



# Εθνικό Μετσόβιο Πολυτεχνείο

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

## Πιθανοτικές Προβλέψεις με Χρήση Νευρωνικών Δικτύων

### Διπλωματική Εργασία

Δημήτριος Δ. Παπαδημητρίου

Επιβλέπων: Βασίλειος Ασημακόπουλος  
Καθηγητής Ε.Μ.Π.

Υπεύθυνοι: Ευάγγελος Σπηλιώτης  
Διδάκτωρ Ε.Μ.Π.

Αρτέμιος Σεμένογλου  
Υποψήφιος Διδάκτωρ Ε.Μ.Π.

ΜΟΝΑΔΑ ΠΡΟΒΛΕΨΕΩΝ ΚΑΙ ΣΤΡΑΤΗΓΙΚΗΣ

Αθήνα, Νοέμβριος 2021





# Εθνικό Μετσόβιο Πολυτεχνείο

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

## Πιθανοτικές Προβλέψεις με Χρήση Νευρωνικών Δικτύων

### Διπλωματική Εργασία

Δημήτριος Δ. Παπαδημητρίου

Επιβλέπων: Βασίλειος Ασημακόπουλος  
Καθηγητής Ε.Μ.Π.

Υπεύθυνοι: Ευάγγελος Σπηλιώτης  
Διδάκτωρ Ε.Μ.Π.

Αρτέμιος Σεμένογλου  
Υποψήφιος Διδάκτωρ Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18η Μαρτίου 2021.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....  
Βασίλειος  
Ασημακόπουλος  
Καθηγητής Ε.Μ.Π.

.....  
Ιωάννης Ψαρράς  
Καθηγητής Ε.Μ.Π.

.....  
Αημήτριος Ασκούνης  
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2021



(Υπογραφή)

.....  
**Δημήτριος Παπαδημητρίου**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © 2021 Δημήτριος Παπαδημητρίου  
,2021. Με 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 τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.



# Ευχαριστίες

Η παρούσα διπλωματική εκπονήθηκε στο πλαίσιο των ερευνητικών δραστηριοτήτων της Μονάδας Προβλέψεων και Στρατηγικής, η οποία υπάγεται στον τομέα Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Καταρχάς θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή κ. Βασίλειο Ασημακόπουλο για την ευκαιρία που μου έδωσε να ασχοληθώ με το ιδιαίτερα ενδιαφέρον θέμα της Μηχανικής Μάθησης στον τομέα των Προβλέψεων. Κυρίως όμως θα ήθελα να του εκφράσω την ευγνωμοσύνη μου για τη δυνατότητα που μου έδωσε να αποτελώ τα τελευταία δύο χρόνια μέλος της Μονάδας και να μετέχω στις εξαιρετικά ενδιαφέρουσες δραστηριότητές της.

Επιπροσθέτως, θα ήθελα να ευχαριστήσω τους Καθηγητές κ.κ. Ιωάννη Ψαρρά και Δημήτριο Ασκούνη για την τιμή που μου έκαναν να συμμετέχουν στην επιτροπή εξέτασης της παρούσας εργασίας.

Ιδιαίτερη μνεία θα ήθελα να κάνω στα υπόλοιπα μέλη της Μονάδας Προβλέψεων και Στρατηγικής. Συγκεκριμένα θα ήθελα να ευχαριστήσω τη διδάκτορα κ. Νικολέττα Ζαμπέτα Λεγάκη, η οποία μέσω της διδασκαλίας του μαθήματος των Τεχνικών Προβλέψεων με έκανε να αγαπήσω τον κλάδο αυτό, θέλοντας να ασχοληθώ περαιτέρω εκτός ακαδημαϊκού πλαισίου. Επιπλέον θα ήθελα να ευχαριστήσω τους Υποψήφιους Διδάκτορες Διαμαντή Κουτσανδρέα, Βαγγέλη Θεοδώρου και Τάσο Καλτσούνη για τις καθημερινές στιγμές που μοιραζόμαστε στο χώρο της Μονάδας. Κυρίως όμως θα ήθελα να ευχαριστήσω το Διδάκτορα κ. Βαγγέλη Σπηλιώτη και τον Υποψήφιο Διδάκτορα κ. Αρτέμιο Σεμένογλου. Πέρα από τις πολύτιμες συμβουλές και υποδείξεις τόσο στο πλαίσιο της παρούσας διπλωματικής εργασίας όσο και των υπολοίπων έργων στα οποία έχω εμπλακεί, αποτελούν για εμένα πρότυπο σκληρής δουλειάς και χαρακτήρα, δίνοντας νέα πνοή στις σπουδές μου.

Τέλος θα ήθελα να ευχαριστήσω τους φίλους μου και ιδιαίτερα τους γονείς μου Δημήτρη και Γιάννα οι οποίοι με τη διακριτική τους παρουσία αποτελούν πάντα στήριγμα.

Δημήτριος Παπαδημητρίου,

Νοέμβριος 2021





# Περίληψη

Η έννοια της πρόβλεψης είναι άρρηκτα συνδεδεμένη με το σύγχρονο επιχειρηματικό και εν γένει επιχειρησιακό περιβάλλον, καθώς η εκτίμηση της μελλοντικής κατάτασης των μεγεθών που επηρεάζουν άμεσα ή έμμεσα τον άνθρωπο, την επιχείρηση ή τον οργανισμό είναι απαραίτητη για την ορθή λήψη αποφάσεων που σχετίζονται με τις ενέργειες του παρόντος. Η πολυπλοκότητα των συστημάτων και η πληθώρα πληροφορίας που καλείται να διαχειριστεί ο εκάστοτε αποφασίζων έχουν οδηγήσει στην ανάγκη αξιοποίησης νέων λύσεων πέρα των αμιγώς στατιστικών, λύσεις οι οποίες λαμβάνονται από άλλους κλάδους όπως επί παραδείγματι τη μηχανική μάθηση.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η εξοικείωση με την εργαλειοθήκη του GluonTS, που επιτρέπει τη χρήση έτοιμων αλλά και την ανάπτυξη νέων μοντέλων μηχανικής και βαθιάς μηχανικής μάθησης, με σκοπό την εξαγωγή πιθανοτικών προβλέψεων. Παράλληλα εξετάζονται παραλλαγές του μοντέλου Απλής Πρόσθιας Τροφοδότησης, με χρήση πέντε διαφορετικών κατανομών με στόχο την εξαγωγή των βέλτιστων πιθανοτικών προβλέψεων σε μηνιαία δεδομένα τα οποία διακρίνονται σε έξι κατηγορίες.

Της παραπάνω μελέτης προηγείται η θεωρητική παρουσίαση και θεμελίωση των εννοιών που άπτονται του κλάδου των Προβλέψεων, καθώς επίσης και του ιδιαίτερα δυναμικού κλάδου της μηχανικής μάθησης. Επιπλέον γίνεται εκτενής περιγραφή της εργαλειοθήκης του GluonTS και των υπερ-παραμέτρων του μοντέλου που χρησιμοποιήθηκε.

Η διπλωματική εργασία περατώνεται με τα συμπεράσματα που εξήχθησαν από την πειραματική διαδικασία, ενώ γίνεται μία προσπάθεια ανάδειξης των εφαρμογών τέτοιου είδους μοντέλων σε πρακτικά προβλήματα.

Λέξεις Κλειδιά

Τεχνικές Προβλέψεων, Μηχανική Μάθηση, Νευρωνικό Δίκτυο, GluonTS, Κατανομές, Πιθανοτικές Προβλέψεις



# Abstract

Forecasting is bonded inseparably to the contemporary corporate and operational environment, due to the fact that the estimation of the future state of the factors which affect every human, business or organisation in a direct or indirect way, is vital for making the best decisions in order to define their present actions. The high complexity as well as the vast amount of the information available to the decider, have created a need to deploy new solutions offered by various fields, such as machine learning.

The aim of this diploma thesis is the familiarization with the GluonTS toolkit, which enables the usage of pre-implemented machine and deep learning models, along with the opportunity to create new ones, in order to produce probabilistic forecasts. Besides, several variations of the SimpleFeedForward are examined, based on five different probability distributions in an effort to define the one with the most accurate probabilistic forecasts.

Before the above, the theoretical background of both fields of Forecasting and Machine Learning, is presented, in addition to the description of the GluonTS toolkit and the hyper-parameters of the model used within the purpose of this thesis.

Ultimately, all conclusions from the experimental process are presented, while basic real life applications of deep learning probabilistic models are featured.

## Keywords

Forecasting Methods, Machine Learning, Neural Network, GluonTs, Distributions, Probabilistic Forecasting



# Περιεχόμενα

Ευχαριστίες.....	vii
Περίληψη.....	ix
Abstract.....	xi
Περιεχόμενα.....	xiii
Κατάλογος Σχημάτων.....	xvii
Κατάλογος Πινάκων.....	xix
<b>1.Εισαγωγή.....</b>	<b>1</b>
1.1 Αντικείμενο της Εργασίας.....	1
1.2 Οργάνωση της Εργασίας.....	2
<b>2. Προβλέψεις και Δείκτες Ακριβείας.....</b>	<b>5</b>
2.1 Χρονοσειρές και Ποιοτικά Χαρακτηριστικά Χρονοσειρών.....	5
2.2 Γραφική Αναπαράσταση Χρονοσειρών.....	6
2.3 Γενικά Περί Προβλέψεων.....	8
2.4 Μοντέλο Χρονοσειρών.....	9
2.4.1 Αιτιοκρατικό Μοντέλο.....	10
2.5 Βασικά Βήματα της Διαδικασίας Πρόβλεψης.....	11
2.6 Πεδία Εφαρμογής των Προβλέψεων.....	12
2.7 Στατιστικές Μέθοδοι Πρόβλεψης.....	12
2.7.1 Αφελής Μέθοδος (Naive) .....	13
2.7.2 Μέθοδος Κινητών Μέσων Όρων (KMO) .....	13
2.7.3 Μέθοδοι Εκθετικής Εξομάλυνσης.....	14
2.7.3.1 Απλή Εκθετική Εξομάλυνση - Μοντέλο Σταθερού Επιπέδου (Simple Exponential Smoothing - SES) .....	15
2.7.3.2 Μοντέλο Γραμμικής Τάσης (Holt Exponential Smoothing) .....	17
2.7.3.3 Μοντέλα Μη γραμμικής Τάσης (Damped Exponential Smoothing) .....	18
2.7.4 Μέθοδοι Παλινδρόμησης.....	19
2.7.4.1 Απλή Γραμμική Παλινδρόμηση.....	19
2.7.4.2 Πολλαπλή Γραμμική Παλινδρόμηση.....	20
2.7.5 Μέθοδος Theta.....	21
2.8 Πιθανοτικές Προβλέψεις.....	23
2.9 Επιπλέον Κατηγορίες Μεθόδων Προβλέψεων.....	26
2.9.1 Μέθοδοι Διακοπτόμενης Ζήτησης.....	26
2.9.2 Κριτικές Προβλέψεις .....	27
2.9.3 Προβλέψεις Προϋπολογισμού.....	28
2.10 Δείκτες Ακριβείας.....	28
2.10.1 Μέσο Σφάλμα (Mean Error - ME) .....	29
2.10.2 Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE) .....	29
2.10.3 Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error - MSE) .....	29

2.10.4 Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error - RMSE) .....	29
2.10.5 Ρίζα Μέσου Τετραγωνικού Κανονικοποιημένου Σφάλματος (Root Mean Squared Scaled Error - RMSSE) .....	30
2.10.6 Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error - MAPE).....	30
2.10.7 Συμμετρικό Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Symmetric Mean Absolute Percentage Error - sMAPE) .....	30
2.10.8 Μέσο Απόλυτο Κανονικοποιημένο Σφάλμα (Mean Absolute Scaled Error - MAsE) .....	31
2.10.9 Pinball Loss (PL) .....	31
2.10.10 Scaled Pinball Loss (SPL).....	32
<b>3. Νευρωνικά Δίκτυα.....</b>	<b>33</b>
3.1 Εισαγωγή.....	33
3.2 Ιδιότητες Νευρωνικών Δικτύων.....	34
3.3 Μοντέλο Νευρώνων.....	36
3.4 Συναρτήσεις Ενεργοποίησης.....	37
3.4.1 Σιγμοειδής Συνάρτηση.....	37
3.4.2 Συνάρτηση Υπερβολικής Εφαπτομένης.....	38
3.4.3 SoftMax Συνάρτηση.....	38
3.4.4 Rectified Linear Unit (ReLU) .....	39
3.5 Αρχιτεκτονικές Νευρωνικών Δικτύων.....	39
3.5.1 Perceptron.....	39
3.5.2 Δίκτυα Πρόσθιας Τροφοδότησης Μοναδικού Επιπέδου.....	42
3.5.3 Δίκτυα Πρόσθιας Τροφοδότησης Πολλαπλών Επιπέδων.....	42
3.5.4 Αναδρομικά Δίκτυα.....	44
3.6 Μηχανική Μάθηση.....	45
3.6.1 Μάθηση με Εκπαιδευτή.....	45
3.6.2 Μάθηση Χωρίς Εκπαιδευτή.....	46
3.6.2.1 Ενισχυτική Μάθηση.....	46
3.6.2.2 Μη Επιβλεπόμενη Μάθηση.....	46
3.6.3 Ο Αλγόριθμος της Οπισθοδιάδοσης (Backpropagation) .....	47
<b>4. Παρουσίαση του GluonTS.....</b>	<b>51</b>
4.1 Εισαγωγή.....	51
4.2 Λειτουργία GluonTS.....	53
4.3 Είσοδος/Εξοδος Δεδομένων και Επεξεργασία.....	53
4.4 Προβλέπτης.....	54
4.5 Κατανομή / Έξοδος.....	54
4.6 Αντικείμενα Πρόβλεψης και Αξιολόγηση.....	54
4.7 Δημιουργία Μοντέλου Χρονοσειρών.....	55
4.8 Μαθηματικό Υπόβαθρο Λειτουργίας του GluonTS.....	56
4.8.1 Generative Μοντέλα.....	58
4.8.2 Discriminative Μοντέλα.....	58
4.8.3 Αυτο-παλινδρομικά Μοντέλα.....	59
4.9 Μαθηματικές Σχέσεις που Διέπουν τη Λειτουργία του SimpleFeedForward.....	60
4.10 Συνάρτηση Απωλειών.....	63
<b>5. Πειραματική Διαδικασία.....</b>	<b>65</b>
5.1 Στόχοι και Δομή Πειραματικής Διαδικασίας.....	65

5.2 Δεδομένα και Προετοιμασία τους.....	66
5.3 Ρύθμιση Υπερ-παραμέτρων και Κατανομής του Μοντέλου.....	67
5.4 Συνάρτηση Απωλειών.....	69
5.5 Κατανομές Πειραμάτων.....	70
5.5.1 Κανονική Κατανομή.....	70
5.5.2 Κατανομή Laplace.....	72
5.5.3 Κατανομή Log-Normal.....	73
5.5.4 Κατανομή Cauchy.....	74
5.5.5 Κατανομή StudentT.....	75
5.6 Εκπαίδευση Μοντέλων και Εξαγωγή Προβλέψεων.....	77
5.7 Αξιολόγηση Μοντέλου.....	78
<b>6. Αποτελέσματα, Συμπεράσματα και Προεκτάσεις.....</b>	<b>79</b>
6.1 Πειραματικό Σύνολο Δεδομένων.....	79
6.2 Σύγκριση Αποτελεσμάτων.....	80
6.2.1 Αποτελέσματα Σημειακών Προβλέψεων.....	80
6.2.2 Αποτελέσματα Πιθανοτικών Προβλέψεων - Σύγκριση Κατανομών.....	82
6.3 Χρόνοι Εκπαίδευσης.....	93
6.4 Συμπεράσματα και Προεκτάσεις.....	93
<b>Βιβλιογραφία.....</b>	<b>95</b>
<b>Παράρτημα Α. Διαγωνισμός M4.....</b>	<b>99</b>
<b>Παράρτημα Β. Παρουσίαση Κώδικα.....</b>	<b>103</b>
Π.Β.1 Υπολογισμός Δεικτών ACF.....	103
Π.Β.2 Test εποχιακότητας.....	103
Π.Β.3 Αποεποχικοποίηση Χρονοσειράς.....	104
Π.Β.4 Δημιουργία Δικτύου Εκπαίδευσης.....	105
Π.Β.4.1 Καθορισμός Κατανομής.....	105
Π.Β.4.2 Καθορισμός Αριθμού και Τύπου Παραμέτρων.....	106
Π.Β.4.3 Καθορισμός Συνάρτησης Απωλειών.....	107
Π.Β.5 Ορισμός Δικτύου.....	107
Π.Β.6 Εκπαίδευση και Εξαγωγή Προβλέψεων.....	107





## Κατάλογος Σχημάτων

Σχήμα 2.1 : Μηνιαίες Εμφανίσεις Ηλιακών Κηλίδων την περίοδο 1934-198.....	7
Σχήμα 2.2 : Εβδομαδιαίες Χορηγήσεις 1ης Δόσης Εμβολίων κατά της COVID-19 στην Ελλάδα για διάστημα 22 εβδομάδων.....	7
Σχήμα 2.3 : Κατηγορίες Μεθόδων Εξομάλυνσης.....	15
Σχήμα 2.4 : Διαίρεση της κατανομής σε ποσοστημόρια.....	24
Σχήμα 2.5 : Μορφή μελλοντικών προβλέψεων με απεικόνιση των prediction intervals (μπλε γραμμές).....	24
Σχήμα 3.1 : Μοντέλο Νευρώνα.....	36
Σχήμα 3.2 : Αρχιτεκτονική του Perceptron.....	40
Σχήμα 3.3 : Το υπερεπίπεδο ως όριο απόφασης για την ταξινόμηση προτύπων με δύο κλάσεις... ..	40
Σχήμα 3.4 : Αρχιτεκτονική Δικτύου Πρόσθιας Τροφοδότησης Ενός Επιπέδου.....	42
Σχήμα 3.5 : Αρχιτεκτονική Δικτύου Πρόσθιας Τροφοδότησης Πολλαπλών Επιπέδων.....	43
Σχήμα 3.6 : Αρχιτεκτονική Αναδρομικών Δικτύων.....	44
Σχήμα 4.1 : Οπτικοποίηση προβλέψεων μέσω του GluonTS .....	55
Σχήμα 4.2: Συγκεντρωτικά οι διαφορές μεταξύ Generative και Discriminative Μοντέλων.....	58
Σχήμα 4.3 : Πρώτο πιθανό παράθυρο.....	62
Σχήμα 4.4 : Δεύτερο πιθανό παράθυρο.....	62
Σχήμα 4.5 : Τελευταίο πιθανό παράθυρο.....	62
Σχήμα 4.6: Μορφή SimpleFeedForward δικτύου που χρησιμοποιήθηκε στην παρούσα διπλωματική εργασία.....	63
Σχήμα 5.1 : Συνάρτηση Πυκνότητας Πιθανότητας Κανονικής Κατανομής.....	71
Σχήμα 5.2 : Συνάρτηση Πυκνότητας Πιθανότητας Κατανομής Laplace.....	72
Σχήμα 5.3 : Συνάρτηση Πυκνότητας Πιθανότητας Κατανομής LogNormal.....	74
Σχήμα 5.4 : Συνάρτηση Πυκνότητας Πιθανότητας Κατανομής Cauchy.....	75
Σχήμα 5.5 : Συνάρτηση Πυκνότητας Πιθανότητας Κατανομής StudentT.....	76
Σχήμα 5.6 : Σχηματική Αναπαράσταση Διαδικασίας Εκπαίδευσης και Πρόβλεψης.....	77
Σχήμα 6.1 : Ποσόστωση μηνιαίων δεδομένων στους επιμέρους κλαδούς.....	79
Σχήμα 6.2 : Bar-Plot δείκτη sMAPE των σημειακών προβλέψεων.....	81
Σχήμα 6.3 : Bar-Plot δείκτη MASE των σημειακών προβλέψεων.....	81
Σχήμα 6.4 : Επεξηγηματικό παράδειγμα πιθανοτικών προβλέψεων.....	82
Σχήμα 6.5 : SPL ανά quantile για κάθε Κατανομή.....	83
Σχήμα 6.6 : Coverage Rate Κατανομής Cauchy.....	86
Σχήμα 6.7 : Coverage Rate Κατανομής Laplace.....	86
Σχήμα 6.8 : Coverage Rate Κατανομής LogNormal.....	87
Σχήμα 6.9 : Coverage Rate Κατανομής Normal.....	87
Σχήμα 6.10 : Coverage Rate Κατανομής StudentT.....	88
Σχήμα 6.11 : RMSSE σφάλματα πιθανοτικών προβλέψεων της κάθε κατανομής ανά quantile....	88
Σχήμα 6.12 : Μεγάλες διαφορές σφαλμάτων μεταξύ ακραίων quantiles και median συνεπάγονται πιο διασκορπισμένες προβλέψεις.....	89
Σχήμα 6.13 : Μικρές διαφορές σφαλμάτων μεταξύ ακραίων quantiles και median συνεπάγονται πιο συμπυκνωμένες προβλέψεις.....	90
Σχήμα 6.14 : Περίπτωση σχετικά συγκεντρωμένων προβλέψεων, με τα ακραία quantiles όμως να αποκλείουν κατά πολύ. ....	90
Σχήμα 6.15 : Περίπτωση όπου το μοντέλο εισάγει συστηματικό σφάλμα στις προβλέψεις.....	91



## Κατάλογος Πινάκων

<i>Πίνακας 5.1: Τιμές Υπερ-Παραμέτρων Μοντέλου.....</i>	<i>69</i>
<i>Πίνακας 6.1 : Κλάδοι Μηνιαίων Δεδομένων διαγωνισμού M4 .....</i>	<i>79</i>
<i>Πίνακας 6.2 : SPL ανά κατανομή και quantile.....</i>	<i>84</i>
<i>Πίνακας 6.3 : RF ανά κατανομή και quantile.....</i>	<i>85</i>
<i>Πίνακας 6.4 : Απόκλιση RF κάθε κατανομής από την ιδανική περίπτωση.....</i>	<i>86</i>
<i>Πίνακας 6.5 : Συγκεντρωτικά αποτελέσματα RMSSE ανά κατανομή και quantile .....</i>	<i>89</i>
<i>Πίνακας 6.6 : Απαίτηση κάθε κατανομής σε χρόνο και μνήμη .....</i>	<i>93</i>
<i>ΠΙΝΑΚΑΣ Π.Α.1: Πλήθος χρονοσειρών διαγωνισμού M4 ανά κατηγορία και χρονικό επίπεδο.</i>	<i>102</i>



# ΚΕΦΑΛΑΙΟ 1

## Εισαγωγή

### *1.1 Αντικείμενο της Εργασίας*

Η έννοια της πρόβλεψης είναι συνυφασμένη με όλο το φάσμα των ανθρώπινων δραστηριοτήτων, καθώς οι αποφάσεις που καλείται να λάβει ο εκάστοτε αποφασίζων απαιτούν μία εκτίμηση της μελλοντικής κατάστασης των μεγεθών που τον ενδιαφέρουν. Η οντότητα του αποφασίζοντος εκτείνεται από το επίπεδο ενός ατόμου ή μίας επιχείρησης έως αυτό ενός οργανισμού ή κράτους.

Δεδομένης της εγγενούς αβεβαιότητας του μέλλοντος και των αγνώστων παραγόντων που επηρεάζουν τις εξελίξεις κάθε ανθρώπινης δραστηριότητας, οι προβλέψεις έχουν καταστεί κεφαλαιώδους σημασίας, ιδιαίτερα τα τελευταία χρόνια κατά τα οποία οι ταχείες και ταυτόχρονα οβιδιακές μεταβολές που συντελούνται, αυξάνουν έτι περαιτέρω την προαναφερθείσα αβεβαιότητα. Δραστηριοποιούμενες σε ένα τέτοιο περιβάλλον, οι επιχειρήσεις καλούνται να λάβουν σημαντικές αποφάσεις με σκοπό τη βιώσιμη λειτουργία τους και την ελαχιστοποίηση του κόστους. Σε αυτό το πλαίσιο, έχει γίνει προσπάθεια ανάπτυξης τεχνικών προβλέψεων που περιλαμβάνουν στατιστικά μοντέλα τα οποία ενδέχεται να συνεπικουρούνται από τις εμπειρικές προβλέψεις, ώστε η τελική πρόβλεψη να συνδυάζει τα οφέλη της μαθηματικής ακρίβειας και την διορατικότητα της ανθρώπινης εμπειρίας.

Ο κλάδος των προβλέψεων θεωρείται ιδιαίτερα δυναμικός, καθώς άμα τη εμφανίσει νέων μεθόδων, γίνονται προσπάθειες προσαρμογής τους σε πλαίσιο που να εξυπηρετεί τις ανάγκες των προβλέψεων. Χαρακτηριστικό παράδειγμα του προαναφερθέντος αποτελεί η ενσωμάτωση της μηχανικής μάθησης, καθώς έπειτα από την αλματώδη πρόοδο της τεχνητής νοημοσύνης κατά τη διάρκεια των τελευταίων ετών, το ενδιαφέρον της επιστημονικής και επιχειρηματικής κοινότητας έχει στραφεί στην εύρεση τρόπων αξιοποίησής της. Ως εκ τούτου γίνονται προσπάθειες ανάπτυξης μη γραμμικών μεθόδων πρόβλεψης, οι οποίες είναι ικανές να υπερβούν τις δυσκολίες που αντιμετωπίζουν οι κλασσικές μέθοδοι.

Οι μέθοδοι μηχανικής μάθησης είχαν καταρχάς αμφισβητηθεί λόγω της υστέρησής τους, από πλευράς προβλεπτικής ακρίβειας, έναντι των κλασσικών μεθόδων, ωστόσο συν τω χρόνω, η απόδοσή τους βελτιώνεται σημαντικά. Η ανάπτυξη δε της βαθιάς μηχανικής μάθησης, έχει δώσει τη δυνατότητα στους ερευνητές να αναπτύσσουν μοντέλα που προσεγγίζουν την έννοια της πρόβλεψης υπό το πρίσμα της πιθανοτικής φύσης τους, με τρόπο αισθητά πιο ευέλικτο σε σχέση με τις δυνατότητες που παρέχουν κλασσικές στατιστικές μέθοδοι, όπως επί παραδείγματι η μέθοδος Theta.

Η πιθανοτική φύση της πρόβλεψης διαφέρει ως προς τη ντετερμινιστική (σημειακή) ως προς το ότι δεν έχουν εξάγουν μία μοναδική τιμή ως το “καλύτερο” αποτέλεσμα, αλλά αναθέτουν σε κάθε πιθανό αποτέλεσμα μία τιμή πιθανότητας να λάβει πράγματι χώρα. Η διασταλτική αυτή προσέγγιση επιτρέπει την εξέταση όλων των πιθανών εναλλακτικών, δίνοντας τη δυνατότητα

στον αποφασίζοντα να λάβει την απόφαση εκείνη που θεωρεί βέλτιστη βάσει των διαφορετικών σεναρίων που αναπτύσσονται.

Η παρούσα διπλωματική επιχειρεί την εξοικείωση του αναγνώστη με την έννοια της πιθανοτικής πρόβλεψης και τη γνωριμία του με ένα σχετικά πρόσφατο εργαλείο του GluonTS (Amazon Web Services - AWS), το οποίο δημιουργήθηκε με σκοπό την ανάπτυξη πιθανοτικών μοντέλων πρόβλεψης που αξιοποιούν την βαθιά μηχανική μάθηση. Παράλληλα, στη βάση της αμφισβήτησης της υπόθεσης της κανονικής κατανομής για την εξαγωγή των πιθανοτικών προβλέψεων, επιχειρείται η σύγκριση των προβλέψεων που εξάγουν τριάντα μοντέλα πρόσθιας τροφοδότησης, έπειτα από την εφαρμογή πέντε διαφορετικών κατανομών.

## ***1.2 Οργάνωση της Εργασίας***

Στο 2ο Κεφάλαιο της παρούσας εργασίας γίνεται η παρουσίαση του κλάδου των προβλέψεων και η απαραίτητη θεωρητική θεμελίωση των εννοιών που αναλύονται στα επόμενα Κεφάλαια. Έτσι εισάγονται απαραίτητα εργαλεία όπως οι χρονοσειρές και οι γραφικές τους παραστάσεις, παρουσιάζονται τα βήματα εξαγωγής μίας πρόβλεψης και αναφέρονται οι πιο κλασσικές στατιστικές και μη μέθοδοι πρόβλεψης. Τέλος, παρουσιάζονται οι βασικότεροι στατιστικοί δείκτες που χρησιμοποιούνται για την αξιολόγηση των προβλέψεων.

Στο 3ο Κεφάλαιο εισάγονται οι έννοιες της μηχανικής μάθησης και των νευρωνικών δικτύων. Αφού αναλυθούν τα κυριότερα χαρακτηριστικά και οι συναρτήσεις ενεργοποίησής τους, παρουσιάζονται οι δημοφιλέστερες αρχιτεκτονικές νευρωνικών δικτύων και εν συνεχεία οι κατηγορίες διαδικασιών μάθησής τους.

Το 4ο Κεφάλαιο αποτελεί την παρουσίαση της εργαλειοθήκης του GluonTS. Αρχικά αναλύονται οι βασικές έννοιες που αφορούν τη βαθιά μηχανική μάθηση και των υπερ-παραμέτρων τέτοιων δικτύων. Ακολούθως, η παρουσίαση επικεντρώνεται το GluonTS, τη λειτουργία του, τον τρόπο δημιουργίας μοντέλων χρονοσειρών και θεμελιώνονται μαθηματικά οι ευρύτερες κατηγορίες στις οποίες δύνανται να διακριθούν τόσο έτοιμα μοντέλα που προσφέρονται μέσω της εργαλειοθήκης αυτής, όσο και τα μοντέλα που αναπτύσσονται μέσω του GluonTS. Στο τέλος του κεφαλαίου γίνεται ιδιαίτερη αναφορά στο μοντέλο Απλής Πρόσθιας Τροφοδότησης (SimpleFeedForward) που αποτέλεσε τον πειραματικό πυρήνα της παρούσας διπλωματικής εργασίας, καθώς παρουσιάζονται οι μαθηματικές σχέσεις που διέπουν τη λειτουργία του.

Στο 5ο Κεφάλαιο παρουσιάζεται η μεθοδολογία που ακολουθήθηκε στο πλαίσιο της πειραματικής μελέτης που διενεργήθηκε, η οποία αφορά την προετοιμασία των δεδομένων εκπαίδευσης, τη ρύθμιση των υπερ-παραμέτρων του δικτύου SimpleFeedForward, τη συνάρτηση απωλειών που χρησιμοποιήθηκε κατά την εκπαίδευσή του και εν τέλει την εξαγωγή των προβλέψεων. Να σημειωθεί ότι στο εν λόγω Κεφάλαιο γίνεται παράλληλα εκτενής αναφορά στις πέντε Κατανομές που εξετάστηκαν, στις ιδιότητες και τις παραμέτρους τους.

Το 6ο Κεφάλαιο περιλαμβάνει την παρουσίαση των ποσοτικών αποτελεσμάτων που προέκυψαν από την πειραματική διαδικασία. Η αρχή του Κεφαλαίου αφορά την παρουσίαση του συνόλου δεδομένων που χρησιμοποιήθηκαν για την εκπαίδευση και την αξιολόγηση των μοντέλων και στη συνέχεια γίνεται ανάλυση τόσο των σημειακών όσο και των πιθανοτικών προβλέψεων που

ελήφθησαν από το μοντέλο. Τέλος βάσει των ανωτέρω αποτελεσμάτων, γίνεται σύγκριση μεταξύ των διαφορετικών κατανομών, με σκοπό τον προσδιορισμό της βέλτιστης προσέγγισης ενός προβλήματος πιθανοτικών προβλέψεων με χρήση της μεθοδολογίας που ακολουθήθηκε στην παρούσα εργασία.

Το 7ο Κεφάλαιο αποσκοπεί στην αποτίμηση των συμπερασμάτων που προέκυψαν από τα ποσοτικά αποτελέσματα και επιχειρεί να παραθέσει ορισμένες σκέψεις αναφορικά με τη χρησιμότητα των πιθανοτικών προβλέψεων στο σύγχρονο περιβάλλον και, καθώς επίσης να δημιουργήσει ορισμένα ερεθίσματα για μελλοντικές μελέτες.

Η παρούσα διπλωματική εργασία ολοκληρώνεται με δύο Παραρτήματα που περιέχουν επικουρικά στοιχεία της εργασίας και στοχεύουν στην πληρότητά της, μέσω της παρουσίασης των διαγωνισμών Μ και την επεξήγηση του κώδικα που χρησιμοποιήθηκε για τη διενέργεια των πειραμάτων.





## ΚΕΦΑΛΑΙΟ 2

### Προβλέψεις και Δείκτες Ακριβείας

#### *2.1 Χρονοσειρές και Ποιοτικά Χαρακτηριστικά Χρονοσειρών*

Οι χρονοσειρές ή χρονολογικές σειρές αποτελούν σύνολα διαδοχικών παρατηρήσεων της τιμής ενός φυσικού ή άλλου μεγέθους, οι οποίες δεν είναι ανεξάρτητες μεταξύ τους, αλλά υπάρχει μία εγγενής σχέση, βάσει της οποίας καθίσταται δυνατός ο προσδιορισμός των μελλοντικών τιμών από τις προηγούμενες. Μία τέτοια διαδικασία ονομάζεται ντετερμινιστική. Το προαναφερθέν εκφράζει μερικώς την πραγματικότητα, καθώς στις πραγματικές χρονοσειρές το παρελθόν καθορίζει μόνο έως ένα βαθμό το μέλλον, καθώς σε όλα τα μεγέθη επενεργούν ποικίλοι εξωγενείς παράγοντες. Δεδομένων των προαναφερθέντων, οι χρονοσειρές αποτελούν στην ουσία στοχαστικές διαδικασίες, αφού ο τυχαίος παράγοντας επηρεάζει σημαντικά την εξέλιξη των μεγεθών.

Η πλειοψηφία των αμιγώς στατιστικών μεθόδων ανάλυσης και πρόβλεψης χρονοσειρών, βασίζεται στην έννοια της αποσύνθεσης, κατά την οποία η αρχική χρονοσειρά, διαχωρίζεται σε επιμέρους, κάθε μία εκ των οποίων φανερώνει ένα χαρακτηριστικό της αρχικής. Πιο συγκεκριμένα, τα χαρακτηριστικά αυτά είναι η τάση, η εποχικότητα, η κυκλικότητα και η διακύμανση με τα σημαντικότερα στοιχεία για καθένα από αυτά να παρουσιάζονται ακολούθως:

- **Τάση:** φανερώνει ενδεχόμενη μακροπρόθεσμη μεταβολή του μέσου επιπέδου των παρατηρήσεων της χρονοσειράς. Η επιλογή κατάλληλου διαστήματος μέσα στο οποίο θα εξεταστεί η μεταβολή, θα πρέπει να γίνεται με προσοχή, λαμβάνοντας υπόψη ότι η κυκλική διακύμανση (που θα εξεταστεί στη συνέχεια), ενδέχεται να υπερβαίνει χρονικά τα διαθέσιμα δεδομένα, με αποτέλεσμα να θεωρηθεί εσφαλμένα ως τάση. Το διάστημα επομένως, λαμβάνεται συναρτήσει του πλήθους των διαθέσιμων παρατηρήσεων αλλά και της φύσης των μεγεθών που αυτές περιγράφουν, όπως επί παραδείγματι στην περίπτωση των χρονοσειρών ζήτησης προϊόντων λιανικής, στις οποίες η πληροφορία μερικών ετών κρίνεται αρκετή για να αποφανθεί κανείς σχετικά με την ύπαρξη τάσης.
- **Κυκλικότητα:** αντιπροσωπεύει μία “κυματοειδή” μεταβολή που οφείλεται σε εξωγενείς συνθήκες και εμφανίζεται κατά περιόδους οι οποίες δεν είναι απαραίτητα σταθερές, ενώ το μήκος τους υπερβαίνει κατά κανόνα το ένα έτος. Η παρουσία κυκλικότητας γίνεται αισθητή στις χρονοσειρές των περισσότερων οικονομικών μεγεθών, όπως λόγου χάριν του Ακαθάριστου Εθνικού Προϊόντος, των τιμών των μετοχών ή των δεικτών βιομηχανικής παραγωγής, καθώς τα μεγέθη αυτά επηρεάζονται από τις εν γένει οικονομικές συνθήκες οι οποίες χαρακτηρίζονται από διαδοχικές ανόδους και υφέσεις, σχηματίζοντας τους λεγόμενους επιχειρηματικούς κύκλους.

- **Εποχικότητα:** αποτελεί μία περιοδική διακύμανση με σταθερό και μικρότερο του έτους μήκος, ενώ είναι συνήθως κατανοητή και προβλέψιμη. Οι αλλαγές που οφείλονται στην εποχικότητα επαναλαμβάνονται κατά τον ίδιο τρόπο ακριβώς κάθε μήκος εποχικότητας, καθιστώντας τις ως εκ τούτου άμεσα εξηγήσιμες και εύκολα μετρήσιμες, με αποτέλεσμα να μπορούν να απομονωθούν από την αρχική χρονοσειρά και να προκύψουν τα αποεποχικοποιημένα δεδομένα. Χαρακτηριστικό παράδειγμα εποχικού μεγέθους είναι οι πωλήσεις παγωτού, οι οποίες εμφανίζουν έντονη άνοδο τους καλοκαιρινούς μήνες, σε αντίθεση με ένα άλλο άκρως εποχικό προϊόν, τα χριστουγεννιάτικα δέντρα, οι πωλήσεις των οποίων εντοπίζονται πριν και κατά την περίοδο των χειμερινών διακοπών.
- **Ασυνέχειες:** αποτελούν μεμονωμένες παρατηρήσεις οι οποίες εμφανίζονται στο γράφημα κάποιας χρονοσειράς ως απότομες αλλαγές στο πρότυπο συμπεριφοράς της χρονοσειράς και δε θα μπορούσαν να προβλεφθούν από τα διαθέσιμα ιστορικά δεδομένα. Οι αλλαγές αυτές μπορούν να διακριθούν σε παροδικού ή μόνιμου χαρακτήρα. Οι ασυνέχειες της πρώτης κατηγορίας αφορούν παρατηρήσεις που αποκλίνουν πολύ από τις υπόλοιπες και σχετίζονται με ειδικά γεγονότα. Ένα τέτοιο outlier (όπως έχει επικρατήσει να ονομάζονται) θα μπορούσε να είναι μία απεργία η οποία προκάλεσε δραματική πτώση της παραγωγής μιας βιομηχανίας. Ο χειρισμός των special events (outliers) πρέπει να γίνεται με προσοχή και απαιτεί άρτια θεωρητική κατάρτιση, κριτική ικανότητα καθώς επίσης και κοινή λογική, ενώ καταλυτικής σημασίας κρίνεται η εμπειρία, δεδομένου του γεγονότος ότι ένας ειδικός του εκάστοτε χώρου μπορεί ευκολότερα να ανιχνεύσει την ύπαρξη ενός special event. Εν αντιθέσει με τον παροδικό χαρακτήρα των outliers, οι αλλαγές επιπέδου (level-shifts) που ανήκουν στη δεύτερη κατηγορία, έχουν μόνιμες συνέπειες στη χρονοσειρά, καθώς πλέον το μέσο επίπεδο των τιμών των παρατηρήσεων μεταβάλλεται με απότομο τρόπο. Χαρακτηριστικό παράδειγμα αλλαγής επιπέδου στην περίπτωση των πωλήσεων μιας εταιρείας αποτελεί η εμφάνιση κάποιας ανταγωνίστριας εταιρείας, εμφάνιση η οποία συνεπάγεται μείωση των πωλήσεων της πρώτης και ως εκ τούτου τη σταθεροποίηση του μέσου επιπέδου της χρονοσειράς πωλήσεων σε ένα χαμηλότερο σημείο.
- **Μη κανονικές διακυμάνσεις:** όταν απομονωθούν η τάση, η εποχικότητα και η κυκλικότητα, που αποτελούν τα συστατικά στοιχεία της χρονοσειράς, απομένουν οι μη κανονικές διακυμάνσεις, οι οποίες ενδέχεται να αντιπροσωπεύουν μία εντελώς στατιστικά τυχαία μεταβλητή ή κάποια ασυνέχεια που να συνδέεται με κάποιο εξαιρετικό γεγονός

## 2.2 Γραφική Αναπαράσταση Χρονοσειρών

Η γραφική αναπαράσταση που αποτελεί ένα ιδιαίτερα σημαντικό εργαλείο τόσο για την ανάλυση της χρονοσειράς όσο και για τη διαδικασία της πρόβλεψης, έγκειται στη δισδιάστατη γραφική απεικόνιση ως προς το χρόνο, των πραγματικών τιμών των διαθέσιμων δεδομένων. Η χρησιμότητα της γραφικής αναπαράστασης είναι προφανής, αν αναλογιστεί κανείς το ότι καθιστά εμφανή τα ποιοτικά χαρακτηριστικά της χρονοσειράς, με αποτέλεσμα τη διευκόλυνση του αναλυτή κατά την επιλογή των κατάλληλων μεθοδολογιών και εργαλείων ώστε να

προκύβουν τα βέλτιστα αποτελέσματα και το μικρότερο δυνατό σφάλμα πρόβλεψης στην εκάστοτε περίπτωση. Πέραν της ανάδειξης των ποιοτικών χαρακτηριστικών, η γραφική αναπαράσταση δύναται επιπλέον να αποκαλύψει εσφαλμένες ή ακραίες τιμές, τις οποίες μπορεί στη συνέχεια να διαχειριστεί κατάλληλα ο αναλυτής. Στα διαγράμματα που ακολουθούν, παρουσιάζονται δύο διαφορετικές περιπτώσεις χρονοσειρών.



Σχήμα 2.1 : Μηνιαίες Εμφανίσεις Ηλιακών Κηλίδων την περίοδο 1934-1983



Σχήμα 2.2 : Εβδομαδιαίες Χορηγήσεις 1ης Δόσης Εμβολίων κατά της COVID-19 στην Ελλάδα για διάστημα 22 εβδομάδων

Στην πρώτη περίπτωση όπου απεικονίζονται οι εμφανίσεις του φαινομένου των ηλιακών κηλίδων σε μηνιαία βάση για την περίοδο 1934 έως και 1983 (οι παρατηρήσεις ελήφθησαν από τους Andrews & Herzberg, 1985), είναι εμφανής η ύπαρξη εποχικότητας στα δεδομένα, καθώς παρατηρείται έξαρση του φαινομένου τους φθινοπωρινούς μήνες. Επιπλέον, ενδιαφέρον

παρουσιάζει το γεγονός ότι σε κάθε μήκος εποχικότητας, η οι παρατηρήσεις είναι αρκετά διαφορετικές ως προς την τιμή τους, διατηρούν ωστόσο το μοτίβο που παρατηρείται στο σύνολο της χρονοσειράς.

Στη δεύτερη περίπτωση παρουσιάζονται τα δεδομένα από το Ευρωπαϊκό Κέντρο Ελέγχου και Πρόληψης Νόσων, που σχετίζονται με τις εβδομαδιαίες χορηγήσεις της πρώτης δόσης του εμβολιαστικού σχήματος στην Ελλάδα για την περίοδο τέλη Δεκεμβρίου 2020 - μέσα Μαΐου 2021. Εμφανές ποιοτικό χαρακτηριστικό της χρονοσειράς αυτής αποτελεί η ανοδική τάση, καθώς συν τω χρόνω, οι εβδομαδιαίοι εμβολιασμοί φαίνεται να αυξάνονται.

### **2.3 Γενικά Περί Προβλέψεων**

Έχοντας πλέον παρουσιάσει ένα από τα βασικότερα εργαλεία της επιστήμης των Προβλέψεων, είμαστε σε θέση να εξετάσουμε ορισμένα γενικά στοιχεία που αφορούν αυτό το νέο, σχετικά, επιστημονικό κλάδο αυτό καθώς και τις εφαρμογές του στην καθημερινότητα, τόσο σε ατομικό επίπεδο, όσο και σε επίπεδο στρατηγικών αποφάσεων μικρών ή μεγάλων επιχειρήσεων και οργανισμών.

Αναφορικά με το πρώτο, είναι πλείστες οι φορές που καλούμαστε να κάνουμε προβλέψεις για θέματα της καθημερινότητας, όπως το αν θα πάρουμε ομπρέλα μαζί μας φεύγοντας από το σπίτι, το οποίο αποτελεί ένα απλό παράδειγμα πρόβλεψης, για την εξαγωγή της οποίας συναξιολογούνται διάφοροι παράγοντες, όπως επί παραδείγματι τα καιρικά δεδομένα. Στον αντίποδα, οι προβλέψεις που καλούνται να εξάγουν οι επιχειρήσεις είναι ζωτικής σημασίας για τη λειτουργία αλλά και την εν γένει βιωσιμότητά τους. Έτσι, ένας αρτοποιός καλείται να προβλέψει τη ζήτηση άρτου της επόμενης μέρας με τον ακριβέστερο δυνατό τρόπο, ώστε αφενός να μην οδηγηθεί σε ζημία λόγω της μη ικανοποίησης της ζήτησης (ill will) και αφετέρου να μην μείνει αδιάθετο εμπόρευμα που θα οδηγούσε σε μεγαλύτερο κόστος παραγωγής και υλικών. Η πολυπλοκότητα των επιχειρησιακών προβλέψεων εταιρειών και οργανισμών μπορεί να είναι ιδιαίτερα μεγάλη, όπως στην περίπτωση της πρόβλεψης της πορείας της εθνικής οικονομίας για το επόμενο διάστημα.

Από τα παραπάνω καθίσταται σαφές ότι στόχο των προβλέψεων αποτελεί η υποστήριξη των αποφάσεων που πρέπει να ληφθούν σε συγκεκριμένες χρονικές στιγμές, γεγονός που αιτιολογεί τη μεγάλη ανάπτυξη που γνωρίζει ο κλάδος τα τελευταία τριάντα χρόνια. Το αυξημένο ενδιαφέρον εντοπίζεται τόσο στην ακαδημαϊκή όσο και στην επιχειρηματική κοινότητα, καθώς η ολοένα και αυξανόμενη αβεβαιότητα σε παγκόσμιο επίπεδο, επιτάσσει μία πιο συστηματική και προσεκτική έρευνα του μέλλοντος. Το οξύμωρο που παρατηρείται είναι ότι το ενδιαφέρον για προβλέψεις αυξάνεται σε χρονικές περιόδους που χαρακτηρίζονται από συνεχείς και απότομες μεταβολές, κατά τις οποίες είναι γνωστό ότι η προβλεπτική ικανότητα είναι αισθητά μειωμένη.

Απότοκο του προαναφερθέντος είναι η έντονη και πολλές φορές δυσμενής κριτική που δέχεται ο κλάδος των Προβλέψεων, λόγω της ανικανότητας έγκαιρης προειδοποίησης για επερχόμενες αλλαγές ή των μεγάλων σφαλμάτων που προκύπτουν. Η αβεβαιότητα αυτή ωστόσο, δημιουργεί μεγαλύτερες ανάγκες για πρόβλεψη, με αποτέλεσμα τη δημιουργία ενός σχήματος κύκλου. Μία πρώτη κατηγοριοποίηση των προβλέψεων μπορεί να γίνει βάσει του είδους τους, κατά την οποία διακρίνονται τριών ειδών προβλέψεις:

- *Στατιστικές*: οι οποίες προκύπτουν με τη χρήση στατιστικών μεθόδων ανάλυσης χρονοσειρών
- *Κριτικές*: οι οποίες αποτελούν εμπειρικές προσεγγίσεις ενός ή μιας ομάδας ατόμων (επιτροπή), που συνήθως είναι ειδικοί στο αντικείμενο που αφορούν τα προς πρόβλεψη μεγέθη
- *Προϋπολογισμού*: στις οποίες εντάσσονται οι στόχοι που τίθενται στο πλαίσιο ενός οργανισμού ή μιας επιχείρησης.

Αξίζει να σημειωθεί ότι στην πράξη είναι δυνατός ο συνδυασμός των ανωτέρω κατηγοριών, για την εξαγωγή μιας “Τελικής Πρόβλεψης”. Στην επόμενη ενότητα του παρόντος κεφαλαίου, θα αναφερθούν οι πιο γνωστές στατιστικές μέθοδοι ανάλυσης χρονοσειρών.

Η διαδικασία εξαγωγής των προβλέψεων είναι δυνατό να περιγραφεί με τη χρήση μοντέλων. Εν γένει η έννοια του μοντέλου αφορά την αναπαράσταση μίας διαδικασίας με αφαιρετικό τρόπο στοχεύοντας σε μία απλοϊκή προσέγγιση του μηχανισμού της. Εν προκειμένω θα εξεταστούν τα μοντέλα ποσοτικών μεθόδων που μπορούν να διακριθούν σε δύο κατηγορίες: τα μοντέλα χρονοσειρών (time series models) και τα αιτιοκρατικά μοντέλα (explanatory models).

## 2.4 Μοντέλο Χρονοσειρών

Τα μοντέλα Χρονοσειρών μπορούν να θεωρηθούν ως ένας από τους θεμέλιους λίθους του κλάδου των Προβλέψεων. Πρόκειται για το πιο διαδεδομένο είδος ποσοτικού μοντέλου που βασίζεται στην υπόθεση ότι η μεταβολή της τιμής του υπό εξέταση μεγέθους ακολουθεί ένα επαναλαμβανόμενο και χρονικά σταθερό πρότυπο (εμφανές ή λανθάνον). Έτσι για την πρόβλεψη των μελλοντικών τιμών, αρκεί η επέκταση του προτύπου αυτού. Είσοδο του συστήματος αποτελεί το σύνολο των διαθέσιμων ιστορικών δεδομένων και έξοδο η ζητούμενη πρόβλεψη. Η σχέση μεταξύ εισόδου και εξόδου είναι γραμμική και έχει τη εξής μορφή:

$$Y = f(X_1, X_2, \dots, X_n)$$

όπου  $X_1, X_2, \dots, X_n$  είναι τα ιστορικά δεδομένα, με  $X_1$  να είναι η αρχαιότερη τιμή και  $X_n$  η πλέον πρόσφατη.

Οι μέθοδοι που περιγράφονται από το συγκεκριμένο μοντέλο συνθέτουν την κατηγορία των μεθόδων χρονοσειρών, η οποία αποτελεί μία ιδιαίτερη κατηγορία των ποσοτικών μεθόδων. Σε αυτή συγκαταλέγονται μεταξύ άλλων η αποσύνθεση (decomposition), η εξομάλυνση (smoothing) και οι αυτοπαλινδρομικές μέθοδοι κινητού μέσου όρου (Autoregressive Moving Average - ARMA), για τις οποίες θα αναφερθούν στη συνέχεια ορισμένα χρήσιμα στοιχεία.

Το εν λόγω μοντέλο κρίνεται ιδιαίτερα αποτελεσματικό όταν η υπό εξέταση χρονοσειρά διατηρεί το μοτίβο συμπεριφοράς της, όπως στην περίπτωση του μελλοντικού ύψους των πωλήσεων ενός προϊόντος, δεδομένου ότι η ποιότητα, η τιμή και το διαφημιστικό του κόστος παραμένουν σταθερά. Ωστόσο, αν επενεργούν εξωγενείς παράγοντες, η αποτελεσματικότητα του μοντέλου μειώνεται καθώς αδυνατεί να συσχετίσει τις προβλέψεις με τις όποιες εξωτερικές

ενέργειες. Συνεπώς, στην περίπτωση της πρόβλεψης των πωλήσεων που αναφέρθηκε προηγουμένως, αν η διαφημιστική πολιτική της εταιρείας διαφοροποιηθεί, οι προβλέψεις δε θα είναι ικανοποιητικές.

Δεδομένων των προαναφερθέντων, είναι αναμενόμενη η θεώρηση του μοντέλου χρονοσειρών ως μαύρου κουτιού (black box), καθώς παρέχει πληροφορία αναφορικά με το τι πρόκειται να συμβεί, χωρίς ωστόσο να δίνει τη δυνατότητα κατανόησης του γιατί θα συμβεί. Μολαταύτα, οι μέθοδοι που περιγράφονται από το μοντέλο των χρονοσειρών επιλέγονται συχνότατα, καθώς σε αρκετές περιπτώσεις το ενδιαφέρον εστιάζεται στην πρόβλεψη του τι θα συμβεί και όχι του γιατί, ενώ παράλληλα η διαδικασία συσχέτισης των εξωγενών παραγόντων και ο προσδιορισμός της ακριβούς αλληλεπίδρασή τους με το υπό εξέταση μέγεθος κρίνεται ασύμφορη λόγω της δυσκολίας της (στις περιπτώσεις που δεν είναι ακόμα και αδύνατη).

### **2.4.1 Αιτιοκρατικό Μοντέλο**

Το αιτιοκρατικό μοντέλο βασίζεται στην υπόθεση ότι υπάρχει σταθερή σχέση αιτίου-αιτιατού μεταξύ των παραμέτρων που επηρεάζουν το υπό πρόβλεψη μέγεθος (ανεξάρτητες μεταβλητές) και του μεγέθους καθ'αυτού (εξαρτημένη μεταβλητή). Η παραπάνω σχέση μπορεί να διατυπωθεί υπό την εξής μορφή:

$$y = f(X_1, X_2, \dots, X_n)$$

όπου  $X_1, X_2, \dots, X_n$  είναι τα διαθέσιμα ιστορικά δεδομένα του εκάστοτε μεγέθους.

Έτσι στο παράδειγμα των πωλήσεων, ως ανεξάρτητες μεταβλητές, θα μπορούσαν να θεωρηθούν η τιμή πώλησης και το διαφημιστικό κόστος, ενώ ως εξαρτημένη το ύψος των πωλήσεων του προϊόντος.

Θεμελιώδης διαφορά των δύο μοντέλων αποτελεί η ο τρόπος που προσδιορίζεται η συνάρτηση που περιγράφει το καθένα από αυτά. Αναλυτικότερα, στα μοντέλα χρονοσειρών η συνάρτηση που περιγράφει το σύστημα θεωρείται προκαθορισμένη και προσδιορίζεται από το πρότυπο που υποθέτουμε ότι ακολουθούν τα ιστορικά δεδομένα, ενώ στα αιτιοκρατικά η συνάρτηση περιγραφής του συστήματος προσδιορίζεται μέσω της παρατήρησης του  $y$  και του συσχετισμού της με τις αντίστοιχες τιμές των παραμέτρων  $X_i$  που θεωρούνται δεδομένες.

Οι αιτιοκρατικές μέθοδοι πρόβλεψης υστερούν στο ότι απαιτούν πολύ περισσότερα δεδομένα σε σχέση με τις μεθόδους χρονοσειρών καθώς απαιτούν περισσότερη πληροφορία για την υπό πρόβλεψη μεταβλητή αλλά και το σύνολο των ανεξάρτητων μεταβλητών, με τα απαιτούμενα ιστορικά δεδομένα να θεωρείται αναγκαίο να είναι διαθέσιμα για τουλάχιστον 50 χρονικές περιόδους στο παρελθόν. Επιπρόσθετα μειονεκτήματά τους αποτελούν η μεγάλη τους ευαισθησία σε οποιαδήποτε αλλαγή της σχέσης μεταξύ ανεξάρτητων μεταβλητών και εξεταζόμενου μεγέθους καθώς επίσης και η ανάγκη χρήσης κάποιας άλλης μεθόδου πρόβλεψης για τον προσδιορισμό των τιμών ορισμένων μεταβλητών εισόδου, γεγονός που αυξάνει σημαντικά το υπολογιστικό τους κόστος.

Στον αντίποδα εμφανίζουν ένα ιδιαίτερα σημαντικό πλεονέκτημα που αφορά τη δυνατότητα που παρέχουν στο χρήστη να προβλέψει τη μελλοντική τιμή κάποιου μεγέθους για διαφορετικούς συνδυασμούς των μεταβλητών εισόδου, καθιστώντας εφικτή τη μελέτη εναλλακτικών σεναρίων με σκοπό την επιλογή του βέλτιστου συνδυασμού αποφάσεων. Συνεπώς στο παράδειγμα της πρόβλεψης του ύψους των πωλήσεων, ο αναλυτής έχει τη δυνατότητα να εκτιμήσει τις πωλήσεις με βάση μια ορισμένη πολιτική, αλλά παράλληλα να προσδιορίσει τη βέλτιστη πολιτική βάσει του επιθυμητού ύψους πωλήσεων.

## **2.5 Βασικά Βήματα της Διαδικασίας Πρόβλεψης**

Οι Μακρυδάκης, Wheelright και Hyndman (1998) έχουν περιγράψει ένα μεθοδολογικό πλαίσιο για την παραγωγή και την αξιολόγηση των προβλέψεων, το οποίο μπορεί να διακριθεί στα ακόλουθα πέντε βήματα:

**ΒΗΜΑ 1:** Καθορισμός του προβλήματος : Συνήθως αποτελεί το δυσκολότερο αλλά συνάμα το πιο σημαντικό μέρος της διαδικασίας, καθώς είναι απαραίτητος ο προσδιορισμός του τι θα προβλεφθεί, με ποιον τρόπο και πώς θα χρησιμοποιηθεί η εξαχθείσα πρόβλεψη

**ΒΗΜΑ 2:** Συλλογή των δεδομένων: Το βήμα αυτό είναι αρκετά χρονοβόρο και κρίνεται μείζονος σημασίας, δεδομένου του γεγονότος ότι θα πρέπει να αφιερωθεί επαρκής χρόνος στην ορθή συλλογή των δεδομένων. Ιδιαίτερης χρησιμότητας είναι επίσης η συλλογή και η αξιοποίηση επιπλέον δεδομένων πέρα των αμιγώς ποσοτικών, που σχετίζονται επί παραδείγματι με την πείρα και την εξειδικευμένη γνώση των εργαζομένων σε μία επιχείρηση για την εξεταζόμενη χρονική περίοδο.

**ΒΗΜΑ 3:** Προετοιμασία των χρονοσειρών: προεπεξεργασία και ανάλυση των δεδομένων. Στο βήμα αυτό εξετάζονται τα ποιοτικά δεδομένα των χρονοσειρών ώστε να διαμορφωθεί μία ολοκληρωμένη εικόνα των διαθέσιμων δεδομένων, αναγνωρίζοντας πιθανά πρότυπα, τάση ή εποχικότητα καθώς επίσης και ασυνήθιστες τιμές. Εν συνεχεία, γίνεται απομόνωση των συνιστωσών της χρονοσειράς όπως και των ασυνήθιστων τιμών, ώστε να προκύψει μία εξομαλυμένη σειρά δεδομένων, ώστε η είσοδος των μοντέλων να έχει καταλληλότερη μορφή.

**ΒΗΜΑ 4:** Επιλογή μεθόδων πρόβλεψης: Βάσει των ιστορικών δεδομένων και των ιδιοτήτων της κάθε μεθόδου, γίνεται επιλογή των κατάλληλων μοντέλων πρόβλεψης και των αντίστοιχων παραμέτρων τους, με σκοπό την παραγωγή προβλέψεων για το ζητούμενο χρονικό ορίζοντα με στατιστική ή κριτική προέκταση των διαθέσιμων δεδομένων.

**ΒΗΜΑ 5:** Χρήση και αξιολόγηση των μοντέλων πρόβλεψης: Το επιλεγμένο μοντέλο πρόβλεψης, τίθεται σε χρήση και εξάγονται οι ζητούμενες προβλέψεις. Κριτήριο που καθορίζει το πόσο ικανοποιητικό είναι το μοντέλο αλλά και οι παραχθείσες προβλέψεις, αποτελεί ο χρόνος, καθώς άμα τη εμφανίσει νέων δεδομένων, γίνεται χρήση τους για την αξιολόγηση και τη μέτρηση της ακρίβειας των προβλέψεων με εξειδικευμένους στατιστικούς δείκτες. Έτσι παρακολουθώντας συν τω χρόνω τα πρότυπα της χρονοσειράς και τα σφάλματα της πρόβλεψης,

καθίσταται δυνατή η έγκαιρη διενέργεια διορθωτικών αλλαγών και η εξάλειψη της προκατάληψης της κάθε μεθόδου.

## **2.6 Πεδία Εφαρμογής των Προβλέψεων**

Όπως έχει αναφερθεί στην παρούσα εργασία, ο επιστημονικός κλάδος των προβλέψεων βρίσκει εφαρμογή σε πληθώρα πεδίων, που καλύπτουν ένα ευρύ φάσμα ανθρώπινων και μη δραστηριοτήτων, όπως ενδεικτικά:

- στην Οικονομία και τα Χρηματοοικονομικά, όπου γίνεται πρόβλεψη γενικών δεικτών χρηματιστηρίων και της μεταβλητότητας αυτών,
- στο Περιβάλλον και το Κλίμα, όπου θέματα που άπτονται της ενεργειακής ζήτησης, της διαχείρισης υδάτινων πόρων, της ρύπανσης κλπ αντιμετωπίζονται με τη χρήση τεχνικών προβλέψεων
- στο Κοινωνικό Περιβάλλον, για πρόβλεψη λόγου χάριν δημογραφικών δεδομένων ή των επιπέδων εγκληματικότητας
- στον Τουρισμό, για την εκτίμηση των συνολικών αφίξεων τουριστών, το ταξιδιωτικό συνάλλαγμα ή τις διανυκτερεύσεις ανά τουριστική περιοχή, που αποτελούν μεγέθη κομβικής σημασίας, ιδιαίτερα για οικονομίες που βασίζονται στην παροχή τουριστικών υπηρεσιών, όπως η Ελλάδα (25.7% - 30.9% του ΑΕΠ το 2018)

Στην επόμενη ενότητα γίνεται αναφορά στις βασικότερες στατιστικές μεθόδους πρόβλεψης.

## **2.7 Στατιστικές Μέθοδοι Πρόβλεψης**

Οι στατιστικές μέθοδοι πρόβλεψης αναφέρονται στην εφαρμογή μοντέλων χρονοσειρών ή αιτιοκρατικών μοντέλων σε μια σειρά δεδομένων, που αποσκοπεί στην αυτοματοποιημένη και συστηματική εξαγωγή προβλέψεων. Η εφαρμογή των στατιστικών μεθόδων πρόβλεψης, δύναται να γίνει σε μεγάλο πλήθος χρονοσειρών (batch forecasting) και μάλιστα με μικρές απαιτήσεις σε χρόνο και υπολογιστικό κόστος, καθώς είναι άμεσα εφαρμόσιμες και αποδεκτά ακριβείς όταν συνδυάζονται με τα κατάλληλα διαστήματα εμπιστοσύνης.

Προϋπόθεση των μεθόδων αυτών αποτελεί η συνέχιση της συμπεριφοράς της χρονοσειράς και στο μέλλον, χωρίς να λαμβάνουν συνεπώς υπόψη ειδικά γεγονότα και ενέργειες. Σημαντικό πλεονέκτημα αποτελεί η ευκολία στη χρήση και την εφαρμογή με τη βοήθεια ειδικών πληροφοριακών συστημάτων, χωρίς την απαίτηση εξειδικευμένης γνώσης. Ως μειονέκτημα, θα μπορούσε να αναφερθεί η χρονοβόρα διαδικασία της συλλογής των δεδομένων, καθώς οι στατιστικές μέθοδοι απαιτούν την ύπαρξη αρκετών ιστορικών δεδομένων.

Ακολούθως παρουσιάζονται οι σημαντικότερες στατιστικές μέθοδοι.



### 2.7.1 Αφελής Μέθοδος (Naive)

Η μέθοδος Naive αποτελεί την πιο απλή στατιστική μέθοδο πρόβλεψης, η οποία χρησιμοποιείται κυρίως ως benchmark για τις πολυπλοκότερες μεθόδους, παρά για την εξαγωγή τελικών προβλέψεων. Σύμφωνα με αυτή, η πρόβλεψη μιας χρονικής στιγμής ισούται με την παρατήρηση της αμέσως προηγούμενης περιόδου, ήτοι:

$$F(t) = Y(t-1)$$

όπου

t η χρονική περίοδος

F(t) η πρόβλεψη κατά τη χρονική περίοδο t

Y(t-1) η τιμή της χρονοσειράς κατά την περίοδο t-1

Σημαντικότερο πλεονέκτημα της συγκεκριμένης μεθόδου είναι το ιδιαίτερα χαμηλό υπολογιστικό κόστος, ωστόσο η απλοϊκή της προσέγγιση οδηγεί σε μικρότερης ακρίβειας προβλέψεις, συνεπώς στην πράξη αξιοποιείται για την πρόβλεψη λίγων περιόδων στο μέλλον. Η απλότητα και το μικρό προαναφερθέν κόστος της είναι οι παράγοντες που την καθιστούν ευρέως χρησιμοποιούμενο benchmark όπως αναφέρθηκε προηγουμένως.

### 2.7.2 Μέθοδος Κινητών Μέσων Όρων (ΚΜΟ)

Οι μέθοδοι αυτές χρησιμοποιούν τον κοινό ορισμό του μέσου όρου χρησιμοποιώντας ένα σταθερό αριθμό δεδομένων, τα οποία συμμετέχουν στον υπολογισμό του μέσου όρου με ίσα βάρη. Όπως περιγράφει και η ονομασία τους, οι κινητοί μέσοι όροι παραλείπουν την πιο παλιά παρατήρηση καθώς γίνεται διαθέσιμη μία νεότερη, διατηρώντας πάντα σταθερό τον αριθμό των πιο πρόσφατων παρατηρήσεων ο μέσος όρος των οποίων θα αποτελέσει την πρόβλεψη για την επόμενη περίοδο. Στην ουσία, κάθε νέα πρόβλεψη, προκύπτει με προσαρμογή της αμέσως προηγούμενης. Η μαθηματική διατύπωση της μεθόδου των κινητών μέσων όρων είναι η ακόλουθη:

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

Στην περίπτωση που το k επιλεγεί ίσο με τη μονάδα, ο ΚΜΟ συμπίπτει με τη μέθοδο Naive. Όσο μεγαλύτερη είναι η τιμή που λαμβάνει η παράμετρος k, τόσο περισσότερα είναι και τα δεδομένα που λαμβάνονται υπόψη και κατά συνέπεια τόσο μικρότερη σημασία δίνεται στη διακύμανση των δεδομένων, αποτελώντας την κατάλληλη πρακτική για τις περιπτώσεις που το στοιχείο της τυχαιότητας είναι έντονο στις ιστορικές παρατηρήσεις. Αντιθέτως, όταν τα ιστορικά δεδομένα παρουσιάζουν αλλαγή επιπέδου, επιλέγονται μικρότερες τιμές της παραμέτρου k ώστε να λαμβάνονται λιγότερες παρατηρήσεις κατά τον υπολογισμό του μέσου όρου, με σκοπό την καλύτερη προσαρμογή του μοντέλου σε αυτή την αλλαγή.

Τέλος στην περίπτωση όπου οι διαθέσιμες παρατηρήσεις παρουσιάζουν εποχικότητα, το  $k$  τίθεται ίσο με το μήκος αυτής, με αποτέλεσμα να εξομαλύνεται η συνιστώσα αυτή. Θα πρέπει να σημειωθεί ωστόσο, ότι το μοντέλο δεν είναι σε θέση να τη συμπεριλάβει κατά την εξαγωγή της πρόβλεψης. Οι μέθοδοι των ΚΜΟ, αδυνατούν επιπλέον να συμπεριλάβουν τη συνιστώσα της τάσης στις προβλέψεις τους. Στα δύο αυτά μειονεκτήματα οφείλεται το γεγονός ότι δε χρησιμοποιούνται ως εργαλεία πρόβλεψης, παρά μόνο για την εξομάλυνση των δεδομένων κατά τη φάση εξομάλυνσής τους, ώστε να προκύψει μία εκτίμηση της γραμμής τάσης-κύκλου.

### 2.7.3 Μέθοδοι Εκθετικής Εξομάλυνσης

Οι μέθοδοι εκθετικής εξομάλυνσης αποτελούν μία “οικογένεια” προβλεπτικών μεθόδων και αναπτύχθηκαν τη δεκαετία του 1950, ούσες έκτοτε ιδιαίτερα δημοφιλείς λόγω της ευκολίας στη χρήση τους, των μικρών απαιτήσεων σε αποθήκευση δεδομένων και του χαμηλού υπολογιστικού κόστους. Πυρήνα τους αποτελεί η προέκταση των προτύπων που εμφανίζουν τα ιστορικά δεδομένα και αφορούν ποιοτικά χαρακτηριστικά όπως τάση, εποχικότητα ή κύκλο. Μετά την εξομάλυνση των στοιχείων αυτών, στόχο αποτελεί η αφαίρεση του θορύβου. Βασική τους αρχή είναι ότι τα πιο πρόσφατα ιστορικά δεδομένα είναι αυτά τα οποία εμπεριέχουν και την περισσότερο χρήσιμη πληροφορία για την πρόβλεψη των μελλοντικών τιμών του υπό εξέταση μεγέθους και ως εκ τούτου η βαρύτητα που αποδίδεται στις τιμές αυτές είναι μεγαλύτερη και φθίνει εκθετικά όσο πιο πίσω πηγαίνουμε στα ιστορικά δεδομένα. Το τελευταίο αποτελεί το λόγο της ευρείας χρήσης τους στις περιπτώσεις βραχυπρόθεσμου σχεδιασμού, ενώ αξίζει να σημειωθεί ότι είναι περισσότερο αποδοτικές όταν εφαρμόζονται σε χρονοσειρές στάσιμες ή με μικρό βαθμό ανάπτυξης. Ενδιαφέρον παρουσιάζει το γεγονός ότι οι μέθοδοι αυτές παρουσιάζουν ικανοποιητικά ποσοστά ακρίβειας σε σχέση με πολυπλοκότερες μεθόδους, το οποίο οφείλεται στο ότι δεν επηρεάζονται από τις όποιες ιδιομορφίες στα πρότυπα που ακολουθούν τα δεδομένα, ούτε από τις ακραίες τιμές που ενδεχομένως να περιέχουν.

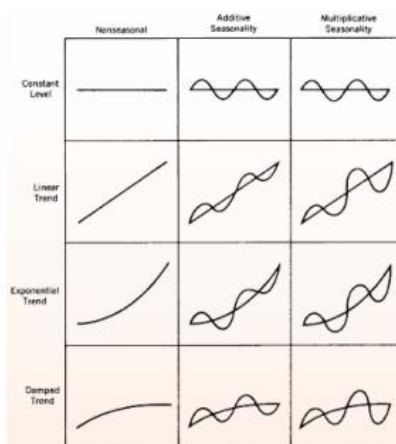
Ανάλογα με τη γραφική παράσταση των ιστορικών δεδομένων συναρτήσει του χρόνου, οι μέθοδοι εκθετικής εξομάλυνσης μπορούν να διακριθούν στις ακόλουθες κατηγορίες μοντέλων :

- *σταθερού επιπέδου (constant level)*, τα οποία υποθέτουν την απουσία τάσης στα δεδομένα, με αποτέλεσμα να θεωρούν ότι οι χρονοσειρές έχουν ένα σχετικά σταθερό μέσο όρο, οπότε η πρόβλεψη μπορεί να προκύψει προεκτείνοντας μια οριζόντια ευθεία γραμμή. Χρησιμοποιούνται ευρέως για προβλέψεις ενός βήματος (*one-step-ahead forecasting*) ή στις περιπτώσεις χρονοσειρών που έχουν έντονο θόρυβο ή τυχαιότητα
- *γραμμικής τάσης (linear trend)* που αποτελούν και την πιο δημοφιλή παραδοχή, θεωρούν ότι οι προβλέψεις για οποιαδήποτε χρονική περίοδο στο μέλλον μπορούν να προκύψουν από την προέκταση μιας ευθείας γραμμής
- *εκθετικής τάσης (exponential trend)*, τα οποία χρησιμοποιούνται για τις περιπτώσεις προβλέψεων των πωλήσεων ενός προϊόντος που βρίσκεται στην αρχή του κύκλου ζωής του, καθώς υποθέτουν ότι το ποσοστό αύξησης των πωλήσεων αυξάνεται συνεχώς. Έχει αποδειχθεί από πληθώρα εμπειρικών μελετών, ότι οι προβλέψεις που εξάγουν τα μοντέλα

εκθετικής και γραμμικής τάσης είναι υπεραισιόδοξες καθώς αυξάνεται ο ορίζοντας πρόβλεψης

- *φθίνουσας τάσης (damped trend)* που αντιμετωπίζουν το προαναφερθέν μειονέκτημα, καθώς υποθέτουν ότι το μέγεθος κατά το οποίο αυξάνονται οι τιμές της χρονοσειρές, μειώνεται βαθμιαία κάθε χρονική περίοδο.

Τα παραπάνω μοντέλα μπορούν να εφαρμοστούν και σε εποχιακές χρονοσειρές, όπως φαίνεται στα διαγράμματα που ακολουθούν, με προσθετική ή πολλαπλασιαστική εποχικότητα. Πιο συγκεκριμένα, στο προσθετικό μοντέλο εποχικότητας (additive seasonality), το εύρος των εποχικών διακυμάνσεων θεωρείται σταθερό και ανεξάρτητο από τις τιμές της ίδιας της χρονοσειράς, σε αντίθεση με το πολλαπλασιαστικό (multiplicative seasonality), στο οποίο οι τιμές είναι ανάλογες, με αποτέλεσμα καθώς η τάση αυξάνει, η εποχικές διακυμάνσεις να γίνονται μεγαλύτερες. Στην πράξη, μεταξύ των δύο προσεγγίσεων, επικρατέστερη είναι η πολλαπλασιαστική εποχικότητα.



Σχήμα 2.3 : Κατηγορίες Μεθόδων Εξομάλυνσης

### 2.7.3.1 Απλή Εκθετική Εξομάλυνση - Μοντέλο Σταθερού Επιπέδου (Simple Exponential Smoothing - SES)

Η απλή εκθετική εξομάλυνση περιγράφεται από τις ακόλουθες εξισώσεις:

$$e_t = Y_t - F_t$$

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

$$F_{t+1} = S_t$$

όπου

t η χρονική περίοδος (το t=1 αναφέρεται στην πρώτη χρονική περίοδο γι την οποία έχουμε διαθέσιμα δεδομένα)

$e_t$  το σφάλμα και αποτελεί την αποκλιση της πρόβλεψης από την πραγματική τιμή

$S_t$  το επίπεδο

$Y_t$  η πραγματική παρατήρηση τη χρονική στιγμή t

$F_t$  η πρόβλεψη τη χρονική στιγμή  $t$   
 $\alpha$  η παράμετρος που αποτελεί

Από τις παραπάνω σχέσεις είναι εμφανής η αναδρομικότητα του μοντέλου, συνεπώς είναι κρίσιμος ο ορισμός ενός αρχικού επιπέδου  $S_0$  προκειμένου να ξεκινήσει η διαδικασία υπολογισμού του μοντέλου πρόβλεψης και των ζητούμενων σημειακών προβλέψεων δεδομένου χρονικού ορίζοντα. Συνήθως για την επιλογή του αρχικού επιπέδου χρησιμοποιείται:

- ο μέσος όρος των παρατηρήσεων
- ο μέσος όρος των  $n$  πρώτων παρατηρήσεων
- η πρώτη παρατήρηση
- το σταθερό επίπεδο από το μοντέλο της απλής γραμμικής παλινδρόμησης

Η επιλογή μίας εκ των παραπάνω εναλλακτικών βασίζεται στα ποιοτικά χαρακτηριστικά της εκάστοτε χρονοσειράς. Ο λόγος που δίνεται τόσο μεγάλη έμφαση στην επιλογή ενός σωστού αρχικού επιπέδου είναι το ενδεχόμενο να μην αντιπροσωπεύει τα δεδομένα θα έχει αρνητικό αντίκτυπο στις τιμές των προβλέψεων και του συντελεστή εξομάλυνσης. Τέλος δεν πρέπει να αγνοηθεί το γεγονός ότι το αρχικό επίπεδο αποτελεί στην ουσία την πρώτη πρόβλεψη ( $F_1$ ).

Οι μαθηματικές σχέσεις που περιγράφουν τη SES, μπορούν να μετασχηματιστούν με τον κάτωθι τρόπο:

$$F_{t+1} = S_t(t) = S_{t-1} + \alpha \cdot e_t$$

$$\xrightarrow{F_t = S_{t-1}} F_{t+1} = F_t + \alpha \cdot e_t$$

$$\xrightarrow{e_t = Y_t - F_t} F_{t+1} = F_t + \alpha \cdot (Y_t - F_t)$$

$$\Rightarrow F_{t+1} = \alpha \cdot Y_t + (1 - \alpha) \cdot F_t$$

Η ανωτέρω διατύπωση αναδεικνύει το γεγονός ότι η πρόβλεψη αποτελεί ένα γραμμικό συνδυασμό της προηγούμενης παρατήρησης και της προηγούμενης πρόβλεψης, με βάρη που καθορίζονται από το συντελεστή εξομάλυνσης  $\alpha$ . Επιπλέον, οι όροι  $(1-\alpha)$  και  $\alpha$  με τους οποίους συμμετέχουν η πρόβλεψη  $F_{t-1}$  και η παρατήρηση  $Y_{t-1}$  αντίστοιχα, καθιστούν εμφανή την εξάρτηση της πρόβλεψης  $F_{t-1}$  από όλες τις προηγούμενες παρατηρήσεις με βάρη που φθίνουν εκθετικά αλλά και από την αρχική πρόβλεψη.

Η επιλογή του βέλτιστου συντελεστή εξομάλυνσης καθορίζεται από δύο αλληλοεξαρτώμενους παράγοντες: το θόρυβο και τη σταθερότητα του μέσου όρου της χρονοσειράς. Στην περίπτωση που ο θόρυβος των δεδομένων της χρονοσειράς είναι υψηλός, επιλέγεται μικρότερη τιμή του συντελεστή εξομάλυνσης ώστε να αποφευχθεί υπερβολική αντίδραση στο θόρυβο. Αντίθετα, η τιμή του  $\alpha$  πρέπει να είναι μεγάλη στην περίπτωση που ο μέσος όρος των τιμών των δεδομένων μεταβάλλεται, ώστε οι προβλέψεις να παρακολουθούν τις μεταβολές που παρουσιάζει η χρονοσειρά. Στην πράξη η επιλογή του βέλτιστου συντελεστή εξομάλυνσης γίνεται αυτόματα

από υπολογιστικά συστήματα και βασίζεται στην ελαχιστοποίηση του Μέσου Τετραγωνικού Σφάλματος (MSE) ή την υλοποίηση ενός επαναληπτικού αλγορίθμου. Οι τιμές που λαμβάνει ο συντελεστής εξομάλυνσης, κυμαίνονται στο διάστημα  $[0,1]$ . Όσο μεγαλύτερη είναι η τιμή του, τόσο μεγαλύτερη βαρύτητα δίνεται στο πιο πρόσφατο παρελθόν, αφού το σφάλμα της τελευταίας περιόδου θα διαδραματίζει σημαντικότερο ρόλο στον υπολογισμό της επόμενης σημειακής πρόβλεψης. Αντίθετα, μικρές τιμές του  $\alpha$ , συνεπάγονται αντίστοιχα μικρές μεταβολές στη νέα πρόβλεψη. Στις ακραίες περιπτώσεις, η SES είναι ταυτίζεται με τη μέθοδο Naive (για  $\alpha=1$ ), ή εξάγει προβλέψεις ίσες με το αρχικό επίπεδο (για  $\alpha=0$ ).

Τέλος χαρακτηριστικό της SES αποτελεί το γεγονός ότι στην περίπτωση που απαιτούνται περισσότερες της μίας σημειακές προβλέψεις, τότε όλες οι επόμενες τίθενται ίσες με την τελευταία υπολογισμένη πρόβλεψη (flat forecasting), καθώς όπως έχει ήδη αναφερθεί, το μοντέλο θεωρεί την ύπαρξη ενός σταθερού επιπέδου.

### 2.7.3.2 Μοντέλο Γραμμικής Τάσης (Holt Exponential Smoothing)

Προέκταση της SES που περιγράφηκε στην προηγούμενη υποενότητα αποτελεί το μοντέλο εξομάλυνσης για γραμμική τάση, που όπως υποδηλώνει η ονομασία της, είναι ικανή να διαχειρίζεται τη συνιστώσα της τάσης και προτάθηκε το 1957 από τον Holt. Οι σχέσεις που την περιγράφουν είναι οι εξής:

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

όπου,

- e δηλώνει το σφάλμα, δηλαδή την απόκλιση της πραγματικής τιμής από την πρόβλεψη,
- S το επίπεδο,
- T την τάση,
- F την πρόβλεψη,
- t τη χρονική περίοδο και m το χρονικό ορίζοντα πρόβλεψης.

Οι παράμετροι  $\alpha$  και  $\beta$  αποτελούν συντελεστές εξομάλυνσης του επιπέδου και της τάσης αντίστοιχα, ενώ λαμβάνουν τιμές στο διάστημα  $[0,1]$ .

Όμοια με την περίπτωση της απλής εκθετικής εξομάλυνσης, είναι αναγκαίο να οριστεί ένα αρχικό επίπεδο  $S_0$  καθώς επίσης και μία αρχική τάση  $T_0$ , ώστε να μπορεί να εκκινήσει η διαδικασία υπολογισμού του μοντέλου πρόβλεψης και των ζητούμενων σημειακών προβλέψεων. Οι εναλλακτικές για την επιλογή του αρχικού επιπέδου είναι οι ίδιες με αυτές που περιγράφηκαν για τη SES, ενώ για την τάση συνήθως χρησιμοποιείται:

- η διαφορά δεύτερης και πρώτης παρατήρησης ( $Y_2 - Y_1$ )
- η διαφορά  $v$ -στής και πρώτης παρατήρησης, διαιρεμένη με  $v-1$ , πχ  $(Y_{10}-Y_1)/9$

- η σταθερά της κλίσης από το μοντέλο της απλής γραμμικής παλινδρόμησης

Επισημαίνεται εκ νέου η ανάγκη ορθής επιλογής του αρχικού επιπέδου αλλά και της αρχικής τάσης, καθώς οι τιμές τους είναι καθοριστικές για τις σημειακές προβλέψεις αλλά και το μοντέλο καθαυτό. Η επιλογή βασίζεται και στην περίπτωση του μοντέλου γραμμικής τάσης βάσει των ποιοτικών χαρακτηριστικών της εκάστοτε χρονοσειράς στην οποία θα εφαρμοστεί το μοντέλο. Αναφορικά με την επιλογή των συντελεστών εξομάλυνσης, η διαδικασία είναι ίδια με αυτή της SES, με την τιμή της παραμέτρου  $\beta$  που αφορά την τάση, να είναι μικρότερη από αυτή του  $\alpha$ , καθώς η τιμή της τάσης είναι συνήθως πολύ μικρότερη από αυτή του επιπέδου. Στην περίπτωση που προστεθεί το ίδιο ποσοστό σφάλματος στο επίπεδο και στην τάση, προκύπτουν ασταθείς προβλέψεις. Τέλος, για την περίπτωση πρόβλεψης μεγάλου πλήθους χρονοσειρών (batch forecasting) συνίσταται η εφαρμογή γραμμικής παλινδρόμησης με ανεξάρτητη μεταβλητή το χρόνο.

### 2.7.3.3 Μοντέλα Μη γραμμικής Τάσης (Damped Exponential Smoothing)

Αν και ιδιαίτερα δημοφιλές, το μοντέλο της γραμμικής τάσης, τείνει να υπερεκτιμά τις προβλέψεις (θετική προκατάληψη), ιδιαίτερα σε μακροπρόθεσμους και μακροπρόθεσμους ορίζοντες. Το πρόβλημα αυτό αντιμετωπίστηκε από τους Gardner και McKenzie το 1985, οι οποίοι πρότειναν μία παραλλαγή του αρχικού μοντέλου ώστε να μπορεί να προσαρμόζεται και στις μη γραμμικές τάσεις, μέσω της προσθήκης μία επιπλέον παραμέτρου, η οποία αποκαλείται παράμετρος διόρθωσης της τάσης (trend-modification parameter) και ελέγχει το ρυθμό αύξησης των τιμών του. Η μαθηματική περιγραφή του μοντέλου μη γραμμικής τάσης είναι η κάτωθι:

$$\begin{aligned} e_t &= Y_t - F_t \\ S_t &= S_{t-1} + T_{t-1} + \alpha \cdot e_t \\ T_t &= T_{t-1} + \beta \cdot e_t \\ F_{t+m} &= S_t + \sum_{i=1}^m \varphi^i \cdot T_t \end{aligned}$$

Οι παραπάνω σχέσεις είναι πανομοιότυπες με την περίπτωση του γραμμικού μοντέλου, με μόνη διαφοροποίηση να αποτελεί ο μη γραμμικός υπολογισμός της τάσης λόγω της παραμέτρου εξομάλυνσης  $\varphi$ , η οποία εν αντιθέσει με τις παραμέτρους  $\alpha$  και  $\beta$  δύναται να λάβει τιμές μεγαλύτερες του μηδενός. Συγκεκριμένα αν:

- $\varphi=0$ : τότε προκύπτει το μοντέλο της απλής εκθετικής εξομάλυνσης, εφόσον η τάση δεν έχει καμία επίδραση στον καθορισμό των στατιστικών προβλέψεων
- $0 < \varphi < 1$  : τότε προκύπτει το μοντέλο της φθίνουσας τάσης
- $\varphi=1$ : τότε προκύπτει το μοντέλο γραμμικής τάσης, καθώς στην εξίσωση υπολογισμού της πρόβλεψης το γινόμενο της μεταβλητής χρονικού ορίζοντα  $m$  και της προηγούμενης τάση  $T_t$  λαμβάνει τη θέση του αθροίσματος
- $\varphi > 1$  : προκύπτει το μοντέλο της εκθετικής τάσης

Παρά το γεγονός ότι δεν υπάρχει κάποιο άνω όριο στην τιμή που μπορεί να λάβει το  $\phi$ , κρίνεται σημαντική η επιβολή άνω και κάτω ορίων ανάλογα με την εκάστοτε περίπτωση.

Ο προσδιορισμός των αρχικών επιπέδων γίνεται με τις διαδικασίες που περιγράφηκαν παραπάνω. Εν κατακλείδι, αξίζει να σημειωθεί ότι το μοντέλο της μη γραμμικής τάσης δίνει ικανοποιητικά αποτελέσματα και χρησιμοποιείται στις περιπτώσεις όπου είναι αδύνατη η επιλογή κάποιου συγκεκριμένου μοντέλου ή όταν απαιτείται ένα πλήρως αυτοματοποιημένο σύστημα παραγωγής προβλέψεων μεγάλου πλήθους χρονοσειρών (batch forecasting).

## 2.7.4 Μέθοδοι Παλινδρόμησης

Μία εξίσου σημαντική “οικογένεια” διαδικασιών και τεχνικών μοντελοποίησης και ανάλυσης διαφόρων μεταβλητών είναι οι μέθοδοι παλινδρόμησης, οι οποίες τυγχάνουν ευρείας αποδοχής στον κλάδο των Προβλέψεων. Η παλινδρόμηση αποσκοπεί στην εύρεση των συσχετίσεων μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών, ενώ πέρα από τη φύση της ως προβλεπτικό εργαλείο, χρησιμοποιείται και ως βοήθημα σε άλλες μεθόδους.

### 2.7.4.1 Απλή Γραμμική Παλινδρόμηση

Η λογική της απλής γραμμικής γραμμικής παλινδρόμησης ή μεθόδου των ελαχίστων τετραγώνων, έγκειται στην ελαχιστοποίηση της απόστασης των πραγματικών παρατηρήσεων  $Y$  από τη βέλτιστη γραμμή παλινδρόμησης. Η ευθεία αυτή προκύπτει από την ελαχιστοποίηση του αθροίσματος των σφαλμάτων, όπως αυτό φαίνεται στις εξισώσεις που ακολουθούν.

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

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

Η μέθοδος της απλής παλινδρόμησης προϋποθέτει την ύπαρξη εξάρτησης της υπό πρόβλεψη μεταβλητής με από την τιμή ή την πρόβλεψη κάποιας άλλης τιμής. Για την εξακρίβωση της προαναφερθείσας ύπαρξης χρησιμοποιείται ο συντελεστής γραμμικής συσχέτισης  $r$ . Ο συντελεστής αυτός λαμβάνει τιμές στο διάστημα  $[-1,1]$ . Στην περίπτωση που η τιμή του είναι θετική, οι δύο μεταβλητές θεωρούνται θετικά συσχετισμένες, ήτοι η μεταβολή της μίας προκαλεί ανάλογη και προς την ίδια κατεύθυνση μεταβολή στην άλλη. Το αντίστροφο ισχύει για τις αρνητικές τιμές, όπου οι μεταβλητές είναι αρνητικά συσχετισμένες, με αποτέλεσμα ενδεχόμενη αύξηση στη μία μεταβλητή να επιφέρει αναλογική μείωση στην άλλη. Όταν ο συντελεστής αυτοσυσχέτισης είναι μηδενικός, οι δύο μεταβλητές θεωρούνται πλήρως ασυσχέτιστες. Ο μαθηματικός τύπος της συσχέτισης μεταξύ δύο μεταβλητών είναι ο ακόλουθος:

Προέκταση του συντελεστή αυτοσυσχέτισης αποτελεί ο συντελεστής  $R^2$  (coefficient of determination), που ισούται με το τετράγωνο της τιμής του συντελεστή αυτοσυσχέτισης,

συνεπώς λαμβάνει τιμές στο διάστημα  $[0,1]$ , ενώ εκφράζει το λόγο της ερμεινευθείσας διακύμανσης της εξαρτημένης μεταβλητής προς τη συνολική της διακύμανση.

Σκοπός της απλής γραμμικής παλινδρόμησης είναι η έκφραση της σχέσης μεταξύ της εξαρτημένης μεταβλητής (υπό πρόβλεψη μέγεθος) και της ανεξάρτητης μεταβλητής με την εξίσωση μιας ευθείας γραμμής της μορφής:

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

Οι συντελεστές  $\alpha$  και  $\beta$  υπολογίζονται με χρήση της μεθόδου ελαχίστων τετραγώνων ως εξής:

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

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

όπου  $\bar{X}$ ,  $\bar{Y}$  οι μέσες τιμές των διανυσμάτων  $X$  και  $Y$ .

Τέλος αξίζει να σημειωθεί το γεγονός ότι η μέθοδος παλινδρόμησης γίνεται να μετασχηματιστεί σε μοντέλο αν ληφθεί υπόψη το ότι ένα μέρος της πραγματικής διαδικασίας δεν μπορεί να ερμηνευθεί από την παλινδρόμηση, συνεπώς εισάγεται ο όρος  $e$  που δηλώνει τις αποκλίσεις αυτές και θεωρείται κανονικά κατανομημένος γύρω από το μηδέν. Ως εκ τούτου, η μαθηματική διατύπωση μετασχηματίζεται ως  $Y_i = \alpha + \beta \cdot X_i + e$ . Στην πράξη βέβαια χρησιμοποιείται και πάλι η σχέση  $Y_i = \alpha + \beta \cdot X_i$  σύμφωνα με τα όσα ειπώθηκαν παραπάνω.

#### **2.7.4.2 Πολλαπλή Γραμμική Παλινδρόμηση**

Στις πραγματικές εφαρμογές, εντοπίζονται περιπτώσεις που απαιτούνται περισσότερες της μίας ανεξάρτητες μεταβλητές, επομένως η απλή γραμμική παλινδρόμηση χρίζει επέκτασης. Έτσι, η μέθοδος της γραμμικής παλινδρόμησης αναδιατυπώνεται ως εξής:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n + e$$

Για τον προσδιορισμό των συντελεστών  $b_0, b_1, b_2, \dots, b_n$  οι οποίοι αποτελούν σταθερές παραμέτρους, απαιτείται ο υπολογισμός των ριζών των μερικών παραγώγων της συνάρτησης σφάλματος για τον καθένα από αυτούς. Στην εν λόγω περίπτωση εξετάζεται και η συσχέτιση της πραγματικής τιμής της μεταβλητής  $Y$  και της υπολογισμένης τιμής  $\hat{Y}$  μέσω του σφάλματος και διατυπώνεται πλέον ως εξής

$$e_i = Y_i - \hat{Y}_i$$



το τετράγωνο της οποίας αποτελεί τον συντελεστή απόφασης (coefficient of determination) και έχει τις ίδιες ιδιότητες με όσα περιγράφηκαν στην περίπτωση της απλής γραμμικής παλινδρόμησης.

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - b_0 - b_1X_1 - b_2X_2 - \dots - b_nX_n)^2$$

Σύμφωνα με τους Μακρυδάκη, Wheelright και Hyndman (1998), κάθε φορά που εφαρμόζεται στην πράξη η πολλαπλή παλινδρόμηση, γίνονται οι εξής τέσσερις βασικές υποθέσεις περί:

- γραμμικής σχέσης ανάμεσα στην εξαρτημένη και τις ανεξάρτητες μεταβλητές. Ακόμα και στην περίπτωση που κάτι τέτοιο δεν ισχύει, γίνεται κατάλληλος μετασχηματισμός των αρχικών μεταβλητών σε νέες που να εμφανίζουν γραμμική σχέση
- σταθερής διακύμανσης των σφαλμάτων παλινδρόμησης (ομοσκεδαστικότητα), η οποία δηλώνει ότι όλα τα σφάλματα πρόβλεψης θα πρέπει να είναι σταθερά καθ'όλο το εύρος των παρατηρήσεων
- ανεξαρτησίας των σφαλμάτων, που συνεπάγεται ότι η τιμή των σφαλμάτων προσαρμογής σε κάθε παρατήρηση των πραγματικών δεδομένων είναι ανεξάρτητη από τις τιμές των προηγούμενων ή των επόμενων
- κανονικής διασποράς των σφαλμάτων προσαρμογής όταν αυτά παρασταθούν γραφικά

### 2.7.5 Μέθοδος Theta

Η μέθοδος Theta αποτελεί μία μονοδιάστατη μέθοδο πρόβλεψης που προτάθηκε από τους Ασημακόπουλο και Νικολόπουλο το 2000 και βασίζεται στη μεταβολή των τοπικών καμπυλοτήτων μιας χρονοσειράς μέσω της παραμέτρου  $\theta$  (theta), η οποία εφαρμόζεται πολλαπλασιαστικά στις διαφορές δεύτερης τάξης των δεδομένων. Η αρχική χρονοσειρά αναλύεται σε δύο ή περισσότερες γραμμές Theta, οι οποίες προσεγγίζουν καλύτερα τη μακροπρόθεσμη συμπεριφορά, ήτοι την τάση, της χρονοσειράς ή τονίζουν τα βραχυπρόθεσμα χαρακτηριστικά της για τιμές της παραμέτρου  $\theta$  μικρότερες ή μεγαλύτερες της μονάδας αντίστοιχα. Κάθε τέτοια γραμμή προεκτείνεται ξεχωριστά στο μέλλον με τη χρήση κάποιας άλλης μεθόδου (συνήθως της απλής εκθετικής εξομάλυνσης), με την τελική πρόβλεψη της μεθόδου Theta να προκύπτει από τον συνδυασμό των επιμέρους προβλέψεων. Η μέθοδος αυτή θα μπορούσε να παρομοιαστεί με ένα μεγεθυντικό φακό, μέσα από τον οποίο οι διακυμάνσεις της χρονοσειράς μεγεθύνονται ή μικραίνουν. Στην περίπτωση που  $\theta=0$ , η χρονοσειρά θα ισοδυναμεί με την ευθεία της απλής γραμμικής παλινδρόμησης (LRL), ενώ για  $\theta=-1$  προκύπτει η συμμετρική της αρχικής χρονοσειράς ως προς την ευθεία της LRL.

Οποιαδήποτε γραμμή Theta υπολογίζεται βάσει της σχέσης

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

όπου

$$Y_t'' = Y_t - 2 \cdot Y_{t-1} + Y_{t-2}$$

με την τιμή της παραμέτρου  $\theta$  να είναι αντιστρόφως ανάλογη του βαθμού μείωσης/αύξησης των καμπυλοτήτων της χρονοσειράς, δηλαδή όσο μικρότερη είναι η τιμή της, τόσο περισσότερο μειώνονται οι καμπύλες της χρονοσειράς. Εν γένει, η διόγκωση των καμπυλοτήτων αποσκοπεί στην ενίσχυση της βραχυπρόθεσμης συμπεριφοράς της χρονοσειράς, ενώ στην αντίθετη περίπτωση ενισχύεται η μακροπρόθεσμη συμπεριφορά της.

Η προαναφερθείσα δυνατότητα διαφορετικών προσεγγίσεων της αρχικής χρονοσειράς μπορεί να αξιοποιηθεί αξιοποιώντας διαφορετικό συνδυασμό γραμμών Theta για διαφορετικό ορίζοντα πρόβλεψης.

Η απλούστερη μορφή της (κλασική μέθοδος Theta) που αποτέλεσε και τη νικητήρια υποβολή στο διαγωνισμό προβλέψεων M3 (Makridakis et al., 2000), αποσυνθέτει την αρχική χρονοσειρά σε δύο γραμμές Theta με τις τιμές της παραμέτρου να είναι  $\theta=0$  και  $\theta=2$ , δηλαδή η τελική πρόβλεψη προκύπτει από το συνδυασμό της LRL και του διπλασιασμού των τοπικών καμπυλοτήτων της αρχικής χρονοσειράς:

$$\text{Theta\_Line}(2) = 2 \cdot Y - \text{LRL}$$

$$Y_t = \frac{1}{2} \cdot (Y_t^{\theta=0} + Y_t^{\theta=2})$$

Η γραμμή Theta  $Y_t^{\theta=0}$  συνήθως προεκτείνεται με τη χρήση της απλής γραμμικής παλινδρόμησης, ενώ η  $Y_t^{\theta=2}$  μέσω της εκθετικής εξομάλυνσης.

Το μεθοδολογικό πλαίσιο της κλασικής μεθόδου Theta μπορεί να περιγραφεί με τα ακόλουθα βήματα:

1. Έλεγχος εποχικότητας
2. Αποεποχικοποίηση με τη χρήση της κλασικής μεθόδου πολλαπλασιαστικής αποσύνθεσης
3. Αποσύνθεση στις δύο γραμμές Theta για  $\theta = 0$  (LRL) και  $\theta=2$
4. Πρόβλεψη μέσω της προέκτασης της LRL και με τη χρήση της SES για την προέκταση της γραμμής Theta με παράμετρο  $\theta=2$
5. Συνδυασμός των δύο προβλέψεων με ίσα βάρη
6. Εποχικοποίηση με τη χρήση των δεικτών που υπολογίστηκαν στο ΒΗΜΑ 2 (για την περίπτωση εποχικών χρονοσειρών)

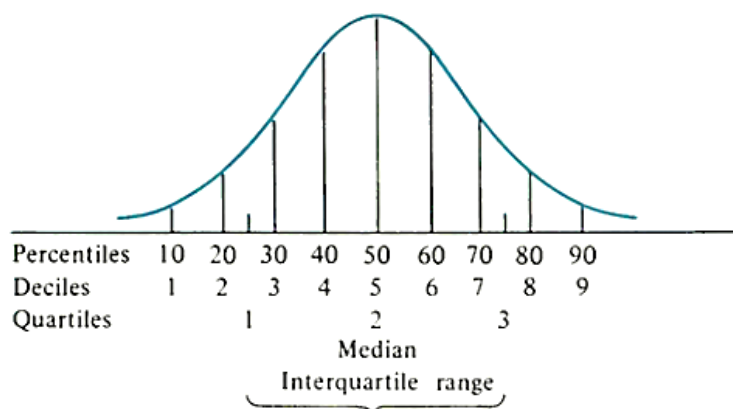
Δεδομένων όλων των προαναφερθέντων καθίσταται σαφές το συγκριτικό πλεονέκτημα της μεθόδου Theta, καθώς μέσω της αποσύνθεσης των αρχικών δεδομένων, αξιοποιείται η πληροφορία που περιέχεται σε αυτά, πληροφορία η οποία χάνεται ή δεν αναγνωρίζεται από τα υπόλοιπα μοντέλα που εφαρμόζονται απευθείας στα αρχικά δεδομένα. Χαρακτηριστικά στην περίπτωση της κλασικής μεθόδου Theta, αξιοποιείται όλη την πληροφορία για την μακροπρόθεσμη τάση της χρονοσειράς με τη γραμμή  $Y_{-}$ , γεγονός που θα ήταν αδύνατο με την εφαρμογή της LRL στα αρχικά δεδομένα, καθώς τότε η μέθοδος θα προσαρμοζόταν στις πιο πρόσφατες τάσεις της χρονοσειράς, ενώ επιπλέον είναι προφανές ότι θα αγνοούσαν όλη η χρήσιμη πληροφορία που σχετίζεται με τις βραχυπρόθεσμες διακυμάνσεις, η οποία πλέον λαμβάνεται μέσω της γραμμής  $Y_{+}$ .

## 2.8 Πιθανοτικές Προβλέψεις

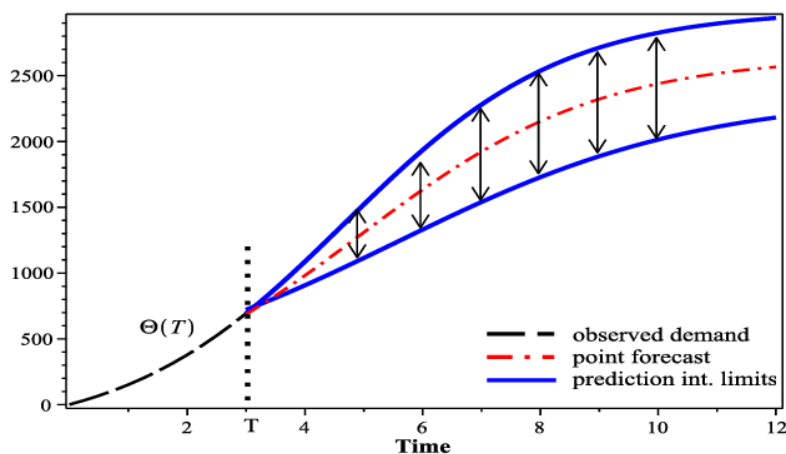
Η παραδοσιακή έννοια της πρόβλεψης αφορά τις σημειακές προβλέψεις, όπου στόχο αποτελεί η πρόβλεψη μίας αναμενόμενης τιμής στο μέλλον  $\hat{y} = E[y|x]$  με μεθόδους όπως αυτές που περιγράφηκαν προηγουμένως. Υπάρχουν ωστόσο και ορισμένες κατηγορίες προβλέψεων, οι οποίες λαμβάνουν υπόψη και τον παράγοντα της αβεβαιότητας, όπως:

- Πιθανοτικές προβλέψεις: προβλέπουν την πυκνότητα της μελλοντικής κατανομής
- Risk Indexes: οι οποίες βρίσκουν το αναμενόμενο επίπεδο του προβλεπτικού σφάλματος
- Scenario Forecasting: οι οποίες δημιουργούν διάφορα σενάρια σημειακών προβλέψεων για την ίδια χρονική περίοδο του μέλλοντος με σκοπό να βρουν τους συσχετισμούς χρονικών διαστημάτων

Για τις ανάγκες αυτών των κατηγοριών προβλέψεων εισάγονται η έννοια των ποσοστημορίων (quantiles) τα οποία όπως φαίνεται στο Σχήμα 2.4 διαιρούν την κατανομή σε διαστήματα, όπως επίσης και η έννοια των διαστημάτων πρόβλεψης (prediction intervals) που αφορούν το ανώτερο και το κατώτερο όριο της διακύμανσης των προβλέψεων σε μία χρονική περίοδο του ορίζοντα πρόβλεψης για δεδομένο διάστημα εμπιστοσύνης, όπως παρουσιάζεται με τις μπλε γραμμές στο Σχήμα 2.5.



Σχήμα 2.4 : Διαιρέση της κατανομής σε ποσοστημόρια



Σχήμα 2.5 : Μορφή μελλοντικών προβλέψεων με απεικόνιση των prediction intervals (μπλε γραμμές). Με μαύρο χρώμα απεικονίζονται οι τιμές των ιστορικών δεδομένων και με κόκκινο οι σημειακές προβλέψεις για τα βήματα του χρονικού οριζόντα

Δεδομένης της αθροιστικής πυκνότητας πιθανότητας  $F$  μιας τυχαίας μεταβλητής  $x$ , το quantile  $q_\tau$  με τιμή στο διάστημα  $[0,1]$ , ορίζεται μοναδικά ως η τιμή του  $x$  τέτοια ώστε  $P(P_t < x) = \tau$  ή ισοδύναμα  $q_\tau^t = F_t^{-1}(\tau)$ .

Αντίστοιχα, το διάστημα πρόβλεψης (prediction interval)  $I_{t+k}$  που παράγεται τη χρονική στιγμή  $t$  για τη χρονική στιγμή  $t+k$ , ορίζεται από το ανώτερο και το κατώτερο όριο που δίνονται από τα ακραία quantiles, δηλαδή  $I_{t+k}^\beta = [\hat{q}_{t+k}^{\alpha_l}, \hat{q}_{t+k}^{\alpha_u}]$ .

Η έννοια της πιθανοτικής πρόβλεψης είναι το αντίθετο της ντετερμινιστικής, κατά την οποία προβλέπεται μία μόνο μελλοντική τιμή του υπό εξέταση μεγέθους, βάσει των ιστορικών δεδομένων που είναι διαθέσιμα. Οι πιθανοτικές προβλέψεις περιλαμβάνουν ένα σύνολο πιθανοτήτων που σχετίζονται με όλες τις πιθανές εκβάσεις της μελλοντικής εξέλιξης του υπό πρόβλεψη μεγέθους, αντί να προσδιορίζουν ένα μόνο αποτέλεσμα ως την πρόβλεψη. Σκοπός τους είναι να συμπεριλάβουν την αβεβαιότητα κατά την πρόβλεψη των μελλοντικών καταστάσεων. Για το λόγο αυτό, η χρήση τους είναι ιδιαίτερα σημαντική στις περιπτώσεις όπου η αβεβαιότητα είναι αδύνατο να εξαλειφθεί ή να περιοριστεί, γεγονός το οποίο αντικατοπτρίζει σχεδόν όλες τις πραγματικές εφαρμογές. Χαρακτηριστικός κλάδος που απαιτεί τη χρήση των πιθανοτικών προβλέψεων είναι ο κλάδος της εφοδιαστικής αλυσίδας, λόγω της αβεβαιότητας της μελλοντικής ζήτησης των αγαθών.

Οι πιθανοτικές προβλέψεις αναδεικνύουν την προτεραιοποίηση των αποφάσεων βάσει των αναμενόμενων, πλην αβέβαιων, αποτελεσμάτων τους. Πληθώρα στατιστικών μοντέλων ή μοντέλων μηχανικής μάθησης (όπως αυτό που χρησιμοποιήθηκε για τους σκοπούς της παρούσας διπλωματικής εργασίας) έχουν τη δυνατότητα να παράγουν πιθανοτικές προβλέψεις. Αντί να παράγουν μία τιμή ως το «καλύτερο» αποτέλεσμα, η πιθανοτική πρόβλεψη αποτελείται, όπως

υποδηλώνει και το όνομά τους, από πιθανότητες που ανατίθενται σε κάθε πιθανό αποτέλεσμα. Στην ουσία όλα τα μελλοντικά συμβάντα παραμένουν πιθανά, χωρίς ωστόσο να είναι ισοπίθανα τα ενδεχόμενα να λάβουν πράγματι χώρα. Έτσι η πρόβλεψη για παράδειγμα της τιμής  $X$  στο ποσοστοτημώριο  $Q$ , υποδηλώνει ότι υπάρχει  $Q\%$  πιθανότητα η πραγματική τιμή του μεγέθους να είναι ίση με  $X$ . Το προαναφερθέν αποτελεί και την ειδοποιό διαφορά τους με τις ντετερμινιστικές προβλέψεις και συνάμα το πλεονέκτημά τους έναντι αυτών, καθώς η πρόβλεψη περιλαμβάνει περισσότερη πληροφορία, δεδομένου του γεγονότος ότι εμπεριέχουν την έννοια των εναλλακτικών. Μία ντετερμινιστική πρόβλεψη είναι δυνατόν να εξαχθεί μέσω της πιθανοτικής, αν ληφθεί το median της τελευταίας, το αντίστροφο ωστόσο, δηλαδή η εξαγωγή της πιθανοτικής από τη ντετερμινιστική πρόβλεψη είναι αδύνατο.

Μολονότι οι πιθανοτικές προβλέψεις θεωρούνται ανώτερες από τις σημειακές, η υιοθέτησή τους έγινε αρκετά αργότερα, αφού η αναγνώρισή τους αυξήθηκε σημαντικά και σταθερά την τελευταία δεκαετία. Κυριότερο παράγοντα που αιτιολογεί το προαναφερθέν, αποτελεί η ανάγκη σημαντικά περισσότερων υπολογιστικών πόρων και ειδικών εργαλείων λογισμικού, τα οποία μέχρι πρότινος δεν ήταν διαθέσιμα.

Για την εξαγωγή πιθανοτικών προβλέψεων είναι δυνατόν να χρησιμοποιηθούν στατιστικά μοντέλα όπως:

- Μοντέλα διακριτής ανάλυσης (discriminant analysis)
- Μοντέλα πολλαπλής γραμμικής παλινδρόμησης, στα οποία οι πιθανοτικές προβλέψεις εξάγονται με δύο τρόπους:
  - Μέσω της χρήσης των prediction intervals, τεχνική κατά την οποία το διάστημα μεταξύ του ανώτερου και του κατώτερου ορίου ολοκληρώνεται και έτσι υπολογίζονται οι πιθανότητες της εκάστοτε εναλλακτικής
  - Μέσω της δημιουργίας ενός πίνακα ενδεχομένων, με την κατανομή των διαφορετικών εναλλακτικών να ορίζεται βάσει της mean πρόβλεψης

αλλά και ML μοντέλα τα οποία εκπαιδεύονται για να βγάλουν βέλτιστη πρόβλεψη για ένα μόνο quantile, στην περίπτωση επί παραδείγματι που αναζητείται η ακριβέστερη πρόβλεψη να συμβεί ένα ενδεχόμενο με πιθανότητα 90%.

Να σημειωθεί ότι το εργαλείο GluonTS που χρησιμοποιήθηκε στο πειραματικό σκέλος της παρούσας διπλωματικής εργασίας, δίνει τη δυνατότητα όπως θα αναφερθεί στο αντίστοιχο κεφάλαιο, της πρόβλεψης για όλο το φάσμα της κατανομής, χωρίς να περιορίζονται τα μοντέλα του σε ένα και μόνο quantile όπως γίνεται στα αντίστοιχα συμβατικά μοντέλα ML.

Καθίσταται σαφές ότι οι πιθανοτικές προβλέψεις είναι ζωτικής σημασίας για τις επιχειρήσεις και τους οργανισμούς, καθώς επιτρέπουν την ποσοτικοποίηση της έννοιας του ρίσκου κατά τη διάρκεια εξαγωγής προβλέψεων. Ως εκ τούτου συνεπικουρούν στη λήψη ορθολογότερων αποφάσεων στις οποίες πέρα από την αμιγή τιμή της πρόβλεψης, καθίσταται δυνατή η συναξιολόγηση της πιθανότητας η πρόβλεψη να αποτελέσει εν τέλει την πραγματική τιμή.

## 2.9 Επιπλέον Κατηγορίες Μεθόδων Προβλέψεων

Μολονότι αντικείμενο της παρούσας διπλωματικής εργασίας αποτελεί η εξαγωγή πιθανοτικών προβλέψεων με τη χρήση νευρωνικών δικτύων, κρίνεται χρήσιμο για λόγους πληρότητας να αναφερθούν συνοπτικά ορισμένα στοιχεία που αφορούν τις μεθόδους πρόβλεψης δεδομένων διακοπτόμενης ζήτησης, τις κριτικές προβλέψεις αλλά και τις προβλέψεις προϋπολογισμού, καθώς αποτελούν σημαντικά πεδία εφαρμογών των τεχνικών προβλέψεων.

### 2.9.1 Μέθοδοι Διακοπτόμενης Ζήτησης

Χαρακτηριστικό της διακοπτόμενης ζήτησης αποτελεί η σποραδική της εμφάνιση, καθώς μεσολαβούν περίοδοι όπου η τιμή της είναι μηδενική, ενώ εντοπίζεται συνήθως μεγάλη διακύμανση στις περιπτώσεις που εμφανίζεται. Τα χαρακτηριστικά αυτά είναι που δυσχεραίνουν το προβλεπτικό έργο και καθιστούν αναγκαία τη διαφορετική αντιμετώπιση των δεδομένων διακοπτόμενης ζήτησης. Τέτοιου είδους δεδομένα παρουσιάζονται σε μονάδες διαχείρισης αποθεμάτων και ανταλλακτικών. Η βελτίωση στο πεδίο των προβλέψεων των δεδομένων διακοπτόμενης ζήτησης είναι πολύ σημαντική, καθώς οδηγεί άμεσα σε μείωση του κόστους των επιχειρήσεων, μέσω των μικρότερων δεσμευμένων κεφαλαίων υπό μορφή αποθεμάτων ή ανταλλακτικών (τα οποία αρκετές φορές παραμένουν αναξιοποίητα και συν τω χρόνω απαξιώνονται) αλλά και τη μείωση των αναγκών για αποθηκευτικούς χώρους που αυξάνει έτι περαιτέρω το όφελος που έχουν οι επιχειρήσεις.

Η ανάγκη εύρεσης διαφορετικών εργαλείων προέκυψε λόγω της αρχικής χρήσης των μεθόδων εκθετικής εξομάλυνσης. Όπως είναι αναμενόμενο, τα σφάλματα που προκύπτουν από την εφαρμογή τους είναι μεγάλα, καθώς λαμβάνοντας υπόψη τη μεγάλη βαρύτητα που αποδίδουν στις πιο πρόσφατες τιμές, οι εκτιμήσεις που εξάγουν είναι μέγιστες μετά την εμφάνιση ζήτησης και ελάχιστες ακριβώς πριν από περιόδους μη μηδενικής ζήτησης.

Οι πιο γνωστές μέθοδοι διακοπτόμενης ζήτησης είναι οι:

- **Croston**, η οποία χωρίζει την εκάστοτε χρονοσειρά σε δύο επιμέρους, με τη μία να αποτελείται από τα χρονικά διαστήματα μεταξύ των μη μηδενικών ζητήσεων (intervals) και την άλλη από το πλήθος των μη μηδενικών ζητήσεων (demands). Στη συνέχεια προεκτείνει τις δύο επιμέρους χρονοσειρές με τη χρήση της SES και με συνηθέστερη τιμή της παραμέτρου εξομάλυνσης  $\alpha=0,05$ . Η τελική πρόβλεψη προκύπτει μέσω της εύρεσης του ηθλικού των δύο ανεξάρτητων προβλέψεων ως εξής:

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

- **SBA (Syntetos & Boylan Approximation)**, η οποία περιορίζει τη θετική προκατάληψη που εμφανίζει η Croston και υπολογίζεται με παρόμοιο τρόπο με αυτή μέσω της σχέσης:

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

όπου  $\alpha$  η παράμετρος της μεθόδου SES

- **ADIDA (Aggregate-Disaggregate Intermittent Demand Approach)**, η οποία προκειμένου να αποφύγει την ασυνέχεια των δεδομένων διακοπτόμενης ζήτησης, τα συναθροίζει σε περιόδους μικρότερης συχνότητας και εν συνεχεία διασπά τις προβλέψεις ώστε να υπολογιστούν οι τελικές προβλέψεις χρονικής συχνότητας ίσης με αυτή του αρχικού επιπέδου. Οι προβλέψεις στο συναθροισμένο επίπεδο εξάγονται με την εφαρμογή μεθόδων για συνεχή δεδομένα, ενώ η διάσπασή τους γίνεται βάσει βαρών. Οι συγγραφείς της μεθόδου ADIDA έχουν προτείνει τις εξής εμπειρικές μεθόδους:
  - *Equal Weights*: απλός ισοβαρής διαχωρισμός που ενδείκνυται για χρονοσειρές με μεγάλη ταχύτητα και χωρίς εποχική συμπεριφορά
  - *Previous Weights*: εφαρμογή των βαρών που έχουν οι  $m$  προηγούμενες παρατηρήσεις, όπου  $m$  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, με  $k \cdot m$  να ισούται με το σύνολο των διαθέσιμων παρατηρήσεων και  $m$  να ισούται με το επίπεδο συνάθροισης. Η προσέγγιση αυτή ενδείκνυται για περιπτώσεις όπου υπάρχει έντονο το στοιχείο της εποχικότητας

## 2.9.2 Κριτικές Προβλέψεις

Η έννοια της πρόβλεψης δεν ισοδυναμεί με προφητεία αλλά με εκτίμηση της μελλοντικής κατάστασης ενός μεγέθους βάσει της υπόθεσης ότι το πρότυπο που φαίνεται να ακολουθούν τα διαθέσιμα δεδομένα του, θα συνεχιστεί και στις μελλοντικές περιόδους, χωρίς να υφίστανται σημαντικές αλλαγές. Ως εκ τούτου, τα στατιστικά μοντέλα δε λαμβάνουν υπόψη ειδικά γεγονότα και ενέργειες, καθώς επίσης αγνοούν τους στρατηγικούς στόχους που θέτει η εκάστοτε επιχείρηση. Απότοκο του προαναφερθέντος αποτελεί η παρουσία σημαντικών σφαλμάτων σε προβλέψεις οι οποίες βασίζονται αποκλειστικά στη στατιστική ανάλυση.

Για την ενσωμάτωση στην τελική πρόβλεψη εξωγενών παραγόντων οι οποίοι επηρεάζουν την εξέλιξη του υπό πρόβλεψη μεγέθους, απαιτείται η εμπλοκή της ανθρώπινης κρίσης και εμπειρίας. Αναγνωρίζοντας στο μεγαλύτερο δυνατό βαθμό τις όποιες μεροληψίες και τους περιορισμούς έχει η κριτική πρόβλεψη, τα αποτελέσματα των στατιστικών μεθόδων μπορούν να προσαρμοστούν κατάλληλα ώστε να γίνει καλύτερη προσέγγιση της μελλοντικής τιμής του υπό εξέταση μεγέθους, καθώς η μαθηματική ορθότητα μπορεί να συνδυαστεί με την ανθρώπινη διαίσθηση και εμπειρική γνώση.

### **2.9.3 Προβλέψεις Προϋπολογισμού**

Οι προβλέψεις στόχου ή προϋπολογισμού αφορούν μία επιθυμητή μελλοντική κατάσταση της επιχείρησης με σκοπό τον καθορισμό των στόχων και της πολιτικής ανάπτυξης της. Οι προβλέψεις αυτές βασίζονται στην εκτίμηση του ρυθμού ανάπτυξης μέσω της προέκτασης του προτύπου που εμφανίζουν τα διαθέσιμα ιστορικά δεδομένα. Η στατιστική εκτίμηση τροποποιείται από τον αντίστοιχο αρμόδιο της εκάστοτε επιχείρησης, με σκοπό να αντικατοπτριστεί με μεγαλύτερη ακρίβεια η επιθυμητή κατάσταση. Γίνεται σαφές επομένως ότι τέτοιου είδους προβλέψεις αποτελούν περισσότερο ένα “wish-future-status” παρά τη πραγματικά προβλεπόμενη πορεία της χρονοσειράς, λόγω της μεροληψίας και της αισιοδοξίας του αποφασίζοντα, με αποτέλεσμα να εμφανίζουν σημαντικά σφάλματα.

### **2.10 Δείκτες Ακρίβειας**

Η βασική στατιστική ανάλυση απαιτεί μόνο τις πραγματικές τιμές των παρατηρήσεων μιας χρονοσειράς. Ωστόσο οι στατιστικοί δείκτες ακριβείας απαιτούν μία σειρά πρόβλεψης, σειρά η οποία προκύπτει από ένα μοντέλο προβλέψεων. Η έννοια του σφάλματος αποκτά επομένως κρίσιμη σημασία και ορίζεται ως η διαφορά της πραγματικής τιμής  $Y$  από την προβλεφθείσα τιμή  $F$  τη χρονική περίοδο  $t$ , δηλαδή:

$$e(t) = Y(t) - F(t)$$

Ένα στατιστικό μοντέλο πρόβλεψης υπολογίζει προβλέψεις για τις τιμές των ιστορικών δεδομένων που αποκρύπτονται για λόγους αξιολόγησης και εκπαίδευσης (insample error) και για το ζητούμενο χρονικό ορίζοντα πρόβλεψης, με το out of sample σφάλμα να υπολογίζεται όταν γίνουν διαθέσιμες οι πραγματικές τιμές του υπό πρόβλεψη μεγέθους, με αποτέλεσμα να αξιολογείται η επίδοση του μοντέλου σε πραγματικές συνθήκες.

Ακολούθως θα παρουσιαστούν οι στατιστικοί δείκτες ακριβείας που χρησιμοποιήθηκαν στο πλαίσιο της παρούσας διπλωματικής με παράλληλη αναφορά στα κυριότερα χαρακτηριστικά του καθενός.



### 2.10.1 Μέσο Σφάλμα (*Mean Error - ME*)

$$ME = \frac{1}{n} \cdot \sum_{t=1}^n (Y(t) - F(t))$$

Το μέσο σφάλμα αποτελεί έναν δείκτη που προσδιορίζει την προκατάληψη (bias) του μοντέλου, καθότι όπως φαίνεται και από την παραπάνω μαθηματική διατύπωση, προκύπτει ως ο προσημασμένος μέσος όρος των σφαλμάτων. Ως εκ τούτου, οι θετικές τιμές του δηλώνουν απαισιοδοξία (εν γένει προβλέπονται μικρότερες τιμές από τις πραγματικές), ενώ οι αρνητικές του αισιοδοξία. Παράλληλα, το μέσο σφάλμα είναι ενδεικτικό και της συστηματικότητας του σφάλματος, δεδομένου του γεγονότος ότι όσο η τιμή του προσεγγίζει το μηδέν, τόσο λιγότερο συστηματικά είναι τα σφάλματα.

### 2.10.2 Μέσο Απόλυτο Σφάλμα (*Mean Absolute Error - MAE*)

$$MAE = \frac{1}{n} \cdot \sum_{t=1}^n |Y(t) - F(t)|$$

Μία παραλλαγή του Μέσου Σφάλματος, αποτελεί το Μέσο Απόλυτο Σφάλμα, το οποίο εκφράζει την ακρίβεια της προβλεπόμενων έναντι των των πραγματικών τιμών, χωρίς να εξετάζεται η μεροληψία του μοντέλου (κατεύθυνση της πρόβλεψης). Η τιμή του δείκτη είναι αντιστρόφως ανάλογη της ακρίβειας της μεθόδου, με τη μικρότερη τιμή να υποδεικνύει μεγαλύτερη ακρίβεια, ενώ διατηρεί τις μονάδες μέτρησης της πραγματικής τιμής.

### 2.10.3 Μέσο Τετραγωνικό Σφάλμα (*Mean Squared Error - MSE*)

$$MSE = \frac{1}{n} \cdot \sum_{t=1}^n (Y(t) - F(t))^2$$

Κυριότερο χαρακτηριστικό του MSE, το οποίο αποτελεί και αυτό ένα μέτρο της ακρίβειας της πρόβλεψης σε σχέση με την πραγματική τιμή, είναι το γεγονός δίνεται μεγαλύτερη βαρύτητα στα μεγαλύτερα σφάλματα, όπως προκύπτει και από την ανωτέρω μαθηματική σχέση στην οποία φαίνεται ο τετραγωνισμός του σφάλματος

### 2.10.4 Ρίζα Μέσου Τετραγωνικού Σφάλματος (*Root Mean Squared Error - RMSE*)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \cdot \sum_{t=1}^n (Y(t) - F(t))^2}$$

Συνέχεια του MSE αποτελεί το RMSE το οποίο προκύπτει ως η τετραγωνική ρίζα του πρώτου, έχοντας τις ίδιες ιδιότητες με αυτό, ώντας όμως εκφρασμένο στις μονάδες της αρχικής χρονοσειράς.

### **2.10.5 Ρίζα Μέσου Τετραγωνικού Κανονικοποιημένου Σφάλματος (Root Mean Squared Scaled Error - RMSSE)**

$$\text{RMSSE} = \sqrt{\frac{1}{h} \cdot \frac{\sum_{t=n+1}^h (Y_t - \hat{Y}_t)^2}{\frac{1}{n-1} \cdot \sum_{t=2}^n (Y_t - Y_{t-1})^2}}$$

Το RMSSE αποτελεί την κανονικοποιημένη παραλλαγή του RMSE ως προς τις διαφορές πρώτου βαθμού της χρονοσειράς, με  $h$  να είναι το μήκος του χρονικού ορίζοντα πρόβλεψης.

### **2.10.6 Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error - MAPE)**

$$\text{MAPE} = \frac{1}{n} \cdot \sum_{t=1}^n \left| \frac{Y(t) - F(t)}{Y(t)} \right| \cdot 100 (\%)$$

Για να είναι δυνατή η αξιολόγηση μιας προβλεπτικής μεθόδου που έχει εφαρμοστεί σε παραπάνω από μία χρονοσειρές, είναι χρήσιμος ο υπολογισμός των σφαλμάτων σε ποσοστιαία μορφή, καθώς το διαφορετικό επίπεδο μέσης τιμής της εκάστοτε χρονοσειράς θα καθιστούσε τη μεταξύ τους σύγκριση δύσκολη, με αποτέλεσμα της εξαγωγή εσφαλμένων συμπερασμάτων. Δεδομένων των προαναφερθέντων, εισάγεται το MAPE, το οποίο είναι εκφρασμένο επί τοις εκατό και λαμβάνει τιμές μεγαλύτερες ή ίσες του μηδενός. Σε αντιστοιχία με το MAE, όσο μικρότερες τιμές λαμβάνει ο δείκτης αυτός, τόσο μεγαλύτερη είναι η ακρίβεια των προβλέψεων. Σημαντικό μειονέκτημα του MAPE είναι η αδυναμία εφαρμογής του σε χρονοσειρές διακοπτόμενης ζήτησης, καθώς οι μηδενικές τιμές παρατηρήσεων του παρονομαστή απειρίζουν το κλάσμα, οδηγώντας σε απροσδιοριστία.

### **2.10.7 Συμμετρικό Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Symmetric Mean Absolute Percentage Error - sMAPE)**

$$\text{sMAPE} = \frac{1}{n} \cdot \sum_{t=1}^n \left| \frac{Y(t) - F(t)}{\frac{Y(t) + F(t)}{2}} \right| \cdot 100 = \frac{1}{n} \cdot \sum_{t=1}^n \left| \frac{2 \cdot (Y(t) - F(t))}{Y(t) + F(t)} \right| \cdot 100 (\%)$$

Αποτελεί φραγμένη παραλλαγή του MAPE, καθώς πλέον το απόλυτο του σφάλματος διαιρείται με το ημίαθροισμα της πραγματικής και της προβλεφθείσας τιμής, εν αντιθέσει με το MAPE όπου στη θέση του παρονομαστή εντοπίζεται μόνο η πραγματική τιμή. Επομένως, οι τιμές που δύναται να λάβει το sMAPE βρίσκονται στο διάστημα [0% , 200%]. Μολονότι χαρακτηρίζεται

συμμετρικός, ο δείκτης αυτός δε μεταχειρίζεται με τον ίδιο τρόπο τις αισιόδοξες και τις απαισιόδοξες προβλέψεις, όπως φαίνεται χαρακτηριστικά από το ακόλουθο παράδειγμα:

- Αισιόδοξη Πρόβλεψη :  $Y_t = 100$  και  $F_t = 110 \rightarrow sMAPE = 4.76\%$
- Απαισιόδοξη Πρόβλεψη :  $Y_t = 100$  και  $F_t = 90 \rightarrow sMAPE = 5.26\%$

### 2.10.8 Μέσο Απόλυτο Κανονικοποιημένο Σφάλμα (Mean Absolute Scaled Error - MAsE)

$$MASE = \frac{\frac{1}{n} \cdot \sum_{t=1}^n |Y(t) - F(t)|}{\frac{1}{n-1} \cdot \sum_{t=2}^n |Y(t) - Y(t-1)|}$$

Ο δείκτης του MAsE προτάθηκε από τους Hyndman και Koehler (2006) σε μια προσπάθεια αντιμετώπισης των περιπτώσεων απροσδιοριστίας που ενδέχεται να προκύψουν με τη χρήση των δεικτών MAPE και sMAPE όπως επίσης και των διαφορετικών βαρών που αποδίδονται ανάλογα με το μέγεθος του σφάλματος από τους δείκτες MSE και RMSE.

Είναι εμφανής η ομοιότητα της μαθηματικής σχέσης του MAsE με αυτή του μέσου απολύτου σφάλματος, με τη διαφορά ότι είναι κανονικοποιημένη με τη μέση τιμή των διαφορών πρώτου βαθμού της χρονοσειράς. Η προαναφερθείσα μέση τιμή είναι στην ουσία το μέσο απόλυτο σφάλμα της αφελούς προβλεπτικής μεθόδου (Naive), στην οποία η πιο πρόσφατη παρατήρηση λαμβάνεται ως πρόβλεψη. Στην περίπτωση που η τιμή του MAsE προκύψει μικρότερη της μονάδας, συνάγεται το συμπέρασμα ότι η μέθοδος πρόβλεψης που εφαρμόστηκε αποδίδει καλύτερα κατά μέσο όρο από τη Naive, ενώ στην περίπτωση που είναι μεγαλύτερη της μονάδας θα ισχύει το αντίθετο. Μόνη περίπτωση απροσδιοριστίας του MAsE είναι εκείνη όπου όλες οι τιμές των διαθέσιμων ιστορικών δεδομένων είναι μεταξύ τους ίσες, περίπτωση η οποία υπό κανονικές συνθήκες θεωρείται μη πιθανή.

Πέρα από τους στατιστικούς δείκτες ακριβείας που παρουσιάστηκαν παραπάνω και με δεδομένο ότι πυρήνα της παρούσας διπλωματικής αποτελεί η μελέτη των πιθανοτικών προβλέψεων, έγινε η χρήση ενός επιπλέον δείκτη, του Pinball Loss, το οποίο δεν απαντάται στις κλασσικές σημειακές προβλέψεις.

### 2.10.9 Pinball Loss (PL)

$$L_{\tau}(y, z) = \begin{cases} (y - z) \cdot \tau & , \text{αν } y \geq z \\ (z - y) \cdot (1 - \tau) & , \text{αν } z > y \end{cases}$$

Το Pinball Loss χρησιμοποιείται για την αξιολόγηση της ακρίβειας πρόβλεψης των quantiles, καθώς εν αντιθέσει με τις σημειακές προβλέψεις, στην προσέγγιση των πιθανοτικών η απλή σύγκριση πραγματικής και προβλεφθείσας τιμής δεν είναι αρκετή. Έτσι σύμφωνα και με τη μαθηματική του διατύπωση, το PL είναι πάντοτε μη αρνητικό και υπολογίζεται για κάθε επιλεγμένο quantile  $\tau$  της πιθανοτικής πρόβλεψης του εκάστοτε βήματος του χρονικού ορίζοντα πρόβλεψης. Η σχέση υπολογισμού του είναι δίκλαδη, με το βάρος που δίνεται στο σφάλμα να διαφέρει ανάλογα με το εάν η αλγεβρική τιμή της προβλεφθείσας τιμής στο εκάστοτε quantile  $\tau$

είναι μικρότερη ή μεγαλύτερη της πραγματικής τιμής. Όσο μικρότερη είναι η τιμή του PL, τόσο περισσότερο ακριβής είναι η πρόβλεψη που έχει εξαχθεί.

### **2.10.10 Scaled Pinball Loss (SPL)**

Το SPL αφορά μία παραλλαγή του PL που παρουσιάστηκε προηγουμένως και η οποία χρησιμοποιήθηκε στο πλαίσιο του διαγωνισμού M5, λόγω της αβεβαιότητας των δεδομένων του διαγωνισμού που προέκυπτε από τη διακοπτόμενη φύση των χρονοσειρών αυτών και κανονικοποιεί το Pinball Loss ως προς τις διαφορές της χρονοσειράς με καθυστέρηση ίση με το μήκος της εποχιακότητας, δηλαδή 12 περιόδων ( $\frac{1}{n-k} \cdot \sum_{t=k+1}^n |Y(t) - Y(t-k)|$ ).

## ΚΕΦΑΛΑΙΟ 3

### Νευρωνικά Δίκτυα

#### 3.1 Εισαγωγή

Όπως διαφαίνεται και από την ονομασία τους, κεντρική ιδέα για την κατασκευή των τεχνητών νευρωνικών δικτύων (ή απλούστερα των νευρωνικών δικτύων) αποτέλεσε ο ανθρώπινος εγκέφαλος. Εν αντιθέσει με τον τρόπο με τον οποίο επιτελεί τους υπολογισμούς ένας ψηφιακός υπολογιστής, ο ανθρώπινος εγκέφαλος χαρακτηρίζεται από μεγαλύτερη πολυπλοκότητα, μη γραμμικότητα, καθώς και από την ικανότητα να οργανώνει τις δικές του δομικές μονάδες (γνωστές και ως νευρώνες), με αποτέλεσμα οι υπολογισμοί να εκτελούνται με μεγαλύτερη ταχύτητα. Η προαναφερθείσα ικανότητα οργάνωσης καθιστά τον ανθρώπινο εγκέφαλο ιδιαίτερα ευπροσάρμοστο στις όποιες μεταβολές του περιβάλλοντός του, με αποτέλεσμα τη δημιουργία κανόνων συμπεριφοράς, οι οποίοι είναι γνωστοί και ως “εμπειρία”. Βάσει του προτύπου λειτουργίας που περιγράφηκε, ένα νευρωνικό δίκτυο σχεδιάζεται έτσι ώστε να μοντελοποιεί τον τρόπο που εκτελείται μία εργασία από τον ανθρώπινο εγκέφαλο. Το ρόλο των δομικών μονάδων που συνάπτονται με συνδέσεις αναλαμβάνουν πολυάριθμα, απλά και διασυνδεδεμένα μεταξύ τους “υπολογιστικά κύτταρα”, τα οποία είναι γνωστά ως νευρωνες ή μονάδες επεξεργασίας.

Η απαρχή των νευρωνικών δικτύων χρονολογείται μεταξύ του τέλους του 19ου και των αρχών του 20ου αιώνα, περίοδο κατά την οποία οι κλάδοι της πληροφορικής και της νευροεπιστήμης γνώρισαν σημαντική ανάπτυξη. Τα πρώτα απτά αποτελέσματα νευρωνικών μοντέλων ήρθαν το 1943 από το νευροφυσιολόγο Warren McCulloch και το μαθηματικό Walter Pitts οι οποίοι στη δημοσίευσή τους με τίτλο “The Logical Calculus of the Ideas Immanent in Nervous Activity” περιέγραψαν τον τρόπο με τον οποίο θα μπορούσαν να λειτουργούν οι νευρώνες. Εν συνεχεία τη δεκαετία του 1960 εμφανίστηκαν οι πρώτοι υπολογιστές των οποίων οι επεξεργαστές λειτουργούσαν με πρότυπο τα νευρωνικά δίκτυα, ωστόσο ελλείψει υπολογιστικών πόρων η περαιτέρω ανάπτυξή τους έμεινε στάσιμη μέχρι τα μέσα της δεκαετίας του 1980 όταν και αναζωπυρώθηκε το ενδιαφέρον για τη νευρωνική υπολογιστική, με πρωτοστάτη τον J.J. Hopfield.

Η υλοποίησή των νευρωνικών δικτύων βασίζεται στη χρήση ηλεκτρονικών μερών ή στην προσομοίωση από εξειδικευμένο λογισμικό, τα οποία μέσω της τροποποίησης των συναπτικών βαρών του δικτύου, προσπαθούν να προσεγγίσουν τον εκάστοτε επιθυμητό στόχο. Η διαδικασία μέσω της οποίας επιτυγχάνεται η προαναφερθείσα μάθηση καλείται αλγόριθμος μάθησης και ένα από τα πλέον σημαντικά πλεονεκτήματά της αποτελεί το γεγονός ότι η προαναφερθείσα τροποποίηση της τοπολογίας του δικτύου μέσω των βαρών που ανατίθενται στους νευρώνες γίνεται δυναμικά, εις τρόπον ώστε να αλλάζει η κατάσταση του καθενός από αυτούς, ήτοι να ενεργοποιούνται ή να απενεργοποιούνται.

Εν κατακλείδι, τα νευρωνικά δίκτυα αποτελούν μία αλγοριθμική, κατανεμημένης αρχιτεκτονικής δομή, η οποία άπτεται του κλάδου της υπολογιστικής νοημοσύνης και είναι ικανή αφενός μεν να παρέχει μεγάλη υπολογιστική ισχύ, αφετέρου δε να γενικευθεί για επεξεργασία εισόδων τις οποίες το εκάστοτε δίκτυο δεν έχει συναντήσει κατά τη διάρκεια της εκπαίδευσής του.

### 3.2 Ιδιότητες Νευρωνικών Δικτύων

Η κατανεμημένη δομή ενός νευρωνικού δικτύου σε συνδυασμό με τη δυνατότητα γενίκευσής του, το καθιστά ικανό να παρέχει αξιόπιστες λύσεις σε ένα ευρύ φάσμα προβλημάτων. Προς επίρρωση του επιχειρήματος υπέρ της χρήσης των νευρωνικών δικτύων, ακολούθως θα παρουσιαστούν ορισμένες επιπρόσθετες ιδιότητές τους.

- **Μη γραμμικότητα**

Ένας νευρώνας έχει τη δυνατότητα είτε να είναι γραμμικός είτε μη γραμμικός, καθιστώντας ως εκ τούτου το νευρωνικό δίκτυο στο σύνολό του γραμμικό ή μη γραμμικό αντίστοιχα. Η ιδιότητα της μη γραμμικότητας, η οποία ειρήσθω εν παρόδω κατανέμεται σε καθ' όλη την έκταση του δικτύου, αποτελεί ένα πολύ σημαντικό πλεονέκτημα που επιτρέπει στα νευρωνικά δίκτυα να επεξεργάζονται μη γραμμικές εισόδους, όπως επί παραδείγματι φυσικούς μηχανισμούς παραγωγής σήματος εξόδου (λχ ομιλία).

- **Αντιστοίχιση Εισόδου - Εξόδου**

Τα νευρωνικά δίκτυα παρέχουν τη δυνατότητα της επιβλεπόμενης μάθησης, η οποία είναι μάλιστα μία από τις δημοφιλέστερες τεχνικές μάθησης. Σύμφωνα με αυτή, εφαρμόζονται στο δίκτυο ορισμένα “παραδείγματα εκπαίδευσης” (ή “παραδείγματα εργασιών”), τα οποία αποτελούν ζεύγη ενός σήματος εισόδου με την αντίστοιχη επιθυμητή απόκριση-στόχο. Επιπροσθέτως παρουσιάζεται στο δίκτυο και ένα τυχαία επιλεγμένο σετ βαρών τα οποία τροποποιούνται με στόχο την ελαχιστοποίηση της διαφοράς μεταξύ εισόδου και εξόδου. Η φύση του κριτηρίου βάσει του οποίου αξιολογείται η σύγκλιση αυτή είναι στατιστική, ενώ η διαδικασία εκπαίδευσης επαναλαμβάνεται για μεγάλο αριθμό παραδειγμάτων ώστε να επιτευχθεί η επιθυμητή κατάσταση ισορροπίας, στην οποία οι μεταβολές που προκύπτουν στα συναπτικά βάρη κρίνονται ως αμελητέες.

- **Προσαρμοστικότητα**

Τα νευρωνικά δίκτυα έχουν τη δυνατότητα να προσαρμόζουν τα συναπτικά βάρη τους ανάλογα με τις αλλαγές που λαμβάνουν χώρα στο περιβάλλον τους ή ακόμα και να επανεκπαιδεύονται σε πραγματικό χρόνο, ώστε να χειρίζονται μη στατικές καταστάσεις, στις οποίες οι εισοδοί που λαμβάνει το δίκτυο μεταβάλλονται συνεχώς, όπως επί παραδείγματι η προσαρμοστική αναγνώριση προτύπων. Ωστόσο, το όφελος που προκύπτει από την ικανότητα προσαρμοστικότητας, ενδέχεται να βαίνει εις βάρος της ευρωστίας του νευρωνικού δικτύου, συνεπώς αναζητείται πάντα η χρυσή τομή μεταξύ ρυθμού ανανέωσης του δικτύου και ευρωστίας, πρόβλημα το οποίο διατυπώνεται και ως δίλημμα σταθερότητας - πλαστικότητας.

- **Ενδεικτική Απόκριση**

Στο πλαίσιο της ταξινόμησης προτύπων, ένα νευρωνικό δίκτυο δύναται να σχεδιαστεί με τρόπο ώστε να παρέχεται, πέραν του προτύπου που επιλέγεται, επιπρόσθετη πληροφορία αναφορικά με το βαθμό εμπιστοσύνης της ληφθείσας απόφασης. Η πληροφορία αυτή είναι μείζονος σημασίας και λαμβάνεται υπόψη σε περιπτώσεις διφορούμενων μοτίβων, ώστε να απορριφθούν ενδεχόμενα ασαφή αποτελέσματα και να βελτιωθεί η συνολική απόδοση του δικτύου.

- **Πληροφορία σχετική με το περιεχόμενο**

Δεδομένων των όσων έχουν ειπωθεί παραπάνω για τα νευρωνικά δίκτυα, συνάγεται το συμπέρασμα ότι η γνώση αντιπροσωπεύεται από την δομή καθαυτή του δικτύου, όπως επίσης και από την κατάσταση ενεργοποίησης των νευρώνων. Το γεγονός ότι κάθε νευρώνας επηρεάζεται ενδεχομένως από τη συνολική δραστηριότητα των υπόλοιπων νευρώνων που απαρτίζουν το δίκτυο, καταδεικνύει το ότι ένα νευρωνικό δίκτυο χειρίζεται με φυσικό τρόπο τη σχετιζόμενη με το περιεχόμενο πληροφορία.

- **Ανοχή σε βλάβες**

Η υλοποίηση του νευρωνικού με υλικό το καθιστά ανεκτικό σε βλάβες, υπό την έννοια ότι η απόδοσή του μειώνεται βαθμιαία. Πιο συγκεκριμένα, λόγω της κατανομημένης φύσης του νευρωνικού δικτύου που έχει περιγραφεί παραπάνω, μια ενδεχόμενη βλάβη θα πρέπει να λάβει μεγάλη έκταση στο εύρος του δικτύου για να είναι αισθητή η διαφορά στην ποιότητα της ανάκλησης προτύπων, καθώς ακόμα και στην περίπτωση που καταστραφεί ένας νευρώνας (ή οι συνδέσεις του), η ποιότητα της ανάκλησης ενός αποθηκευμένου προτύπου θα μειωθεί τοπικά στην περιοχή του κατεστραμμένου νευρώνα.

- **Γενίκευση στη σχεδίαση και την ανάλυση**

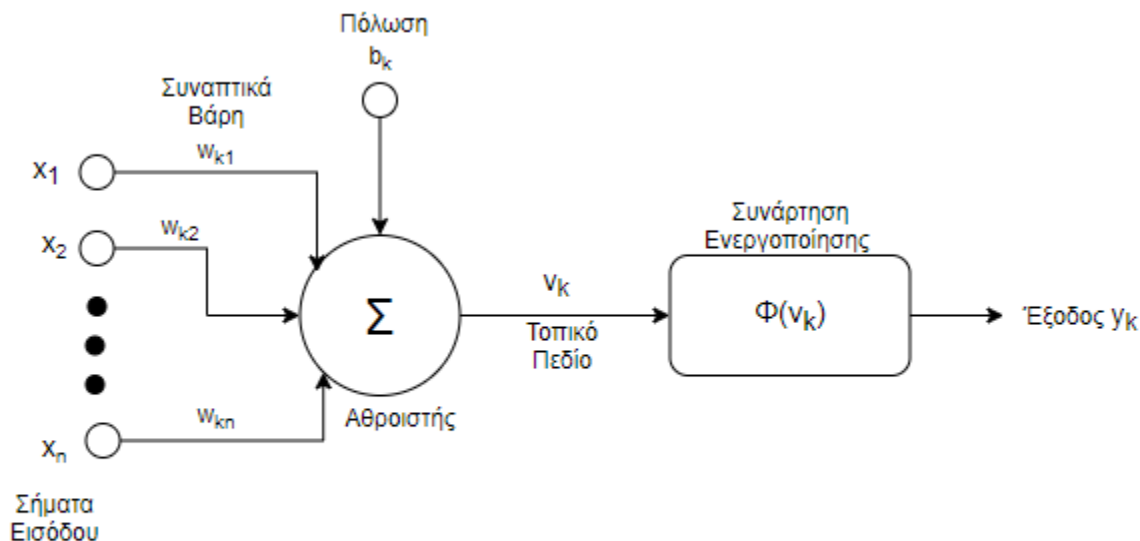
Η γενίκευση των νευρωνικών δικτύων αποτελεί την ιδιότητα στην οποία οφείλουν την καθολική τους αποδοχή σε πληθώρα επιστημονικών πεδίων. Αναλυτικότερα, τα νευρωνικά δίκτυα δύναται να χρησιμοποιηθούν άμεσα σε διαφορετικές εφαρμογές, χωρίς να διαφοροποιηθούν η σημειογραφία και ο τρόπος εφαρμογής τους, με αποτέλεσμα να καθίστανται κοινώς αποδεκτά ως επεξεργαστές πληροφοριών, επιτρέποντας τη χρήση ίδιων θεωριών και αλγορίθμων μάθησης σε διαφορετικές εφαρμογές και γνωστικά αντικείμενα, καθώς και την απρόσκοπτη ενοποίηση επιμέρους λειτουργικών μονάδων (modules) για το σχηματισμό σπονδυλωτών δικτύων.

- **Αναλογία με τη Φυσιολογία του Εγκεφάλου**

Η φιλοσοφία των νευρωνικών δικτύων εδράζεται στη δομή και τις αρχές λειτουργίας του ανθρώπινου εγκεφάλου που αποτελεί την επιτομή της εύρωστης και παράλληλης επεξεργασίας, με κυριότερα χαρακτηριστικά τη μεγάλη ταχύτητα και ισχύ. Η αναλογία αυτή καθιστά τα νευρωνικά δίκτυα ιδιαίτερα χρήσιμο εργαλείο των νευροβιολόγων, οι οποίοι τα αξιοποιούν για τη μελέτη ποικίλων νευροβιολογικών φαινομένων. Στον αντίποδα, οι μηχανικοί αντλούν από το πεδίο της νευροβιολογίας ιδέες για νέες τεχνικές σχεδίασης ώστε να δημιουργήσουν νέα δίκτυα που να ανταποκρίνονται σε σύνθετα προβλήματα που είναι αδύνατο να επιλυθούν με τις συμβατικές τεχνικές σχεδίασης.

### 3.3 Μοντέλο Νευρώνων

Όπως περιγράφηκε παραπάνω, δομική μονάδα των νευρωνικών δικτύων αποτελεί ο νευρώνας. Ένα βασικό μοντέλο νευρώνα παρουσιάζεται στο ακόλουθο



Σχήμα 3.1 : Μοντέλο Νευρώνα

Τα συστατικά του στοιχείου είναι τρία και συγκεκριμένα:

- οι συνάψεις ή διασυνδέσεις, κάθε μία εκ των οποίων έχει το δικό της βάρος με το οποίο πολλαπλασιάζονται οι εισοδοί πριν εισέλθουν στον αθροιστή. Έτσι, μία είσοδος  $x_i$  θα πολλαπλασιαστεί με βάρος  $w_{ki}$  από το νευρώνα  $k$  (εν γένει στη βιβλιογραφία επιλέγεται ο πρώτος δείκτης του συμβολισμού της βαρύτητας να αναφέρεται στο νευρώνα). Να σημειωθεί ότι το βάρος κάθε νευρώνα δύναται να λάβει τόσο αρνητικές όσο και θετικές τιμές
- ο αθροιστής, στον οποίο αθροίζονται τα σταθμισμένα σήματα εισόδου
- η συνάρτηση ενεργοποίησης (activation function) ή συνάρτηση περιορισμού (squashing function), η οποία περιορίζει το εύρος του σήματος εξόδου εντός ορισμένων ορίων. Τυπικά το εύρος τιμών του πλάτους εξόδου κανονικοποιείται στο διάστημα  $[0,1]$  ή  $[-1,1]$ .

Σημαντικό στοιχείο του νευρώνα αποτελεί επιπλέον η πόλωση (bias), η οποία ανάλογα με την τιμή της (θετική ή αρνητική) οδηγεί σε αύξηση ή μείωση της διέγερσης της συνάρτησης ενεργοποίησης.



Συνοπτικά, η μαθηματική περιγραφή της λειτουργίας ενός νευρώνα είναι η εξής:

$$u_k = \sum_{j=1}^m w_{kj} x_j$$

$$y_k = \varphi(u_k + b_k)$$

όπου

$x_j$	το εκάστοτε σήμα εισόδου
$w_{kj}$	το συναπτικό βάρος του νευρώνα $k$ όταν λαμβάνει την είσοδο $x_j$
$u_k$	η έξοδος του γραμμικού συνδυαστή
$b_k$	η πόλωση
$\varphi$	η συνάρτηση ενεργοποίησης
$y_k$	η τελική έξοδος του νευρώνα $k$

### 3.4 Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης (activation functions - AF) καθορίζουν την ενεργοποίηση ή μη ενός νευρώνα βάσει του σταθμισμένου αθροίσματος των σημάτων εισόδου και της προκατάληψης. Χειρίζονται κλιμακωτά τα δεδομένα και παράγουν έξοδο που περιέχει τις παραμέτρους των δεδομένων. Ανάλογα με τη συνάρτηση που περιγράφουν, οι AF μπορεί να είναι γραμμικές ή μη γραμμικές και χρησιμοποιούνται για τον έλεγχο των εξόδων νευρωνικών δικτύων που αξιοποιούνται σε πληθώρα εφαρμογών, από αναγνώριση ομιλίας μέχρι ανίχνευση δακτυλικών αποτυπωμάτων.

Οι ευρύτερα χρησιμοποιούμενες συναρτήσεις ενεργοποίησης είναι οι εξής:

#### 3.4.1 Σιγμοειδής Συνάρτηση

Πρόκειται για μία μη γραμμική, αύξουσα συνάρτηση η οποία χρησιμοποιείται ως επί το πλείστον σε νευρωνικά δίκτυα πρόσθιας τροφοδότησης και λαμβάνει τιμές στο διάστημα  $[0,1]$ . Το όνομά της οφείλεται στη γραφική της παράσταση που θυμίζει το αγγλικό γράμμα “S”, ενώ είναι φραγμένη και ορίζεται για πραγματικές τιμές εισόδου. Υπάρχουν αρκετές παραλλαγές της, με τη λογιστική που είναι και η πιο γνωστή να έχει την ακόλουθη μορφή:

$$\varphi(u) = \frac{1}{1 + e^{-au}}$$

Στον παραπάνω τύπο,  $u$  είναι η έξοδος του γραμμικού συνδυαστή, ή τοπικό πεδίο και η παράμετρος  $a$  ονομάζεται παράμετρος κλίσης. Όσο μεγαλύτερη είναι η τιμή της παραμέτρου

αυτής, τόσο πιο απότομη γίνεται η κλίση της καμπύλης της σιγμοειδούς συνάρτησης. Επικρατέστερη τιμή της παραμέτρου κλίσης είναι η μονάδα.

Η σιγμοειδής συνάρτηση εμφανίζεται στα επίπεδα εξόδου των νευρωνικών δικτύων μηχανικής και βαθιάς μάθησης και το σημαντικό της πλεονέκτημα είναι η ευκολία κατανόησης της λειτουργίας της καθώς και η αποτελεσματικότητά της σε “ρηγά” δίκτυα. Στον αντίποδα εμφανίζει ορισμένα μειονεκτήματα, όπως η απότομη απόσβεση των τιμών κατά την ανάστροφη διάδοση και ο κορεσμός.

### 3.4.2 Συνάρτηση Υπερβολικής Εφαπτομένης

Εν αντιθέσει με τη σιγμοειδή συνάρτηση, η υπερβολική εφαπτομένη είναι κεντραρισμένη γύρω από το 0 λαμβάνοντας τιμές στο διάστημα  $[-1,1]$  και προτιμάται σε σχέση με την πρώτη λόγω του γεγονότος ότι δίνει καλύτερα αποτελέσματα στην εκπαίδευση νευρωνικών δικτύων πολλών επιπέδων. Ο τύπος της συνάρτησης αυτής είναι ο ακόλουθος:

$$\varphi(v) = \tanh(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}}$$

Η ανωτέρω συνάρτηση προτιμάται έναντι της σιγμοειδούς λόγω της καλύτερης απόδοσης που εμφανίζει κατά την εκπαίδευση νευρωνικών δικτύων πολλαπλών επιπέδων και η χρήση της εντοπίζεται σε εφαρμογές που σχετίζονται με αναδρομικά δίκτυα και επεξεργασία φυσικής γλώσσας.

Παράλληλα, το γεγονός ότι μπορεί να λάβει μηδενικές τιμές επικουρεί το έργο της διαδικασίας του backpropagation (το οποίο θα αναλυθεί σε επόμενη ενότητα της παρούσας εργασίας). Αξιοσημείωτο μειονέκτημά της ωστόσο αποτελεί το γεγονός ότι στην περίπτωση που λάβει την τιμή 0 (για  $x=0 \rightarrow f(0) = 0$ ) η κλίση της γίνεται ίση με τη μονάδα, με αποτέλεσμα να εμφανίζεται η κατάσταση νεκρού νευρώνα, κατά την οποία το συναπτικό βάρος δε χρησιμοποιείται.

### 3.4.3 SoftMax Συνάρτηση

Η SoftMax συνάρτηση ενεργοποίησης χρησιμοποιείται για τον υπολογισμό της κατανομής πιθανότητας ενός διανύσματος πραγματικών αριθμών. Η έξοδός της λαμβάνει τιμές στο διάστημα  $[0,1]$  με τις υπολογιζόμενες πιθανότητες να αθροίζουν στη μονάδα. Δίνεται από τη σχέση:

$$\varphi(v) = \frac{e^{v_i}}{\sum_{j=1}^K e^{v_j}}$$

Ενώ η Σιγμοειδής συνάρτηση χρησιμοποιείται σε εφαρμογές δυαδικής φύσης, η Softmax απαντάται σε εφαρμογές πολυπαραγοντικής κατηγοριοποίησης, στις οποίες επιστρέφεται η πιθανότητα της κάθε κλάσης. Έτσι η κλάση η οποία συγκεντρώνει τη μεγαλύτερη πιθανότητα αποτελεί την κλάση-στόχο.

### 3.4.4 Rectified Linear Unit (ReLU)

Η πιο ευρέως χρησιμοποιούμενη συνάρτηση ενεργοποίησης είναι η ReLU, η οποία προτάθηκε από τους Nair και Hinton (2010). Η ειδοποιός διαφορά της σε σχέση με τις άλλες συναρτήσεις είναι το γεγονός ότι δεν προκαλεί προβλήματα στα βαθιά δίκτυα (δίκτυα με πολλά επίπεδα), όπως επί παραδείγματι το μηδενισμό των παραγώγων που συνεπάγεται την απενεργοποίηση ορισμένων νευρώνων κατά τη διαδικασία της εκπαίδευσης. Ορίζεται ως εξής:

$$\varphi(v) = \max(0, v)$$

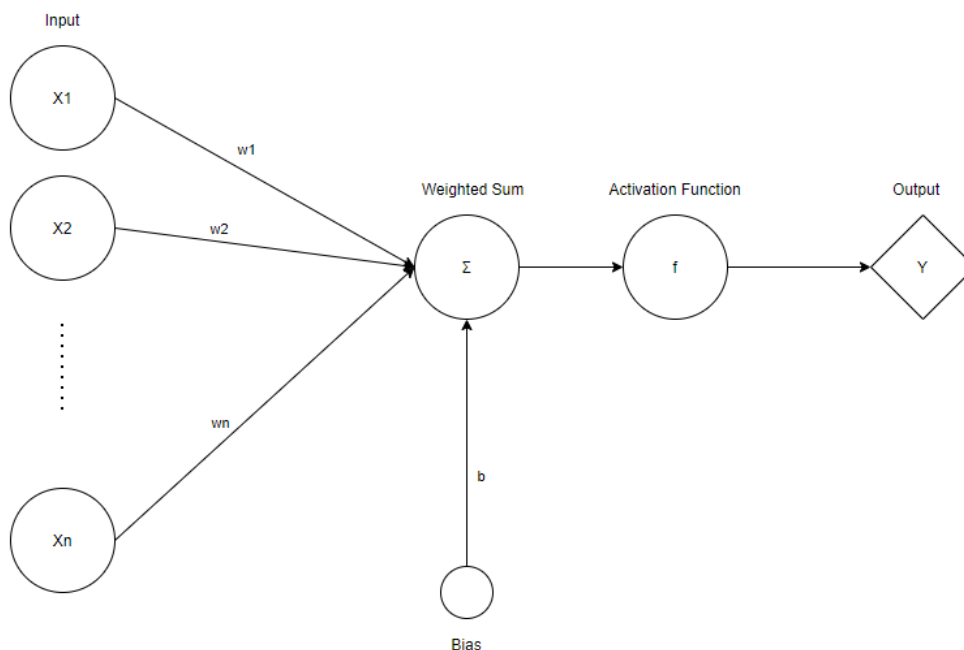
Όπως διαπιστώνεται από τη μορφή της, η ReLU έχει μία σχεδόν γραμμική μορφή διατηρώντας τις ιδιότητες των γραμμικών μοντέλων σε περίπτωση που εφαρμοστεί σε αυτά. Πέραν της αποφυγής του μηδενισμού των παραγώγων των νευρώνων, επιπρόσθετο όφελος αποτελεί το μικρότερο κόστος της από πλευράς χρόνου. Το χρονικά μικρότερο κόστος οφείλεται στο ότι δεν απαιτούνται σύνθετοι υπολογισμοί εκθετικών μεγεθών. Στον αντίποδα ωστόσο, εγκυμονεί ο κίνδυνος της υπερπροσαρμογής (overfitting) στα δεδομένα, για την αποφυγή της οποίας έχει υιοθετηθεί η τεχνική της εγκατάλειψης (dropout technique) με αποτέλεσμα τη βελτίωση της απόδοσης των βαθέων νευρωνικών δικτύων.

## 3.5 Αρχιτεκτονικές Νευρωνικών Δικτύων

Η αρχιτεκτονική ενός νευρωνικού δικτύου περιγράφει τον τρόπο με τον οποίο είναι δομημένοι οι νευρώνες μέσα σε αυτό, ενώ παράλληλα σχετίζεται άμεσα με τον αλγόριθμο μάθησης (περισσότερα για τους αλγορίθμους μάθησης αναφέρονται στην αντίστοιχη παράγραφο) που χρησιμοποιείται για την εκπαίδευσή του. Εν συνεχεία παρουσιάζονται οι μερικές από τις πλέον δημοφιλείς αρχιτεκτονικές.

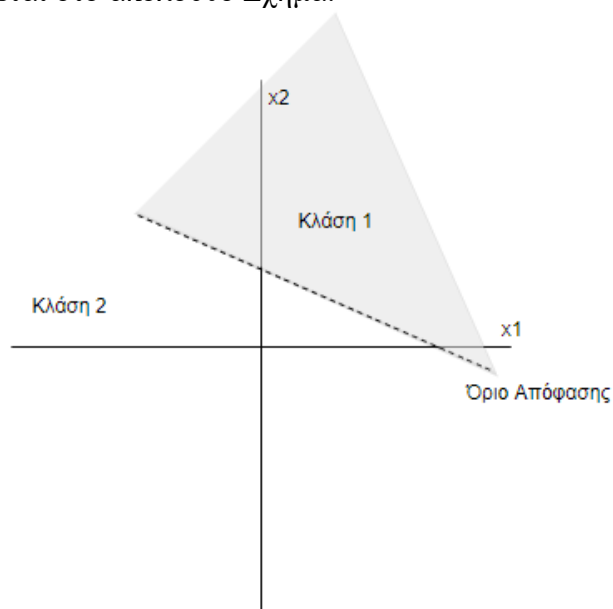
### 3.5.1 Perceptron

Το Perceptron αποτελεί την απλούστερη μορφή νευρωνικού δικτύου. Η χρήση του αφορά την ταξινόμηση προτύπων που είναι γραμμικά διαχωρίσιμα, ήτοι μπορούν να κατηγοριοποιηθούν σε δύο διακριτές κλάσεις τα δείγματα των οποίων βρίσκονται σε αντίθετες πλευρές ενός υπερεπιπέδου. Η δομή ενός Perceptron αποτελείται από ένα γραμμικό νευρώνα και μία μη γραμμική συνάρτηση ενεργοποίησης και παρουσιάζεται στο ακόλουθο σχήμα:



Σχήμα 3.2 : Αρχιτεκτονική του Perceptron

Η συγκεκριμένη αρχιτεκτονική προτάθηκε από τον Rosenblatt, ο οποίος απέδειξε ότι ο αλγόριθμος του Perceptron δύναται να συγκλίνει και μάλιστα μπορεί να τοποθετήσει ένα διαχωριστικό υπερεπίπεδο απόφασης μεταξύ δύο κλάσεων. Οι κλάσεις αυτές θα προκύπτουν από το γραμμικό διαχωρισμό των δειγμάτων που χρησιμοποιούνται στο πλαίσιο της εκπαίδευσης, όπως φαίνεται στο ακόλουθο Σχήμα:



Σχήμα 3.3 : Το υπερεπίπεδο ως όριο απόφασης για την ταξινόμηση προτύπων με δύο κλάσεις

Ο νευρώνας βασίζεται στο μοντέλο των McCulloch-Pitts, το οποίο αποτελεί το απλούστερο μαθηματικά μοντέλο δυαδικής εξόδου. Πιο συγκεκριμένα τα σήματα εισόδου σταθμίζονται με τα αντίστοιχα βάρη και στην περίπτωση που το σταθμισμένο άθροισμα υπερβαίνει μία ορισμένη τιμή κατωφλίου, ο νευρώνας ενεργοποιείται. Συνεπώς η έξοδος του είναι +1 ανάλογα με την ενεργοποίηση ή μη του νευρώνα, ενώ η μαθηματική διατύπωση της εξόδου του γραμμικού τμήματος (που αποτελεί ταυτόχρονα την είσοδο του περιοριστή) να είναι η ακόλουθη:

$$u_k = \sum_{j=1}^m w_j x_j + b$$

Υπενθυμίζεται από την παρουσίαση του μοντέλου νευρώνων ότι:

- $x_j$  οι είσοδοι του νευρώνα
- $w_j$  τα συναπτικά βάρη και
- $b$  η πόλωση

Για το διαχωρισμό των σημάτων εισόδου σε δύο κλάσεις που αναφέρθηκε παραπάνω, ο νευρώνας αντιστοιχίζει την τιμή +1 στη μία (έστω  $C_1$ ) και την τιμή -1 στην άλλη (έστω  $C_2$ ). Καθίσταται προφανές ότι η περίπτωση που

$$\sum_{j=1}^m w_j x_j + b = 0$$

αποτελεί το όριο απόφασης. Αναφορικά με την πόλωση, οποιαδήποτε μεταβολή στην τιμή της, μετατοπίζει το υπερεπίπεδο σε σχέση με την αρχή των αξόνων.

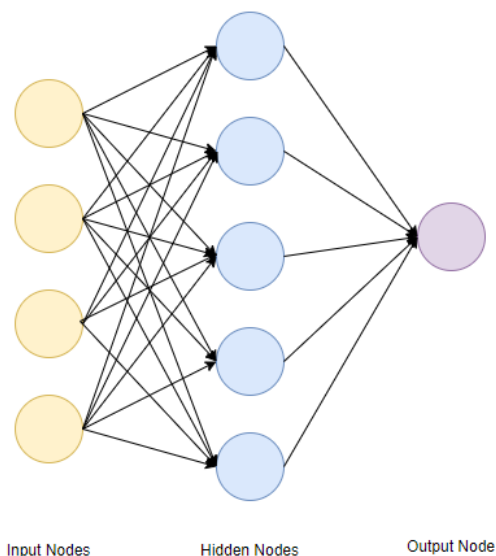
Για να είναι εφικτά τα προαναφερθέντα, θα πρέπει ο Perceptron να παράγει έγκυρη έξοδο, γεγονός που επιτυγχάνεται μόνο αν τα πρότυπα που ανήκουν σε διαφορετικές κλάσεις έχουν επαρκείς διαφορές μεταξύ τους ώστε να είναι δυνατός ο προσδιορισμός ενός διαχωριστικού επιπέδου.

Η εκπαίδευση του Perceptron αποτελεί μία επαναληπτική διαδικασία, μέσω της οποίας προσαρμόζονται τα συναπτικά βάρη έτσι ώστε τα δείγματα που παρουσιάζονται από διαφορετικές κλάσεις να ταξινομηθούν στην αντίστοιχη επιθυμητή κλάση, βάσει ενός κανόνα διόρθωσης σφαλμάτων ο οποίος αποτελεί και τον αλγόριθμο σύγκλισης του perceptron.

Τέλος ως μέτρο απόδοσης χρησιμοποιείται η συνάρτηση απόδοσης που προτάθηκε από τον Shynk (1990) με σκοπό την αντιμετώπιση της δυσκολίας που θα παρουσίαζε, στην αναλυτική παραγωγή του αλγορίθμου, η χρήση της μέσης πιθανότητας του λάθους ταξινόμησης, ήτοι της μέσης πιθανότητας του perceptron να λάβει απόφαση υπέρ μιας κλάσης όταν το διάνυσμα εισόδου ανήκει σε άλλη κλάση.

### 3.5.2 Δίκτυα Πρόσθιας Τροφοδότησης Μοναδικού Επιπέδου

Τα δίκτυα εμπρόσθιας τροφοδότησης (feedforward networks) αποτελούν μία ιδιαίτερα δημοφιλή κατηγορία αρχιτεκτονικών, ενώ αποτελούν και την αρχιτεκτονική των δικτύων που χρησιμοποιήθηκαν για τους σκοπούς της παρούσας διπλωματικής εργασίας. Στην κατηγορία αυτή ανήκουν τα δίκτυα που οργανώνονται σε επίπεδα, με τους κόμβους (νευρώνες) ενός επιπέδου να συνδέονται απευθείας με τους κόμβους του επόμενου, χωρίς να είναι δυνατό το αντίστροφο. Στην περίπτωση των δικτύων πρόσθιας τροφοδότησης ενός επιπέδου, που είναι και η απλούστερη υλοποίηση της feedforward αρχιτεκτονικής, ένα επίπεδο εισόδου αποτελούμενο από πηγαίους κόμβους, που αναπαριστούν τις συνιστώσες των διανυσμάτων εισόδου, συνδέεται μέσω ενός επιπέδου κόμβων υπολογισμού με το επίπεδο νευρώνων εξόδου. Όπως συμπεραίνεται και από το κάτωθι σχήμα ενός απλού δικτύου πρόσθιας τροφοδότησης με τέσσερις μεταβλητές εισόδου, πέντε κόμβους στο κρυφό επίπεδο και έναν κόμβο εξόδου, ο χαρακτηρισμός του ενός επιπέδου αναφέρεται στα επίπεδα υπολογισμού, καθώς δεν προσμετράται το επίπεδο εισόδου.



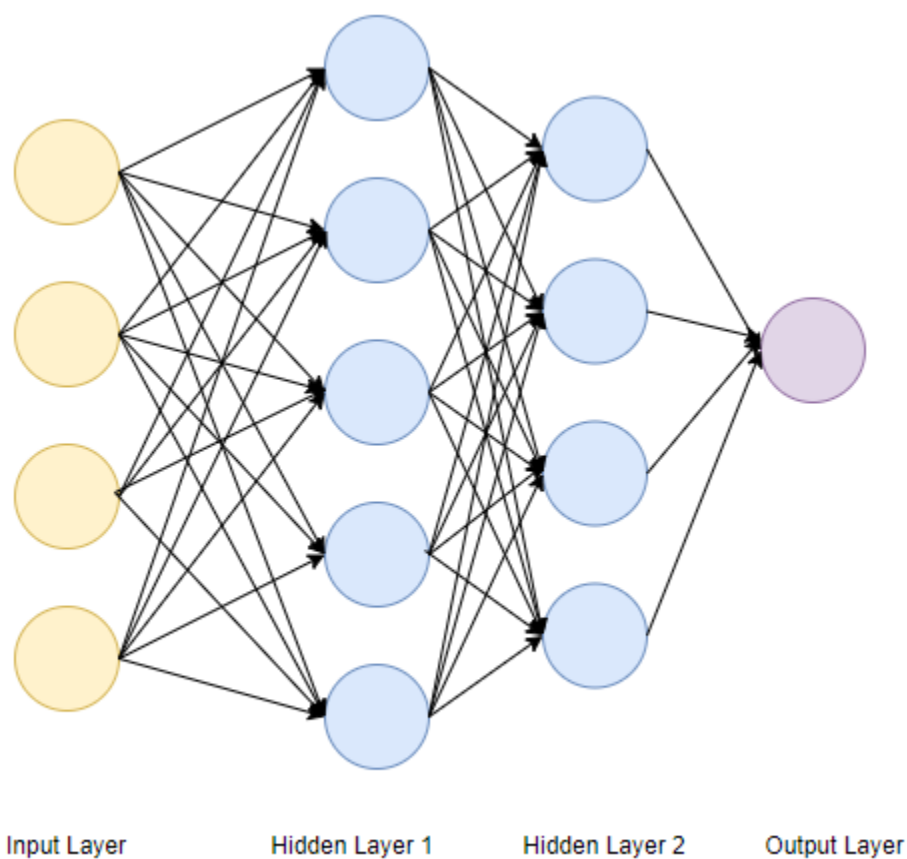
Σχήμα 3.4 : Αρχιτεκτονική Δικτύου Πρόσθιας Τροφοδότησης Ενός Επιπέδου

### 3.5.3 Δίκτυα Πρόσθιας Τροφοδότησης Πολλαπλών Επιπέδων

Επέκταση της προαναφερθείσας κατηγορίας Δικτύων Πρόσθιας Τροφοδότησης Μοναδικού Επιπέδου αποτελούν τα δίκτυα πολλαπλών επιπέδων. Πιο συγκεκριμένα, αποτελούνται από ένα ή περισσότερα “κρυφά” επίπεδα, οι νευρώνες (κόμβοι) των οποίων αποκαλούνται επίσης “κρυφοί”. Ο όρος “κρυφός” υποδηλώνει την ιδιότητα του μέρους αυτού του δικτύου να μην είναι άμεσα ορατό από την είσοδο ή την έξοδο, παρά μόνο μεσολαβούν μεταξύ της εξωτερικά προερχόμενης εισόδου και της εξόδου του δικτύου. Κάθε “κρυφό” επίπεδο συνδέεται με το αμέσως επόμενο, χωρίς να λαμβάνει χώρα η αντίστροφη διασύνδεση. Έτσι η έξοδος των

υπολογιστικών κόμβων του πρώτου επιπέδου αποτελεί την είσοδο των νευρώνων του δεύτερου, ακολούθως η έξοδος του δεύτερου κρυφού επιπέδου είναι η είσοδος του τρίτου κ.ο.κ. Το σύνολο των σημάτων εξόδου στο επίπεδο εξόδου του δικτύου συνιστά και την τελική απόκριση του δικτύου. Η περιγραφή ενός τέτοιου δικτύου γίνεται βάση του αριθμού των επιπέδων και των κόμβων που εντοπίζεται σε καθένα από αυτά, δηλαδή ένα δίκτυο  $i$  πηγαίων κόμβων (είσοδος),  $h_1$  και  $h_2$  νευρώνες στο πρώτο και το δεύτερο κρυφό επίπεδο αντίστοιχα και ο κόμβους στο επίπεδο εξόδου, αναφέρεται ως δίκτυο  $i - h_1 - h_2 - o$ .

Κατά την υλοποίηση της εν λόγω αρχιτεκτονικής, δίνεται η δυνατότητα σύνδεσης κάθε κόμβου του εκάστοτε επιπέδου με όλους τους κόμβους του επόμενου (προς τα εμπρός) επιπέδου, ή με ορισμένους από αυτούς. Στην πρώτη περίπτωση γίνεται λόγος για πλήρως συνδεδεμένο δίκτυο, όπως επί παραδείγματι το δίκτυο του κάτωθι Σχήματος, ενώ στην περίπτωση απουσίας ορισμένων συναπτικών συνδέσεων, το δίκτυο θεωρείται μερικώς συνδεδεμένο. Τέλος, όσο περισσότερα κρυφά επίπεδα προστίθενται στο δίκτυο, τόσο ποιοτικά υψηλότερα είναι τα χαρακτηριστικά και τα στατιστικά που εξάγει.

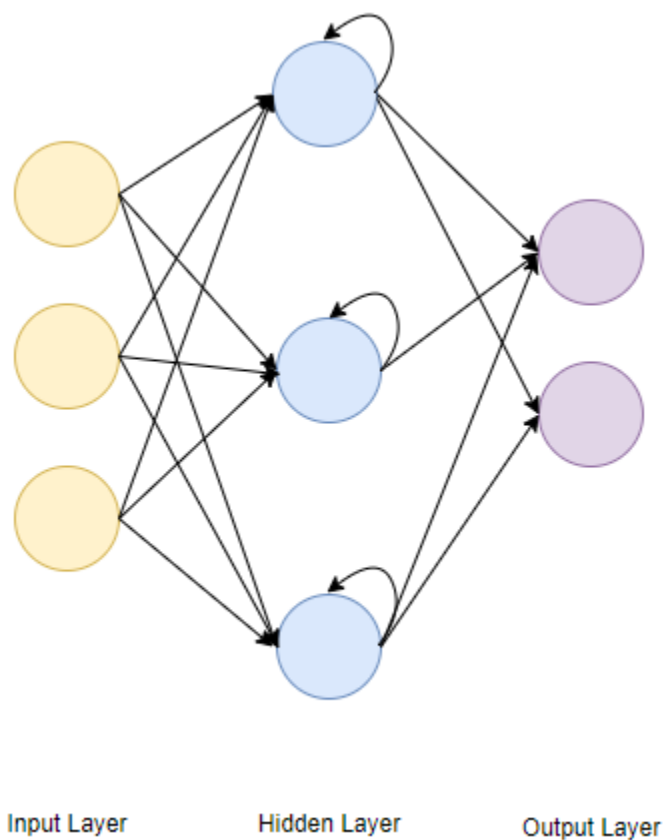


Σχήμα 3.5 : Αρχιτεκτονική Δικτύου Πρόσθιας Τροφοδότησης Πολλαπλών Επιπέδων

### 3.5.4 Αναδρομικά Δίκτυα

Τα αναδρομικά νευρωνικά δίκτυα (recurrent neural networks) ακολουθούν διαφορετική προσέγγιση της αρχιτεκτονικής των δικτύων πρόσθιας τροφοδότησης. Εισήχθησαν για πρώτη φορά τη δεκαετία του 1990 και έχουν σχεδιαστεί για εφαρμογές σε σειριακά ή χρονικά εξαρτώμενα πρότυπα. Εν αντιθέσει με τα δίκτυα πρόσθιας τροφοδότησης που περιγράφηκαν παραπάνω, καθιστούν ικανή την τροφοδότηση του σήματος εξόδου προς τα πίσω, διαθέτοντας τουλάχιστον ένα βρόχο ανάδρασης. Επιπλέον, δίνεται η δυνατότητα αυτοανάδρασης, δηλαδή η έξοδος ενός νευρώνα να λαμβάνεται και ως είσοδός του. Η λογική αναφορικά με τα κρυφά επίπεδα παραμένει η ίδια με αυτή των δικτύων πρόσθιας τροφοδότησης, συνεπώς μπορεί να γίνει διάκριση ανάμεσα σε δίκτυα μοναδικού επιπέδου και δίκτυα πολλαπλών επιπέδων, ενώ αντίστοιχη διάκριση μπορεί να γίνει βάσει πλήρους ή μερικούς διασύνδεσης των κόμβων μεταξύ τους.

Τα αναδρομικά δίκτυα έχουν λιγότερη απήχηση σε σύγκριση με τα πρόσθιας τροφοδότησης, λόγω του γεγονότος ότι οι αλγόριθμοι μάθησης που εφαρμόζονται σε αυτά είναι λιγότερο ισχυροί από αυτούς των feedforward δικτύων, διατηρούν ωστόσο έντονο το ενδιαφέρον της επιστημονικής κοινότητας, καθώς είναι πιο κοντά στον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου και θεωρείται ότι μπορούν να αντιμετωπίσουν προβλήματα στα οποία τα feedforward δίκτυα αντιμετωπίζουν δυσκολίες.



Σχήμα 3.6 : Αρχιτεκτονική Αναδρομικών Δικτύων



### **3.6 Μηχανική Μάθηση**

Η ομοιότητα των νευρωνικών δικτύων με τον ανθρώπινο εγκέφαλο εντοπίζεται και στον τρόπο με τον οποίο μαθαίνουν να αναγνωρίζουν πρότυπα στα δεδομένα που λαμβάνουν ως είσοδο. Οι διαδικασίες μέσω των οποίων επιτυγχάνεται η προαναφερθείσα μάθηση μπορούν να διακριθούν σε δύο κατηγορίες βάσει της ύπαρξης ή μη εκπαιδευτή.

#### **3.6.1 Μάθηση με Εκπαιδευτή**

Η μάθηση με εκπαιδευτή αναφέρεται και ως επιβλεπόμενη μάθηση (supervised learning) και αποτελείται από δύο μέρη: τον εκπαιδευτή και το νευρωνικό δίκτυο. Ο εκπαιδευτής θεωρείται ότι αντιπροσωπεύει τη γνώση που σχετίζεται με το περιβάλλον μέσω ενός συνόλου παραδειγμάτων εισόδου-εξόδου, περιβάλλον το οποίο αγνοείται από το νευρωνικό δίκτυο. Ο ρόλος του εκπαιδευτή συνίσταται στην παροχή στο νευρωνικό δίκτυο μιας επιθυμητής απόκρισης για το διάνυσμα που τίθεται ως είσοδος σε αυτό. Παράλληλα, το δίκτυο καθαυτό παράγει τη δική του πραγματική απόκριση για την απόκριση που δέχεται ως είσοδο. Κατά τη διαδικασία της εκπαίδευσης, το σήμα εξόδου του δικτύου τροποποιείται μέσω της προσαρμογής των συναπτικών βαρών του, τα οποία λειτουργούν ως ελεύθερες παράμετροι, σύμφωνα πάντα με την επιρροή του διανύσματος εκπαίδευσης και ενός σήματος σφάλματος. Το σήμα σφάλματος ορίζεται ως η διαφορά μεταξύ της πραγματικής απόκρισης του δικτύου και της επιθυμητής απόκρισης (έξοδος του εκπαιδευτή).

Η προαναφερθείσα διαδικασία εκτελείται επαναληπτικά, με στόχο την παραγωγή πραγματικής εξόδου όμοιας με την επιθυμητή. Στην ουσία πρόκειται για μία διαδικασία μεταφοράς γνώσης από τον εκπαιδευτή προς το δίκτυο η οποία αποθηκεύεται με τη μορφή των συναπτικών βαρών και αντιπροσωπεύει τη μακροπρόθεσμη μνήμη. Το κριτήριο για τον καθορισμό της επίτευξης ή μη της σύγκλισης, είναι στατιστικής φύσης, όπως επί παραδείγματι το μέσο τετραγωνικό σφάλμα. Με την επίτευξη της σύγκλισης, το νευρωνικό δίκτυο αποδεσμεύεται από τον εκπαιδευτή, καθώς θεωρείται πλέον ικανό να αντιμετωπίσει οποιεσδήποτε μεταβολές οφείλονται σε εξωγενείς παράγοντες του περιβάλλοντός του.

Δεδομένων των προαναφερθέντων, γίνεται σαφές το γεγονός ότι η επιβλεπόμενη μάθηση πρόκειται για μία διαδικασία μάθησης μέσω διόρθωσης σφαλμάτων και μπορεί να οριστεί ως μία συνάρτηση των ελευθέρων βαρών του δικτύου, οι οποίες αποτελούν τις ανεξάρτητες μεταβλητές. Η εν λόγω διαδικασία μάθησης μπορεί να αποδοθεί με τη μορφή μιας πολυδιάστατης επιφάνειας σφάλματος, τα σημεία της οποίας αναπαριστούν μία συγκεκριμένη λειτουργία του συστήματος. Στόχος είναι η διαδοχική μετατόπιση των βαρών με τρόπο ώστε το σύστημα να κινείται προς ένα τοπικό ή ολικό ελάχιστο (η κατάσταση ολικού ελαχίστου προτιμάται σε σχέση με αυτή του τοπικού). Η πληροφορία που σχετίζεται με την κατεύθυνση που πρέπει να ακολουθηθεί για την προσέγγιση του παραπάνω ελαχίστου, λαμβάνεται μέσω της κλίσης (gradient) της επιφάνειας σφάλματος που αντιστοιχεί στην τρέχουσα κατάσταση του συστήματος. Στην πράξη, η ακριβής τιμή της κλίσης αγνοείται, με αποτέλεσμα να χρησιμοποιείται μία εκτίμησή της, αλλά με δεδομένη τη χρήση ενός ειδικά σχεδιασμένου

αλγορίθμου, την επάρκεια ζευγών εισόδου-εξόδου και ικανό αριθμό επαναλήψεων, η προσέγγιση αγνώστων ζευγών εισόδου-εξόδου γίνεται με ιδιαίτερα ικανοποιητικό τρόπο. Υπάρχει πληθώρα τεχνικών επιβλεπόμενης μάθησης όπως λόγου χάριν η Naive Bayes ή τα Support Vector Machines.

### ***3.6.2 Μάθηση Χωρίς Εκπαιδευτή***

Όπως υποδηλώνει και το όνομα αυτού του είδους μάθησης, η διαδικασία δεν επιβλέπεται από κάποιον συντελεστή που να έχει το ρόλο του εκπαιδευτή. Η ειδοποιός αυτή διαφορά συνεπάγεται την απουσία των χαρακτηριστικών παραδειγμάτων επιθυμητής απόκρισης βάσει των οποίων εκπαιδύεται το δίκτυο, με αποτέλεσμα σχεδόν κάθε διαδικασία μάθησης χωρίς εκπαιδευτή να μπορεί να θεωρηθεί ως ένα πιθανοτικό μοντέλο των δεδομένων. Η κατηγορία αυτή, μπορεί με τη σειρά της να διακριθεί σε δύο υποκατηγορίες: την ενισχυτική και τη μη επιβλεπόμενη μάθηση, οι οποίες θα περιγραφούν ακολούθως.

#### ***3.6.2.1 Ενισχυτική Μάθηση***

Κατά την ενισχυτική μάθηση, η εκπαίδευση λαμβάνει χώρα πάνω σε ένα ζεύγος εισόδου-εξόδου μέσω της συνεχούς αλληλεπίδρασης του δικτύου με το περιβάλλον του και αποσκοπεί στην ελαχιστοποίηση ενός βαθμωτού δείκτη απόδοσης. Το ρόλο του εκπαιδευτή στην περίπτωση αυτή παίζει ένας μηχανισμός ο οποίος λειτουργεί ως κριτής. Αναλυτικότερα, ο κριτής παρατηρεί μία χρονική ακολουθία ερεθισμάτων του εξωτερικού περιβάλλοντος (σήμα) την οποία και μετατρέπει σε ένα σήμα υψηλότερης ποιότητας που καλείται ευρετικό σήμα ενίσχυσης. Στόχος της διαδικασίας είναι η εύρεσης μιας ακολουθίας βημάτων ικανής να ελαχιστοποιήσει τη συνάρτηση τρέχοντος κόστους. Η συνάρτηση τρέχοντος κόστους δύναται να οριστεί ως πρόβλεψη του αθροιστικού κόστους των ενεργειών που εκτελούνται σε μια αλληλουχία βημάτων. Από την ανωτέρω περιγραφή, αναδεικνύεται ένα σημαντικό μειονέκτημα της μεθόδου ενισχυτικής μάθησης, που αφορά την εισαγωγή χρονικής καθυστέρησης, καθώς η ανατροφοδότηση (ενίσχυση) του ευρετικού σήματος είναι φύσει αδύνατον να γίνει χωρίς χρονικό κόστος. Συνεπώς το δίκτυο θα πρέπει να αντιμετωπίσει ένα επιπρόσθετο πρόβλημα που σχετίζεται με την ανάθεση εμπιστοσύνης, ήτοι θα πρέπει να βρίσκεται σε θέση να καθορίζει το βαθμό επιτυχίας για κάθε ενέργεια της ακολουθίας βημάτων που εφαρμόστηκε, με τον κύριο μηχανισμό ενίσχυσης να περιορίζεται στην αποτίμηση μόνο του τελικού αποτελέσματος.

#### ***3.6.2.2 Μη Επιβλεπόμενη Μάθηση***

Η διαδικασία της μη επιβλεπόμενης μάθησης, ελλείψει εξωτερικού εκπαιδευτή ή κριτή, βασίζεται στην ύπαρξη ενός ανεξάρτητου μέτρου ποιότητας το οποίο καλείται να μάθει το νευρωνικό δίκτυο και σύμφωνα με το οποίο βελτιστοποιεί τις ελεύθερες παραμέτρους του. Έτσι, είναι σε θέση να σχηματίζει εσωτερικές αναπαραστάσεις που κωδικοποιούν τα δεδομένα εκπαίδευσης και να βρίσκει τα χαρακτηριστικά τους με σκοπό τη δημιουργία νέων κλάσεων. Για την υλοποίηση της προσέγγισης αυτής, ακολουθείται η λογική της ανταγωνιστικής μάθησης, με τη χρήση για παράδειγμα δύο επιπέδων στο δίκτυο, εκ των οποίων το ένα αποτελεί το επίπεδο

εισόδου και το δεύτερο περιλαμβάνει νευρώνες οι οποίοι ανταγωνίζονται μεταξύ τους για να αποκτήσουν τη δυνατότητα να αποκριθούν στα δεδομένα εισόδου. Πρόκειται για στρατηγική τύπου “winner takes all”, καθώς μόνο ένας νευρώνας επικρατεί και ενεργοποιείται κάθε φορά, γεγονός που συνεπάγεται την απενεργοποίηση όλων των υπολοίπων.

### 3.6.6 Ο αλγόριθμος της Οπισθοδιάδοσης (*Backpropagation*)

Ο αλγόριθμος του Backpropagation είναι μία από τις δημοφιλέστερες μεθόδους μάθησης, λόγω της απλότητας αλλά και συνάμα των δυνατοτήτων του, ενώ αποτελεί ιδανική επιλογή για μη γραμμικά δίκτυα με τυχαίες συνδέσεις μεταξύ των κόμβων, δίκτυα τα οποία απαντώνται στην πλειοψηφία των προβλημάτων που καλούνται να αντιμετωπίσουν τα νευρωνικά δίκτυα. Η ονομασία του αλγορίθμου προέκυψε από τον όρο που χρησιμοποίησε ο Rosenblatt, ο οποίος έπειτα από τη δημιουργία του Perceptron, προσπάθησε τις δεκαετίες του 1960 και 1970, να γενικεύσει τη λογική του απλού αυτού δικτύου, σε περισσότερα επίπεδα.

Εν γένει, η διαδικασία με την οποία εκπαιδεύεται ένα δίκτυο πρόσθιας τροφοδότησης όπως αυτό που μελετήθηκε για τους σκοπούς της παρούσας διπλωματικής εργασίας, συνίσταται στην ελαχιστοποίηση του σφάλματος μεταξύ πραγματικής και επιθυμητής εξόδου, που αποτελεί τη συνάρτηση απώλειας της εκπαίδευσης. Για τον προσδιορισμό των παραμέτρων που ελαχιστοποιούν τη συνάρτηση απώλειας, τα συναπτικά βάρη προσαρμόζονται παράδειγμα προς παράδειγμα. Η καινοτομία του αλγορίθμου Backpropagation έγκειται στο γεγονός ότι παράλληλα με την προσαρμογή των βαρών, καθιστά εφικτό τον υπολογισμό του σφάλματος που οφείλεται στα βάρη κάθε νευρώνα, συμπεριλαμβανομένων και των νευρώνων που ανήκουν στα “κρυφά” επίπεδα του δικτύου. Η προαναφερθείσα διαδικασία γίνεται με αναδρομικό τρόπο, και δύναται να διακριθεί σε δύο φάσεις ανάλογα με την κατεύθυνση προς την οποία εξελίσσονται οι ενέργειες:

- Προς τα εμπρός, φάση κατά την οποία τα συναπτικά βάρη παραμένουν σταθερά και το σήμα εισόδου διαδίδεται μέσω των διαδοχικών επιπέδων του δικτύου έως ότου φτάσει στην έξοδο, με αποτέλεσμα οι αλλαγές να περιορίζονται στα δυναμικά ενεργοποίησης και τις εξόδους των νευρώνων. Σε κάθε επίπεδο υπολογίζονται τα λειτουργικά σήματα του δικτύου και λαμβάνεται το σήμα εξόδου από το τελευταίο επίπεδο. Συνοπτικότερα, το δiάνυσμα εισόδου παρουσιάζεται στο πρώτο κρυφό επίπεδο, διασχίζει όλο το εύρος του δικτύου και καταλήγει στην έξοδο όπου υπολογίζεται το σφάλμα κάθε νευρώνα σε σχέση με την επιθυμητή απόκριση
- Προς τα πίσω, φάση κατά την οποία γίνεται η παραγωγή του σήματος σφάλματος. Το σήμα αυτό είναι αποτέλεσμα της σύγκρισης μεταξύ της τιμής εξόδου του δικτύου και της επιθυμητής απόκρισης στο σήμα εισόδου. Η διάδοση του σήματος σφάλματος γίνεται επίσης επίπεδο προς επίπεδο, ακολουθώντας όμως κατεύθυνση προς τα πίσω. Εν συνεχεία το δίκτυο αξιοποιεί το σήμα σφάλματος για την προσαρμογή των συναπτικών βαρών των κόμβων του.

Η διόρθωση των βαρών μεταξύ δύο νευρώνων που περιγράφηκε προηγουμένως γίνεται βάσει του αποκαλούμενου κανόνα Δέλτα, η μαθηματική διατύπωση του οποίου είναι η ακόλουθη:

$$\Delta w_{ji}(n) = \eta \cdot \delta_j(n) \cdot y_i(n)$$

όπου

- $\Delta w_{ji}(n)$  η διόρθωση του βάρους κατά τη n-οστή επανάληψη
- $\eta$  η παράμετρος του ρυθμού μάθησης
- $\delta_j(n)$  η τοπική κλίση του νευρώνα j
- $y_i(n)$  το σήμα εισόδου j, το οποίο αποτελεί στην ουσία το σήμα εξόδου του νευρώνα i

Το τοπικό πεδίο ενός νευρώνα που αποτελεί μέρος ενός ευρύτερου νευρωνικού δικτύου δίνεται από τη σχέση:

$$u_j(n) = \sum_{i=0}^m w_{ji}(n) y_i(n)$$

ενώ η έξοδος του περιγράφεται μέσω του ακόλουθου τύπου:

$$y_j(n) = \phi_j(u_j(n))$$

Από τον ορισμό της διόρθωσης βαρών που πραγματοποιεί ο αλγόριθμος Backpropagation, γίνεται εμφανής η αναλογία τους με τη μερική παράγωγο της ενέργειας σφάλματος του νευρώνα. Η τελευταία μπορεί να θεωρηθεί ως ένας συντελεστής ευαισθησίας ο οποίος υποδεικνύει την κατεύθυνση αναζήτησης του συναπτικού βάρους του εκάστοτε κόμβου στο χώρο βαρών.

Όσον αφορά την τοπική κλίση, υπάρχει άμεση εξάρτησή της με το είδος του νευρώνα και πιο συγκεκριμένα το αν πρόκειται για κρυφό κόμβο ή για κόμβο εξόδου. Στην πρώτη περίπτωση η τοπική κλίση θα ισούται με το γινόμενο του σήματος σφάλματος  $e_j(n)$  και της παραγώγου  $\phi'_j(u_j(n))$ , ενώ το σφάλμα εξόδου ορίζεται ως η διαφορά της επιθυμητής απόκρισης από την πραγματική, ήτοι  $e_j(n) = d_j(n) - y_j(n)$ . Στην περίπτωση όμως που ο νευρώνας είναι κόμβος εξόδου, τότε το σφάλμα ορίζεται με αναδρομικό τρόπο, βάσει των σφαλμάτων όλων των νευρώνων με τους οποίους συνδέεται άμεσα ο κόμβος j. Ως εκ τούτου, η τοπική κλίση υπολογίζεται από τη σχέση :

$$\delta_j(n) = \phi'_j(u_j(n)) \sum_k \delta_k(n) w_{kj}(n)$$

δηλαδή από το γινόμενο της σχετιζόμενης παραγώγου με το άθροισμα των τοπικών κλίσεων όλων των νευρώνων που ανήκουν στο αμέσως βαθύτερο επίπεδο και συνδέονται άμεσα με το νευρώνα j.

Για να είναι εφαρμόσιμες οι ανωτέρω σχέσεις, απαραίτητη προϋπόθεση που πρέπει να πληροί η συνάρτηση ενεργοποίησης είναι να είναι συνεχής. Ο λόγος που η διαφορισιμότητα της συνάρτησης ενεργοποίησης αποτελεί καθοριστικό παράγοντα είναι ότι η γνώση της παραγώγου

της είναι αναγκαία για τη χρήση της στους υπολογισμούς των μεγεθών που αναφέρθηκαν παραπάνω.

Αντίστοιχης βαρύτητας είναι η ορθή επιλογή της παραμέτρου μάθησης ( $\eta$ ), καθώς αυτή καθορίζει την ταχύτητα της μάθησης. Αναλυτικότερα, όσο μικρότερες τιμές λαμβάνει η παράμετρος, τόσο αναλογικά μικρότερες θα είναι οι αλλαγές στα συναπτικά βάρη του δικτύου έπειτα από κάθε επανάληψη και τόσο αντιστρόφως ανάλογα θα αυξάνεται το χρονικό κόστος του αλγορίθμου. Ωστόσο, η “τροχιά” αναζήτησης και επιλογής των βέλτιστων συναπτικών βαρών από το χώρο των βαρών θα είναι αρκετά ομαλότερη. Στον αντίποδα, η επιλογή μεγάλης παραμέτρου μάθησης ενδέχεται να οδηγήσει το σύστημα σε αστάθεια εξαιτίας των μεγάλων αλλαγών που προκύπτουν με το πέρασ κάθε επανάληψης. Συνεπώς, επιλέγεται η χρυσή τομή μεταξύ ομαλούς σύγκλισης και ταχύτητας μάθησης.

Η σύγκλιση του αλγορίθμου δεν είναι δυνατόν να αποδειχθεί, καθώς επίσης δεν υπάρχουν καλά ορισμένα κριτήρια για τον τερματισμό του. Ελλείψει των προαναφερθέντων, χρησιμοποιούνται ορισμένα λογικά κριτήρια βάσει των οποίων τερματίζεται η διαδικασία προσαρμογής των συναπτικών βαρών των κόμβων του δικτύου. Ο προσδιορισμός των κριτηρίων τερματισμού γίνεται αξιοποιώντας τις ιδιότητες των τοπικών και του ολικού ελαχίστου της εκάστοτε επιφάνειας σφάλματος.

Έστω ότι το διάνυσμα βαρών  $w^*$  συμβολίζει ένα τοπικό ή ολικό ελάχιστο. Απαραίτητη συνθήκη για να ισχύει το προαναφερθέν είναι η κλίση της επιφάνειας σφάλματος ως προς τα βάρη, ήτοι η μερική παράγωγος πρώτης τάξης, να είναι μηδενική στο συγκεκριμένο σημείο, με αποτέλεσμα μία διατύπωση ενός κριτηρίου τερματισμού να είναι η εξής:

*“Ο αλγόριθμος BK θεωρείται ότι συγκλίνει όταν η Ευκλείδεια νόρμα του διανύσματος κλίσης φτάσει σε ένα επαρκώς μικρό κατώφλι”*

Το παραπάνω κριτήριο έχει το μειονέκτημα των μεγάλων χρόνων μάθησης για την περίπτωση των επιτυχημένων δοκιμών, καθώς και της απαίτησης υπολογισμού των διανυσμάτων κλίσης. Για το λόγο αυτό, θα μπορούσε να αξιοποιηθεί μία άλλη ιδιότητα των ελαχίστων και συγκεκριμένα το γεγονός ότι η συνάρτηση κόστους στο σημείο αυτό είναι στατική. Ως εκ τούτου, μπορεί να προταθεί το ακόλουθο κριτήριο τερματισμού:

Ο αλγόριθμος BK θεωρείται ότι συγκλίνει όταν ο ρυθμός μεταβολής της συνάρτησης κόστους ανά εποχή είναι επαρκώς μικρός.

Ο χαρακτηρισμός “επαρκώς μικρός” περιγράφει μεταβολές που κυμαίνονται από 0.1-1% ανά εποχή. Τέλος, το μειονέκτημα του κριτηρίου αυτού είναι ο ενδεχόμενος πρώιμος τερματισμός της διαδικασίας μάθησης.

Στο πλαίσιο των πειραμάτων που διενεργήθηκαν, έγινε χρήση μάθησης με εκπαιδευτή. Περισσότερες πληροφορίες αναφορικά με την υλοποίησή του παρατίθενται στην αντίστοιχη ενότητα της παρούσας εργασίας.



## ΚΕΦΑΛΑΙΟ 4

### Παρουσίαση του GluonTS

#### 4.1 Εισαγωγή

Τα τελευταία χρόνια, μία υποκατηγορία της μηχανικής μάθησης, τα δίκτυα βαθιάς μάθησης (deep learning) έχουν σημειώσει αλματώδη πρόοδο με τη χρήση τους να γίνεται όλο και πιο δημοφιλής. Το deep learning συνίσταται από νευρωνικά δίκτυα δύο ή τριών επιπέδων, τα οποία μιμούμενα τη λειτουργία του ανθρώπινου εγκεφάλου, μαθαίνουν από μεγάλη ποσότητα δεδομένων. Η ικανότητα διαχείρισης ενός τέτοιου όγκου δεδομένων και η εκμάθηση από αυτά, τα καθιστά ιδανικά για εφαρμογές που σχετίζονται με τεχνολογίες αιχμής όπως αναγνώριση εικόνων, λόγου και κειμένου, καθώς και στην αυτόνομη οδήγηση, χωρίς την παρέμβαση του ανθρώπινου παράγοντα.

Η ειδοποιός διαφορά της βαθιάς μάθησης από τη μηχανική, είναι ο τύπος των δεδομένων που διαχειρίζεται καθώς επίσης και οι μέθοδοι με τις οποίες μαθαίνουν τα δίκτυα που ανήκουν σε αυτή. Οι κατηγορίες μεθόδων εκμάθησης στα δίκτυα μηχανικής μάθησης που περιγράφηκαν στην αντίστοιχη παράγραφο της παρούσας διπλωματικής εργασίας, απαιτούν εν γένει μία προεπεξεργασία των δεδομένων για να οργανωθούν σε μία δομημένη μορφή, όπως επί παραδείγματι αυτής ενός πίνακα.

Η βαθιά μάθηση είναι ικανή να διαχειριστεί και να επεξεργαστεί μη δομημένα δεδομένα, όπως κείμενο ή εικόνα, αυτοματοποιώντας την εξαγωγή χαρακτηριστικών για τα δεδομένα αυτά. Χαρακτηριστικό παράδειγμα που αναδεικνύει την προαναφερθείσα ικανότητα, αποτελεί η διάκριση ενός σετ φωτογραφιών κατοικιδίων ανάλογα με το ζώο που απεικονίζεται σε αυτές, όπως “σκύλος”, “γάτα”, “χάμστερ” κλπ. Στην περίπτωση αυτή, οι αλγόριθμοι βαθιάς μάθησης είναι ικανοί να αποφανθούν αναφορικά με το ποια χαρακτηριστικά είναι ικανά να διαχωρίσουν το ένα ζώο από το άλλο, όπως λόγου χάριν τα αυτιά. Η χρήση αλγορίθμων μηχανικής μάθησης σε αυτή την εφαρμογή θα απαιτούσε τον ρητό καθορισμό της ιεραρχίας των χαρακτηριστικών αυτών από τον άνθρωπο.

Τα παραπάνω οφέλη από τη χρήση της βαθιάς μάθησης έχουν το τίμημα της απαίτησης εξαιρετικά μεγάλης υπολογιστικής ισχύος και αντίστοιχα μεγάλης διαθεσιμότητας δεδομένων. Από πλευράς υλικού, αντί για τις CPUs (Central Processing Units), χρησιμοποιούνται οι GPUs (Graphical Processing Units), οι οποίες είναι ικανές να διαχειριστούν μεγάλο όγκο υπολογισμών σε πολλαπλούς πυρήνες, παρέχοντας παράλληλα άφθονη μνήμη. Είναι προφανές ότι οι προαναφερθείσες απαιτήσεις σε πόρους, μπορούν να οδηγήσουν σε πολύ υψηλά κόστη.

Η ικανότητα αναγνώρισης προτύπων που διαθέτουν οι αλγόριθμοι βαθιάς μάθησης, τα οδήγησε στην επέκτασή τους σε κλάδους όπως οι τεχνικές προβλέψεων. Η χρήση τους ωστόσο σε εφαρμογές του “πραγματικού κόσμου” όπως η πρόβλεψη χρονοσειρών παρέμεινε περιορισμένη εξαιτίας της έλλειψης των απαιτούμενων πόρων αλλά κυρίως λόγω της απουσίας κατάλληλων και φιλικών προς το χρήστη πακέτων, με το σκηνικό να αλλάζει τα τελευταία χρόνια έπειτα και από την εισαγωγή τους στην κοινότητα. Η χρήση ενός τέτοιου πακέτου, του GluonTS, το οποίο θα αναλυθεί εκτενέστερα σε επόμενη ενότητα του παρόντος κεφαλαίου, αποτέλεσε τον πυρήνα του πειραματικού μέρους της παρούσας διπλωματικής εργασίας.

Εν γένει, κριτήρια επιλογής του καταλληλότερου πακέτου για τους σκοπούς της πρόβλεψης των χρονοσειρών θεωρούνται τα ακόλουθα:

- η απαιτούμενη εμπειρία σε ανάπτυξη μοντέλων βαθιάς μάθησης και η προσπάθεια που χρειάζεται ο πειραματισμός με τις υπερπαραμέτρους τους
- η ευχέρεια που παρέχεται στο χρήστη να μεταβάλλει τις υπερπαραμέτρους δικτύου και εκπαίδευσης
- οι λειτουργικότητες και τα επιπρόσθετα χαρακτηριστικά που παρέχονται και αφορούν την πρόβλεψη χρονοσειρών

Η πολυπλοκότητα των μοντέλων βαθιάς μάθησης απαιτεί αναπόφευκτα ένα trade-off μεταξύ της ευκολίας χρήσης τους και της ευελιξίας τους όσον αφορά τη ρύθμιση των υπερπαραμέτρων τους. Δεδομένου του προαναφερθέντος, τα διαθέσιμα εμπορικά πακέτα μπορούν να διακριθούν αφενός στην προσέγγιση “μαύρου κουτιού”, στα οποία περιλαμβάνονται εργαλεία όπως το Auto-Keras (DataLab , Texas A&M University) και το AutoML (Google), στα οποία ο χρήστης αγνοεί την εσωτερική δομή και λειτουργία και αφετέρου στην προσέγγιση της ευκολίας εξατομίκευσης, όπως το Tensorflow (Google) και το CNTK (Microsoft) τα οποία απαιτούν εμπειρία και χρόνο από το χρήστη για την κατά το δοκούν ρύθμιση των μοντέλων. Αναφορικά με το τρίτο κριτήριο επιλογής, αξίζει να αναφερθεί το γεγονός ότι παρότι αρκετά πακέτα παρέχουν προεκπαιδευμένα μοντέλα και επιπρόσθετες λειτουργικότητες, έτοιμα σετ δεδομένων και συναρτήσεις απωλειών για ποικίλες εφαρμογές, δεν τα παρέχουν για εφαρμογές πρόβλεψης χρονοσειρών, απαιτώντας περαιτέρω ενέργειες από πλευράς χρήστη. Παρακάτω, θα αναφερθεί ο τρόπος με τον οποίο αντιμετωπίστηκαν τέτοιου είδους δυσκολίες κατά το στήσιμο της πειραματικής διαδικασίας στο πλαίσιο της εν λόγω διπλωματικής εργασίας.

Εξετάζοντας τους προαναφερθέντες παράγοντες, το πακέτο που επιλέχθηκε να μελετηθεί είναι το GluonTS, ένα σχετικά πρόσφατο εργαλείο (πρώτη κυκλοφορία το Μάρτιο του 2019) που αναπτύχθηκε από την Amazon Web Services με σκοπό την πλήρωση της ανάγκης ενός εύκολου προς χρήση πακέτου πρόβλεψης χρονοσειρών, μέσω της παροχής έτοιμων μοντέλων πρόβλεψης και της δυνατότητας ανάπτυξης νέων εξατομικευμένων μοντέλων. Το GluonTS διανέμεται ως βιβλιοθήκη της γλώσσας Python και έχει δημιουργηθεί για πιθανοτικές προβλέψεις, ήτοι αποτελεί ιδανική λύση για το σκοπό της προκειμένης διπλωματικής εργασίας, ενώ παράλληλα



παρέχει τη δυνατότητα για σημειακές προβλέψεις. Παρά τη χρονικά μικρή παρουσία του, έγινε γρήγορα αποδεκτό από την κοινότητα.

## 4.2 Λειτουργία GluonTS

Η λειτουργία του GluonTS βασίζεται στη διάδραση των συστατικών μερών του, τα οποία θα περιγραφούν στην παρούσα ενότητα. Εν συντομία, το GluonTS απαιτεί συγκεκριμένη δομή των δεδομένων που να περιλαμβάνουν πληροφορίες για την κάθε χρονοσειρά, δεδομένα τα οποία αποτελούν την είσοδο ενός μοντέλου που ενδέχεται να είναι ήδη υλοποιημένο.. Εν συνέχεια εξάγει μία κατανομή τιμών για καθένα από τα βήματα του χρονικού ορίζοντα πρόβλεψης για όλες τις χρονοσειρές. Η έξοδος αυτή δεν αποτελεί την κατανομή των παραμέτρων καθαυτών, αλλά ένα set τυχαία επιλεγμένων δειγμάτων από την προβλεφθείσα κατανομή.

Όπως αναφέρθηκε παραπάνω, τα μοντέλα του GluonTS αξιοποιούν πληροφορία από δύο set δεδομένων τα οποία αφορούν την εκπαίδευση και την αξιολόγηση (testing), ‘dataset.train’ και ‘dataset.test’ αντίστοιχα. Επιπλέον, αξιοποιείται ένα σετ δεδομένων, το “dataset.metadata” που περιλαμβάνει πληροφορίες που σχετίζονται με τις χρονοσειρές των δύο προαναφερθέντων σετ δεδομένων. Οι πληροφορίες αυτές μπορούν επί παραδείγματι να αφορούν τη συχνότητα της κάθε χρονοσειράς, το μήκος της πρόβλεψης (prediction length) ή το μήκος πλαισίου (context length) (οι παράμετροι αυτές θα αναλυθούν εκτενέστερα στη συνέχεια). Τα σετ δεδομένων πρέπει να είναι σε μορφή jsonlines ή parquet.

Ο τρόπος με τον οποίο είναι δομημένο το GluonTS, του επιτρέπει να διαχειρίζεται ταυτόχρονα πολλαπλές χρονοσειρές, γεγονός που συνεπάγεται αύξηση του χρόνου εκπαίδευσης στην περίπτωση που λάβει ως είσοδο μία μόνο χρονοσειρά πολύ μεγάλου μήκους σε σχέση με την περίπτωση όπου το σετ δεδομένων αποτελείται από πολλές χρονοσειρές μικρότερου μήκους.

## 4.3 Είσοδος/Έξοδος Δεδομένων και Επεξεργασία

Η διαχείριση των δεδομένων γίνεται σε σειρές (streams ή Python iterators) δομών λεξιλογίων (dictionaries). Κατά την επεξεργασία των χαρακτηριστικών των χρονοσειρών, ένας DatasetReader επαναλαμβάνεται καθ όλη την έκταση του dataset και παράγει με τη σειρά του τα δικά του streams από dictionaries. Ως εκ τούτου, η επεξεργασία των χαρακτηριστικών συνίσταται από μία σειρά μετασχηματισμών (Transformations) οι οποίοι δρουν διαδοχικά στα παραχθέντα streams των dictionaries. Καθένας από τους μετασχηματισμούς αυτούς έχει τη δυνατότητα να προσθέσει ή να αφαιρέσει κλειδιά, να τροποποιήσει τιμές και να φιλτράρει ή να προσθέσει νέα αντικείμενα στο stream. Το GluonTS παρέχει ποικιλία μετασχηματισμών, όπως τον Box-Cox μετασχηματισμό, ενώ παράλληλα δίνει τη δυνατότητα στο χρήστη να αναπτύξει δικό του εξατομικευμένο μετασχηματισμό και να τον συμπεριλάβει στο pipeline, ανάλογα με το σκοπό που θέλει να επιτύχει.

## 4.4 Προβλέπτης

Όμοια με το scikit-learn και το SparkML, το GluonTS ακολουθεί τη λογική του stateless API. Έτσι ένας εκτιμητής (Estimator) περιλαμβάνει μία μέθοδο εκπαίδευσης, η οποία λαμβάνει ένα dataset εκπαίδευσης και επιστρέφει έναν προβλέπτη (Predictor). Με τη σειρά του ο Predictor εφαρμόζεται σε ένα test dataset και παράγει προβλέψεις. Τα κλασσικά προβλεπτικά μοντέλα τα οποία εκτιμούν παραμέτρους ανά χρονοσειρά προτού εξαγάγουν τις προβλέψεις, αποτελούν αμιγείς προβλέπτες στο εν λόγω API.

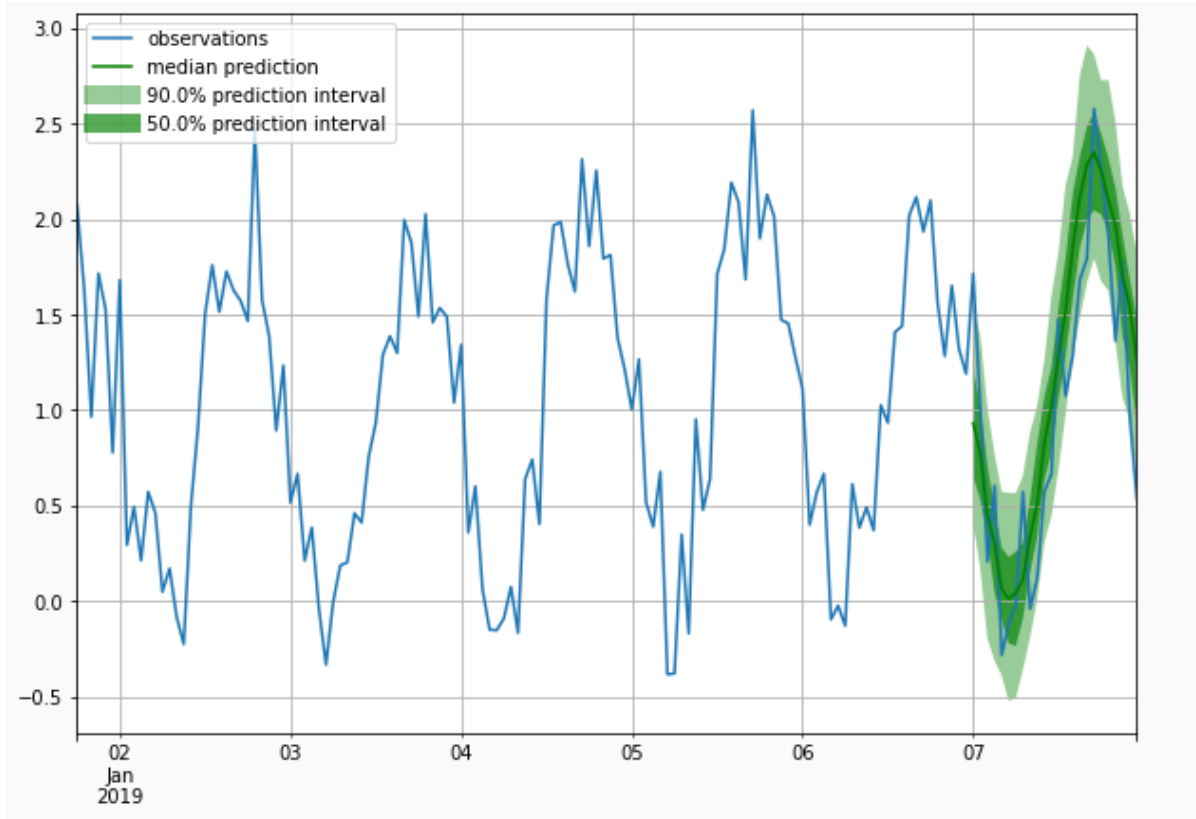
## 4.5 Κατανομή / Έξοδος

Όπως έχει ήδη αναφερθεί, αντικείμενο της εργασίας αυτής είναι οι πιθανοτικές προβλέψεις. Το GluonTS αποτελεί ένα ιδανικό εργαλείο για εξαγωγή πιθανοτικών προβλέψεων, καθώς παρέχει μια ευρεία γκάμα κατανομών (και πυκνοτήτων). Στα περισσότερα από τα μοντέλα που περιλαμβάνει (και τα οποία θα αναλυθούν εκτενέστερα στη συνέχεια), η έξοδος του νευρωνικού δικτύου είναι ένα διάνυσμα τιμών οι οποίες μετασχηματίζονται σε ένα διάνυσμα παραμέτρων της επιλεγμένης κατανομής. Η τελική έξοδος είναι τιμές η οποίες έχουν προκύψει από δειγματοληψία της παραμετροποιημένης, βάσει των παραμέτρων που προέκυψαν στο προηγούμενο βήμα, κατανομής.

## 4.6 Αντικείμενα Πρόβλεψης και Αξιολόγηση

Τα εκπαιδευμένα μοντέλα πρόβλεψης (Predictors) επιστρέφουν αντικείμενα πρόβλεψης (Forecast objects) με τη μορφή προβλέψεων οι οποίες περιέχουν έναν χρονικό δείκτη (που αφορά τη αρχή και το τέλος της χρονοσειράς) και μία αναπαράσταση της πιθανότητας κατανομής των τιμών για κάθε τιμή του χρονικού δείκτη. Ο υπολογισμός και η αναπαράσταση της από κοινού αυτής κατανομής ποικίλουν από μοντέλο σε μοντέλο.

Η ποσοτική αξιολόγηση της ακρίβειας των μοντέλων του GluonTS γίνεται με τη βοήθεια του Εκτιμητή (Evaluator), ο οποίος δύναται να εφαρμοστεί σε ένα stream από Forecast αντικείμενα και ένα stream από τιμές-στόχους (σε μορφή pandas dataframe). Ο Evaluator διαπερνά τα διανύσματα των πραγματικών τιμών και των προβλέψεων και υπολογίζει μετρικές (metrics) όπως την απώλεια ποσοστημορίου (quantile loss), με το MASE, MAPE και το sMAPE, που αποτελούν δείκτες ακριβείας που περιγράφονται στην ενότητα 2.10 της παρούσας διπλωματικής εργασίας. Τα metrics αυτά επιστρέφονται από τον Evaluator σε μορφή pandas dataframe, καθιστώντας εύκολη την οπτικοποίηση αλλά και την περαιτέρω επεξεργασία του από το χρήστη. Η οπτικοποίηση αυτή μπορεί να γίνει και μέσα από το εργαλείο του GluonTS καθώς περιλαμβάνει μεθόδους που οπτικοποιούν χρονοσειρές χρησιμοποιώντας τη βιβλιοθήκη matplotlib, όπως φαίνεται στο παρακάτω διάγραμμα.



Σχήμα 4.1 : Οπτικοποίηση προβλέψεων μέσω του GluonTS

Εκτός των προαναφερθέντων metrics τα οποία υπολογίζονται ανά τιμή κάθε φορά, ο Evaluator προσφέρει τη δυνατότητα υπολογισμού συναθροισμένων στατιστικών για όλο το dataset, όπως για παράδειγμα του Σταθμισμένου Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (weighted Mean Absolute Percentage Error - wMAPE) ή της σταθμισμένης απώλειας ποσοστημορίου (weighted quantile loss), metrics τα οποία είναι χρήσιμα για τις περιπτώσεις αυτόματου συντονισμού των μοντέλων.

Ιδιαίτερα χρήσιμο χαρακτηριστικό του GluonTS είναι ότι επιτρέπει τη δοκιμή των μοντέλων πρόβλεψης στα ιστορικά δεδομένα (Backtesting). Οι μέθοδοι με τις οποίες υπολογίζονται τα metrics κατά το backtesting (backtest\_metrics methods), αξιοποιούν έναν προ-εκπαιδευμένο Predictor ή εκπαιδεύουν έναν Estimator πάνω σε ένα dataset εκπαίδευσης και εν συνεχεία αξιολογούν το μοντέλο στο test dataset. Κατά τη διαδικασία αυτή, ένα επιπλέον εργαλείο, ο διαχωριστής (Splitter), χωρίζει το διαθέσιμο dataset σε τρία μέρη: το εκπαίδευσης (train), το αξιολόγησης (validation) και το δοκιμής (test), τα οποία μπορεί να αποτελούν απλά χρονικά σημεία ή κυλόμενα παράθυρα (rolling windows).

#### 4.7 Δημιουργία Μοντέλου Χρονοσειρών

Στο πλαίσιο της πειραματικής διαδικασίας της παρούσας διπλωματικής εργασίας, έγινε χρήση ενός τροποποιημένου μοντέλου από αυτά που παρέχει η εργαλειοθήκη του GluonTS (περαιτέρω περί της πειραματικής προσέγγισης στην αντίστοιχη ενότητα). Συνεπώς, αξιοποιήθηκε η

ευελιξία του GluonTS, του οποίου κυρίαρχος σκοπός είναι η δημιουργία νέων μοντέλων, απλά ακολουθώντας πιστά τη λογική του Estimator/Predictor API, αξιοποιώντας ένα σκελετό-πρότυπο μοντέλου. Έτσι η δημιουργία, για παράδειγμα, ενός προβλεπτικού μοντέλου βαθιάς μάθησης βασισμένου στον GluonForecastEstimator, μπορεί απλά να γίνει εφαρμόζοντας και ρυθμίζοντας τα ακόλουθα μέρη και μεθόδους:

- `training_network.cls`: το οποίο θα αποτελεί ένα νευρωνικό δίκτυο που επιστρέφει την απώλεια εκπαίδευσης του μοντέλου και καλείται κατά τη διάρκεια της εκπαίδευσης
- `pred_network.cls`: το οποίο χρησιμοποιεί τις ίδιες παραμέτρους με το δίκτυο εκπαίδευσης (`training_network`) αλλά όπως υποδηλώνει το όνομά του, θα επιστρέφει τα δείγματα πρόβλεψης
- `create_transformation`: που αποτελεί μέθοδο η οποία επιστρέφει το pipeline του μετασχηματισμού
- `forecast_cls`: το οποίο αποτελεί την κλάση πρόβλεψης που εμπλέκεται στην έξοδο του δικτύου πρόβλεψης

#### 4.8 Μαθηματικό Υπόβαθρο Λειτουργίας του GluonTS

Έστω  $Z = \{z_{i,1:T_i}\}_{i=1}^N$  ένα σετ χρονοσειρών με  $z_{i,1:T_i} := (z_{i,1}, z_{i,2}, \dots, z_{i,T_i})$  και  $z_{i,t} \in \mathbb{R}$  η τιμή της  $i$ -οστής χρονοσειράς τη χρονική στιγμή  $t$ . Θεωρούμε, χωρίς βέβαια αυτό να είναι απαραίτητο, ότι οι διαδοχικές παρατηρήσεις της εκάστοτε χρονοσειρές είναι ισαπέχουσες, ενώ ο μονάδες χρόνου (ώρες, ημέρες, εβδομάδες κλπ) μπορεί να διαφέρουν από χρονοσειρά σε χρονοσειρά. Επιπλέον οι χρονοσειρές δεν είναι απαραίτητο να είναι ευθυγραμμισμένες, υπό την έννοια ότι η πρώτη παρατήρηση (για  $t=1$ ) μπορεί να αναφέρεται σε διαφορετικό χρονικό σημείο για την περίπτωση της κάθε χρονοσειράς.

Έστω ότι  $X = \{x_{i,1:T_i}\}_{i=1}^N$  ένα σετ χρονικά μεταβαλλόμενων συντελεστών με  $x_{i,t} \in \mathbb{R}$ . Όπως αναφέρουν χαρακτηριστικά οι Hyndman και Athanasopoulos (2017), ο στόχος του forecasting είναι η πρόβλεψη της κατανομής πιθανότητας των μελλοντικών τιμών  $z_{i,T_{i+1}:T_{i+\tau}}$ , δεδομένων των παρελθοντικών τιμών της χρονοσειράς  $z_{i,1:T_i}$ , των συντελεστών  $x_{i,1:T_{i+\tau}}$  και των παραμέτρων  $\Phi$  του προβλεπτικού μοντέλου, δηλαδή

$$p(z_{i,T_{i+1}:T_{i+\tau}} \mid z_{i,1:T_i}, x_{i,1:T_{i+\tau}}; \Phi)$$

Στην περίπτωση του GluonTS τα σύνολα  $Z$  και  $X$ , αποτελούν τα δεδομένα εκπαίδευσης των μοντέλων. Εν γένει τα κλασσικά μοντέλα αποτελούν με ορισμένες εξαιρέσεις (Chapados, 2014) τοπικά μοντέλα, δηλαδή μαθαίνουν τις προαναφερθείσες παραμέτρους  $\Phi$  για κάθε χρονοσειρά ανεξάρτητα. Τα τελευταία χρόνια έχουν προταθεί άλλα μοντέλα, τα οικουμενικά (global) τα οποία μαθαίνουν τις παραμέτρους καθ' όλη την έκταση του dataset που λαμβάνουν ως είσοδο, εξάγοντας χαρακτηριστικά υψηλότερης τάξης, με αποτέλεσμα να είναι σε θέση να αναγνωρίζουν

πολυπλοκότερα μοτίβα ακόμα και από dataset τα οποία περιέχουν μη προεπεξεργασμένα δεδομένα. Βάσει του τρόπου με τον οποίο μοντελοποιείται ο στόχος  $Z$ , τα μοντέλα μπορούν να διακριθούν σε δύο κύριες κατηγορίες που περιλαμβάνουν τα:

- γεννητικά μοντέλα (generative models) με τη μοντελοποίηση που υλοποιούν να διατυπώνεται μαθηματικά  $p(z_{i,1:T_{i+\tau}} | x_{i,1:T_{i+\tau}}; \Phi)$
- μοντέλα διάκρισης (discriminative models) τα οποία μοντελοποιούν το στόχο βάσει της σχέσης  $p(z_{i,T_{i+1}:T_{i+\tau}} | z_{i,1:T_i}, x_{i,1:T_{i+\tau}}; \Phi)$

Αναφορικά με την πρώτη κατηγορία, γίνεται η υπόθεση ότι οι δοθείσες χρονοσειρές παράγονται από μία στοχαστική διαδικασία  $p(Z|X; \Phi)$  δεδομένων των συντελεστών  $X$ , με τη διαδικασία αυτή να έχει μία παραμετρική δομή με άγνωστες τις παραμέτρους  $\Phi$  του μοντέλου. Χαρακτηριστικά παραδείγματα μοντέλων που ανήκουν στην κατηγορία των generative models, αποτελούν τα ARIMA και το κλασικό ETS. Ο υπολογισμός των αγνώστων παραμέτρων γίνεται μέσω της μεγιστοποίησης της πιθανότητας των τιμών των παρατηρήσεων με την εφαρμογή τους στο μοντέλο, δεδομένων των συντελεστών του συνόλου  $X$ . Με την εκμάθηση των παραμέτρων  $\Phi$ , η πιθανότητα της εξίσωσης  $p(z_{i,T_{i+1}:T_{i+\tau}} | z_{i,1:T_i}, x_{i,1:T_{i+\tau}}; \Phi)$  εξάγεται μέσω της σχέσης  $p(Z|X; \Phi)$ . Εν αντιθέσει με τα στατιστικά μοντέλα (ARIMA, ETS κλπ) τα οποία μαθαίνουν τις παραμέτρους  $\Phi$  ανεξάρτητα για κάθε χρονοσειρά, τα νευρωνικά generative μοντέλα, εκφράζουν τις παραμέτρους  $\Phi$  ως συνάρτηση των βαρών που μοιράζονται μεταξύ όλων των χρονοσειρών του συνόλου των δεδομένων εκπαίδευσης.

Στον αντίποδα, τα discriminative μοντέλα μοντελοποιούν την κατανομή της εξίσωσης  $p(z_{i,T_{i+1}:T_{i+\tau}} | z_{i,1:T_i}, x_{i,1:T_{i+\tau}}; \Phi)$  απευθείας μέσω ενός νευρωνικού δικτύου. Συγκριτικά με τα generative, τα μοντέλα της κατηγορίας αυτής είναι περισσότερο ευέλικτα, καθώς κάνουν λιγότερες δομικές υποθέσεις και ως εκ τούτου μπορούν να χρησιμοποιηθούν σε μία ευρύτερη γκάμα προβλημάτων και εφαρμογών. Στην ουσία, τα generative μοντέλα υποθέτουν μία υποκείμενη διαδικασία παραγωγής δεδομένων, βρίσκοντας παραμέτρους  $\Phi$  ικανές να αναπαραστήσουν τις παρατηρήσεις, ενώ τα discriminative μοντέλα μοντελοποιούν απευθείας την υπό συνθήκη κατανομή του μέλλοντος βάσει των παρελθοντικών τιμών, βρίσκοντας κατά την εκπαίδευση παραμέτρους  $\Phi$  έτσι ώστε οι προβλέψεις να ταιριάζουν με τις μελλοντικές τιμές δεδομένων των παρελθοντικών (βλ *Σχήμα 4.2*). Τέλος τα discriminative μοντέλα μπορούν να διακριθούν σε περαιτέρω υποκατηγορίες όπως τα αυτοπαλινδομικά και τα δίκτυα πρόσθιας τροφοδότησης.



Σχήμα 4.2: Συγκριτικά οι διαφορές μεταξύ Generative και Discriminative Μοντέλων

Δεδομένων των προαναφερθέντων, ακολουθεί αναφορά σε ορισμένα έτοιμα μοντέλα που παρέχει η εργαλειοθήκη του GluonTS.

#### 4.8.1 Generative Μοντέλα

Τα Generative Μοντέλα του GluonTS μπορούν να διακριθούν σε δύο υποκατηγορίες: τα State Space Models (SSMs) και τα Deep State Space Models (γνωστά και ως DeepState). Όσον αφορά τα SSMs, παρέχουν ένα πλαίσιο μοντελοποίησης και εκμάθησης χρονοσειρών και μοντελοποιούν τη δομή των δεδομένων μέσω μία λανθάνουσας κατάστασης που μπορεί να χρησιμοποιηθεί για την κωδικοποίηση των ποιοτικών χαρακτηριστικών της εκάστοτε χρονοσειράς, όπως επί παραδείγματι της τάσης ή της εποχιακότητας.

Τα DeepState μοντέλα αποτελούν μία πιθανοτική προσέγγιση των προβλέψεων και βασίζονται στο συνδυασμό των SSMs μοντέλων με το deep learning. Αναλυτικότερα, παραμετροποιούν το γραμμικό SSM χρησιμοποιώντας ένα αναδρομικό νευρωνικό δίκτυο (Recurrent Neural Network - RNN), με τα βάρη μάθησης να εξάγονται από ένα σύνολο μη προπεξεργασμένων δεδομένων και των σχετικών τους συντελεστών, οι οποίοι μπορεί να είναι ανεξάρτητοι ή μη του χρόνου.

#### 4.8.2 Discriminative Μοντέλα

Τα Discriminative Μοντέλα αποτελούνται από δύο κύρια μέρη: τον κωδικοποιητή (encoder) και τον αποκωδικοποιητή (decoder) και βασίζονται στην προσέγγιση μάθησης σειράς προς σειρά (sequence-to-sequence). Το δίκτυο του encoder διαβάζει τις χρονοσειρές και κωδικοποιεί την πληροφορία που σχετίζεται με την ακολουθία σε μία λανθάνουσα κατάσταση. Στη συνέχεια, η κωδικοποιημένη λανθάνουσα πληροφορία μεταφέρεται στο δίκτυο του decoder όπου συνδυάζεται με τα χαρακτηριστικά που έχουν εξαχθεί για το εύρος πρόβλεψης, με αποτέλεσμα

την παραγωγή της πρόβλεψης. Η προσέγγιση αυτή μολονότι είναι πολυπλοκότερη και πιο απαιτητική όσον αφορά τη διαθεσιμότητα δεδομένων συγκριτικά με τα απλά αυτο-παλινδρομικά δίκτυα, προσφέρει στο χρήστη αρκετά πλεονεκτήματα. Καταρχάς τα μοντέλα αυτά είναι ικανά να διαχειρίζονται αλλαγές στο σύνολο των χαρακτηριστικών τόσο στο εύρος εκπαίδευσης όσο και στο εύρος πρόβλεψης, επιτρέποντας με αυτό τον τρόπο την αξιοποίηση συντελεστών οι τιμές των οποίων δεν είναι διαθέσιμες. Επιπρόσθετα, το γεγονός ότι δομικά τα μοντέλα αυτά δεν ακολουθούν την αυτο-παλινδρομική αρχιτεκτονική, τους προσφέρει τη δυνατότητα της μη συσσώρευσης σφάλματος στην περίπτωση μεγάλου χρονικού ορίζοντα πρόβλεψης, καθώς η πρόβλεψη του προηγούμενου βήματος δεν αποτελεί χαρακτηριστικό του παρόντος βήματος.

Συγκριτικά με τα αυτο-παλινδρομικά μοντέλα τα οποία αναπαριστούν πολύ ικανοποιητικά τη σειριακή φύση των χρονοσειρών, η προσέγγιση sequence-to-sequence ταιριάζει περισσότερο στην πολυπαραγοντική παλινδρόμηση. Το προαναφερθέν συνεπάγεται την ανάγκη εκ των προτέρων δημιουργίας του ορίζοντα πρόβλεψης στα sequence-to-sequence μοντέλα και την εκ νέου εκπαίδευση για την περίπτωση που είναι επιθυμητή η εξαγωγή προβλέψεων που υπερβαίνουν χρονικά τον ορίζοντα των  $\tau$  βημάτων. Γνωστά μοντέλα που εμπίπτουν στο πλαίσιο της sequence-to-sequence προσέγγισης αποτελούν τα:

- **Νευρωνικά Μοντέλα Παλινδρόμησης Τεταρτημορίου (Neural Quantile Regression Models - NQRM)** που αποτελούν τεχνική απευθείας πρόβλεψης συγκεκριμένου quantile μίας εξαρτημένης μεταβλητής. Πιο αναλυτικά, εντός του sequence-to-sequence πλαισίου, υλοποιείται ένας decoder η έξοδος του οποίου είναι ένα σύνολο quantiles για κάθε βήμα του ορίζοντα πρόβλεψης, ενώ η εκπαίδευση ενός τέτοιου μοντέλου γίνεται αξιοποιώντας τη συνάρτηση quantile απωλειών.
- **Transformer**, που αποτελεί ένα μοντέλο που έχει γνωρίσει μεγάλη απήχηση στις εφαρμογές αναγνώρισης φυσικής γλώσσας. Είναι ικανό να αναγνωρίζει τις εξαρτήσεις και να επιτρέπει την παραλληλοποίηση των υπολογισμών καθώς αυτοί γίνονται ανεξαρτήτων όλων των χρονικών βημάτων, μέσω της εξάλειψης του σειριακού υπολογισμού. Η υλοποίηση του Transformer στο GluonTS ακολουθεί τη γνήσια αρχιτεκτονική (Vaswani et al., 2017), έχοντας όμως αντικαταστήσει την softmax κατανομή διακριτών τιμών με άλλες κατανομές σύμφωνα με τα όσα συζητήθηκαν στην ενότητα (Κατανομή/Έξοδος) ώστε να ταιριάζει καλύτερα στο πλαίσιο των πιθανοτικών προβλέψεων.

### 4.8.3 Αυτο-παλινδρομικά Μοντέλα

Η οικογένεια των αυτο-παλινδρομικών μοντέλων, περιορίζει την πρόβλεψη σειράς σε ένα βήμα κάθε φορά. Τα μοντέλα εκπαιδεύονται σειρά-σειρά με ορίζοντα πρόβλεψης ενός βήματος κάθε φορά, ενώ το σφάλμα συναθροίζεται για όλη τη σειρά. Γνωστά μοντέλα που εντάσσονται σε αυτή την οικογένεια είναι:

- **Μη παραμετρικός Προβλέπτης Χρονοσειρών (Non-Parametric Time Series forecaster - NPITS)** που χρησιμοποιεί παρελθούσες παρατηρήσεις ως πρόβλεψη για το παρόν βήμα. Εν αντιθέσει με τη μέθοδο Naïve και τη seasonal Naïve που χρησιμοποιούν ένα καθορισμένο χρονικό δείκτη (προηγούμενος δείκτης T-1 ή προηγούμενη εποχή T-τ) ως πρόβλεψη για το χρονικό σημείο T, το NPITS λαμβάνει μέσω τυχαίας δειγματοληψίας έναν χρονικό δείκτη  $t \in \{0, \dots, T-1\}$  από το παρελθόν για να παραγάγει πρόβλεψη για το παρόν βήμα T. Η κατανομή δειγματοληψίας χρησιμοποιεί βάρη που φθίνουν εκθετικά καθώς μεταβαίνει όλο και πιο πίσω στο παρελθόν, σύμφωνα με τη σχέση

$$q_t(t) = \frac{e^{-\alpha|T-t|}}{\sum_{t'=0}^{T-1} e^{-\alpha|T-t'|}}$$

όπου το  $\alpha$  αποτελεί μία υπερπαράμετρο που προσαρμόζεται στα εκάστοτε δεδομένα. Ως εκ τούτου, η πιθανότητα δειγματοληψίας των πλέον πρόσφατων δεδομένων είναι μεγαλύτερη από αυτή των παλαιότερων.

- **DeepAR**, το οποίο αποτελείται από ένα RNN το οποίο λαμβάνει ως εισόδους τα προηγούμενα χρονικά σημεία και τις αντίστοιχες μεταβλητές, που αποτελούν χρονικά καθυστερημένες τιμές που σχετίζονται με τη δοθείσα συχνότητα. Έτσι λόγω χάρη, η καθυστέρηση στα ωριαία δεδομένα μπορεί να είναι ίση με μία ώρα (προηγούμενο χρονικό σημείο), μία ημέρα (1 \* 24 ώρες), μία εβδομάδα (7 \* 24 ώρες) κ.ο.κ. Τέλος, είτε υπολογίζει τις παραμέτρους μίας παραμετρικής κατανομής, είτε μίας συνάρτησης quantile απωλειών.
- **Wavenet** που αποτελεί ένα αυτο-παλινδρομικό νευρωνικό δίκτυο που αποτελείται από στιβασμένα διεσταλμένα και απλά συνελκτικά στρώματα, που διασφαλίζουν ότι κάθε προβλεφθείσα τιμή είναι ανεξάρτητη από τις μελλοντικές τιμές. Δεδομένου του γεγονότος ότι το Wavenet είχε αρχικώς αναπτυχθεί για την παραγωγή κυματομορφών για εφαρμογές όπως παραγωγή λόγου ή σύνθεση μουσικής, έχουν γίνει η απαραίτητες αλλαγές εντός του GluonTS, ώστε το μοντέλο να είναι κατάλληλο για πρόβλεψη χρονοσειρών.

Από τα διαθέσιμα μοντέλα του GluonTS, έγινε επιλογή του SimpleFeedForward, το θεωρητικό υπόβαθρο του οποίου έχει αναλυθεί στην Ενότητα 3.5. Ακολούθως παρουσιάζεται η μαθηματική διατύπωση της λειτουργίας του.

#### 4.9 Μαθηματικές Σχέσεις που Διέπουν τη Λειτουργία του SimpleFeedForward

Το SimpleFeedForward εντάσσεται στην κατηγορία των Discriminative μοντέλων, καθώς δεδομένων παρελθοντικών τιμών μήκους K, προβλέπει μελλοντική κατανομή για τ χρονικά βήματα. Η έξοδος του είναι οι παράμετροι του πιθανοτικού μοντέλου, ήτοι στην περίπτωση που η επιλεγμένη κατανομή εξόδου είναι η Gaussian, η έξοδος του θα είναι  $\theta = \{\mu,$



σ.}. Βασική υπόθεση που λαμβάνει το μοντέλο είναι ότι η από κοινού κατανομή είναι χρονικά ανεξάρτητη, ενώ τέλος η εκπαίδευσή του είναι τύπου global, δηλαδή το ίδιο δίκτυο με διάνυσμα βαρών  $W$ , χρησιμοποιείται καθ' όλη την έκταση του συνόλου δεδομένων.

Θεωρώντας ένα διάστημα  $[x_{i,a}, \dots, x_{i,c-1}]$  της  $i$ -οστής χρονοσειράς, στόχο αποτελεί η εκτίμηση της κατανομής του μελλοντικού παραθύρου  $[x_{i,b}, \dots, x_{i,c-1}]$  (γνωστό και ως prediction target) δεδομένων των ιστορικών της παρατηρήσεων  $[x_{i,a}, \dots, x_{i,b-1}]$  (context target), κατανομή η οποία εκφράζεται όπως έχει αναφερθεί σε προηγούμενη ενότητα ως εξής:

$$p_{\theta}(x_{i,b}, \dots, x_{i,c-1} \mid x_{i,a}, \dots, x_{i,b-1})$$

με  $\theta$  να είναι οι παράμετροι του δικτύου. Η ανωτέρω κατανομή μπορεί να παραγοντοποιηθεί σύμφωνα με την ακόλουθη σχέση:

$$p_{\theta}(x_{i,b}, \dots, x_{i,c-1} \mid x_{i,a}, \dots, x_{i,b-1}) = \prod_{t=b}^{c-1} p_{\theta}(x_{i,t} \mid x_{i,a}, \dots, x_{i,b-1})$$

Ορίζουμε  $NN_{\theta}$  ένα νευρωνικό δίκτυο που παραμετροποιείται με το  $\theta$ . Ο χώρος της υπόθεσης ορίζεται ως  $\{h : \mathbb{R}^{b-a} \rightarrow \mathbb{R}^{k(c-b)}\}$  όπου  $k$  είναι ο αριθμός των παραμέτρων που απαιτούνται για να οριστεί η κατανομή της εξόδου. Στην περίπτωση που ως κατανομή εξόδου υποτεθεί η Gaussian, η παράμετρο  $k$  λαμβάνει την τιμή 2.

Αναφορικά με την αξιολόγηση της κατανομής, αυτής μπορεί να περιγραφεί ως:

$$[\mu_{\theta i,b}, \dots, \mu_{\theta i,c-1}, \sigma_{\theta i,b}, \dots, \sigma_{\theta i,c-1}] = NN_{\theta}(x_{i,a}, \dots, x_{i,b-1})$$

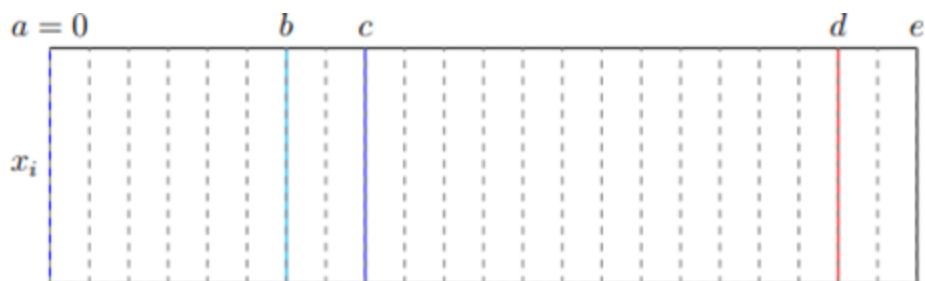
Δεδομένων των προαναφερθέντων, η εκτιμώμενη κατανομή της μεταβλητής  $x$  του μελλοντικού παραθύρου ( $t \in [b, c-1]$ ) με  $\Phi_{\mu,\sigma} : \mathbb{R} \rightarrow \mathbb{R}$  εκφράζεται ως μία παραμετροποιημένη Gaussian κατανομή με παραμέτρους  $\mu$  και  $\sigma$

$$p_{\theta}(x_{i,t} \mid x_{i,a}, \dots, x_{i,b-1}) = \varphi_{\mu_{\theta i,t}, \sigma_{\theta i,t}}(x_{i,t})$$

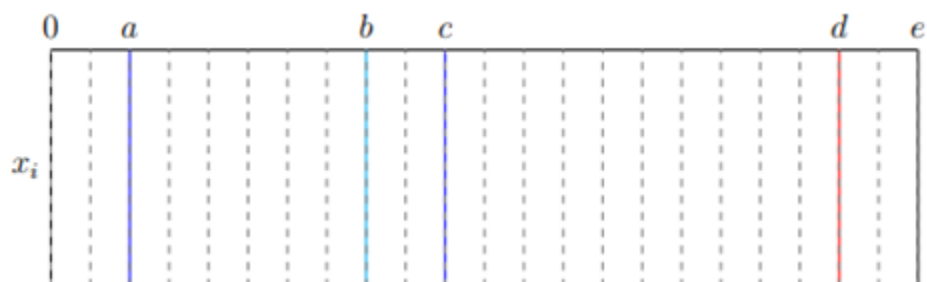
Αξιοποιώντας την εξίσωση  $p_{\theta}(x_{i,b}, \dots, x_{i,c-1} \mid x_{i,a}, \dots, x_{i,b-1}) = \prod_{t=b}^{c-1} p_{\theta}(x_{i,t} \mid x_{i,a}, \dots, x_{i,b-1})$ , η ανωτέρω σχέση μπορεί να μετασχηματιστεί στην εξής παραγοντοποιημένη μορφή:

$$p_{\theta}(x_{i,b}, \dots, x_{i,c-1} \mid x_{i,a}, \dots, x_{i,b-1}) = \prod_{t=b}^{c-1} \varphi_{\mu_{\theta i,t}, \sigma_{\theta i,t}}$$

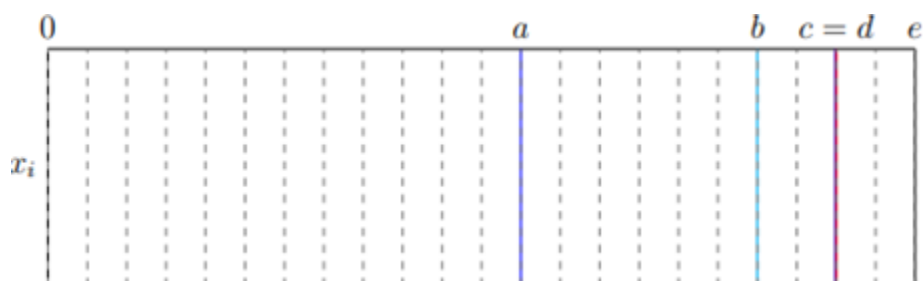
Η εκτίμηση αυτή γίνεται για διαφορετικά διαστήματα-παράθυρα  $[x_{i,a}, \dots, x_{i,c-1}]$ , από το πρώτο δυνατό παράθυρο για τιμή του  $b$  ίση με το context\_length έως το τελευταίο δυνατό, όπου  $b = d - \text{predict\_length}$ , όπως φαίνεται στα Σχήματα 4.3 - 4.5. Το μήκος του παραθύρου είναι μεγαλύτερο από το prediction length που αναφέρθηκε παραπάνω. Συνήθως λαμβάνεται μήκος διπλάσιο από τον χρονικό ορίζοντα πρόβλεψης.



Σχήμα 4.3 : Πρώτο πιθανό παράθυρο



Σχήμα 4.4 : Δεύτερο πιθανό παράθυρο



Σχήμα 4.5 : Τελευταίο πιθανό παράθυρο

Έτσι μπορεί να προκύψει ο τελικός υπολογισμός της απώλειας της εκτίμησης, ως το άθροισμα για κάθε χρονοσειρά και κάθε χρονικό σημείο της συνάρτησης απωλειών  $L$ :

$$\text{loss}(\theta) = \sum_{i=1}^I \sum_{t=b-a}^{d-1} L(x_{i,t}, \varphi_{\mu_{\theta_{i,t}}, \sigma_{\theta_{i,t}}}(x_{i,t}))$$

Να σημειωθεί ότι η απώλεια υπολογίζεται στο prediction target.

#### 4.10 Συνάρτηση Απωλειών

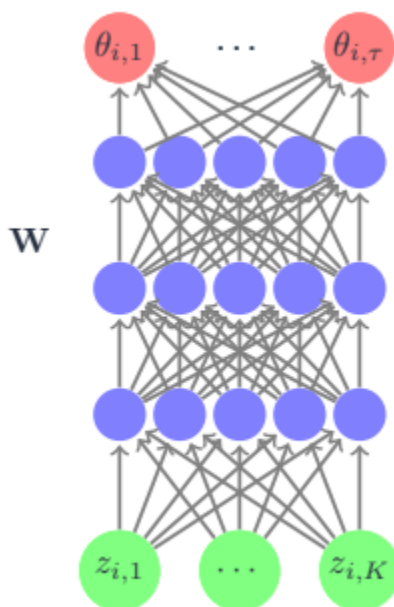
Οι συναρτήσεις απωλειών όπως αυτή που αναφέρθηκε στην προηγούμενη ενότητα, αποτελούν ένα ιδιαίτερα σημαντικό συστατικό μέρος της διαδικασίας πρόβλεψης με τη χρήση του deep learning, καθώς μέσα από αυτές γίνεται η αξιολόγηση της ποιότητας της πρόβλεψης κατά τη διάρκεια της εκπαίδευσης, με σκοπό την ενημέρωση του μοντέλου στη συνέχεια. Για την πλειονότητα των έτοιμων μοντέλων του GluonTS, η συνάρτηση απωλειών είναι το αντίστροφο του λογαρίθμου της πυκνότητας πιθανότητας (log-density) που δίνεται από τη σχέση:

$$L_{\text{default}}(y, \Phi(x)) = -\ln(\Phi(x = y))$$

όπου  $\Phi(x)$  είναι η κατανομή και  $y$  η πραγματική τιμή. Για τις περιπτώσεις μοντέλων για τα οποία η παραπάνω σχέση δεν ευσταθεί μαθηματικά λόγω της διαφορετικής τους αρχιτεκτονικής, ορίζεται διαφορετική συνάρτηση απωλειών. Χαρακτηριστικά παραδείγματα τέτοιων μοντέλων είναι τα MQRNN και MQCNN, τα οποία χρησιμοποιούν ένα συνδυασμό διαφορετικών απωλειών τεταρτημορίων (quantile losses) ως συνάρτηση απωλειών.

Η χρήση της παραπάνω εξίσωσης εκφράζει το πόσο αντιπροσωπευτική της πραγματικής κατανομής είναι η προβλεφθείσα. Να σημειωθεί ότι η πραγματική κατανομή αγνοείται, ωστόσο μέσω της δειγματοληψίας των διαθέσιμων ιστορικών παρατηρήσεων είναι δυνατόν να προσεγγιστεί.

Στο πλαίσιο της παρούσας διπλωματικής εργασίας, έγινε χρήση διαφορετικής συνάρτησης απωλειών, όπως θα παρουσιαστεί στην αντίστοιχη ενότητα του κεφαλαίου της προτεινόμενης μεθοδολογίας.



Σχήμα 4.6: Μορφή SimpleFeedForward δικτύου που χρησιμοποιήθηκε στην παρούσα διπλωματική εργασία



## ΚΕΦΑΛΑΙΟ 5

### Πειραματικής Διαδικασίας

#### 5.1 Στόχοι και Δομή Πειραματικής Διαδικασίας

Η παρούσα διπλωματική εργασία έχει διττό σκοπό. Αφενός επιχειρεί να βοηθήσει στην εξοικείωση με ένα νέο εργαλείο στον κλάδο των προβλέψεων, αυτό του GluonTS και αφετέρου να διερευνήσει την ακρίβεια των προβλέψεων που εξάγονται με τη χρήση του. Αναφορικά με τη διερεύνηση της ακρίβειας, αυτή δύναται να διακριθεί σε δύο περιπτώσεις:

- στη σύγκριση των σημειακών προβλέψεων που εξάγονται από το μοντέλο της απλής πρόσθιας τροφοδότησης (SimpleFeedForward) με άλλες πιο συμβατικές στατιστικές και ML μεθόδους και μοντέλα
- στη σύγκριση των πιθανοτικών προβλέψεων που εξάγονται από το επιλεγθέν μοντέλο του SimpleFeedForward διαφοροποιώντας την κατανομή της εξόδου.

Η υλοποίηση των δικτύων που χρησιμοποιήθηκαν και η πραγματοποίηση των πειραμάτων έγιναν σε γλώσσα προγραμματισμού Python (v.3.10). Δεδομένης της ευχέρειας που προσφέρει το GluonTS, αντί του έτοιμου μοντέλου SimpleFeedForward, έγινε χρήση μίας παραλλαγής του που αναπτύχθηκε από τους Lorenzo Stella και Tim Janushowski στο πλαίσιο του 40- Διεθνούς Συμποσίου Forecasting. Η ανάπτυξη του μοντέλου έγινε πάνω στο Gluon API του MXNet χρησιμοποιώντας το PyTorch.

Τα βασικότερα βήματα της προτεινόμενης μεθοδολογίας είναι τα ακόλουθα:

- *Προετοιμασία και αποεποχικοποίηση δεδομένων:* το στάδιο αυτό περιλαμβάνει τη δημιουργία ενός dataset σε μορφή κατάλληλη για εισαγωγή στο μοντέλο και εν συνεχεία την ανάπτυξη αλγορίθμου ο οποίος εξομαλύνει τη συνιστώσα της εποχιακότητας στα δεδομένα εκπαίδευσης
- *Ρύθμιση υπερπαραμέτρων και κατανομής του μοντέλου:* στην παραλλαγή του SimpleFeedForward πάνω στην οποία βασίστηκε το πειραματικό σκέλος της παρούσας διπλωματικής εργασίας, γίνονται οι αλλαγές που κρίθηκαν απαραίτητες ώστε το τελικό μοντέλο να εξυπηρετεί βέλτιστα τους σκοπούς της μελέτης που διενεργήθηκε
- *Εκπαίδευση και εξαγωγή Προβλέψεων:* το εν λόγω στάδιο περιλαμβάνει την εκπαίδευση του μοντέλου όπως έχει παραμετροποιηθεί βάσει του προηγούμενου βήματος και εν συνεχεία την εξαγωγή των πιθανοτικών προβλέψεων που χρησιμοποιούνται για την αξιολόγηση της επίδοσής του. Η εξαγωγή έγινε με τη μέθοδο των συνολικών

προβλέψεων (ensembles), δηλαδή επιλέχθηκε η median τιμή των προβλέψεων κάθε χρονοσειράς και κάθε βήματος από 30 μοντέλα SimpleFeedForward.

- *Αξιολόγηση του Μοντέλου:* στο τελευταίο στάδιο υπολογίζονται οι ποσοτικές μετρικές που καθορίζουν την ακρίβεια του μοντέλου που χρησιμοποιείται. Ο υπολογισμός αυτός λαμβάνει χώρα πάνω στα δεδομένα αξιολόγησης (test dataset).

## 5.2 Δεδομένα και Προετοιμασία τους

Λαμβάνοντας υπόψη το γεγονός ότι ως μοντέλο πρόβλεψης χρησιμοποιείται νευρωνικό δίκτυο, απαιτείται η προετοιμασία των δεδομένων ώστε να λάβουν μορφή κατάλληλη για εισαγωγή στο μοντέλο. Εν αντιθέσει με τα κλασσικά μοντέλα ML, τα οποία λαμβάνουν ως είσοδο εκπαίδευσης δύο σύνολα δεδομένων, το `X_train` και το `Y_train` που αποτελούν πίνακες με τα διανύσματα εκπαίδευσης και τις επιθυμητές εξόδους των διανυσμάτων αυτών (με την ίδια σειρά που τοποθετούνται στον πίνακα `X_train`) αντίστοιχα, στο GluonTS απαιτείται απλά ο καθορισμός του training dataset που περιέχει τα δεδομένα εκπαίδευσης και το test dataset που περιλαμβάνει τα δεδομένα για την αξιολόγηση των προβλέψεων. Επιπλέον, δεδομένης της δομής των δεδομένων που απαιτεί η εργαλειοθήκη του GluonTS, ανατίθεται μία χρονική ετικέτα σε κάθε παρατήρηση της εκάστοτε χρονοσειράς, ορίζοντας ένα starting date. Κατά τον πειραματισμό με τις χρονικές ετικέτες, διαπιστώθηκε ότι χωρίς βλάβη της γενικότητας είναι δυνατό να οριστεί κοινή ημερομηνία αρχής για κάθε χρονοσειρά, η οποία όπως φαίνεται και από το ακόλουθο χωρίο κώδικα ορίστηκε ως η 01/01/2019

```
freq = "M"
start = pd.Timestamp("01-01-2019", freq=freq)

train_ds = ListDataset([{'target': x, 'start': start}
                        for x in insample_list],
                        freq=freq)
```

Έτσι ήταν πλέον δυνατή η δημιουργία του αντικειμένου ListDataset του GluonTS που είναι η απαραίτητη μορφή εισόδου δεδομένων στα μοντέλα του και περιλαμβάνει τις παρατηρήσεις, τη χρονική ταμπέλα και τη συχνότητα των δεδομένων. Να σημειωθεί ότι τα ιστορικά δεδομένα πρέπει να είναι της μορφής np.array().

Του προαναφερθέντος προσδιορισμού του training dataset, προηγείται η αποεποχικοποίηση του συνόλου των δεδομένων εκπαίδευσης και η εξαγωγή των δεικτών εποχικότητας. Καταρχάς διενεργείται έλεγχος σημαντικής εποχιακής συμπεριφοράς για κάθε χρονοσειρά, μέσω του ελέγχου αυτοσυσχέτισης δεδομένων με περίοδο καθυστέρησης (k) ίση με τον αριθμό των περιόδων ενός κύκλου εποχιακότητας (pos) σε σύγκριση με τις αυτοσυσχετίσεις περιόδου καθυστέρησης έως και μίας μονάδας μικρότερης από τον αριθμό των περιόδων ενός κύκλου εποχιακότητας. Όπως θα αναφερθεί στη συνέχεια, τα δεδομένα που χρησιμοποιήθηκαν είναι μηνιαία, επομένως ο αριθμός των περιόδων εποχιακότητας είναι ίσος με 12. Μία χρονοσειρά θεωρείται εποχιακή αν και μόνο αν

$$|ACF_{pos}| > \text{Limit}$$

όπου

$$ACF_k = \frac{\sum_{i=1+k}^n [(Y_i - \bar{Y}) \cdot (Y_{i-k} - \bar{Y})]}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad \text{και} \quad \text{Limit} = t_{\text{critical}} \sqrt{\frac{1+2 \cdot (ACF_1 + \sum_{i=2}^{pos-1} ACF_i^2)}{n}}$$

Η τιμή του  $t_{\text{critical}}$  εξαρτάται από το διάστημα εμπιστοσύνης που επιλέγεται. Εν προκειμένω έγινε επιλογή ενός διαστήματος εμπιστοσύνης 95%, που αποτελεί και τη συνηθέστερη τιμή και ως εκ τούτου η τιμή του  $t_{\text{critical}}$  είναι 1,96. Μετά τον υπολογισμό των δεικτών ACF, λαμβάνει χώρα το τεστ εποχιακότητας και στην περίπτωση που μία χρονοσειρά κριθεί εποχιακή, γίνεται εξομάλυνση της συνιστώσας της εποχιακότητας με τη μέθοδο των κινητών μέσων όρων, ενώ οι συντελεστές εποχιακότητας σώζονται σε ξεχωριστό αρχείο για να αξιοποιηθούν στην εποχικοποίηση των προβλέψεων που εξάγει το μοντέλο. Στο Παράρτημα Β παρατίθεται ο αλγόριθμος που επιτελεί τις προαναφερθείσες διαδικασίες.

Σημειώνεται ότι δε λαμβάνει χώρα κάποια επιπλέον επεξεργασία των δεδομένων, καθώς το ίδιο το μοντέλο κανονικοποιεί εσωτερικά τα δεδομένα. Ωστόσο, για λόγους πληρότητας έγινε δοκιμή κανονικοποίησης των δεδομένων βάσει της σχέσης

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

η οποία είχε σαν αποτέλεσμα μία μικρή επι τα χείρω μεταβολή της ακρίβειας των προβλέψεων.

Τέλος, τα δεδομένα που χρησιμοποιήθηκαν είναι τα ωριαία δεδομένα του διαγωνισμού προβλέψεων M4, πληροφορίες για τον οποίο παρατίθενται στο Παράρτημα Α.

### 5.3 Ρύθμιση Υπερ-παραμέτρων και Κατανομής του Μοντέλου

Η επιτυχής εφαρμογή των ML μοντέλων γενικά αλλά και πιο συγκεκριμένα των DL μοντέλων, όπως αυτό που χρησιμοποιήθηκε εν προκειμένω, εξαρτάται σε πολύ μεγάλο βαθμό από την ορθή επιλογή και ρύθμιση των κατάλληλων υπερ-παραμέτρων και όσων παραμέτρων αφορούν τη διαδικασία της εκπαίδευσης. Μέσω της επιλογής των βέλτιστων αυτών υπερ-παραμέτρων, το μοντέλο είναι σε θέση να αναγνωρίζει καλύτερα τα χαρακτηριστικά της εκάστοτε χρονοσειράς που προβλέπει. Δεδομένου του γεγονότος ότι το GluonTS εστιάζει στην πρόβλεψη χρονοσειρών, οι απαιτούμενες ενέργειες από πλευράς χρήστη όσον αφορά τη ρύθμιση των υπερ-παραμέτρων είναι λιγότερες συγκριτικά με άλλα DL πακέτα.

Για την περίπτωση του SimpleFeedForward, αξιοποιήθηκαν τα πορίσματα που είχαν προκύψει από την ερευνητική δραστηριότητα της Μονάδας Προβλέψεων και Στρατηγικής του Εθνικού Μετσόβιου Πολυτεχνείου από. Αναλυτικότερα η βελτιστοποίηση του δικτύου εστιάστηκε στις ακόλουθες υπερ-παραμέτρους:

- αριθμός κρυφών επιπέδων (*hidden layers*) και του αντίστοιχου μεγέθους τους (*num\_hidden\_dimensions*). Αφού εξετάστηκαν δίκτυα έως και τριών *hidden layers* με πλήθος κόμβων από 2 έως 128 κόμβων, βέλτιστη επιλογή της υπερπαραμέτρου αυτής κρίθηκε η να είναι *hidden\_dimensions* = [64 , 32 , 16] , ήτοι το δίκτυο αποτελείται από τρία *hidden layers*, με το πρώτο να αποτελείται από 64 κόμβους, το δεύτερο από 32 και το τρίτο από 16.
- *prediction\_length* τέθηκε ίση με 18, δεδομένου του ότι ο ζητούμενος ορίζοντας πρόβλεψης των μηνιαίων δεδομένων στο πλαίσιο του διαγωνισμού M4 ήταν μήκους 18 βημάτων
- *context\_length*: έγινε διερεύνηση αναφορικά με την τιμή που θα έπρεπε να λάβει. Ως υπερ-παραμέτρος, περιγράφει τον αριθμό των χρονικών σημείων των δεδομένων εκπαίδευσης που λαμβάνει το μοντέλο πριν κάνει την πρόβλεψη. Στη βιβλιογραφία, μία συνήθης τιμή του *context\_length* κυμαίνεται στο ίδιο επίπεδο με το μήκος του *prediction\_length*, ωστόσο δοκιμάζοντας τιμές ίσες με ακέραια πολλαπλάσια του μήκους του ορίζοντα πρόβλεψης, προέκυψε ότι η βέλτιστη τιμή του *context\_length* στο πλαίσιο των πειραμάτων που διενεργήθησαν ήταν ίση με 36 (διπλάσιο μήκος από *prediction\_length*).
- *time\_freq*: όπως και στην περίπτωση του *prediction\_length*, η υπερ-παραμέτρος αυτή εξαρτάται αποκλειστικά από τη φύση των δεδομένων. Έγκυρες τιμές θεωρείται οποιοσδήποτε ακέραιος αριθμός ακολουθούμενος από τα γράμματα M, W, D, H η min τα οποία αναφέρονται σε μήνες, εβδομάδες, ημέρες κ.ο.κ. Εν προκειμένω, το *time\_freq* έλαβε την τιμή “M”, δεδομένου του ότι τα δεδομένα που χρησιμοποιήθηκαν ήταν μηνιαία.
- *epochs*: η υπερ-παραμέτρος αυτή αφορά το μέγιστο αριθμό περασμάτων που κάνει το μοντέλο στα δεδομένα εκπαίδευσης και η βέλτιστη τιμή της εξαρτάται από μέγεθος των δεδομένων και το ρυθμό μάθησης. Οι τυπικές τιμές εποχών κυμαίνονται από 10-1000. Δεδομένου του γεγονότος ότι η παραλλαγή που χρησιμοποιήθηκε δεν προσφέρει τη δυνατότητα του *early\_stopping\_patience*, που αποτελεί μία υπερ-παραμέτρο που διακόπτει την εκπαίδευση όταν δεν παρατηρείται σημαντική πρόοδος εντός του προκαθορισμένου αριθμού εποχών, καθώς και του ότι χρησιμοποιήθηκε η τεχνική των *ensembles* από 30 μοντέλα η οποία αυξάνει σημαντικά τη ζήτηση υπολογιστικών πόρων, ο αριθμός των εποχών τέθηκε ίσος με 20.
- *batch\_size*: η υπερ-παραμέτρος αυτή καθορίζει τον αριθμό των δειγμάτων που διαδίδεται μέσα στο δίκτυο. Για παράδειγμα αν είναι διαθέσιμα 1050 δείγματα εκπαίδευσης και το *batch\_size* είναι 100, τότε ο αλγόριθμος λαμβάνει τα πρώτα 100 δείγματα (1° -100° ) του συνόλου εκπαίδευσης και εκπαιδεύει το δίκτυο. Στη συνέχεια λαμβάνει τα επόμενα 100 δείγματα (101° - 200° ) και επαναλαμβάνει την προαναφερθείσα δείγματα. Στην περίπτωση όπου ο μέγεθος των δειγμάτων δεν είναι ακέραιο πολλαπλάσιο του *batch\_size* (όπως στο παράδειγμα που παρατέθηκε), στο τέλος λαμβάνονται τα υπολειπόμενα δείγματα. Εν γένει προτιμάται τιμή του *batch\_size* μικρότερη ή ίση με το πλήθος των δειγμάτων εκπαίδευσης, καθώς μειώνεται η απαίτηση σε μνήμη ( η εκπαίδευση γίνεται



με λιγότερα δείγματα, άρα χρειάζεται λιγότερη μνήμη) ενώ παράλληλα αυξάνεται η ταχύτητα στα περισσότερα δίκτυα, καθώς μετά από κάθε προώθηση ενημερώνονται τα βάρη του δικτύου και αναπροσαρμόζονται καλύτερα. Στο πλαίσιο των πειραμάτων που έτρεξαν για τους σκοπούς της παρούσας διπλωματικής εργασίας, η τιμή που έλαβε η υπερ-παράμετρος του `batch_size` ήταν 128 (όσο και το μήκος της μικρότερης χρονοσειράς του `training dataset`).

- `num_batches_per_epoch`: όπως υποδηλώνει και το όνομά της, η υπερ-παράμετρος αυτή αφορά τον αριθμό των `batches` που λαμβάνει το μοντέλο σε κάθε εποχή. Στον αλγόριθμο που χρησιμοποιήθηκε, η τιμή της τέθηκε ίση με 128.

Να σημειωθεί ότι η εύρεση των βέλτιστων τιμών έγινε με τη χρήση του αλγορίθμου `Tree-of-Parzen-Estimators (TPE)` όπως είναι υλοποιημένη στη βιβλιοθήκη `HyperOpt` της `Python`, αξιοποιώντας τα δεδομένα του διαγωνισμού `M3`. Οι βέλτιστες τιμές προέκυψαν βάσει του τελευταίου παραθύρου των ιστορικών δεδομένων που χρησιμοποιήθηκε ως σύνολο αξιολόγησης, το μήκος του οποίου ήταν ίσο με τον ορίζοντα πρόβλεψης. Στον ακόλουθο Πίνακα παρουσιάζονται συγκεντρωτικά οι βέλτιστες τιμές των προαναφερθέντων παραμέτρων:

Υπερ-παράμετρος	Τιμή
<code>hidden layers</code>	3
<code>num_hidden_dimensions</code>	[64, 32, 16]
<code>prediction_length</code>	18
<code>context_length</code>	36
<code>time_freq</code>	M
<code>epochs</code>	20
<code>batch_size</code>	128
<code>num_batches_per_epoch</code>	128

Πίνακας 5.1: Τιμές Υπερ-Παραμέτρων Μοντέλου

## 5.4 Συνάρτηση Απωλειών

Η συνάρτηση απωλειών που χρησιμοποιείται αποτελεί μία εξατομικευμένη συνάρτηση για την περίπτωση της κατανομή `StudentT` του κώδικα που αποτέλεσε τη βάση για την παρούσα διπλωματική και δίνεται μαθηματικά από τη σχέση:

$$loss = -distr.log\_prob(target / scale) + torch.log(scale)$$

όπου η παράμετρος `target` αποτελεί το διάνυσμα-στόχο κατά την εκπαίδευση και η παράμετρος `scale` είναι το κανονικοποιημένο διάνυσμα των παρελθοντικών τιμών που λαμβάνει ως είσοδο το μοντέλο σε κάθε εποχή εκπαίδευσης.

Ένα από τα βασικά ερωτήματα που θέτει η παρούσα διπλωματική είναι το κατά πόσο η υπόθεση που λαμβάνεται συνήθως αναφορικά με την κατανομή που ακολουθούν οι πιθανοτικές προβλέψεις δίνει πράγματι το μικρότερο σφάλμα. Από τη βιβλιογραφία, προκύπτει ότι η συχνότερα χρησιμοποιούμενη κατανομή είναι η Κανονική (Normal Distribution). Στο πλαίσιο των πειραμάτων που διενεργήθησαν, εξετάστηκαν τέσσερις ακόμα κατανομές πέρα από την Κανονική και συγκεκριμένα η Laplace, η LogNormal, η StudentT και η Cauchy, η υλοποίηση των οποίων παρουσιάζεται ακολούθως μαζί με ορισμένα χαρακτηριστικά τους. Να σημειωθεί ότι λόγω των ιδιαίτερων χαρακτηριστικών της κάθε κατανομής, πέρα από την αλλαγή της δηλωσης του τύπου κατανομή που λαμβάνει το μοντέλο, απαιτούνται επιπλέον ενέργειες που αφορούν τον αριθμό των παραμέτρων τους και της συνάρτησης απώλειας κατά την εκπαίδευση.

## **5.5 Κατανομές πειραμάτων**

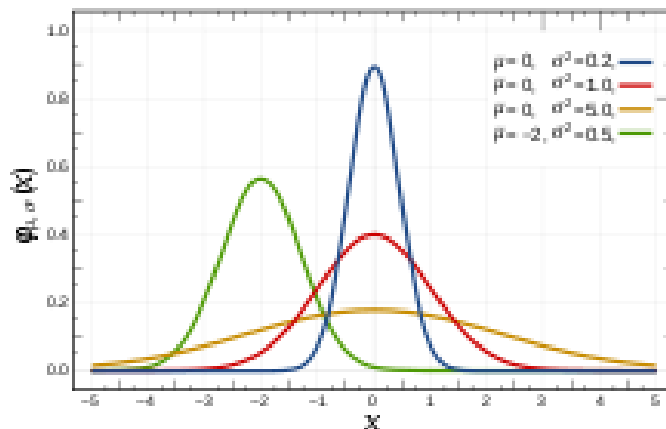
Στην ενότητα αυτή θα αναλυθούν οι πέντε Κατανομές που μελετήθηκαν κατά τη διάρκεια των πειραμάτων που διενεργήθηκαν.

### **5.5.1 Κανονική Κατανομή**

Η Κανονική κατανομή (γνωστή και ως Gaussian) αποτελεί μία συνεχή κατανομή της οποίας η συνάρτηση πυκνότητας πιθανότητας δίνεται από τη σχέση:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

όπου  $\mu$  είναι η αναμενόμενη ή μέση τιμή (στην περίπτωση της κατανομής αυτής ταυτίζεται και με τη median τιμή), ενώ το  $\sigma$  είναι η τυπική απόκλιση. Καθώς μεταβάλλεται η τιμή του  $\mu$ , μετατοπίζεται η καμπύλη προς τα αριστερά ή προς τα δεξιά, ενώ η μεταβολή στην τυπική απόκλιση (και κατά συνέπεια στη διακύμανση η οποία ισούται με το τετράγωνο της τυπικής απόκλισης), μεταβάλλει αντίστοιχα το ύψος της κορυφής της καμπύλης, όπως παρουσιάζεται στο κάτωθι διάγραμμα:



Σχήμα 5.1 : Συνάρτηση Πυκνότητας Πιθανότητας Κανονικής Κατανομής

Στο περιβάλλον του PyTorch στο οποίο υλοποιήθηκε το μοντέλο που χρησιμοποιήθηκε, η κλίση της κλάσης της Κανονικής κατανομής γίνεται απλά με την εντολή

```
torch.distributions.normal.Normal(loc,scale,validate_args=None)
```

με την παράμετρο `loc` να είναι το  $\mu$  και το `scale` το  $\sigma$ , ήτοι απαιτούνται δύο παράμετροι για τον προσδιορισμό της (βλ. αντίστοιχη παράγραφο Παραρτήματος Β). Όπως προαναφέρθηκε, η παραλλαγή του `SimpleFeedForward` που χρησιμοποιήθηκε, χρησιμοποιούσε την κατανομή `StudentT`. Λόγω της πιο “απλωμένης” φύσης της κανονικής κατανομής ωστόσο, η συνάρτηση απωλειών που είχε δημιουργηθεί εμφάνιζε εξαιρετικά μεγάλες τιμές, καθώς για τα ακραία ποσοστοτημώρια η συνάρτηση `log_prob` όπως είναι υλοποιημένη στον πηγαίο κώδικα της Κανονικής Κατανομής του Torch και φαίνεται στο ακόλουθο στιγμιότυπο, λάμβανε μεγάλες τιμές.

```
[docs] def log_prob(self, value):
    if self._validate_args:
        self._validate_sample(value)
    # compute the variance
    var = (self.scale ** 2)
    log_scale = math.log(self.scale) if isinstance(self.scale, Real) else self.scale.log()
    return -((value - self.loc) ** 2) / (2 * var) - log_scale - math.log(math.sqrt(2 *
math.pi))
```

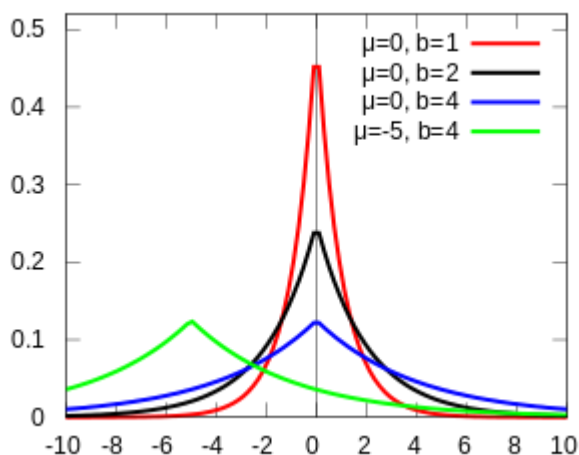
Για την αποφυγή του προβλήματος αυτού, τροποποιήθηκε το διάνυσμα που αποτελεί την έξοδο της συνάρτησης απωλειών, εξαιρώντας το top 20% των σφαλμάτων (εξαιρείται το top, καθώς ενώ η έξοδος της `log_prob` της κατανομής είναι αρνητική, το πρόσημο αντιστρέφεται λόγω της συνάρτησης απωλειών).

### 5.5.2 Κατανομή Laplace

Η κατανομή Laplace είναι μία συνεχής κατανομή πιθανότητας η οποία είναι γνωστή λόγω της μορφής της και ως διπλοεκθετική, ενώ η συνάρτηση πυκνότητας πιθανότητάς της δίνεται από τη σχέση:

$$f(x | \mu, b) = \frac{1}{2b} e^{-\left(\frac{|x-\mu|}{b}\right)} = \begin{cases} e^{-\left(\frac{\mu-x}{b}\right)} & , \text{αν } x < \mu \\ e^{-\left(\frac{x-\mu}{b}\right)} & , \text{αν } x \geq \mu \end{cases}$$

όπου  $\mu$  είναι η παράμετρος θέσης (location parameter) και καθορίζει τη θέση της κατανομής και  $b > 0$  είναι μία παράμετρος κλίμακας (scale parameter). Όσο μεγαλύτερες τιμές λαμβάνει η παράμετρος  $b$ , τόσο πιο “απλωμένη” είναι η κατανομή. Οι μεταβολές της καμπύλης της κατανομής ανάλογα με τις διαφορετικές τιμές των παραμέτρων  $\mu$  και  $b$ , φαίνονται χαρακτηριστικά στο ακόλουθο διάγραμμα:



Σχήμα 5.2 : Συνάρτηση Πυκνότητας Πιθανότητας Κατανομής Laplace

Η κλιση της κλάσης της κατανομή Laplace γίνεται με την εντολή:

```
torch.distributions.laplace.Laplace(loc,scale,validate_args=None)
```

και όμοια με την περίπτωση της Κανονικής κατανομής, απαιτούνται δύο παράμετροι για τον προσδιορισμό της, το `loc` που αντιστοιχεί στο μέσο της κατανομής (mean) και το `scale` που αφορά την κλίμακά της, δηλαδή το πόσο απλωμένη θα είναι.

Όπως γίνεται εμφανές από τον πηγαίο κώδικα της κατανομής Laplace που παρατίθεται στη συνέχεια, ο υπολογισμός της `log_prob` της, καθιστά αναγκαία την επέμβαση στο διάνυσμα

απωλειών, με την προτεινόμενη μεθοδολογία να είναι η ίδια με την περίπτωση της Κανονικής Κατανομής (εξαιρέση ακραίων top 20% των τιμών).

```
def log_prob(self, value):
    if self._validate_args:
        self._validate_sample(value)
    return -torch.log(2 * self.scale) - torch.abs(value - self.loc) / self.scale
```

### 5.5.3 Κατανομή Log-Normal

Η κατανομή Log-Normal αποτελεί μία συνεχή κατανομή μίας τυχαίας μεταβλητής της οποίας ο αλγόριθμος είναι κανονικά καταταμημένος, και ως εκ τούτου αν η τυχαία μεταβλητή  $X$  ακολουθεί την κατανομή Laplace, τότε η τυχαία μεταβλητή  $Y = \ln(X)$  είναι κανονικά καταταμημένη, ή ισοδύναμα αν η τυχαία μεταβλητή  $Y$  ακολουθεί κανονική κατανομή, τότε η  $X = \exp(Y)$  ακολουθεί Log-Normal κατανομή.

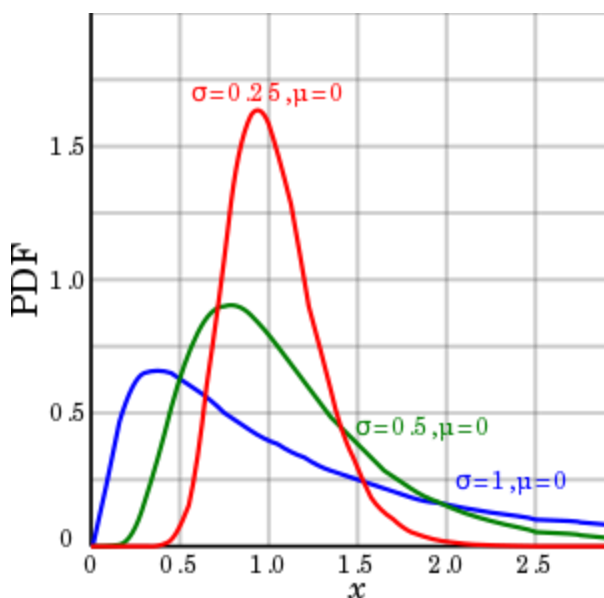
Μία τυχαία θετική μεταβλητή ακολουθεί Log-Normal κατανομή, λχ  $X \sim \text{Lognormal}(\mu_x, \sigma_x^2)$  αν ο φυσικός λογάριθμος του  $X$  είναι κανονικά καταταμημένος με μέση τιμή  $\mu$  και διακύμανση  $\sigma$ . Έτσι η συνάρτηση πυκνότητας πιθανότητας της Log-Normal κατανομής, με  $\Phi$  τη συνάρτηση αθροιστικής κατανομής και  $\phi$  τη συνάρτηση πυκνότητας κατανομής, δίνεται από τη σχέση:

$$\begin{aligned} f_X(x) &= \frac{d}{dx} \Pr(X \leq x) = \frac{d}{dx} \Pr(\ln X \leq \ln x) = \frac{d}{dx} \Phi\left(\frac{\ln x - \mu}{\sigma}\right) = \\ &= \phi\left(\frac{\ln x - \mu}{\sigma}\right) \frac{d}{dx} \left(\frac{\ln x - \mu}{\sigma}\right) = \phi\left(\frac{\ln x - \mu}{\sigma}\right) \frac{1}{\sigma x} = \\ &= \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \end{aligned}$$

Οι παράμετροι  $\mu$ ,  $\mu_x$  και  $\sigma$ ,  $\sigma_x$  συνδέονται μεταξύ τους με τις ακόλουθες σχέσεις:

$$\mu = \ln\left(\frac{\mu_x^2}{\sqrt{\mu_x^2 + \sigma_x^2}}\right) \quad \text{και} \quad \sigma = \ln\left(1 + \frac{\sigma_x^2}{\mu_x^2}\right)$$

Η καμπύλη της Log-Normal είναι χαρακτηριστική σε σχέση με τις καμπύλες των υπόλοιπων κατανομών που εξετάστηκαν, υπό την έννοια ότι δίνει μεγάλο βάρος στα μικρά percentiles, γεγονός που είναι εμφανές από το ότι η κορυφή της καμπύλης είναι μετατοπισμένη προς τα αριστερά σε σχέση με τις καμπύλες των υπόλοιπων διαγραμμάτων.



Σχήμα 5.3 : Συνάρτηση Πυκνότητας Πιθανότητας Κατανομής LogNormal

Η κλίση από το PyTorch γίνεται όμοια με τις προαναφερθείσες περιπτώσεις κατανομών μέσω της εντολής

```
torch.distributions.log_normal.LogNormal(loc, scale, validate_args=None)
```

Για τον προσδιορισμό της είναι αρκετές και πάλι οι δύο παράμετροι loc και scale, με τη διαφορά ότι το loc αναφέρεται πλέον στη μέση τιμή (mean) του λογαρίθμου της κατανομής, καθώς επίσης και η τυπική απόκλιση (scale) αναφέρεται στο λογάριθμο της κατανομής.

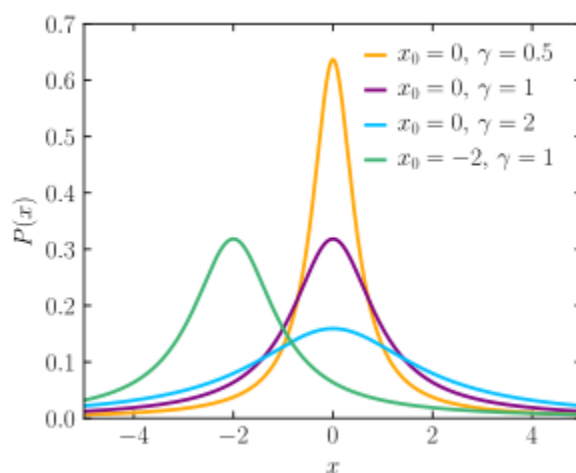
### 5.5.4 Κατανομή Cauchy

Η Κατανομή Cauchy με συνάρτηση πυκνότητας πιθανότητας:

$$f(x ; x_0 , \gamma) = \frac{1}{\pi\gamma [1+(\frac{x-x_0}{\gamma})^2]} = \frac{1}{\pi\gamma} \left[ \frac{\gamma^2}{(x-x_0)^2 + \gamma^2} \right]$$

αποτελεί μία συνεχή κατανομή, η οποία αποτελεί την κατανομή του λόγου δύο ανεξάρτητων κανονικά κατανεμημένων τυχαίων μεταβλητών με μηδενική μέση τιμή. Στην παραπάνω εξίσωση το  $x_0$  αποτελεί την παράμετρο θέση (location parameter) που καθορίζει τη θέση της κορυφής της καμπύλης και η παράμετρος  $\gamma$  είναι η παράμετρος κλίμακας που καθορίζει το μέσο πλάτος στο μέσο (half-width at half-maximum - HWHM).

Η μετατόπιση της καμπύλης ανάλογα με τις τιμές των παραμέτρων  $x_0$  και  $\gamma$  είναι παρόμοια με όσα έχουν περιγραφεί για την περίπτωση των προηγούμενων κατανομών και χαρακτηριστικά παραδείγματα παρουσιάζονται στο ακόλουθο διάγραμμα:



Σχήμα 5.4 : Συνάρτηση Πυκνότητας Πιθανότητας Κατανομής Cauchy

Η κλίση της από το περιβάλλον PyTorch για τις ανάγκες της παρούσας διπλωματικής εργασίας γίνεται και σε αυτή την περίπτωση με την εντολή

```
torch.distributions.cauchy.Cauchy(loc,scale,validate_args=None)
```

με τις παραμέτρους loc και scale να αφορούν τη μέση τιμή και την τυπική απόκλιση αντίστοιχα.

### 5.5.5 Κατανομή StudentT

Η κατανομή StudentT αποτελεί μία οικογένεια συνεχών κατανομών πιθανότητας, που προκύπτει όταν εκτιμάται ένας κανονικά κατανομημένος πληθυσμός δειγμάτων, το πλήθος του οποίου όμως είναι μικρό και η τυπική του απόκλιση άγνωστη.

Η συνάρτηση πυκνότητας πιθανότητας περιγράφεται μαθηματικά ως εξής:

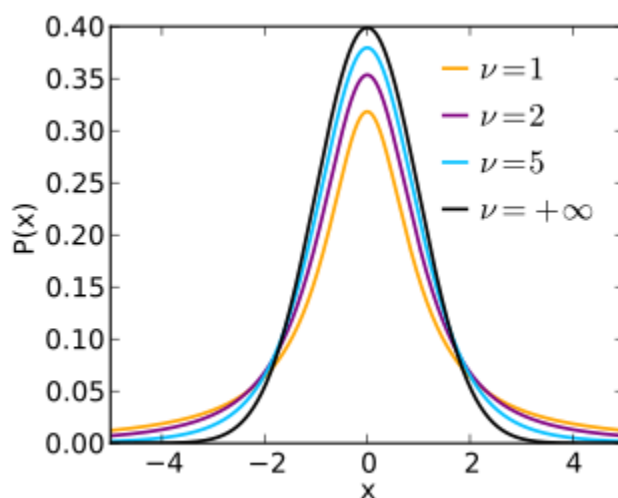
$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

όπου  $\nu$  είναι οι βαθμοί ελευθερίας και  $\Gamma$  η συνάρτηση Γάμμα. Ισοδύναμα η συνάρτηση κατανομής πιθανότητας μπορεί να γραφτεί ως:

$$f(t) = f(t) = \frac{1}{\sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

όπου  $B$  είναι η συνάρτηση  $B$ .

Η συνάρτηση πυκνότητας πιθανότητας της StudentT είναι συμμετρική και η μορφή της μοιάζει με τη μορφή καμπάνας της κανονικής κατανομής με μέση τιμή 0 και διακύμανση 1, με τη διαφορά ότι η κορυφή της είναι λίγο χαμηλότερη και η γραφική της παράσταση ευρύτερη. Καθώς η παράμετρος  $\nu$  των βαθμών ελευθερίας αυξάνει, τόσο περισσότερο προσεγγίζει την κανονική κατανομή μέσης τιμής 0 και διακύμανσης 1. Για το λόγο αυτό η παράμετρος  $\nu$  είναι γνωστή και ως παράμετρος κανονικότητας (normality parameter). Η μεταβολή της καμπύλης ανάλογα με την τιμή της παραμέτρου  $\nu$  φαίνεται στο ακόλουθο διάγραμμα:



Σχήμα 5.5 : Συνάρτηση Πυκνότητας Πιθανότητας Κατανομής StudentT

Η κλίση της γίνεται με την εντολή

```
torch.distributions.studentT.StudentT(df, loc=0.0, scale=1.0, validate_args=None)
```

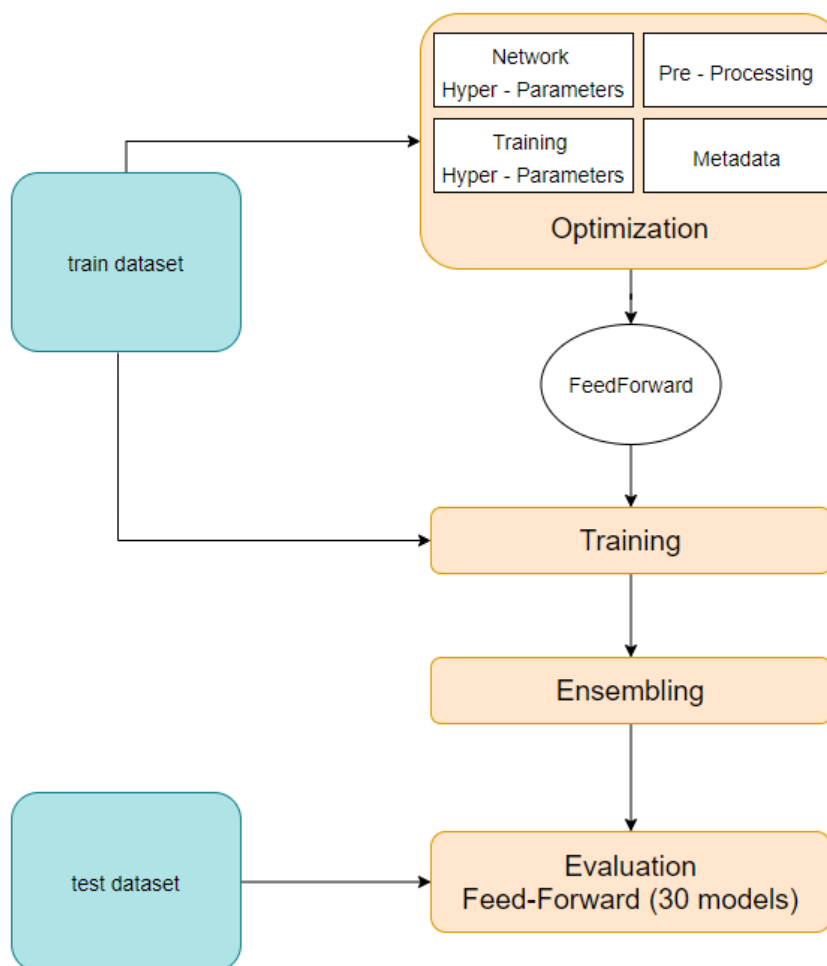
Είναι εμφανές ότι εν αντιθέσει με τις υπόλοιπες τέσσερις κατανομές που απαιτούσαν δύο παραμέτρους για τον προσδιορισμό τους, στη StudentT εισάγεται και μία τρίτη, η  $df$  που αφορά τους βαθμούς ελευθερίας (παράμετρος  $\nu$  συνάρτησης πυκνότητας πιθανότητας).

Ο τρόπος με τον οποίο δηλώνονται και δημιουργούνται οι παράμετροι στο δίκτυο εκπαίδευσης του SimpleFeedForward μοντέλου παρουσιάζεται αναλυτικά στο Παράρτημα Β όπου γίνεται ανάλυση του κώδικα που χρησιμοποιήθηκε για το πειραματικό σκέλος της εργασίας.



## 5.6 Εκπαίδευση Μοντέλων και Εξαγωγή Προβλέψεων

Έχοντας προσδιορίσει τις τιμές των υπερ-παραμέτρων του μοντέλου, καθώς επίσης την κατανομή που είναι επιθυμητή στην έξοδο τού, είναι πλέον δυνατή η εκπαίδευση του SimpleFeedForward. Η εκπαίδευση των μοντέλων στο GluonTS χρησιμοποιεί όλα τα δεδομένα του training dataset, ακολουθώντας συνεπώς μία “cross-learning” προσέγγιση, αφού το μοντέλο μαθαίνει από πολλαπλές χρονοσειρές αντί για μία. Εν συνεχεία, το εκπαιδευμένο μοντέλο χρησιμοποιείται για την εξαγωγή προβλέψεων. Αντί όμως η τελική πρόβλεψη να αποτελεί την έξοδο ενός και μόνο μοντέλου, επιλέχθηκε να αξιοποιηθούν οι προβλέψεις από πολλαπλά μοντέλα (συνολικά 30 στον αριθμό), των οποίων οι προβλέψεις εκτιμήθηκαν και τελικώς ελήφθη η median τιμή των 30 διαφορετικά εκπαιδευμένων (λόγω τυχαίας και διαφορετικής αρχικοποίησης των βαρών τους) για καθένα από τα 18 βήματα του ορίζοντα πρόβλεψης, τεχνική γνωστή και ως ensembles. Ο λόγος που επιλέχθηκε η προαναφερθείσα προσέγγιση αφορά την προσπάθεια ενίσχυσης της προβλεπτικής ακρίβειας της προτεινόμενης μεθοδολογίας. Σχηματικά η διαδικασία της εκπαίδευσης και πρόβλεψης παρουσιάζεται στο ακόλουθο διάγραμμα ροών



Σχήμα 5.6 : Σχηματική Αναπαράσταση Διαδικασίας Εκπαίδευσης και Πρόβλεψης

Τέλος αξίζει να υπενθυμιστεί η διαφορά μεταξύ της λήψης της median και της mean τιμής, η οποία εν ολίγοις εντοπίζεται στο ότι η mean τιμή αποτελεί τον κλασσικό αριθμητικό μέσο όρο, ή αλλιώς την αναμενόμενη τιμή, ενώ το median αποτελεί το κέντρο του διάνυσματος (στην περίπτωση που το πλήθος των στοιχείων του είναι περιττό) ή το ημίαθροισμα των δύο μεσαίων στοιχείων του διάνυσματος (στην περίπτωση που το διάνυσμα είναι άρτιου μήκους), έχοντας όμως πρώτα διατάξει σε αύξουσα σειρά τις τιμές του. Έτσι για παράδειγμα, αν δίνεται το διάνυσμα  $[10,10,20,40,70]$ , το mean θα είναι ίσο με  $(10+10+20+40+70)/5 = 30$ , ενώ το median θα είναι το 20. Εν γένει η λήψη της median τιμής θεωρείται ότι αυξάνει την ακρίβεια, δεδομένου του γεγονότος ότι δεν επηρεάζεται από ακραίες τιμές που ενδεχομένως να προκύψουν, γεγονός που οδήγησε στην απόφαση να χρησιμοποιηθεί για τα ensembles των προβλέψεων.

### **5.7 Αξιολόγηση Μοντέλου**

Όπως έχει αναφερθεί παραπάνω, η αξιολόγηση των ensembled προβλέψεων συνίσταται σε δύο μέρη: τη σύγκριση των σημειακών προβλέψεων που παράγονται σε σχέση με άλλες πιο συμβατικές μεθόδους και μοντέλα και τη σύγκριση των πιθανοτικών προβλέψεων μεταξύ των διαφορετικών κατανομών που επιλέχθηκαν να εξεταστούν. Αμφότερων των συγκρίσεων προηγήθηκε η εποχικοποίηση των median προβλέψεων που ελήφθησαν από τα 30 μοντέλα, με τη χρήση των δεικτών εποχιακότητας που υπολογίστηκαν κατά τη διαδικασία της αποεποχικοποίησης της χρονοσειράς.

Αναφορικά με τη σύγκριση των σημειακών προβλέψεων, έγινε χρήση των στατιστικών δεικτών sMAPE (Ενότητα 2.10.7) και το MAsE (Ενότητα 2.10.8).

Τέλος για τη σύγκριση των διαφορετικών κατανομών χρησιμοποιήθηκε ο δείκτης του RMSSE (Ενότητα 2.10.6) και το Scaled Pinball Loss (Ενότητα 2.10.10).

## ΚΕΦΑΛΑΙΟ 6

### Αποτελέσματα, Συμπεράσματα και Προεκτάσεις

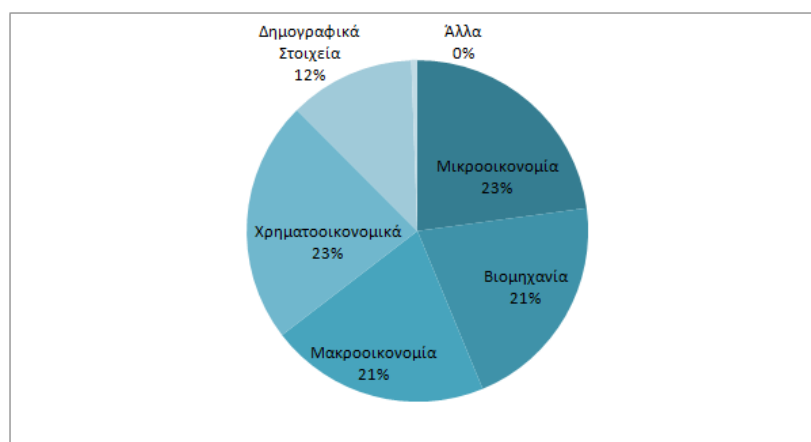
#### 6.1 Πειραματικό Σύνολο Δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση και την αξιολόγηση της προτεινόμενης μεθοδολογίας είναι τα μηνιαία δεδομένα του διαγωνισμού M4 (περισσότερες πληροφορίες για το διαγωνισμό στο Παράρτημα Α). Το σύνολο των δεδομένων αποτελείται από 48000 χρονοσειρές από διαφορετικούς κλάδους όπως φαίνεται στον ακόλουθο Πίνακα:

ΔΕΔΟΜΕΝΑ	Μικροοικονομία	Βιομηχανία	Μακροοικονομία	Χρηματοοικονομικά	Δημογραφικά Στοιχεία	Άλλα	Σύνολο
Μηνιαία	10975	10017	10016	10987	5728	277	48000

Πίνακας 6.1 : Κλάδοι Μηνιαίων Δεδομένων διαγωνισμού M4

Με τον επιμερισμό των συνολικών δεδομένων στις επιμέρους κατηγορίες να φαίνεται στο ακόλουθο διάγραμμα:



Σχήμα 6.1 : Ποσόστωση μηνιαίων δεδομένων στους επιμέρους κλάδους

Ο λόγος που επιλέχθηκαν τα συγκεκριμένα πειράματα για τη διενέργεια των πειραμάτων είναι ότι το ιδιαίτερα μεγάλο πλήθος τους είναι ιδανικό για εφαρμογές μηχανικής μάθησης, οι οποίες όπως έχει εξηγηθεί στο αντίστοιχο κεφάλαιο της παρούσας διπλωματικής εργασίας, απαιτούν διαθεσιμότητα πολλών δεδομένων.

Τα δεδομένα του M4 είναι διαθέσιμα σε δύο σύνολα: ένα που αφορά την εκπαίδευση και ένα την αξιολόγηση. Το δεύτερο στην ουσία περιέχει τις 18 τιμές της κάθε χρονοσειράς που το μοντέλο αγνοεί κατά την εκπαίδευσή του. Να σημειωθεί ότι η εσωτερική αξιολόγηση κατά τη διάρκεια της εκπαίδευσης πραγματοποιείται από το ίδιο το μοντέλο, το οποίο απομονώνει τις 18 τελευταίες παρατηρήσεις σε κάθε χρονοσειρά του συνόλου εκπαίδευσης, χωρίς να απαιτείται κάποια ενέργεια από πλευράς χρήστη.

Όλα τα δεδομένα του συνόλου εκπαίδευσης υπέστησαν την επεξεργασία που αναφέρθηκε στο προηγούμενο Κεφάλαιο με σκοπό την αφαίρεση της συνιστώσας της εποχιακότητας ώστε να ενισχυθεί η ακρίβεια των αποτελεσμάτων.

## 6.2 Σύγκριση Αποτελεσμάτων

### 6.2.1 Αποτελέσματα Σημειακών Προβλέψεων

Όπως έχει αναφερθεί οι σημειακές προβλέψεις των μοντέλων που υλοποιήθηκαν στο πλαίσιο της παρούσας εργασίας, λαμβάνονται ως το mean των πιθανοτικών προβλέψεων που εξάγονται. Αναφορικά με τη σύγκριση των σημειακών προβλέψεων που προκύπτουν από τις πέντε διαφορετικές κατανομές, αξιοποιήθηκαν οι δείκτες ακρίβειας sMAPE και το κανονικοποιημένο MASE ως προς τις διαφορές με καθυστέρηση 12 περιόδων (όσο και το μήκος εποχιακότητας). Στην επιλογή των δύο αυτών δεικτών συνετέλεσαν οι εξής παράγοντες:

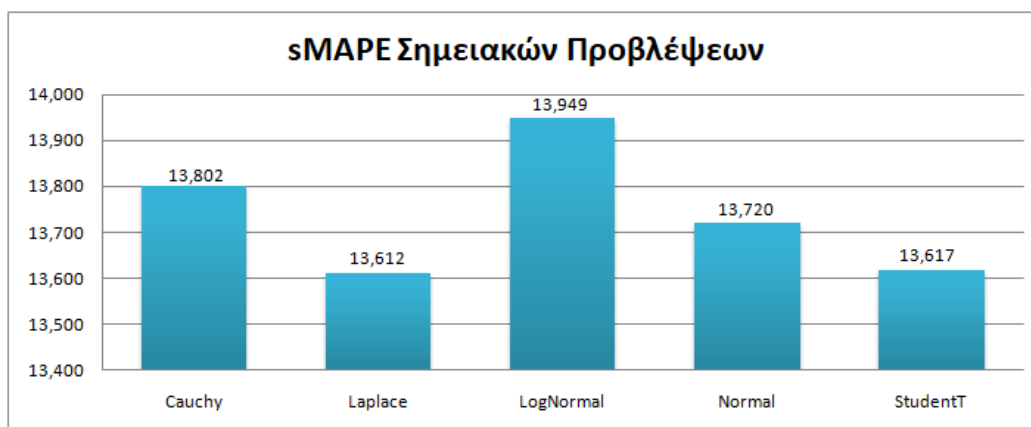
#### sMAPE

- είναι ανεξάρτητο της κλίμακας της εκάστοτε χρονοσειράς και επιτρέπει τη σύγκριση μεταξύ χρονοσειρών με διαφορετικό μέσο επίπεδο
- δε λαμβάνει απροσδιόριστες τιμές όπως το MAPE
- είναι φραγμένο μεταξύ των ορίων [0%, 200%]

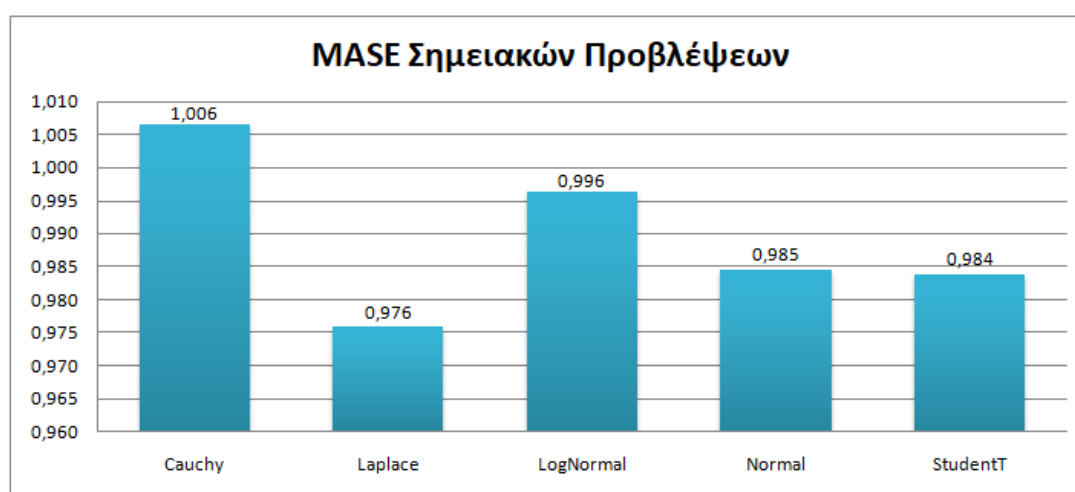
#### MASE

- δε δίνει απροσδιόριστες τιμές
- επιτρέπει τη σύγκριση μεταξύ διαφορετικών χρονοσειρών
- εμπεριέχει την έννοια του σημείου αναφοράς (benchmarking), δεδομένου του ότι για τιμές μικρότερες της μονάδας υποδηλώνει ότι η επιλεγμένη μέθοδος είναι καλύτερη από τη Naive (ή την sNaive εν προκειμένω που έχει εφαρμοστεί το lag)

Η αξιολόγηση των προβλέψεων που εξήχθησαν από τα 30 ensembled μοντέλα, αφού προηγήθηκε η εποχικοποίησή τους βάσει των δεικτών εποχιακότητας που είχαν υπολογιστεί (βλ. Ενότητα 5.2) παρατίθεται στα ακόλουθα διαγράμματα:



Σχήμα 6.2 : Bar-Plot δείκτη sMAPE των σημειακών προβλέψεων



Σχήμα 6.3 : Bar-Plot δείκτη MASE των σημειακών προβλέψεων

Όπως προκύπτει από τα ανωτέρω Σχήματα, η κατανομή Laplace παρουσιάζει την καλύτερη επίδοση σε αμφότερους τους δείκτες ακριβείας, με το sMAPE της να ανέρχεται στο 13.612% και το MAsE στο 0.976 . Η καλύτερη επίδοση της Κατανομής Laplace έναντι της Normal (sMAPE 13.720%, MAsE 0.985) οφείλεται στο γεγονός ότι ο υπολογισμός του *maxlikelihood* στην περίπτωση της πρώτης βελτιστοποιείται για το MAE, το οποίο έχει την ίδια συμπεριφορά με τους δείκτες ακρίβειας MAsE και SMAPE. Ανίθετα, η Normal Κατανομή βελτιστοποιείται βάσει του MSE. Επιπλέον επιβεβαιώνεται και πειραματικώς η κοινή συμπεριφορά των κατανομών Normal και StudentT, καθώς η δεύτερη εντοπίζει την Κανονική κατανομή σε μικρά δείγματα. Πιο συγκεκριμένα το MAsE τους είναι σχεδόν ίδιο (0.985 έναντι 0.984 αντίστοιχα), με την StudentT όμως να υπερτερεί στην ακρίβεια του δείκτη sMAPE των σημειακών προβλέψεων κατά 0.103% λόγω της στενότερης «καμπάνας» που έχει γύρω από το mean σε σχέση με την Κανονική κατανομή. Τέλος οι κατανομές Cauchy και LogNormal η συνάρτηση πυκνότητας πιθανότητας των οποίων εμφανίζει μεγαλύτερες ουρές και μικρότερη καμπάνα, φαίνεται να παρουσιάζουν αισθητά χειρότερα αποτελέσματα στις σημειακές προβλέψεις έναντι των υπολοίπων τριών κατανομών.

## 6.2.2 Αποτελέσματα Πιθανοτικών Προβλέψεων - Σύγκριση Κατανομών

Για τη σύγκριση των πιθανοτικών προβλέψεων μεταξύ των πέντε διαφορετικών κατανομών που εξετάστηκαν και συγκεκριμένα της Cauchy, της Laplace, της LogNormal, της Normal και της StudentT, αξιοποιήθηκαν δύο στατιστικοί δείκτες ακρίβειας: το κανονικοποιημένο ως προς τις διαφορές με καθυστέρηση 12 περιόδων (όσο και το μήκος εποχιακότητας) Pinball Loss (Scaled Pinball Loss - SPL) και η Ρίζα Μέσου Τετραγωνικού Κανονικοποιημένου Σφάλματος (RMSSE). Αμφότεροι οι δείκτες που επιλέχθηκαν είναι κανονικοποιημένοι, καθώς όπως προκύπτει από τη βιβλιογραφία, οι μετρήσεις που βασίζονται σε κανονικοποιημένους δείκτες εμφανίζουν τις πλέον προτιμητέες στατιστικές ιδιότητες.

Μολονότι στη βιβλιογραφία το Pinball Loss χρησιμοποιείται συνήθως ως συνάρτηση απωλειών κατά την εκπαίδευση των νευρωνικών δικτύων, θεωρήθηκε ενδιαφέρον στην παρούσα διπλωματική εργασία να χρησιμοποιηθεί ως δείκτης αξιολόγησης. Λόγω του μαθηματικού του ορισμού

$$L(y,z) = \begin{cases} (y - z) \cdot \tau & , \text{αν } y \geq z \\ (z - y) \cdot (1 - \tau) & , \text{αν } z > y \end{cases}$$

όπου

y η πραγματική τιμή

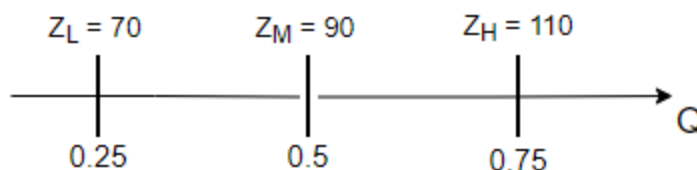
z η πρόβλεψη

$\tau$  το quantile

το Pinball Loss διαχειρίζεται με διαφορετικό τρόπο τα σφάλματα εισάγοντας τον παράγοντα  $\tau$  ή  $(1-\tau)$  για να “τιμωρήσει” με διαφορετικό συντελεστή το σχετικό σφάλμα, ανάλογα με το να η πρόβλεψη κινείται ή όχι στη σωστή κατεύθυνση.

Στην ουσία, κάθε φορά που εξάγεται μία quantile πρόβλεψη Q για ένα quantile  $\tau$ , λόγου χάρη το 0.05, υποτίθεται ότι η αναμενόμενη τιμή θα είναι ίση ή μεγαλύτερη από αυτή που προβλέφθηκε (Q) το 95% των φορών και μικρότερη το 5% των φορών.

Έστω για παράδειγμα ότι η πραγματική τιμή ενός μεγέθους που θέλουμε να προβλέψουμε είναι  $Y=100$ , και οι προβλέψεις που εξάγονται στα quantiles 0.25 , 0.5 και 0.75 είναι οι ακόλουθες:



Σχήμα 6.4 : Επεξηγηματικό παράδειγμα πιθανοτικών προβλέψεων

Τότε από το μαθηματικό ορισμό του Pinball Loss προκύπτει ότι τα σφάλματα στις τρεις περιπτώσεις θα είναι:

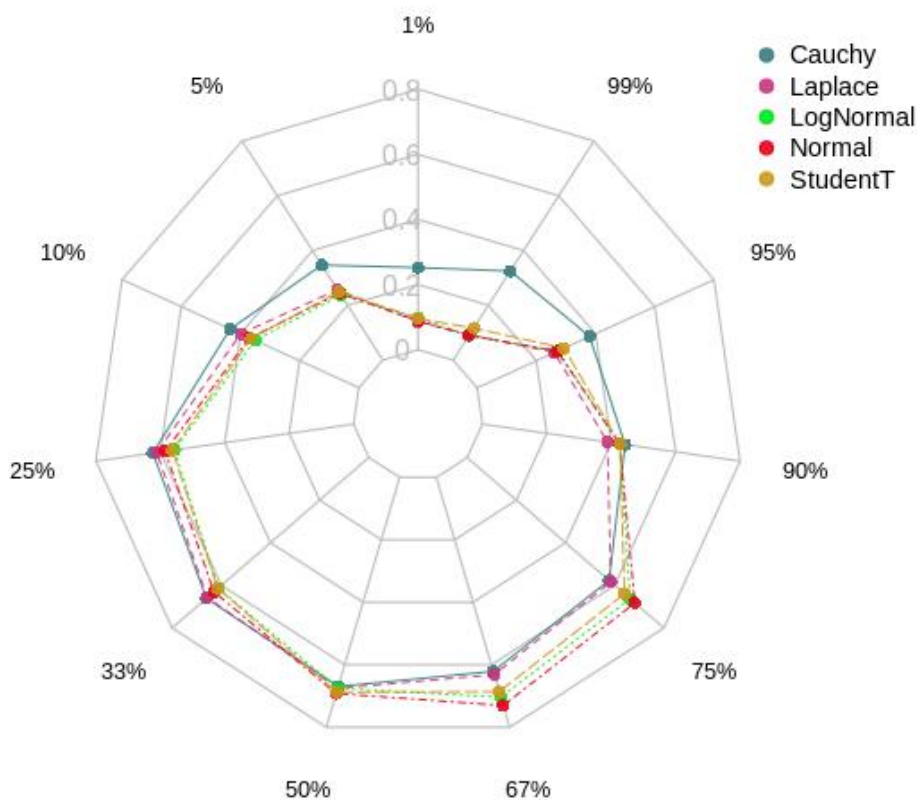
$$L_L = (100-70) * 0.25 = 7.5$$

$$L_M = (100-90) * 0.5 = 5$$

$$L_H = (110-100) * (1-0.75) = 2.5$$

Στις περιπτώσεις M και H, παρατηρούμε ότι ενώ το σχετικό σφάλμα είναι ίδιο ( $100-90 = 110-100$ ), το μικρότερο σφάλμα δίνεται στη δεύτερη περίπτωση καθώς η πρόβλεψη δίνεται προς τη σωστή κατεύθυνση.

Το διάγραμμα του Σχήματος 6.5 οπτικοποιεί την επίδοση της κάθε κατανομής ανά quantile. Η πληροφορία που εξάγεται αφορά τη σύγκριση των κατανομών σε επίπεδο quantile και όχι τη συνολική συμπεριφορά της εκάστοτε κατανομής, λόγω του διαφορετικού τρόπου με τον οποίο επιβάλλεται πέναλτυ στο σφάλμα ανάλογα με την κατεύθυνση της πρόβλεψης. Όσο μικρότερη είναι η τιμή του σφάλματος, ήτοι όσο πιο κοντά στο κέντρο του διαγράμματος κινούμαστε, τόσο πιο ακριβής είναι η πρόβλεψη που εξάγεται.



Σχήμα 6.5 : SPL ανά quantile για κάθε Κατανομή

Τα SPL παρατίθενται συγκεντρωτικά στον ακόλουθο Πίνακα:

	1%	5%	10%	25%	33%	50%	67%	75%	90%	95%	99%	Average
cauchy	0,252	0,347	0,435	0,623	0,661	<b>0,669</b>	<b>0,621</b>	<b>0,577</b>	0,443	0,380	0,325	0,485
laplace	<b>0,085</b>	0,257	0,399	0,611	0,657	0,676	0,631	0,584	<b>0,388</b>	<b>0,261</b>	0,093	<b>0,422</b>
lognormal	0,099	<b>0,238</b>	<b>0,350</b>	<b>0,557</b>	<b>0,612</b>	0,676	0,704	0,663	0,427	0,273	<b>0,090</b>	0,426
normal	0,089	0,244	0,369	0,585	0,630	0,692	0,730	0,682	0,426	0,270	0,090	0,437
studentT	0,096	0,249	0,364	0,563	0,614	0,688	0,686	0,640	0,427	0,293	0,116	0,431

Πίνακας 6.2 : SPL ανά κατανομή και quantile

Παρατηρούμε ότι όπως και στην περίπτωση των σημειακών προβλέψεων, η κατανομή Laplace παρουσιάζει την καλύτερη μέση επίδοση για το σύνολο των quantiles, με το μέσο SPL να ανέρχεται στο 0.422. Αξιοσημείωτη είναι και η επίδοση της κατανομής LogNormal, η οποία παράγει τις πιο ακριβείς προβλέψεις στα μικρά quantiles, γεγονός που αποδίδεται στη μορφή της καμπύλης της, καθώς η «καμπάνα» είναι μετατοπισμένη προς τα αριστερά. Επιπλέον οι κατανομές Normal και StudentT παρουσιάζουν παρόμοια μεταξύ τους συμπεριφορά, το οποίο είναι αναμενόμενο δεδομένου του ότι η τελευταία χρησιμοποιείται για την εκτίμηση της μέσης τιμής ενός πληθυσμού που κατανέμεται κανονικά σε περίπτωση που το δείγμα είναι μικρό και η τυπική απόκλιση άγνωστη.

Είναι προφανές ότι η ακρίβεια των προβλέψεων της εκάστοτε κατανομής σε κάθε quantile, είναι σε συμφωνία με τη μορφή της συνάρτησης πυκνότητας πιθανότητάς της. Πιο συγκεκριμένα, κατανομές οι οποίες έχουν μεγαλύτερες ουρές, όπως επί παραδείγματι η LogNormal που αναφέρθηκε παραπάνω, είναι σε θέση να προβλέπουν με μεγαλύτερη ακρίβεια τα ακραία quantiles, ενώ όσες κατανομές είναι στενότερες και έχουν εντονότερη καμπάνα με κέντρο τα μεσαία quantiles, τείνουν να εμφανίζουν καλύτερη επίδοση στις σημειακές προβλέψεις.

Στα Σχήματα 6.6 – 6.10 που ακολουθούν στη συνέχεια παρουσιάζονται διαγράμματα με τα coverage rates της κάθε κατανομής. Σκοπός της μελέτης των coverage rate είναι να ελεγχθεί εάν στη μέση περίπτωση για το μήκος του χρονικού ορίζοντα πρόβλεψης έχει γίνει υπέρ ή υποεκτίμηση του quantile που ζητείται κάθε φορά, σε σχέση με την ιδανική περίπτωση (ευθεία γραμμή των διαγραμμάτων). Το coverage ratio λαμβάνει τιμή ίση με τη μονάδα στην περίπτωση που η προβλεφθείσα τιμή είναι μικρότερη ή ίση της πραγματικής, ενώ στην αντίθετη περίπτωση η τιμή της είναι μηδενική. Η μαθηματική διατύπωση του RF είναι:

$$RF = \frac{1}{h} \sum_{t=1}^h \{1 | (F_t \leq Y_t) + 0 | (Y_t < F_t)\}$$

με

h μήκος χρονικού ορίζοντα πρόβλεψης

$F_t$  η προβλεφθείσα τιμή

$Y_t$  η πραγματική τιμή

Συνεπώς στο εκάστοτε διάγραμμα, εάν τα σημεία βρίσκονται κάτω από την ευθεία  $y=x$ , εξάγεται το συμπέρασμα ότι στο συγκεκριμένο ποσοστημόριο, η κατανομή υποεκτίμησε το quantile, ενώ



το αντίθετο συμβαίνει όταν η θέση των σημείων βρίσκεται πάνω από την εν λόγω ευθεία. Στον Πίνακα 6.3 παρατίθενται αναλυτικά τα αποτελέσματα της κάθε κατανομής ανά quantile, ενώ στον Πίνακα 6.4 έχουν υπολογιστεί οι διαφορές της επίδοσης κάθε κατανομής από την ιδανική περίπτωση. Αρνητικό αποτέλεσμα συνεπάγεται θετική προκατάληψη των προβλέψεων, αφού η προβλεφθείσα τιμή στο υπό εξέταση quantile είναι μεγαλύτερη από την πραγματική, ενώ θετικό αποτέλεσμα υποδηλώνει αρνητική προκατάληψη. Τέλος όσο πιο κοντά στο 0 είναι η τιμή σε κάθε κελί του Πίνακα 6.4, τόσο πιο συνεπής είναι η κατανομή στο αντίστοιχο quantile.

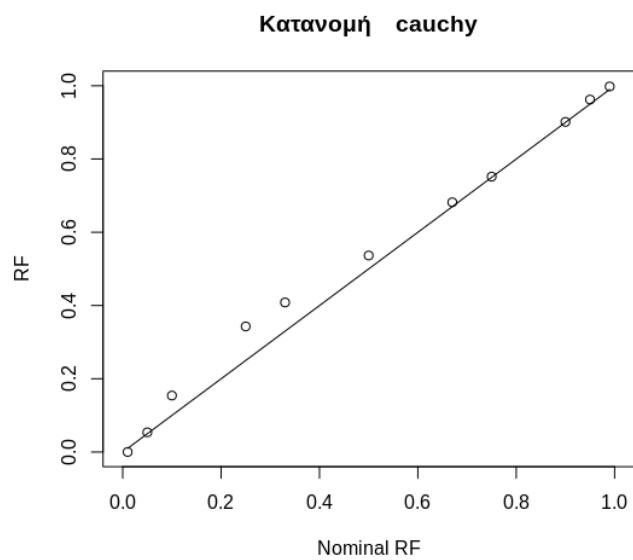
distribution	1%	5%	10%	25%	33%	50%	67%	75%	90%	95%	99%
optimal	<b>0,010</b>	<b>0,050</b>	<b>0,100</b>	<b>0,250</b>	<b>0,330</b>	<b>0,500</b>	<b>0,670</b>	<b>0,750</b>	<b>0,900</b>	<b>0,950</b>	<b>0,990</b>
cauchy	0,008	0,069	0,142	0,310	0,396	0,591	0,751	0,808	0,909	0,949	0,988
laplace	0,059	0,111	0,157	0,284	0,363	0,584	0,778	0,831	0,908	0,932	0,959
lognormal	0,079	0,130	0,174	0,293	0,364	0,560	0,763	0,826	0,913	0,939	0,967
normal	0,047	0,123	0,189	0,341	0,408	0,542	0,688	0,757	0,874	0,916	0,962
studentT	0,000	0,054	0,154	0,343	0,408	0,537	0,682	0,752	0,901	0,962	0,998

Πίνακας 6.3 : RF ανά κατανομή και quantile

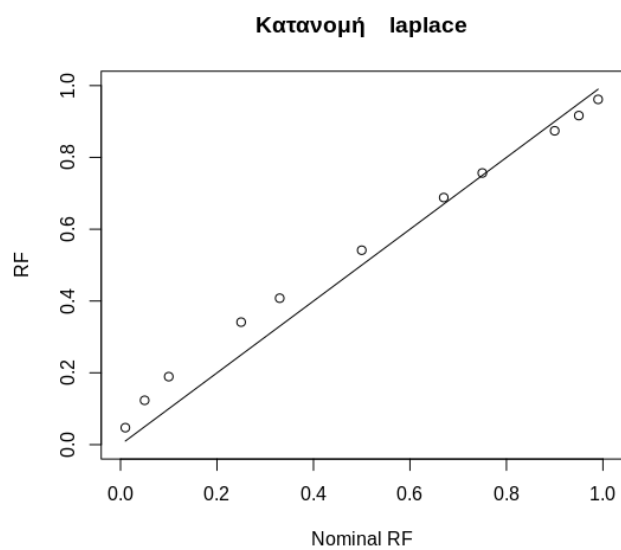
distribution	1%	5%	10%	25%	33%	50%	67%	75%	90%	95%	99%	Average
cauchy	<b>0,002</b>	-0,019	<b>-0,042</b>	-0,060	-0,066	-0,091	-0,081	-0,058	-0,009	<b>0,001</b>	<b>0,002</b>	-0,038
laplace	-0,049	-0,061	-0,057	<b>-0,034</b>	<b>-0,033</b>	-0,084	-0,108	-0,081	-0,008	0,018	0,031	-0,042
lognormal	-0,069	-0,080	-0,074	-0,043	-0,034	-0,060	-0,093	-0,076	-0,013	0,011	0,023	-0,046
normal	-0,037	-0,073	-0,089	-0,091	-0,078	-0,042	-0,018	-0,007	0,026	0,034	0,028	-0,032
studentT	0,010	<b>-0,004</b>	-0,054	-0,093	-0,078	<b>-0,037</b>	<b>-0,012</b>	<b>-0,002</b>	<b>-0,001</b>	-0,012	-0,008	<b>-0,026</b>

Πίνακας 6.4 : Απόκλιση RF κάθε κατανομής από την ιδανική περίπτωση

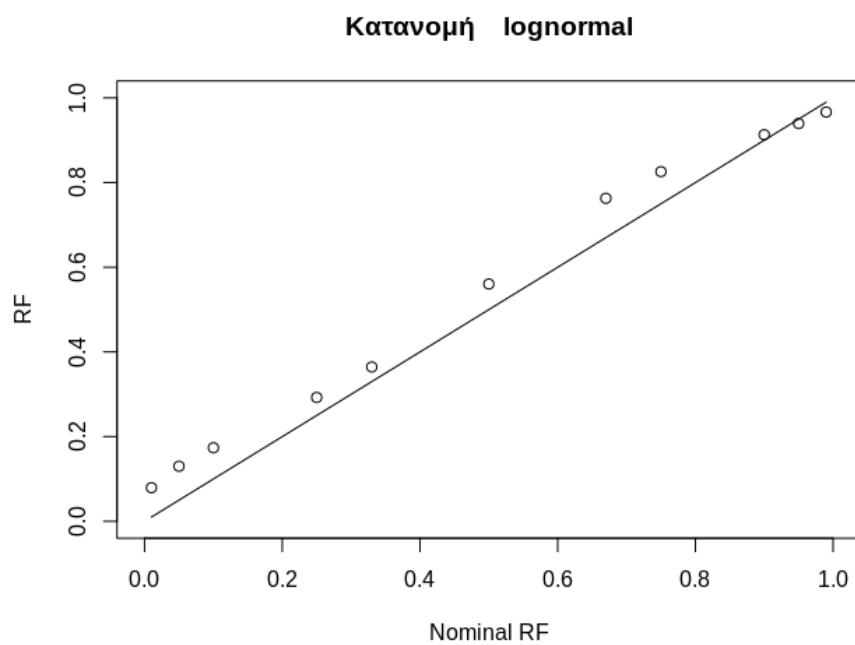
Από τον Πίνακα 6.4 φαίνεται ότι εν γένει όλες οι κατανομές υπερεκτιμούν την πραγματική τιμή, καθώς η μέση σχετική διαφορά σε κάθε περίπτωση είναι αρνητική. Εξαιρετικό ενδιαφέρον παρουσιάζει η καλή επίδοση της κατανομής StudentT η οποία στις μισές περιπτώσεις (5 στα 11 quantiles) φαίνεται να επιστρέφει το quantile που της ζητήθηκε, καθώς επίσης και η συνέπεια της κατανομής Cauchy στα ακραία quantiles.



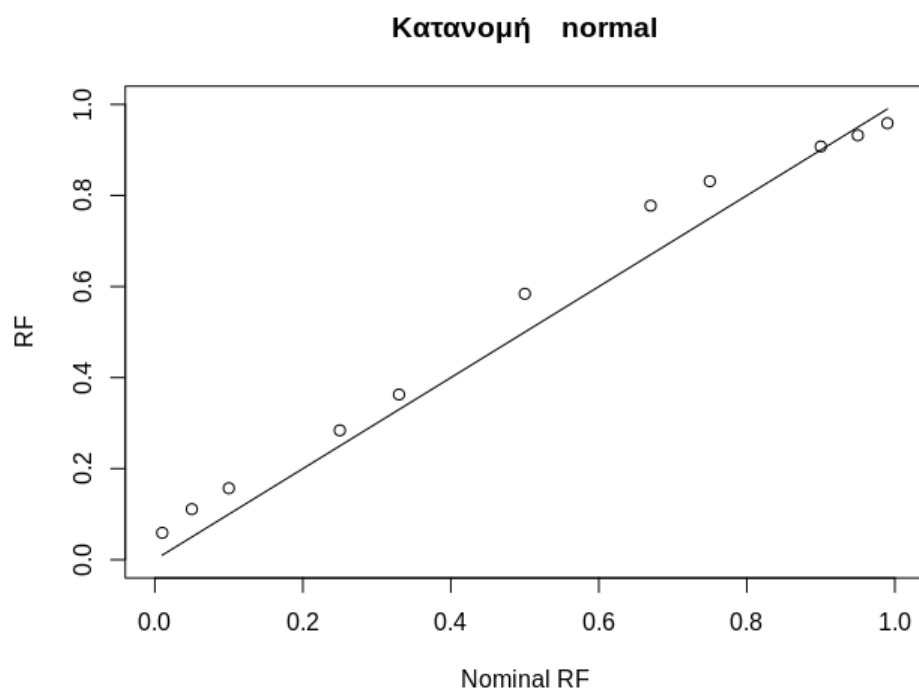
*Σχήμα 6.6 : Coverage Rate Κατανομής Cauchy*



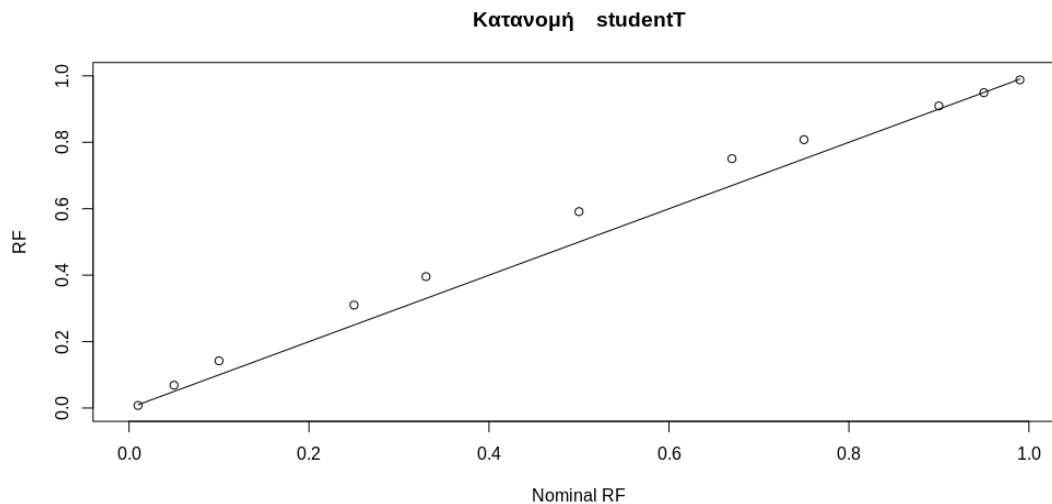
*Σχήμα 6.7 : Coverage Rate Κατανομής Laplace*



*Σχήμα 6.8 : Coverage Rate Κατανομής LogNormal*

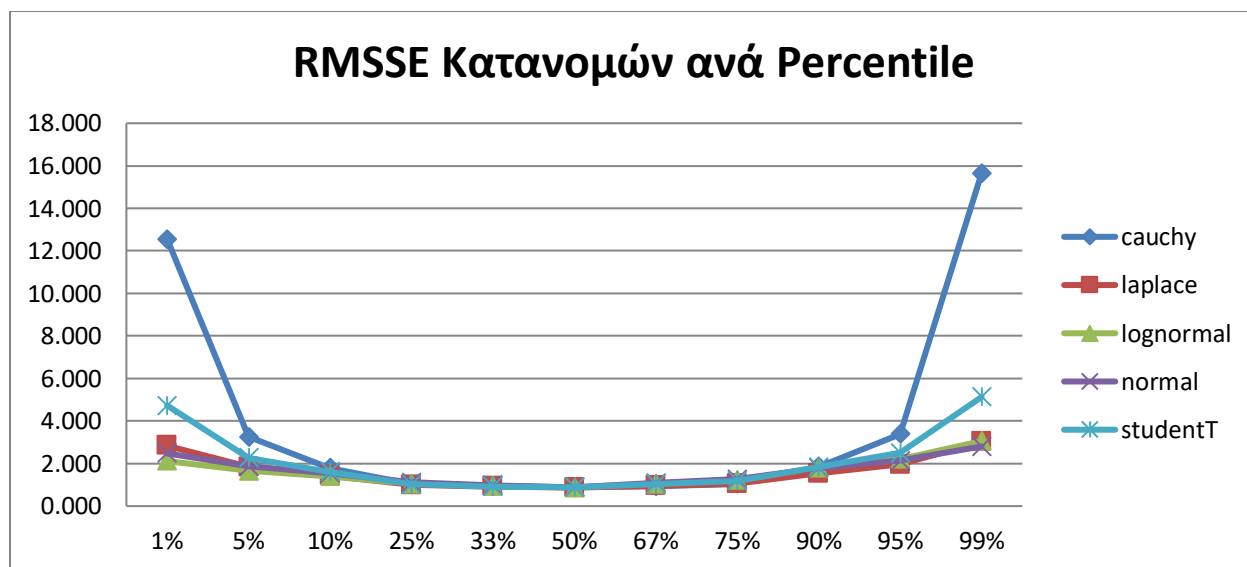


*Σχήμα 6.9 : Coverage Rate Κατανομής Normal*



Σχήμα 6.10 : Coverage Rate Κατανομής StudentT

Η εξέταση των παραπάνω διαγραμμάτων πρέπει να συνδυαστεί με την εξέταση του δείκτη ακριβείας RMSSE, καθώς διαφορετική προσέγγιση ενδέχεται να οδηγήσει σε παραπλανητικά αποτελέσματα. Ο δείκτης RMSSE αναδεικνύει το πόσο μακριά είναι η τελική πρόβλεψη από την πραγματική, χωρίς να διαχειρίζεται με διαφορετικό τρόπο το σχετικό σφάλμα όπως γίνεται στην περίπτωση του SPL.



Σχήμα 6.11 : RMSSE σφάλματα πιθανοτικών προβλέψεων της κάθε κατανομής ανά quantile

	1%	5%	10%	25%	33%	50%	67%	75%	90%	95%	99%	Average
cauchy	12,556	3,256	1,797	1,025	0,939	0,880	<b>0,941</b>	<b>1,043</b>	1,859	3,384	15,650	3,939
laplace	2,849	1,834	1,431	<b>0,997</b>	0,923	<b>0,876</b>	0,953	1,058	<b>1,535</b>	<b>1,963</b>	3,019	1,585
lognormal	<b>2,126</b>	<b>1,650</b>	<b>1,405</b>	1,053	0,953	0,877	1,050	1,218	1,799	2,208	3,057	<b>1,582</b>
normal	2,487	1,861	1,551	1,111	0,972	0,883	1,088	1,258	1,787	2,134	<b>2,805</b>	1,631
studentT	4,732	2,278	1,609	1,030	<b>0,921</b>	0,883	1,042	1,193	1,832	2,505	5,132	2,105

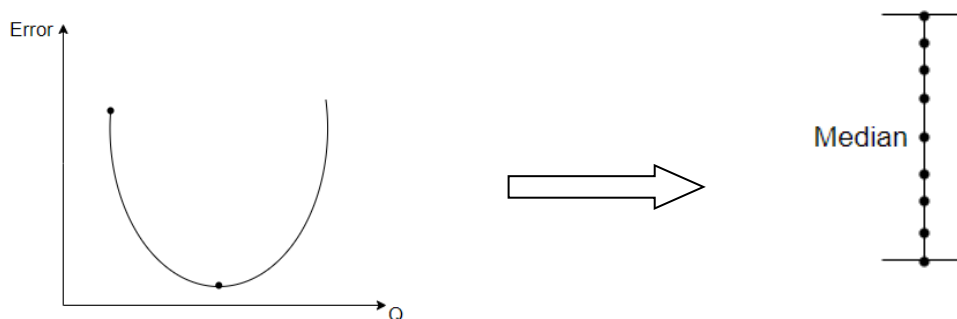
Πίνακας 6.5 : Συγκεντρωτικά αποτελέσματα RMSSE ανά κατανομή και quantile

Από το παραπάνω Σχήμα καθίσταται προφανής ο λόγος για τον οποίο τα διαγράμματα RF εξετάζονται μαζί με το RMSSE. Πιο συγκεκριμένα ενώ η κατανομή Cauchy είναι η πλέον συνεπής στην περίπτωση που ζητούνται ακραία quantiles, όπως επί παραδείγματι στο 1%, στο 95% και στο 99%, η πρόβλεψη καθαυτή αποκλίνει κατά πολύ από την πραγματική τιμή.

Όπως και στην περίπτωση του SPL, έτσι και στο RMSSE, παρατηρούμε ότι η ακρίβεια των προβλέψεων της κάθε κατανομής είναι σε συμφωνία με τη μορφή της καμπύλης πυκνότητας πιθανότητάς της. Συνεπώς, η κατανομή LogNormal, της οποίας η κορυφή είναι μετατοπισμένη προς τα αριστερά, εμφανίζει καλύτερη επίδοση στα μικρότερα quantiles. Αντιστοίχως, στην περιοχή των ενδιάμεσων quantile καλύτερες επιδόσεις εμφανίζουν κατανομές με έντονη την κορυφή στο median, όπως επί παραδείγματι η Laplace.

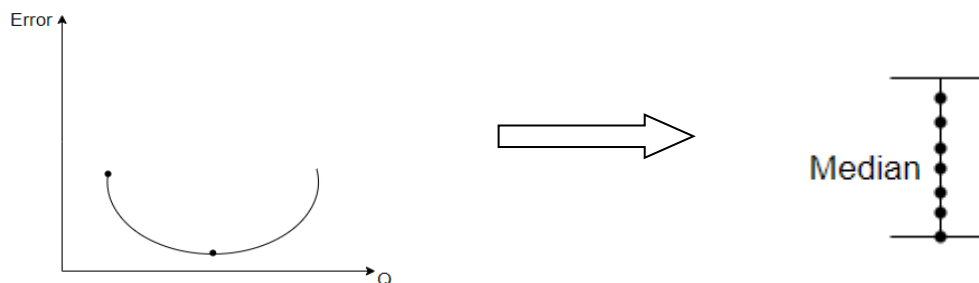
Πέρα από το αμιγώς ποσοτικό μέρος, στο οποίο η μεγαλύτερη προβλεπτική ακρίβεια αντιστοιχεί στην καμπύλη που βρίσκεται χαμηλότερα στον κατακόρυφο άξονα, η μορφή καθαυτή της καμπύλης των RMSSE δίνει επιπρόσθετη πληροφορία σχετικά με τη συμπεριφορά των προβλέψεων που εξάγει η κάθε κατανομή, μέσω της εξέτασης της κυρτότητάς της.

Όσο πιο κυρτή είναι η καμπύλη, δηλαδή όσο μεγαλύτερη είναι η κατακόρυφη απόσταση των ακραίων quantile από το median, τόσο μεγαλύτερο είναι το prediction interval και άρα πιο “απλωμένες” οι προβλέψεις που εξάγονται σε σχέση με τη σημειακή πρόβλεψη (median - quantile: 50%) για τη ζητούμενη χρονική στιγμή του ορίζοντα πρόβλεψης, όπως παρουσιάζεται στο



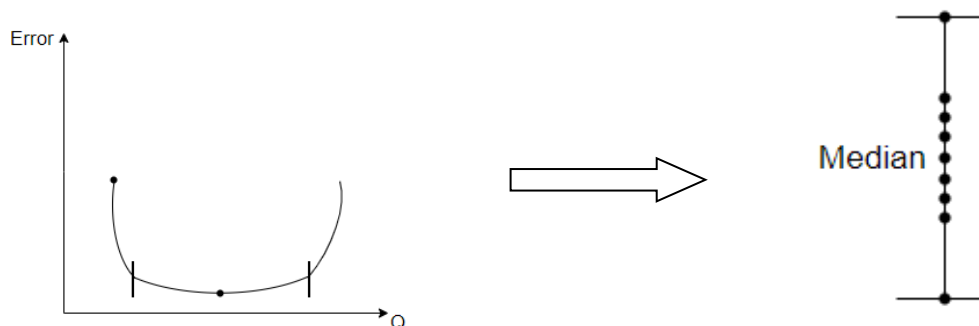
Σχήμα 6.12 : Μεγάλες διαφορές σφαλμάτων μεταξύ ακραίων quantiles και median συνεπάγονται πιο διασκορπισμένες προβλέψεις

Αντίθετα, όσο μικρότερη είναι η κυρτότητα της καμπύλης, τόσο πιο “συγκεντρωμένες” γύρω από το median θα είναι οι προβλέψεις για τη δεδομένη χρονική στιγμή του ορίζοντα πρόβλεψης.



Σχήμα 6.13 : Μικρές διαφορές σφαλμάτων μεταξύ ακραίων quantiles και median συνεπάγονται πιο συμπυκνωμένες προβλέψεις

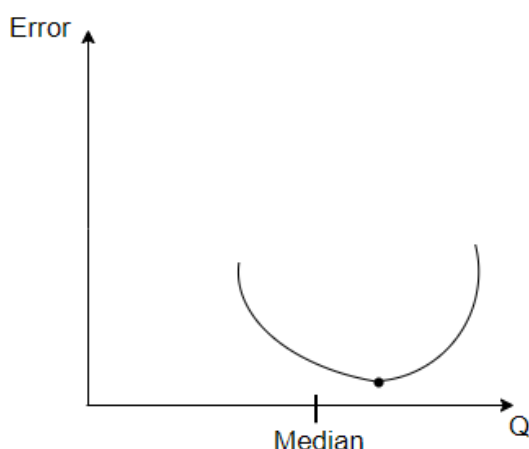
Επιπλέον υπάρχει η περίπτωση η κλίση της καμπύλης στα ακραία quantiles να είναι μεγάλη, γεγονός που υποδηλώνει ότι ενώ οι προβλέψεις κυμαίνονται σε μικρή απόσταση γύρω από το median, τα ακραία quantiles ξεφεύγουν αρκετά.



Σχήμα 6.14 : Περίπτωση σχετικά συγκεντρωμένων προβλέψεων, με τα ακραία quantiles όμως να αποκλείουν κατά πολύ.

Σε όλες τις παραπάνω περιπτώσεις η ελάχιστη τιμή της καμπύλης σφάλματος είναι στο median (quantile 50%). Σε περίπτωση που κάτι τέτοιο δεν ισχύει, όπως επί παραδείγματι στο ακόλουθο Σχήμα 6.8 όπου το ολικό ελάχιστο είναι μετατοπισμένο προς τα δεξιά (έστω στο 0.6 quantile),

τότε συνάγεται το συμπέρασμα ότι το μοντέλο εμφανίζει ένα συστηματικό σφάλμα προκατάληψης.



Σχήμα 6.15 : Περίπτωση όπου το μοντέλο εισάγει συστηματικό σφάλμα στις προβλέψεις

Σύμφωνα με τα αποτελέσματα που προέκυψαν από την πειραματική διαδικασία που ακολουθήθηκε, οι κατανομές Laplace, LogNormal και Normal είναι αυτές που παράγουν τις λιγότερο διασκορπισμένες προβλέψεις, με ικανοποιητική ακρίβεια και στα ακραία quantiles, γεγονός που δεν παρατηρείται στην περίπτωση της StudentT, η οποία ενώ παράγει σχετικά ακριβείς προβλέψεις στα ενδιάμεσα quantiles, έχει αισθητά χειρότερη επίδοση στο για τα ποσοστημόρια του 1% και του 99%. Τέλος, αναφορικά με την κατανομή Cauchy ενώ στο διάστημα [25%,75%] εξάγει ακριβείς προβλέψεις (μάλιστα στο 65% και το 75% έχει τις πιο ακριβείς προβλέψεις), η επίδοσή της στα ακραία quantiles είναι εξαιρετικά ανεπαρκής.

Επιπλέον, παρατηρείται μία αντίθετη συμπεριφορά της καμπύλης του RMSSE σε όλες τις κατανομές σε σχέση με αυτή του SPL, καθώς κατά την κίνηση προς το mean το σφάλμα φαίνεται να αυξάνει στη δεύτερη περίπτωση. Το προαναφερθέν οφείλεται στο διαφορετικό τρόπο με τον οποίο προκύπτει το SPL στα ακραία quantile από τα ενδιάμεσα, για το λόγο αυτό επισημάνθηκε παραπάνω το γεγονός ότι το διάγραμμα του SPL διαβάζεται σε επίπεδο quantile και όχι κατανομής.

Δεδομένων όλων των προαναφερθέντων, διαπιστώνεται η επιβεβαίωση της αμφισβήτησης που επιχειρεί να εγείρει η παρούσα διπλωματική εργασία αναφορικά με την υπόθεση της Κανονικής Κατανομής κατά την εξαγωγή των πιθανοτικών προβλέψεων. Αν γίνει χρήση της Κανονικής Κατανομής, τότε η εκτίμηση της μέγιστης πιθανοφάνειας της παραμέτρου (maximum likelihood estimate) της διασποράς δίνεται από τη σχέση:

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu}_{y,t})^2$$

όπου  $T$  είναι το μέγεθος του δείγματος και  $\hat{\mu}_{y,t}$  η εκτίμηση της μέσης τιμής  $\mu_{y,t}$ . Γίνεται εμφανές το γεγονός ότι η παραπάνω μαθηματική διατύπωση συμπίπτει με το Μέσο Τετραγωνικό Σφάλμα –MSE (Ενότητα 2.10.3), με αποτέλεσμα όταν χρησιμοποιείται για την εκτίμηση μέγιστης πιθανοφάνειας κατά την εκπαίδευση ενός μοντέλου, να δίνει τις ίδιες παραμέτρους που θα έδινε η χρήση της ελαχιστοποίησης του MSE. Έτσι το log-likelihood που περιγράφηκε στην αντίστοιχη ενότητα της παρούσας εργασίας λαμβάνει την ακόλουθη μορφή:

$$\ell(Y|\theta, \sigma^2) = -\frac{T}{2} \log(2\pi\sigma^2) - \sum_{t=1}^T \frac{(y_t - \hat{\mu}_{y,t})^2}{2\sigma^2}$$

όπου  $\theta$  είναι το διάνυσμα των εκτιμώμενων παραμέτρων του μοντέλου και  $\log$  ο φυσικός αλγόριθμος.

Στον αντίποδα, η χρήση της κατανομής Laplace για την εκτίμηση της μέγιστης πιθανοφάνειας συμπίπτει με τη χρήση της ελαχιστοποίησης του Μέσου Απολύτου Σφάλματος – MAE (Ενότητα 2.10.2), καθώς η παράμετρος κλίμακας (Ενότητα 5.2.2), εκτιμάται μέσω της σχέσης

$$\hat{s}^2 = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{\mu}_{y,t}|$$

Έτσι η διασπορά της τυχαίας μεταβλητής που εκτιμάται θα είναι  $\sigma^2 = 2s^2$ .

Τέλος το γεγονός ότι η LogNormal εμφανίζει καλύτερη συμπεριφορά στα ακραία quantiles, οφείλεται στο γεγονός ότι η διαφορά εν γένει των Log-παραλλαγών των πυκνοτήτων κατανομών έναντι των αρχικών συνοψίζεται ως εξής:

$$f_{\log}(\log(y_t)) = \frac{1}{y_t} f(\log(y_t))$$

Απότοκο του προαναφερθέντος αποτελεί το γεγονός ότι η καμπύλη τους θα εμφανίζει διαφορετική δεξιά ουρά, ανάλογα με τις παραμέτρους του location, του scale και της μορφής οι οποίες περιγράφηκαν στις αντίστοιχες ενότητες της εργασίας αυτής.



### 6.3 Χρόνοι Εκπαίδευσης

Τα πειράματα διενεργήθηκαν σε τοπικό server της Μονάδας Προβλέψεων και Στρατηγικής του Εθνικού Μετσόβιου Πολυτεχνείου, ο οποίος προσφέρει δύο Κεντρικές Μονάδες Επεξεργασίας (CPU) Intel Xeon Gold 5118 @ 2.30 GHz συνολικής μνήμης RAM 64GB και δύο κάρτες γραφικών (GPU) Quadro P5000 με συνολική μνήμη RAM μεγέθους 32 GB.

Σε κάθε πείραμα διαφορετικής κατανομής, τα 30 μοντέλα εκπαιδεύτηκαν για 20 εποχές έκαστο στη CPU, με τις απαιτήσεις σε χρόνο και υλικό να φαίνονται στον ακόλουθο Πίνακα:

# models	# epochs	distribution	RAM (GB)	total execution time
30	20	StudentT	20	01:16:13
30	20	Normal	15,8	00:42:27
30	20	Cauchy	13	01:08:21
30	20	LogNormal	10,4	00:44:01
30	20	Laplace	9,8	00:33:26

Πίνακας 6.6 : Απαιτηση κάθε κατανομής σε χρόνο και μνήμη

### 6.4 Συμπεράσματα και Προεκτάσεις

Από τα αποτελέσματα που προέκυψαν και τα οποία παρουσιάστηκαν παραπάνω, γίνεται εμφανής η δυναμική της χρήσης των αλγορίθμων μηχανικής και βαθιάς μηχανικής μάθησης στον κλάδο των Προβλέψεων και δη των πιθανοτικών προβλέψεων. Λαμβάνοντας δε υπόψη την ύπαρξη εργαλειοθηκών όπως αυτή του GluonTS, που μελετήθηκε στο πλαίσιο της παρούσας εργασίας, εργαλειοθήκες οι οποίες παρέχουν έτοιμα μοντέλα αλλά και πλαίσια ανάπτυξης νέων εξατομικευμένων μοντέλων, η προαναφερθείσα δυναμική ενισχύεται έτι περαιτέρω.

Αναφορικά με τα ποσοτικά αποτελέσματα, εξάγονται τα κάτωθι συμπεράσματα:

- στις σημειακές προβλέψεις η κατανομή Laplace παρουσιάζει τις πιο ακριβείς προβλέψεις (sMAPE = 13.612% και MASE = 0.976), καθώς λόγω φύσης της συνάρτησης πυκνότητας πιθανότητας, προβλέπει καλύτερα την αναμενόμενη τιμή (mean), γεγονός που την καθιστά τη βέλτιστη επιλογή σε περιπτώσεις όπου η σημειακή πρόβλεψη εξάγεται από το σύνολο των πιθανοτικών που έχουν παραχθεί
- Για τις εφαρμογές πιθανοτικών προβλέψεων αποδεικνύεται χρήσιμη η αξιοποίηση διαφορετικών Κατανομών ανά quantile και πιο συγκεκριμένα στην αριστερή ουρά της LogNormal, στη μέση της Laplace και στη δεξιά ουρά της Cauchy
- Επιβεβαιώνεται η βασική αμφισβήτηση που επιχειρήσε να αναδείξει η παρούσα διπλωματική, ότι η χρήση της κανονικής κατανομής για των προσδιορισμό των

παραμέτρων της κατανομής των πιθανοτικών προβλέψεων, δεν έχει απαραίτητα ως αποτέλεσμα τα βέλτιστα αποτελέσματα

Δεδομένου του γεγονότος ότι ο παράγοντας της αβεβαιότητας λαμβάνεται πλέον υπόψη κατά τη λήψη οποιασδήποτε επιχειρησιακής απόφασης, καθίσταται σαφής η σημασία των πιθανοτικών προβλέψεων στο σύγχρονο περιβάλλον. Μέσω των τελευταίων, ο αποφασίζων έχει τη δυνατότητα να εξετάσει όλες τις πιθανές εναλλακτικές έχοντας παράλληλα στη διάθεσή του και την πιθανότητα να λάβει χώρα το εκάστοτε ενδεχόμενο.

Παράλληλα, η γενικότερη τάση για στροφή σε Ανανεώσιμες Πηγές Ενέργειας σε ευρωπαϊκό αλλά και παγκόσμιο επίπεδο θα ενισχύσει έτι περαιτέρω τις πιθανοτικές προβλέψεις. Ιδιαίτερα για την περίπτωση της λειτουργίας και του ελέγχου των αιολικών πάρκων, η ακριβής και αξιόπιστη πρόβλεψη είναι ζωτικής σημασίας. Δεδομένης της αστάθειας των χρονοσειρών της αιολικής ενέργειας, οι παραδοσιακές πιθανοτικές προβλέψεις αποτυγχάνουν να προβλέψουν τη μελλοντική κατάσταση, γεγονός που εγκυμονεί ρίσκο για τη διαχείριση του Συστήματος Ηλεκτρικής Ενέργειας. Συνεπώς, μοντέλα όπως αυτά που χρησιμοποιήθηκαν στο πλαίσιο της παρούσας διπλωματικής εργασίας, είναι δυνατόν να βρουν άμεσα εφαρμογή, ειδικά στον ελλαδικό χώρο ο οποίος πληροί τις προϋποθέσεις ανάπτυξης αιολικών πάρκων.

Επιπρόσθετη εφαρμογή στο χώρο της ενέργειας των μοντέλων που εξετάστηκαν, αποτελεί η πρόβλεψη της ωριαίας τιμής για την επόμενη ημέρα διαπραγμάτευσης στο χρηματιστήριο ενέργειας. Η φύση των αγορών αυτών καθιστά αδήριτη την ανάγκη να ληφθούν υπόψη πολλαπλές επεξηγηματικές μεταβλητές, γεγονός που αυξάνει την πολυπλοκότητα του προβλήματος και ενισχύει το πλεονέκτημα των μοντέλων βαθιάς μηχανικής μάθησης έναντι των πιο συμβατικών μεθόδων.

## BIBΛΙΟΓΡΑΦΙΑ

- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai and Soumith Chintala. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019
- Alexander Alexandrov, Konstantinos Benidis, Michael Bohlke-Schneider, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Danielle C. Maddix, Syama Rangapuram, David Salinas, Jasper Schulz, Lorenzo Stella, Ali Caner Turkmen, and Yuyang Wang. *GluonTS: Probabilistic Time Series Models in Python*. 2019
- Can Wan , Zhao Xu, Pierre Pinson, Zhao Yang Dong and Kit Po Wong. *Probabilistic forecasting of wind power generation using extreme learning machine*. 2014
- Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. *Activation Functions: Comparison of Trends in Practice and Research for Deep Learning*. 2018
- DeepAR Hyperparameters ,  
[https://docs.aws.amazon.com/sagemaker/latest/dg/deepar\\_hyperparameters.html](https://docs.aws.amazon.com/sagemaker/latest/dg/deepar_hyperparameters.html).  
Accessed on [2021-10-18].
- Everette Gardner. *Exponential smoothing: The state of the art*. 1985
- Fei-Fei Li and Jia Li. *Cloud AutoML: Making AI accessible to every business*. 2018
- Haifeng Jin, Qingquan Song and Xia Hu. *Auto-Keras: An Efficient Neural Architecture Search System*. 2019
- [https://github.com/lostella/ISF-2020-Deep-Learning-Workshop/blob/main/notebooks/GluonTS\\_ISF\\_2020\\_workshop.ipynb](https://github.com/lostella/ISF-2020-Deep-Learning-Workshop/blob/main/notebooks/GluonTS_ISF_2020_workshop.ipynb). Accessed on [2021-10-18].
- <https://github.com/Mcompetitions/M4-methods/blob/master/Evaluation%20and%20Ranks.xlsx> Accessed on [2021-10-18].
- <https://pytorch.org/docs/stable/distributions.html>. Accessed on [2021-10-18].
- <https://www.ibm.com/cloud/learn/deep-learning> , Accessed on [2021-10-18].
- <https://www.lokad.com/probabilistic-forecasting-definition>. Accessed on [2021-10-18].

- Ivan Svetunkov. *Statistics for Business Analytics*. 2021
- Jie Yan, Yongqian Liu, Shuang Han, Yimei Wang and Shuanglei Feng. *Reviews on uncertainty analysis of wind power forecasting*. 2015
- José Andrade, Jorge Filipe, Marisa Reis and Ricardo Bessa. *Probabilistic Price Forecasting for Day-Ahead and Intraday Markets: Beyond the Statistical Model*. 2017
- Kostas Hatalis. *Probabilistic Forecasting: How and Why?*. <https://www.slideshare.net/KostasHatalis/probabilistic-forecasting-how-and-why>. Accessed on [2021-10-18].
- Kurt Hornik, Maxwell Stinchcombe and Halbert White. *Multilayer feedforward networks are universal approximators*. 1989
- Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. *TensorFlow: a system for large-scale machine learning*. 2016
- Olivier Bousquet, Ulrike von Luxburg and Gunnar Rätsch. *Advanced Lectures on Machine Learning*. 2004
- Pádraig Cunningham, Matthieu Cord and Sarah Delany. *Supervised Learning*. 2008
- Rich Caruana and Alexandru Niculescu-Mizil. *An empirical comparison of supervised learning algorithms*. 2006
- Rob Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. 2018
- Simon Mason and Gillian Mimmack. *Comparison of Some Statistical Methods of Probabilistic Forecasting of ENSO*. 2002
- Simon Haykin. *Νευρωνικά Δίκτυα και Μηχανική Μάθηση*. 2010
- Spyros Makridakis, Evangelos Spiliotis and Vassilios Assimakopoulos. *The M4 Competition: 100,000 time series and 61 forecasting methods*. 2020
- Spyros Makridakis, Evangelos Spiliotis and Vassilios Assimakopoulos. *The M5 Accuracy competition: Results, findings and conclusions*. 2020
- Spyros Makridakis, Evangelos Spiliotis and Vassilios Assimakopoulos. *Statistical and Machine Learning forecasting methods: Concerns and ways forward*. 2018

- Théo Stassen. *Master's Thesis : Comparison of probabilistic forecasting deep learning models in the context of renewable energy production*. 2020
- Vassilios Assimakopoulos and Konstantinos Nikolopoulos. *The theta model: a decomposition approach to forecasting*. 2000
- Warren McCulloch and Walter Pitts. *A logical calculus of the ideas immanent in nervous activity*. 1943
- Yao Zhang, Jianxue Wang and Xifan Wang. *Review on probabilistic forecasting of wind power generation*. 2014
- Yves Chauvin, and David Rumelhart. *Backpropagation: Theory, Architectures, and Applications*. 1994
- Φώτιος Πετρόπουλος και Βασίλειος Ασημακόπουλος. *Επιχειρησιακές Προβλέψεις. Συμμετρία*. 2013



## ΠΑΡΑΡΤΗΜΑ Α

### ΔΙΑΓΩΝΙΣΜΟΣ Μ4

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση και την αξιολόγηση της απόδοσης των νευρωνικών δικτύων που χρησιμοποιήθηκαν για τους σκοπούς της παρούσας διπλωματικής εργασίας, ελήφθησαν από το διαγωνισμό Μ4. Οι διαγωνισμοί Μ ή Makridakis Competitions αποτελούν μία σειρά ανοικτών διαγωνισμών, οι οποίοι διοργανώνονται από τον Καθηγητή Σπύρο Μακρυδάκη. Η αρχή έγινε το 1982 με το διαγωνισμό Μ1, ενώ ο τελευταίος διαγωνισμός (Μ5) πραγματοποιήθηκε το 2020. Σημαντικό ρόλο στη διοργάνωση των δύο τελευταίων διαγωνισμών (Μ4 και Μ5) διαδραμάτισε η Μονάδα Προβλέψεων και Στρατηγικής του Εθνικού Μετσόβιου Πολυτεχνείου.

Σκοπός των διαγωνισμών αυτών είναι η μελέτη, η αξιολόγηση και η σύγκριση διαφορετικών προβλεπτικών μεθόδων και μοντέλων, τα οποία υποβάλλουν ειδικοί, ερευνητές, ακόμα και φοιτητές οι οποίοι συμμετέχουν στον εκάστοτε διαγωνισμό. Οι υποβολές των διαγωνιζόμενων αξιολογούνται βάσει των πραγματικών δεδομένων, άλλων μεθόδων-σημείων αναφοράς (benchmark μέθοδοι) αλλά και των υπολοίπων υποβολών. Απώτερο της όλης διαδικασίας είναι η εξαγωγή συμπερασμάτων αναφορικά με τις βέλτιστες μεθόδους πρόβλεψης ποικίλων μεταβλητών ενδιαφέροντος, γεγονός που επιβεβαιώνεται από το μεγάλο ενδιαφέρον που δείχνουν για το διαγωνισμό τόσο η ακαδημαϊκή κοινότητα, όσο και επαγγελματίες του χώρου. Όπως χαρακτηριστικά αναφέρει ο Rob J. Hyndman “...ο Σπύρος αξίζει συγχαρητήρια για την αλλαγή στον ερευνητικό κλάδο των προβλέψεων μέσω της σειράς των διαγωνισμών”, καθώς “...εστιάζουν την προσοχή στο ποια μοντέλα παράγουν καλές προβλέψεις, αντί των μαθηματικών ιδιοτήτων τους”.

Ο διαγωνισμός Μ4, τα μηνιαία δεδομένα του οποίου χρησιμοποιήθηκαν στο πλαίσιο της μελέτης που διενεργήθηκε για τους σκοπούς της παρούσας διπλωματικής εργασίας, διενεργήθηκε το διάστημα Ιανουάριος-Μάιος 2018. Ειδοποιός διαφορά του σε σχέση με τους τρεις προηγούμενους διαγωνισμούς ήταν ο μεγάλος αριθμός χρονοσειρών (100,000 στο σύνολο), οι οποίες μάλιστα χαρακτηρίζονταν από έντονη ποικιλία, καθώς σκοπός του διαγωνισμού ήταν η εύρεση τρόπου βελτίωσης της ακρίβειας των προβλέψεων και ο προσδιορισμός των καταλληλότερων μεθόδων για την κάθε περίπτωση. Επιπλέον, σημαντική διαφοροποίηση έναντι των προηγούμενων διαγωνισμών αποτέλεσαν η επέκταση του αριθμού των προβλεπτικών μεθόδων που χρησιμοποιήθηκαν καθώς και το γεγονός ότι η αξιολόγηση των υποβολών έγινε τόσο στις σημειακές προβλέψεις, όσο και στα διαστήματα εμπιστοσύνης. Αναφορικά με τις μεθόδους και τα μοντέλα που χρησιμοποιήθηκαν, θα πρέπει να σημειωθεί ότι στο διαγωνισμό αυτό εισήχθησαν μοντέλα και μέθοδοι που βασίζονταν στα νευρωνικά δίκτυα, πέρα των απλών στατιστικών που υποβλήθησαν στους τρεις προηγούμενους διαγωνισμούς.

Τα δεδομένα του διαγωνισμού (100,000 χρονοσειρές) κατηγοριοποιήθηκαν σε έξι κατηγορίες βάσει συχνότητας δεδομένων και έξι κατηγορίες βάσει κλάδου εφαρμογής, με την αξιολόγηση των σημειακών προβλέψεων και των διαστημάτων εμπιστοσύνης να γίνεται για κάθε μία από τις

προαναφερθείσες κατηγορίες. Το πλήθος των χρονοσειρών ανά κατηγορία παρατίθεται στον κάτωθι πίνακα:

Χρονικά διαστήματα μεταξύ διαδοχικών παρατηρήσεων	Μικροοικονομία	Βιομηχανία	Μακροοικονομία	Χρηματοοικονομικά	Δημογραφικά Στοιχεία	Άλλα	ΣΥΝΟΛΟ
Ετήσια	6,538	3,716	3,903	6,519	1,088	1,236	<b>23,000</b>
Τετραμηνιαία	6,020	4,637	5,315	5,305	1,858	865	<b>24,000</b>
Μηνιαία	10,975	10,017	10,016	10,987	5,728	277	<b>48,000</b>
Εβδομαδιαία	112	6	41	164	24	12	<b>359</b>
Ημερήσια	1,476	422	127	1,559	10	633	<b>4,227</b>
Ωριαία	0	0	0	0	0	414	<b>414</b>
<b>ΣΥΝΟΛΟ</b>	<b>25,121</b>	<b>18,798</b>	<b>19,402</b>	<b>24,534</b>	<b>8,708</b>	<b>3,437</b>	<b>100,000</b>

*ΠΙΝΑΚΑΣ Π.Α.1: Πλήθος χρονοσειρών διαγωνισμού M4 ανά κατηγορία και χρονικό επίπεδο*

Τα παραπάνω δεδομένα αντλήθηκαν από το ForeDeCk, μία βάση δεδομένων που έχει αναπτυχθεί από το Εθνικό Μετσόβιο Πολυτεχνείο που περιλαμβάνει 900,000 συνεχείς χρονοσειρές οι οποίες έχουν αντληθεί με τη σειρά τους από δημόσια διαθέσιμες πηγές. Για να αποφευχθεί η πρόβλεψη αρνητικών τιμών, ή τιμών μικρότερων του 10, οι οποίες θα επέφεραν προβλήματα κατά τον υπολογισμό των σφαλμάτων, έγινε κανονικοποίηση των δεδομένων μέσω της πρόσθεσης μίας σταθεράς στις χρονοσειρές ώστε η ελάχιστη τιμή τους να είναι το 10. Εν συνεχεία τα δεδομένα χωρίστηκαν σε δύο μέρη: το σετ εκπαίδευσης (training set) και το σετ αξιολόγησης (test set), το οποίο είχε αποκρυφθεί καθ' όλη τη διάρκεια του διαγωνισμού και βάσει του οποίου έγινε η αξιολόγηση των υποβολών. Τέλος οποιαδήποτε πληροφορία ικανή να καταστήσει αναγνωρίσιμη τη χρονοσειρά, όπως επί παραδείγματι η ημερομηνία αρχής της, αποκρύφθηκε και έγινε και αυτή διαθέσιμη με το πέρας του διαγωνισμού.

Για την αξιολόγηση των υποβολών χρησιμοποιήθηκαν ποικίλες στατιστικές μέθοδοι όπως η SES, η Theta και η Holt, αλλά και άλλες βασισμένες σε νευρωνικά δίκτυα όπως η αρχιτεκτονική MLP και του RNN. Τέλος για την αξιολόγηση των διαστημάτων πρόβλεψης (prediction intervals) αξιοποιήθηκαν μέθοδοι όπως το ETS και τα ARIMA.

Ο χρονικός ορίζοντας πρόβλεψης που ζητήθηκε για κάθε κατηγορία συχνότητας δεδομένων βασίστηκε στη φύση των αποφάσεων που καλούνται να λάβουν υπό πραγματικές συνθήκες οι εταιρείες ή οι οργανισμοί. Χαρακτηριστικά παραδείγματα αποτελούν οι ετήσιες προβλέψεις, οι οποίες συνήθως αφορούν μακροπρόθεσμες αποφάσεις που σχετίζονται με το στρατηγικό σχεδιασμό για τα επόμενα 1-5 χρόνια, ενώ στον αντίποδα τα τριμηνιαία ή μηνιαία δεδομένα, αφορούν σχεδιασμό προϋπολογισμού και ο επιθυμητός ορίζοντας δεν υπερβαίνει συνήθως τα



δύο χρόνια (8 τρίμηνα- 24 μήνες αντίστοιχα). Αναλυτικότερα, ο χρονικός ορίζοντας πρόβλεψης ανά κατηγορία διαμορφώθηκε ως εξής:

- 6 για τα ετήσια δεδομένα
- 8 για τα τριμηνιαία
- 18 για τα μηνιαία
- 13 για τα εβδομαδιαία
- 14 για τα ημερήσια
- 48 για τα ωριαία

Η αξιολόγηση των σημειακών προβλέψεων έγινε βάσει των δεικτών σφάλματος sMAPE και MASE (βλ. αντίστοιχη ενότητα, υπολογίζοντας το συνολικό σταθμισμένο μέσο όρο τους ώστε να επιτευχθεί η μέγιστη δυνατή αντικειμενικότητα. Η λήψη του σταθμισμένου μέσου όρου, γεγονός που δεν έγινε στον προηγούμενο διαγωνισμό (M3), βασίστηκε στο γεγονός ότι οι υποκατηγορίες του σετ των δεδομένων με μεγαλύτερο πλήθος χρονοσειρών και ορίζοντα πρόβλεψης, επηρεάζουν αναλογικά περισσότερο τον υπολογισμό της ακρίβειας. Ως εκ τούτου, στον M4 λαμβάνεται ο μέσος όρος στο επίπεδο της χρονοσειράς, με αποτέλεσμα η βαρύτητα κάθε χρονοσειράς να είναι ίδια σε όλο το σετ δεδομένων.

Αναφορικά με το διάστημα εμπιστοσύνης, έγινε επιλογή της τιμής του 95%, ούσα η κοινά χρησιμοποιούμενη στον επιχειρηματικό και επιχειρησιακό κόσμο, ενώ για την αξιολόγηση των υποβολών χρησιμοποιήθηκε η Μέση Κανονικοποιημένη Βαθμολογία Διαστημάτων (Mean Scaled Interval Score - MSIS), η μαθηματική διατύπωση του οποίου παρουσιάζεται ακολούθως:

$$MSIS = \frac{1}{h} \cdot \frac{\sum_{t=n+1}^{n+h} (U_t - L_t) + \frac{2}{\alpha} (L_t - Y_t) 1_{Y_t < L_t} + \frac{2}{\alpha} (Y_t - U_t) 1_{Y_t > U_t}}{\frac{1}{n-m} \sum_{t=m+1}^n |Y_t - Y_{t-m}|}$$

όπου:

$L_t$  και  $U_t$  είναι το κατώτερο και ανώτερο όριο των διαστημάτων πρόβλεψης

$Y_t$  είναι οι μελλοντικές τιμές της χρονοσειράς

$\alpha$  το επίπεδο εμπιστοσύνης, η τιμή του οποίου εν προκειμένω είναι 0,05, δεδομένου του ότι το διάστημα εμπιστοσύνης είναι της τάξης του 95%

Συνολικά ο διαγωνισμός απέφερε πληθώρα οφελών, καθώς ανέδειξε τις νέες προοπτικές που διανοίγονται στον κλάδο των προβλέψεων. Τα σημαντικότερα πορίσματα που προέκυψαν είναι το ακόλουθα:

**Πόρισμα 1:** Ο συνδυασμός μεθόδων είναι αποτελεσματικότερος

Οι μέθοδοι που έλαβαν τις κορυφαίες θέσεις του διαγωνισμού αποτελούσαν συνδυασμό μεθόδων (στατιστικών ή/και ML μεθόδων), τόσο για τις σημειακές προβλέψεις, όσο και για τα διαστήματα εμπιστοσύνης. Χαρακτηριστικά αξίζει να αναφερθεί ότι μόνο μία εκ των υποβολών που κατέλαβαν τις πρώτες δέκα θέσεις στις σημειακές προβλέψεις και μία εκ των πέντε κορυφαίων υποβολών στα διαστήματα εμπιστοσύνης αφορούσε αμιγώς στατιστική ή ML μέθοδο.

**Πόρισμα 2:** Πολυπλοκότερες μέθοδοι μπορούν να οδηγήσουν σε μεγαλύτερη προβλεπτική ακρίβεια

Σε αντίθεση με τους προηγούμενους M διαγωνισμούς, στον M4 διαφάνηκε ότι η υπολογιστική ισχύς μπορεί να οδηγήσει σε πιο ακριβείς προβλέψεις, χωρίς ωστόσο το προαναφερθέν να λαμβάνεται ως τεκμήριο, καθότι απλές αμιγώς στατιστικές μέθοδοι είχαν καλύτερα αποτελέσματα από σύνθετες μεθόδους βασισμένες σε νευρωνικά δίκτυα.

**Πόρισμα 3:** Χρήση πληροφορίας από πολλαπλές χρονοσειρές για την πρόβλεψη άλλης

Η συνάθροιση χρονοσειρών στο ίδιο επίπεδο συχνότητας και η αξιοποίησή τους κατά την εκπαίδευση των μεθόδων ή των μοντέλων, οδήγησε σε ορθότερη πρόβλεψη ή επιλογή των βαρών που πρέπει να χρησιμοποιηθούν κατά των συνδυασμό των μεθόδων/μοντέλων που λαμβάνονται υπόψη για την εξαγωγή της τελικής πρόβλεψης, γεγονός που υποδεικνύει τα οφέλη της cross-learning προσέγγισης.

## ΠΑΡΑΡΤΗΜΑ Β

### Παρουσίαση Κώδικα

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση και την αξιολόγηση της απόδοσης των νευρωνικών δικτύων που υλοποιήθηκαν στο πλαίσιο της παρούσας διπλωματικής εργασίας παρουσιάστηκαν στην Ενότητα 6.1.

Ακολούθως θα παρουσιαστεί ο κώδικας, σε γλώσσα Python, του δικτύου, καθώς και όσες επεξηγήσεις κρίνονται απαραίτητες για την κατανόηση της λειτουργίας του.

#### ***ΠΒ.1 Υπολογισμός Δεικτών ACF***

```
def acf(data, k):
    """
    Autocorrelation function
    :param data: time series
    :param k: lag
    :return:
    """
    m = np.mean(data)
    s1 = 0
    for i in range(k, len(data)):
        s1 = s1 + ((data[i] - m) * (data[i - k] - m))

    s2 = 0
    for i in range(0, len(data)):
        s2 = s2 + ((data[i] - m) ** 2)

    return float(s1 / s2)
```

Για τον υπολογισμό των αυτοσυσχετίσεων των δεδομένων που περιγράφηκε στην ενότητα της προετοιμασίας των δεδομένων του Κεφαλαίου 5, χρησιμοποιήθηκε η παραπάνω συνάρτηση, η οποία καλείται από τη συνάρτηση του test εποχιακότητας, λαμβάνει ως παραμέτρους τα δεδομένα της χρονοσειράς ως διάνυσμα και τον αριθμό k που υποδηλώνει τη χρονική καθυστέρηση και επιστρέφει μία μεταβλητή τύπου float που αποτελεί το ACFk

#### ***Π.Β.2 Test εποχιακότητας***

```
def seasonality_test(original_ts, ppy, tcrit):
    """
    Seasonality test
    :param original_ts: time series
    :param ppy: periods per year
    :param tcrit: seasonality critical cutoff
    :return: boolean value: whether the TS is seasonal
    """
    s = acf(original_ts, 1)
```

```

for i in range(2, ppy):
    s = s + (acf(original_ts, i) ** 2)

limit = tcrit * (sqrt((1 + 2 * s) / len(original_ts)))

return abs(acf(original_ts, ppy)) > limit

```

Για να αποφανθούμε σχετικά με το ποιες χρονοσειρές είναι εποχιακές, καλείται η ανωτέρω συνάρτηση, η οποία λαμβάνει ως παραμέτρους την αρχική χρονοσειρά (δεδομένα ως έχουν, χωρίς κάποια επεξεργασία), το μήκος της εποχιακότητας (εν προκειμένω 12) και την τιμή της παραμέτρου `tcritical` που όπως έχει αναφερθεί στη σχετική ενότητα τίθεται ίση με 1.96. Η συνάρτηση επιστρέφει μία boolean μεταβλητή τύπου TRUE/FALSE για το αν η υπό εξέταση χρονοσειρά είναι αντίστοιχα εποχιακή ή όχι

### Π.Β.3 Αποεποχικοποίηση Χρονοσειράς

Η συνάρτηση `deseasonalize(original_ts, ppy, tcrit)` λαμβάνει ως παραμέτρους τη χρονοσειρά, το μήκος εποχιακότητας και την παράμετρο `tcritical`. Με τη σειρά της καλεί τις συναρτήσεις που περιγράφηκαν στις ενότητες Π.Β.2 και Π.Β.3 και στην περίπτωση που αποφανθεί ο αλγόριθμος ότι είναι μία σειρά είναι εποχιακή, καλεί τη συνάρτηση , η οποία εξομαλύνει τη συνιστώσα της εποχιακότητας της χρονοσειράς με τη χρήση της μεθόδου των Κινητών Μέσων Όρων

Όσες χρονοσειρές έχουν χαρακτηριστεί ως εποχιακές, αποεποχικοποιούνται με την κλήση της συνάρτησης `deseasonalize(original_ts, ppy, tcrit)` η οποία εξομαλύνει τη συνιστώσα της εποχιακότητας με τη μέθοδο των κινητών μέσων όρων `moving_averages`. Τέλος επιστρέφει τους δείκτες εποχιακότητας (θέτει 0 τους δείκτες σε περίπτωση όπου η χρονοσειρά δεν είναι εποχιακή).

```

def deseasonalize(original_ts, ppy, tcrit):
    """
    Calculates and returns seasonal indices
    :param original_ts: original data
    :param ppy: periods per year
    :param tcrit: seasonality critical cutoff
    :return:
    """
    # === get in-sample data
    original_ts = original_ts[:-out_of_sample]
    """
    if seasonality_test(original_ts, ppy, tcrit):
        # ==== get moving averages
        ma_ts = moving_averages(original_ts, ppy)

        # ==== get seasonality indices
        le_ts = original_ts * 100.0 / ma_ts
        le_ts = np.hstack((le_ts, np.full((ppy - (len(le_ts) % ppy)), np.nan)))
        le_ts = np.reshape(le_ts, (-1, ppy))
        si = np.nanmean(le_ts, 0)
        norm = np.sum(si) / (ppy * 100.0)
        si = si / norm

```

```

else:
    si = np.full(ppy, 100.0)

return si

def moving_averages(ts_init, window):
    """
    Calculates the moving averages for a given TS
    :param ts_init: the original time series
    :param window: window length
    :return: moving averages ts
    """
    """
    As noted by Professor Isidro Lloret Galiana:
    line 82:
    if len(ts_init) % 2 == 0:

    should be changed to
    if window % 2 == 0:

    This change has a minor (less than 0.05%) impact on the calculations of the seasonal indices
    In order for the results to be fully replicable this change is not incorporated into the code below
    """

    # if len(ts_init) % 2 == 0:
    #     ts_ma = pd.rolling_mean(ts_init, window, center=True)
    #     ts_ma = pd.rolling_mean(ts_ma, 2, center=True)
    #     ts_ma = np.roll(ts_ma, -1)
    # else:
    #     ts_ma = pd.rolling_mean(ts_init, window, center=True)

    if window % 2 == 0:
        ts_ma = pd.Series(ts_init).rolling(window, center=True).mean()
        ts_ma = ts_ma.rolling(2, center=True).mean()
        ts_ma = np.roll(ts_ma.values, -1)
    else:
        ts_ma = pd.Series(ts_init).rolling(window, center=True).values

    return ts_ma

```

## Π.Β.4 Δημιουργία Δικτύου Εκπαίδευσης

Στην κλάση του δικτύου εκπαίδευσης TrainingFeedForwardNetwork κρίνονται άξια αναφοράς τα ακόλουθα σημεία:

### Π.Β.4.1 Καθορισμός Κατανομής

```
distr_type = torch.distributions.Cauchy
```

Ο καθορισμός της Κατανομής γίνεται με μία απλή εντολή, η οποία «τραβάει» από το Torch την επιθυμητή κατανομή.

### Π.Β.4.2 Καθορισμός Αριθμού και Τύπου Παραμέτρων

Όπως έχει αναφερθεί στο Κεφάλαιο 5, όλες οι κατανομές πλην της Laplace απαιτούν δύο παραμέτρους για τον προσδιορισμό τους, και συγκεκριμένα το loc (μέση τιμή  $\mu$ ) και στο scale (τυπική απόκλιση  $\sigma$ ). Συνεπώς η μορφή τους καθορίζεται με την ακόλουθη συνάρτηση αρχικοποίησης του δικτύου, κατά την οποία έχουν τη μορφή γραμμικού επιπέδου:

```
def __init__(
    self,
    prediction_length: int,
    context_length: int,
    hidden_dimensions: List[int],
    batch_norm: bool = False,
    scaling: Callable = mean_abs_scaling,
) -> None:
    super().__init__()

    assert prediction_length > 0
    assert context_length > 0
    assert len(hidden_dimensions) > 0

    self.prediction_length = prediction_length
    self.context_length = context_length
    self.hidden_dimensions = hidden_dimensions
    self.batch_norm = batch_norm
    self.scaling = scaling

    dimensions = [context_length] + hidden_dimensions[:-1]

    modules = []
    for in_size, out_size in zip(dimensions[:-1], dimensions[1:]):
        modules += [self._linear_layer(in_size, out_size), nn.ReLU()]
        if batch_norm:
            modules.append(nn.BatchNorm1d(units))
    modules.append(self._linear_layer(dimensions[-1], prediction_length * hidden_dimensions[-1]))
    self.nn = nn.Sequential(*modules)
    # CHANGE IN DISTRIBUTUION -> CHANGE IN NUMBER AND TYPE OF PARAMETERS
    #self.df_proj = nn.Sequential(self._linear_layer(hidden_dimensions[-1], 1), nn.Softplus())
    self.loc_proj = self._linear_layer(hidden_dimensions[-1], 1)
    self.scale_proj = nn.Sequential(self._linear_layer(hidden_dimensions[-1], 1), nn.Softplus())
```

ενώ η δημιουργία τους από τη συνάρτηση

```
def distr_and_scale(self, context):
    scale = self.scaling(context)
    scaled_context = context / scale
    nn_out = self.nn(scaled_context)
    nn_out_reshaped = nn_out.reshape(-1, self.prediction_length, self.hidden_dimensions[-1])

    distr_args = (
        #1.5 + self.df_proj(nn_out_reshaped).squeeze(dim=-1),
        #2.0 + self.df_proj(nn_out_reshaped).squeeze(dim=-1),
        self.loc_proj(nn_out_reshaped).squeeze(dim=-1),
```

```

        self.scale_proj(nn_out_reshaped).squeeze(dim=-1),
    )
    distr = net.distr_type(*distr_args)

    return distr, scale

```

### Π.Β.4.3 Καθορισμός Συνάρτησης Απωλειών

Όπως αναφέρθηκε στο Κεφάλαιο 5 κατά την παρουσίαση της εργαλειοθήκης του GluonTs, είναι δυνατή η δημιουργία εξατομικευμένων συναρτήσεων απωλειών, όπερ και εγένετο μέσω της ακόλουθης συνάρτησης που ανήκει στην κλάση `TrainingFeedForwardNetwork`:

```

def forward(self, context, target):
    assert context.shape[-1] == self.context_length
    assert target.shape[-1] == self.prediction_length

    distr, scale = self.distr_and_scale(context)
    loss = (-distr.log_prob(target / scale) + torch.log(scale)).mean(dim=1)
    # loss = (-distr.log_prob(target / scale)).mean(dim=1)
    # return distr.log_prob(target / scale).min(1), (target / scale).min(1)
    return loss

```

### Π.Β.5 Ορισμός Δικτύου

Δεδομένων των προαναφερθέντων, δηλώνεται το δίκτυο και οι βασικές υπερπαραμέτροι αυτού που αφορούν το μήκος του χρονικού ορίζοντα πρόβλεψης (`prediction_length`), το `context_length`, δηλαδή το μήκος του παραθύρου των δεδομένων που βλέπει σε κάθε επανάληψη το μοντέλο και τον αριθμό των κρυφών επιπέδων (`hidden_layers`). Τέλος να σημειωθεί ότι το μοντέλο εφαρμόζει κανονικοποίηση στα δεδομένα που λαμβάνει (`mean_abs_scaling`).

```

net = TrainingFeedForwardNetwork(
    prediction_length=prediction_length,
    context_length=context_length,
    hidden_dimensions=hidden_dimensions,
    batch_norm=False,
    scaling=mean_abs_scaling,
)

```

### Π.Β.6 Εκπαίδευση και εξαγωγή προβλέψεων

Όπως φαίνεται και από το τμήμα του αλγορίθμου που ακολουθεί, η εκπαίδευση του κάθε μοντέλου γίνεται για 20 εποχές, με το διάνυσμα των τιμών των απωλειών σε κάθε εποχή να καθορίζεται από την τιμή στόχο και την προβλεφθείσα τιμή κάθε φορά. Σημειώνεται ότι για λόγους εξάλειψης των ακραίων τιμών που προκύπτουν λόγω του μαθηματικού ορισμού του `log_prob` των κατανομών (ο κώδικας είχε αρχικά αναπτυχθεί αποκλειστικά για κατανομή `StudentT`), εφαρμόζεται `mean` επί του διανύσματος των απωλειών αφού έχει προηγηθεί η εξαίρεση του top 20% των τιμών. Τέλος ως `optimizer` έχει επιλεγεί ο Adam.

```

for epoch_no in range(20):
    sum_epoch_loss = 0.0
    for batch_no, batch in enumerate(data_loader, start=1):
        optimizer.zero_grad()
        # CHANGE IN DISTRIBUTION -> CHANGE IN LOSS
        loss_vec = net(context=batch["past_target"], target=batch["future_target"])
        #print(loss_vec)
        #print(type(loss_vec))
        #loss_vec_new = loss_vec.detach().numpy()
        #print (loss_vec_new)
        quantile=torch.quantile(loss_vec,0.80)
        #print(quantile)
        loss_vec_new=loss_vec[(loss_vec<quantile)]
        #print(loss_vec_new)
        #print (stop)
        loss=loss_vec_new.mean()
        #loss = loss_vec.mean()
        loss.backward()

        optimizer.step()

    sum_epoch_loss += loss.detach().numpy().item()

print(f"{epoch_no}: {sum_epoch_loss / num_batches_per_epoch}")

```

Τέλος η πρόβλεψη γίνεται εφαρμόζοντας τη συνάρτηση predict με όρισμα το σύνολο των δεδομένων εκπαίδευσης (train\_ds) στο feedforward δίκτυο κλάσης PyTorchPredictor.

```

pred_net = SamplingFeedForwardNetwork(
    prediction_length=net.prediction_length,
    context_length=net.context_length,
    hidden_dimensions=net.hidden_dimensions,
    batch_norm=net.batch_norm,
)
copy_parameters(net, pred_net)

feedforward = PyTorchPredictor(
    prediction_length=prediction_length, freq=freq,
    input_names=["past_target"], prediction_net=pred_net, batch_size=32, input_transform=transformation,
    device=None
)

predictions = feedforward.predict(train_ds)

```