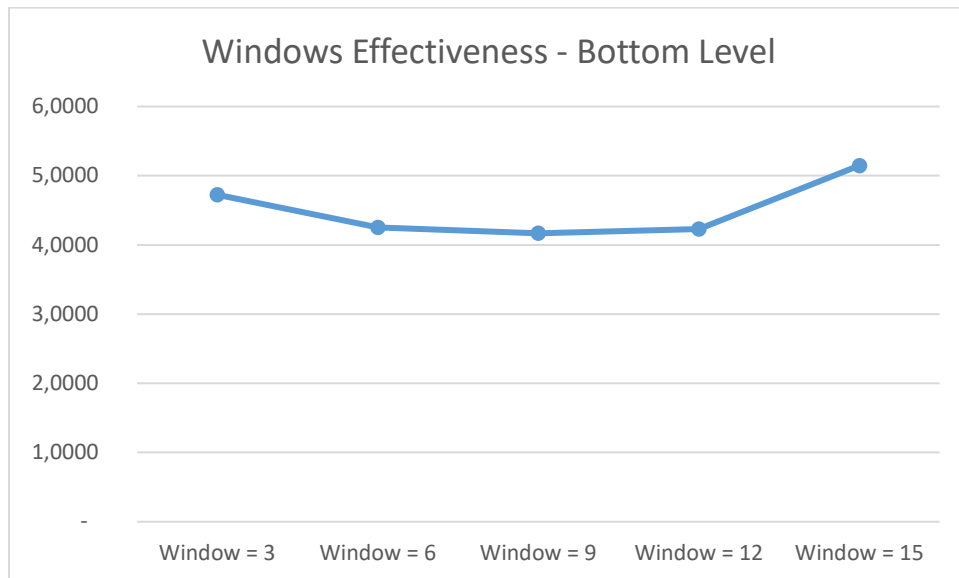
Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

στα συναθροισμένα επίπεδα (στην περίπτωση μας στα επίπεδα 0 και 1), αλλά όσον αφορά στο επίπεδο 2 τα αποτελέσματα δεν είναι αρκετά βελτιωμένα συγκριτικά με τις υπάρχουσες μεθόδους πρόβλεψης.

Στα διαγράμματα που ακολουθούν φαίνεται η επίδραση του μήκους των παραθύρων που επιλέγουμε και των παρατηρήσεων που κρατάμε για κάθε πίνακα στα σφάλματα της μεθόδου. Συγκεκριμένα χρησιμοποιούμε το κανονικοποιημένο MSE, όπως προτάθηκε στο κεφάλαιο 5.2.

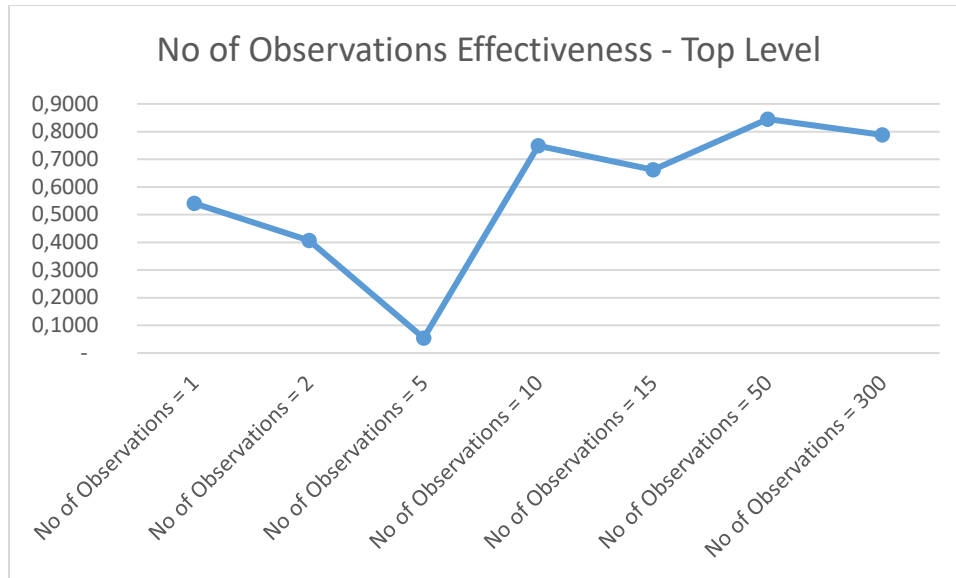


Σχήμα 6.16 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Top Level)

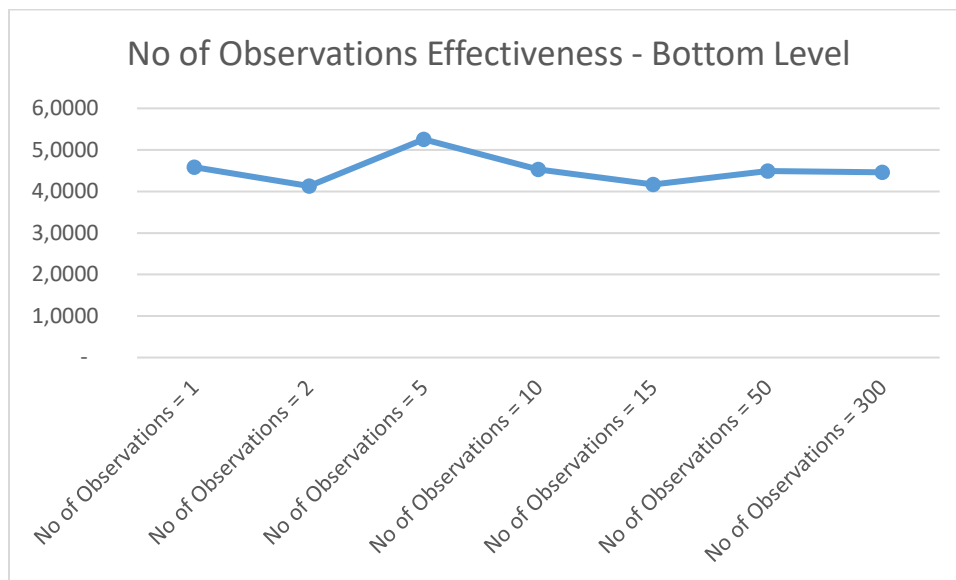


Σχήμα 6.17 Πως το μήκος παραθύρου επηρεάζει το σφάλμα (MSE Bottom Level)

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα



Σχήμα 6.18 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Top Level)

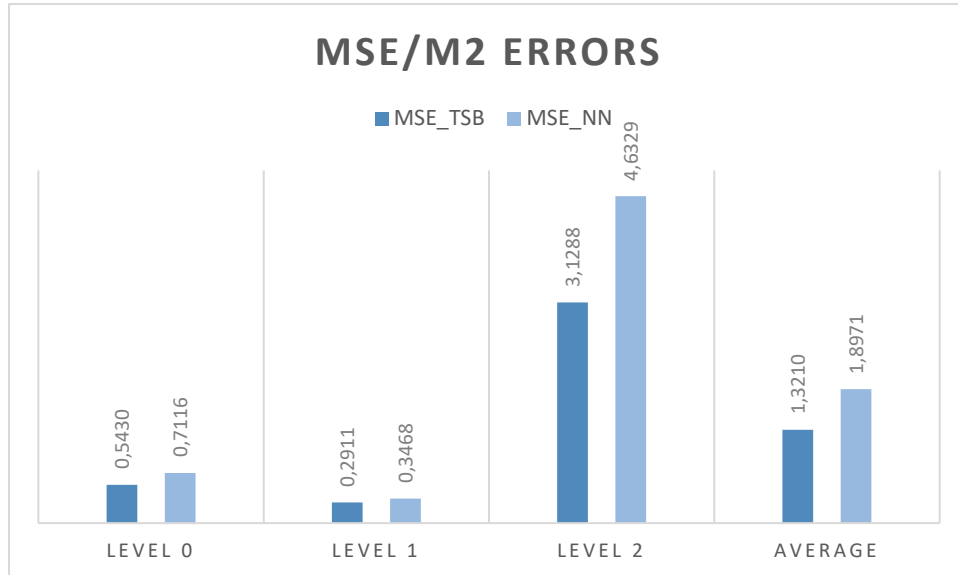


Σχήμα 6.19 Πως το πλήθος παρατηρήσεων επηρεάζει το σφάλμα (MSE Bottom Level)

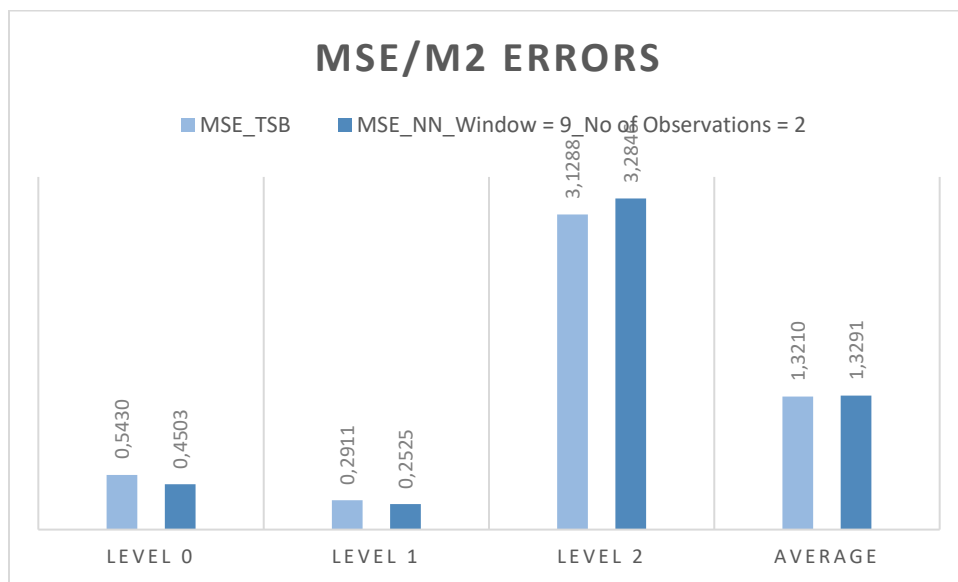
Αν οπτικοποιήσουμε, λοιπόν, τη συμπεριφορά των σφαλμάτων της μεθόδου και των παραλλαγών αυτής συγκριτικά με τα σφάλματα της μεθόδου TSB λαμβάνουμε τους πίνακες που ακολουθούν. Στο πρώτο γράφημα για κάθε μέθοδο εμφανίζεται ένας μέσος όρος των σφαλμάτων που παρήχθησαν από τις διάφορες δοκιμές στις παραμέτρους (Window, No of Observations), ενώ στο δεύτερο παρουσιάζονται τα σφάλματα που παράγει η μέθοδος με συγκεκριμένες παραμέτρους ώστε να ελαχιστοποιούνται.

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

Αρχική Μέθοδος



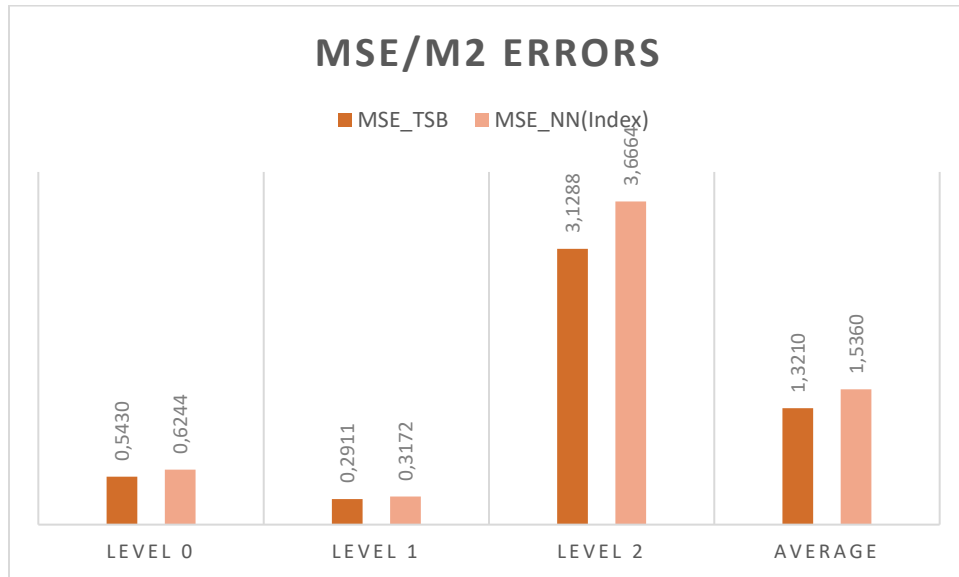
Σχήμα 6.20 Σφάλματα της μεθόδου κατά μέσο όρο σε σύγκριση με τα σφάλματα της TSB



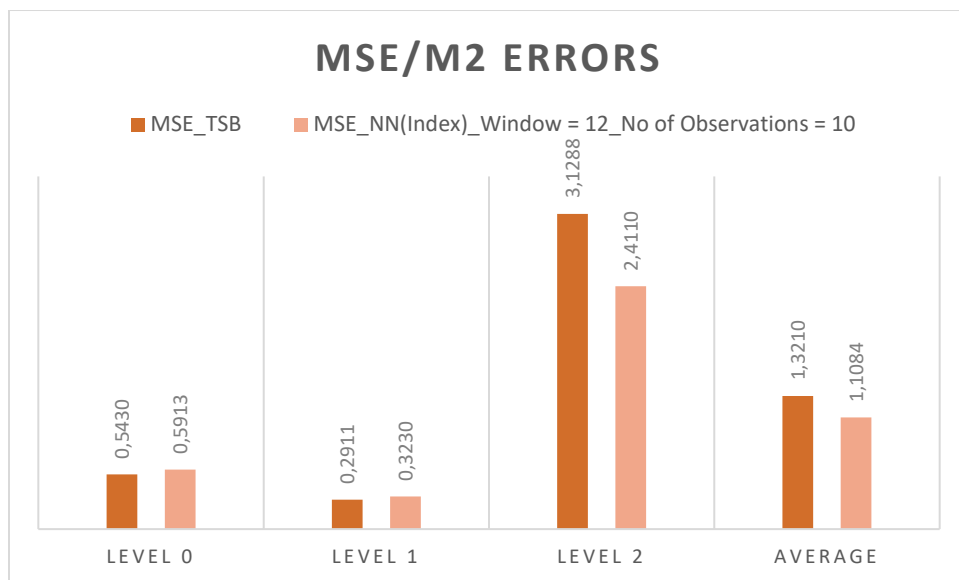
Σχήμα 6.21 Σφάλματα της μεθόδου ($W=9$, $NoO=2$) σε σύγκριση με τα σφάλματα της TSB

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

Παραλλαγή με χρήση Index



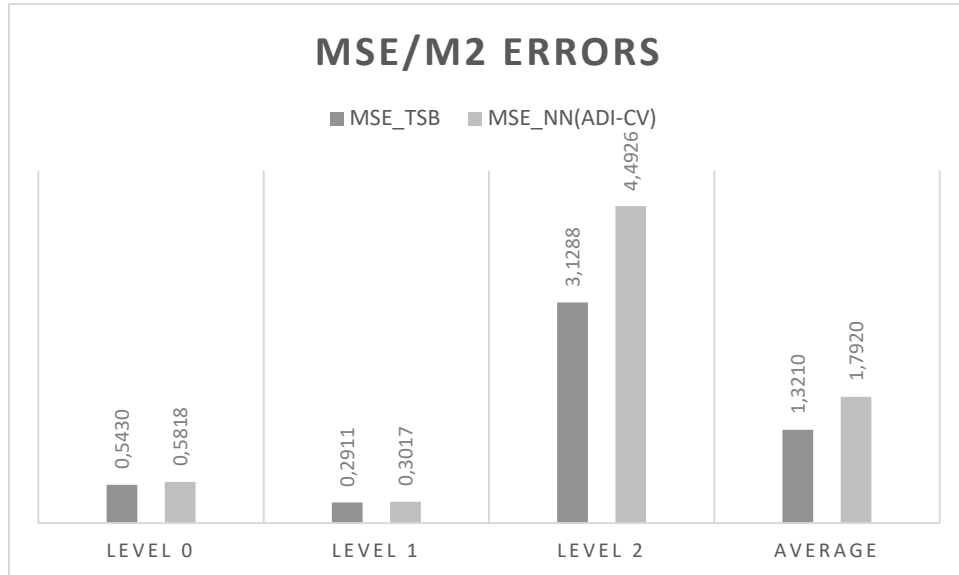
Σχήμα 6.22 Σφάλματα της μεθόδου (Index) κατά μέσο όρο σε σύγκριση με τα σφάλματα της TSB



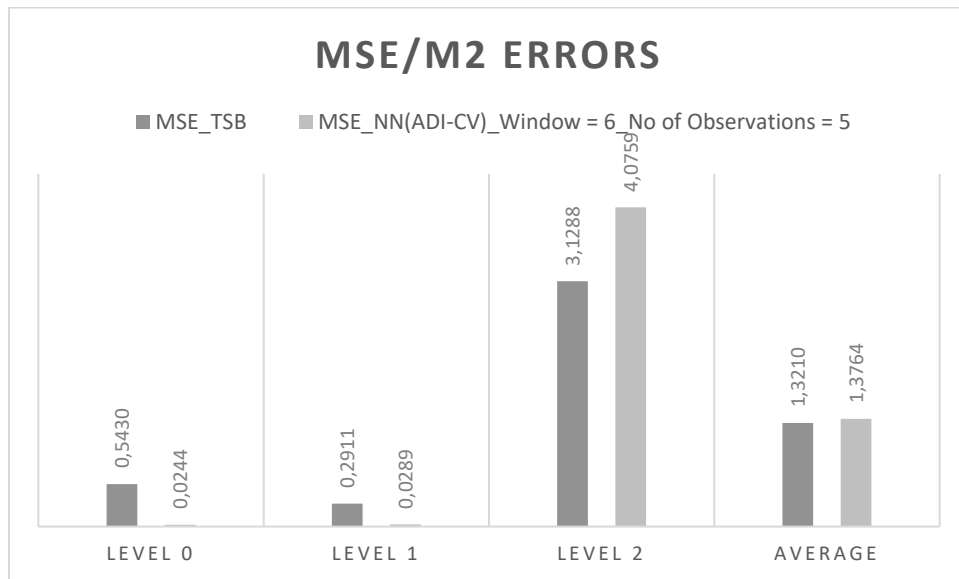
Σχήμα 6.23 Σφάλματα της μεθόδου (Index) ($W=12$, $NoO=10$) σε σύγκριση με τα σφάλματα της TSB

Κεφάλαιο 6. Δεδομένα και Αποτελέσματα

Παραλλαγή με χρήση ADI-CV²



Σχήμα 6.24 Σφάλματα της μεθόδου (ADI-CV²) κατά μέσο όρο σε σύγκριση με τα σφάλματα της TSB



Σχήμα 6.25 Σφάλματα της μεθόδου(ASI-CV²) (W=6, NoO=5) σε σύγκριση με τα σφάλματα της TSB

Στο παράρτημα παρατίθενται ενδεικτικοί πίνακες σφαλμάτων από κάθε μέθοδο αλλά και πίνακες που εμφανίζουν τις μεταβολές που παρουσιάζει η κάθε μέθοδος όσο μεταβάλλουμε τις παραμέτρους της – το μήκος του διανύσματος εκπαίδευσης ή τον αριθμό γραμμών που κρατάμε από κάθε πίνακα εκπαίδευσης αναλυτικά.

Κεφάλαιο 7 .Συμπεράσματα και Προεκτάσεις

7.1. Αποτίμηση των αποτελεσμάτων - Συμπεράσματα

Με μία πρώτη ματιά στα αποτελέσματα της μεθόδου γίνεται εύκολα σαφές ότι ένα νευρωνικό δίκτυο, και πιο συγκεκριμένα ένα Νευρωνικό Δίκτυο πολλών επιπέδων, μπορεί να αποτελέσει ιδιαίτερα αποδοτικό μέσο παραγωγής προβλέψεων, αν και εφόσον ρυθμιστούν κατάλληλα συγκεκριμένες παράμετροι. Στη μέθοδο που αναπτύχθηκε στην παρούσα διπλωματική η έμφαση δόθηκε στον τρόπο με τον οποίον παρέχουμε στο νευρωνικό τα δεδομένα που έχουμε στη διάθεσή μας προκειμένου η εκπαίδευση να αποφέρει εν τέλει ακρίβεια στην πρόβλεψη.

Ξεκινώντας από τα παράθυρα που χρησιμοποιούμε στην εκπαίδευση, περνάμε στην καθορισμό των γραμμών των πινάκων εκπαίδευσης, δηλαδή στις παρατηρήσεις, που χρειάζεται να διατηρήσουμε για την εκπαίδευση του δικτύου και καταλήγουμε στις περαιτέρω πληροφορίες που μπορούμε να παρέχουμε στο νευρωνικό δίκτυο προκειμένου να παρουσιάσει ακόμα μεγαλύτερη ακρίβεια. Όλα τα βήματα αυτά είναι αναγκαία και καθορίζουν τη λειτουργία του δικτύου.

Όσον αφορά τα αποτελέσματα, αυτά καθ' αυτά, παρατηρούμε ότι:

- Χωρίς την σωστή προετοιμασία του νευρωνικού δεν παράγουμε προβλέψεις καλύτερες από αυτές των συμβατικών μεθόδων πρόβλεψης.
- Μπορούμε να ρυθμίσουμε τις διαφορετικές παραμέτρους της μεθόδου και ως εκ τούτου καταλήγουμε σε πολλαπλές «εκδόσεις» της μεθόδου (ανάλογα με τις παραμέτρους που θα αποφασίσουμε να χρησιμοποιήσουμε) με αποτέλεσμα ορισμένες από αυτές να εμφανίζουν προβλέψεις πιο ακριβείς – ορισμένες αρκετά πιο ακριβείς- από αυτές των μεθόδων που θέσαμε ως benchmarks.

Επίσης, συμπεραίνουμε από το σύνολο της μελέτης ότι:

- Εμφανίζεται μία βελτιωμένη απόδοση των μεθόδων στα ανώτερα επίπεδα της ιεραρχίας όταν κρατάμε 5 γραμμές για κάθε πίνακα εκπαίδευσης, δηλαδή όταν No of Observations = 5.
- Παρατηρούμε ότι σε γενικές γραμμές αύξηση του μήκους των παραθύρων που χρησιμοποιούμε επιφέρει μικρότερα σφάλματα, αλλά χωρίς τεράστιες μεταβολές από το παράθυρο μήκους 9 και πάνω.

Τα αποτελέσματα και τα σφάλματα που έχουμε στη διάθεσή μας αφορούν στο συγκεκριμένο dataset, όπως αυτό περιεγράφηκε στο κεφάλαιο 6. Αυτό υποδηλώνει ότι :

- Κάποιες από τις παραλλαγές της μεθόδου που προτάθηκαν, αλλά και η ίδια η αρχική μέθοδος, παρά το γεγονός ότι δεν εμφανίζουν ακρίβεια στις προβλέψεις μεγαλύτερη από αυτήν των μεθόδων-benchmarks δεν συνεπάγεται αντίστοιχη αδυναμία σε άλλα dataset με διαφορετικά χαρακτηριστικά από αυτό που μελετάται στην παρούσα εργασία.
- Από την άλλη, ο καθορισμός των παραμέτρων της μεθόδου στις τιμές που φαίνεται να παράγουν πολύ καλύτερες προβλέψεις από τις αρχικές δεν εγγυάται αντίστοιχη αποδοτικότητα σε άλλα datasets. Ωστόσο, σε datasets με χαρακτηριστικά όπως αυτό που μελετήθηκε (διακοπτόμενη ζήτηση – ιεραρχία – lumpy και erratic χρονοσειρές), αναμένουμε αντίστοιχα καλή απόδοση της μεθόδου.

Κλείνοντας, ως τελικό σχόλιο, ας ειπωθεί ότι η πρόβλεψη με αλγορίθμους μηχανικής μάθησης φαίνεται να είναι ιδιαίτερα αποδοτική και ίσως σταδιακά αντικαταστήσει τις συμβατικές αλγοριθμικές μεθόδους. Η δυνατότητα για παραγωγή διαφορετικών μεθόδων που θα προσαρμόζονται στο dataset, αφού θα εκπαιδεύονται από αυτό κατάλληλα, υποδεικνύει ότι με σωστό χειρισμό είναι πιθανό να μπορούν να παραχθούν προβλέψεις που θα ξεπερνούν σε ακρίβεια αυτές που παράγουμε με τις τωρινές στατιστικές μεθόδους προβλέψεων.

7.2. Επίλογος

Η παρούσα διπλωματική, όπως διατυπώνεται και παραπάνω, αποσκοπεί στην ανάπτυξη μιας μεθοδολογίας παραγωγής προβλέψεων με βασικό εργαλείο ένα Νευρωνικό Δίκτυο πολλών επιπέδων. Σε αντίθεση με παλαιότερες απόπειρες πρόβλεψης με νευρωνικά δίκτυα, χρησιμοποιούμε για την εκπαίδευσή του όλο το dataset που έχουμε στη διάθεσή μας με σκοπό την πλήρη αξιοποίηση των πληροφοριών που αυτό εμπεριέχει.

Ξεκινάμε με τη μελέτη των υπάρχουσών τεχνικών προβλέψεων προκειμένου να μελετήσουμε την απόδοσή τους στο dataset που έχουμε στη διάθεσή μας, έτσι ώστε να αποτελέσουν benchmarks για τη μέθοδο που πρόκειται να αναπτύξουμε. Ύστερα, μελετάμε τα κύρια χαρακτηριστικά των νευρωνικών δικτύων που θα χρειαστεί να λάβουμε υπόψιν στην ανάπτυξη της μεθοδολογίας που προτείνεται. Με καταγεγραμμένη τη θεωρητική βάση τόσο των μεθόδων, όσο και του τρόπου λειτουργίας των νευρωνικών δικτύων και συγκεκριμένα του νευρωνικού δικτύου πολλών επιπέδων (MLP), το οποίο και χρησιμοποιούμε, μπορούμε να περάσουμε στην επεξεργασία του dataset και την εκπαίδευση του MLP.

Δομούμε μέσω των χρονοσειρών του dataset πίνακες εκπαίδευσης, χρησιμοποιώντας παράθυρα μήκους που καθορίζουμε κατάλληλα, δημιουργώντας έτσι πρότυπα με βάση τα οποία εκπαιδεύεται το νευρωνικό δίκτυο. Εν συνεχεία,

προσθέτουμε πληροφορίες της επιλογής μας προκειμένου να δράσουν επικουρικά στα δεδομένα που ήδη παρέχουμε στο δίκτυο με σκοπό την βελτιστοποίηση της απόδοσής του. Οι πληροφορίες αυτές που περιγράφονται και στα παραπάνω κεφάλαια είναι είτε ένας δείκτης που υποδεικνύει τη θέση της εκάστοτε χρονοσειράς στο dataset, είτε τα μεγέθη ADI και CV^2 όπως αυτά ορίζονται και αναλύονται στο κεφάλαιο 5.

Εν συνεχεία, μελετάμε τη συμπεριφορά του νευρωνικού δικτύου σε αλλαγές, δηλαδή πως μεταβάλλεται η ακρίβεια των προβλέψεων που αυτό παράγει όσο εμείς μεταβάλλουμε παραμέτρους που εμπλέκονται στην εκπαίδευσή του. Ενδεικτικά, δίνουμε διαφορετικές τιμές στο μήκος παραθύρου που χρησιμοποιούμε και το πλήθος παρατηρήσεων που κρατάμε από κάθε χρονοσειρά και καταγράφουμε τα αποτελέσματα.

Κλείνοντας, καταλήγουμε σε μία γενικότερη εικόνα για την μεθοδολογία που αναπτύχθηκε παρατηρώντας τις επιδράσεις των παραμέτρων και τα σφάλματα της μεθόδου και των παραλλαγών και είμαστε σε θέση να κάνουμε αποτίμηση των αποτελεσμάτων, που σε γενικές γραμμές παρουσιάζει αποτελέσματα καλύτερα του benchmarking που κάναμε, αλλά εμφανίζει και σημεία τα οποία δυνητικά μπορεί να παρουσιάσουν βελτίωση με περαιτέρω μελέτη.

Επιγραμματικά αυτά συνοψίζονται ως εξής:

- Μελέτη στατιστικών μεθόδων πρόβλεψης και καθορισμός σφαλμάτων και benchmarks
- Μελέτη νευρωνικών δικτύων και χαρακτηριστικών αυτών για την κατανόηση του δικτύου που θα χρησιμοποιήσουμε
- Επεξεργασία του dataset για την εκπαίδευση του δικτύου
- Προσθήκη βοηθητικών παραμέτρων στην αρχική μέθοδο για βελτίωση της απόδοσης
- Συνολική καταγραφή των αποτελεσμάτων και συγκριτική αποτίμηση τόσο μεταξύ των διαφορετικών παραλλαγών όσο και μεταξύ της μεθόδου και των benchmarks
- Συνολική αποτίμηση της μεθόδου

Βασιζόμενοι στο γεγονός ότι παρατηρούμε ύπαρξη χώρου για βελτιστοποίηση της μεθόδου και περαιτέρω μελέτη θεωρούμε ότι μπορούν να γίνουν κάποια βήματα. Συγκεκριμένα:

- Αντί νευρωνικού δικτύου θα μπορούσαμε να δομήσουμε μεθοδολογία για να προβλέπει ιεραρχικά δομημένες χρονοσειρές με άλλο εργαλείο, για παράδειγμα με δέντρα απόφασης (decision trees) ή με δίκτυο βοηθητικών διανυσμάτων (Support-vector machine).
- Οι παραλλαγές που προτείνουμε περιορίζονται στη χρήση δείκτη ή του ζεύγους δεικτών ADI-CV². Αντί αυτών θα μπορούσαν να χρησιμοποιηθούν άλλα μεγέθη προκειμένου να εκπαιδευτεί καλύτερα το νευρωνικό δίκτυο. Η μέγιστη και η ελάχιστη τιμή κάθε χρονοσειράς θα μπορούσαν να αποτελέσουν

Κεφάλαιο 7. Συμπεράσματα και Προεκτάσεις

ένα τέτοιο παράδειγμα ή η μέση τιμή της κάθε χρονοσειράς αντί των δεικτών που χρησιμοποιήσαμε ίσως επιφέρει καλύτερα αποτελέσματα.

Η παρούσα διπλωματική, λοιπόν, μπορεί να αποτελέσει μία καλή βάση για την πρόβλεψη σε ιεραρχικές δομές που δομούνται με χρονοσειρές διακοπτόμενης ζήτησης. Επίσης, η καινοτομία στη χρήση όλων των δεδομένων μπορεί να αποτελέσει βάση και σε αυτό το κομμάτι καθώς γίνεται παράθεση ενός τρόπου επεξεργασίας όλου του πλήθους και όχι μονομερούς επεξεργασίας που περιλαμβάνει αποκοπή πληροφορίας.

Βιβλιογραφία

- Adya, M., Collopy, F., 1998. How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting* 17 (5-6), 481–495.
- Assimakopoulos, V. and Nikolopoulos, K. The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, 16(4):521–530, Oct 2000.
- Becker, S., 1991. “Unsupervised learning procedures for neural networks”, *International Journal of Neural Systems*, issue 2, p. 17-33.
- Bernstein, D.S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Princeton, New Jersey.
- Churchland, P.S. & T.J. Sejnowski, 1992. *The Computational Brain*, Cambridge, MA: MIT Press.
- Croston, J.D., (1972).” Forecasting and stock control for intermittent demand.” *Operational Research Quarterly*, 23, 289-303.
- de Vos, N. J. (2003), *Rainfall – Runoff modeling using artificial neural networks*, M. Sc. Thesis Report, Civil Engineering Informatics Group and Section of Hydrology & Ecology, Subfaculty of Civil Engineering, Delft University of Technology.
- Demand Forecasting. TSB Method.[online] Available at: <http://warwickdf.weebly.com/id--tsb.html> [Accessed 24 Jan. 2020]
- Forecasting & Strategy Unit. *Σημειώσεις Μαθήματος Τεχνικών Προβλέψεων*, 2017.
- Gutierrez, R. S., Solis, A. O., Mukhopadhyay, S., 2008. Lumpy demand forecasting using neural networks. *International Journal of Production Economics* 111 (2), 409 – 420.
- Haykin, S., 2002. *Adaptive Filter Theory*, 4th Edition, Englewood Cliffs, NJ: Prentice Hall
- Haykin, S., «Νευρωνικά Δίκτυα και Μηχανική Μάθηση», 3^η Έκδοση. Εκδόσεις Παπασωτηρίου, Αθήνα 2010.
- Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. *Neural Networks* 4 (2), 251–257.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2 (5), 359–366.
- Kerlirzin, P. & F. Vallet, 1993. “Robustness in multilayer perceptrons”, *Neural Computation*, issue 5, p. 473-482
- Kourentzes, N., 2013. “Intermittent Demand Forecasts with Neural Networks”, *International Journal of Production Economics*, 143: 198-206.
- Kramer, A.H. & Sangiovanni-Vincentelli, 1989. “Efficient parallel learning algorithms for neural networks”, *Advances in neural Information Processing Systems*, issue 1, p. 40-48, San Mateo, CA: Morgan Kaufmann.

Βιβλιογραφία

- Markham, I. S., Rakes, T. R., April 1998. The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression. *Computers & Operations Research* 25, 251–263.
- Minsky, M.L. & S.A. Papert, 1969. *Perceptrons*, Cambridge, MA: MIT Press.
- Nikolopoulos, K. and Assimakopoulos, V. Generalizing the Theta model. Unknown, 2004.
- Nikolopoulos, K., Syntetos A.A., Boylan J.E., Petropoulos F., Assimakopoulos V.(2011) “An aggregate disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis”. *Journal of the Operational Research Society*; 62:544–54.
- NumPy Developers. *Numpy*, 2019.
- Petropoulos F., Nikolopoulos K. , Assimakopoulos V.,(2008). “An evaluation of the accuracy of intermittent demand forecasting approaches”. *Proceedings of the 28th International Symposium on Forecasting, Nice (France)*
- Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679 – 688, 2006.
- Rob J. Hyndman*, Roman A. Ahmed, George Athanasopoulos, Han Lin Shang. Optimal combination forecasts for hierarchical time series. March 2008
- scikit-learn developers. *scikit-learn*, 2019.
- Syntetos, A.A. and Boylan, J.E., (2001). “On the bias of intermittent demand estimates. *International Journal of Production Economics*”, 71, 457-466.
- Syntetos, A.A. and Boylan, J.E., (2005). The accuracy of intermittent demand estimates”. *International Journal of Forecasting*, 21, 303-314.
- Syntetos, A.A., Boylan, J.E., Croston, J.D., 2005. On the categorization of demand patterns. *Journal of the Operational Research Society* 56, 495-503.
- Teunter, R. and Sani, B., (2009). “On the bias of Croston’s forecasting method.” *European Journal of Operational Research*, 194, 177-183.
- Tian, Y., Wiens, D.P., 2006. On equality and proportionality of ordinary least squares, weighted least squares and best linear unbiased estimators in the general linear model. *Statistics and Probability Letters* 76 (12), 1265–1272.
- Werbos, P.J., 1992. “Neural networks and the human mind: New mathematics fits humanistic insight”, *IEEE International Conference on Systems, Man, Cybernetics*, issue 1, p. 78-83, Chicago.
- Widrow , B & S.D. Stearns, 1985. *Adaptive Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall
- Zhang, G., Patuwo, B. E., Hu, M. Y., 1998. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting* 14 (1), 35 – 62.
- Φώτιος Πετρόπουλος, Βασίλειος Ασημακόπουλος, «Επιχειρησιακές Προβλέψεις». Εκδόσεις Συμμετρία, Αθήνα 2013

Παράρτημα

Μέρος 1. Πίνακες Σφαλμάτων

Παρουσιάζονται ενδεικτικά κάποιοι πίνακες σφαλμάτων των μεθόδων που αναπτύχθηκαν χρησιμοποιώντας το νευρωνικό δίκτυο καθώς και πίνακες που παρουσιάζουν τη μεταβολή των σφαλμάτων, άρα και των αποτελεσμάτων της κάθε μεθόδου, όσο μεταβάλουμε ορισμένα χαρακτηριστικά της.

Αρχική Μέθοδος

Ενδεικτικοί Πίνακες Σφαλμάτων για Συγκεκριμένες Παραμέτρους

Μήκος Διανύσματος Εκπαίδευσης (Window) = 9
Γραμμές από κάθε πίνακα X_i (No_of_observations) = 2

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,4455	0,2504	3,2919	1,3293
(MAE/M)	0,6674	0,5004	0,5944	0,5874
(ME/M)	0,6674	0,5004	0,5004	0,5561

Πίνακας 0.1 Πίνακας Σφαλμάτων (MLP) W=9, NoO = 2

Μήκος Διανύσματος Εκπαίδευσης (Window) = 6
Γραμμές από κάθε πίνακα X_i (No_of_observations) = 5

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,1585	0,1125	5,4025	1,8912
(MAE/M)	0,3981	0,3321	0,7736	0,5013
(ME/M)	0,3981	0,3321	0,3320	0,3541

Πίνακας 0.2 Πίνακας Σφαλμάτων (MLP) W=6, NoO= 5

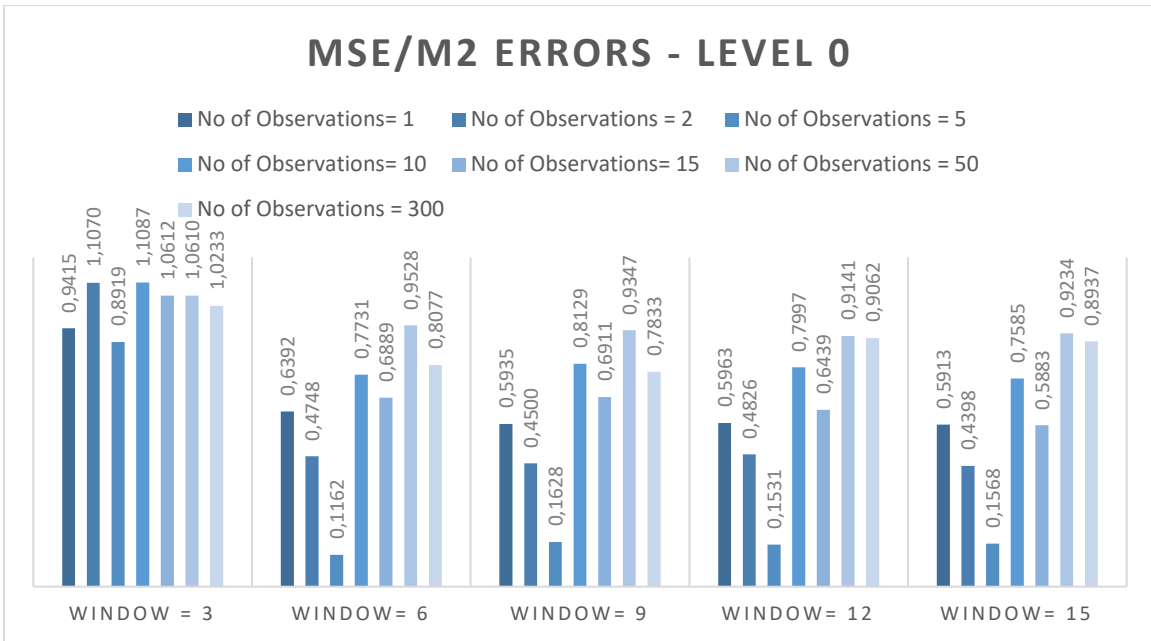
Μήκος Διανύσματος Εκπαίδευσης (Window) = 15
Γραμμές από κάθε πίνακα X_i (No_of_observations) = 2

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,4484	0,2697	3,7022	1,4734
(MAE/M)	0,6696	0,5016	0,5940	0,5884
(ME/M)	0,6696	0,5016	0,5016	0,5576

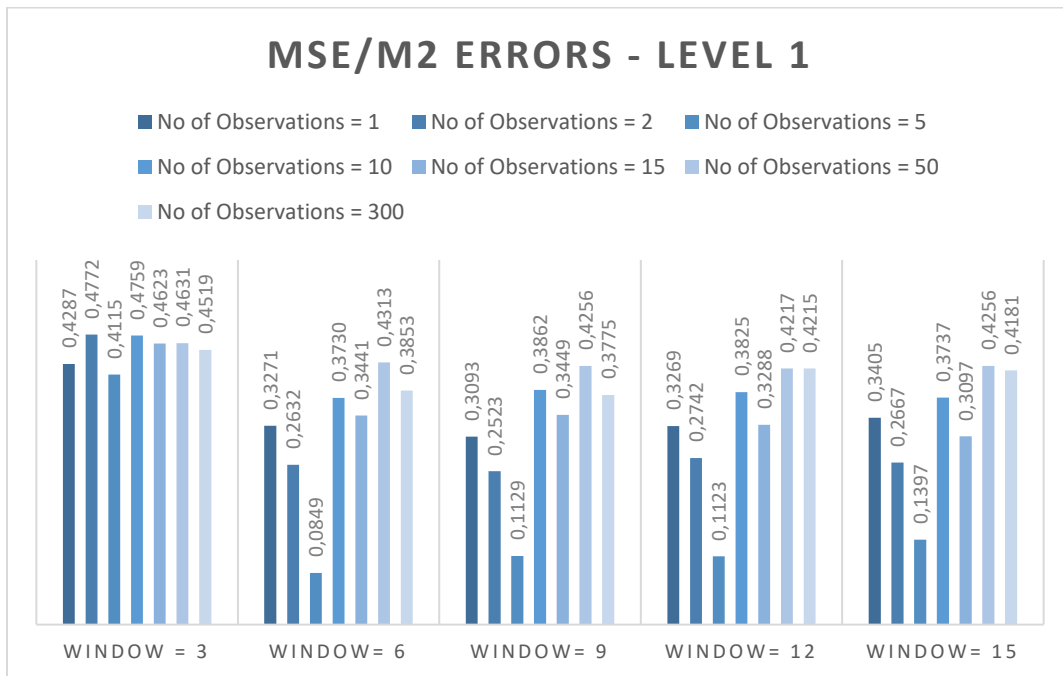
Πίνακας 0.3 Πίνακας Σφαλμάτων (MLP) W=15, NoO = 2

Μεταβολή Σφαλμάτων με μεταβολή των παραμέτρων της μεθόδου

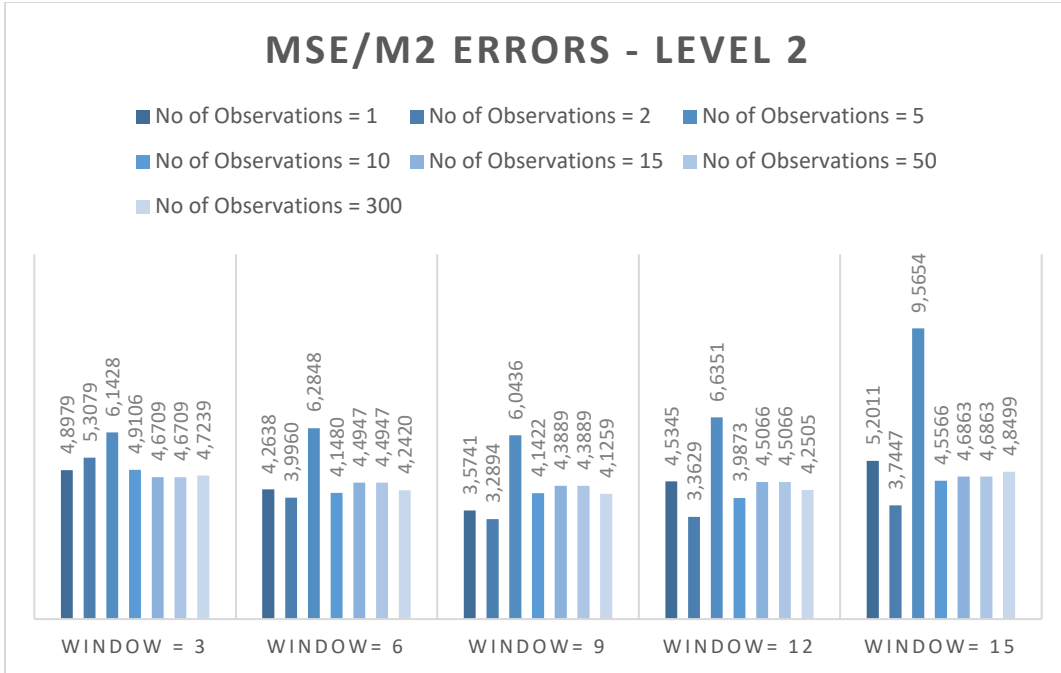
MSE/M²



Σχήμα 0.1 MSE/M² Errors – Level 0 (MLP)

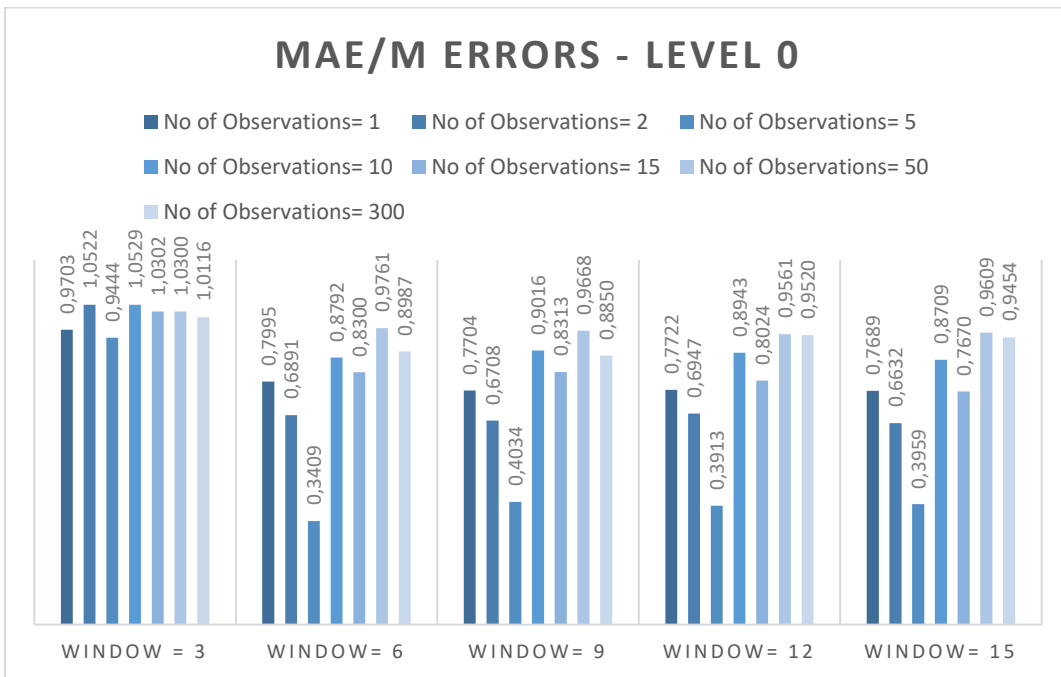


Σχήμα 0.2 MSE/M² Errors – Level 1 (MLP)

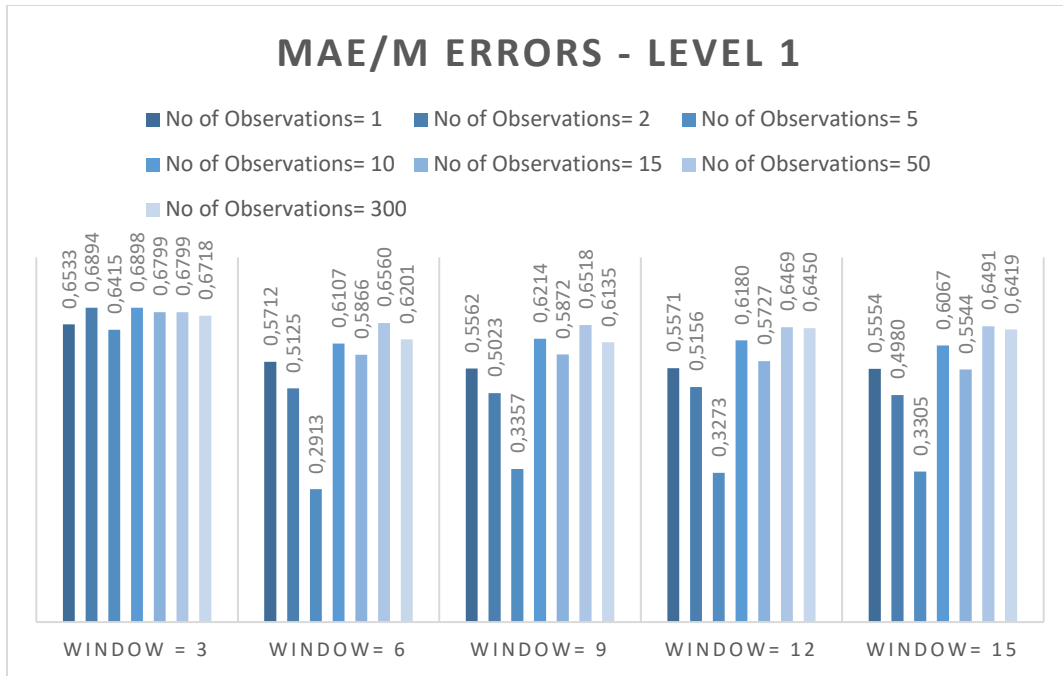


Σχήμα 0.3 MSE/M² Errors – Level 2 (MLP)

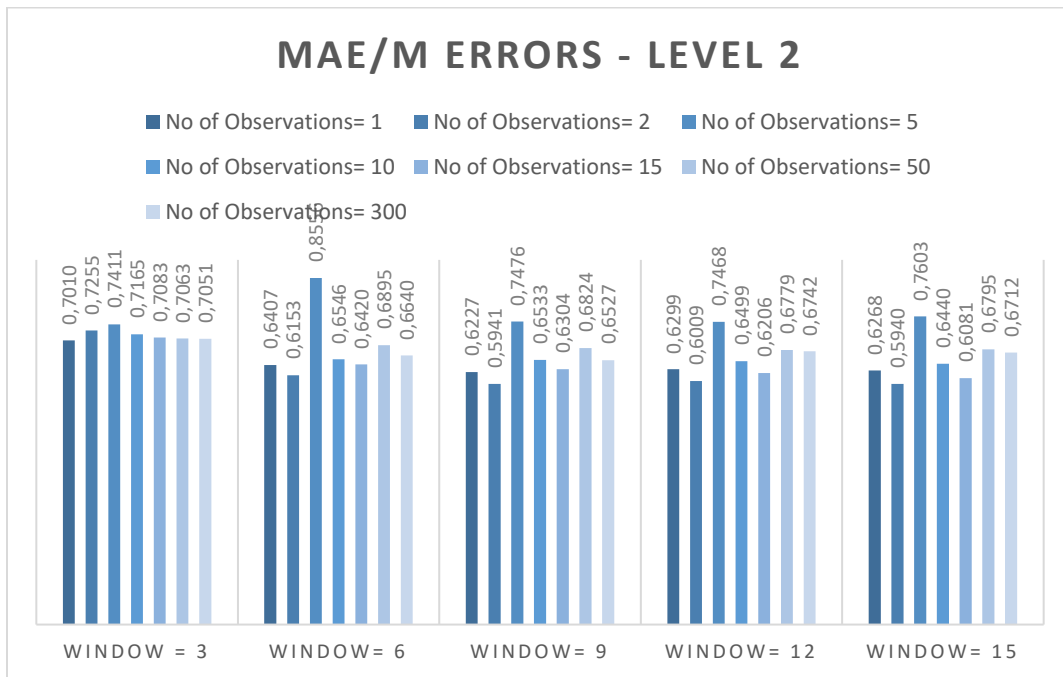
MAE/M



Σχήμα 0.4 MAE/M Errors – Level 0 (MLP)

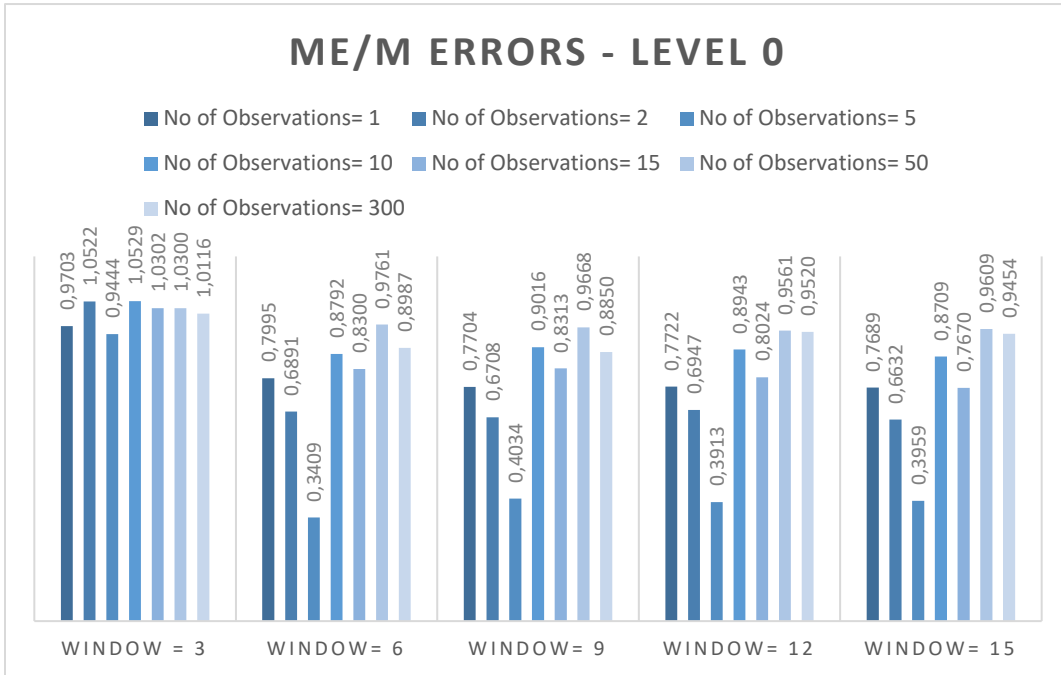


Σχήμα 0.5 MAE/M Errors – Level 1 (MLP)

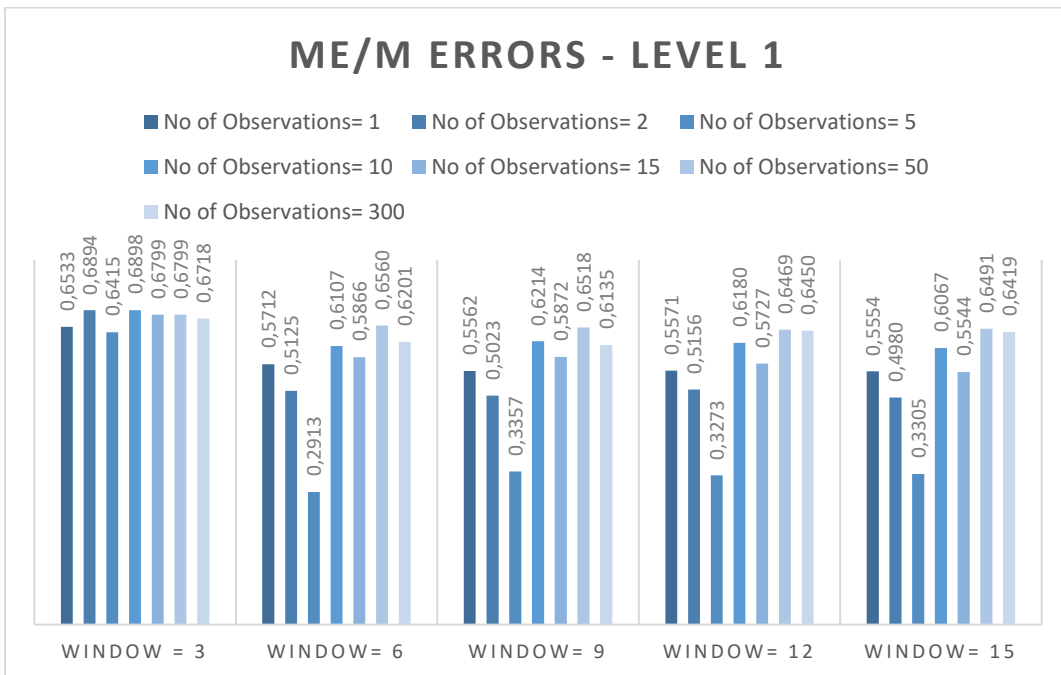


Σχήμα 0.6 MAE/M Errors – Level 2 (MLP)

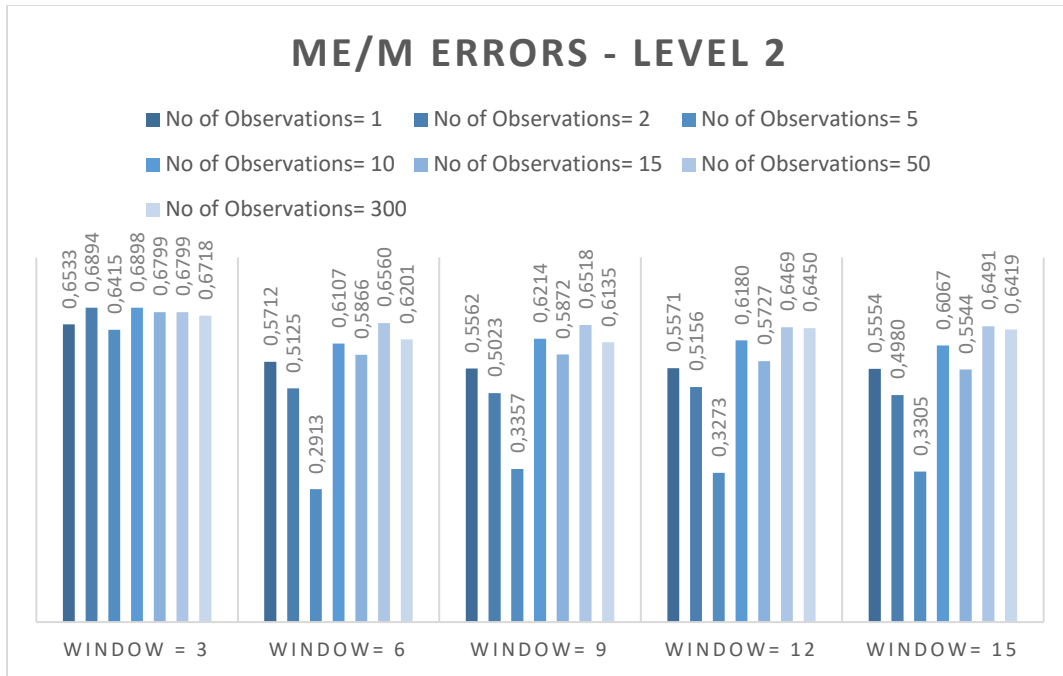
ME/M



Σχήμα 0.7 ME/M Errors – Level 0 (MLP)



Σχήμα 0.8 ME/M Errors – Level 1 (MLP)



Σχήμα 0.9 ME/M Errors – Level 2 (MLP)

Παραλλαγή Μεθόδου με χρήση Index

Ενδεικτικοί Πίνακες Σφαλμάτων για Συγκεκριμένες Παραμέτρους

Μήκος Διανύσματος Εκπαίδευσης (Window) = 12

Γραμμές από κάθε πίνακα X_i (No of Observations) = 10

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,5913	0,3230	2,411	1,1084
(MAE/M)	0,7689	0,5554	0,6013	0,6419
(ME/M)	0,7689	0,5554	0,5554	0,6266

Πίνακας 0.4 Πίνακας Σφαλμάτων (Index) W=12, NoO = 10

Μήκος Διανύσματος Εκπαίδευσης (Window) = 6

Γραμμές από κάθε πίνακα X_i (No of Observations) = 5

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,0946	0,0833	4,7587	1,6455
(MAE/M)	0,3076	0,2665	0,8126	0,4623
(ME/M)	0,3076	0,2665	0,2665	0,2802

Πίνακας 0.5 Πίνακας Σφαλμάτων (Index) W=6, NoO = 5

Παράρτημα

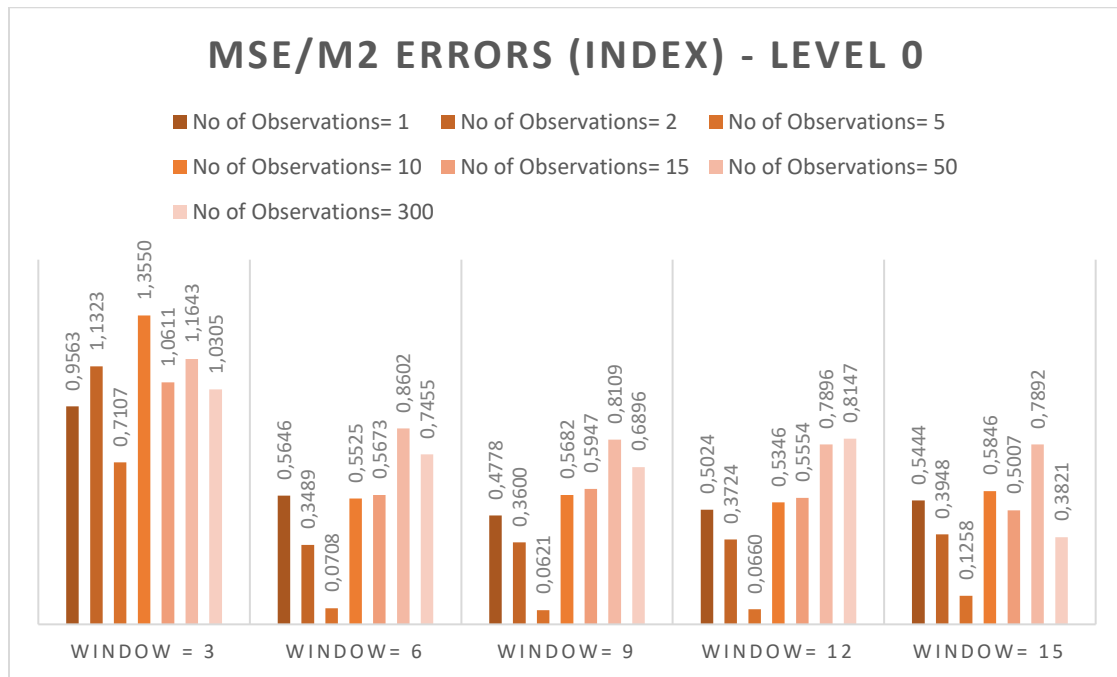
Μήκος Διανύσματος Εκπαίδευσης (Window) = 9
Γραμμές από κάθε πίνακα X_i (No of Observations) = 5

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,0917	0,1452	3,1604	1,1324
(MAE/M)	0,3029	0,2756	0,7013	0,4266
(ME/M)	0,3029	0,2630	0,2630	0,2763

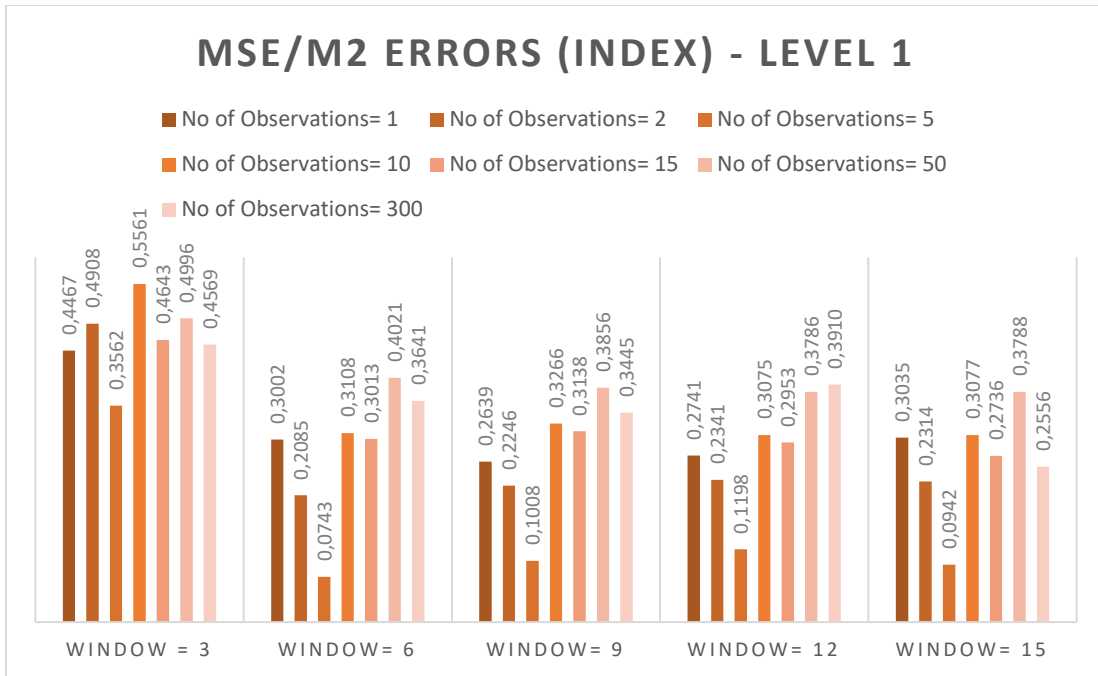
Πίνακας 0.6 Πίνακας Σφαλμάτων (Index) W=9, NoO = 5

Μεταβολή Σφαλμάτων με μεταβολή των παραμέτρων της μεθόδου

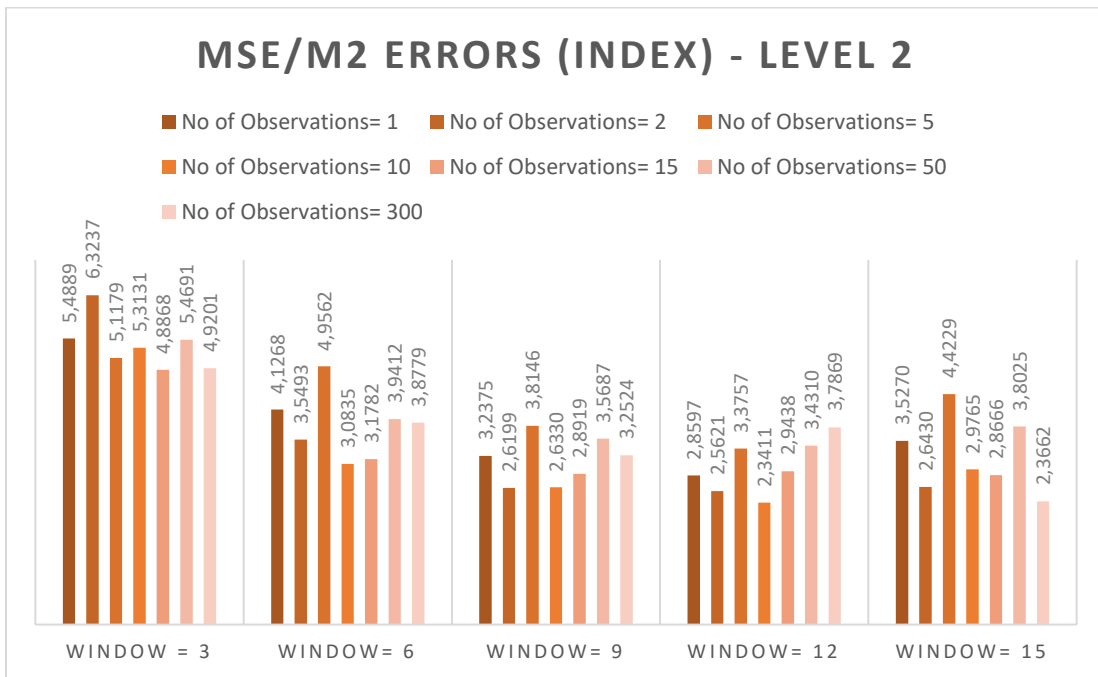
MSE/M²



Σχήμα 0.10 MSE/M² Errors – Level 0 (Index)

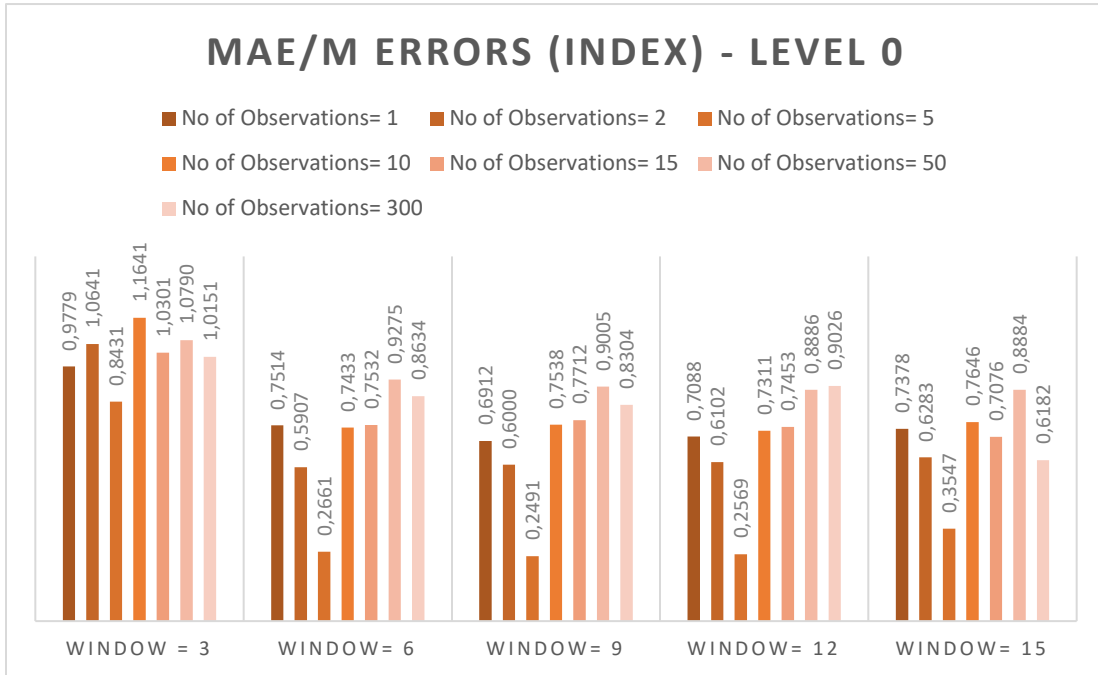


Σχήμα 0.11 MSE/M² Errors - Level 1(Index)

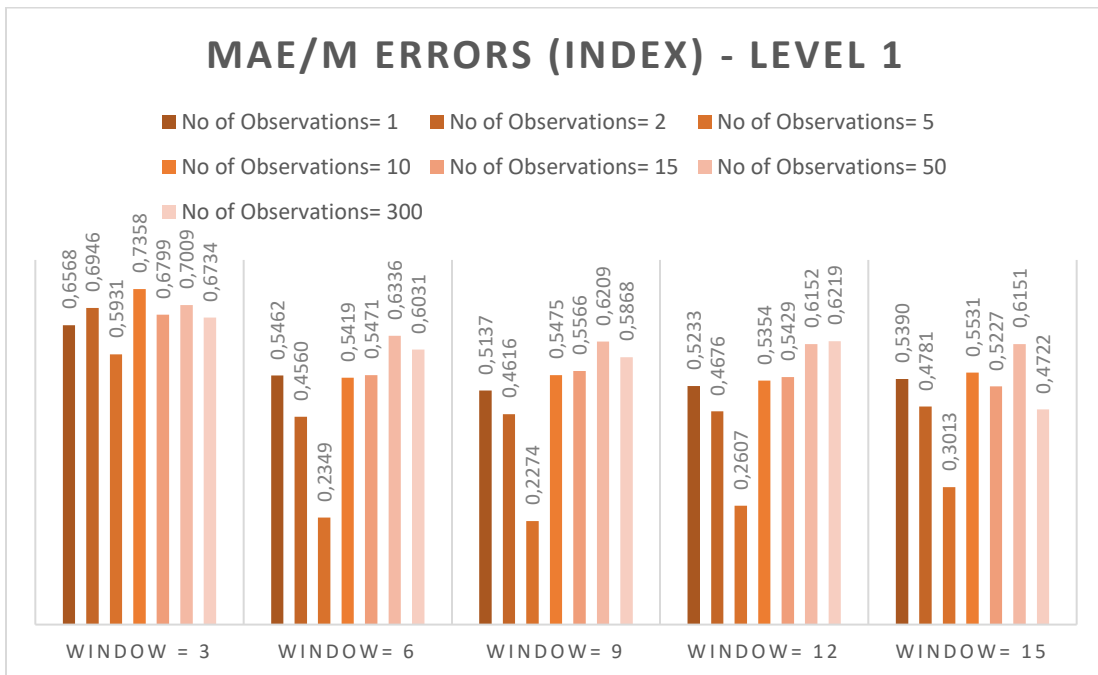


Σχήμα 0.12 MSE/M² Errors - Level 2(Index)

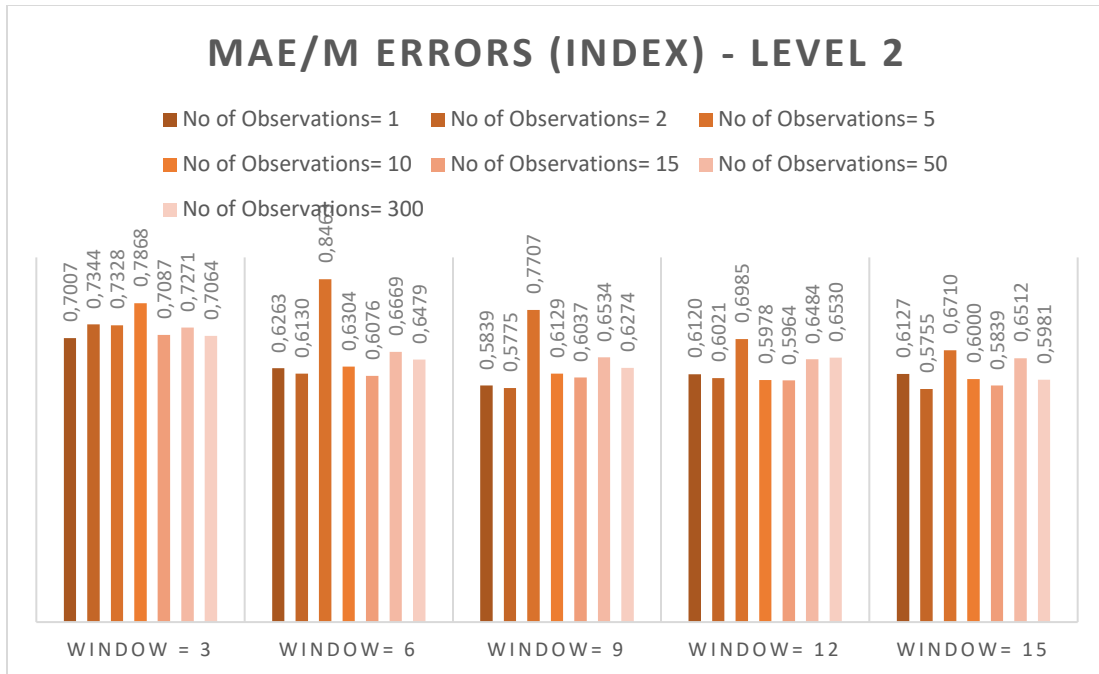
MAE/M



Σχήμα 0.13 MAE/M Errors - Level 0(Index)

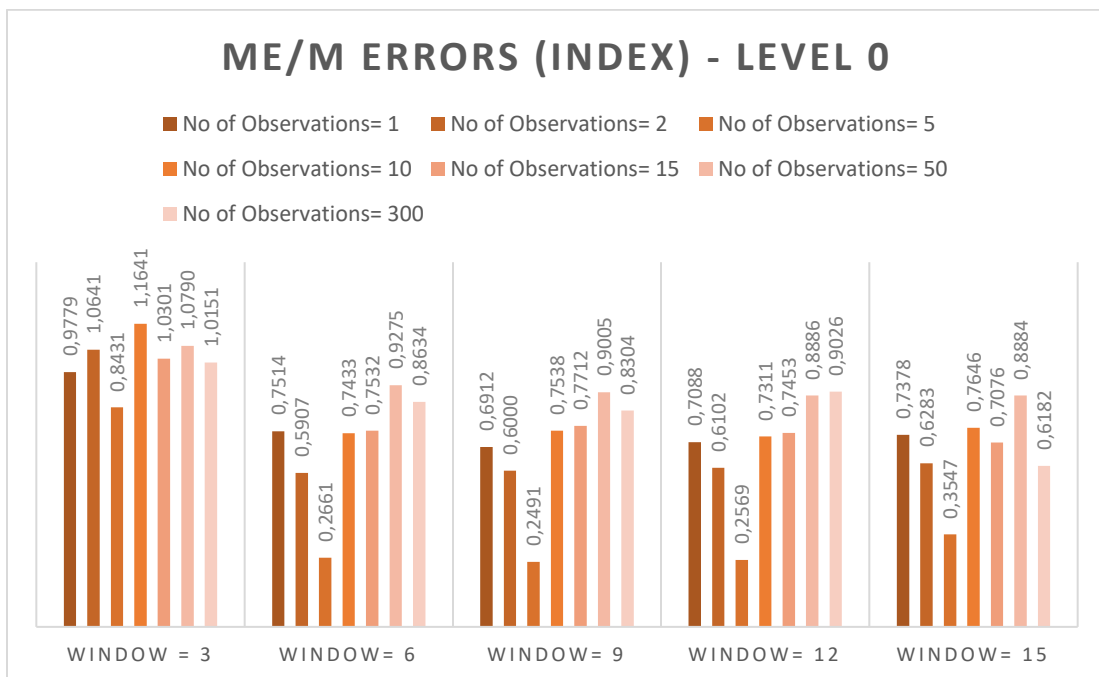


Σχήμα 0.14 MAE/M Errors - Level 1(Index)

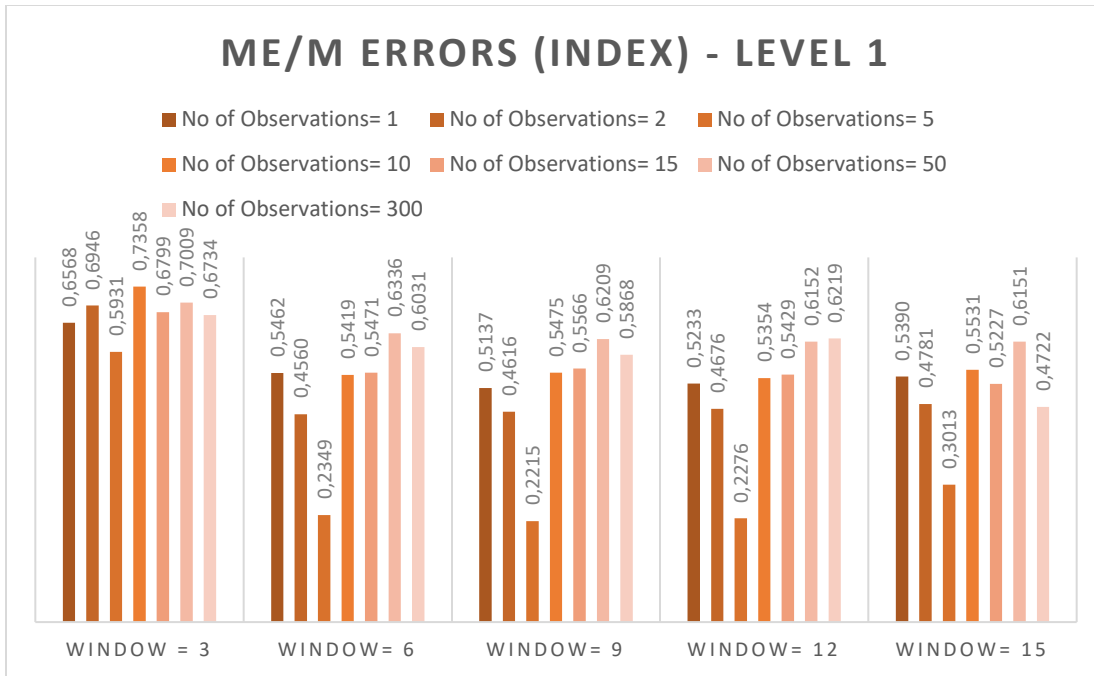


Σχήμα 0.15 MAE/M Errors - Level 2(Index)

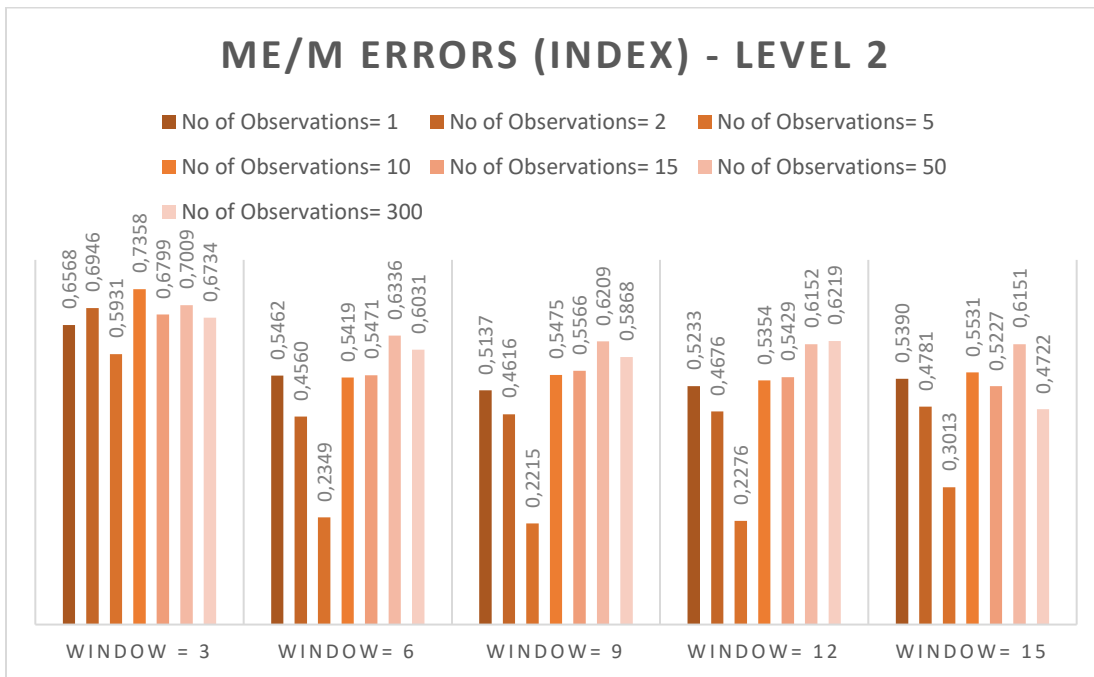
ME/M



Σχήμα 0.16 ME/M Errors - Level 0(Index)



Σχήμα 0.17 ME/M Errors - Level 1(Index)



Σχήμα 0.18 ME/M Errors - Level 2(Index)

Παραλλαγή Μεθόδου με χρήση ADI-CV²

Ενδεικτικοί Πίνακες Σφαλμάτων για Συγκεκριμένες Παραμέτρους

Μήκος Διανύσματος Εκπαίδευσης (Window) = 6

Γραμμές από κάθε πίνακα X_i (No of Observations) = 5

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,0244	0,0289	4,0759	1,3764
(MAE/M)	0,1563	0,1450	0,7220	0,3411
(ME/M)	0,1563	0,1450	0,1450	0,1488

Πίνακας 0.7 Πίνακας Σφαλμάτων (ADI-CV²) W=6, NoO = 5

Μήκος Διανύσματος Εκπαίδευσης (Window) = 12

Γραμμές από κάθε πίνακα X_i (No of Observations) = 2

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,3659	0,2343	3,6870	1,4291
(MAE/M)	0,6049	0,4644	0,5886	0,5527
(ME/M)	0,6049	0,4644	0,4644	0,5113

Πίνακας 0.8 Πίνακας Σφαλμάτων (ADI-CV²) W=12, NoO= 2

Μήκος Διανύσματος Εκπαίδευσης (Window) = 15

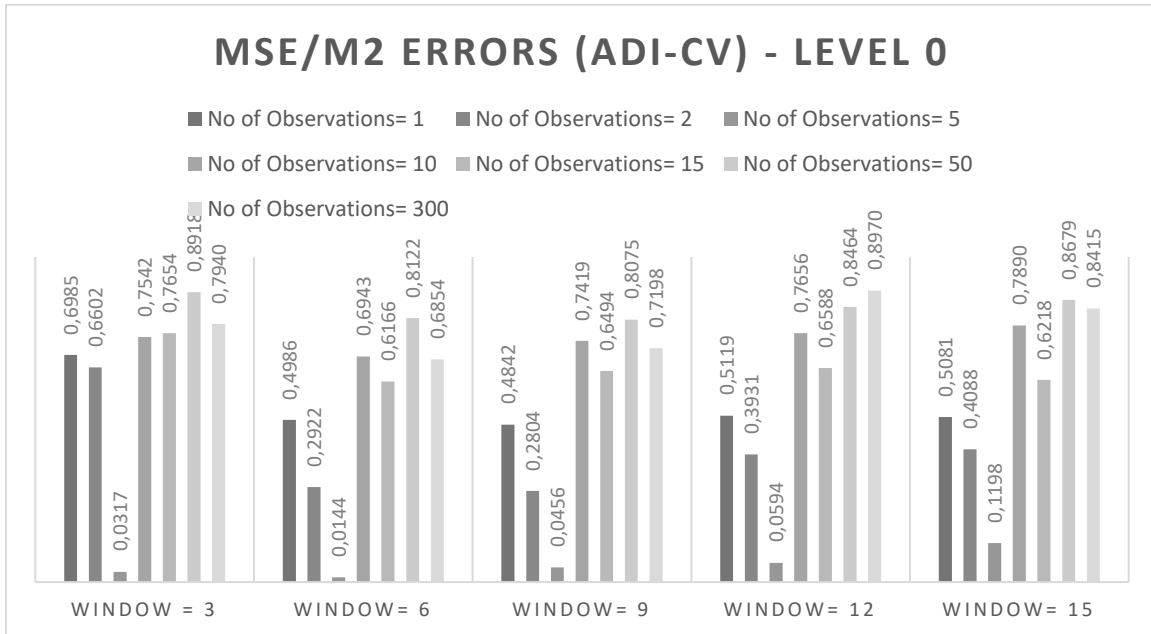
Γραμμές από κάθε πίνακα X_i (No of Observations) = 5

	Level 0	Level 1	Level 2	Average
(MSE/M ²)	0,1256	0,1365	8,8088	3,0237
(MAE/M)	0,3544	0,3011	0,7898	0,4818
(ME/M)	0,3544	0,3011	0,3011	0,3189

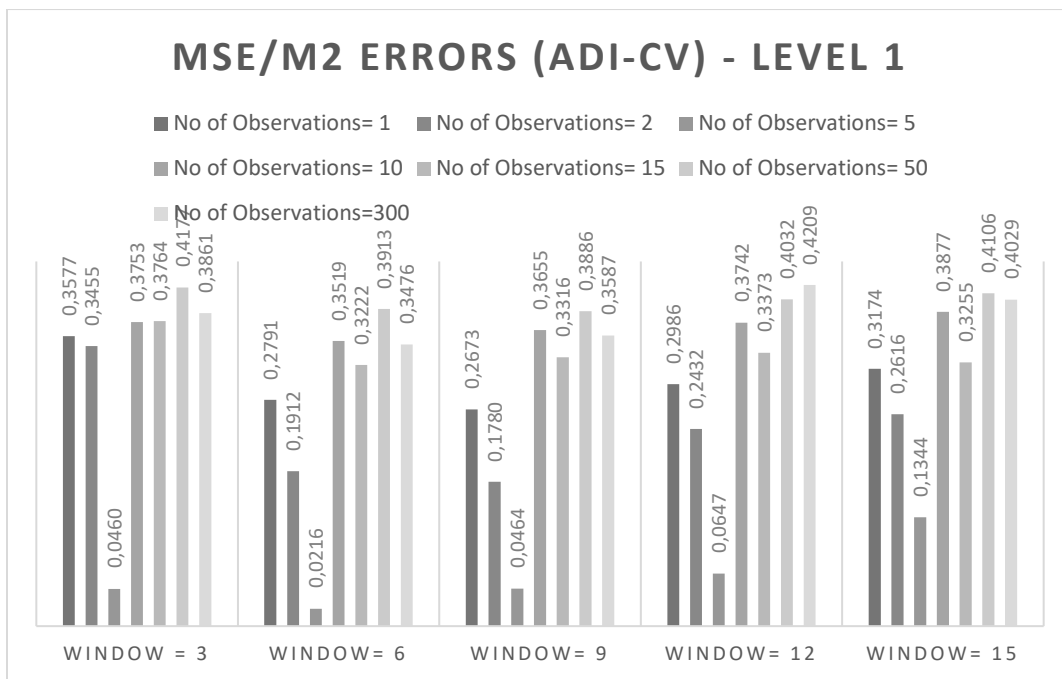
Πίνακας 0.9 Πίνακας Σφαλμάτων (ADI-CV²) W=15, NoO= 5

Μεταβολή Σφαλμάτων με μεταβολή των παραμέτρων της μεθόδου

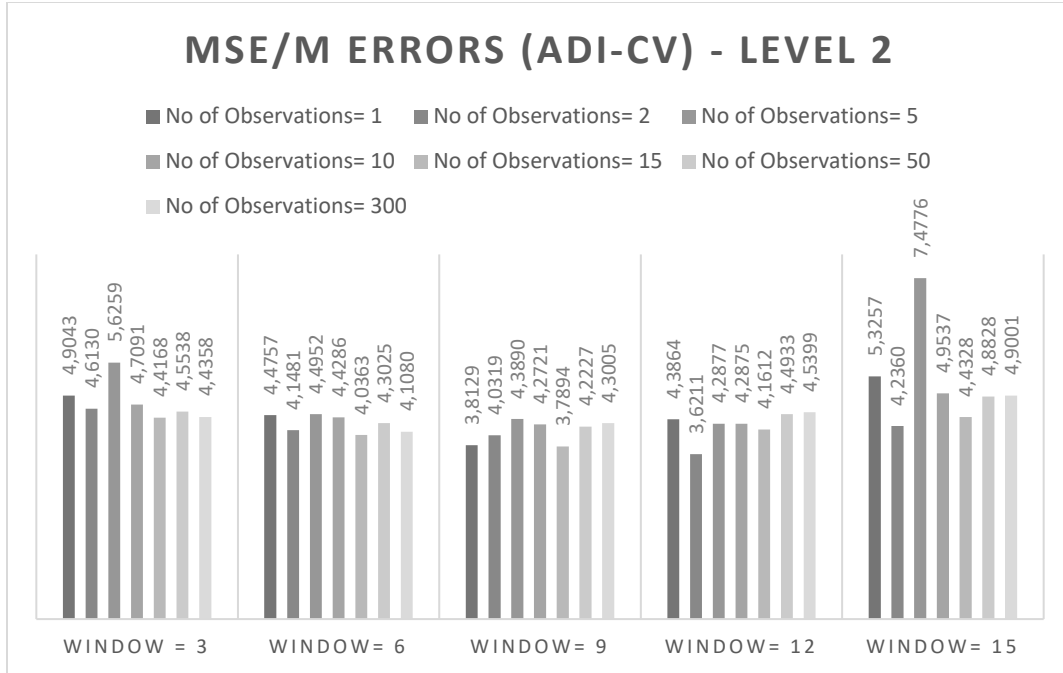
MSE/M²



Σχήμα 0.19 MSE/M² Errors - Level 0(ADI-CV²)

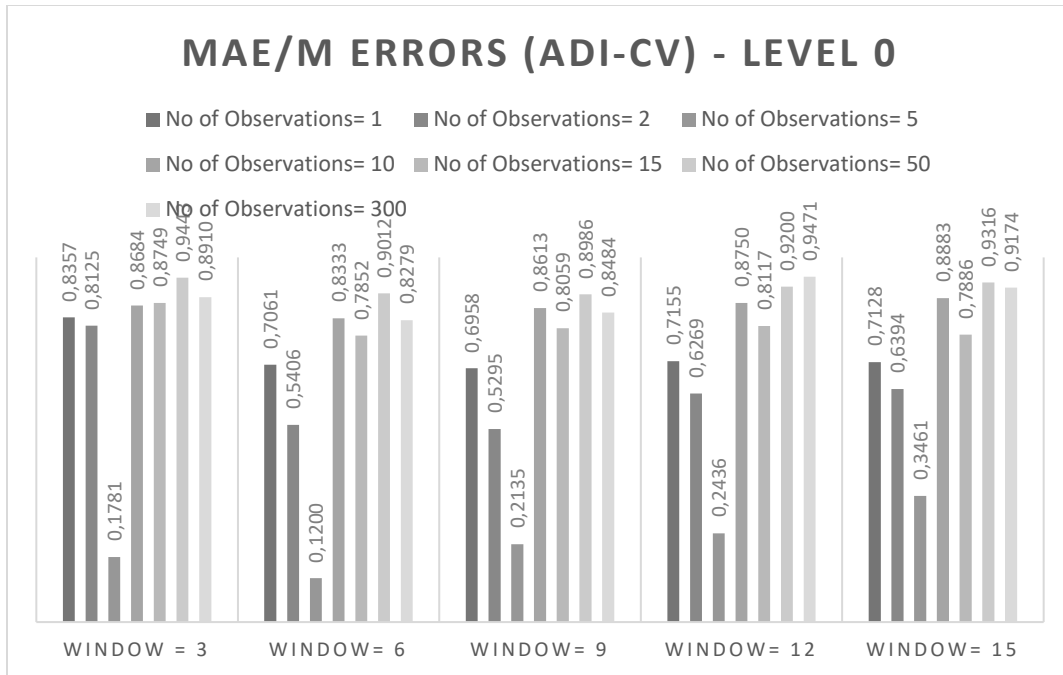


Σχήμα 0.20 MSE/M² Errors - Level 1(ADI-CV²)

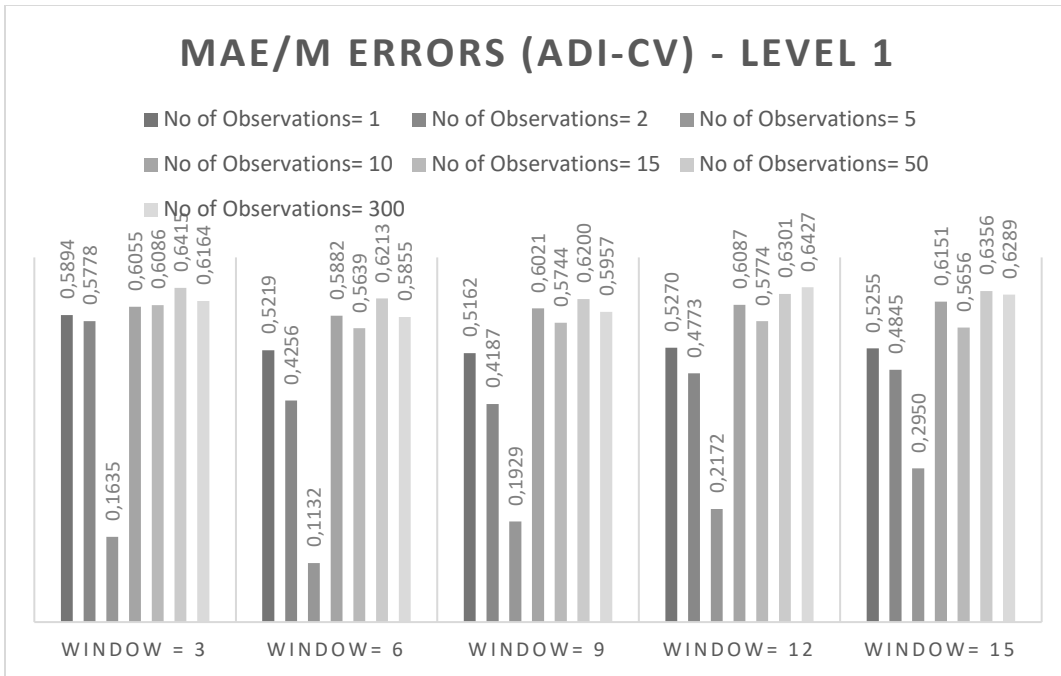


Σχήμα 0.21 MSE/M² Errors - Level 2(ADI-CV²)

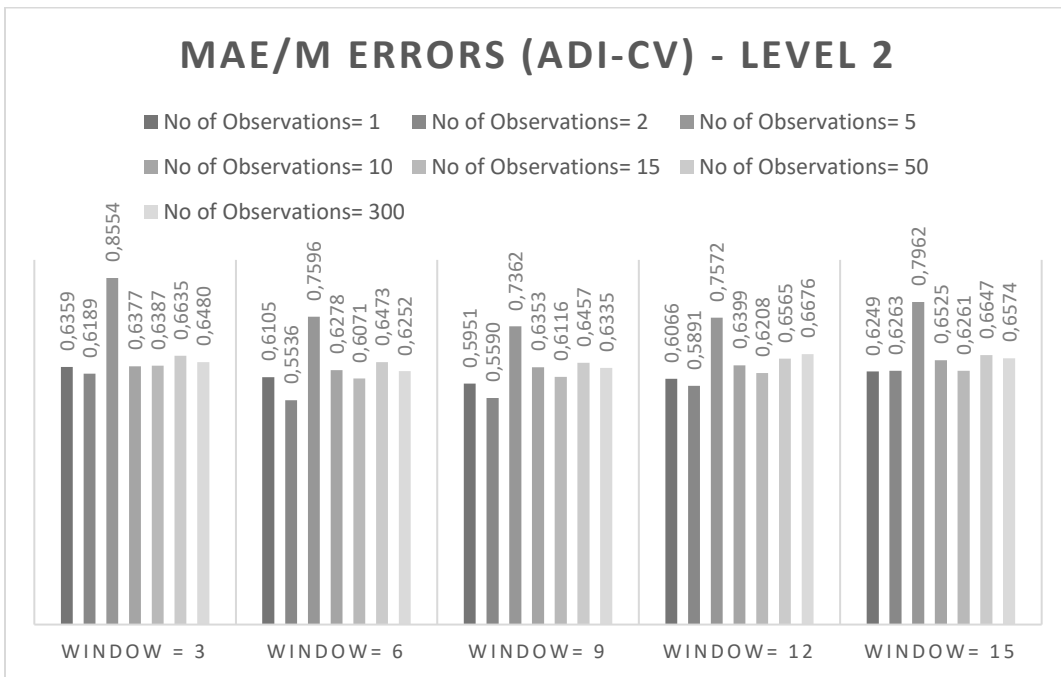
MAE/M



Σχήμα 0.22 MAE/M Errors - Level 0(ADI-CV²)

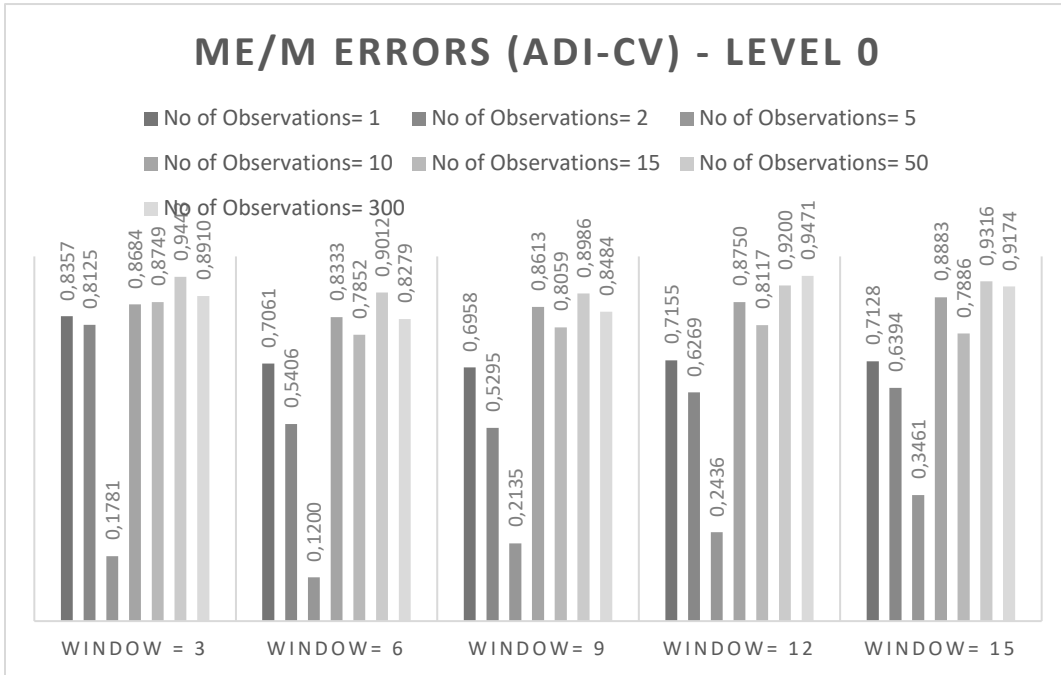


Σχήμα 0.23 MAE/M Errors - Level 1(ADI-CV²)

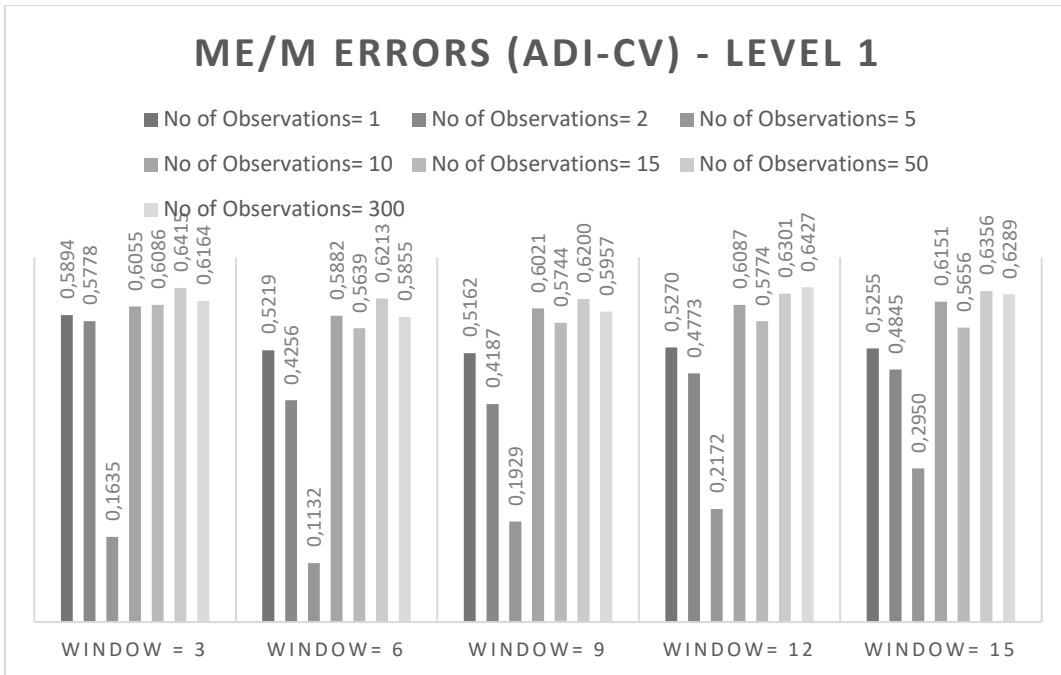


Σχήμα 0.24 MAE/M Errors - Level 2(ADI-CV²)

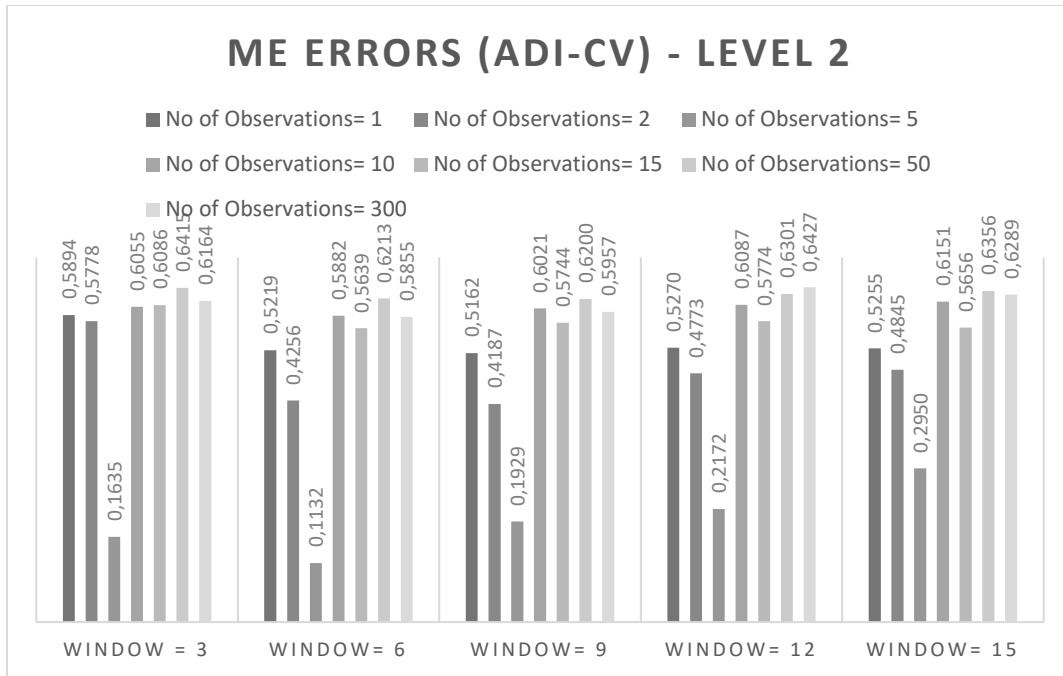
ME/M



Σχήμα 0.25 ME/M Errors - Level 0(ADI-CV²)



Σχήμα 0.26 ME/M Errors - Level 1(ADI-CV²)



Σχήμα 0.27 ME/M Errors - Level 2(ADI-CV²)

Μέρος 2. Παρουσίαση Κώδικα

Για την εκτέλεση των πειραμάτων και τον υπολογισμό των αποτελεσμάτων αναπτύχθηκαν οι αντίστοιχοι αλγόριθμοι στη γλώσσα προγραμματισμού: Python V.3.7.1. Για την πληρότητα της εργασίας, παρατίθενται οι σημαντικές αλγοριθμικές διαδικασίες.

Π.2.1.Για την μέθοδο SES:

```
def SES(m, a):  
    e = 0  
    F = 0  
    S = Mean_Value(m)  
  
    for i in range(len(m)):  
        F = S  
        e = m[i] - F  
        S = S + a*e  
  
    F = S  
    return F
```

Η ρουτίνα για τη μέθοδο SES λαμβάνει ως εισόδους δύο μεταβλητές: ένα διάνυσμα m μήκους t που πρόκειται για την χρονοσειρά στην οποία θέλουμε να εφαρμόσουμε τη μέθοδο, και μία μεταβλητή float που πρόκειται για το συντελεστή εξομάλυνσης της μεθόδου. Επιστρέφει μια μεταβλητή float που πρόκειται για την πρόβλεψη της μεθόδου για τη χρονική περίοδο $t+1$.

Π.2.2.Για τη μέθοδο Croston:

```
def Croston(timeseries):  
  
    d = 1  
    demands = []  
    intervals = []  
  
    for i in range(len(timeseries)):  
        if timeseries[i] == 0:  
            d = d+1  
        else:  
            intervals.append(d)  
            demands.append(timeseries[i])  
            d = 1  
  
    demands_SES = SES(demands, 0.1)
```

Παράρτημα

```
intervals_SES = SES(intervals,0.1)
C = demands_SES / intervals_SES

return C
```

Η ρουτίνα για τη μέθοδο Croston λαμβάνει ως εισόδους μία μεταβλητή: ένα διάνυσμα (timeseries) μήκους τα το οποίο πρόκειται για την χρονοσειρά στην οποία θέλουμε να εφαρμόσουμε τη μέθοδο. Εντός της ρουτίνας καλείται η SES για τον υπολογισμό πρόβλεψης για τα διανύσματα demands και intervals με συντελεστή εξομάλυνσης 0.1. Επιστρέφει μια μεταβλητή float που πρόκειται για την πρόβλεψη της μεθόδου για τη χρονική περίοδο t+1.

Π.2.3.Για την μέθοδο SBA:

```
def SBA(C_forecast, SBAa):
    F = (1 - SBAa)*C_forecast

    return F
```

Η ρουτίνα για τη μέθοδο SBA λαμβάνει ως εισόδους δύο μεταβλητές: μία μεταβλητή τύπου float που πρόκειται για την πρόβλεψη της χρονοσειράς στην οποία θέλουμε να εφαρμόσουμε τη μέθοδο τη χρονική περίοδο t+1, και μία μεταβλητή float που πρόκειται για το συντελεστή της μεθόδου. Επιστρέφει μια μεταβλητή float που πρόκειται για την πρόβλεψη της μεθόδου για τη χρονική περίοδο t+1.

Π.2.4.Για την μέθοδο TSB:

```
def TSB(m, a, b):

    pdot = 1
    zdot = 0
    Ydot = 0

    for i in range(len(m)):
        if m[i] == 0:
            pdot = (1-b) * pdot
        else:
            pdot = b + (1-b)*pdot
            zdot = zdot + a*(m[i]-zdot)
            Ydot = pdot*zdot

    F = Ydot
    return F
```


Παράρτημα

Η ρουτίνα για τη μέθοδο TSB λαμβάνει ως εισόδους τρεις μεταβλητές: ένα διάνυσμα m μήκους t το οποίο πρόκειται για την χρονοσειρά στην οποία θέλουμε να εφαρμόσουμε τη μέθοδο, μία μεταβλητή α τύπου float που πρόκειται για την μία από τις δύο παραμέτρους της μεθόδου, και μία μεταβλητή β τύπου float που πρόκειται για τη δεύτερη παράμετρο της μεθόδου. Επιστρέφει μια μεταβλητή float που πρόκειται για την πρόβλεψη της μεθόδου για τη χρονική περίοδο $t+1$.

Π.2.5.Για την μέθοδο optimal reconciliation:

```
def Hyndman_Method(forecast):

    S1 = np.ones((1, 209))
    ones_array1 = np.ones((1,104))
    ones_array2 = np.ones((1,105))
    zeros_array1 = np.zeros((1,105))
    zeros_array2 = np.zeros((1,104))
    S2 = np.concatenate((ones_array1, zeros_array1), axis = 1)
    S3 = np.concatenate((zeros_array2, ones_array2), axis = 1)
    I209 = np.identity(209)
    S = np.concatenate((S1, S2, S3, I209))                #Summing matrix S

    veta = []                #  $\beta = S(S'S)^{-1}S'$ 
    Strans = np.transpose(S)                #  $S'$ 

    A = np.dot(Strans,S)                #  $A = S'S$ 
    Ainv = np.linalg.inv(A)                #  $A^{-1} = (S'S)^{-1}$ 
    B = np.dot(S,Ainv)                #  $B = S(S'S)^{-1}$ 
    veta = np.dot(B,Strans)
    Opt_Forecast = np.dot(veta,forecast)
    Opt_Forecast.reshape(1,212)
    Opt_Forecast_list = []
    for i in range(len(Opt_Forecast)):
        Opt_Forecast_list.append(Opt_Forecast[i])

    return Opt_Forecast_list
```

Η ρουτίνα για την μέθοδο της optimal reconciliation λαμβάνει μία μεταβλητή ως είσοδο: ένα διάνυσμα(forecast) το οποίο πρόκειται για το διάνυσμα με όλες τις προβλέψεις της μεθόδου που έχουμε χρησιμοποιήσει για να λάβουμε τις προβλέψεις βάσης (base forecasts) για τη χρονική περίοδο $t+1$ ιεραρχικά τοποθετημένες, με την πρόβλεψη του επιπέδου 0 να βρίσκεται στη πρώτη θέση του διανύσματος, τις προβλέψεις του επιπέδου 1 να ακολουθούν και τα υπόλοιπα στοιχεία του διανύσματος να αποτελούν τις προβλέψεις για κάθε επιμέρους χρονοσειρά ξεχωριστά. Η ρουτίνα δημιουργεί τον πίνακα άθροισης S της μεθόδου και ύστερα με κατάλληλες αλγεβρικές πράξεις πινάκων με τη βοήθεια της βιβλιοθήκης numpy, παράγει ένα διάνυσμα που αποτελεί τις ανανεωμένες προβλέψεις της μεθόδου. Οι

Παράρτημα

προβλέψεις βρίσκονται στοιχισμένες κατά τον ίδιο τρόπο που στοιχίσαμε το διάνυσμα εισόδου. Από τη ρουτίνα λαμβάνουμε ως έξοδο το διάνυσμα με τις ανανεωμένες τιμές των προβλέψεών μας.

Π.2.6.Για την εύρεση του MSE/M²:

```
def MSE(forecast, actual):
    F = 1.0
    if type(forecast) == list:
        n = len(forecast)
        m = 0
        kappa = Mean_Value(forecast)
        for i in range(n):
            m = m + (forecast[i] - actual[i])**2
        mse = m / n
        mean_squared = kappa**2
        F = mse / mean_squared

    elif type(forecast) != list:
        sq = (forecast + actual) / 2
        mse = (forecast - actual)**2
        F = mse / sq**2

    return F
```

Η ρουτίνα για τον υπολογισμό του σφάλματος MSE/M² λαμβάνει ως είσοδο 2 μεταβλητές: ένα διάνυσμα(forecast) που πρόκειται για το διάνυσμα με τις προβλέψεις για τη χρονική περίοδο t+1 που παρήχθησαν από τη μέθοδο την οποία αξιολογούμε, ιεραρχικά τοποθετημένες, με την πρόβλεψη του επιπέδου 0 να βρίσκεται στη πρώτη θέση του διανύσματος, τις προβλέψεις του επιπέδου 1 να ακολουθούν και τα υπόλοιπα στοιχεία του διανύσματος να αποτελούν τις προβλέψεις για κάθε επιμέρους χρονοσειρά ξεχωριστά και ένα διάνυσμα(actual) στο οποίο είναι τοποθετημένες οι πραγματικές τιμές των χρονοσειρών τη χρονική περίοδο t+1 τοποθετημένες ιεραρχικά με τρόπο ανάλογο της δομής του διανύσματος forecast. Η ρουτίνα υπολογίζει το μέσο τετραγωνικό σφάλμα των προβλέψεων και εν συνεχεία το διαιρεί με το τετράγωνο της μέσης τιμής των προβλέψεων. Επιστρέφει ως έξοδο μία μεταβλητή float που αποτελεί το εν λόγω σφάλμα της μεθόδου που εξετάζουμε.

Π.2.7.Για την εύρεση του MAE/M:

```
def MAE(forecast, actual):
    F = 1.0
    if type(forecast) == list:
        n = len(forecast)
        diff = 0
        kappa = Mean_Value(forecast)
        for i in range(n):
            diff = diff + abs(forecast[i] - actual[i])
        mae = diff / n
        F = mae / kappa

    elif type(forecast) != list:
        sq = (forecast + actual)/2
        diff = abs(forecast - actual)
        F = diff / sq
    return F
```

Η ρουτίνα για τον υπολογισμό του σφάλματος MAE/M λαμβάνει ως είσοδο 2 μεταβλητές: ένα διάνυσμα(forecast) που πρόκειται για το διάνυσμα με τις προβλέψεις για τη χρονική περίοδο t+1 που παρήχθησαν από τη μέθοδο την οποία αξιολογούμε, ιεραρχικά τοποθετημένες, με την πρόβλεψη του επιπέδου 0 να βρίσκεται στη πρώτη θέση του διανύσματος, τις προβλέψεις του επιπέδου 1 να ακολουθούν και τα υπόλοιπα στοιχεία του διανύσματος να αποτελούν τις προβλέψεις για κάθε επιμέρους χρονοσειρά ξεχωριστά και ένα διάνυσμα(actual) στο οποίο είναι τοποθετημένες οι πραγματικές τιμές των χρονοσειρών τη χρονική περίοδο t+1 τοποθετημένες ιεραρχικά με τρόπο ανάλογο της δομής του διανύσματος forecast. Η ρουτίνα υπολογίζει το μέσο απόλυτο σφάλμα των προβλέψεων και εν συνεχεία το διαιρεί με τη μέση τιμή των προβλέψεων. Επιστρέφει ως έξοδο μία μεταβλητή float που αποτελεί το εν λόγω σφάλμα της μεθόδου που εξετάζουμε.

Π.2.8.Για την εύρεση του ME/M:

```
def ME(forecast, actual):
    F = 1.0
    if type(forecast) == list:
        n = len(forecast)
        diff = 0
        kappa = Mean_Value(forecast)
        for i in range(n):
            diff = diff + forecast[i] - actual[i]
        diff = abs(diff)
        me = diff / n
```

Παράρτημα

$$F = me / kappa$$

```
elif type(forecast) != list:
    sq = (forecast + actual)/2
    diff = abs(forecast - actual)
    F = diff / sq

return F
```

Η ρουτίνα για τον υπολογισμό του σφάλματος ME/M λαμβάνει ως είσοδο 2 μεταβλητές: ένα διάνυσμα(forecast) που πρόκειται για το διάνυσμα με τις προβλέψεις για τη χρονική περίοδο t+1 και ένα διάνυσμα(actual) στο οποίο είναι τοποθετημένες οι πραγματικές τιμές των χρονοσειρών τη χρονική περίοδο t+1 τοποθετημένες με τρόπο ανάλογο της δομής του διανύσματος forecast. Η ρουτίνα υπολογίζει το μέσο σφάλμα των προβλέψεων και εν συνεχεία το διαιρεί με τη μέση τιμή των προβλέψεων. Επιστρέφει ως έξοδο μία μεταβλητή float που αποτελεί το εν λόγω σφάλμα της μεθόδου που εξετάζουμε.

Π.2.9.Για την απεικόνιση του πίνακα σφαλμάτων:

```
def Metrics_Matrix(f_vector, act_val_vector):

    Matrix = array([[0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0]])
    Matrix = reshape(Matrix, (3,4))

    Matrix[0][0] = MSE(f_vector[0], act_val_vector[0])
    Matrix[1][0] = MAE(f_vector[0], act_val_vector[0])
    Matrix[2][0] = ME(f_vector[0], act_val_vector[0])

    Matrix[0][1] = MSE(f_vector[1:3], act_val_vector[1:3])
    Matrix[1][1] = MAE(f_vector[1:3], act_val_vector[1:3])
    Matrix[2][1] = ME(f_vector[1:3], act_val_vector[1:3])

    Matrix[0][2] = MSE(f_vector[3:], act_val_vector[3:])
    Matrix[1][2] = MAE(f_vector[3:], act_val_vector[3:])
    Matrix[2][2] = ME(f_vector[3:], act_val_vector[3:])

    Matrix[0][3] = (Matrix[0][0] + Matrix[0][1] + Matrix[0][2]) / 3
    Matrix[1][3] = (Matrix[1][0] + Matrix[1][1] + Matrix[1][2]) / 3
    Matrix[2][3] = (Matrix[2][0] + Matrix[2][1] + Matrix[2][2]) / 3

    return Matrix
```

Η ρουτίνα για την απεικόνιση του πίνακα σφαλμάτων, όπως αυτός περιεγράφηκε προηγουμένως, λαμβάνει ως εισόδους δύο μεταβλητές: ένα διάνυσμα (f_vector) που πρόκειται για το διάνυσμα με τις προβλέψεις για τη χρονική περίοδο

Παράρτημα

t+1 που παρήχθησαν από τη μέθοδο την οποία αξιολογούμε, ιεραρχικά τοποθετημένες, με την πρόβλεψη του επιπέδου 0 να βρίσκεται στη πρώτη θέση του διανύσματος, τις προβλέψεις του επιπέδου 1 να ακολουθούν και τα υπόλοιπα στοιχεία του διανύσματος να αποτελούν τις προβλέψεις για κάθε επιμέρους χρονοσειρά ξεχωριστά και ένα διάνυσμα(`act_val_vector`) στο οποίο είναι τοποθετημένες οι πραγματικές τιμές των χρονοσειρών τη χρονική περίοδο t+1 τοποθετημένες ιεραρχικά με τρόπο ανάλογο της δομής του διανύσματος `f_vector`. Η ρουτίνα διασπά τα δύο διανύσματα στα επίπεδα της ιεραρχίας και καλεί τις ρουτίνες για τον υπολογισμό σφαλμάτων σε κάθε επίπεδο. Τα αποτελέσματα τοποθετούνται σε έναν πίνακα διαστάσεων 3 επί 4 που δομείται όπως περιεγράφηκε παραπάνω. Η ρουτίνα επιστρέφει τον πίνακα αυτόν.

Π.2.10. Για τον υπολογισμό των Μεγεθών-Δεικτών ADI και CV²

```
def ADI_Calculator(vector):  
  
    index = 1  
    idi = []  
  
    for i in range(len(vector)):  
        if vector[i] == 0:  
            index = index + 1  
        else:  
            idi.append(index)  
            index = 1  
  
    if index != 1:  
        idi.append(index)  
  
    adi = Mean_Value(idi)  
  
    return adi  
  
def CVsquared_Calculator(vector):  
  
    m = Mean_Value(vector)  
    wantedsum = 0  
  
    for i in range(len(vector)):  
        wantedsum = wantedsum + (vector[i] - m)**2  
  
    sd = wantedsum / (len(vector) - 1)  
    m = m**2  
    cv = sd / m  
  
    return cv
```

Παράρτημα

Οι δύο ρουτίνες αυτές λαμβάνουν ως είσοδο τη χρονοσειρά για την οποία επιθυμούμε να υπολογίσουμε τους δύο δείκτες. Ως έξοδο λαμβάνουμε τους δείκτες.

Π.2.11.Για την κατασκευή των πινάκων εκπαίδευσης (X_i)

```
def X_Matrix(data_vector, training_set_index):
```

```
    n = len(data_vector)
    d = n - training_set_index
    tr = np.empty((d, training_set_index))
```

```
    for i in range(d):
        for j in range(training_set_index):
            tr[i][j] = data_vector[i+j]
```

```
    return tr
```

Η ρουτίνα για την κατασκευή του πίνακα εκπαίδευσης που αντιστοιχεί σε κάθε χρονοσειρά λαμβάνει ως εισόδους δύο μεταβλητές. Η πρώτη (`data_vector`) αποτελεί τη χρονοσειρά για την οποία θέλουμε να κατασκευάσουμε τον πίνακα, ενώ η δεύτερη αποτελεί το μήκος διανυσμάτων εκπαίδευσης που έχουμε επιλέξει για την μέθοδο την οποία πραγματοποιούμε. Η έξοδος της ρουτίνας αποτελεί τον πίνακα εκπαίδευσης X_i κατασκευασμένο όπως περιγράφεται στο κεφάλαιο 5.

Λόγω της ύπαρξης παραλλαγών της αρχικής μεθόδου, η ρουτίνα μορφοποιείται κατάλληλα ως εξής:

Index Version

```
def X_Matrix(data_vector, training_set_index, set_index):
```

```
    n = len(data_vector)
    d = n - training_set_index
    tr = np.empty((d, training_set_index))
```

```
    for i in range(d):
        for j in range(training_set_index):
            tr[i][j] = data_vector[i+j]
```

```
    indexmatrix = np.zeros((d, 1))
```

```
    for i in range(d):
```

```
        indexmatrix[i] = set_index
```

```
    new_tr = np.concatenate((tr, indexmatrix), axis = 1)
```

```
    return new_tr
```

Παράρτημα

Στην παραλλαγή αυτή της ρουτίνας εισάγουμε μία ακόμα μεταβλητή στις μεταβλητές εισόδου, την `set_index`, η οποία αποτελεί το δείκτη της κάθε χρονοσειράς για την οποία κατασκευάζουμε τον πίνακα X_i και την οποία προσθέτουμε ως επιπλέον στήλη στα διανύσματα από τα οποία αποτελείται ο πίνακας.

ADI-CV² Version

```
def X_Matrix(data_vector, training_set_index):

    n = len(data_vector)
    d = n - training_set_index
    tr = np.empty((d, training_set_index))

    for i in range(d):
        for j in range(training_set_index):
            tr[i][j] = data_vector[i+j]

    adimeasure = ADI_Calculator(data_vector)
    cvmeasure = CVsquared_Calculator(data_vector)

    indexmatrix = np.zeros((d, 2))
    for i in range(d):
        indexmatrix[i][0] = adimeasure
        indexmatrix[i][1] = cvmeasure

    new_tr = np.concatenate((tr, indexmatrix), axis = 1)

    return new_tr
```

Η παραλλαγή αυτή δεν έχει επιπλέον μεταβλητές εισόδου, αλλά υπολογίζουμε εντός της ρουτίνας τα μεγέθη ADI και CV² της χρονοσειράς για την οποία κατασκευάζουμε τον πίνακα X_i και τα τοποθετούμε ως έξτρα στήλες στα διανύσματα εκπαίδευσης από τα οποία αυτός αποτελείται.

Π.2.12. Για την κατασκευή των διανυσμάτων προτύπων (y_i)

```
def y_Matrix(data_vector, training_set_index):

    n = len(data_vector)
    d = n - training_set_index
    tr = []

    for i in range(d):
        tr.append(data_vector[i+training_set_index])

    F = np.array(tr)
```

```
return F
```

Η ρουτίνα αυτή λαμβάνει ως εισόδους δύο μεταβλητές. Η πρώτη (`data_vector`) αποτελεί τη χρονοσειρά για την οποία θέλουμε να κατασκευάσουμε τον πίνακα, ενώ η δεύτερη αποτελεί το μήκος διανυσμάτων εκπαίδευσης που έχουμε επιλέξει για την μέθοδο την οποία πραγματοποιούμε. Η έξοδος της ρουτίνας αποτελεί το διάνυσμα προτύπων y_i κατασκευασμένο όπως περιγράφεται στο κεφάλαιο 5.

Π.2.13. Για την προετοιμασία των διανυσμάτων εισόδου

```
def setforprediction(vector, trindex):  
  
    wanted = np.array(vector[-trindex:])  
    wanted = np.reshape(wanted, (1,trindex))  
  
    return wanted
```

Η ρουτίνα αυτή λαμβάνει ως εισόδους δύο μεταβλητές. Η πρώτη αποτελεί τη χρονοσειρά για την οποία θέλουμε να προβλέψουμε την τιμή που θα λάβει τη χρονική περίοδο $t+1$, ενώ η δεύτερη αποτελεί το μήκος που έχουμε επιλέξει για τα διανύσματα εκπαίδευσης. Η έξοδος είναι το διάνυσμα που θα θέσουμε ως είσοδο στο νευρωνικό ώστε να παράγει πρόβλεψη για τη χρονική περίοδο $t+1$.

Λόγω της ύπαρξης παραλλαγών της αρχικής μεθόδου, η ρουτίνα μορφοποιείται κατάλληλα ως εξής:

Index Version

```
def setforprediction(vector, index, trindex):  
  
    wanted = np.array(vector[-trindex:])  
    wanted = np.append(wanted, index)  
    wanted = np.reshape(wanted, (1,trindex+1))  
  
    return wanted
```

Στην παραλλαγή αυτή της ρουτίνας εισάγουμε μία ακόμα μεταβλητή στις μεταβλητές εισόδου, την `index`, η οποία αποτελεί το δείκτη της κάθε χρονοσειράς για την οποία κατασκευάζουμε το διάνυσμα που θα τεθεί ως είσοδος στο εκπαιδευμένο νευρωνικό.

ADI-CV² Version

```
def setforprediction(vector,index):  
  
    a = ADI_Calculator(vector)
```


Παράρτημα

```
cv = CVsquared_Calculator(vector)

wanted = np.array(vector[-index:])
wanted = np.append(wanted, a)
wanted = np.append(wanted, cv)
wanted = np.reshape(wanted, (1,index+2))

return wanted
```

Στην παραλλαγή αυτή της ρουτίνας δεν εισάγουμε επιπλέον μεταβλητή στις μεταβλητές εισόδου, αλλά υπολογίζουμε εντός της ρουτίνας τους δείκτες ADI και CV^2 της χρονοσειράς για την οποία κατασκευάζουμε το διάνυσμα που θα τεθεί ως είσοδος στο εκπαιδευμένο νευρωνικό.

