



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

Ανάπτυξη αρχιτεκτονικών συνελκτικών
νευρωνικών δικτύων για πρόβλεψη
χρονοσειρών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νικόλαος Χ. Δίκαιος

Επιβλέπων : Ασημακόπουλος Βασίλειος

Καθηγητής Ε.Μ.Π.

Υπεύθυνος : Αρτέμιος-Ανάργυρος Σεμένογλου

Υπ. Διδάκτωρ Ε.Μ.Π.

Αθήνα, Ιούλιος 2022



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

Ανάπτυξη αρχιτεκτονικών συνελκτικών
νευρωνικών δικτύων για πρόβλεψη
χρονοσειρών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νικόλαος Χ. Δίκαιος

Επιβλέπων : Ασημακόπουλος Βασίλειος

Καθηγητής Ε.Μ.Π.

Υπεύθυνος : Αρτέμιος-Ανάργυρος Σεμένογλου

Υπ. Διδάκτωρ Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12^η Ιουλίου 2022.

.....

.....

.....

Βασίλειος Ασημακόπουλος

Ιωάννης Ψαρράς

Δημήτριος Ασκούνης

Αθήνα, Ιούλιος 2022



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

.....
Δίκαιος Νικόλαος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Ηλεκτρονικών Υπολογιστών

Copyright © Νικόλαος Δίκαιος, 2022.

Με την επιφύλαξη παντός δικαιώματος. All rights reserved

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

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

Στη μνήμη του Πατέρα μου

Περίληψη

Ο τομέας των προβλέψεων, και ειδικότερα των επιχειρησιακών, παρουσιάζει μία αξιοσημείωτη ανάπτυξη τα τελευταία χρόνια, η οποία λαμβάνει χώρα παράλληλα με την επανάσταση της επιστήμης των δεδομένων.

Οι παραδοσιακές στατιστικές μέθοδοι προβλέψεων που κυριαρχούσαν στον χώρο κατά τα πρώτα χρόνια της νέας χιλιετίας, αρχίζουν να δέχονται έντονο ανταγωνισμό από τις μεθόδους μηχανικής μάθησης και ιδιαίτερα από τα νευρωνικά δίκτυα, προς τα οποία το ενδιαφέρον της επιστημονικής κοινότητας, πέρασε από τον ενθουσιασμό κατά την δεκαετία του 80, στον παροπλισμό, και κατόπιν στην φρενήρη υιοθέτησή τους κατά τη δεύτερη δεκαετία του 21^{ου} αιώνα.

Στο πρώτο τμήμα της παρούσας διπλωματικής γίνεται μία εκτενής παρουσίαση των παραδοσιακών μεθόδων και τεχνικών παραγωγής προβλέψεων. Κατόπιν, η εργασία εστιάζει στις μεθόδους μηχανικής μάθησης και ιδιαίτερα στα νευρωνικά δίκτυα και τον τρόπο χρήσης τους στον χώρο των προβλέψεων.

Ακολουθεί η περιγραφή της πειραματικής διαδικασίας που υλοποιήθηκε με τη γλώσσα προγραμματισμού `python`. Στα πλαίσιά της υλοποιήθηκαν και διερευνήθηκαν πολλές αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων (CNN) προκειμένου να εκπονηθούν σημειακές προβλέψεις για τις ετήσιες χρονοσειρές του διαγωνισμού M4.

Μέσα από μια διαδικασία βελτιστοποίησης υπερπαραμέτρων, επιλέχθηκε η βέλτιστη αρχιτεκτονική, και οι επιδόσεις της συγκρίθηκαν με ένα Multi Layer Perceptron που αναπτύχθηκε ως benchmark μέθοδος για τις ανάγκες της παρούσας, την μέθοδο Theta που αποτελεί καθιερωμένη στατιστική μέθοδο με εξαιρετικές επιδόσεις σε προηγούμενους διαγωνισμούς M, καθώς και με την μέθοδο ES-RNN (Exponential Smoothing Recurrent Neural Network), που ούσα υβριδική μέθοδος (συνδυασμός εκθετικής εξομάλυνσης με αναδρομικά νευρωνικά δίκτυα) αναδείχθηκε νικήτρια του διαγωνισμού M4.

Η εργασία ολοκληρώνεται με προτάσεις περαιτέρω βελτιστοποίησης της επιλεγείσας αρχιτεκτονικής συνελκτικού δικτύου, καθώς και δημιουργίας μίας δικτυακής εφαρμογής, προκειμένου τα απότοκα της παρούσας να τεθούν σε παραγωγική λειτουργία και στην διάθεση του κοινού. Τέλος, γίνεται συζήτηση σχετικά με το ενεργειακό αποτύπωμα των μεθόδων που χρησιμοποιήθηκαν.

Λέξεις κλειδιά:

Τεχνικές Προβλέψεων, Μηχανική Μάθηση, Συνελκτικά Νευρωνικά δίκτυα, Python, Keras.

Abstract

Forecasting, and in particular business forecasting, is a field that has shown remarkable growth in recent years, a growth taking place in parallel with the data science revolution.

The traditional statistical forecasting methods that prevailed in the field during the first years of the new millennium are beginning to face intense competition from machine learning methods and especially from neural networks, to which, the interest of the scientific community, transitioned from the enthusiasm during the 80s, in decommissioning, and then again in a frantic adoption during the second decade of 21st century.

The present dissertation's first part comprises of an extensive presentation of traditional forecasting methods and techniques. Then, its focus shifts to machine learning methods and especially to neural networks and their usage in the field of predictions.

What follows, is a description of the experimental process implemented with the python programming language. As part of this, many convolutional neural network (CNN) architectures were implemented and explored in order to get point predictions for the M4 competition's, yearly time series data.

Through a process of hyper-parameter optimization, the optimal architecture was selected, and its performance was compared with that of a Multi-Layer Perceptron (developed as a benchmark method for the needs of the dissertation), of the Theta method which is an established statistical method with excellent performance in previous M competitions, and of the ESRNN (Exponential Smoothing Recurrent Neural Network), which, being a purely hybrid method (combination of exponential smoothing with recurrent neural networks) emerged as the winner of the M4 competition.

Our work is concluded with proposals for further optimization of the selected network architecture, as well as the creation of a web application, in order for the findings of our research to be put into production and become available to the public. Finally, a discussion is made about the energy footprint of our methods.

Keywords:

Forecasting Techniques, Machine Learning, Convolutional Neural Networks, Python, Keras.

Ευχαριστίες

Στο σημείο αυτό θα ήθελα να δώσω τις ευχαριστίες μου στον Καθηγητή του τομέα Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων κ. Βασίλειο Ασημακόπουλο ο οποίος μου έδωσε την ευκαιρία να εκπονήσω υπό την επίβλεψή του την παρούσα εργασία, συνδυάζοντας επιστημονικά αντικείμενα που από χρόνια έχουν κεντρίσει το ενδιαφέρον μου. Θα ήθελα επίσης να ευχαριστήσω τον συντονιστή της Μονάδας Προβλέψεων και στρατηγικής Δρ. Ευάγγελο Σπηλιώτη για την υποστήριξή του, καθώς και τον υπεύθυνο της παρούσας εργασίας υπ. Διδάκτορα Αρτέμιο-Ανάργυρο Σεμένογλου, του οποίου η συστηματική κι επιμελής καθοδήγηση σε όλα τα στάδια της μελέτης υπήρξε καταλυτική, και μετέτρεψε την εκπόνηση της εργασίας σε μία γόνιμη εμπειρία αναζήτησης και ανακάλυψης νέων γνώσεων και δεξιοτήτων.

Θα ήθελα να ευχαριστήσω επίσης τη μητέρα μου Άννα, τα αδέρφια μου Γιάννη και Αγγελική και φυσικά τη Σοφία, χωρίς την αμέριστη υποστήριξη των οποίων δεν θα ήταν δυνατό να επιτευχθεί η ολοκλήρωση των σπουδών μου όντας παράλληλα εργαζόμενος πλήρους απασχόλησης.

Δε θα μπορούσα να παραλείψω να αναφερθώ και στον, εδώ και πολλά χρόνια κολλητό μου φίλο Γιώργο, που σε μεγάλο βαθμό ενέπνευσε το ενδιαφέρον μου για τα αντικείμενα του προγραμματισμού υπολογιστών και της τεχνητής νοημοσύνης τα οποία με την πάροδο των χρόνων θα μπορούσα να πω ότι έφτασαν σε σημείο να με καθορίζουν ως άνθρωπο, καθώς και για τη γνώση και διαίσθηση που μου μετέδωσε μέσα από τις πολύωρες συζητήσεις μας.

Τέλος ένα μεγάλο ευχαριστώ θα ήθελα να απευθύνω και στους συνεργάτες-συνοδοιπόρους σε όλη αυτή τη μαραθώνια διαδικασία ολοκλήρωσης των σπουδών στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ. Άκη, Γιάννη, Γιώργο, Ευδοκία, Παύλο, Στέλιο, εύχομαι η τύχη να τα φέρει έτσι, ώστε να συνεργαστούμε και εκτός φοιτητικών πλαισίων.

Αθήνα, Ιούλιος 2022

Νικόλαος Χ. Δίκαιος

Συντομογραφίες

AI	Artificial Intelligence
ANN	Artificial Neural Network
ARIMA	Auto Regressive Integrated Moving Average
CNN	Convolutional Neural Network
DL	Deep Learning
GPU	Graphic Processing Unit
ML	Machine Learning
MLP	Multilayer Perceptron
SES	Simple Exponential Smoothing
TCN	Temporal Convolutional Network
TPU	Tensor Processing Unit

Περιεχόμενα

Περιεχόμενα	xvii
Κατάλογος εικόνων	xxiii
Κατάλογος πινάκων.....	xxv
Κατάλογος διαγραμμάτων.....	xxvii
1 Εισαγωγή.....	2
1.1 Γενικά	2
1.2 Οργάνωση της εργασίας	3
2 Γενικά περί προβλέψεων	3
2.1 Χρονοσειρές	6
2.2 Ποιοτικά χαρακτηριστικά χρονοσειρών	7
2.2.1 Τάση (Trend).....	7
2.2.2 Εποχικότητα (Seasonality)	8
2.2.3 Στασιμότητα (Stationary)	9
2.2.4 Κυκλικότητα (Cyclical).....	9
2.2.5 Ασυνέχειες.....	10
2.2.6 Τυχειότητα	11
2.3 Βασικά Βήματα σε μια Διαδικασία Πρόβλεψης.....	11
2.3.1 Καθορισμός Προβλήματος (Problem Definition)	11
2.3.2 Συγκέντρωση πληροφοριών (Gathering Information)	11
2.3.3 Προετοιμασία χρονοσειρών (data and adjustments).....	11
2.3.4 Επιλογή & Προσαρμογή Μοντέλου (Choosing & Fitting models)	12
2.3.5 Χρήση και αποτίμηση του μοντέλου πρόβλεψης (forecasting and monitoring)	12
2.4 Βασική Στατιστική ανάλυση	12
2.4.1 Μέση τιμή (Average)	13
2.4.2 Μέγιστη και ελάχιστη τιμή (Maximum και Minimum).....	13
2.4.3 Τυπική απόκλιση (Standard Deviation)	13

2.4.4	Διακύμανση (Variance).....	13
2.4.5	Συνδιακύμανση (Covariance).....	14
2.4.6	Συντελεστής γραμμικής συσχέτισης (Linear Correlation Coefficient).....	14
2.4.7	Συντελεστής αυτοσυσχέτισης (Autocorrelation Coefficient).....	15
2.4.8	Συντελεστής μεταβλητότητας (Coefficient of Variation).....	16
2.4.9	Μέση τιμή διαστήματος μεταξύ ζητήσεων (intermittent Demand Interval).....	16
2.5	Στατιστικές Μέθοδοι προβλέψεων.....	16
2.5.1	Naïve.....	16
2.5.2	Απλή γραμμική παλινδρόμηση.....	17
2.5.3	Πολλαπλή γραμμική παλινδρόμηση.....	18
2.5.4	Μέθοδοι Εκθετικής Εξομάλυνσης.....	19
2.5.5	Μοντέλο σταθερού επιπέδου (Simple Exponential Smoothing (SES)).....	19
2.5.6	Μοντέλο γραμμικής τάσης (Holt Exponential Smoothing).....	20
2.5.7	Damped.....	20
2.5.8	Winters.....	21
2.5.9	Μέθοδος Theta.....	23
2.5.10	Ολοκληρωμένα Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου Όρου (ARIMA). 25	
2.6	Ανάλυση ακρίβειας προβλέψεων.....	27
2.6.1	Μέσο σφάλμα (Mean Error).....	27
2.6.2	Μέσο απόλυτο σφάλμα (Mean Absolute Error).....	28
2.6.3	Μέσο τετραγωνικό σφάλμα (Mean Squared Error).....	28
2.6.4	Ρίζα μέσου τετραγωνικού Σφάλματος (Root Mean Squared Error).....	28
2.6.5	Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error).....	28
2.6.6	Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error).....	29

2.6.7	Σχετικοί δείκτες.....	30
2.6.8	Μέσο απόλυτο κανονικοποιημένο σφάλμα (Mean Absolute Scaled Error).	30
2.7	Διαγωνισμοί προβλέψεων.....	31
2.7.1	Διαγωνισμός M1.....	31
2.7.2	Διαγωνισμός M2.....	32
2.7.3	Διαγωνισμός M3.....	33
2.7.4	Διαγωνισμός M4.....	34
2.7.5	Διαγωνισμός M5.....	37
2.7.6	Διαγωνισμός M6.....	38
3	Τεχνητή Νοημοσύνη – Μηχανική Μάθηση.....	41
3.1	Τεχνητή νοημοσύνη.....	41
3.2	Μηχανική Μάθηση.....	42
3.3	Επιβλεπόμενη μάθηση.....	43
3.3.1	Μηχανές διανυσμάτων υποστήριξης (Support Vector Machines).....	44
3.3.2	Δένδρα απόφασης.....	45
3.3.3	Τεχνικές παλινδρόμησης.....	46
3.3.4	Νευρωνικά δίκτυα.....	48
3.4	Μη επιβλεπόμενη μάθηση.....	48
3.4.1	Ανάλυση σε κύριες συνιστώσες (Principal Component Analysis).	49
3.4.2	Συσταδοποίηση (Clustering).....	50
4	Νευρωνικά Δίκτυα - Βαθιά Μηχανική Μάθηση.....	55
4.1	Νευρωνικά δίκτυα.....	57
4.1.1	Τεχνητοί νευρώνες.....	58
4.1.2	Perceptron.....	58
4.1.3	Μάθηση.....	62
4.1.4	Στοχαστική κατάβαση δυναμικού (stochastic gradient descent).....	62

4.1.5	Οπίσθια διάδοση (Backpropagation).....	65
4.1.6	Ζητήματα εκπαίδευσης.....	66
4.1.7	Βελτιστοποίηση υπερπαραμέτρων	68
4.2	Multi Layer Perceptron.....	70
4.3	Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)	73
4.3.1	Η συνέλιξη.....	76
4.3.2	Η έννοια του γεμίσματος (padding)	77
4.3.3	Η έννοια του βήματος συνέλιξης (convolution stride).....	78
4.3.4	Βαθμός διαστολής (dilation rate)	79
4.3.5	Η διαδικασία επιλογής μεγίστων (max pooling).....	80
4.3.6	Επιπέδωση (flattening)	81
4.3.7	Πανόραμα της αρχιτεκτονικής ενός συνελκτικού νευρωνικού δικτύου.	81
4.4	Χρονικά (temporal) συνελκτικά δίκτυα (TCN).....	83
4.4.1	Αιτιατή συνέλιξη (Causal Convolution)	83
4.4.2	Εκμετάλλευση του dilation rate	84
5	Πειραματική διαδικασία.....	87
5.1	Εισαγωγή	87
5.2	Περιγραφή και ανάλυση του συνόλου δεδομένων	88
5.3	Προεπεξεργασία δεδομένων	90
5.4	Δημιουργία συνελκτικών δικτύων (CNN).....	92
5.4.1	Τεχνική τερματισμού εκπαίδευσης και βελτιστοποιητής.....	92
5.4.2	Εκπαίδευση πολλαπλών μοντέλων (ensembling) και διάμεσες (median) προβλέψεις.	92
5.4.3	Βελτιστοποίηση υπερπαραμέτρων	93
5.4.4	Τεχνολογίες υλοποίησης.	96
5.5	Δείκτες αξιολόγησης.....	97

5.6	Επιλογή μεθόδων αναφοράς (benchmarks).....	97
6	Μελέτη και ανάλυση των Αποτελεσμάτων.....	99
6.1	Επιλογή βέλτιστου μοντέλου και σύγκριση επιδόσεων με μεθόδους αναφοράς. ...	100
6.2	Στατιστική ανάλυση και διαστήματα εμπιστοσύνης.....	100
6.3	Βέλτιστο πλήθος εκπαιδευόμενων μοντέλων (ensembling).....	101
6.4	Επίδραση των βασικών παραμέτρων στην ακρίβεια του μοντέλου.	102
6.4.1	Παρατηρήσεις εισόδου.....	103
6.4.2	μέγεθος δεσμίδας (batch size).....	104
6.4.3	Ρυθμός μάθησης (learning rate).....	105
6.4.4	Ποσοστό συνόλου ελέγχου (test size).....	106
6.4.5	Αριθμός φίλτρων.....	107
6.4.6	Αριθμός νευρώνων «πυκνού» επιπέδου (dense neurons).....	108
6.4.7	Αριθμός συνελκτικών επιπέδων (convolutional layers).....	109
6.4.8	Αριθμός πυκνών επιπέδων (dense layers).....	110
6.4.9	Βαθμός διαστολής (dilation rate).....	111
6.4.10	Απεικόνιση ενδιάμεσων εξόδων των συνελκτικών επιπέδων.....	112
6.5	Συσχέτιση του σφάλματος των προβλέψεων με τα χαρακτηριστικά της χρονοσειράς 116	
6.6	Συσταδοποίηση των χρονοσειρών (clustering) και σύγκριση επιδόσεων των μεθόδων αναφοράς ανά συστάδα.....	117
7	Συμπεράσματα – Μελλοντική έρευνα.....	121
8	Βιβλιογραφικές αναφορές.....	127

Κατάλογος εικόνων

Εικόνα 1. Εξέλιξη του παγκόσμιου πληθυσμού από το 1600 μΧ.	8
Εικόνα 2. Εποχικές διακυμάνσεις του ποσοστού θετικών τεστ για κορωνοϊούς στην Σουηδία.8	
Εικόνα 3. Κυλιόμενη 10ετής απόδοση του δείκτη S&P 500.	10
Εικόνα 4. Τα αποτελέσματα των μεθόδων του διαγωνισμού M3 σύμφωνα με τη μετρική sMAPE	34
Εικόνα 5. Το παράδειγμα της μηχανικής μάθησης σε σχέση με τον προστακτικό προγραμματισμό [32].	43
Εικόνα 6. Κατηγοριοποίηση δεδομένων από Μηχανές Διανυσμάτων Υποστήριξης [36])	44
Εικόνα 7. Εφαρμογή δένδρων απόφασης σε πρόβλημα χαρτογραφικής επιλογής (selection) [40]	46
Εικόνα 8. Χρήση λογιστικής και γραμμικής παλινδρόμησης σε πρόβλημα δυαδικής κατηγοριοποίησης [41].	47
Εικόνα 9. Ανάλυση κύριων συνιστωσών σε δεδομένα πωλήσεων φαγητού.	49
Εικόνα 10. Τα βήματα του αλγορίθμου K-means σε δισδιάστατα δεδομένα [42].	52
Εικόνα 11. Δημιουργία δενδρογράμματος ιεραρχικής συσταδοποίησης [42]	54
Εικόνα 12. Αναπαράσταση του βιολογικού νευρώνα.	57
Εικόνα 13. Η λειτουργία του τεχνητού νευρώνα [46].	58
Εικόνα 14. Ο κανόνας μάθησης perceptron [48]	59
Εικόνα 15. Συνηθισμένες συναρτήσεις ενεργοποίησης στα Νευρωνικά δίκτυα [49].	60
Εικόνα 16. Συνάρτηση μίας μεταβλητής και η κλίση της [32]	64
Εικόνα 17. Κατάβαση δυναμικού σε μία επιφάνεια δύο διαστάσεων (2 παράμετροι) [32] ...	65
Εικόνα 18. Η λειτουργία του αλγορίθμου backpropagation με τη χρήση του κανόνα της αλυσίδας [53].	66
Εικόνα 19. Διαδικασία εκπαίδευσης με τη μέθοδο early stopping [54].	67
Εικόνα 20. Σύγκριση των μεθόδων της αναζήτησης πλέγματος και της τυχαίας αναζήτησης [55]	69
Εικόνα 21. Σύγκριση αναζήτησης πλέγματος και Bayesian optimization [56].	70
Εικόνα 22. Η δομή ενός Multi Layer Perceptron [57]	71
Εικόνα 23. Χρήση MLP για την πρόβλεψη χρονοσειρών [60].	73
Εικόνα 24. Αποδόμηση εικόνας σε τοπικά μοτίβα όπως ακμές, υφές κλπ. [32]	74

Εικόνα 25. Κατηγοριοποίηση εικόνας με CNN με την ιεραρχική εκμάθηση ολοένα και πιο σύνθετων χαρακτηριστικών από τα διαδοχικά επίπεδα του δικτύου [42]	76
Εικόνα 26. Διατήρηση των διαστάσεων 5 x 5 του χάρτη χαρακτηριστικών με τη χρήση padding [32].	78
Εικόνα 27. Συνέλιξη με διαστάσεις φίλτρου 3 x 3 και βήμα (stride) με διαστάσεις 2 x 2 [32]. Η διάσταση της εξόδου θα είναι 2 x 2.	79
Εικόνα 28. Συνέλιξη δύο διαστάσεων με τη χρήση φίλτρου διάστασης 3 x 3 και dilation rate 2 χωρίς padding [62].	80
Εικόνα 29. Κατηγοριοποίηση εικόνας με τη χρήση συνελκτικών νευρωνικών δικτύων [42]	81
Εικόνα 30. Μονοδιάστατο συνελκτικό νευρωνικό δίκτυο για πρόβλεψη πολλαπλών βημάτων χρονοσειράς [63]	82
Εικόνα 31. Διασφάλιση της αιτιατότητας με padding μόνο στην αριστερή πλευρά της χρονοσειράς [64]	84
Εικόνα 32. Αύξηση μεγέθους του πεδίου λήψης με χρήση δύο συνελκτικών επιπέδων [64].	85
Εικόνα 33. Η αρχιτεκτονική του wavenet που δημιουργεί ένα χρονικό συνελκτικό δίκτυο [65].	86

Κατάλογος πινάκων

Πίνακας 1. Χώρος παραμέτρων αρχιτεκτονικής, πρώτης σειράς πειραμάτων.	93
Πίνακας 2. Χώρος παραμέτρων εκπαίδευσης, πρώτης σειράς πειραμάτων.	94
Πίνακας 3. Σύγκριση του CNN με τις μεθόδους αναφοράς.....	100
Πίνακας 4. sMAPE με βάση τον αριθμό των εκπαιδευμένων μοντέλων (χρήση διάμεσων προβλέψεων)	102
Πίνακας 5. Αποτελέσματα πειραμάτων με μεταβολή των παρατηρήσεων εισόδου.....	103
Πίνακας 6. Αποτελέσματα πειραμάτων με μεταβολή του μεγέθους δεσμίδας.....	104
Πίνακας 7. Αποτελέσματα πειραμάτων με μεταβολή του ρυθμού μάθησης.....	105
Πίνακας 8. Αποτελέσματα πειραμάτων με μεταβολή του ποσοστού του συνόλου ελέγχου .	106
Πίνακας 9. Αποτελέσματα πειραμάτων με μεταβολή του αριθμού φίλτρων.....	107
Πίνακας 10. Αποτελέσματα πειραμάτων με μεταβολή των νευρώνων πυκνού επιπέδου	108
Πίνακας 11. Αποτελέσματα πειραμάτων με μεταβολή του αριθμού συνελκτικών επιπέδων.	109
Πίνακας 12. Αποτελέσματα πειραμάτων με μεταβολή του αριθμού πυκνών επιπέδων.	110
Πίνακας 13. Αποτελέσματα πειραμάτων με μεταβολή του βαθμού διαστολής.....	111
Πίνακας 14. Αποτελέσματα γραμμικής παλινδρόμησης για τη συσχέτιση της επίδοσης των συνελκτικών δικτύων με τα 8 βασικά χαρακτηριστικά.	116
Πίνακας 15. sMAPE των μεθόδων ανά τεταρτημόριο.....	118
Πίνακας 16. Ποσοστιαία βελτίωση του sMAPE από το CNN ως προς τις μεθόδους αναφοράς.	119

Κατάλογος διαγραμμάτων

Διάγραμμα 1. Αναγωγή των 8 βασικών χαρακτηριστικών σε χώρο 2 διαστάσεων, με ανάλυση PCA	89
Διάγραμμα 2. Χάρτης θερμότητας των κυρίων συνιστωσών σε σχέση με τα 8 βασικά χαρακτηριστικά.	90
Διάγραμμα 3. Σχηματική αναπαράσταση της πειραματικής διαδικασίας.....	95
Διάγραμμα 4. Ραβδόγραμμα του sMAPE ως προς τις παρατηρήσεις εισόδου (διάμεσος 30 μοντέλων).....	103
Διάγραμμα 5. Ραβδόγραμμα του sMAPE ως προς το μέγεθος δεσμίδας (διάμεσος 30 μοντέλων).....	104
Διάγραμμα 6. Ραβδόγραμμα του sMAPE ως προς τον ρυθμό μάθησης (διάμεσος 30 μοντέλων).....	105
Διάγραμμα 7. Ραβδόγραμμα του sMAPE ως προς το ποσοστό του συνόλου ελέγχου (διάμεσος 30 μοντέλων).....	106
Διάγραμμα 8. Ραβδόγραμμα του sMAPE ως προς τον αριθμό φίλτρων (διάμεσος 30 μοντέλων).....	107
Διάγραμμα 9. Ραβδόγραμμα του sMAPE ως προς τον αριθμό νευρώνων πυκνού επιπέδου (διάμεσος 30 μοντέλων).....	108
Διάγραμμα 10. Ραβδόγραμμα του sMAPE ως προς τον αριθμό συνελκτικών επιπέδων (διάμεσος 30 μοντέλων).....	109
Διάγραμμα 11. Ραβδόγραμμα του sMAPE ως προς τον αριθμό πυκνών επιπέδων (διάμεσος 30 μοντέλων).....	110
Διάγραμμα 12. Ραβδόγραμμα του sMAPE ως προς τον βαθμό διαστολής (διάμεσος 30 μοντέλων).....	111
Διάγραμμα 13. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά με τάση, χωρίς εποχικότητα ή τυχαιότητα.....	112
Διάγραμμα 14. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά , με αλλαγή μοτίβου (από στασιμότητα αποκτάει τάση).....	113
Διάγραμμα 15. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά με κυκλικότητα.....	114
Διάγραμμα 16. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά με πτωτική τάση και κυκλικότητα.....	114

Διάγραμμα 17. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά με τυχαιότητα χωρίς τάση.	115
Διάγραμμα 18. Διαχωρισμός των χρονοσειρών σε συστάδες (clusters).....	118

1 Εισαγωγή

1.1 Γενικά

Στόχος της παρούσας διπλωματικής εργασίας είναι η διερεύνηση της αποτελεσματικότητας της χρήσης Συνελκτικών Νευρωνικών Δικτύων (CNN) στην πρόβλεψη χρονοσειρών χρησιμοποιώντας δεδομένα που προέρχονται από τον διαγωνισμό προβλέψεων M4.

Παραδοσιακά ο χώρος των προβλέψεων επιχειρησιακών χρονοσειρών κυριαρχείται από τις στατιστικές μεθόδους. Ωστόσο τα τελευταία χρόνια, ένας νέος παίκτης έχει εισέλθει δυναμικά στο χώρο, και δεν είναι άλλος από τη Μηχανική Μάθηση (ML). Συγκεκριμένα, για να αναφερθούμε στο χώρο των διαγωνισμών προβλέψεων, σε παλαιότερους διαγωνισμούς (M3) η απόδοση των μεθόδων αυτών δεν κρίθηκε ικανοποιητική έτσι ώστε να ανταγωνιστούν τις παραδοσιακές στατιστικές μεθόδους. Ωστόσο στους τελευταίους διαγωνισμούς, η χρήση αλλά και η αποτελεσματικότητά τους έχει βελτιωθεί σημαντικά. Η μεγάλη έκπληξη που σημειώθηκε στο διαγωνισμό M4 ήταν η νίκη μίας «υβριδικής» προσέγγισης η οποία έκανε χρήση στατιστικών αλλά και ML χαρακτηριστικών.

Το μεγάλο «μειονέκτημα» της μηχανικής μάθησης (ειδικότερα της βαθιάς μάθησης), είναι εμφανώς οι μεγάλοι χρόνοι εκπαίδευσης που απαιτούνται και κατ' επέκταση το ενεργειακό της αποτύπωμα. Η παρούσα διπλωματική εστιάζει συν τοις άλλοις στο να αξιολογήσει το χρόνο και συνεπώς και την ενεργειακή συμπεριφορά των συνελκτικών νευρωνικών δικτύων.

Οι χρονοσειρές που χρησιμοποιήθηκαν αντλήθηκαν από τα δεδομένα του διαγωνισμού προβλέψεων M4. Για την επεξεργασία των δεδομένων χρησιμοποιήθηκε το οικοσύστημα τεχνολογιών της γλώσσας python και για την δημιουργία των συνελκτικών δικτύων, χρησιμοποιήθηκαν οι βιβλιοθήκες tensorflow και keras.

Σαν κριτήριο αξιολόγησης των μοντέλων χρησιμοποιήθηκε το sMAPE.

Ως μέθοδοι αναφοράς χρησιμοποιήθηκαν, ένα Multi-Layer Perceptron (MLP) που εκπαιδεύτηκε αποκλειστικά για τις ανάγκες της παρούσας εργασίας, η καθιερωμένη στατιστική μέθοδος Theta, καθώς και η νικητήρια μέθοδος του διαγωνισμού M4 (ES-RNN).

1.2 Οργάνωση της εργασίας

Στο *δεύτερο* κεφάλαιο, γίνεται παρουσίαση της διαδικασίας παραγωγής προβλέψεων και η απαραίτητη θεμελίωση των σχετικών εννοιών. Παρουσιάζονται οι πιο κλασικές στατιστικές μέθοδοι πρόβλεψης και τα μέτρα ακρίβειας προβλέψεων. Γίνεται επίσης αναφορά στην ιστορία, τα αποτελέσματα και τα συμπεράσματα των διαγωνισμών M (Makridakis Competitions).

Στο *τρίτο* κεφάλαιο παρουσιάζονται οι κυριότερες έννοιες της μηχανικής μάθησης, η οποία διακρίνεται σε επιβλεπόμενη και μη επιβλεπόμενη και περιγράφονται οι βασικές μέθοδοι της κάθε κατηγορίας.

Στο *τέταρτο* κεφάλαιο γίνεται αναλυτική παρουσίαση των νευρωνικών δικτύων και των δομικών τους συστατικών. Ιδιαίτερη έμφαση δίνεται στα συνελκτικά νευρωνικά δίκτυα (CNN) και στην πολλά υποσχόμενη αρχιτεκτονική τους, που ακούει στο όνομα Χρονικά Συνελκτικά δίκτυα (Temporal Convolutional Networks), η οποία απευθύνεται και έχει εμφανίσει εξαιρετικές επιδόσεις στην ανάλυση χρονολογικών δεδομένων.

Στο *πέμπτο* Κεφάλαιο παρουσιάζεται η πειραματική διαδικασία που ακολουθήθηκε: Δεδομένα, προεπεξεργασία, τεχνολογίες και αρχιτεκτονικές των μοντέλων, επιλογή μετρικών αξιολόγησης και μεθόδων αναφοράς, καθώς και βελτιστοποίηση των υπερπαραμέτρων που επιστρατεύτηκαν προκειμένου να επιλεγεί η καλύτερη αρχιτεκτονική.

Στο *έκτο* κεφάλαιο γίνεται ανάλυση και παρουσίαση των πειραματικών αποτελεσμάτων. Διερευνώνται τα όρια του στατιστικού σφάλματος των μοντέλων, και προσδιορίζεται το βέλτιστο πλήθος που πρέπει να εκπαιδευτεί και να συνδυαστεί (ensembling) προκειμένου να έχουμε τα βέλτιστα αποτελέσματα.

Διερευνάται επίσης η επίδραση της κάθε παραμέτρου εκπαίδευσης στην ακρίβεια των προβλέψεων, γίνεται διαχωρισμός των χρονοσειρών σε clusters με βάση τις δύο κύριες συνιστώσες τους και αναλύεται η επίδοση του βέλτιστου μοντέλου σε σχέση με τις μεθόδους αναφοράς που επιλέχθηκαν. Τέλος, εκτελείται γραμμική παλινδρόμηση, η οποία συσχετίζει την ακρίβεια των προβλέψεων που επιτεύχθηκε από το επιλεγέν μοντέλο, με τις τιμές που παίρνει η κάθε χρονοσειρά για καθένα από τα βασικά της χαρακτηριστικά.

Στο *έβδομο* κεφάλαιο, παρουσιάζονται τα συμπεράσματα της μελέτης, προτάσεις για περαιτέρω βελτίωση των επιδόσεων του μοντέλου και γίνεται πρόταση για δημιουργία μιας εφαρμογής διαδικτύου (web application) μέσω της οποίας θα τεθούν σε παραγωγική λειτουργία και στη διάθεση του κοινού τα αποτελέσματα της παρούσας εργασίας. Γίνεται επίσης συζήτηση σχετικά με το ενεργειακό αποτύπωμα των μεθόδων.

Τέλος, στο *όγδοο* κεφάλαιο, παρατίθενται οι βιβλιογραφικές αναφορές στις οποίες ανέτρεξε ο μελετητής για την εκπόνηση της παρούσας.

2 Γενικά περί προβλέψεων

Λέγοντας πρόβλεψη (forecasting) εννοούμε τη διαδικασία κατά την οποία προσπαθούμε να προσεγγίσουμε μελλοντικές τιμές ενός φαινομένου, βασιζόμενοι κατά κύριο λόγο σε δεδομένα του παρελθόντος και του παρόντος. Από τον ορισμό που δόθηκε, αντιλαμβανόμαστε πως οι προβλέψεις εκτός του ότι βρίσκουν εφαρμογή σε μία μεγάλη γκάμα ανθρωπίνων δραστηριοτήτων, ταυτόχρονα χρησιμοποιούνται και στην καθημερινότητά μας πολλές φορές ασυναίσθητα (π.χ. προγραμματισμός των δραστηριοτήτων μας με βάση μετεωρολογικές προβλέψεις).

Ξεκινάμε λοιπόν, αναφέροντας ενδεικτικά ορισμένους βασικούς τομείς της επιστήμης και της ανθρωπογενούς δραστηριότητας, στους οποίους η αξιόπιστη παραγωγή προβλέψεων διαδραματίζει κομβικό ρόλο:

- **Οικονομία & Χρηματοοικονομικά:** Τεχνικές προβλέψεων των χρηματιστηριακών δεικτών, της μεταβλητότητάς τους, καθώς και της βέλτιστης επιλογής χαρτοφυλακίου. Είναι γνωστό πως η σωστή πρόβλεψη των τιμών των μετοχών αποτελεί κάτι σαν «άγιο δισκοπότηρο» για όσους ασχολούνται με τον χρηματοοικονομικό τομέα.
- **Περιβάλλον & Κλίμα:** Πρόβλεψη ενεργειακής ζήτησης, διαχείριση υδάτινων πόρων, μετεωρολογία και ρύπανση κ.λπ.. Η ενέργεια είναι ως γνωστόν η κινητήριος δύναμη των οικονομιών, και ως εκ τούτου η ενεργειακή ζήτηση συναρτάται με πολλά άλλα φαινόμενα που αφορούν τους πολίτες, όπως είναι η αύξηση των τιμών των αγαθών (στασιμοπληθωρισμός).
- **Κοινωνικό περιβάλλον:** Πρόβλεψη της πληθυσμιακής εξέλιξης, των επιδημιών, της εγκληματικότητας κ.λπ. Για παράδειγμα η πρόβλεψη της ηλικιακής σύνθεσης του ελληνικού πληθυσμού, διαδραματίζει κομβικό ρόλο στη διαμόρφωση πολιτικών και στρατηγικών βιωσιμότητας του ασφαλιστικού συστήματος, ενώ η πρόβλεψη της εξέλιξης μιας πανδημίας καθορίζει σε μεγάλο βαθμό τα μέτρα που καλείται να λάβει μία κυβέρνηση.
- **Τουρισμός:** Εκτίμηση αφίξεων τουριστών, διανυκτερεύσεων και ταξιδιωτικού συναλλάγματος. Η σωστή πρόβλεψη της τουριστικής κίνησης είναι ζωτικής σημασίας για μια χώρα όπως η Ελλάδα (της οποίας ένα μεγάλο ποσοστό του ΑΕΠ εξαρτάται

από τον τομέα αυτόν), όχι μόνο σε κυβερνητικό επίπεδο, για την κατάρτιση του κρατικού προϋπολογισμού, αλλά και στη σφαίρα της ιδιωτικής οικονομίας.

- **Μεταφορές & Μετακινήσεις:** Πρόβλεψη κυκλοφοριακού φόρτου κατά τον σχεδιασμό έργων οδοποιίας, εκτίμηση αυτοκινητιστικών ατυχημάτων κ.λπ.
- **Ακίνητα & κτηματικές περιουσίες:** Αξιόπιστα μοντέλα πρόβλεψης των αντικειμενικών ή και πραγματικών αξιών ακινήτων θα μπορούσαν να διαδραματίσουν σημαντικό ρόλο στη χάραξη της κρατικής φορολογικής και κοινωνικής πολιτικής, και να βοηθήσουν και τη διαδικασία λήψης αποφάσεων των ιδιωτών που δραστηριοποιούνται στο χώρο.

Τα τελευταία χρόνια, το ενδιαφέρον και η σημασία των προβλέψεων έχουν αυξηθεί κατακόρυφα, τόσο από τον ακαδημαϊκό κόσμο όσο και από τους πρακτικά ενασχολούμενους με την επιχειρησιακή έρευνα. Οι σχεδιαστές της κυβερνητικής πολιτικής, οι διευθύνοντες των επιχειρήσεων και πολλοί άλλοι, προσπαθώντας να αντιμετωπίσουν την εγγενή αβεβαιότητα στους τομείς της οικονομίας της πολιτικής κ.λπ., αναζητούν ολοένα και πιο αξιόπιστη παραγωγή προβλέψεων.

Ο κίνδυνος όμως και η αβεβαιότητα, αποτελούν κομβικά χαρακτηριστικά στο ολοένα και πιο ευμετάβλητο ανθρωπογενές περιβάλλον όπου διαβιούμε και ως εκ τούτου στην διαδικασία της πρόβλεψης. Ο καθηγητής Μακρινδάκης και οι συνεργάτες του στο βιβλίο «Χορεύοντας με την τύχη», κατηγοριοποιούν την καθημερινή αβεβαιότητα σε όλα τα επίπεδα της ζωής μας, κάνοντας λόγο για την «αβεβαιότητα του μετρό» και την «αβεβαιότητα της καρύδας». Στην πρώτη περίπτωση αναφέρονται σε συνεχείς και μικρές τυχαίες διακυμάνσεις, τις οποίες υφίσταται για παράδειγμα η χρονική διάρκεια της μετάβασης του μετρό ανάμεσα σε δύο απομακρυσμένους σταθμούς, ενώ στην δεύτερη σε εντελώς απρόσμενα και σπάνια γεγονότα που μπορεί να έχουν σημαντικές επιδράσεις και συνέπειες, όπως είναι οι φυσικές καταστροφές, οι επιδημίες κ.λπ. [1].

Ο τομέας των προβλέψεων έχει δεχθεί δυσμενείς κριτικές και μεγάλη δυσαρέσκεια εξαιτίας της αποτυχίας των μεθόδων για έγκαιρη προειδοποίηση σχετικά με επερχόμενες αλλαγές, αλλά και εξαιτίας των μεγάλων σφαλμάτων που έχουν παρατηρηθεί στις προβλέψεις. Αποτελεί ειρωνεία βέβαια το γεγονός, πως όταν επικρατούν συνθήκες ομαλότητας, τότε τα σφάλματα των μεθόδων πρόβλεψης ελαττώνονται μεν αλλά ταυτόχρονα ελαττώνονται και οι

ανάγκες για παραγωγή προβλέψεων. Αντιθέτως, όταν επικρατούν απρόοπτες συνθήκες (π.χ. πόλεμοι, επιδημίες), οπότε και αυξάνεται κατακόρυφα η ανάγκη για αξιόπιστες προβλέψεις, τότε είναι που οι μέθοδοι των προβλέψεων αποτυγχάνουν ως επί το πλείστον.

Γενικά θεωρείται καλή πρακτική από τους ασχολούμενους με τις προβλέψεις, το να καταδεικνύουν τον βαθμό αβεβαιότητας που συνεπάγεται η μέθοδος και η φύση των δεδομένων που χρησιμοποίησαν. Πολύς λόγος έχει γίνει για την ακρίβεια των μοντέλων που έχουν αναπτυχθεί, τα οποία ιστορικά έχουν χαρακτηριστεί από μεγάλα σφάλματα καθώς και από την αδυναμία τους να υποδείξουν επερχόμενες αλλαγές που αιφνιδίασαν ολόκληρο τον επιχειρηματικό κόσμο. Ωστόσο, η βασικότερη αιτία που πυροδοτεί την κριτικές αυτές, είναι οι εσφαλμένες απαιτήσεις των χρηστών των προβλέψεων, καθώς συνήθως αυτό που απαιτείται δεν είναι μία πρόβλεψη αλλά μία «προφητεία». Αυτό όμως απέχει πολύ από την πραγματικότητα καθώς η πρόβλεψη δεν είναι υποκατάστατο της προφητείας, και φυσικά τα σφάλματα είναι αναμενόμενα και αναπόφευκτα.

Είναι λοιπόν κομβικής σημασίας να ενημερώνονται οι τελικοί χρήστες των προβλέψεων, όσον αφορά τα ρεαλιστικά πλεονεκτήματα αλλά και τα όρια των αντίστοιχων μεθόδων, αφού μόνο αν ληφθούν αυτά υπ' όψη, μπορούν επωφελώς να χρησιμοποιηθούν κατά τον σχεδιασμό και τη λήψη των αποφάσεων.

Βασικό διακύβευμα σε κάθε διαδικασία πρόβλεψης, είναι η ύπαρξη δεδομένων όσο το δυνατόν πιο επικαιροποιημένων και υψηλής ποιότητας, ούτως ώστε να λάβουμε και τις πιο ακριβείς προβλέψεις. Σε κάποιες περιπτώσεις ακόμα και τα δεδομένα που χρησιμοποιούνται προκειμένου να προβλεφθεί η μεταβλητή ενδιαφέροντος, προκύπτουν μέσω μιας διαδικασίας πρόβλεψης. Είναι όμως αναπόφευκτο, πολλές φορές, να μην υπάρχει πρόσβαση σε ακριβή και αξιόπιστα δεδομένα, ενώ ακόμα και αν αυτά υπάρχουν, μπορεί να μην αποτελούν τον βέλτιστο οδηγό για μια επέκταση στο μέλλον. Αυτό μπορεί να οφείλεται είτε στο ότι αφορούν πολύ παλιότερες χρονικές στιγμές, που δεν έχουν πλέον κάποια προβλεπτική αξία, είτε στο ότι η φύση του φαινομένου που προσπαθούμε να προσεγγίσουμε είναι τέτοια, ώστε να επηρεάζεται από απρόβλεπτους εξωτερικούς παράγοντες. Για παράδειγμα οι προβλέψεις στον χρηματοπιστωτικό - επιχειρηματικό τομέα, είναι πολύ επιρρεπείς σε παράγοντες όπως είναι οι νομοθετικές ρυθμίσεις (πχ regulations κρυπτονομισμάτων από την Αμερικανική κυβέρνηση), η πολιτική (ξαφνικές αλλαγές κυβέρνησης που μπορούν να επηρεάσουν δυσμενώς ή ευμενώς

κάποιον οικονομικό τομέα εξαιτίας της ιδεολογικής τοποθέτησης της νέας κυβέρνησης), οι φυσικές καταστροφές, οι πανδημίες (είναι σε όλους γνωστές οι συνέπειες του COVID19 σε δημογραφικά, οικονομικά μεγέθη και δείκτες ευημερίας των περισσότερων κρατών της γης), καθώς και έκτακτα γεγονότα (π.χ η ανακοίνωση νέου προϊόντος) που μπορούν να επηρεάσουν θετικά ή αρνητικά την εξέλιξη του υπό παρατήρηση φαινομένου, και που φυσικά μία μέθοδος πρόβλεψης δεν μπορεί να γνωρίζει εκ των προτέρων.

Τέλος πρέπει να αναφερθεί, πως τα μοντέλα πρόβλεψης συνήθως υφίστανται μία διαδικασία βελτιστοποίησης, προκειμένου να προσαρμοστούν όσο το δυνατόν καλύτερα στα υπάρχοντα δεδομένα. Ωστόσο, η προσαρμογή αυτή κάθε άλλο παρά εγγυάται πιο ποιοτική παραγωγή προβλέψεων, εκτός και αν ισχύει η υπόθεση της σταθερότητας. Η υπόθεση αυτή σημαίνει πως δεν πρέπει να συμβαίνουν αλλαγές στα ποιοτικά χαρακτηριστικά των χρονοσειρών. Ωστόσο, εγγύηση για τα τη σταθερότητα είναι προφανές πως δεν υπάρχει, και στην πραγματικότητα λαμβάνουν καθημερινά χώρα δομικές αλλαγές στην οικονομία, αλλαγές θέσεων και σκέψεων, πολιτικές κινήσεις αντίθετες προς τις υπάρχουσες τάσεις, καθώς και τεχνολογικές βελτιώσεις οι οποίες διαταράσσουν τα υπάρχοντα πρότυπα. Έτσι λοιπόν η επιστήμη των προβλέψεων, είναι καταδικασμένη να αντιμάχεται τα αενάως μεταβαλλόμενα πρότυπα συμπεριφοράς των χρονοσειρών.

2.1 Χρονοσειρές

Λέγοντας χρονοσειρά εννοούμε ένα σύνολο ιστορικών δεδομένων που απαρτίζονται από διαδοχικές παρατηρήσεις εντός ενός χρονικού διαστήματος. Συνήθως μία χρονοσειρά περιλαμβάνει παρατηρήσεις που έχουν ληφθεί με σταθερό χρονικό βήμα. Συνηθισμένα βήματα καταγραφής παρατηρήσεων είναι:

- **Ημερήσια:** τιμές της ελάχιστης και μέγιστης ημερήσιας θερμοκρασίας κ.λπ.
- **Εβδομαδιαία:** Εβδομαδιαίες πωλήσεις μιας επιχείρησης κ.λπ.
- **Μηνιαία:** Ο δείκτης εμπιστοσύνης καταναλωτή (CCI) που υπολογίζεται σε μηνιαία βάση κ.λπ.
- **Τριμηνιαία:** Ακαθάριστο Εθνικό Προϊόν (ΑΕΠ) και δείκτες ανεργίας υπολογισμένα σε τριμηνιαία βάση κ.λπ.

- **Ετήσια:** Τα προηγούμενα μεγέθη υπολογισμένα σε ετήσια βάση κ.λπ

Σημαντικό χαρακτηριστικό των χρονοσειρών είναι ότι οι διαδοχικές παρατηρήσεις δεν είναι ανεξάρτητες μεταξύ τους, δεν είναι όμως και πλήρως εξαρτημένες. Στην περίπτωση που υπήρχε μία πλήρης εξάρτηση, τότε οι μελλοντικές τιμές θα μπορούσαν να προσδιοριστούν ντετερμινιστικά από τις προηγούμενες. Ωστόσο, όπως αναφέραμε αυτό δεν συμβαίνει, και οι πραγματικές μελλοντικές τιμές καθορίζονται μόνο μερικώς από τις παρελθούσες. Στον πραγματικό κόσμο, τα περισσότερα φαινόμενα επηρεάζονται σε μεγάλο βαθμό από την τύχη και ως εκ τούτου τις περισσότερες φορές αντιμετωπίζουμε τις χρονοσειρές ως στοχαστικές διαδικασίες¹.

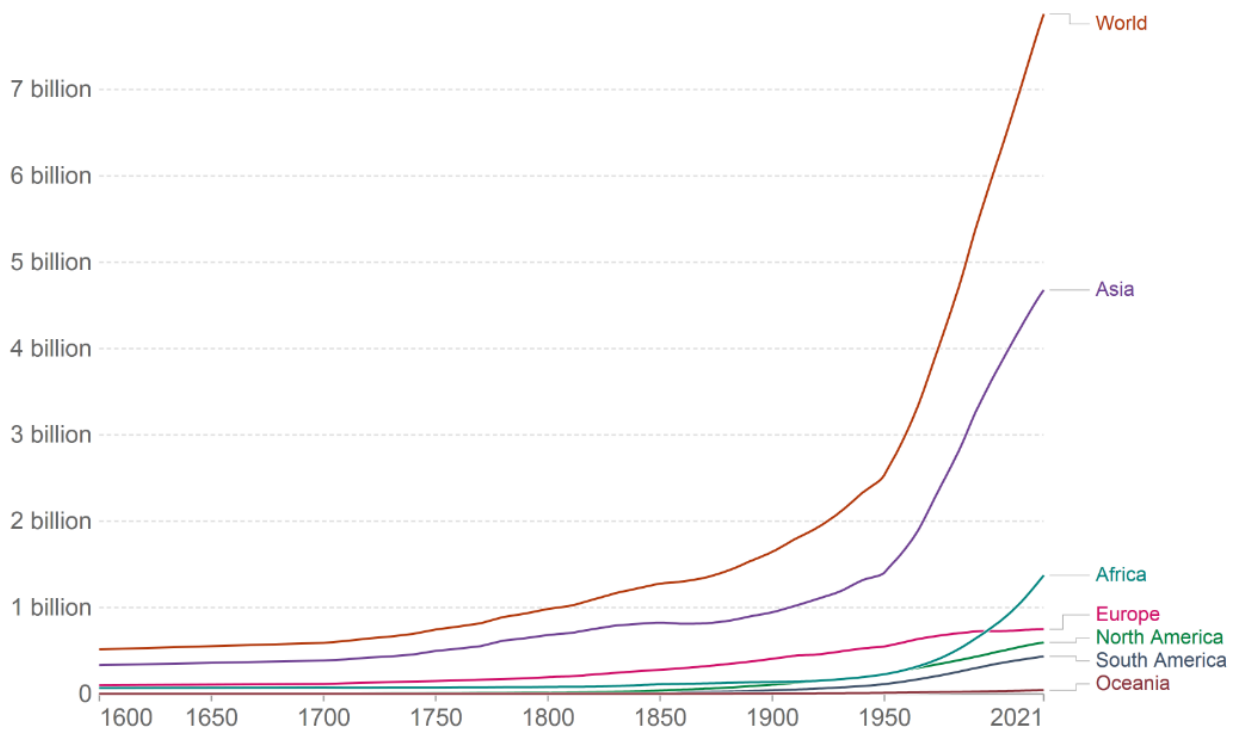
2.2 Ποιοτικά χαρακτηριστικά χρονοσειρών

Στη ενότητα αυτή παραθέτουμε συνοπτικά τα κυριότερα ποιοτικά χαρακτηριστικά των χρονοσειρών, που λαμβάνονται υπ' όψη στις διαδικασίες πρόβλεψης με σκοπό την παραγωγή αξιόπιστων αποτελεσμάτων.

2.2.1 Τάση (Trend)

Μία χρονοσειρά λέγεται ότι έχει τάση, όταν υπάρχει μία μακροχρόνια αύξηση ή ελάττωση στις τιμές των δεδομένων. Η αύξηση και η μείωση αυτή δεν είναι απαραίτητο να είναι γραμμική [2]. Χαρακτηριστικό παράδειγμα χρονοσειράς με τάση αποτελεί ο παγκόσμιος πληθυσμός, του οποίου την εξέλιξη από το 1600 μ.Χ. έως τις μέρες μας βλέπουμε στο ακόλουθο διάγραμμα [3] και που όπως παρατηρούμε παρουσιάζει εκθετική τάση.

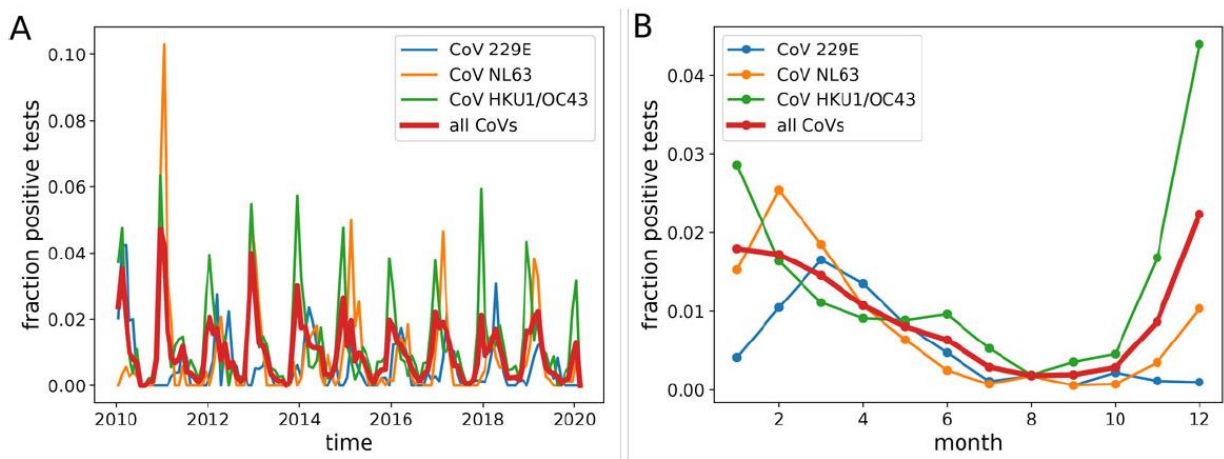
¹Ως στοχαστική ή τυχαία διαδικασία ορίζεται μία συλλογή τυχαίων μεταβλητών που δεικτοδοτούνται από κάποιο μαθηματικό σύνολο, που σημαίνει ότι κάθε τυχαία μεταβλητή της διαδικασίας σχετίζεται με μοναδικό τρόπο με ένα τυχαίο στοιχείο του συνόλου [15].



Εικόνα 1. Εξέλιξη του παγκόσμιου πληθυσμού από το 1600 μΧ.

2.2.2 Εποχικότητα (Seasonality)

Μία χρονοσειρά λέμε ότι έχει εποχικό μοτίβο, όταν επηρεάζεται από παράγοντες, όπως είναι η εποχή του χρόνου, ή η μέρα της εβδομάδας. Ως παράδειγμα τέτοιων χρονοσειρών, μπορούμε να αναφέρουμε τις κρατήσεις σε τουριστικά καταλύματα, ή τον αριθμό κρουσμάτων της εποχικής γρίπης.



Εικόνα 2. Εποχικές διακυμάνσεις του ποσοστού θετικών τεστ για κορωνοϊούς στην Σουηδία.

Στο τμήμα Α του διαγράμματος βλέπουμε το ποσοστό των θετικών τεστ κορονοϊών ανάμεσα στα έτη 2010 και 2019, ενώ στο τμήμα Β βλέπουμε τα ανά μήνα αθροισμένα ποσοστά για όλο το χρονικό διάστημα. Παρατηρούμε πως τα ποσοστά εμφανίζουν μεγάλη πτώση κατά τη διάρκεια του καλοκαιριού και τις αρχές του φθινοπώρου, ενώ οι ιοί HKU/OC43 παρουσιάζουν κορύφωση το διάστημα Δεκεμβρίου-Ιανουαρίου και οι NL63 και 229E κατά το διάστημα Φεβρουαρίου-Μαρτίου [4].

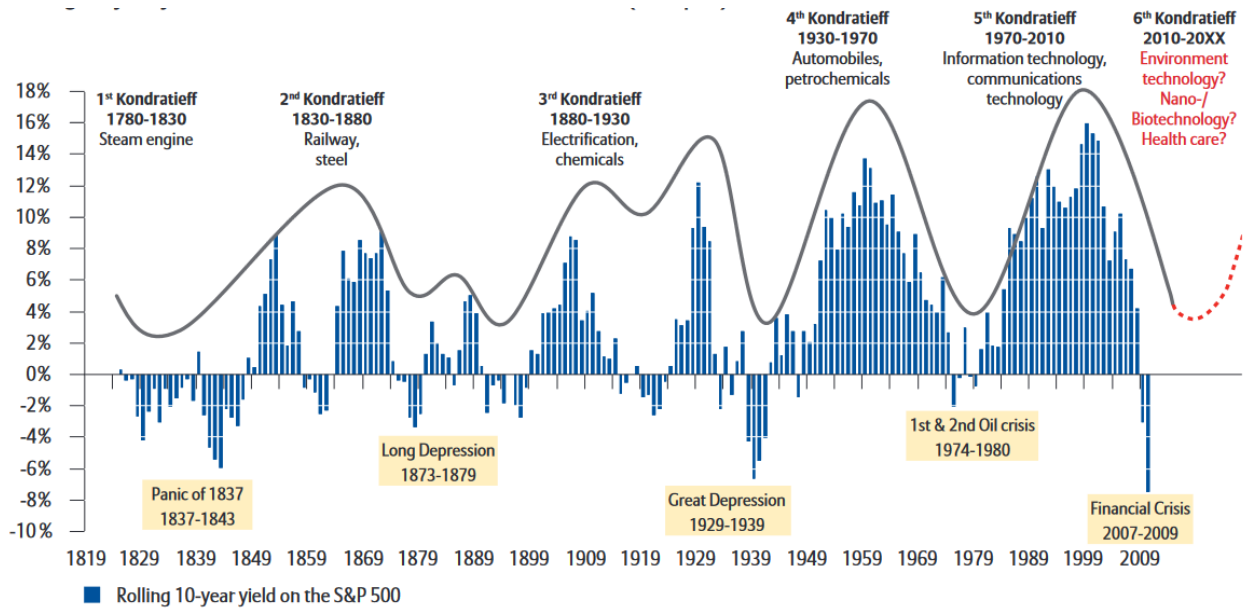
2.2.3 Στασιμότητα (Stationary)

Στάσιμη χρονοσειρά λέγεται αυτή της οποίας οι ιδιότητες δεν εξαρτώνται από τη χρονική στιγμή κατά την οποία την παρατηρούμε [2].

2.2.4 Κυκλικότητα (Cyclical)

Κύκλος λέμε ότι έχει μία χρονοσειρά όταν εμφανίζει ανόδους και πτώσεις οι οποίες δεν έχουν σταθερή συχνότητα. Αυτές οι διακυμάνσεις συνήθως οφείλονται στις οικονομικές συνθήκες που επικρατούν, και συχνά σχετίζονται με τον «επιχειρηματικό κύκλο».

Η διάρκεια των διακυμάνσεων είναι συνήθως μεγαλύτερη των δύο χρόνων [2]. Ως παράδειγμα θα μπορούσαμε να αναφέρουμε τους κύκλους ή αλλιώς κύματα Kondratieff (που επίσης ονομάζονται supercycles, great surges, long waves, K-waves ή the long economic cycle) οι οποίοι αναφέρονται στα κυκλικά φαινόμενα που διέπουν την παγκόσμια σύγχρονη οικονομία και σχετίζονται στενά με τον κύκλο ζωής της τεχνολογίας. Αναφέρεται πως η περίοδος ενός κύματος κυμαίνεται από 50 έως 60 χρόνια, και οι κύκλοι περιλαμβάνουν εναλλασσόμενα διαστήματα μεγάλης ανάπτυξης ακολουθούμενα από διαστήματα ύφεσης ή σχετικά χαμηλής ανάπτυξης [5].



Εικόνα 3. Κυλιόμενη 10ετής απόδοση του δείκτη S&P 500.

Στην εικόνα 3 παρατηρούμε τους κύκλους ή αλλιώς κύματα Kondratieff όπως αποτυπώνονται από τις αποδόσεις του χρηματιστηριακού δείκτη S&P 500 κατά τη διάρκεια των ετών 1814 έως 2009 [6].

2.2.5 Ασυνήθειες

- Ασυνήθιστες τιμές (outliers)

Ασυνήθιστες τιμές σε μία χρονοσειρά, είναι αυτές που διαφέρουν σημαντικά από τα μοτίβα (patterns) και τις τάσεις που ακολουθούν οι υπόλοιπες. Σαν παράδειγμα outlier τιμών θα μπορούσαν να αναφερθούν τα τροχαία ατυχήματα κατά τη διάρκεια μιας καταιγίδας σε σχέση με τις τιμές που λαμβάνουν υπό κανονικές συνθήκες. Μία συχνή πηγή εμφάνισης outlier τιμών είναι τα τυπογραφικά λάθη κατά τη διάρκεια καταχώρησης των δεδομένων. Η επισήμανση των τιμών αυτών είναι πολύ σημαντική κατά τη διαδικασία των προβλέψεων, καθώς είναι ικανές να αλλοιώσουν την προβλεπτική δυνατότητα του μοντέλου [7].

- Αλλαγή επιπέδου (level shift)

Όταν υφίσταται μία απότομη άνοδος ή πτώση στις τιμές μιας χρονοσειράς, από μία χρονική στιγμή και μετά.

2.2.6 Τυχειότητα

Μια χρονοσειρά χαρακτηρίζεται από τυχειότητα, όταν οι παρατηρήσεις είναι πιθανοκρατικά ανεξάρτητες μεταξύ τους. Παρά την ύπαρξη μιας πλειάδας μοντέλων και αλγορίθμων για την πρόβλεψη χρονοσειρών, είναι αδύνατο να προβλεφθεί η κατανομή της τιμής στόχου, αν είναι λευκός θόρυβος ή ακολουθεί αυτό που ονομάζουμε τυχαίο περίπατο (random walk) [8].

2.3 Βασικά Βήματα σε μια Διαδικασία Πρόβλεψης

Η συνήθης μεθοδολογία παραγωγής προβλέψεων ακολουθεί κάποια συγκεκριμένα βήματα. Θα αναφερθούμε συνοπτικά σε αυτά.

2.3.1 Καθορισμός Προβλήματος (Problem Definition)

Αποτελεί γενικά το πιο δύσκολο κομμάτι στην διαδικασία της πρόβλεψης. Στο στάδιο αυτό αναμένεται να αποσαφηνιστεί το τι ακριβώς θα προβλέψουμε, με ποιο τρόπο θα χρησιμοποιηθούν οι προβλέψεις καθώς και ποιοι θα είναι οι τελικοί χρήστες τους. Ιδιαίτερη σημασία πρέπει να δοθεί και να γίνουν συζητήσεις με τους αρμόδιους για τη συλλογή των στοιχείων και τη διαχείριση των βάσεων δεδομένων [2].

2.3.2 Συγκέντρωση πληροφοριών (Gathering Information)

Η ορθή συλλογή και συντήρηση των δεδομένων είναι επίσης κεφαλαιώδους σημασίας. Συνήθως είναι απαραίτητη η συγκέντρωση δύο ειδών πληροφοριών. Αφενός των στατιστικών που είναι ως επί το πλείστον αριθμητικά και αφετέρου δεδομένων που σχετίζονται με την κρίση, την πείρα και την εμπειρία του προσωπικού που ασχολείται με την υπό μελέτη αγορά για την εξεταζόμενη χρονική περίοδο [2].

2.3.3 Προετοιμασία χρονοσειρών (data and adjustments)

Κατά το βήμα αυτό, επιδιώκουμε να αποκτήσουμε μία γενική “αίσθηση” των δεδομένων και συγκεκριμένα να διαπιστώσουμε αν η χρονοσειρά ακολουθεί κάποια πρότυπα, αν υπάρχει τάση, εποχικότητα, ή ασυνήθιστες τιμές.

Συνήθως η διαδικασία ξεκινάει με απεικονίσεις των δεδομένων [2], και κατόπιν γίνονται κάποιες προσαρμογές έτσι ώστε να διαχειριστούμε τις κενές και τις μηδενικές τιμές και να γίνουν οι απαραίτητες ημερολογιακές προσαρμογές που θα λαμβάνουν υπ’ όψη τις εργάσιμες ημέρες, τις αργίες κ.λπ.

Κατόπιν, γίνεται απομόνωση των επί μέρους συνιστωσών της χρονοσειράς (τάση, κύκλος, εποχικότητα) και αναγνώριση ειδικών γεγονότων (special events) προκειμένου να καταλήξουμε σε μία εξομαλυμένη σειρά, στην οποία θα εφαρμοστούν τα επιλεγμένα μοντέλα.

2.3.4 Επιλογή & Προσαρμογή Μοντέλου (Choosing & Fitting models)

Στο βήμα αυτό και χρησιμοποιώντας τη γνώση που έχουμε αποκομίσει από τα προηγούμενα, καλούμαστε να επεκτείνουμε τη χρονοσειρά στο μέλλον, κάνοντας χρήση μοντέλων, ή/και κριτικής πρόβλεψης. Σημαντική είναι σε αυτό το στάδιο η επιλογή των σωστών παραμέτρων του μοντέλου, επιλογή που εξαρτάται τόσο από τις ιδιότητες της χρονοσειράς, όσο και από το μοντέλο. Όσον αφορά το ποιο είναι το βέλτιστο μοντέλο, εξαρτάται από τα ιστορικά δεδομένα που είναι διαθέσιμα και το πόσο ισχυρές είναι οι συσχετίσεις μεταξύ της μεταβλητής που θέλουμε να προβλέψουμε και των επεξηγηματικών μεταβλητών (regressors) [2].

2.3.5 Χρήση και αποτίμηση του μοντέλου πρόβλεψης (forecasting and monitoring)

Στο τελευταίο αυτό στάδιο χρησιμοποιούμε το επιλεγμένο μοντέλο με τις αντίστοιχες παραμέτρους που προέκυψαν, για την παραγωγή προβλέψεων. Καθώς ο χρόνος περνάει και νέα δεδομένα έρχονται, μπορούμε να πραγματοποιήσουμε μία ποσοτική αξιολόγηση του μοντέλου που επιλέξαμε, με βάση εξειδικευμένους στατιστικούς δείκτες [2]. Ταυτόχρονα, παρακολουθώντας το πρότυπο της χρονοσειράς και τα σφάλματα των προβλέψεων, μπορούμε να προβούμε σε διορθωτικές αλλαγές και να εξαλείψουμε κάποιες από τις μεροληψίες των μεθόδων μας. Κάτι τέτοιο συνεπάγεται ενδεχομένως την επανάληψη κάποιων από τα προηγούμενα βήματα της διαδικασίας.

2.4 Βασική Στατιστική ανάλυση

Βασική προεργασία για την μελέτη των χρονοσειρών και την διεξαγωγή της διαδικασίας των προβλέψεων είναι ο υπολογισμός των βασικών στατιστικών δεικτών (θεωρούμε πως μελετούμε μία χρονοσειρά Y η οποία έχει μέγεθος n παρατηρήσεων).

2.4.1 Μέση τιμή (Average)

Η μέση τιμή αναφέρεται στον απλό γραμμικό μέσο όρο των τιμών των παρατηρήσεων. Μας δείχνει το επίπεδο γύρω από το οποίο κινούνται οι πραγματικές τιμές της χρονοσειράς.

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

2.4.2 Μέγιστη και ελάχιστη τιμή (Maximum και Minimum)

Αποτελούν τις ακραίες τιμές της χρονοσειράς, και μας δίνουν μία ένδειξη της διακύμανσης των δεδομένων, καθώς και της τυχαιότητας που περιέχεται σε αυτά.

2.4.3 Τυπική απόκλιση (Standard Deviation)

Η τυπική απόκλιση αποτελεί ένα μέτρο του βαθμού κατά τον οποίο οι παρατηρήσεις είναι διεσπαρμένες γύρω από τη μέση τιμή καθώς και της μεταβλητότητας αυτών. Δίνεται από διαφορετικό τύπο, ανάλογα με το αν πρόκειται για δείγμα ή για πληθυσμό. Μία χαμηλή τυπική απόκλιση μας δείχνει ότι οι τιμές τείνουν να βρίσκονται σε κοντινό διάστημα γύρω από την μέση τιμή, και αντιστρόφως.

Σε περίπτωση πληθυσμού η τυπική απόκλιση δίνεται από την εξίσωση:

$$\sigma_{\text{πληθυσμού}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}}$$

Και αντίστοιχα σε περίπτωση δείγματος²:

$$\sigma_{\text{δείγματος}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}}$$

2.4.4 Διακύμανση (Variance)

Ως Διακύμανση ορίζουμε το τετράγωνο της τυπικής απόκλισης. Συνεπώς ο τύπος της διακρίνεται επίσης ανάλογα με το αν αναφερόμαστε σε δείγμα ή πληθυσμό.

² Στην περίπτωση που λαμβάνουμε τα μέτρα διασποράς (τυπική απόκλιση, διακύμανση) ενός δείγματος αντί ολόκληρου του πληθυσμού, κάνουμε τη διαίρεση με n-1 αντί για n, προκειμένου να λάβουμε μία αμερόληπτη εκτίμηση της διακύμανσης του συνολικού πληθυσμού. Αυτή η διόρθωση είναι γνωστή ως Bessel's Correction.

Σε περίπτωση πληθυσμού έχουμε την εξίσωση:

$$S^2_{\text{πληθυσμού}} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}$$

Ενώ περίπτωση δείγματος:

$$S^2_{\text{δείγματος}} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}$$

2.4.5 Συνδιακύμανση (Covariance)

Στην περίπτωση που μελετούμε δύο διαφορετικές ανεξάρτητες μεταβλητές, X και Y, ορίζουμε ως συνδιακύμανση, το μέτρο που μας δείχνει πόσο αυτές συμμεταβάλλονται, προκειμένου να καθορίσουμε τελικά το κατά πόσον μεταβάλλονται ανάλογα, αντιστρόφως ανάλογα, ή είναι ασυσχέτιστες μεταξύ τους. Στην πρώτη περίπτωση λέμε ότι έχουμε θετική συνδιακύμανση, στην δεύτερη αρνητική, ενώ όταν είναι ασυσχέτιστες οι μεταβλητές έχουμε τιμές συνδιακύμανσης κοντά στο 0. Κατά τη μελέτη των χρονοσειρών υπολογίζουμε τη συνδιακύμανσή τους ανάλογα με τον αύξοντα αριθμό της χρονικής περιόδου.

Δίνεται από την ακόλουθη σχέση:

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n [(X_i - \bar{X})[(Y_i - \bar{Y})]$$

2.4.6 Συντελεστής γραμμικής συσχέτισης (Linear Correlation Coefficient).

Ο συντελεστής αυτός, εκφράζει τη συγκέντρωση των σημείων ενός διαγράμματος διασποράς γύρω από την ευθεία παλινδρόμησης, και μας δίνει έτσι ένα μέτρο της γραμμικής συσχέτισης μεταξύ των δύο μεταβλητών που μελετάμε [9]. Οι τιμές του βρίσκονται στο διάστημα [-1,1].

Συμβολίζεται με r και αναφέρεται και ως συντελεστής Pearson.

Αν $r = \pm 1$, υπάρχει τέλεια γραμμική συσχέτιση μεταξύ των δύο μεταβλητών

Αν $-0.3 < r < 0.3$ δεν υπάρχει κάποια γραμμική συσχέτιση μεταξύ των δύο μεταβλητών, ωστόσο, μπορεί να υπάρχει κάποια άλλου είδους μη γραμμική συσχέτιση.

Αν $-0.5 < r \leq -0.3$ ή $0.3 \leq r \leq 0.5$ υπάρχει μία ασθενής γραμμική συσχέτιση (αντιστρόφως ανάλογη ή ανάλογη αντιστοίχως)

Αν $-0.7 < r \leq -0.5$ ή $0.5 \leq r \leq 0.7$ υπάρχει μέση γραμμική συσχέτιση (αντιστρόφως ανάλογη ή ανάλογη αντιστοίχως)

Αν $-0.8 < r \leq -0.7$ ή $0.7 \leq r < 0.8$, υπάρχει ισχυρή γραμμική συσχέτιση.

Αν $-1 < r \leq -0.8$ ή $0.8 \leq r < 1$, υπάρχει πολύ ισχυρή γραμμική συσχέτιση.

Στην περίπτωση που μελετούμε χρονοσειρές, η συσχέτιση γίνεται όπως και στην περίπτωση της συνδιακύμανσης, θεωρώντας ως μία μεταβλητή το χρόνο.

Ο τύπος του συντελεστή Pearson είναι ο εξής:

$$r_{XY} = \frac{\sum_{i=1}^n [(X_i - \bar{X})(Y_i - \bar{Y})]}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Παρατηρούμε πως ο συντελεστής Pearson αποτελεί ουσιαστικά το ηλίκο του συντελεστή συσχέτισης προς το γινόμενο των τυπικών αποκλίσεων των μεταβλητών [10].

2.4.7 Συντελεστής αυτοσυσχέτισης (Autocorrelation Coefficient)

Ο συντελεστής αυτοσυσχέτισης εκφράζει τη συσχέτιση μεταξύ παρατηρήσεων της ίδιας μεταβλητής με χρονική υστέρηση k περιόδους. Οι τιμές που παίρνει κυμαίνονται επίσης στο διάστημα $[0, 1]$. Εάν η τιμή είναι κοντά στο μηδέν, η συσχέτιση είναι ανύπαρκτη, ενώ αν η τιμή είναι κοντά στη μονάδα, τότε έχουμε σημαντική συσχέτιση. Η χρήση του δείκτη αυτού αποκτάει ιδιαίτερη σημασία όταν θέλουμε να εντοπίσουμε αλγεβρικά την εποχική συμπεριφορά μιας σειράς παρατηρήσεων [9].

Δίνεται από την ακόλουθη σχέση:

$$ACF_k = \frac{\sum_{i=1+k}^n [(Y_i - \bar{Y})(Y_{i-k} - \bar{Y})]}{\sum_{i=1}^n [(Y_i - \bar{Y})^2]}$$

2.4.8 Συντελεστής μεταβλητότητας (Coefficient of Variation)

Ο Συντελεστής μεταβλητότητας αποτελεί ένα κανονικοποιημένο μέτρο της διασποράς των παρατηρήσεων ενός δείγματος ή ενός πληθυσμού. Σε σύγκριση με την απλή τυπική απόκλιση έχει το πλεονέκτημα ότι είναι απαλλαγμένος από την επίδραση του επιπέδου των παρατηρήσεων και συνήθως εκφράζεται ως ποσοστό. Ωστόσο ο υπολογισμός του είναι αδύνατος όταν η μέση τιμή ισούται με μηδέν [9].

Δίνεται από την ακόλουθη σχέση:

$$c_v = \frac{\sigma}{\bar{Y}} \cdot 100(\%)$$

2.4.9 Μέση τιμή διαστήματος μεταξύ ζητήσεων (intermittent Demand Interval).

Το μέτρο αυτό αναφέρεται σε χρονοσειρές διακοπτόμενης ζήτησης και εκφράζει τη μέση τιμή των αποστάσεων διαδοχικών περιόδων με μη μηδενική τιμή. Εάν η τιμή του δείκτη αυτού είναι ίση με τη μονάδα, τότε έχουμε μία χρονοσειρά συνεχούς ζήτησης, ενώ στην περίπτωση διακοπτόμενης ζήτησης έχουμε τιμές μεγαλύτερες της μονάδας. Μία μεγάλη τιμή του δείκτη συνεπάγεται μεγάλα κατά μέσο όρο μεσοδιαστήματα μεταξύ μη μηδενικών παρατηρήσεων.

2.5 Στατιστικές Μέθοδοι προβλέψεων

Στην ενότητα αυτή θα παρουσιάσουμε τις κυριότερες στατιστικές μεθόδους πρόβλεψης.

2.5.1 Naïve

Αποτελεί την πιο απλή στατιστική μέθοδο. Επί της ουσίας η πρόβλεψη της για την επόμενη περίοδο αντιστοιχεί στην παρατηρούμενη τιμή της τρέχουσας περιόδου. Επειδή η μέθοδος αυτή προβλέπει βέλτιστα όταν τα δεδομένα ακολουθούν το μοντέλο του τυχαίου περιπάτου, αποκαλείται συχνά και random walk forecast [2].

Η μέθοδος αυτή δεν χρησιμοποιείται γενικά για την παραγωγή προβλέψεων, αλλά έχει αξία ως benchmark προκειμένου να αξιολογηθεί η ποιότητα άλλων μεθόδων.

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

2.5.2 Απλή γραμμική παλινδρόμηση

Η απλή γραμμική παλινδρόμηση αποτελεί εφαρμογή της μεθόδου ελαχίστων τετραγώνων. Βασική της υπόθεση είναι πως δύο μεταβλητές X και Y εκ των οποίων η X ονομάζεται ανεξάρτητη και η Y εξαρτημένη, συνδέονται μεταξύ τους με γραμμική σχέση της μορφής:

$$Y = a + \beta \cdot X$$

Θεωρούμε πως έχουμε n παρατηρήσεις (X_i, Y_i) και ζητούμε να προσδιορίσουμε την καλύτερη ευθεία που περιγράφει τη μεταξύ τους σχέση.

Θεωρούμε επίσης πως η ανεξάρτητη μεταβλητή X είναι πάντοτε στη σωστή της θέση, συνεπώς η απόσταση του σημείου από την ευθεία δίνεται ως η απόσταση της εξαρτημένης μεταβλητής Y από αυτήν:

$$d_i = Y_i - a - \beta X_i$$

Αποδεικνύεται θεωρητικά ότι η πιθανότερη ευθεία που αντιστοιχεί στις παρατηρήσεις είναι αυτή που δίνεται από τις τιμές του a και του β που ελαχιστοποιούν το άθροισμα των ελαχίστων τετραγώνων των αποκλίσεων των σημείων από την ευθεία [11].

$$S = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n (Y_i - a - \beta X_i)^2 = \text{ελάχιστο}$$

Οι αναγκαίες συνθήκες για την ελαχιστοποίηση του S είναι οι:

$$\frac{dS}{da} = 0 \text{ και } \frac{dS}{d\beta} = 0$$

Από τις οποίες προκύπτουν οι σχέσεις:

$$a n + \beta \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i$$

και:

$$a \sum_{i=1}^n X_i + \beta \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i \cdot Y_i$$

Οι δύο αυτές σχέσεις μπορούν να λυθούν ως προς α και β και τελικά παίρνουμε:

$$\beta = \frac{\sum_{i=1}^n [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

και

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

Όπου \bar{X} και \bar{Y} είναι οι μέσες τιμές των διανυσμάτων X και Y .

Με αυτό τον τρόπο, και έχοντας προσδιορίζει την ευθεία παλινδρόμησης, μπορούμε να τη χρησιμοποιήσουμε ως μοντέλο πρόβλεψης, δηλαδή να προσδιορίσουμε την τιμή του Y για μία μελλοντική τιμή του X .

Εξαιτίας της υπόθεσης γραμμικής συσχέτισης την οποία κάναμε, υπάρχουν προφανώς πολλές περιπτώσεις στις οποίες η εφαρμογή της απλής γραμμικής παλινδρόμησης είναι ακατάλληλη ως μοντέλο [9]. Παράδειγμα τέτοιας περίπτωσης είναι οι πωλήσεις ενός προϊόντος ή μιας υπηρεσίας που εξαρτώνται από τις εποχές του χρόνου, όπως για παράδειγμα η πρόβλεψη πωλήσεων παγωτών που εξαρτώνται σχεδόν αποκλειστικά από την εποχή του έτους.

2.5.3 Πολλαπλή γραμμική παλινδρόμηση

Σε πολλές περιπτώσεις υπάρχει μια γραμμική σχέση μεταξύ μιας εξαρτημένης μεταβλητής Y και περισσότερων της μίας ανεξάρτητων μεταβλητών X . Η γενική μορφή της πολλαπλής παλινδρόμησης είναι :

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

Όπου το e αναφέρεται στο τυχαίο σφάλμα το οποίο θεωρείται πως ακολουθεί κανονική κατανομή με μέση τιμή το μηδέν [9].

Στην περίπτωση της πολλαπλής γραμμικής παλινδρόμησης έχουμε μία γραμμική σχέση ως προς τους συντελεστές β , τους οποίους μπορούμε να υπολογίσουμε με τη μέθοδο των ελαχίστων τετραγώνων επί του συνόλου δεδομένων που διαθέτουμε, με μεθοδολογία αντίστοιχη της απλής γραμμικής παλινδρόμησης.

Πρέπει να επισημάνουμε εδώ πως στην περίπτωση που το Y εξαρτάται μόνο από μία μεταβλητή (απλή γραμμική παλινδρόμηση), τότε μπορούμε να το παραστήσουμε με μία ευθεία στο επίπεδο, όταν εξαρτάται από δύο μεταβλητές μπορεί να παρασταθεί με ένα επίπεδο, ενώ αν έχουμε παραπάνω από δύο μεταβλητές, τότε το Y παριστάνεται σε επιφάνεια με περισσότερες από δύο διαστάσεις γνωστή και ως υπερεπίπεδο [9].

2.5.4 Μέθοδοι Εκθετικής Εξομάλυνσης

Αναπτύχθηκαν την δεκαετία του 50 και έγιναν από τις πιο δημοφιλείς μεθόδους προβλέψεων στον κόσμο των επιχειρήσεων κυρίως λόγω της ευκολίας τους, της ελάχιστης απαίτησης σε υπολογιστικό χρόνο, καθώς και την ανάγκη ύπαρξης λίγων αναγκαίων παρατηρήσεων προκειμένου να παραχθούν προβλέψεις. Οι μέθοδοι αυτές είναι κατάλληλες για βραχυπρόθεσμες προβλέψεις ενός μεγάλου όγκου χρονοσειρών. Αποδίδουν καλύτερα σε δεδομένα που παρουσιάζουν στασιμότητα ή μικρό ρυθμό ανάπτυξης ή μείωσης ως προς το χρόνο.

Οι κυριότερες μέθοδοι της κατηγορίας αυτής είναι οι :

2.5.5 Μοντέλο σταθερού επιπέδου (Simple Exponential Smoothing (SES))

Η πιο απλή μέθοδος εκθετικής εξομάλυνσης, η οποία είναι κατάλληλη για την πρόβλεψη χρονοσειρών που δεν εμφανίζουν ξεκάθαρη τάση ή εποχικότητα [9]. Το μοντέλο περιγράφεται από τις εξισώσεις:

$$e_t = Y_t - F_t$$

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

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

Η τιμή e δηλώνει την απόκλιση της πραγματικής τιμής από την πρόβλεψη (σφάλμα), S_t είναι το επίπεδο της χρονοσειράς κατά την χρονική στιγμή t και F_t η αντίστοιχη πρόβλεψη. Η παράμετρος a είναι ο συντελεστής εξομάλυνσης της μεθόδου και μπορεί να πάρει τιμές στο διάστημα $[0,1]$.

2.5.6 Μοντέλο γραμμικής τάσης (Holt Exponential Smoothing)

Η μέθοδος Holt που είναι γνωστή και ως γραμμική εκθετική εξομάλυνση, είναι μία δημοφιλής μέθοδος εξομάλυνσης για την πρόβλεψη δεδομένων που εμφανίζουν τάση. Η επέκταση προτάθηκε από τον Holt το 1957. Το μοντέλο της περιγράφεται από τις ακόλουθες εξισώσεις:

$$e_t = Y_t - F_t$$

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

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

$$F_{t+m} = S_t + m \cdot T_t$$

Κατ' αντιστοιχία και με την απλή εκθετική εξομάλυνση, e είναι το σφάλμα, S το επίπεδο της χρονοσειράς, F η πρόβλεψη, t η χρονική περίοδος, και m ο χρονικός ορίζοντας στον οποίο εκτείνουμε την πρόβλεψη. Τέλος, με a συμβολίζεται ο συντελεστής εξομάλυνσης του επιπέδου, ενώ με β ο συντελεστής εξομάλυνσης της τάσης. [9].

Παρατηρούμε εδώ, ότι η συνάρτηση πρόβλεψης δεν είναι πλέον επίπεδη αλλά με τάση. Η πρόβλεψη για m βήματα μπροστά, ισούται με την τελευταία τιμή της συνάρτησης του επιπέδου, αυξημένη κατά m φορές την τιμή της τελευταίας εκτίμησης της τάσης. Συνεπώς οι προβλέψεις που προκύπτουν είναι μία γραμμική συνάρτηση του m [2].

2.5.7 Damped

Οι προβλέψεις που παράγονται από την γραμμική μέθοδο Holt, προβάλλουν μία συνεχή τάση της χρονοσειράς προς το μέλλον. Ωστόσο εμπειρικές μελέτες έχουν δείξει πως αυτές οι μέθοδοι τείνουν να «υπέρ-προβλέπουν», ειδικά όταν πρόκειται για μακρύτερους χρονικούς ορίζοντες. αυτήν την αδυναμία έρχεται να διορθώσει η μέθοδος Damped (φθίνουσας ή μη γραμμικής τάσης) με τροποποιήσεις στο μοντέλο Holt (Gardner και McKenzie 1985). Στη μέθοδο αυτή εκτελείται μία απλή γραμμική παλινδρόμηση με το χρόνο ως ανεξάρτητη μεταβλητή:

$$X = A + B \cdot t$$

ορίζεται ως αρχικό επίπεδο η σταθερά A της παλινδρόμησης, ως αρχική τάση η κλίση B της παλινδρόμησης, κι επιλέγεται ένας damping factor φ .

Οι εξισώσεις που περιγράφουν το εν λόγω μοντέλο είναι οι :

$$e_t = Y_t - F_t$$

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

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

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

Παρατηρούμε πως υπάρχει ομοιότητα των εξισώσεων με αυτές του γραμμικού μοντέλου, εκτός από την τελευταία, όπου αντί για μια γραμμική αύξηση μέσω του συντελεστή m , ουσιαστικά πραγματοποιείται ένας μη γραμμικός υπολογισμός μέσω του συντελεστή εξομάλυνσης φ .

Αν ο συντελεστής αυτός είναι ίσος με 0 τότε η μέθοδος εκφυλίζεται σε σταθερού επιπέδου (SES) ενώ αν είναι ίσος με 1 τότε ταυτίζεται με την μέθοδο γραμμικής τάσης (Holt). Οι συνήθεις τιμές που παίρνει ο damping factor είναι μεταξύ 0 και 1, οπότε και προκύπτει το λεγόμενο μοντέλο φθίνουσας τάσης, ενώ τέλος αν το φ πάρει τιμές μεγαλύτερες τις μονάδας, προκύπτει το μοντέλο της εκθετικής τάσης [9].

2.5.8 Winters

Τα μοντέλα που περιγράφηκαν μέχρι στιγμής είναι μη εποχικά μοντέλα και είναι κατάλληλα για χρονοσειρές που εμφανίζουν μόνο πρότυπα τάσεων. Σε περίπτωση όμως που έχουμε έντονη εποχική συνιστώσα, μπορούμε να τα επεκτείνουμε με την προσθήκη ενός εξομαλυσμένου εποχικού παράγοντα για κάθε χρονική περίοδο σε διάστημα ενός έτους, οποίος αποσκοπεί στην διόρθωση των προβλέψεων ανάλογα με την αναμενόμενη εποχική διακύμανση.

Οι εποχικές διακυμάνσεις μπορεί να είναι είτε προσθετικού είτε πολλαπλασιαστικού χαρακτήρα. Στην πρώτη περίπτωση ο εποχικός παράγοντας, ορίζεται ως η διαφορά κάθε σημείου της χρονοσειράς από τον μέσο όρο των τιμών της για όλο το έτος. Συνεπώς με τη

χρήση του παράγοντα αυτού προκύπτουν οι τιμές από τις οποίες έχει αφαιρεθεί το εποχικό πρότυπο.

Αποεποχικοποιημένες τιμές = πραγματικές τιμές – εποχικός παράγοντας

Αντιθέτως στη δεύτερη περίπτωση, ο εποχικός παράγοντας ορίζεται ως ο λόγος κάθε τιμής της χρονοσειράς προς το μέσο όρο των τιμών της για όλο το έτος. Κατ' αντιστοιχία έχουμε ότι:

Αποεποχικοποιημένες τιμές = πραγματικές τιμές / εποχικός παράγοντας.

Η μέθοδος Winters χρησιμοποιεί τις συνιστώσες του επιπέδου, της τάσης και της εποχικότητας για την παραγωγή προβλέψεων. Σε κάθε περίοδο γίνεται χρήση τριών βαρών (ή παραμέτρων εξομάλυνσης) προκειμένου να ενημερωθούν οι τιμές της εκάστοτε συνιστώσας. Το μοντέλο αυτό [12], είναι σταθερού επιπέδου με πολλαπλασιαστική εποχικότητα και περιγράφεται από τις ακόλουθες εξισώσεις:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + \frac{a \cdot e_t}{I_{t-p}}$$

$$I_t = I_{t-p} + \frac{\gamma \cdot e_t}{S_{t-1}}$$

$$F_{t+m} = S_t \cdot I_{t-p+m}$$

Στις εξισώσεις αυτές, συμβολίσαμε με I τον εξομαλυμένο εποχικό παράγοντα. Θεωρούμε πως το έτος αποτελείται από p (periods) περιόδους και συνεπώς θα υπάρχουν p εποχικοί παράγοντες, ένας δηλαδή για κάθε περίοδο. Συνεπώς ο όρος I_t αναφέρεται στην περίοδο t ενώ ο όρος I_{t-p} αναφέρεται στον εποχικό παράγοντα πριν από p περιόδους. Η πρόβλεψη της μεθόδου προκύπτει από το γινόμενο του αποεποχικοποιημένου επιπέδου με τον κατάλληλο εποχικό παράγοντα για τη χρονική περίοδο στο μέλλον που αφορά η πρόβλεψη [9].

2.5.9 Μέθοδος Theta

Η μέθοδος θ [13] είναι μια μονοδιάστατη μέθοδος πρόβλεψης. Βασίζεται στην μεταβολή των τοπικών καμπυλοτήτων μιας χρονοσειράς μέσα από την παράμετρο θ (Theta), η οποία εφαρμόζεται απευθείας (πολλαπλασιαστικά) στις διαφορές δεύτερης τάξης των δεδομένων. Συγκεκριμένα:

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

Όπου:

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

Με τον τρόπο αυτό δημιουργείται μια νέα χρονοσειρά που διατηρεί την μέση τιμή και την κλίση παλινδρόμησης της αρχικής χρονοσειράς αλλά όχι και τις τοπικές καμπυλότητες. Οι χρονοσειρές που παράγονται με αυτή την διαδικασία ονομάζονται γραμμές θ (Theta Lines). Βασικό ποιοτικό χαρακτηριστικό αυτών των γραμμών είναι η καλύτερη προσέγγιση της μακροπρόθεσμης συμπεριφοράς και τάσης των δεδομένων και η ανάδειξη-τονισμός των βραχυπρόθεσμων χαρακτηριστικών, ανάλογα με την τιμή της παραμέτρου θ .

Όσο μικρότερη είναι η τιμή της παραμέτρου, τόσο μεγαλύτερος είναι και ο βαθμός μείωσης των καμπυλοτήτων. Αντίθετα, σε περίπτωση που έχουμε τιμές του θ μεγαλύτερες του 1, τότε προκύπτουν χρονοσειρές διογκωμένες σε σχέση με την αρχική, κάτι που έχει ως αποτέλεσμα την ενίσχυση της βραχυπρόθεσμης συμπεριφοράς της. Στην οριακή περίπτωση όπου $\theta = 0$ η χρονοσειρά που δημιουργείται είναι αυτή της απλής γραμμικής παλινδρόμησης.

Η μέθοδος αποσυνθέτει (διαχωρίζει) την αρχική χρονοσειρά σε δύο ή περισσότερες γραμμές Theta. Η κάθε γραμμή Theta προεκτείνεται στο μέλλον ξεχωριστά (με την ίδια ή και με διαφορετικές μεθόδους πρόβλεψης) και οι παραγόμενες προβλέψεις συνδυάζονται για να προκύψει η τελική πρόβλεψη.

Τα βασικά βήματα που ακολουθούνται στη διαδικασία της κλασικής μεθόδου Theta είναι τα εξής:

- **Έλεγχος εποχικότητας:** Ελέγχουμε τη χρονοσειρά για στατιστικά σημαντική εποχική συμπεριφορά

- **Αποεποχικοποίηση:** Εφ' όσον διαπιστώσουμε τέτοια συμπεριφορά από τον έλεγχο του προηγούμενου βήματος, αποεποχικοποιούμε τη χρονοσειρά μέσω της μεθόδου πολλαπλασιαστικής αποσύνθεσης.
- **Αποσύνθεση:** Αποσυνθέτουμε τη χρονοσειρά σε δύο γραμμές Theta. Αυτές είναι η ευθεία γραμμικής παλινδρόμησης ($\theta = 0$) και η γραμμή Theta που προκύπτει για την τιμή $\theta = 2$.
- **Πρόβλεψη:** Λαμβάνουμε υπ' όψη της μελλοντικές τιμές της ευθείας γραμμικής παλινδρόμησης, καθώς και τις τιμές που προκύπτουν από την δεύτερη γραμμή αν εφαρμόσουμε σε αυτήν απλή εκθετική εξομάλυνση.
- **Συνδυασμός:** Παίρνουμε τον απλό μέσο όρο των προβλέψεων των δύο γραμμών (ίσα βάρη).
- **Εποχικοποίηση:** Εποχικοποιούμε τις τελικές προβλέψεις χρησιμοποιώντας τους δείκτες εποχικότητας που υπολογίστηκαν στο αντίστοιχο βήμα.

Στην περίπτωση που θέλουμε να συμπεριλάβουμε περισσότερες από δύο γραμμές Theta, πρέπει είτε να επιλύσουμε ένα σύστημα γραμμικών εξισώσεων, είτε να υπολογίσουμε μια απλή παλινδρόμηση.

Παραθέτουμε παρακάτω έναν εναλλακτικό τρόπο υπολογισμού της γραμμής Theta για $\theta=2$ έχοντας προσδιορίσει τους συντελεστές της απλής γραμμικής παλινδρόμησης [9]:

$$Y_t = \frac{1}{2} \cdot (Y_t^{\theta=1+a} + Y_t^{\theta=1-a}) \stackrel{a=1}{\Leftrightarrow}$$

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

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

$$2 \cdot Y_t = LRL_t + Y_t^{\theta=2} \Leftrightarrow$$

$$Y_t^{\theta=2} = LRL_t - 2 \cdot Y_t$$

Παρατηρούμε από τα παραπάνω, πως με αυτό τον τρόπο οι υπολογισμοί για τον υπολογισμό της γραμμής Theta για $\theta=2$ απλοποιούνται πολύ.

Ο απλός συνδυασμός δύο γραμμών Theta, για $\theta=0$ (ευθεία γραμμή) και $\theta=2$ (διπλασιασμός των τοπικών καμπυλοτήτων) χρησιμοποιήθηκε για την παραγωγή προβλέψεων για τις 3003 χρονοσειρές του διεθνούς διαγωνισμού προβλέψεων M3 [14].

Η μέθοδος παρήγαγε πολύ καλά αποτελέσματα, με μικρά σφάλματα προβλεπτικής ακρίβειας, ιδιαίτερα για τις μηνιαίες χρονοσειρές και τα μικροοικονομικά δεδομένα.

2.5.10 Ολοκληρωμένα Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου Όρου (ARIMA).

Τα ολοκληρωμένα αυτοπαλινδρομικά μοντέλα κινητών μέσων όρων ARIMA (Auto Regressive Integrated Moving Average) (Box Jenkins 1970) είναι στοχαστικά μοντέλα τα οποία μας βοηθάνε να περιγράψουμε το μηχανισμό εξέλιξης ενός μεγέθους ενδιαφέροντος και να προβλέψουμε τις μελλοντικές τιμές του. Αποτελούν γενίκευση του μοντέλου αυτοπαλινδρομικού κινητού μέσου όρου (ARMA) και επί της ουσίας και τα δύο εκτελούν μία προσαρμογή (fitting) στην χρονοσειρά προκειμένου να γίνει καλύτερη κατανόηση των δεδομένων.

Όπως και άλλα στατιστικά μοντέλα πρόβλεψης, για παράδειγμα αυτά της εκθετικής εξομάλυνσης, τα μοντέλα ARIMA παράγουν προβλέψεις βασισμένα αποκλειστικά στις ιστορικές παρατηρήσεις του υπό εξέταση μεγέθους. Αυτό αποτελεί πλεονέκτημα, ειδικά σε περιπτώσεις που δεν γνωρίζουμε τις εξωτερικές μεταβλητές που το επηρεάζουν ή δεν μπορούμε να τις αξιοποιήσουμε άμεσα (βλ. ντετερμινιστικά μοντέλα).

Κάθε μοντέλο ARIMA εκφράζει ένα διαφορετικό μηχανισμό εξέλιξης και η επιλογή του καταλληλότερου για την προέκταση μιας χρονοσειράς, γίνεται εξετάζοντας παράγοντες όπως η σχέση μεταξύ διαδοχικών παρατηρήσεων (αυτοσυσχέτιση), η ύπαρξη τάσης, η ύπαρξη εποχικότητας και το σφάλμα πρόβλεψης.

Στην γενική τους μορφή τα μοντέλα ARIMA αποτελούνται από τον τυχαίο παράγοντα (παράγοντας MA), τις τιμές που εμφανίστηκαν σε προηγούμενες περιόδους (παράγοντας AR και I) και άλλες στοχαστικές μεταβλητές. Πιο συγκεκριμένα, κάθε μοντέλο ARIMA μπορεί να εκφραστεί ως γραμμικός συνδυασμός των παραπάνω παραγόντων και στόχος μας είναι να ανακαλύψουμε εκείνον που παράγει τις καλύτερες προβλέψεις. Αν το μοντέλο περιλαμβάνει αποκλειστικά παράγοντες αυτοπαλινδρόμησης αναφέρεται ως AR(p), αν περιλαμβάνει

αποκλειστικά παράγοντες κινητών μέσω όρων ως MA(q), και αν περιλαμβάνει και τους δύο ως ARMA(p,q), όπου τα p και q δηλώνουν την τάξη του μοντέλου ανά παράγοντα. Ο παράγοντας I(d) αναφέρεται στη διαφορίση της χρονοσειράς πριν την εφαρμογή ενός μοντέλου ARMA(p,q) και έχει ως στόχο την αφαίρεση της τάσης από τα δεδομένα.

Για την εφαρμογή των μοντέλων ARIMA είναι αναγκαίες οι ακόλουθες προϋποθέσεις:

- **Διακριτότητα.** Η χρονοσειρά πρέπει να αποτελείται από τιμές που αναφέρονται σε ισαπέχουσες χρονικά παρατηρήσεις $Y_t, Y_{t+\tau} \dots Y_{t+\tau}$ με τ ακέραιο μεγαλύτερο του μηδενός.
- **Στασιμότητα.** Η προϋπόθεση αυτή συνεπάγεται πως η μέση τιμή (μ), η διακύμανση σ^2 καθώς και η συνάρτηση αυτοσυσχέτισης (ACF) διατηρούνται σταθερές στο πέρασμα του χρόνου. Όταν η χρονοσειρά δεν πληροί την συγκεκριμένη προϋπόθεση, τότε μπορούμε να την μετατρέψουμε σε στάσιμη, κάνοντας χρήση διαφορίσης που πραγματοποιείται με τον υπολογισμό των διαφορών μεταξύ διαδοχικών παρατηρήσεων.

$$y'_t = y_t - y_{t-1}$$

Για να εξακριβωθεί με αυτόματο και στατιστικό τρόπο το αν υπάρχει ανάγκη διαφορίσης, υπάρχουν διάφοροι έλεγχοι υποθέσεων που κάνουν την αρχική υπόθεση πως η χρονοσειρά δεν είναι στάσιμη, και ανάλογα με τα αποτελέσματα αποδέχονται ή απορρίπτουν την αρχική υπόθεση. Ο πιο διαδεδομένος από τους ελέγχους αυτούς είναι ο έλεγχος Dickey-Fuller.

- **Παραγωγή βραχυχρόνιων προβλέψεων.** Επειδή τα μοντέλα ARIMA αποτελούν γραμμικούς συνδυασμούς ιστορικών παρατηρήσεων και στοχαστικών παραγόντων, η παραγωγή προβλέψεων για μελλοντικές χρονικές στιγμές απαιτεί τη γνώση των προηγούμενων τιμών. Συνεπώς, αν επιδιώκουμε να προβλέψουμε χρονικές στιγμές πολύ μεταγενέστερες, οι προβλέψεις αυτές θα πρέπει αναγκαστικά να βασιστούν σε αυτές που έγιναν για πιο βραχείς χρονικούς ορίζοντες, με αποτέλεσμα η προσδοκώμενη ακρίβεια αναγκαστικά να ελαττώνεται, αφού οι προβλέψεις στις οποίες βασιζόμαστε περιέχουν εγγενώς και σφάλματα.

Θα αναφερθούμε ακολούθως σε ορισμένες γνωστές περιπτώσεις μοντέλων ARIMA που είτε προκύπτουν με φυσικό τρόπο, είτε είναι μαθηματικώς αντίστοιχες με άλλα δημοφιλή μοντέλα πρόβλεψης:

- Το μοντέλο ARIMA (0,1,0) ή αλλιώς I(1) δίνεται από τη σχέση: $X_t = X_{t-1} - \varepsilon_t$, όπου ε_t είναι το σφάλμα, και επί της ουσίας αποτελεί έναν τυχαίο περίπατο.
- Το ARIMA (0,0,0) είναι ένα μοντέλο λευκού θορύβου
- Το ARIMA (0,1,2) αντιστοιχεί στο μοντέλο Damped Holt, δηλαδή εκθετική εξομάλυνση με μη γραμμική τάση.
- Το ARIMA (0,1,1) χωρίς σταθερά, αποτελεί ένα μοντέλο απλής εκθετικής εξομάλυνσης [15].

2.6 Ανάλυση ακρίβειας προβλέψεων

Θα αναφερθούμε ακολούθως, στις κυριότερες μετρικές που χρησιμοποιούνται για την αξιολόγηση της ακρίβειας των προβλέψεων και στις βασικές τους ιδιότητες.

2.6.1 Μέσο σφάλμα (Mean Error).

Υπολογίζεται από τον απλό προσημασμένο μέσο όρο των σφαλμάτων και εκφράζει ένα μέτρο συστηματικότητας του σφάλματος. Όσο η τιμή του είναι κοντά στο μηδέν, τα σφάλματα τείνουν να είναι τυχαία και όχι συστηματικά. Όταν ο δείκτης λαμβάνει θετικές τιμές, τότε οι προβλέψεις χαρακτηρίζονται από απαισιοδοξία (εφ' όσον οι προβλέψεις ήταν μικρότερες από τις πραγματικές) και αντίστροφα. Αναφέρεται συχνά και ως bias (μεροληψία).

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)$$

2.6.2 Μέσο απόλυτο σφάλμα (Mean Absolute Error).

Ο δείκτης αυτός δηλώνει ένα μέσο μέτρο αστοχίας της πρόβλεψης λαμβάνοντας τον μέσο όρο των απολύτων τιμών των διαφορών του επιπέδου από την πρόβλεψη. Όσο μεγαλύτερη είναι η τιμή του δείκτη, τόσο μικρότερη είναι και η ακρίβεια της μεθόδου. Παρατηρούμε πως με τη χρήση του δείκτη αυτού δεν μπορούμε να αποκομίσουμε εικόνα σχετικά με το αν οι προβλέψεις μας είναι αισιόδοξες ή απαισιόδοξες.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - F_i|$$

2.6.3 Μέσο τετραγωνικό σφάλμα (Mean Squared Error)

Όπως και το MAE, το MSE, αποτελεί μέτρο ακρίβειας της πρόβλεψης χωρίς να μας δίνει πληροφορία σχετικά με το αν είναι αισιόδοξη ή απαισιόδοξη. Η διαφοροποίησή του σε σχέση με το MAE έγκειται στο γεγονός ότι το εν λόγω δίνει πολύ μεγαλύτερο βάρος στα μεγάλα σφάλματα (αφού λαμβάνεται υπ' όψη το τετράγωνο των σφαλμάτων) ενώ αντίστοιχα μικρότερο βάρος δίνεται στα μικρότερα σφάλματα. Βρίσκει μεγάλη εφαρμογή στον υπολογισμό των βέλτιστων παραμέτρων εξομάλυνσης.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2$$

2.6.4 Ρίζα μέσου τετραγωνικού Σφάλματος (Root Mean Squared Error)

Υπολογίζεται από την τετραγωνική ρίζα του MSE. Έχει τις ίδιες ιδιότητες με αυτό, όντας εκπεφρασμένο στις μονάδες της αρχικής χρονοσειράς.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2}$$

2.6.5 Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error).

Στις περιπτώσεις όπου είναι πιο χρήσιμος ο υπολογισμός σφαλμάτων πρόβλεψης σε καθαρά ποσοστιαία μορφή, όπως για παράδειγμα στην περίπτωση που θέλουμε να συγκρίνουμε την ακρίβεια μιας μεθόδου πρόβλεψης που έχει εφαρμοστεί σε παραπάνω από μία χρονοσειρές οι οποίες έχουν διαφορετικά επίπεδα μέσης τιμής μεταξύ τους, χρησιμοποιούμε τη

συγκεκριμένη μετρική. Μεγάλη χρησιμότητα εμφανίζει επίσης στις περιπτώσεις όπου οι πραγματικές τιμές (Y), είναι ιδιαίτερα υψηλές. Ο δείκτης MAPE είναι εκπεφρασμένος επί τοις εκατό και λαμβάνει τιμές μεγαλύτερες ή ίσες του μηδενός. Είναι εμφανές πως οι μικρότερες τιμές του δείκτη υποδηλώνουν και καλύτερη απόδοση της μεθόδου πρόβλεψης.

Ωστόσο ο δείκτης δεν μπορεί να εφαρμοστεί σε χρονοσειρές διακοπτόμενης ζήτησης, καθώς καταλήγει σε απροσδιοριστία.

Υπολογίζεται από την ακόλουθη εξίσωση:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{Y_i} \right| \cdot 100(\%)$$

2.6.6 Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error).

Αποτελεί μια παραλλαγή του μέσου απόλυτου ποσοστιαίου σφάλματος. Στον εν λόγω δείκτη [16], η απόλυτη τιμή του σφάλματος δεν διαιρείται με την πραγματική τιμή αλλά με το ημιάθροισμα της πραγματικής τιμής και της πρόβλεψης. Με την τροποποίηση αυτή, ο δείκτης αποκτά ένα άνω όριο με αποτέλεσμα να λαμβάνει τιμές στο διάστημα [0% - 200%].

Ωστόσο βασικό πρόβλημα του εν λόγω δείκτη, καθώς δεν είναι απόλυτα συμμετρικός, όπως αναφέρεται στο όνομά του [17], έγκειται στο γεγονός ότι οι αισιόδοξες και οι απαισιόδοξες προβλέψεις δεν τυγχάνουν αντίστοιχης μεταχείρισης.

Αυτό γίνεται εμφανές στο εξής παράδειγμα :

Αισιόδοξη πρόβλεψη:

$$Y_t = 100 \text{ και } F_t = 110 \rightarrow sMAPE = 4.76\%$$

Απαισιόδοξη πρόβλεψη:

$$Y_t = 100 \text{ και } F_t = 90 \rightarrow sMAPE = 5.76\%$$

Ο δείκτης υπολογίζεται από την ακόλουθη εξίσωση:

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{\frac{(Y_i + F_i)}{2}} \right| \cdot 100(\%) = \frac{1}{n} \sum_{i=1}^n \left| \frac{2 \cdot (Y_i - F_i)}{Y_i + F_i} \right| \cdot 100(\%)$$

2.6.7 Σχετικοί δείκτες.

Στις περιπτώσεις που εφαρμόζουμε περισσότερες από μία μεθόδους πρόβλεψης, υπολογίζοντας για την καθεμία κάποιον από τους δείκτες σφάλματος που προαναφέρθηκαν, αποφαινόμαστε για την καταλληλότητα τους ως προς τις συγκεκριμένες χρονοσειρές, βάσει της σωστής προσαρμογής του μοντέλου. Η σύγκριση μπορεί να γίνει βάσει κάποιας μεθόδου που θεωρείται ως μέθοδος αναφοράς (benchmark). Ορίζουμε λοιπόν το σχετικό σφάλμα της μεθόδου, ως το πηλίκο του σφάλματος της μεθόδου προς το σφάλμα της μεθόδου αναφοράς όπως φαίνεται στην ακόλουθη εξίσωση. Έστω για παράδειγμα ότι ορίζουμε σαν δείκτη σφάλματος το MAE. Τότε η σχετική τιμή του μέτρου της μεθόδου i σε σχέση με την benchmark μέθοδο, θα δίνεται από την ακόλουθη εξίσωση:

$$RelMAE = \frac{MAE_i}{MAE_{benchmark}}$$

Σχετικό σφάλμα μπορεί να εφαρμοστεί για όλους τους δείκτες που προαναφέραμε.

2.6.8 Μέσο απόλυτο κανονικοποιημένο σφάλμα (Mean Absolute Scaled Error).

Οι Hyndman και Koehler (2006) [18], προκειμένου να αντιμετωπίσουν τις περιπτώσεις απροσδιοριστίας των δεικτών MAPE και sMAPE αλλά και να δώσουν την ίδια βαρύτητα στα μικρά και τα μεγάλα σφάλματα, σε αντίθεση με τους δείκτες MSE και RMSE, πρότειναν τον συσχετιστικό στατιστικό δείκτη MAsE για την μέτρηση της ακρίβειας των προβλέψεων. Ο δείκτης δίνεται από την ακόλουθη εξίσωση:

$$MAsE = \frac{\frac{1}{n} \sum_{i=1}^n |Y_i - F_i|}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|}$$

Η εξίσωση του, θυμίζει αρκετά το μέσο απόλυτο σφάλμα, το οποίο είναι κανονικοποιημένο με τη μέση τιμή των διαφορών πρώτου βαθμού της χρονοσειράς. Ο παρονομαστής ουσιαστικά αποτελεί το μέσο απόλυτο σφάλμα του μοντέλου πρόβλεψης, όταν ως μέθοδος πρόβλεψης έχει εφαρμοσθεί η μέθοδος Naïve, η οποία δίνει ως πρόβλεψη της επόμενης χρονικής περιόδου, την παρατηρηθείσα τιμή της τρέχουσας. Όταν το MAsE υπολογισθεί

μικρότερο της μονάδας, τότε συμπεραίνουμε πως η μέθοδος που χρησιμοποιήσαμε έχει κατά μέσο όρο καλύτερη απόδοση από την μέθοδο Naïve, και αντίστροφα. Η μόνη περίπτωση κατά την οποία ο δείκτης MAsE δίνει απροσδιόριστα σφάλματα είναι όταν όλες οι τιμές των δεδομένων που διαθέτουμε είναι ίσες μεταξύ τους.

2.7 Διαγωνισμοί προβλέψεων

Η σειρά διαγωνισμών Makridakis Competitions (επίσης γνωστοί ως M-Competitions) είναι μία σειρά ανοιχτών διαγωνισμών, που αποσκοπούν στο να αξιολογήσουν και να συγκρίνουν την ακρίβεια διαφόρων μοντέλων πρόβλεψης χρονοσειρών. Οργανώνονται από ομάδες, που καθοδηγούνται από τον ερευνητή του τομέα των προβλέψεων Σπύρο Μακρινδάκη. Ο πρώτος διαγωνισμός της σειράς διεξήχθη το 1982. Το συνεχώς αυξανόμενο ενδιαφέρον για τη σειρά των διαγωνισμών, είναι εμφανές από το γεγονός ότι ενώ στον πρώτο διαγωνισμό συμμετείχαν μόνο οι Makridakis και Hibbon, στον πέμπτο διαγωνισμό (M5) συμμετείχαν συνολικά 5.507 ομάδες με 7.092 συμμετέχοντες από 101 χώρες, ενώ αντίστοιχη αύξηση παρουσίασαν τα σύνολα δεδομένων των χρονοσειρών, τα οποία από 1.001 χρονοσειρές στον M1 έφτασαν σε 42.840 χρονοσειρές στον διαγωνισμό M5.

2.7.1 Διαγωνισμός M1

Στον πρώτο διαγωνισμό που διεξήχθη το 1982 [19], το σύνολο δεδομένων αποτελούνταν από 1001 χρονοσειρές οι οποίες αφορούσαν μικροοικονομικά, μακροοικονομικά, βιομηχανικά και δημογραφικά μεγέθη. Από αυτές χρησιμοποιήθηκε ένα υποσύνολο των 111 χρονοσειρών στις περιπτώσεις όπου ήταν πολύ δύσκολο να τρέξουν όλες οι 1.001. Οι διαγωνιζόμενοι χρησιμοποίησαν 15 βασικές μεθόδους πρόβλεψης καθώς και 9 παραλλαγές.

Το βασικό συμπέρασμα που προέκυψε από την διεξαγωγή του διαγωνισμού, ήταν πως οι στατιστικά πολύπλοκες και εξεζητημένες μέθοδοι δεν παράγουν απαραίτητα και ακριβέστερες προβλέψεις σε σχέση με τις πιο απλές. Παρατηρήθηκε επίσης πως η κατάταξη της απόδοσης των διαφόρων μεθόδων διαφοροποιήθηκε ανάλογα με το κριτήριο ακρίβειας του οποίου γίνεται χρήση, ενώ ο συνδυασμός απλών μεθόδων πρόβλεψης συνήθως οδήγησε σε καλύτερη ακρίβεια σε σχέση με αυτή των επιμέρους μεθόδων που χρησιμοποιήθηκαν.

Η ακρίβεια πρόβλεψης όπως ήταν αναμενόμενο φάνηκε πως εξαρτάται από την έκταση του ορίζοντα πρόβλεψης και συγκεκριμένα όσο αυξάνεται ο ορίζοντας πρόβλεψης τόσο ελαττώνεται και η ακρίβεια των προβλέψεων.

Τα ευρήματα της μελέτης του διαγωνισμού έχουν επαληθευτεί και αναπαραχθεί μέσω της χρήσης νέων μεθόδων σε μεταγενέστερες έρευνες [20].

2.7.2 Διαγωνισμός M2

Ο Διαγωνισμός διεξήχθη το 1993 σε ευρύτερη κλίμακα σε σχέση με τον πρώτο, και μάλιστα ανοιχτή πρόσκληση δημοσιεύτηκε στο *International Journal of Forecasting*. Το σύνολο δεδομένων αποτελούνταν από 29 χρονοσειρές, στις οποίες περιλαμβάνονταν 23 σειρές από τις 4 συνεργαζόμενες εταιρείες καθώς και 6 σειρές με μακροοικονομικά δεδομένα. Δηλαδή αναλύθηκαν πολύ λιγότερες από τις 1001 χρονοσειρές του διαγωνισμού M1. Ωστόσο, η διαφοροποίηση σε σχέση με τον προηγούμενο, έγκειτο στο ότι υπήρξε συνεχής επικοινωνία και δυνατότητα των συμμετεχόντων να ζητούν δεδομένα από τις επιχειρήσεις προκειμένου να βελτιωθεί η προβλεπτική ακρίβεια των μεθόδων. Στον διαγωνισμό εφαρμόστηκαν συνολικά 16 μέθοδοι πρόβλεψης καθώς και 3 συνδυασμοί των μεθόδων αυτών.

Σκοπός του διαγωνισμού ήταν να προσομοιωθεί η διαδικασία των προβλέψεων που γίνεται σε πραγματικές συνθήκες και να εφαρμοστεί ένας συνδυασμός στατιστικών μεθόδων με κριτικές προβλέψεις που προέκυψαν από την εξωτερική πληροφόρηση κι εμπειρία των συμμετεχόντων. Οι αναλυτές μπορούσαν να βελτιώσουν τη γνώση τους υποβάλλοντας τις προβλέψεις τους για τις διαδοχικές δοκιμασίες πρόβλεψης (καθώς ο διαγωνισμός διεξήχθη σε κυλιόμενα χρονικά διαστήματα), και λαμβάνοντας μετά από κάθε περίοδο, κάποια πληροφόρηση από την κάθε εταιρία, την οποία χρησιμοποιούσαν στην επόμενη περίοδο.

Εκτός των δημοσιευθέντων αποτελεσμάτων του διαγωνισμού, πολλοί από τους συμμετέχοντες έγραψαν μικρά άρθρα όπου περιέγραφαν τις εμπειρίες που αποκόμισαν στο διαγωνισμό. Ωστόσο, παρά τα εύσημα που δέχθηκαν οι διοργανωτές για τις προσπάθειες που κατέβαλαν, υπήρξε η εντύπωση πως οι συμμετέχοντες εξακολουθούσαν να μην έχουν αρκετή εκ των έσω πρόσβαση στα δεδομένα των εταιρειών, κάτι που θα ίσχυε εν πολλοίς και εάν προσπαθούσαν να διεξάγουν προβλέψεις υπό κανονικές συνθήκες [21]. Άλλωστε, παρά τις επιπλέον πληροφορίες που δόθηκαν, οι βελτιώσεις στην ακρίβεια των προβλέψεων ήταν μικρές έως ανύπαρκτες [22].

2.7.3 Διαγωνισμός M3

Ο τρίτος διαγωνισμός της σειράς διεξήχθη το έτος 2000. Σκοπός του διαγωνισμού, ήταν κυρίως να αναπαραχθούν και να επεκταθούν τα χαρακτηριστικά των δύο προηγούμενων διαγωνισμών, με το να συμπεριληφθούν περισσότερες μέθοδοι πρόβλεψης, περισσότεροι ερευνητές (κυρίως από τον χώρο των νευρωνικών δικτύων) καθώς και περισσότερες χρονοσειρές [14]. Η δημοσίευση των αποτελεσμάτων έγινε στο *International Journal of Forecasting*.

Τα δεδομένα αποτελούνταν από 3003 χρονοσειρές που περιλάμβαναν ετήσιες, τριμηνιαίες, μηνιαίες καθώς και άλλων συχνοτήτων, αναφερόμενες σε μικροοικονομικά, μακροοικονομικά, βιομηχανικά, δημογραφικά δεδομένα κ.λπ.

Προκειμένου να εξασφαλισθεί η ύπαρξη επαρκών δεδομένων για την ανάπτυξη μοντέλων πρόβλεψης υψηλής ακρίβειας, τέθηκαν κατώτατα όρια για τον αριθμό των παρατηρήσεων που θα περιλάμβανε κάθε χρονοσειρά: 14 για τις ετήσιες, 16 για τις τριμηνιαίες, 48 για τις μηνιαίες και 60 για τις υπόλοιπες χρονοσειρές [14]. Συνολικά εφαρμόστηκαν 24 μέθοδοι πρόβλεψης.

Κατά τη διεξαγωγή του διαγωνισμού αυτού, επιβεβαιώθηκαν τα βασικά συμπεράσματα του διαγωνισμού M1 και έγιναν δημοσιεύσεις με διαφορετικές αναλύσεις επί του συνόλου δεδομένων [18]. Επίσης κατασκευάστηκε ένα μεγάλο σετ δεδομένων το οποίο αποτελεί ακόμα και σήμερα σημείο αναφοράς. Σύμφωνα με τον Rob J. Hyndman, τα δεδομένα του διαγωνισμού εξακολούθησαν να χρησιμοποιούνται από το 2000 έως και σήμερα για τον έλεγχο της ακρίβειας νέων μεθόδων, σε τέτοιο βαθμό ώστε αν μία νέα μέθοδος δεν καταστεί ανταγωνιστική στο συγκεκριμένο σετ δεδομένων, είναι δύσκολο να δημοσιευθεί στο *IJF*.

Τέλος, στον διαγωνισμό αυτό καθιερώθηκε η μέθοδος Theta ως μία νέα και ακριβής μέθοδος εκπόνησης προβλέψεων. Στον ακόλουθο πίνακα παρατίθενται συνοπτικά τα αποτελέσματα των μεθόδων που διαγωνίστηκαν.

Μέθοδος Πρόβλεψης	SMAPE των 1428 μηνιαίων χρονοσειρών	SMAPE στο σύνολο των χρονοσειρών (3003)
THETA	13.85	13.01
ForecastPro	13.86	13.19
ForcX	14.45	13.49
COMB S-H-D	14.48	13.52
DAMPEN	14.59	13.63
THETAsm	14.66	13.88
RBF	14.77	13.75
B-J automatic	14.81	14.01
AutomatANN	14.93	14.11
SMARTFCS	15.03	14.13

Εικόνα 4. Τα αποτελέσματα των μεθόδων του διαγωνισμού M3 σύμφωνα με τη μετρική sMAPE

2.7.4 Διαγωνισμός M4

Ο τέταρτος διαγωνισμός της σειράς (M4) ανακοινώθηκε το Νοέμβριο του 2017 και διεξήχθη καθ' όλη τη διάρκεια του 2018 [23].

Ο διαγωνισμός επέκτεινε και αναπαρήγαγε τα αποτελέσματα των προηγούμενων τριών διαγωνισμών, χρησιμοποιώντας ένα εκτεταμένο και χαρακτηριζόμενο από ποικιλία σύνολο χρονοσειρών. Δόθηκε έτσι η ευκαιρία να συγκριθεί η ακρίβεια που παρουσιάζουν διαφορετικές μέθοδοι πρόβλεψης και κυρίως να συγκριθούν οι επιδόσεις μεταξύ μηχανικής μάθησης και στατιστικής. Η σύγκριση αυτή έγινε με τη χρήση απλών μεθόδων αναφοράς, ενώ το μεγάλο πλήθος των δεδομένων αξιοποιήθηκε προκειμένου να εξακριβωθεί η στατιστική σημαντικότητα των αποτελεσμάτων.

- **Σύνολο δεδομένων**

Το σύνολο δεδομένων, όπως προαναφέρθηκε περιλάμβανε 100.000 χρονολογικές σειρές, οι οποίες υποδιαιρούνται σε 6 κατηγορίες χρονικής συχνότητας, και σε 6 βασικούς τομείς.

Οι 100.000 χρονοσειρές αντλήθηκαν από μία βάση δεδομένων ονόματι Fore-DeCk, η οποία δημιουργήθηκε στο Εθνικό Μετσόβιο Πολυτεχνείο και περιλαμβάνει 900.000 σειρές συνεχούς χρόνου, που αντλήθηκαν από πολλές διαφορετικές και δωρεάν προσβάσιμες πηγές.

Η βάση αυτή εστιάζει σε εφαρμογές επιχειρησιακών προβλέψεων, περιλαμβάνοντας σειρές από διαφορετικούς τομείς της οικονομίας, όπως βιομηχανία, υπηρεσίες, τουρισμός, εισαγωγές εξαγωγές κ.λπ. [24].

- **Προεπεξεργασία**

Η επιλογή των χρονοσειρών έγινε με τυχαία δειγματοληψία από τη βάση δεδομένων. Κατόπιν, έγινε προσαρμογή τους προκειμένου να εξαλειφθούν οι αρνητικές τιμές, καθώς και αυτές που ήταν μικρότερες του 10, προκειμένου να αποφευχθούν πιθανά προβλήματα κατά τον υπολογισμό των μετρικών απόδοσης. Η προσαρμογή έγκειτο στην απλή πρόσθεση μίας σταθεράς στις τιμές τους, έτσι ώστε η ελάχιστη τιμή να είναι το 10 [23].

- **Καθιέρωση μεθόδων αναφοράς**

Μία καινοτομία του διαγωνισμού ήταν η καθιέρωση διαφόρων μεθόδων αναφοράς, οι οποίες αντλούν από τη στατιστική και τη μηχανική μάθηση. Έτσι λοιπόν, για να αξιολογηθούν οι βελτιωμένες επιδόσεις των υποβολών του M4 έναντι των απλών τυποποιημένων μεθόδων, και για να εξακριβωθούν οι λόγοι που συντέλεσαν στις βελτιωμένες αυτές επιδόσεις, αποφασίστηκε να συμπεριληφθούν 10 διαφορετικές μέθοδοι αναφοράς, με διαφορετικά χαρακτηριστικά η κάθε μία. Για παράδειγμα, η μέθοδος Naïve 2 μπορεί να εντοπίζει μόνο τις εποχικότητες, η Single Exponential Smoothing (DSES) εντοπίζει το επίπεδο, η Holt προβλέπει κάνοντας χρήση γραμμικής τάσης κ.ο.κ [23].

- **Αριθμός παρατηρήσεων και ορίζοντας πρόβλεψης**

Ο αριθμός των ελάχιστων παρατηρήσεων που περιείχαν οι χρονοσειρές του συνόλου δεδομένων, ήταν διαφορετικός για κάθε κατηγορία χρονικής συχνότητας, και συγκεκριμένα, ήταν (μαζί με τον αριθμό των απαιτούμενων προβλέψεων):

- 13 και 6 για τις ετήσιες,
- 16 και 8 για τις τριμηνιαίες,
- 42 και 18 για τις μηνιαίες,
- 80 και 13 για τις εβδομαδιαίες
- 93 και 48 για τις ημερήσιες.
- 700 και 48 για τις ωριαίες.

Οι σειρές περιείχαν κατά μέσο όρο πολύ περισσότερες παρατηρήσεις σε σχέση με τον διαγωνισμό M3, προκειμένου να δοθεί στους συμμετέχοντες η ευκαιρία να δοκιμάσουν περίπλοκες μεθόδους οι οποίες απαιτούν την ύπαρξη μεγάλου όγκου δεδομένων για την κατάλληλη εκπαίδευση. Οι ορίζοντες πρόβλεψης καθορίστηκαν λαμβάνοντας υπ' όψη τη φύση των αποφάσεων στις οποίες συνήθως συμμετέχουν οι χρονοσειρές βάσει της χρονικής τους συχνότητας [24].

- **Συμπεράσματα**

Τα κυριότερα συμπεράσματα που προέκυψαν από τη διεξαγωγή του διαγωνισμού συνοψίζονται στα παρακάτω [24]:

- Ο συνδυασμός μεθόδων πρόβλεψης συνεχίζει να είναι η αποτελεσματικότερη στρατηγική παραγωγής προβλέψεων υψηλής ακρίβειας. Συγκεκριμένα, εκ των 17 επικρατέστερων μεθόδων, οι 12 ήταν συνδυασμοί ως επί το πλείστον στατιστικών μεθόδων.
- Η χρήση «υβριδικών μεθόδων» που συνδυάζουν μεθόδους Μηχανικής Μάθησης με στατιστικές, απέδωσε βέλτιστα. Τη μεγάλη έκπληξη του διαγωνισμού, αποτέλεσε η προσέγγιση που υπέβαλλε ο Slawlek Smyl, επιστήμονας δεδομένων στην Uber Technologies. Σύμφωνα με το μέτρο αξιολόγησης sMAPE, η μέθοδος παρουσίασε μία βελτίωση της τάξεως του 10% σε σχέση με τον συνδυασμό μεθόδων «ορόσημα» (Comb) του διαγωνισμού.
- Το σύνολο των μεθόδων υποεκτιμά την αβεβαιότητα του μέλλοντος. Ωστόσο, υπάρχουν κάποιες που την εκτιμούν με υψηλή ακρίβεια.
- Πιο πολύπλοκες μέθοδοι ενδέχεται να οδηγούν σε υψηλότερη ακρίβεια. Οι τρεις προηγούμενοι διαγωνισμοί, πιθανότατα και άλλες μελέτες [25] [26] κατέληξαν στο ότι η χρήση πιο πολύπλοκων μεθόδων και ως εκ τούτου και η δαπάνη περισσότερου υπολογιστικού χρόνου δεν ανταμείφτηκε με αντίστοιχη βελτίωση της ακρίβειας των αποτελεσμάτων, γεγονός που αντιστράφηκε στον M4.
- Η αξιοποίηση πληροφοριών που προέρχονται από πολλές χρονοσειρές (cross learning), οδηγεί σε καλύτερα αποτελέσματα σε σχέση με αυτά που παράγονται από προβλέψεις βασισμένες σε μία μόνο χρονοσειρά.

- Οι μέθοδοι πρόβλεψης που βασίζονται αποκλειστικά σε αλγορίθμους Μηχανικής Μάθησης δεν παράγουν ακριβείς προβλέψεις.

2.7.5 Διαγωνισμός M5

Ο τελευταίος ολοκληρωμένος διαγωνισμός της σειράς M [27] διεξήχθη το 2020. Ως dataset, χρησιμοποιήθηκαν πραγματικά δεδομένα που αντλήθηκαν από τη Walmart και αφορούσαν 3,049 διαφορετικούς κωδικούς προϊόντων. Ο διαγωνισμός διεξήχθη στην δικτυακή πλατφόρμα Kaggle [28] και προσέφερε σημαντικά βραβεία της τάξεως των 100.000\$ στους νικητές.

Συνολικά δόθηκαν προς πρόβλεψη 42.840 ημερήσιες χρονοσειρές, ιεραρχικά δομημένες, που ξεκινούσαν από το επίπεδο της αποθήκης (SKU) και κλιμακώνονταν έως και την συνολική ζήτηση σε μία μεγάλη γεωγραφική περιοχή. Συμμετείχαν μέθοδοι στατιστικές, μηχανικής μάθησης και συνδυασμός αυτών .

Πέρα από τα δεδομένα πωλήσεων, υπήρχαν επίσης εξωγενείς μεταβλητές σχετικά με τις τιμές, προωθητικές ενέργειες, στοκ στις αποθήκες, την ημέρας της εβδομάδας στην οποία αναφέρονται τα δεδομένα, γιορτές και αργίες.

Οι διαγωνιζόμενοι υπέβαλλαν τόσο σημειακές προβλέψεις όσο και πιθανοτικές (50%, 67%, 95%, 99% και median). Δόθηκε έμφαση σε δεδομένα διακοπτόμενης ζήτησης. Οι συμμετέχοντες ήταν σε μεγαλύτερο βαθμό επαγγελματίες του χώρου και λιγότερο ακαδημαϊκοί.

Βασικά συμπεράσματα που προέκυψαν από την αξιολόγηση του διαγωνισμού [29] ήταν τα ακόλουθα:

- Οι μέθοδοι μηχανικής μάθησης κατάφεραν να ξεπεράσουν τις παραδοσιακές στατιστικές μεθόδους. Συγκεκριμένα όλες οι κορυφαίες μέθοδοι που υποβλήθηκαν χρησιμοποίησαν καθαρά μηχανική μάθηση.
- Ο συνδυασμός μεθόδων ακόμα και απλών οδήγησε σε βελτίωση της ακρίβειας των προβλέψεων

- Η αξία του cross learning, δηλαδή η αξιοποίηση του γεγονότος ότι οι χρονοσειρές του διαγωνισμού ήταν συσχετισμένες.
- Ενώ στον διαγωνισμό M2 οι εξωγενείς πληροφορίες δεν βελτίωσαν σημαντικά την ακρίβεια των προβλέψεων, αντίθετα στον M5 η 2^η και 5^η σε ακρίβεια μέθοδοι χρησιμοποίησαν τέτοιου είδους προσαρμογές για την ενίσχυσή τους.
- Η προστιθέμενη αξία των στρατηγικών επικύρωσης (Cross Validation). Συγκεκριμένα οι 4 κορυφαίες υποβολές του διαγωνισμού και η πλειονότητα των 50 κορυφαίων υποβολών, χρησιμοποίησαν τέτοιες στρατηγικές.
- Η σημασία των εξωγενών επεξηγηματικών μεταβλητών. Γενικά οι μέθοδοι που λαμβάνουν αποκλειστικά υπ' όψη τις ιστορικές παρατηρήσεις, αποτυγχάνουν να αποτυπώσουν την επίδραση αργιών, ειδικών γεγονότων, εκπτώσεων, τιμών και πιθανόν και του καιρού. Στον διαγωνισμό όλες οι νικητήριες μέθοδοι αξιοποίησαν την σχετική εξωγενή πληροφορία.

2.7.6 Διαγωνισμός M6

Ο διαγωνισμός M6 διεξάγεται κατά τη διάρκεια του 2022 και έως τις αρχές του 2023. Θα είναι παρόμοιος με τους προηγούμενους 5 διαγωνισμούς, ενώ εστιάζει στην πρόβλεψη τιμών μετοχών (αποδόσεις) και κινδύνου. Η καινοτομία του συγκεκριμένου διαγωνισμού έγκειται στο γεγονός πως εστιάζει εξ' ίσου στην λήψη επενδυτικών αποφάσεων που βασίζονται στις προαναφερθείσες προβλέψεις. Από τους συμμετέχοντες ζητείται να διεξάγουν προβλέψεις και να λάβουν επενδυτικές αποφάσεις από ένα σύμπαν 50 S&P500 μετοχών, και 50 διεθνών διαπραγματεύσιμων αμοιβαίων κεφαλαίων (ETFs), που καλύπτουν μία ευρεία γκάμα χωρών και κεφαλαιουχικών στοιχείων (assets).

Βασικό σκεπτικό σχεδίασης της δομής του διαγωνισμού, αποτέλεσε το εξής παράδοξο:

Σύμφωνα με την “υπόθεση της αποδοτικής αγοράς” (EMH), οι τιμές των μετοχών, αντανακλούν όλη την υπάρχουσα σχετική πληροφόρηση, κάτι που συνεπάγεται ότι δεν είναι εφικτή η κατ' εξακολούθηση “επικράτηση” των επενδυτών επί της αγοράς. Από την άλλη πλευρά, θρυλικοί επενδυτές όπως ο Warren Buffet, καθώς και εταιρείες όπως η Blackstone, έχουν επιτύχει αξιοθαύμαστα αποτελέσματα επί μακρών χρονικών διαστημάτων και έχουν συσσωρεύσει κέρδη που είναι αδύνατο να αποδοθούν σε απλή τύχη. Το γεγονός αυτό θέτει την αξιοπιστία της EMH υπό τη σκιά της αμφιβολίας [30].

Ο βασικός στόχος του Μ6 διαγωνισμού, είναι να μελετηθεί το παραπάνω παράδοξο και να εξεταστεί η ΕΜΗ υπό ένα διαφορετικό πρίσμα.

Σκοπός είναι λοιπόν, μέσω του διαγωνισμού Μ6, να εξακριβωθεί εάν αποδόσεις μεγαλύτερες του μέσου όρου μπορούν να επιτευχθούν, από έναν παράγοντα ή από συνδυασμό των ακόλουθων παραγόντων.

- Την δυνατότητα ακριβούς πρόβλεψης των συνολικών αποδόσεων της αγοράς, ή μεμονωμένων μετοχών/ αμοιβαίων κεφαλαίων.
- Την δυνατότητα κατάλληλης δημιουργίας μοντέλων αβεβαιότητας και κινδύνου του συνόλου της αγοράς ή μεμονωμένων μετοχών/αμοιβαίων κεφαλαίων.
- Τη δυνατότητα να συνδυαστούν οι προβλέψεις ακρίβειας και αβεβαιότητας με επενδυτικές αποφάσεις.
- Τη δυνατότητα χρήσης κριτικής ικανότητας κατά την παραγωγή προβλέψεων και επενδυτικών αποφάσεων, με σκοπό την “επικράτηση” επί της αγοράς.
- Τη σημασία που έχει μία σταθερή επενδυτική στρατηγική.
- Τη σημασία που έχουν άλλοι παράγοντες, όπως οι κριτικές προβλέψεις ή αυτές που βασίζονται σε μοντέλα, αλλά και οι προκαταλήψεις κατά τη λήψη επενδυτικών αποφάσεων και άλλες αδυναμίες, οι οποίες μπορούν να τύχουν εκμετάλλευσης προκειμένου να επιτευχθούν άνω του μέσου όρου αποδόσεις.

3 Τεχνητή Νοημοσύνη – Μηχανική Μάθηση

Στο κεφάλαιο αυτό θα περιγράψουμε, τον δεύτερο μεγάλο τομέα πέραν της στατιστικής, από τον οποίο μπορούν να αντληθούν μεθοδολογίες χρήσιμες στον τομέα των προβλέψεων. Ο τομέας αυτός είναι η Μηχανική Μάθηση που αποτελεί παρακλάδι της τεχνητής νοημοσύνης, και ο οποίος όπως είδαμε και στην ενότητα των διαγωνισμών M, τα τελευταία χρόνια έχει αποκτήσει κυρίαρχο ρόλο στους διαγωνισμούς αλλά και στο χώρο των προβλέψεων γενικά.

3.1 Τεχνητή νοημοσύνη

Ο όρος τεχνητή νοημοσύνη (AI) έχει εμπνεύσει εδώ και πολλά χρόνια τη φαντασία του ανθρώπου έχοντας σαν απότοκο την δημιουργία πολλών βιβλίων και ταινιών επιστημονικής φαντασίας, σε πολλά μάλιστα από τα οποία, οι μηχανές καταφέρνουν με κάποιο τρόπο να αυτονομηθούν από τον άνθρωπο και να ενεργήσουν με σκοπό την επικράτηση και κυριαρχία τους επί του ανθρώπινου είδους.

Οι ορισμοί που έχουν δοθεί ποικίλλουν, ωστόσο σύμφωνα με τους Russel & Norvig [31], όλοι οι ορισμοί περιλαμβάνουν μηχανές/υπολογιστές που έχουν τις ακόλουθες ιδιότητες: Είτε ενεργούν λογικά ή/και ανθρώπινα είτε σκέφτονται λογικά ή/και ανθρώπινα. Ένας άλλος σύντομος ορισμός που δίνει ο Francois Chaullet [32], είναι: «Η τεχνητή νοημοσύνη συνίσταται στην προσπάθεια αυτοματοποίησης πνευματικών εργασιών που υπό κανονικές συνθήκες διεκπεραιώνουν άνθρωποι».

Αποτελεί έναν από τους πιο σύγχρονους επιστημονικούς τομείς. Ξεκίνησε να αναπτύσσεται μετά τον 2^ο παγκόσμιο πόλεμο και το όνομα «Τεχνητή Νοημοσύνη» καθιερώθηκε το 1956. Περιλαμβάνει δε, μία τεράστια ποικιλία επιμέρους πεδίων, που εκτείνονται από πιο γενικά (μάθηση και αντίληψη) σε πιο ειδικά [31], όπως τα παιχνίδια, η απόδειξη μαθηματικών θεωρημάτων, η αυτόματη οδήγηση οχημάτων και η διάγνωση ασθενειών. Επόμενο ήταν, πως και ο χώρος των προβλέψεων θα επιχειρούσε να επωφεληθεί από την ανάπτυξη του τομέα αυτού, επιχειρώντας τη δημιουργία μοντέλων πρόβλεψης που θα υπερτερούν των «παραδοσιακών» στατιστικών μεθόδων.

Από το 1956 που θεμελιώθηκε η τεχνητή νοημοσύνη ως επιστημονικό πεδίο, έως και σήμερα, η πορεία της κάθε άλλο παρά γραμμική μπορεί να χαρακτηριστεί. Αντιθέτως, χαρακτηρίζεται από κύματα αισιοδοξίας και πίστης στις προοπτικές του τομέα, που ακολουθούνται από

περιόδους απογοήτευσης και οικονομικής στενότητας λόγω της έλλειψης χρηματοδότησης, τις οποίες διαδέχονται ξανά περιόδους αισιοδοξίας λόγω της εύρεσης νέων μεθόδων και τεχνολογικών ανακαλύψεων [31] [33].

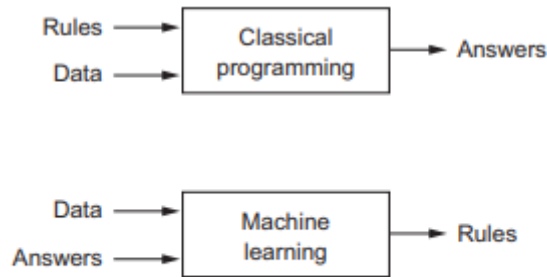
Στα πλαίσια της τεχνητής νοημοσύνης, έχουν δοκιμαστεί πολλές διαφορετικές προσεγγίσεις, συμπεριλαμβανομένης της εξομοίωσης του ανθρώπινου εγκεφάλου, της μοντελοποίησης του ανθρώπινου τρόπου εύρεσης λύσης σε προβλήματα, τυπικής λογικής, μεγάλων βάσεων δεδομένων, συστημάτων γνώσης κ.λπ. Κατά τη διάρκεια του 21^{ου} αιώνα, παρατηρούμε την επικράτηση του επιμέρους πεδίου της μηχανικής μάθησης, το οποίο βασίζεται σε μεγάλο βαθμό σε μαθηματικές και στατιστικές μεθόδους και αποδείχθηκε αρκετά αποτελεσματικό στην επίλυση αρκετών δύσκολων προβλημάτων είτε σε ακαδημαϊκά, είτε σε βιομηχανικά πλαίσια [31].

3.2 Μηχανική Μάθηση

Όπως αναφέρθηκε και στην προηγούμενη ενότητα η μηχανική μάθηση αποτελεί επί μέρους πεδίο του ευρύτερου τομέα της τεχνητής νοημοσύνης. Ουσιαστικά περιλαμβάνει τη μελέτη υπολογιστικών αλγορίθμων οι οποίοι έχουν τη δυνατότητα να βελτιώνονται αυτόματα μέσω της εμπειρίας και της χρήσης διαθέσιμων δεδομένων [34].

Η μηχανική μάθηση πηγάζει από το εξής ερώτημα που θέτει ο Francois Chaullet [32]: «θα μπορούσε ένας υπολογιστής να προχωρήσει ένα βήμα παραπέρα από τον γνωστό τρόπο που γνωρίζουμε ότι λειτουργεί και να μάθει μόνος του τον τρόπο να επιτελεί μια συγκεκριμένη εργασία; Θα μπορούσε δηλαδή, αντί να δέχεται και να εκτελεί κανόνες που του δίνονται με προστακτικό τρόπο από τους προγραμματιστές, να «μάθει» μόνος του τους κανόνες αυτούς, εξετάζοντας τα διαθέσιμα δεδομένα;»

Το ερώτημα αυτό δημιουργεί τους ορίζοντες για ένα εντελώς καινούριο προγραμματιστικό παράδειγμα. Στον κλασικό προγραμματισμό, ο άνθρωπος τροφοδοτεί τον υπολογιστή με κανόνες και δεδομένα, και αναμένει από αυτόν να επεξεργαστεί τα δεδομένα αυτά σύμφωνα με τους κανόνες, ώστε να προκύψουν οι ζητούμενες απαντήσεις. Αντιθέτως στη μηχανική μάθηση, ο υπολογιστής τροφοδοτείται με δεδομένα αλλά και με τις απαντήσεις-συμπεράσματα που αναμένονται από τα δεδομένα, και μέσω αυτών εξάγει τους κανόνες [32]. Οι κανόνες αυτοί μπορούν έπειτα να χρησιμοποιηθούν για την εξαγωγή απαντήσεων από νέα δεδομένα που θα προκύψουν.



Εικόνα 5. Το παράδειγμα της μηχανικής μάθησης σε σχέση με τον προστακτικό προγραμματισμό [32].

Οι αλγόριθμοι της μηχανικής μάθησης, χρησιμοποιούνται πλέον σε μία ευρεία γκάμα εφαρμογών, όπως αναφέρθηκε και στην προηγούμενη ενότητα. Στις περισσότερες από τις εφαρμογές αυτές, θα ήταν πολύ δύσκολο έως και αδύνατο να αναπτυχθούν αποτελεσματικοί συμβατικοί αλγόριθμοι [35].

Η μηχανική μάθηση, χωρίζεται σε δύο μεγάλες κατηγορίες. Την επιβλεπόμενη και την μη επιβλεπόμενη.

3.3 Επιβλεπόμενη μάθηση

Στην επιβλεπόμενη μάθηση, κάνοντας χρήση ιστορικών δεδομένων, και γνωρίζοντας τα ζητούμενα αποτελέσματα, αποσκοπούμε στο να συσχετίσουμε τα δεδομένα της εισόδου με κάποιο επιθυμητό αποτέλεσμα το οποίο αποτελεί και την έξοδο του μοντέλου που θα δημιουργήσουμε. Επί της ουσίας η επιβλεπόμενη μάθηση έγκειται στην διαδικασία «εκμάθησης» μίας συνάρτησης που αντιστοιχίζει την είσοδο σε μία έξοδο, με βάση τα ήδη υπάρχοντα δείγματα εισόδου - εξόδου [31].

Ένας αλγόριθμος επιβλεπόμενης μάθησης λοιπόν, μέσω της ανάλυσης των δεδομένων εκπαίδευσης (όπως αποκαλούνται τα δεδομένα για τα οποία διαθέτουμε την ζητούμενη τιμή εξόδου), συμπεραίνει (infer) μία συνάρτηση η οποία κατόπιν χρησιμοποιείται για την εφαρμογή της σε δεδομένα των οποίων την έξοδο δεν γνωρίζουμε και στόχος είναι η καλύτερη δυνατή προσέγγισή αυτής της εξόδου από τη συνάρτηση.

Τα μοντέλα της επιβλεπόμενης μάθησης συνήθως επιλύουν δύο βασικές κατηγορίες προβλημάτων:

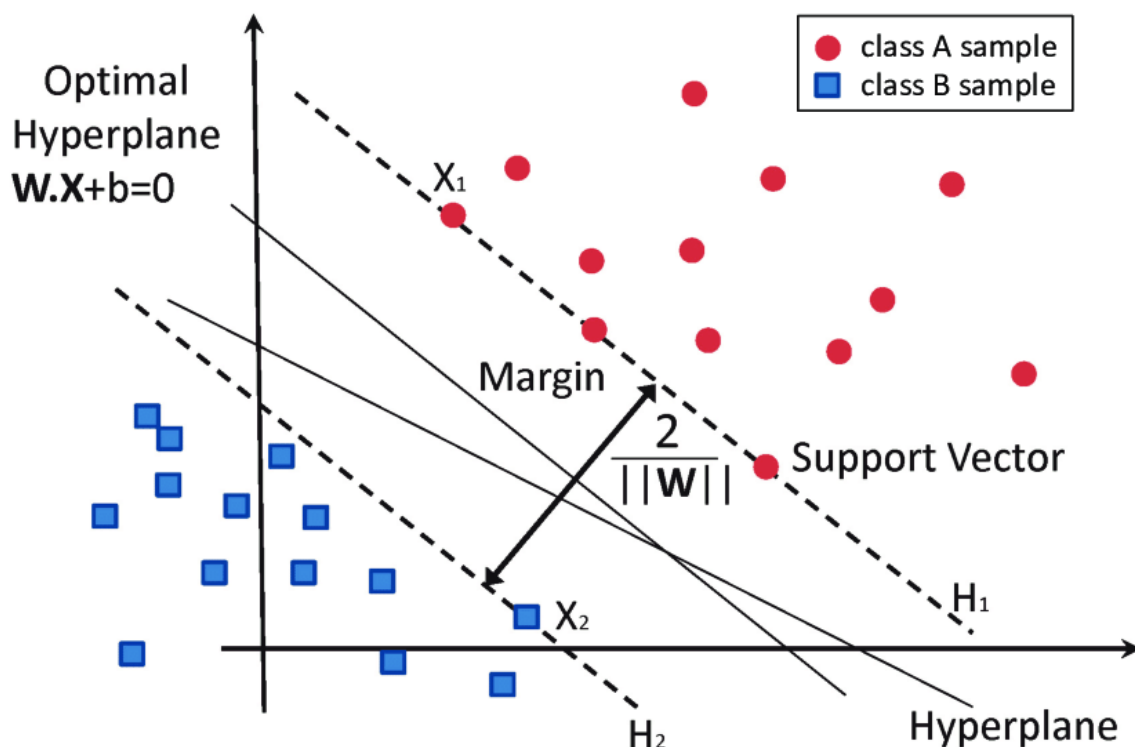
Κατηγοριοποίηση (Classification), δηλαδή ταξινόμηση των δειγμάτων της εισόδου σε δύο ή περισσότερες κλάσεις. Παράδειγμα προβλήματος κατηγοριοποίησης είναι ο διαχωρισμός των εισερχόμενων e-mails σε ανεπιθύμητα (spam) και σε μη ανεπιθύμητα (not spam).

Παλινδρόμηση (Regression), δηλαδή εκτίμηση της τιμής μίας ή περισσότερων εξαρτημένων μεταβλητών. Παράδειγμα προβλήματος παλινδρόμησης είναι η εκτίμηση των μελλοντικών τιμών μίας χρονοσειράς.

Μία πλειάδα αλγορίθμων εμπίπτουν στην κατηγορία της επιβλεπόμενης μάθησης και θα αναφέρουμε ενδεικτικά μερικούς:

3.3.1 Μηχανές διανυσμάτων υποστήριξης (Support Vector Machines)

Η μέθοδος αυτή συνίσταται στην κατασκευή ενός υπέρ-επιπέδου (hyperplane) στον χώρο των δεδομένων, το οποίο διαχωρίζει βέλτιστα τις διαφορετικές κλάσεις δεδομένων όπως φαίνεται στο ακόλουθο σχήμα [36], όπου σημεία δύο διαστάσεων κατηγοριοποιούνται με τον προσδιορισμό μίας ευθείας γραμμής (βέλτιστο υπέρ-επίπεδο).



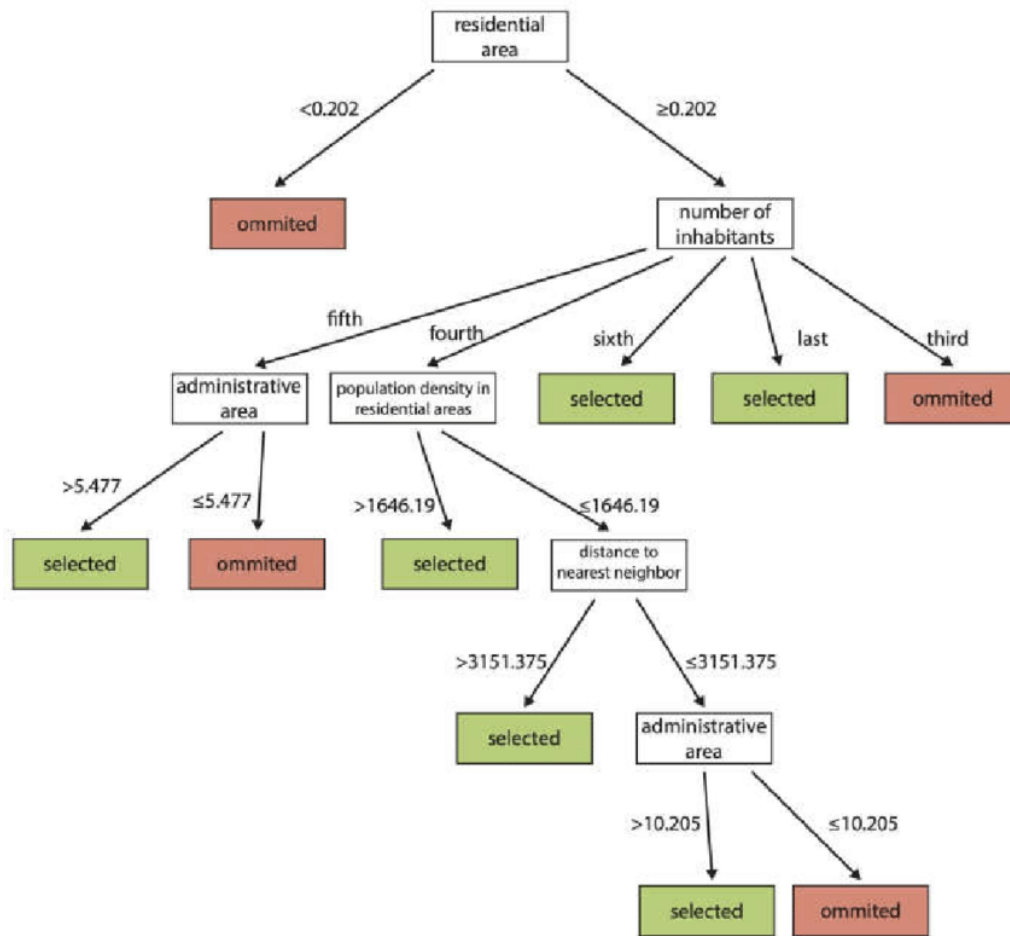
Εικόνα 6. Κατηγοριοποίηση δεδομένων από Μηχανές Διανυσμάτων Υποστήριξης [36]

Η μέθοδος αυτή έχει το πλεονέκτημα ότι μπορεί να προσεγγίζει προβλήματα κατηγοριοποίησης, έχοντας ανάγκη από πολύ λιγότερο υπολογιστικό χρόνο σε σχέση με τα νευρωνικά δίκτυα, ενώ χρησιμοποιείται και σε προβλήματα παλινδρόμησης. Οι μηχανές διανυσμάτων υποστήριξης έχουν χρησιμοποιηθεί ευρύτατα σε προβλήματα επεξεργασίας φυσικής γλώσσας, κατηγοριοποίησης κειμένου [37] [38], αναγνώρισης εικόνας, ενώ έχει επιχειρηθεί και η χρήση τους στον τομέα της πρόβλεψης χρονοσειρών [39].

3.3.2 Δένδρα απόφασης

Τα δένδρα απόφασης, παράγουν προβλέψεις κατασκευάζοντας απλούς κανόνες απόφασης με βάση τα διαθέσιμα δεδομένα, μπορούν να χρησιμοποιηθούν σε προβλήματα κατηγοριοποίησης, όπως επίσης και σε προβλήματα παλινδρόμησης.

Τα δένδρα αυτά χρησιμοποιούνται κυρίως σαν εργαλεία υποστήριξης αποφάσεων προκειμένου να προσδιοριστεί η στρατηγική που είναι πιο πιθανό να οδηγήσει στον επιθυμητό στόχο, αποτελούν όμως κι ένα δημοφιλές εργαλείο στον τομέα της μηχανικής μάθησης.

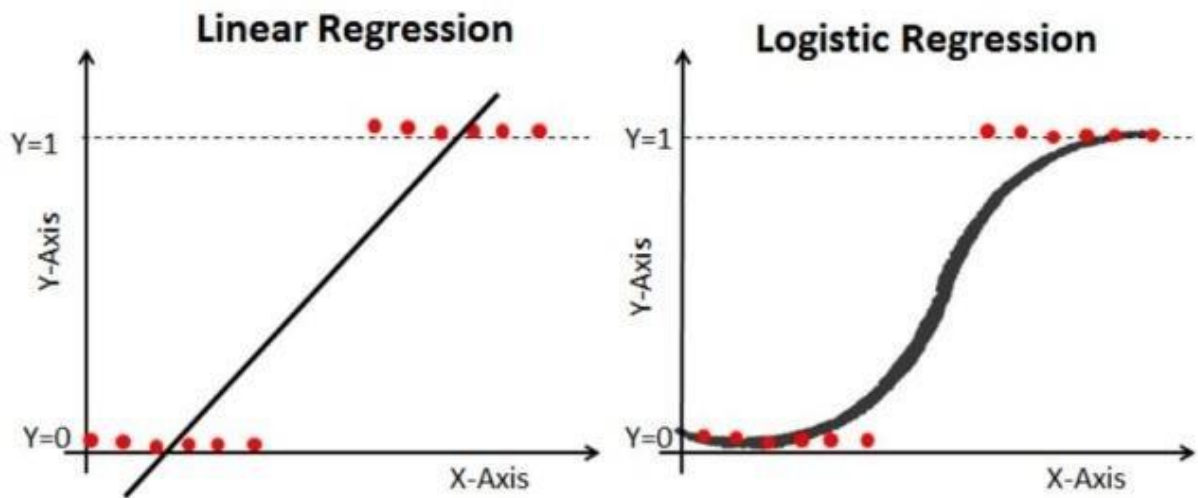


Εικόνα 7. Εφαρμογή δένδρων απόφασης σε πρόβλημα χαρτογραφικής επιλογής (selection) [40]

Στην παραπάνω εικόνα βλέπουμε ένα δένδρο απόφασης προκειμένου να ληφθεί η απόφαση επιλογής της απεικόνισης ή όχι, στοιχείων σε χάρτες μικρής κλίμακας [40].

3.3.3 Τεχνικές παλινδρόμησης

Σε προηγούμενη ενότητα αναφερθήκαμε στην απλή γραμμική παλινδρόμηση, με τη χρήση της οποίας μπορούμε να προσεγγίσουμε και να προβλέψουμε τις μελλοντικές τιμές μίας χρονοσειράς μέσω μίας ευθείας γραμμής αλλά και στην πολλαπλή γραμμική παλινδρόμηση, η οποία δημιουργεί μία γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών.



Εικόνα 8. Χρήση λογιστικής και γραμμικής παλινδρόμησης σε πρόβλημα δυαδικής κατηγοριοποίησης [41].

Μία ακόμα μορφή παλινδρόμησης που χρησιμοποιείται συχνά στα πλαίσια της μηχανικής μάθησης είναι η λογιστική παλινδρόμηση, η οποία χρησιμοποιεί την λογιστική καμπύλη και χρησιμοποιείται συνήθως σε δυαδικά προβλήματα κατηγοριοποίησης. Στην λογιστική παλινδρόμηση, αντί να μοντελοποιούμε την τιμή της εξαρτημένης μεταβλητής Y , μοντελοποιούμε την πιθανότητα να ανήκει το Y σε μία συγκεκριμένη κατηγορία [42]. Η χρήση της μπορεί ωστόσο να επεκταθεί σε προβλήματα κατηγοριοποίησης, πέραν των δύο κατηγοριών.

Στην εικόνα 8 παρατηρούμε τη μοντελοποίηση της πιθανότητας μιας εξαρτημένης μεταβλητής με τη χρήση γραμμικής παλινδρόμησης στο αριστερό σχήμα και λογιστικής παλινδρόμησης στο δεξί σχήμα.

$$p(X) = \Pr(Y = 1|X)$$

Η σχέση μεταξύ $p(X)$ και X , στην περίπτωση της γραμμικής παλινδρόμησης αναπαρίσταται ως εξής:

$$p(X) = \beta_0 + \beta_1 X$$

Εάν χρησιμοποιηθεί αυτή η προσέγγιση για να προβλεφθεί η πιθανότητα, υπάρχει το εμφανές πρόβλημα πως σε περιοχές κοντά στο μηδέν η πιθανότητα παίρνει αρνητικές τιμές, κάτι που φυσικά δεν μπορεί να υφίσταται καθώς η πιθανότητες παίρνουν τιμές από 0 έως 1 [42].

Για να αποφευχθεί αυτό το πρόβλημα, πρέπει να μοντελοποιήσουμε την $p(X)$ με μία συνάρτηση της οποίας το πεδίο τιμών είναι μεταξύ του 0 και 1 για όλες τις δυνατές τιμές του X . Μία από αυτές τις τιμές είναι και η λογιστική συνάρτηση που περιγράφεται από την ακόλουθη εξίσωση [42].

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

Η εύρεση των συντελεστών γίνεται με τη μέθοδο της μέγιστης πιθανοφάνειας.

3.3.4 Νευρωνικά δίκτυα.

Τα νευρωνικά δίκτυα είναι αλγόριθμοι που επιχειρούν να προσομοιώσουν, τον τρόπο λειτουργίας των βιολογικών νευρωνικών δικτύων που απαρτίζουν τους εγκεφάλους των ζώωντων οργανισμών και καθώς αποτελούν το κυρίως αντικείμενο της παρούσας διπλωματικής θα παρουσιαστούν αναλυτικά σε επόμενη ενότητα.

3.4 Μη επιβλεπόμενη μάθηση

Εν αντιθέσει με την επιβλεπόμενη μάθηση όπου όπως αναφέρθηκε διαθέτουμε ένα σύνολο από δείγματα, καθώς και την απάντηση – πρόβλεψη που αντιστοιχεί σε αυτά, στην μη επιβλεπόμενη μάθηση, επιχειρούμε την επίλυση του προβλήματος στο οποίο δεν υπάρχουν απαντήσεις – προβλέψεις, για τα δείγματα του συνόλου δεδομένων. Σκοπός του προβλήματος είναι αντ' αυτού, η ανακάλυψη μοτίβων και συμπερασμάτων που αφορούν το σύνολο των δειγμάτων που διαθέτουμε. Οι δύο κυριότερες μορφές μη επιβλεπόμενης μάθησης είναι η ανάλυση σε κύριες συνιστώσες (Principal Component Analysis) και η συσταδοποίηση (clustering).

Η μη επιβλεπόμενη μάθηση εμφανίζει συνήθως πολύ περισσότερες προκλήσεις στον ερευνητή σε σχέση με την επιβλεπόμενη, κάτι που γίνεται εμφανές από το γεγονός ότι δεν υπάρχει κάποιος συγκεκριμένος στόχος στην ανάλυσή, όπως είναι μια πρόβλεψη ή μια

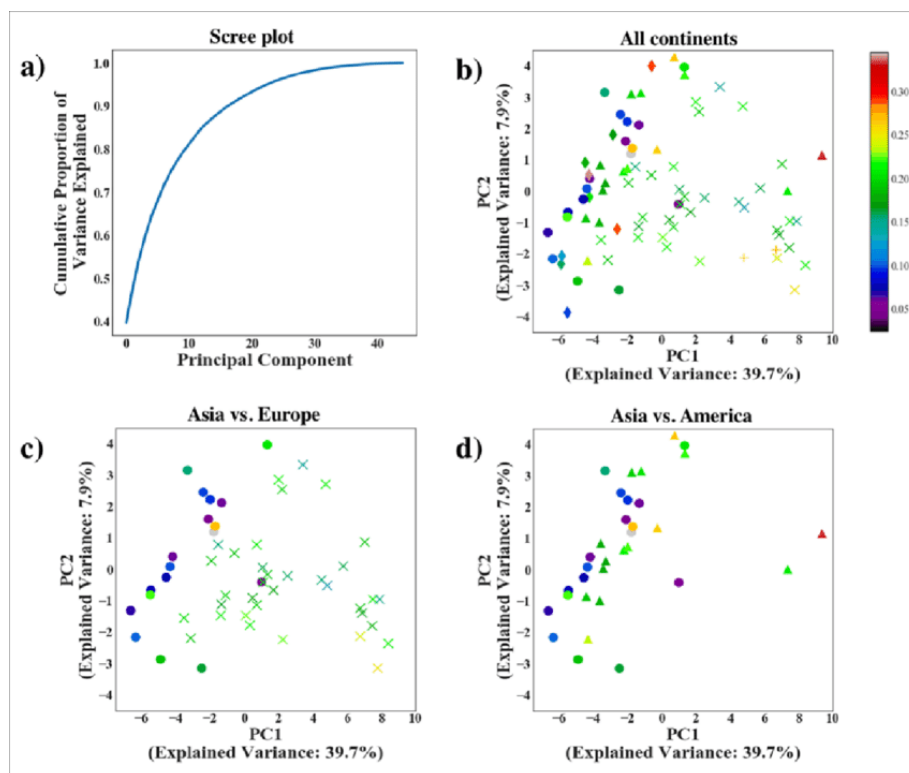
απάντηση [42] και συχνά χρησιμοποιείται στα προκαταρκτικά στάδια της επιβλεπόμενης μάθησης.

3.4.1 Ανάλυση σε κύριες συνιστώσες (Principal Component Analysis).

Η μέθοδος αυτή χρησιμοποιείται όταν υπάρχει η ανάγκη για οπτική αναπαράσταση των δεδομένων, καθώς και μείωση της διάστασης των χαρακτηριστικών, προτού εφαρμοστούν οι τεχνικές επιβλεπόμενης μάθησης.

Όταν διαθέτουμε ένα σύνολο δεδομένων με μεγάλη διάσταση (μεγάλος αριθμός features), με μεταβλητές που συσχετίζονται μεταξύ τους, οι κύριες συνιστώσες μας βοηθούν στο να τις συμπτύξουμε, σε έναν μικρότερο αριθμό μεταβλητών, οι οποίες συλλογικά μπορούν να ερμηνεύσουν το μεγαλύτερο μέρος της μεταβλητότητας (variability) του αρχικού συνόλου [42].

Η ανάλυση σε κύριες συνιστώσες αναφέρεται στη διαδικασία με την οποία υπολογίζονται οι κύριες συνιστώσες, καθώς και στην μετέπειτα χρήση των συνιστωσών αυτών, προκειμένου να αποκτηθεί μία καλύτερη κατανόηση των δεδομένων [42].



Εικόνα 9. Ανάλυση κύριων συνιστωσών σε δεδομένα πωλήσεων φαγητού.

Στην εικόνα βλέπουμε τα αποτελέσματα της ανάλυσης κύριων συνιστωσών σε δεδομένα πωλήσεων φαγητού. Το επάνω αριστερά τεταρτημόριο απεικονίζει το ποσοστό της μεταβλητότητας που ερμηνεύεται από τις κύριες συνιστώσες, ως συνάρτηση του αριθμού τους. Στα υπόλοιπα τρία απεικονίζονται τα δεδομένα με βάση τις δύο πρώτες κύριες συνιστώσες. Τα διαφορετικά σύμβολα αντιστοιχούν σε διαφορετικές χώρες και η διαβάθμιση των χρωμάτων γίνεται με βάση τα ποσοστά της παχυσαρκίας [43].

3.4.2 Συσταδοποίηση (Clustering)

Η συσταδοποίηση αναφέρεται σε μία ευρεία γκάμα μεθόδων και τεχνικών με τη βοήθεια των οποίων μπορούμε να εντοπίσουμε υποομάδες και συστάδες σε ένα σύνολο δεδομένων. Σκοπός της συσταδοποίησης των δεδομένων είναι ο διαχωρισμός τους με τέτοιο τρόπο, ώστε οι παρατηρήσεις στα πλαίσια της κάθε ομάδας να παρουσιάζουν ομοιότητα μεταξύ τους, ενώ οι παρατηρήσεις μεταξύ διαφορετικών ομάδων να είναι διαφορετικές [42]. Ανάγκη για συσταδοποίηση μπορεί να εμφανιστεί σε πολλές περιπτώσεις, όπως για παράδειγμα στον κατακερματισμό της αγοράς (διαχωρισμός πελατών σε ομάδες με παρόμοια χαρακτηριστικά), στην ανάλυση κοινωνικών δικτύων, στην ιατρική (συσταδοποίηση κυττάρων από καρκινικούς όγκους με βάση τις μετρήσεις τους) κ.α.

Τόσο η συσταδοποίηση, όσο και η ανάλυση σε κύριες συνιστώσες, επιδιώκουν να απλοποιήσουν την ερμηνεία των δεδομένων μέσω μίας πιο συνοπτικής παρουσίασης, όμως οι μηχανισμοί που χρησιμοποιούν αυτές τεχνικές είναι διαφορετικοί:

- Η ανάλυση σε κύριες συνιστώσες επιδιώκει την αναπαράσταση των δεδομένων σε ένα χώρο λιγότερων διαστάσεων από αυτόν των παρατηρήσεων, με τέτοιο τρόπο ώστε να επεξηγείται ένα σημαντικό κομμάτι της μεταβλητότητας.
- Η συσταδοποίηση επιδιώκει να εντοπίσει ομογενείς υποομάδες μέσα στις παρατηρήσεις [42].

Συσταδοποίηση με τον αλγόριθμο K-Means

Ο αλγόριθμος K-means αποτελεί έναν απλό και κομψό τρόπο προκειμένου να διαχωρίσουμε ένα σύνολο δεδομένων σε K διακριτές και μη επικαλυπτόμενες συστάδες (clusters). Προκειμένου να εφαρμόσουμε την εν λόγω τεχνική πρέπει αρχικά να αποφασιστεί ο αριθμός

των συστάδων K . Κατόπιν ο αλγόριθμος θα αναθέσει κάθε παρατήρηση σε ακριβώς μία από τις K συστάδες.

Επιδίωξη του αλγορίθμου είναι να διαμορφώσει τις συστάδες με τέτοιο τρόπο ώστε το άθροισμα των διαφοροποιήσεων για κάθε συστάδα μεταξύ των μελών της να γίνει ελάχιστο. Η διαφοροποίηση αυτή συνήθως ορίζεται ως η ευκλείδεια απόσταση [42]. Τελικά το πρόβλημα ανάγεται σε ένα πρόβλημα βελτιστοποίησης, το οποίο αναπαρίσταται από την ακόλουθη εξίσωση:

$$\underset{C_1, \dots, C_k}{\text{minimize}} \left\{ \sum_{k=1}^K \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right\}$$

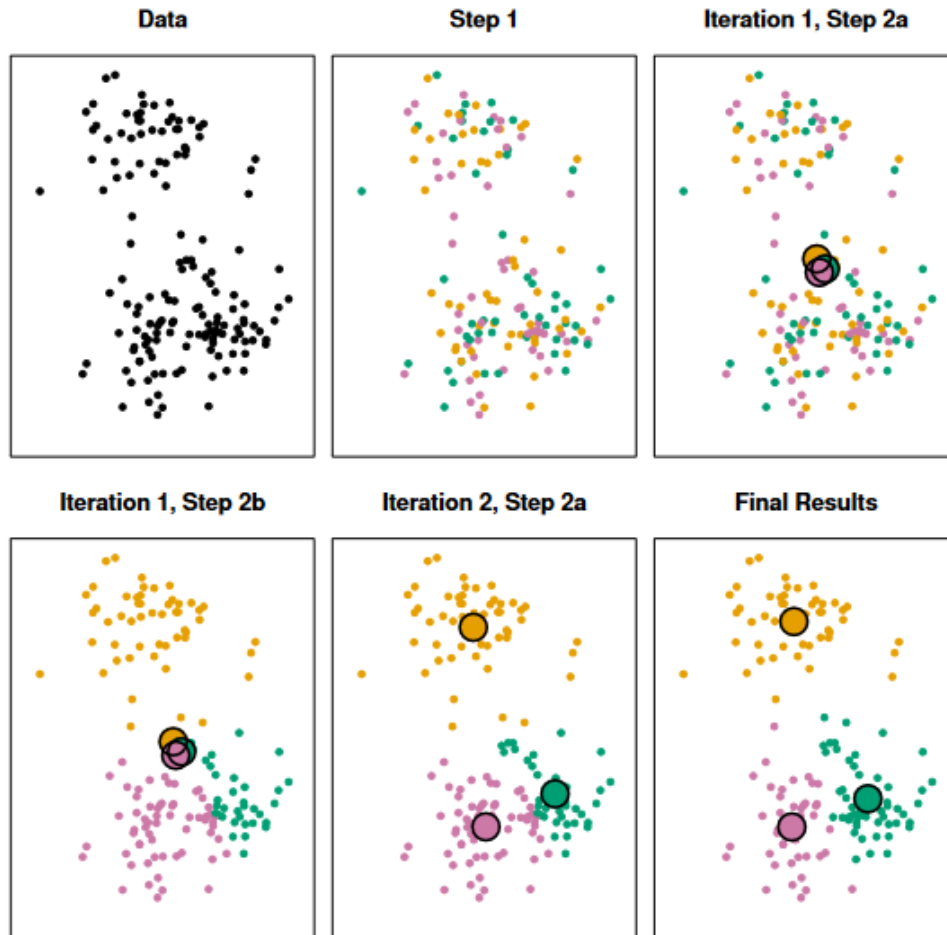
Όπου με $|C_k|$ συμβολίζεται ο αριθμός των παρατηρήσεων που ανήκουν στο K -οστό cluster. Δηλαδή με άλλα λόγια η διαφοροποίηση εντός του K -οστού cluster είναι το ανά ζεύγη άθροισμα των διαφορών μεταξύ των παρατηρήσεων δια τον συνολικό αριθμό των παρατηρήσεων στη συστάδα αυτή [42].

Το πρόβλημα αυτό είναι δυσεπίλυτο από αλγοριθμικής απόψεως (NP-hard) ακόμα και αν επιδιώκουμε να διαχωρίσουμε τα δεδομένα σε 2 συστάδες.

Υπάρχει ωστόσο ένας απλός αλγόριθμος ο οποίος έχει αποδειχθεί πως μπορεί να μας δώσει ένα τοπικό βέλτιστο στο πρόβλημα και παρουσιάζεται στα παρακάτω βήματα [42]:

1. Αποδίδουμε με τυχαίο τρόπο έναν αριθμό από το 1 έως το K σε κάθε παρατήρηση, δηλαδή την αναθέτουμε σε μία από τις συστάδες.
2. Επαναλαμβάνουμε τα ακόλουθα βήματα έως ότου οι αναθέσεις των συστάδων παραμείνουν αμετάβλητες.
 - A. Για κάθε μία από τις K συστάδες, υπολογίζουμε το κεντροειδές της, το οποίο συνίσταται στο διάνυσμα που έχει ως χαρακτηριστικά (features), τον μέσο όρο των αντίστοιχων χαρακτηριστικών των παρατηρήσεων που ανήκουν στην εν λόγω συστάδα.
 - B. Αναθέτουμε κάθε παρατήρηση στην συστάδα της οποίας το κεντροειδές είναι πλησιέστερα προς αυτήν (με το κριτήριο της ευκλείδειας απόστασης).

Ο αλγόριθμος αυτός εγγυάται την ελάττωση των τιμών της αντικειμενικής συνάρτησης του προβλήματος σε κάθε βήμα. Η λειτουργία του αλγορίθμου παρουσιάζεται εποπτικά στο ακόλουθο σχήμα.



Εικόνα 10. Τα βήματα του αλγορίθμου K-means σε διδιάστατα δεδομένα [42].

Τέλος πρέπει να αναφέρουμε, ότι η επιλογή του αριθμού των συστάδων δεν αποτελεί απλή διαδικασία, και επιφέρει αρκετές προκλήσεις στον μελετητή, καθώς μία λανθασμένη επιλογή του, μπορεί να οδηγήσει και σε λάθος συμπεράσματα σχετικά με τα δεδομένα.

Ιεραρχική συσταδοποίηση:

Η ιεραρχική συσταδοποίηση είναι μία μέθοδος που επιδιώκει την δημιουργία μίας ιεραρχίας συστάδων. Υπάρχουν δύο διαφορετικές στρατηγικές προκειμένου να επιτευχθεί αυτός ο στόχος [44]:

- Η συσσωματική (Anglomerative) :

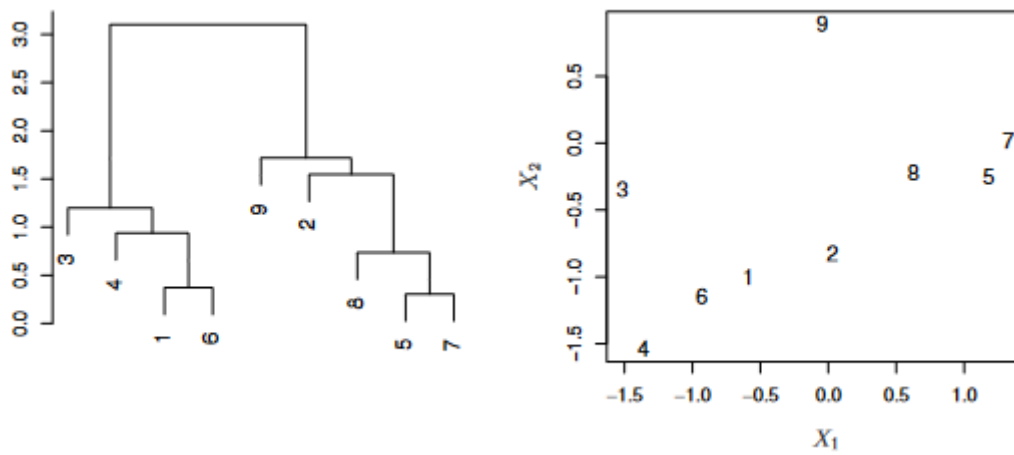
Αποτελεί προσέγγιση τύπου «από κάτω προς τα πάνω» (bottom-up). Κάθε παρατήρηση θεωρείται αρχικά ως ένα cluster, και διαδοχικά όσο προχωράει ο αλγόριθμος, ζεύγη παρατηρήσεων συνενώνονται.

- **Η διαιρετική (divisive):**

Αποτελεί προσέγγιση τύπου «από πάνω προς τα κάτω» (top-down), στην οποία όλες οι παρατηρήσεις θεωρούνται αρχικά ως μία συστάδα, και επιτελούνται διαδοχικές αναδρομικές διαμερίσεις αυτής όσο εξελίσσεται ο αλγόριθμος

Παραγόμενο της διαδικασίας ιεραρχικής συσταδοποίησης αποτελεί ένα δενδρόγραμμα. Στην περίπτωση που χρησιμοποιείται η συσσωματική στρατηγική, το δενδρόγραμμα αυτό λαμβάνεται με τη χρήση του εξής απλού αλγορίθμου [42]:

Ξεκινάμε καθορίζοντας ένα μέτρο διαφορετικότητας μεταξύ κάθε ζεύγους παρατηρήσεων. Στις περισσότερες των περιπτώσεων χρησιμοποιείται η ευκλείδεια απόσταση. Όπως αναφέρθηκε, ξεκινώντας ο αλγόριθμος αντιμετωπίζει την κάθε παρατήρηση ως μία ξεχωριστή συστάδα. Οι δύο παρατηρήσεις που είναι πιο συναφείς μεταξύ τους, συνενώνονται, κι έτσι αν είχαμε αρχικά n συστάδες, πλέον έχουμε $n-1$. Ο αλγόριθμος συνεχίζει επαναληπτικά με το επόμενο ζεύγος και πλέον έχουμε $n-2$ παρατηρήσεις κ.ο.κ. Εν τέλει μετά την ολοκλήρωσή του απομένει μόνο μία συστάδα και το δενδρόγραμμα έχει ολοκληρωθεί, όπως φαίνεται στο ακόλουθο σχήμα:



Εικόνα 11. Δημιουργία δενδρογράμματος ιεραρχικής συσταδοποίησης [42]

4 Νευρωνικά Δίκτυα - Βαθιά Μηχανική Μάθηση

Διανύουμε την τρίτη δεκαετία του 21^{ου} αιώνα, και ο όρος βαθιά μηχανική μάθηση, ακούγεται πλέον με εκθετικά αυξανόμενους ρυθμούς. Η βάση πάνω στην οποία θεμελιώθηκε ο συγκεκριμένος τομέας είναι τα τεχνητά Νευρωνικά δίκτυα (ANN), τα οποία θα μπορούσε κανείς να ορίσει ως υπολογιστικά συστήματα, των οποίων η δομή και η λειτουργία είναι εμπνευσμένη από τα βιολογικά νευρωνικά δίκτυα που απαρτίζουν τον εγκέφαλο των ζώων οργανισμών. Η δημοφιλία και η χρήση τους έχουν εκτιναχθεί τα τελευταία χρόνια και οι εφαρμογές τους άπτονται πλέον της καθημερινότητας όλων μας. Ωστόσο δεν ήταν πάντοτε έτσι.

Παρ' όλο που οι κεντρικές ιδέες πίσω από τα νευρωνικά δίκτυα, είχαν ερευνηθεί από την δεκαετία του 1950, η προσέγγιση αυτή χρειάστηκε δεκαετίες για να ξεκινήσει να εφαρμόζεται στην πράξη. Για ένα μεγάλο χρονικό διάστημα, ο χαμένος συνδετικός κρίκος ήταν ο αποτελεσματικός τρόπος εκπαίδευσής τους. Αυτό άλλαξε κατά τη διάρκεια της δεκαετίας του 80, όταν «επανεφευρέθηκε» ο αλγόριθμος οπίσθιας τροφοδότησης (backpropagation) και ξεκίνησε να εφαρμόζεται στα νευρωνικά δίκτυα [32]. Το χρονικό διάστημα αυτό υπήρξε ένας μεγάλος ενθουσιασμός για τη συγκεκριμένη προσέγγιση, μέχρι το σημείο που εμφανίστηκαν οι Μηχανές Διανυσμάτων Υποστήριξης (SVM), η τεχνική boosting και τα random forests, τα οποία απομάκρυναν τα νευρωνικά δίκτυα από το προσκήνιο. Αυτό οφειλόταν εν μέρει στο γεγονός ότι τα νευρωνικά δίκτυα απαιτούσαν μεγάλη παραμετροποίηση ενώ οι νέες αυτές μέθοδοι λειτουργούσαν σχεδόν “out of the box”. Οφειλόταν επίσης και στο γεγονός ότι οι νέες μέθοδοι υπερτερούσαν και ως προς τα αποτελέσματα [42].

Η κατάσταση αυτή εξακολούθησε και κατά την πρώτη δεκαετία του 21^{ου} αιώνα, όμως από το 2010 και μετά, τα νευρωνικά δίκτυα επανήλθαν στο προσκήνιο με την νέα πλέον ονομασία «βαθιά μηχανική μάθηση» έχοντας να επιδείξουν μεγάλες επιτυχίες σε προβλήματα όπως αναγνώριση εικόνας και βίντεο, επεξεργασία κειμένου και ομιλίας κ.λπ. [42].

Ωστόσο, οι βασικές αρχές των συνελκτικών νευρωνικών δικτύων (CNN), ο αλγόριθμος backpropagation και ο αλγόριθμος Long Short Term Memory (LSTM) που χρησιμοποιείται ευρέως για την πρόβλεψη χρονοσειρών, ήταν ήδη γνωστά από τα τέλη του 1990. Η βαθιά

μάθηση όμως, απογειώθηκε την δεκαετία του 2010, εξαιτίας των εξής κυρίως παραγόντων [32] :

- **Hardware και υπολογιστική ικανότητα**

Μεταξύ του 1990 και του 2010 οι επεξεργαστές του εμπορίου επιταχύνθηκαν κατά έναν παράγοντα της τάξεως του 5000 περίπου, με αποτέλεσμα πλέον, μικρά μοντέλα βαθιάς μηχανικής μάθησης να μπορούν να εκπαιδευτούν σε έναν οικιακό φορητό υπολογιστή (laptop) κάτι που θα ήταν αδύνατο κατά τα τέλη της δεκαετίας του 1990.

- **Δεδομένα**

Σύμφωνα με τον Francois Chollet [32] η τεχνητή νοημοσύνη θεωρείται από κάποιους η νέα βιομηχανική επανάσταση. Αν λοιπόν η βαθιά μηχανική μάθηση αποτελεί τη лоκομοτίβα της επανάστασης αυτής, τότε τα δεδομένα αποτελούν την καύσιμη ύλη της. Χωρίς αυτά τίποτα δεν θα ήταν εφικτό. Αναφορικά με τα δεδομένα, πέραν της εκθετικής (σύμφωνα με τον νόμο του Moore) αύξησης της αποθηκευτικής ικανότητας τα τελευταία 20 χρόνια, πραγματικό καταλύτη αποτέλεσε και το διαδίκτυο, το οποίο κατέστησε εφικτή τη συλλογή και διανομή μεγάλων συνόλων δεδομένων προκειμένου αυτά να χρησιμοποιηθούν στην μηχανική μάθηση.

- **Βελτιώσεις στους αλγορίθμους**

Εκτός από τους περιορισμούς στην υπολογιστική ισχύ και στα δεδομένα, μέχρι τα τέλη της δεκαετίας του 2000 δεν υπήρχαν και αποτελεσματικοί τρόποι να εκπαιδευτούν τα πολύ βαθιά νευρωνικά δίκτυα. Αυτό άλλαξε γύρω στο 2010 με την έλευση σημαντικών αλγοριθμικών βελτιώσεων, που αφορούσαν σε συναρτήσεις ενεργοποίησης (activation functions) των νευρωνικών δικτύων, σε καλύτερες μεθόδους αρχικοποίησης των βαρών, καθώς και σε μεθόδους βελτιστοποίησης όπως η Adam.

Ταυτόχρονα με αυτά, υπήρξε αναζωπύρωση του ενδιαφέροντος των επενδυτών προς την βαθιά μάθηση με αποτέλεσμα την τελευταία δεκαετία τα ποσά που διοχετεύτηκαν προς αυτή την κατεύθυνση να αυξηθούν εκθετικά.

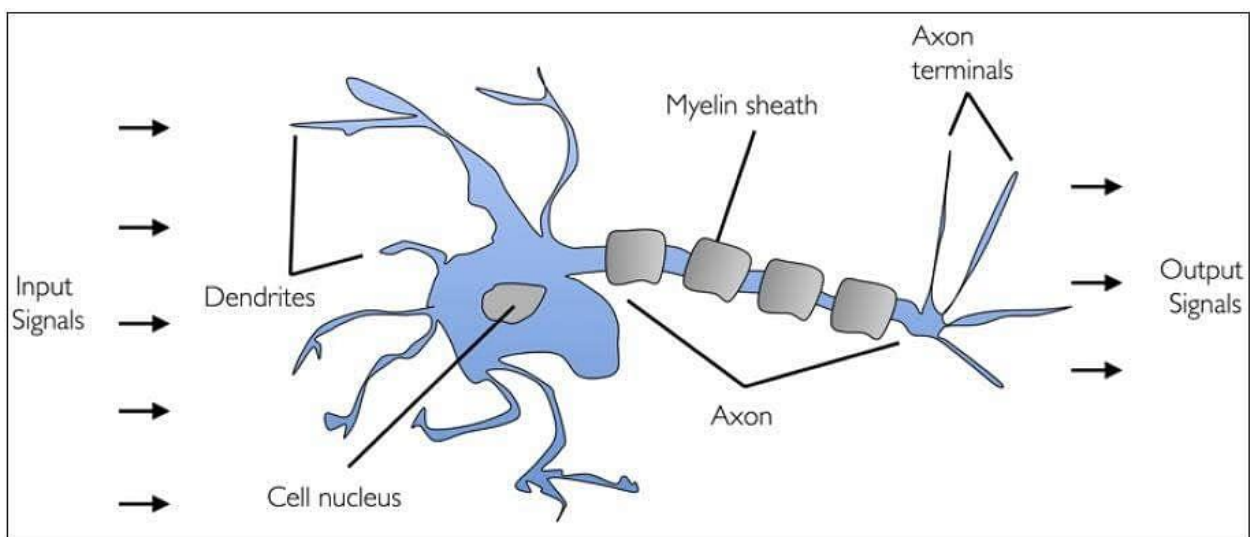
Τέλος, δεν πρέπει να παραβλέψουμε τον «εκδημοκρατισμό» που επιτελέστηκε. Κατά τα πρώτα χρόνια, για να ασχοληθεί κάποιος με τη βαθιά μάθηση, θα έπρεπε να έχει σημαντική εμπειρία στις γλώσσες C/C++ καθώς και στην πλατφόρμα CUDA. Πλέον επαρκούν βασικές

γνώσεις της γλώσσας rython, προκειμένου να επιτελέσει κάποιος προχωρημένη έρευνα στην βαθιά μάθηση [32].

4.1 Νευρωνικά δίκτυα

Η πεμπουσία των νευρωνικών δικτύων αποτελεί το Multi Layer Perceptron καθώς αποτελεί το χαρακτηριστικότερο παράδειγμά τους. Για την κατανόηση της λειτουργίας του μοντέλου αυτού, θα παρουσιάσουμε μερικά βασικά στοιχεία για τους τεχνητούς νευρώνες, το Perceptron και τη διαδικασία μάθησής του.

Ο ανθρώπινος εγκέφαλος αποτελείται από δισεκατομμύρια νευρώνων. Οι νευρώνες είναι, διασυνδεδεμένα μεταξύ τους νευρικά κύτταρα τα οποία είναι υπεύθυνα για την επεξεργασία και μετάδοση χημικών και ηλεκτρικών σημάτων. Αποτελούνται από τους δενδρίτες, τον πυρήνα και τον άξονα, όπως φαίνεται στην ακόλουθη εικόνα.

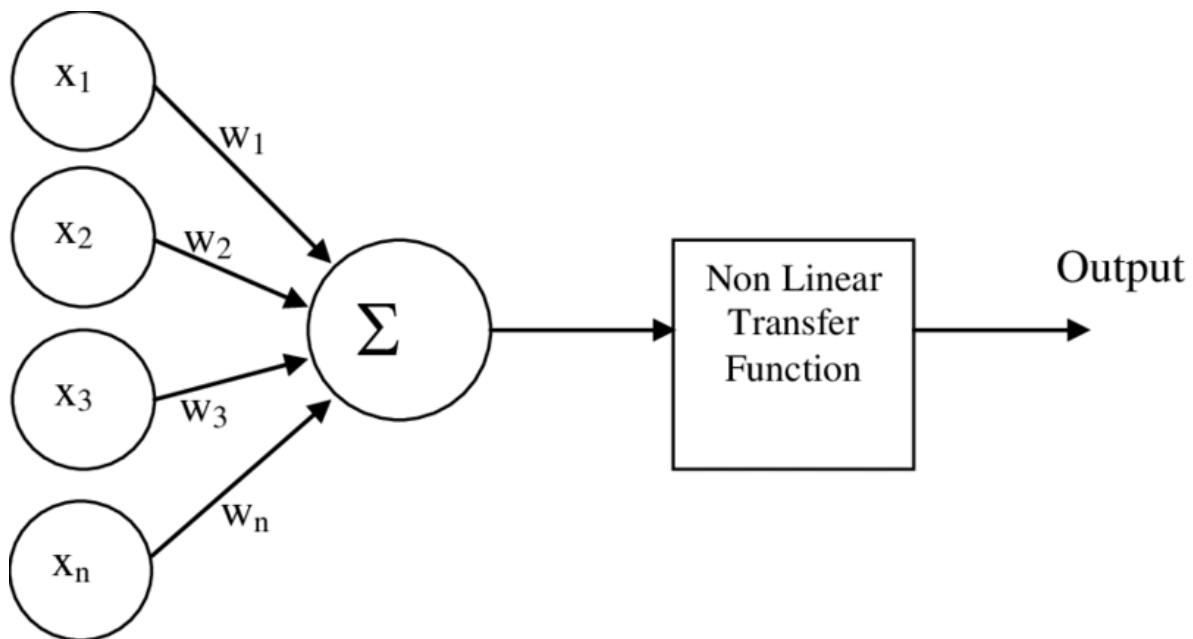


Εικόνα 12. Αναπαράσταση του βιολογικού νευρώνα.

Οι δενδρίτες (είσοδος του νευρώνα) είναι διακλαδώσεις οι οποίες λαμβάνουν πληροφορία από άλλους νευρώνες. Ο πυρήνας είναι υπεύθυνος για την επεξεργασία των σημάτων που λαμβάνονται, ενώ τέλος ο άξονας μπορεί να γίνει αντιληπτός σαν ένα καλώδιο που έχει σαν αποστολή τη μεταφορά του επεξεργασμένου σήματος προς άλλους νευρώνες [45]. Ο άξονας, όπως φαίνεται και στην εικόνα, καταλήγει σε διακλαδώσεις - «τερματικά» οι οποίες συνδέονται με τους δενδρίτες των άλλων νευρώνων. Οι συνδέσεις μεταξύ του άξονα και των άλλων νευρώνων ονομάζονται συνάψεις.

4.1.1 Τεχνητοί νευρώνες.

Η πρώτη αναφορά και σύλληψη του τεχνητού νευρώνα έγινε από τους McCulloch-Pitts (1943) και το αποτέλεσμα ονομάστηκε McCulloch-Pitts (MCP) neuron. Ο τεχνητός νευρώνας είναι επί της ουσίας μία μαθηματική συνάρτηση που βασίζεται σε ένα μοντέλο βιολογικών νευρώνων. Κάθε νευρώνας δέχεται έναν αριθμό n εισόδων, αποδίδει ένα συγκεκριμένο βάρος στην καθεμία και τροφοδοτεί το άθροισμά τους σε μία μη γραμμική συνάρτηση η οποία παρέχει μία έξοδο. Η διαδικασία αυτή απεικονίζεται στο ακόλουθο σχήμα [46].

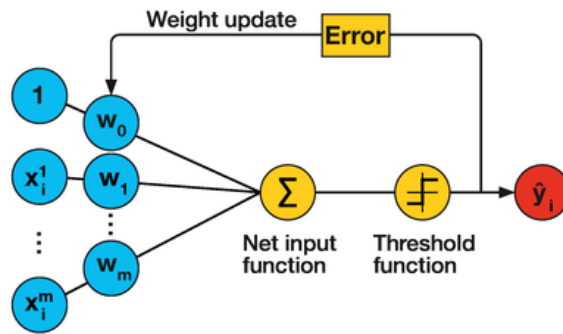


Εικόνα 13. Η λειτουργία του τεχνητού νευρώνα [46]

4.1.2 Perceptron

Η ιδέα του perceptron διατυπώθηκε από τον Rosenblatt (1957) [47]. Πρότεινε έναν κανόνα μάθησης βασισμένο στην αρχιτεκτονική του αρχικού MCP νευρώνα (1943). Το perceptron αποτελεί αλγόριθμο επιβλεπόμενης μάθησης για δυαδικούς κατηγοριοποιητές. Ο αλγόριθμος του Perceptron καθιστά τους νευρώνες ικανούς να μαθαίνουν και να επεξεργάζονται τα στοιχεία, με τα οποία τροφοδοτούνται από το σύνολο δεδομένων εκπαίδευσης, εξετάζοντάς τα, ένα κάθε φορά.

Το perceptron ουσιαστικά, εντοπίζει τα βέλτιστα βάρη που πρέπει να έχουν τα σήματα εισόδου, με σκοπό να σχηματιστεί ένα γραμμικό όριο απόφασης, το οποίο καθιστά δυνατό το διαχωρισμό των δεδομένων σε δύο γραμμικά διαχωριζόμενες κλάσεις.



Εικόνα 14. Ο κανόνας μάθησης perceptron [48]

Λαμβάνοντας πολλαπλά σήματα εισόδου, το perceptron τα αθροίζει πολλαπλασιάζοντάς τα με τα βάρη w_i . Αν το άθροισμα υπερβαίνει ένα συγκεκριμένο κατώφλι-όριο, επιστρέφει ή δεν επιστρέφει αντίστοιχα μία έξοδο. Σε ένα πρόβλημα επιβλεπόμενης μάθησης, αυτή η έξοδος μπορεί επί παραδείγματι να χρησιμοποιηθεί για να προβλέψει την κλάση στην οποία ανήκει ένα δείγμα.

Η συνάρτηση που χρησιμοποιεί το perceptron προκειμένου να αντιστοιχήσει τις εισόδους του με την έξοδο ορίζεται ως εξής:

$$f(x) = \begin{cases} 1, & w \cdot x + b > 0 \\ 0, & w \cdot x \leq 0 \end{cases}$$

Στην εξίσωση αυτή, συμβολίζουμε:

Με w το διάνυσμα βαρών των εισόδων το οποίο παίρνει τιμές στο σύνολο των πραγματικών αριθμών.

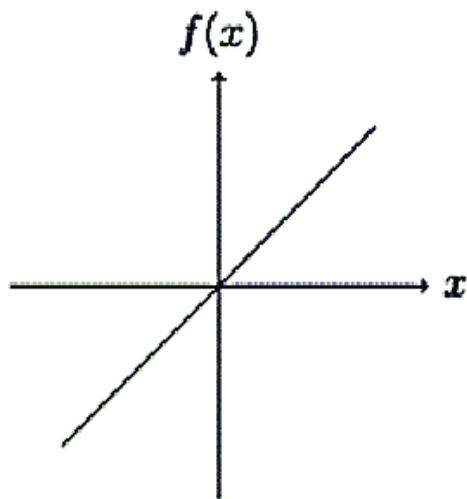
Με b το λεγόμενο bias, δηλαδή ένα στοιχείο το οποίο μετατοπίζει την ευθεία του ορίου έτσι ώστε να μην περνάει από την αρχή των αξόνων και η οποία δεν εξαρτάται από τις τιμές εισόδου.

Με x το διάνυσμα των εισόδων.

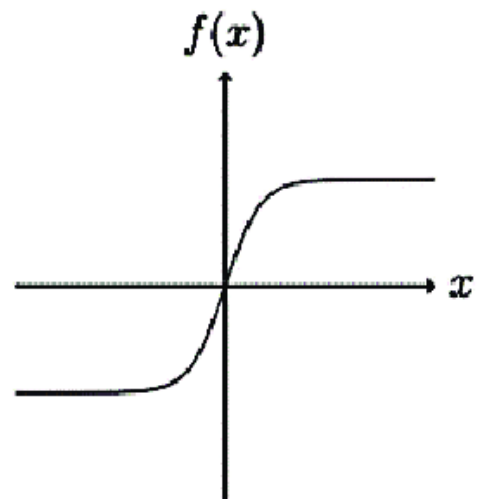
Συνεπώς, ισχύει ότι:

$$w \cdot x = \sum_{i=1}^n w_i x_i$$

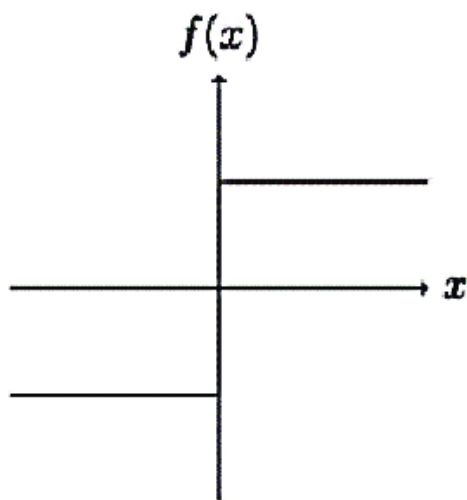
Όπου n ο αριθμός των εισόδων όπως έχουμε ήδη αναφέρει.



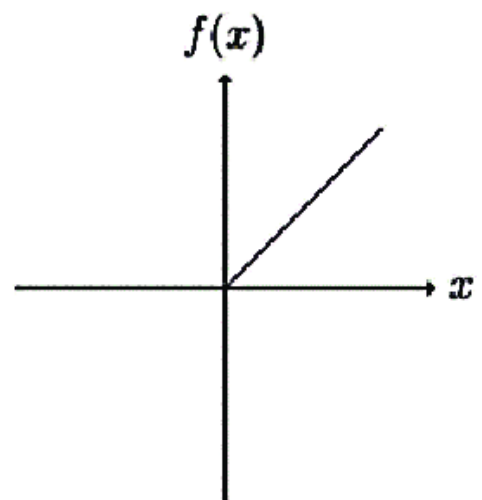
Linear function



Sigmoid function



lump function



Rectified linear function

Εικόνα 15. Σνηθισμένες συναρτήσεις ενεργοποίησης στα Νευρωνικά δίκτυα [49].

Η δυαδική αυτή έξοδος μπορεί να έχει είτε τη μορφή 0 και 1 είτε τη μορφή -1 και 1 ανάλογα με το ποια συνάρτηση ενεργοποίησης θα επιλεγεί. Τα τελευταία χρόνια και ειδικότερα στις πιο πολύπλοκες αρχιτεκτονικές βαθιάς μάθησης, έχει γίνει πολύ δημοφιλής η συνάρτηση ReLU (Rectified Linear Unit) η οποία δίνεται από την ακόλουθη εξίσωση:

$$f(x) = x^+ = \max(0, x)$$

Η ReLU, ξεκίνησε να εφαρμόζεται για την εξαγωγή οπτικών χαρακτηριστικών (features) σε αρχιτεκτονικές ιεραρχικών νευρωνικών δικτύων ξεκινώντας κατά την δεκαετία του 1960, και

στην πορεία εμφανίστηκε ισχυρή επιχειρηματολογία για το ότι έχει έντονη διασύνδεση με βιολογικούς μηχανισμούς καθώς και μαθηματική αιτιολόγηση περί αυτού [50] [51]. Το 2011 παρατηρήθηκε ότι οδηγεί σε καλύτερη εκπαίδευση αλγορίθμων της βαθιάς μηχανικής μάθησης σε σχέση με τις μέχρι τότε κοινώς χρησιμοποιούμενες συναρτήσεις ενεργοποίησης [52], όπως για παράδειγμα η λογιστική σιγμοειδής συνάρτηση, η οποία είναι εμπνευσμένη από τη θεωρία των πιθανοτήτων και για την οποία έγινε λόγος σε προηγούμενη ενότητα, καθώς και η πιο πρακτική στη χρήση της συνάρτηση της υπερβολικής εφαπτομένης.

Τα πλεονεκτήματά της έγκεινται σε:

- **Αραιή ενεργοποίηση:** Σε ένα νευρωνικό δίκτυο με τυχαία αρχικοποιημένα βάρη, μόνο το 50% των νευρώνων κρυφών επιπέδων ενεργοποιούνται.
- **Καλύτερη διάδοση κλίσης (gradient propagation).** Το πρόβλημα εξαφανιζόμενης κλίσης (vanishing gradient) εμφανίζεται λιγότερο συχνά σε σχέση με τις σιγμοειδείς συναρτήσεις [52].
- **Αποδοτικός υπολογισμός:** Υπολογίζεται μόνο μέσω συγκρίσεων, προσθέσεων και πολλαπλασιασμών.
- **Κλιμακωτά αμετάβλητη:** Ισχύει η σχέση :

$$\max(0, ax) = a \max(0, x) \forall a \geq 0$$

Κάποια πιθανά προβλήματα που μπορεί να προκύψουν από τη χρήση της είναι:

- **Μη διαφορίσιμη στο μηδέν.** Ωστόσο είναι διαφορίσιμη οπουδήποτε αλλού και η τιμή της παραγώγου στο μηδέν μπορεί να επιλεγεί αυθαιρέτως σε 0 ή 1.
- **Μη συμμετρική ως προς το μηδέν.**
- **Μη περιορισμένη**
- **Το πρόβλημα της «αποθνήσκουσας» ReLU (dying ReLU problem).** Σε κάποιες περιπτώσεις οι νευρώνες μπορεί να περιέλθουν σε καταστάσεις όπου πλέον δεν ενεργοποιούνται για καμία δυνατή τιμή εισόδου.

Δεδομένου ότι η εν λόγω συνάρτηση χρησιμοποιείται σχεδόν αποκλειστικά από το 2017 στις περισσότερες εφαρμογές μηχανικής μάθησης που αφορούν υπολογιστική όραση, αναγνώριση ομιλίας και υπολογιστικής νευροεπιστήμης, επιλέχθηκε και για την παρούσα διπλωματική ως συνάρτηση ενεργοποίησης για τους νευρώνες, με εξαίρεση το επίπεδο εξόδου.

4.1.3 Μάθηση

Όπως αναφέραμε και προηγουμένως, η βασική ιδιότητα του Perceptron έγκειται στον κανόνα μάθησης, μέσω του οποίου τα τυχαίως αρχικοποιημένα βάρη, βελτιστοποιούνται, έτσι ώστε η έξοδος του νευρώνα να δίνει τη σωστή λύση για το πρόβλημα.

Είναι προφανές πως οι αρχικές έξοδοι του Perceptron, δεν μπορούν να δώσουν κοντινές στην πραγματικότητα τιμές. Προκειμένου να συμβεί αυτό, λαμβάνει χώρα μία σταδιακή τροποποίηση των βαρών, η οποία βασίζεται στην ανατροφοδότηση που λαμβάνουμε, συγκρίνοντας την έξοδο με τα ιδανικά αποτελέσματα που θα έπρεπε να προκύψουν. Αυτή η διαδικασία συνιστά και την λεγόμενη εκπαίδευση από την οποία παίρνει το όνομά της η μηχανική μάθηση.

Τα βήματα του επαναληπτικού αλγορίθμου που εφαρμόζονται είναι τα εξής [32]:

1. Θεωρούμε ένα τμήμα (batch) δειγμάτων εκπαίδευσης και τις αντίστοιχες τιμές – στόχους.
2. Λαμβάνουμε τις προβλέψεις που μας δίνει το δίκτυο για το συγκεκριμένο τμήμα.
3. Υπολογίζουμε με ένα μέτρο απώλειας (loss function) την απόσταση των προβλέψεων από τις τιμές στόχους.
4. Ενημερώνουμε τα βάρη κατά τρόπο τέτοιο ώστε να ελαττώσουμε ελαφρώς την απόσταση του τμήματος αυτού από τις τιμές στόχους.

Εφαρμόζοντας τον αλγόριθμο αυτό αρκετές φορές στα δεδομένα εκπαίδευσης, θα καταλήξουμε σε ένα διάνυσμα βαρών με τη χρήση του οποίου, η συνάρτηση απώλειας θα παίρνει πολύ χαμηλές τιμές στα δεδομένα εκπαίδευσης.

4.1.4 Στοχαστική κατάβαση δυναμικού (stochastic gradient descent)

Στην προηγούμενη ενότητα αναφερθήκαμε στην επαναληπτική διαδικασία βάσει της οποίας ενημερώνουμε τα βάρη, όμως ένα σημαντικό ερώτημα που τίθεται, είναι «με ποιο τρόπο γίνεται σωστά και αποδοτικά αυτή η ενημέρωση»; Μία πρώτη αφελής προσέγγιση θα ήταν να διατηρούμε κάθε φορά σταθερά όλα τα βάρη πλην ενός, και να δοκιμάζουμε διαφορετικές τιμές για το ένα το οποίο μπορούμε να μεταβάλλουμε (μία αύξηση και μία μείωση), μέχρις ότου διαπιστώσουμε ότι η τιμή της συνάρτησης απώλειας ελαττώνεται. Τη διαδικασία αυτή θα έπρεπε να επαναλάβουμε για ολόκληρο το διάνυσμα βαρών. Μία τέτοια προσέγγιση όμως

θα ήταν εξαιρετικά δαπανηρή από πλευράς υπολογιστικού χρόνου, καθώς θα έπρεπε να εκτελέσουμε δύο περάσματα σε κάθε μεμονωμένο βάρος, ο αριθμός των οποίων στα πιο σύνθετα μοντέλα μπορεί να φτάσει τις χιλιάδες ή και τα εκατομμύρια.

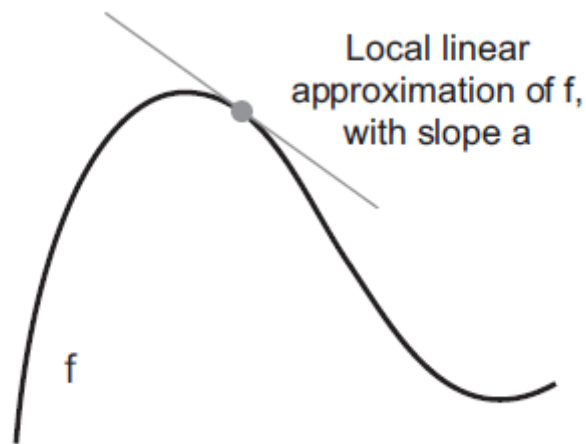
Μία πιο αποδοτική προσέγγιση βασίζεται στο γεγονός ότι όλες οι συναρτήσεις που χρησιμοποιούνται είναι παραγωγίσιμες και είμαστε σε θέση να υπολογίσουμε τις μερικές παραγώγους τους ως προς τους συντελεστές (βάρη) του δικτύου. Μπορούμε, αφού γίνει αυτό να μεταβάλλουμε τα βάρη προς την αντίθετη κατεύθυνση από αυτήν της κλίσης (παραγώγου) προκειμένου να ελαττώσουμε την τιμή της συνάρτησης κόστους [32].

Με τον όρο gradient (βαθμίδα) εννοούμε την παράγωγο ενός τανυστή (tensor) και αποτελεί τη γενίκευση της έννοιας της παραγώγου με εισόδους (δηλαδή συναρτήσεις) πολλών διαστάσεων [32]. Όπως ήδη αναφέρθηκε για συναρτήσεις μίας μεταβλητής, αντίστοιχα μπορούμε να μεταβάλλουμε και την τιμή μίας πολυμεταβλητής συνάρτησης, μεταβάλλοντας τις ανεξάρτητες μεταβλητές της, που εν προκειμένω είναι τα βάρη, προς την αντίθετη κατεύθυνση από αυτήν της παραγώγου ως προς το συγκεκριμένο βάρος.

Ένας διαισθητικός τρόπος για να γίνουν καλύτερα αντιληπτά τα παραπάνω είναι μέσω του ακόλουθου σχήματος που παρουσιάζει την λογική αυτή με τη βοήθεια μιας συνάρτησης μίας μεταβλητής. Όπως παρατηρούμε στο σχήμα, στο σημείο της συνάρτησης όπου έχει σχεδιαστεί η κλίση της, είναι αρνητική, δηλαδή η συνάρτηση είναι γνησίως φθίνουσα στο σημείο αυτό (ελαττώνεται η τιμή της για αύξηση της ανεξάρτητης μεταβλητής). Συνεπώς, για να ελαττώσουμε την τιμή της συνάρτησης, πρέπει να μεταβάλλουμε την ανεξάρτητη μεταβλητή προς την αντίθετη κατεύθυνση της κλίσης, δηλαδή να την αυξήσουμε.

Πρέπει να επισημανθεί εδώ, πως η αύξηση ή η μείωση των βαρών δεν πρέπει να γίνεται αυθαίρετα αλλά βάσει ενός μικρού και προκαθορισμένου βήματος (step ή αλλιώς learning rate) κάθε φορά, το οποίο χρησιμοποιείται σαν πολλαπλασιαστικός παράγοντας της βαθμίδας (gradient) επειδή η τελευταία, αποτυπώνει την καμπυλότητα της συνάρτησης μόνο σε μία κοντινή περιοχή γύρω από το σημείο που μελετάμε. Έτσι λοιπόν αν συμβολίσουμε με W_0 το αρχικό διάνυμα βαρών, τότε αυτό που θα λάβουμε από την πρώτη επανάληψη της διαδικασίας W_1 , θα γίνεται από τη σχέση [32] :

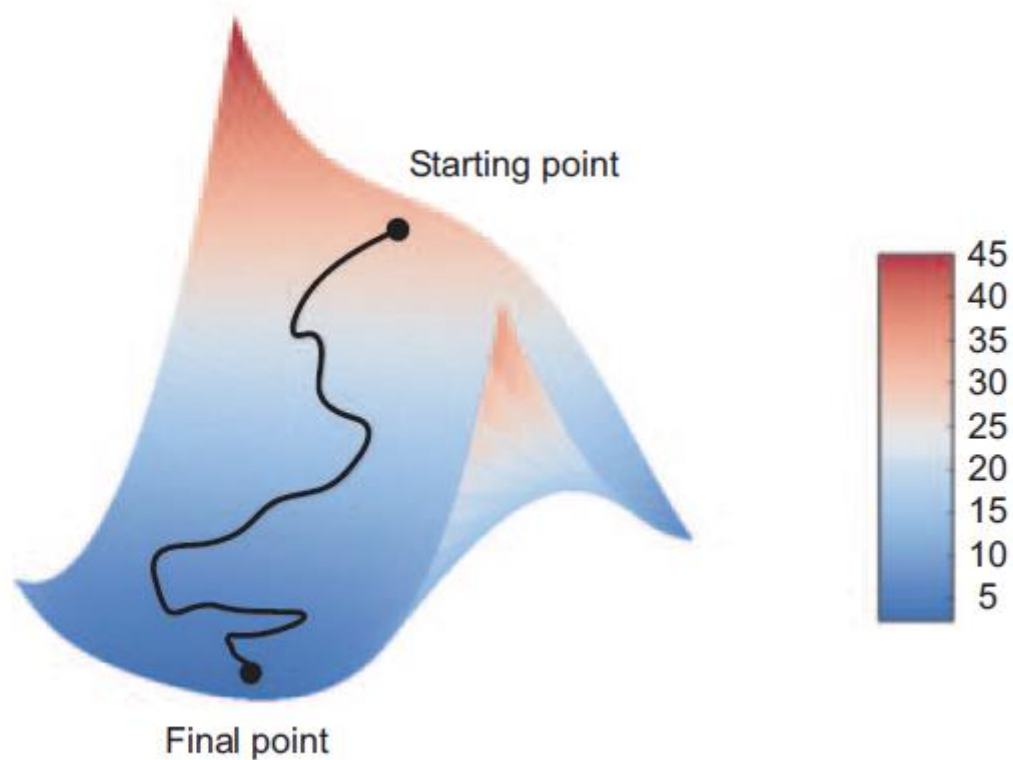
$$W1 = W0 - step * grad(f)(W0)$$



Εικόνα 16. Συνάρτηση μίας μεταβλητής και η κλίση της [32]

Έχοντας μία παραγωγίσιμη συνάρτηση, είναι δυνατό να υπολογιστεί το ολικό ελάχιστό της με αναλυτικό τρόπο, βρίσκοντας τα σημεία όπου μηδενίζεται η παράγωγός της. Όμως σε ένα νευρωνικό δίκτυο με χιλιάδες ή και εκατομμύρια παραμέτρους κάτι τέτοιο είναι πρακτικά ανέφικτο. Ως εκ τούτου, αντί να υπολογιστεί μονομιάς το ελάχιστο, υπολογίζεται σταδιακά με μικρές αλλαγές στο διάνυσμα των βαρών.

Κεφαλαιώδους σημασίας είναι η επιλογή μιας λογικής τιμής για το βήμα (step) προσαρμογής. Μία πολύ μικρή τιμή μπορεί να «εγκλωβίσει» τα βάρη σε ένα τοπικό ελάχιστο της συνάρτησης, ενώ μία πολύ μεγάλη τιμή μπορεί να μεταβάλλει τα βάρη σε εντελώς τυχαίες τιμές επί της καμπύλης [32].



Εικόνα 17. Κατάβαση δυναμικού σε μία επιφάνεια δύο διαστάσεων (2 παράμετροι) [32]

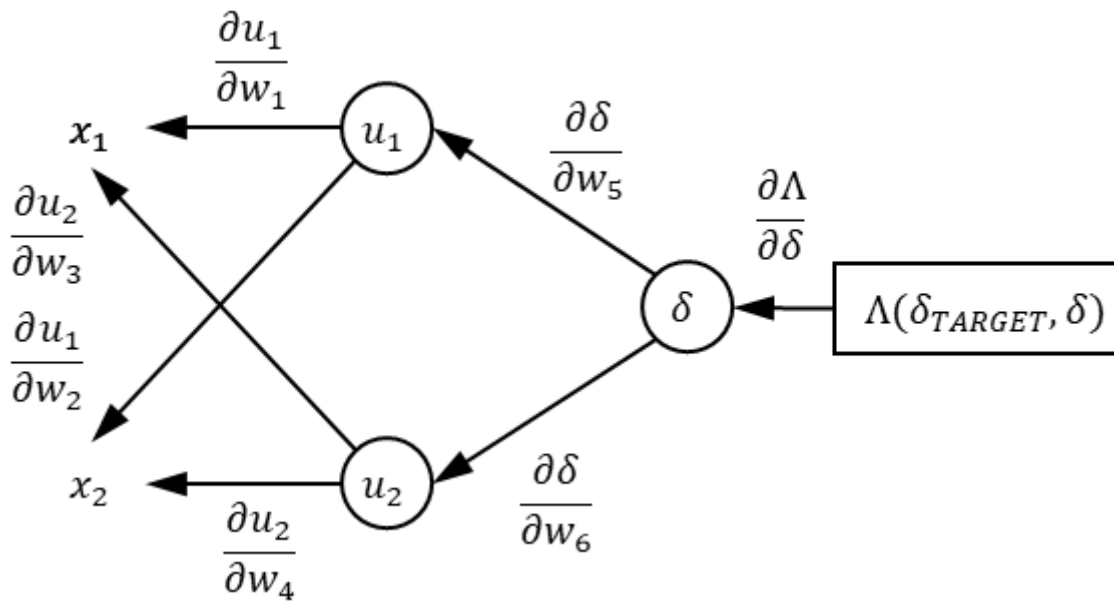
Η περιγραφείσα τεχνική αποτελεί την στοχαστική κατάβαση δυναμικού, την οποία μπορούμε να αντιληφθούμε οπτικά με το παραπάνω σχήμα που την απεικονίζει σε μία επιφάνεια δύο ανεξάρτητων παραμέτρων, καθώς είναι αρκετά δύσκολο να την αντιληφθεί κανείς σε έναν χώρο με εκατομμύρια παραμέτρους, όπως συμβαίνει με τα νευρωνικά δίκτυα που χρησιμοποιούνται στην πράξη.

4.1.5 Οπίσθια διάδοση (Backpropagation)

Τελευταίο κομμάτι στην αλυσίδα της εκπαίδευσης των νευρωνικών δικτύων αποτελεί ο αλγόριθμος οπίσθιας διάδοσης ή backpropagation. Στη πράξη, και λόγω του γεγονότος ότι στις συνήθεις περιπτώσεις, τα νευρωνικά δίκτυα περιλαμβάνουν πολλές συναρτήσεις με τανυστές (tensors), οι οποίες έχουν την ιδιότητα να έχουν παράγωγο που υπολογίζεται με απλό τρόπο, και η κάθε μία αποτελεί σύνθεση της άλλης, κάνουμε χρήση του λεγόμενου κανόνα της αλυσίδας προκειμένου να υπολογίσουμε τις ζητούμενες παραγώγους. Ο κανόνας της αλυσίδας δίνεται από τη σχέση:

$$f(g(x)) = f'(g(x)) * g'(x)$$

Η εφαρμογή του αλγορίθμου backpropagation στα νευρωνικά δίκτυα, ξεκινάει με τον υπολογισμό τις τιμές της συνάρτησης απώλειας για το τελευταίο επίπεδο και κινείται προς τα πίσω φτάνοντας τελικά στα αρχικά επίπεδα, εφαρμόζοντας διαδοχικά τον κανόνα της αλυσίδας προκειμένου να υπολογιστεί η συνεισφορά της κάθε παραμέτρου στην συνάρτηση απώλειας [32].



Εικόνα 18. Η λειτουργία του αλγορίθμου backpropagation με τη χρήση του κανόνα της αλυσίδας [53].

Η διαδικασία αυτή απεικονίζεται στο παραπάνω σχήμα.

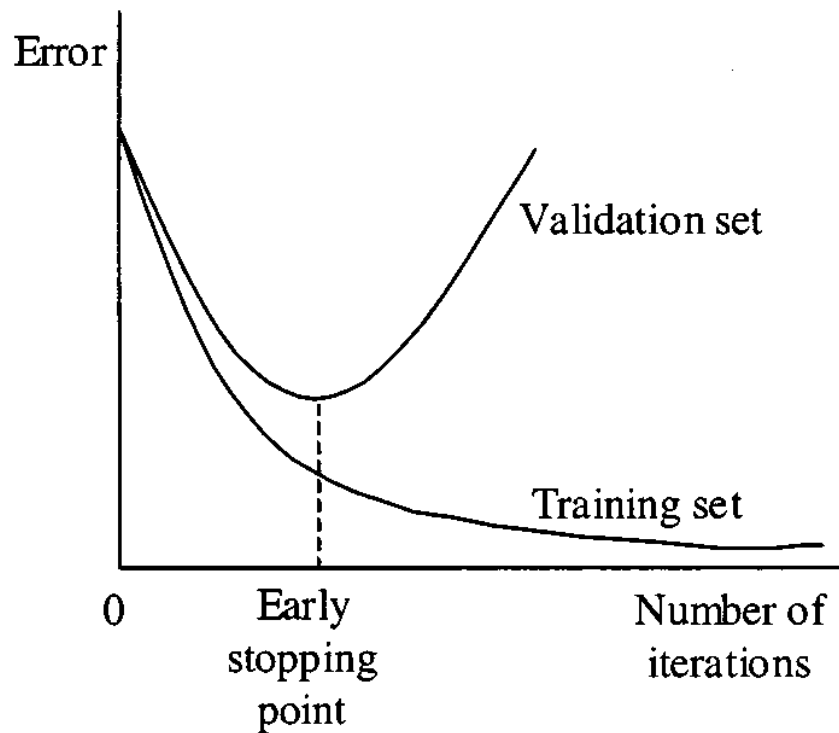
4.1.6 Ζητήματα εκπαίδευσης

Αντί του καθορισμού ενός συγκεκριμένου πλήθους εποχών³ για την εκπαίδευση ενός νευρωνικού δικτύου, είναι συνηθισμένη πρακτική η επιλογή της μεθοδολογίας early stopping η οποία σταματάει την εκπαίδευση, όταν οι προσαρμογές των βαρών, σταματάνε να

³Λέγοντας εποχή εννοούμε έναν κύκλο εκπαίδευσης των δεδομένων με τη χρήση όλου του συνόλου εκπαίδευσης μία φορά. Μία εποχή αποτελείται από μία ή περισσότερες δεσμίδες (batches) όπου ως δεσμίδα αναφέρεται το σύνολο των δειγμάτων που χρησιμοποιούνται για εκπαίδευση έως ότου λάβει χώρα μία ενημέρωση των παραμέτρων του μοντέλου. Έτσι, εποχή είναι ουσιαστικά η πλήρης επεξεργασία των δεδομένων που συμπεριλαμβάνει την εκπαίδευση σε ένα ή περισσότερα batches.

επιφέρουν βελτιώσεις στο σύνολο δεδομένων ελέγχου. Επί της ουσίας αυτό που συμβαίνει με το early stopping, είναι ότι αποθηκεύονται και ενημερώνονται οι τρέχουσες βέλτιστες παράμετροι κατά τη διάρκεια της εκπαίδευσης και όταν μετά από κάποιες επαναλήψεις σταματήσει να παρατηρείται βελτίωση, τερματίζεται η εκπαίδευση και χρησιμοποιείται το σύνολο των τελευταίων βέλτιστων παραμέτρων που προέκυψαν.

Η όλη διαδικασία παρουσιάζεται σχηματικά στο ακόλουθο διάγραμμα:



Εικόνα 19. Διαδικασία εκπαίδευσης με τη μέθοδο early stopping [54].

Όπως παρατηρούμε στην εικόνα, όσο αυξάνονται οι επαναλήψεις της εκπαίδευσης, το μοντέλο προσαρμόζεται ολοένα και περισσότερο στα δεδομένα εκπαίδευσης, και ως εκ τούτου το σφάλμα του ελαττώνεται προσεγγίζοντας ασυμπτωτικά στο μηδέν. Αντιθέτως, στα δεδομένα ελέγχου υπάρχει ένα σημείο επαναλήψεων εκπαίδευσης στο οποίο ελαχιστοποιείται, και από το σημείο εκείνο και μετά, εμφανίζεται το φαινόμενο του overfitting με αποτέλεσμα να αρχίσει να αυξάνεται. Η προσφορά του early stopping έγκειται στον προσδιορισμό αυτού του σημείου ελαχιστοποίησης του σφάλματος. Ταυτόχρονα, με τη χρήση του, εξοικονομείται χρόνος εκπαίδευσης σε περίπτωση που είχε χρησιμοποιηθεί μεγαλύτερος αριθμός εποχών από τον αναγκαίο.

Συχνά, για την πιο αποτελεσματική εκπαίδευση του δικτύου χρησιμοποιείται επίσης ο αλγόριθμος (optimizer) Adam, ο οποίος βελτιστοποιεί την τεχνική της κατάβασης δυναμικού.

Ο Adam χρησιμοποιεί έναν συνδυασμό δύο μεθοδολογιών κατάβασης δυναμικού, λαμβάνοντας υπόψη τον εκθετικά σταθμισμένο μέσο όρο των κλίσεων. Με τη χρήση των μέσων όρων επιτυγχάνεται η ταχύτερη σύγκλιση του στην ελάχιστη τιμή. Είναι ιδιαίτερα χρήσιμος όταν επιχειρείται να δοθεί λύση σε μεγάλα προβλήματα που περιλαμβάνουν μεγάλο πλήθος παραμέτρων.

4.1.7 Βελτιστοποίηση υπερπαραμέτρων

Όπως έχει αναφερθεί, ένας από τους λόγους για τους οποίους τα Νευρωνικά δίκτυα έχασαν την δημοτικότητά τους κατά τη δεκαετία του 2000 είναι και το γεγονός ότι η εκπαίδευση απαιτεί τον καθορισμό πολλών παραμέτρων. Ένα ζήτημα κεφαλαιώδους σημασίας, προκειμένου να υπάρξουν ποιοτικά αποτελέσματα είναι η σωστή επιλογή των τιμών των παραμέτρων, μία διαδικασία που ονομάζεται βελτιστοποίηση υπερπαραμέτρων ή «hyperparameter optimization». Λέγοντας υπερπαραμέτρους, εννοούμε εκείνες των οποίων η τιμή δεν προκύπτει μέσω της διαδικασίας εκπαίδευσης αλλά αντιθέτως πρέπει να καθορισθεί εκ των προτέρων από τον ερευνητή.

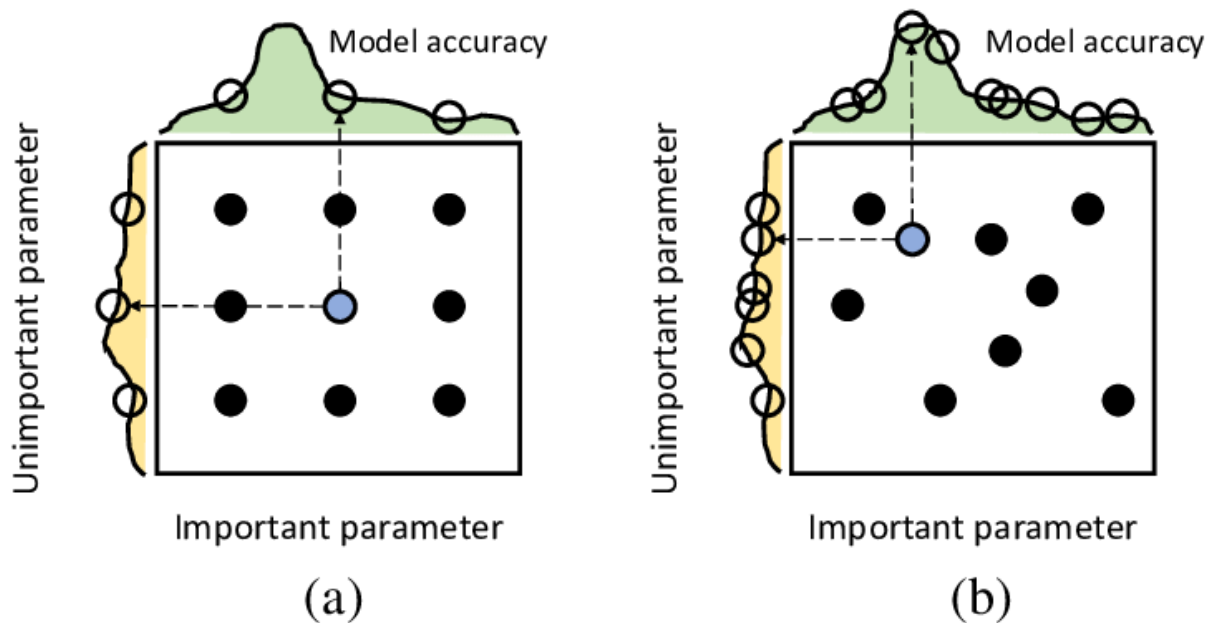
Ο παραδοσιακός τρόπος για την εκτέλεση της βελτιστοποίησης, είναι η λεγόμενη αναζήτηση πλέγματος (grid search) η οποία συνίσταται σε μία εξαντλητική αναζήτηση των βέλτιστων τιμών, μέσα από ένα επιλεγμένο από τον ερευνητή χώρο υπερπαραμέτρων. Για κάθε συνδυασμό παραμέτρων, εκπαιδεύεται το μοντέλο και κάνοντας χρήση μίας μετρικής, επιλέγεται τελικά αυτός ο οποίος θα επιφέρει το καλύτερο αποτέλεσμα.

Η προσέγγιση αυτή ωστόσο, πάσχει από την λεγόμενη «κατάρρα των διαστάσεων» (curse of dimensionality). Επειδή οι παράμετροι που πρέπει να εξεταστούν είναι πάρα πολλές, και το μέγεθος του χώρου υπερπαραμέτρων αυξάνεται εκθετικά με την προσθήκη επιπλέον παραμέτρων, τελικά προκύπτει ένα τεράστιο πλήθος εκπαιδεύσεων που πρέπει να γίνουν, κάτι που συνεπάγεται και μεγάλες ανάγκες σε χρόνο και υπολογιστικούς πόρους .

Μία δεύτερη λιγότερο απαιτητική τεχνική είναι η λεγόμενη τυχαία αναζήτηση. Στην περίπτωση αυτή, ο ερευνητής καθορίζει όπως και πριν τον χώρο των υπερπαραμέτρων αλλά εδώ ορίζεται επίσης κι ένας ορισμένος αριθμός εκπαιδεύσεων που θα λάβουν χώρα, κατά τις

οποίες, δημιουργείται με τυχαίο τρόπο ένας συνδυασμός υπερπαραμέτρων και τελικά επιλέγεται αυτός που έδωσε τα καλύτερα αποτελέσματα.

Στην ακόλουθη εικόνα παρουσιάζονται συγκριτικά οι δύο αυτές μέθοδοι:

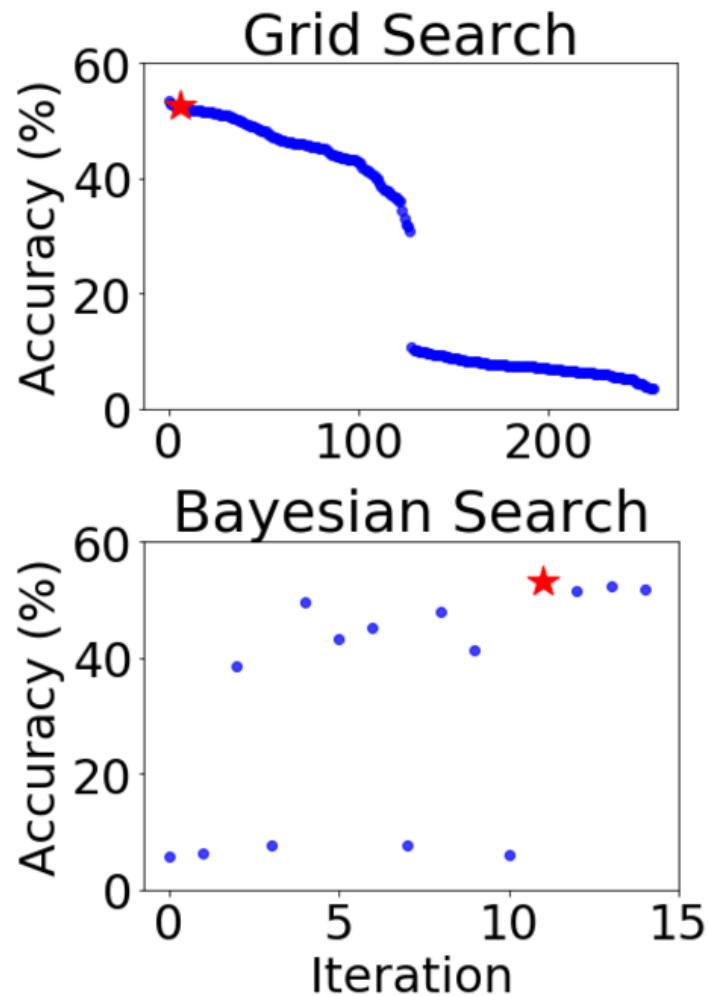


Εικόνα 20. Σύγκριση των μεθόδων της αναζήτησης πλέγματος και της τυχαίας αναζήτησης [55]

Όπως παρατηρούμε, αν η αναζήτηση πλέγματος είναι πολύ αραιή, δεν εγγυάται βέλτιστα αποτελέσματα, και για να γίνει πιο πυκνή απαιτεί εξαιρετικά μεγάλο αριθμό δοκιμών.

Τέλος, υπάρχει και μία τρίτη μέθοδος βελτιστοποίησης, η οποία αποκαλείται Bayesian (με χρήση του εκτιμητή TreeParzen ή TPE) της οποίας η λειτουργία συνοψίζεται ως εξής: Δημιουργείται ένα πιθανοκρατικό μοντέλο της αντικειμενικής συνάρτησης που επιχειρείται να βελτιστοποιηθεί, το οποίο χρησιμοποιείται προκειμένου να επιλεγούν οι πιο «υποσχόμενοι» συνδυασμοί παραμέτρων.

Στην ακόλουθη εικόνα παρουσιάζεται η προσέγγιση του Bayesian optimization συγκριτικά με την αναζήτηση πλέγματος.



Εικόνα 21. Σύγκριση αναζήτησης πλέγματος και Bayesian optimization [56].

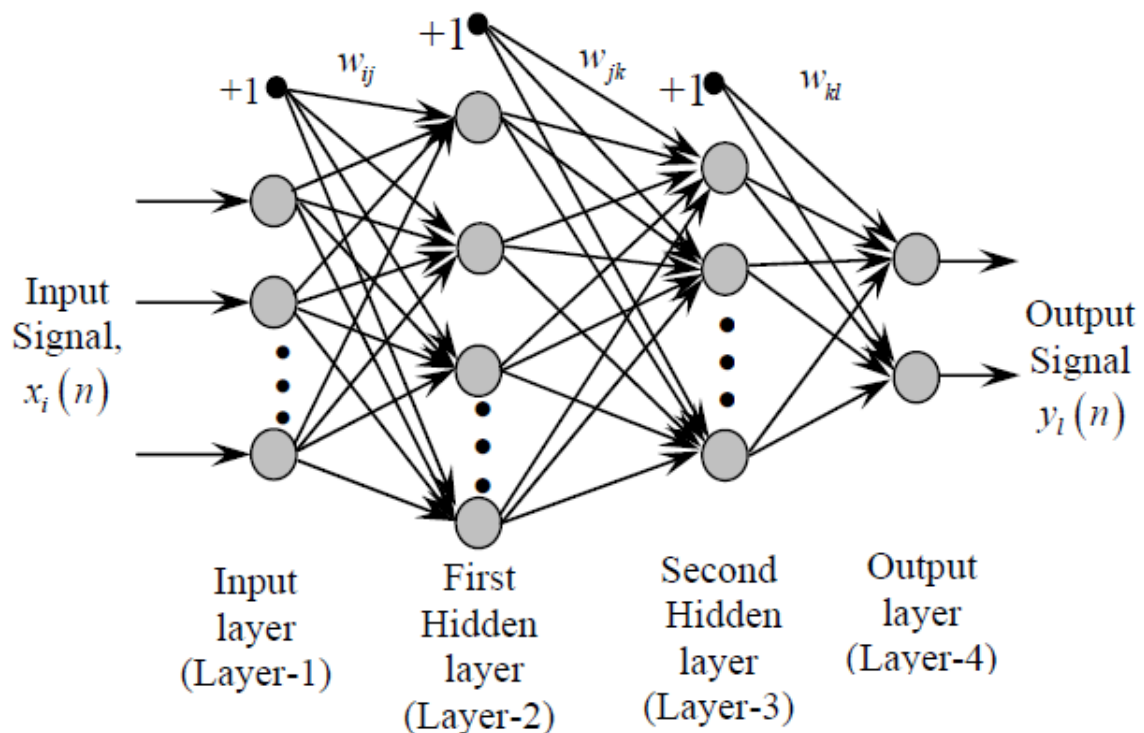
4.2 Multi Layer Perceptron

Το Multi Layer Perceptron ή αλλιώς MLP είναι η πιο γνωστή και βασική διάταξη νευρωνικών δικτύων και ουσιαστικά αποτελεί σύνθεση των απλών νευρώνων που παρουσιάστηκαν στις προηγούμενες ενότητες σε πολλαπλά αλληλένδετα επίπεδα. Ο όρος χρησιμοποιείται γενικά με διαφορετικό τρόπο και πολλές φορές γίνεται χρήση του, προκειμένου να περιγραφεί οποιοδήποτε τεχνητό νευρωνικό δίκτυο πρόσθιας τροφοδότησης, ενώ άλλες φορές για να περιγραφούν αρχιτεκτονικές δικτύων που αποτελούνται από πολλαπλά επίπεδα νευρώνων.

Ένα MLP περιλαμβάνει τουλάχιστον τρία επίπεδα κόμβων, ένα επίπεδο εισόδου, ένα κρυμμένο επίπεδο κι ένα επίπεδο εξόδου. Εκτός των κόμβων της εισόδου οι υπόλοιποι συνιστούν έναν νευρώνα ο οποίος χρησιμοποιεί μία συνάρτηση ενεργοποίησης, συνήθως μη γραμμικής.

Για την εκπαίδευση του MLP χρησιμοποιείται η τεχνική backpropagation όπως περιγράφηκε στην προηγούμενη ενότητα. Κάθε νευρώνας πυροδοτείται από κάποια συνάρτηση ενεργοποίησης, όπως αυτές που παρουσιάστηκαν στην αντίστοιχη ενότητα, με πιο συχνά χρησιμοποιούμενη τα τελευταία χρόνια τη συνάρτηση ReLU με τη χρήση της οποίας αποφεύγονται τα αριθμητικά προβλήματα που εμφανίζουν οι σιγμοειδείς συναρτήσεις.

Η αρχιτεκτονική του MLP παρουσιάζεται στο ακόλουθο σχήμα.



Εικόνα 22. Η δομή ενός Multi Layer Perceptron [57]

Χαρακτηριστικό της παραπάνω αρχιτεκτονικής αποτελεί το γεγονός ότι οι έξοδοι όλων των νευρώνων της εισόδου καθώς και των ενδιάμεσων (γνωστών και ως κρυφά επίπεδα) επιπέδων

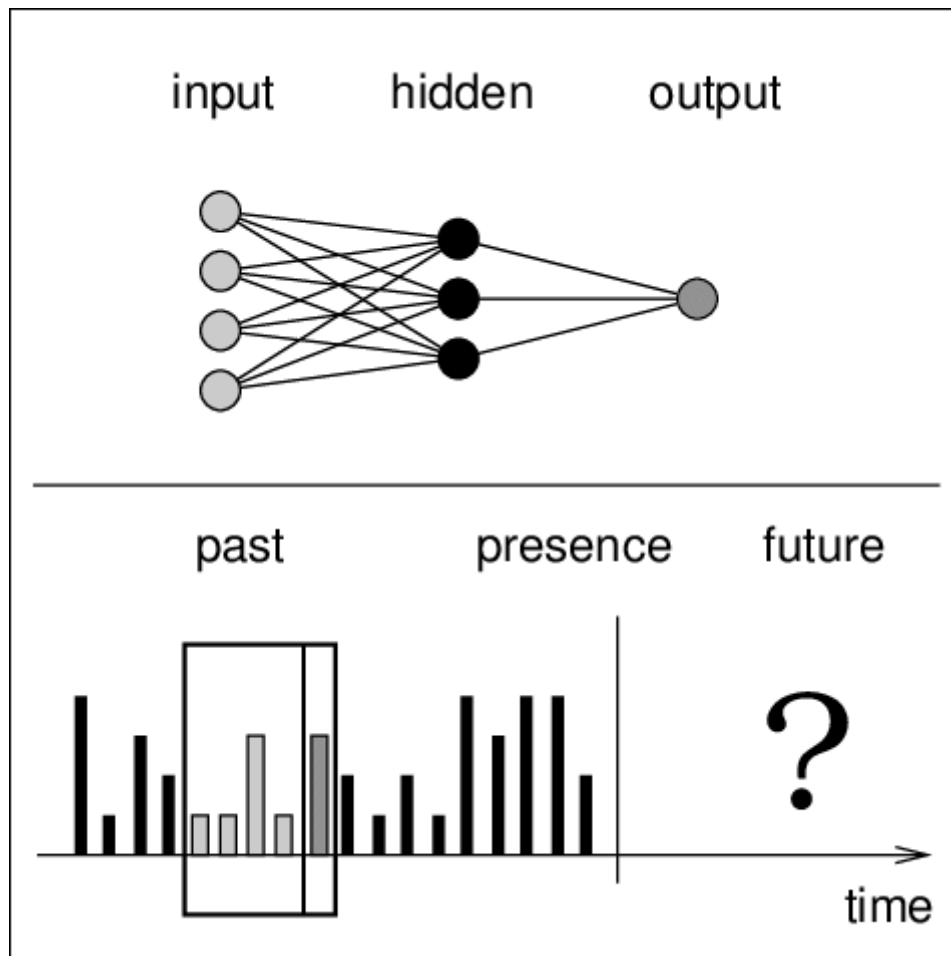
συνδέονται με τις εισόδους των νευρώνων του επόμενου επιπέδου. Εξ ου και η συνήθης ονομασία τους ως πυκνά (dense) επίπεδα. Στην περίπτωση που περιλαμβάνουν μόνον ένα κρυφό επίπεδο, αναφέρονται συχνά και ως vanilla neural networks [58].

Παρατηρούμε στην εικόνα 22, ότι οι νευρώνες πλην αυτών της εισόδου, λαμβάνουν και μία είσοδο η οποία δεν εξαρτάται από τα δεδομένα, γνωστή ως bias και η οποία πολλαπλασιάζεται επίσης με ένα συντελεστή βάρους, υπολογιζόμενο κατά τη διαδικασία εκπαίδευσης όπως και όλοι οι άλλοι συντελεστές των εισόδων.

Τα πολλαπλά επίπεδα καθώς και η μη γραμμική συνάρτηση ενεργοποίησης των MLP που τα διακρίνουν από το απλό γραμμικό perceptron έχουν ως αποτέλεσμα το να είναι σε θέση το MLP να μπορεί να διακρίνει δεδομένα τα οποία δεν είναι γραμμικώς διαχωρίσιμα. Συγκεκριμένα, όπως έχει αποδειχθεί από το θεώρημα Cybenko [59], έχουν την ιδιότητα να μπορούν να προσομοιώσουν οποιαδήποτε συνάρτηση (universal function approximators).

Τα MLP αποτέλεσαν εξαιρετικά δημοφιλή λύση κατά την δεκαετία του 80, με εφαρμογές σε πεδία όπως αναγνώριση ομιλίας και εικόνας, καθώς και αυτόματης μετάφρασης [58], αλλά έκτοτε παραγκωνίστηκαν από τις πολύ πιο απλές και παρόμοιες στη χρήση τους μηχανές διανυσμάτων υποστήριξης (support vector machines). Τα τελευταία ωστόσο χρόνια το ενδιαφέρον προς αυτά έχει αναζωπυρωθεί δεδομένων και των μεγάλων επιτυχιών που έχει σημειώσει η βαθιά μηχανική μάθηση.

Η χρήση των Multi Layer Perceptron παρ' ότι δεν είναι πολύ διαδεδομένη στο χώρο των προβλέψεων χρονοσειρών δεδομένου ότι υπάρχουν αρχιτεκτονικές αναδρομικών νευρωνικών δικτύων (RNN) που θεωρούνται καταλληλότερες γι' αυτό το σκοπό, έχει εν τούτοις δοκιμαστεί, και μάλιστα ένα MLP χρησιμοποιήθηκε και ως benchmark στον διαγωνισμό M4. Στην ακόλουθη εικόνα παρουσιάζεται η μεθοδολογία πρόβλεψης μίας univariate χρονοσειράς με τη χρήση MLP.



Εικόνα 23. Χρήση MLP για την πρόβλεψη χρονοσειρών [60].

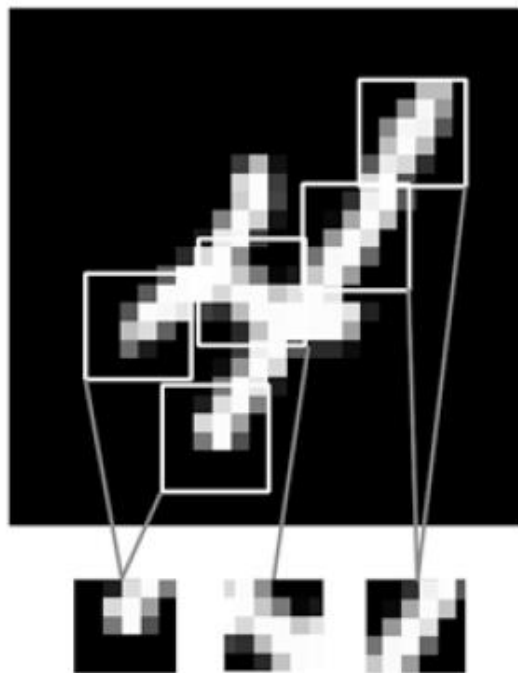
Όπως παρατηρούμε στην εικόνα, το δίκτυο δέχεται ως είσοδο 4 παρελθοντικές τιμές της χρονοσειράς, και δίνει ως έξοδο την πρόβλεψη για την αμέσως επόμενη χρονική περίοδο. Το σύνολο δεδομένων για την εκπαίδευσή του, δημιουργείται από το σύνολο των παραθύρων 5 συνεχόμενων τιμών που περιέχονται στα ιστορικά δεδομένα, εκ των οποίων οι 4 πρώτες τιμές αποτελούν την είσοδο και η 5^η την τιμή στόχο.

4.3 Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)

Τα συνελικτικά νευρωνικά δίκτυα αποτελούν μία εξειδικευμένη κατηγορία νευρωνικών δικτύων με σκοπό την επεξεργασία δεδομένων που έχουν γνωστή τοπολογία τύπου πλέγματος. Παραδείγματα τέτοιου τύπου δεδομένων αποτελούν οι χρονοσειρές, τις οποίες μπορούμε να αντιληφθούμε σαν ένα μονοδιάστατο πλέγμα το οποίο αποτελείται από δείγματα παρμένα σε τακτά χρονικά διαστήματα, καθώς και οι εικόνες οι οποίες μπορούν να γίνουν αντιληπτές ως πλέγματα δύο διαστάσεων που αποτελούνται από pixels. Το όνομά τους

υπονοεί πως τα δίκτυα αυτά κάνουν χρήση της μαθηματικής λειτουργίας που ονομάζεται συνέλιξη, και η οποία είναι μία εξειδικευμένη μορφή γραμμικού μετασχηματισμού. Τα συνελκτικά δίκτυα λοιπόν, μπορούν να οριστούν ως νευρωνικά δίκτυα που κάνουν χρήση της συνέλιξης, αντί του απλού πολλαπλασιασμού πινάκων, σε τουλάχιστον ένα από τα επίπεδά τους [61].

Η θεμελιώδης διαφορά μεταξύ ενός πυκνά συνδεδεμένου δικτύου και ενός συνελκτικού δικτύου είναι ότι τα πυκνά νευρωνικά δίκτυα (όπως τα MLP) μαθαίνουν μοτίβα από το σύνολο των χαρακτηριστικών εισόδου (global patterns), ενώ τα συνελκτικά δίκτυα είναι σε θέση να μαθαίνουν τοπικά μοτίβα (local patterns) [32]. Στην περίπτωση που τα δεδομένα μας αποτελούνται από εικόνες, τα τοπικά αυτά μοτίβα εντοπίζονται σε μικρά παράθυρα δύο διαστάσεων. Αντίστοιχα στην περίπτωση των χρονοσειρών, τα παράθυρα αφορούν μία μικρή περιοχή της χρονοσειράς. Στην επόμενη εικόνα παρουσιάζεται η απομόνωση τοπικών μοτίβων από ένα συνελκτικό νευρωνικό δίκτυο που διενεργεί αναγνώριση εικόνας.



Εικόνα 24. Αποδόμηση εικόνας σε τοπικά μοτίβα όπως ακμές, υφές κλπ. [32]

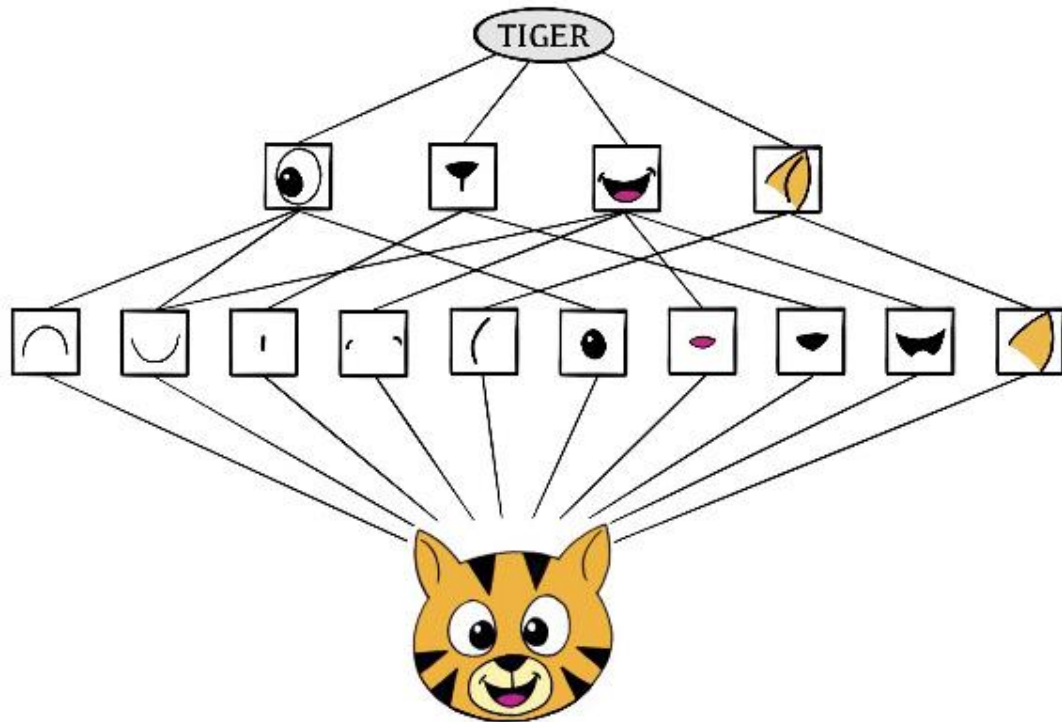
Θα μπορούσαμε να πούμε ότι ως ένα βαθμό, τα CNNs προσομοιάζουν τον τρόπο με τον οποίο ο ανθρώπινος εγκέφαλος ταξινομεί τις εικόνες, αναγνωρίζοντας συγκεκριμένα

χαρακτηριστικά ή μοτίβα σε όλο το εύρος της εικόνας βάση των οποίων καθίσταται διακριτή η κάθε κλάση [42].

Δύο πολύ βασικά χαρακτηριστικά των εν λόγω δικτύων είναι τα εξής:

- Αφότου “μάθουν” ένα συγκεκριμένο τοπικό μοτίβο, μπορούν να το αναγνωρίσουν σε οποιοδήποτε άλλο σημείο, εν αντιθέσει με τα πυκνά συνδεδεμένα νευρωνικά δίκτυα τα οποία αν το μοτίβο εμφανιστεί σε μία άλλη θέση, πρέπει να το ξαναμάθουν από την αρχή. Για παράδειγμα σε μία εικόνα, αν αναγνωρίσουν μία ακμή στο άνω αριστερό μέρος της, τότε μπορούν πλέον να το αναγνωρίζουν και σε οποιοδήποτε άλλο σημείο της. Η ιδιότητα αυτή είναι πολύ σημαντική καθώς τα καθιστά ικανά να εκπαιδούνται με μικρότερο αριθμό δειγμάτων [32].
- Είναι σε θέση να μαθαίνουν ιεραρχίες μοτίβων: Έτσι σε μία εικόνα, θα μπορούσε το πρώτο συνελικτικό επίπεδο να μάθει να αναγνωρίζει μικρά τοπικά μοτίβα, όπως είναι οι ακμές, ένα δεύτερο επίπεδο μπορεί να μάθει πιο σύνθετα μοτίβα που σχηματίζονται από τα απλά του πρώτου επιπέδου κ.ο.κ. Αυτή είναι και η ιδιότητα που τους επιτρέπει να μαθαίνουν με αποδοτικό τρόπο ολοένα και πιο σύνθετες και αφηρημένες έννοιες [32].

Η δεύτερη αυτή ιδιότητα παρουσιάζεται σχηματικά στη ακόλουθη εικόνα, η οποία παρουσιάζει τον τρόπο με τον οποίο ένα συνελικτικό δίκτυο προκειμένου να αναγνωρίσει και να κατηγοριοποιήσει την εικόνα μίας τίγρης, μαθαίνει στο πρώτο επίπεδο τα στοιχειώδη σχήματα που εμφανίζονται σε αυτήν, κατόπιν τα συνδυάζει στο δεύτερο επίπεδο προκειμένου να αναγνωρίζει τα μέρη του σώματός (μάτια, μύτη, στόμα) και εν τέλει καταλήγει στην κατηγοριοποίηση.



Εικόνα 25. Κατηγοριοποίηση εικόνας με CNN με την ιεραρχική εκμάθηση ολόενα και πιο σύνθετων χαρακτηριστικών από τα διαδοχικά επίπεδα του δικτύου [42]

4.3.1 Η συνέλιξη

Ένα τυπικό συνελικτικό επίπεδο αποτελείται από ένα μεγάλο αριθμό φίλτρων συνέλιξης. Κάθε φίλτρο αποτελεί και ένα πρότυπο βάσει του οποίου καθορίζεται αν ένα συγκεκριμένο τοπικό μοτίβο εμφανίζεται ή όχι στο δείγμα. Η μαθηματική λειτουργία που επιτελείται, ονομάζεται όπως προαναφέρθηκε συνέλιξη και η οποία συνίσταται στον επαναλαμβανόμενο πολλαπλασιασμό μεταξύ των στοιχείων πινάκων και πρόσθεσης των αποτελεσμάτων.

Για την κατανόηση του πώς λειτουργεί το φίλτρο της συνέλιξης, παραθέτουμε το ακόλουθο απλό παράδειγμα μίας εικόνας διαστάσεων 4 x 3 [42].

$$\text{Αρχική εικόνα} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \\ j & k & l \end{bmatrix}$$

Θεωρούμε επίσης ένα φίλτρο 2 x 2 το οποίο έχει την ακόλουθη μορφή.

$$\text{Φίλτρο συνέλιξης} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$$

Μετά την διενέργεια της συνέλιξης, θα έχουμε το ακόλουθο αποτέλεσμα:

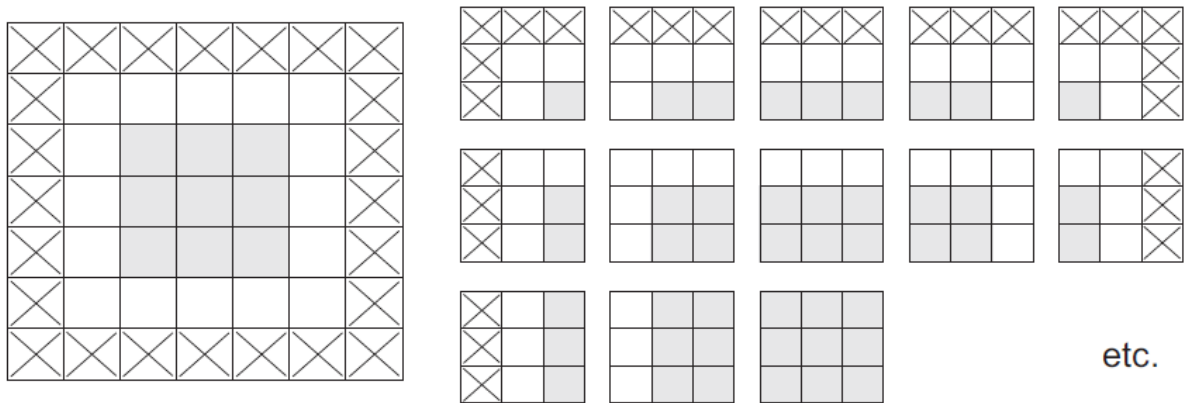
$$\text{Συνελιγμένη εικόνα} = \begin{bmatrix} a\alpha + b\beta + d\gamma + e\delta & b\alpha + c\beta + e\gamma + f\delta \\ d\alpha + e\beta + g\gamma + h\delta & e\alpha + f\beta + h\gamma + i\delta \\ g\alpha + h\beta + j\gamma + k\delta & h\alpha + i\beta + k\gamma + l\delta \end{bmatrix}$$

Έτσι για παράδειγμα το πάνω αριστερά στοιχείο της συνελιγμένης εικόνας, προέρχεται από τον πολλαπλασιασμό του κάθε στοιχείου του φίλτρου συνέλιξης, με το αντίστοιχο στοιχείο του πάνω αριστερά τμήματος της αρχικής εικόνας με διαστάσεις 2×2 και πρόσθεση των αποτελεσμάτων. Τα υπόλοιπα στοιχεία προέρχονται με την μετατόπιση του 2×2 τμήματος της αρχικής εικόνας το οποίο συνελίσσεται, ούτως ώστε να ληφθούν υπόψη και να συνελιχθούν όλες οι 2×2 περιοχές της αρχικής εικόνας. Με αυτόν τον τρόπο η συνελιγμένη εικόνα καταλήγει να έχει διαστάσεις 3×2 .

4.3.2 Η έννοια του γεμίσματος (padding)

Όπως παρατηρήσαμε στο προηγούμενο παράδειγμα, η διενέργεια της συνέλιξης είχε ως αποτέλεσμα τη συρρίκνωση του χάρτη χαρακτηριστικών (feature map) δηλαδή της αρχικής εικόνας. Η συρρίκνωση έγινε κατά μία μονάδα στις γραμμές και μία μονάδα στις στήλες. Αντίστοιχα, αν το φίλτρο ήταν μεγέθους 3×3 θα είχαμε συρρίκνωση κατά 2 μονάδες στις γραμμές και 2 μονάδες στις στήλες κ.ο.κ.

Στην περίπτωση που επιθυμούμε το αποτέλεσμα της συνέλιξης να έχει τις ίδιες διαστάσεις με αυτές της αρχικής εικόνας τότε μπορούμε να χρησιμοποιήσουμε την διαδικασία του padding. Αυτή, συνίσταται στην προσθήκη του κατάλληλου αριθμού γραμμών και στηλών σε κάθε πλευρά του χάρτη χαρακτηριστικών, προκειμένου να καθίσταται δυνατό το κεντράρισμα του φίλτρου σε κάθε ένα από τα στοιχεία της εικόνας μας [32]. Η διαδικασία παρουσιάζεται σχηματικά στην κάτωθι εικόνα, όπου έχουμε έναν αρχικό χάρτη χαρακτηριστικών διαστάσεων 5×5 και εφαρμόζουμε ένα φίλτρο διαστάσεων 3×3 . Χωρίς τη χρήση του padding, η τελική διάσταση μετά τη συνέλιξη θα ήταν 3×3 . Αντιθέτως, με τη χρήση του padding έχουμε τη δυνατότητα να διατηρήσουμε τη διάσταση 5×5 προσθέτοντας συμμετρικά, περιμετρικά της εικόνας 2 γραμμές και δύο στήλες.

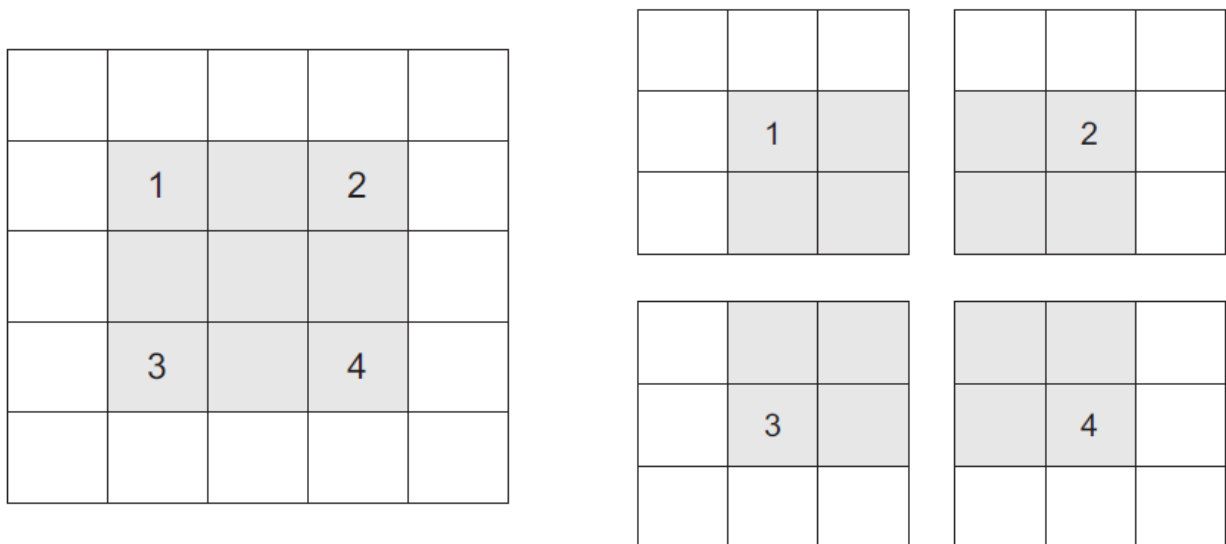


Εικόνα 26. Διατήρηση των διαστάσεων 5 x 5 του χάρτη χαρακτηριστικών με τη χρήση padding [32].

Τα τετράγωνα που είναι επισημασμένα με X, είναι αυτά που προστέθηκαν με το padding.

4.3.3 Η έννοια του βήματος συνέλιξης (convolution stride)

Ένας ακόμα παράγοντας που επηρεάζει τις διαστάσεις της εξόδου είναι το βήμα συνέλιξης. Μέχρι στιγμής στα παραδείγματα που παρουσιάστηκαν, η μετακίνηση του φίλτρου κατά το πλάτος και το ύψος της εικόνας γίνεται μετακινώντας το, κατά ένα στοιχείο κάθε φορά. Η απόσταση μεταξύ δύο διαδοχικών τμημάτων που αποκόπτονται από την εικόνα προκειμένου να συνελιχθούν με το φίλτρο, αποτελεί το βήμα της συνέλιξης [32]. Στις περιπτώσεις που συναντήσαμε, η τιμή του βήματος αυτού ισούται με 1. Είναι όμως δυνατό να χρησιμοποιήσουμε και τιμές μεγαλύτερες του 1 όπως φαίνεται στο παρακάτω σχήμα.

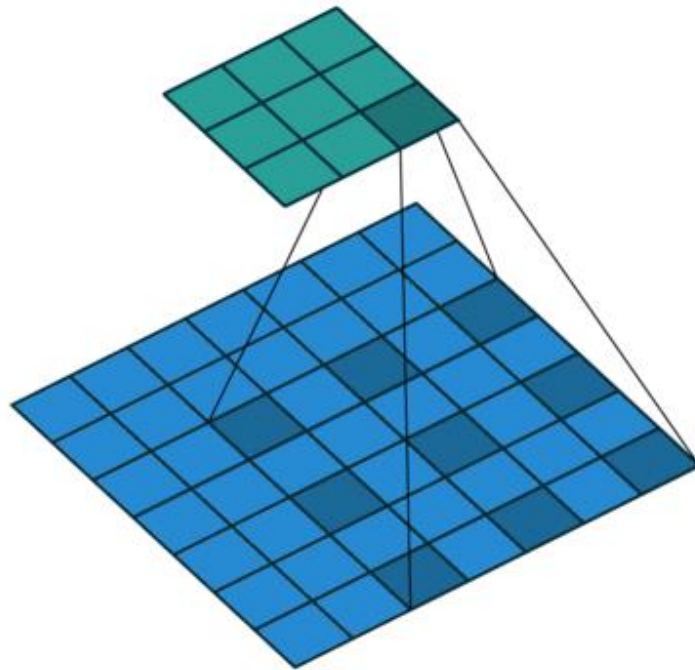


Εικόνα 27. Συνέλιξη με διαστάσεις φίλτρου 3 x 3 και βήμα (stride) με διαστάσεις 2 x 2 [32]. Η διάσταση της εξόδου θα είναι 2 x 2.

Ουσιαστικά χρησιμοποιώντας τιμή βήματος ίση με 2, γίνεται υποδειγματοληψία του χάρτη χαρακτηριστικών κατά ένα παράγοντα ίσο με 2. Γενικότερα πάντως, οι περιπτώσεις στις οποίες χρησιμοποιείται η τεχνική του βήματος συνέλιξης είναι περιορισμένες [32].

4.3.4 Βαθμός διαστολής (dilation rate)

Μία ακόμα παράμετρος που μπορεί να τεθεί στα συνελκτικά επίπεδα, αποτελεί ο βαθμός διαστολής ή dilation rate. Η παράμετρος αυτή, καθορίζει την ύπαρξη ενός διάκενου μεταξύ των τιμών του χάρτη χαρακτηριστικών, που λαμβάνονται προκειμένου να προκύψει το τμήμα του, που θα συνελίσσεται κάθε φορά με το φίλτρο. Έτσι για παράδειγμα ένα φίλτρο διαστάσεων 3 x 3 με ένα dilation rate 2, θα «εποπτεύει» την ίδια έκταση του χάρτη με ένα φίλτρο διαστάσεων 5 x 5, χρησιμοποιώντας όμως μόνο 9 παραμέτρους. Ουσιαστικά υλοποιείται λαμβάνοντας ένα φίλτρο διαστάσεων 5 x 5 και διαγράφοντας κάθε δεύτερη σειρά και στήλη.



Εικόνα 28. Συνέλιξη δύο διαστάσεων με τη χρήση φίλτρου διάστασης 3×3 και dilation rate 2 χωρίς padding [62].

Με αυτόν τον τρόπο, προκύπτει ένα μεγαλύτερο εύρος εποπτείας του φίλτρου, με το ίδιο υπολογιστικό κόστος [62].

Η τεχνική αυτή μπορεί να γίνει ιδιαίτερος χρήσιμη για κάποιους τύπους μοντέλων, όπως είναι τα χρονικά (temporal) συνελκτικά νευρωνικά δίκτυα που θα συναντήσουμε στη συνέχεια.

4.3.5 Η διαδικασία επιλογής μεγίστων (max pooling)

Μία συνηθισμένη πρακτική, μετά την εφαρμογή ενός συνελκτικού επιπέδου είναι η εφαρμογή του max pooling. Στην περίπτωση αυτή εφαρμόζουμε ένα επίπεδο επιλογής (pooling layer), το οποίο επί της ουσίας παρέχει ένα τρόπο συμπύκνωσης ενός μεγάλου χάρτη χαρακτηριστικών σε ένα μικρότερο «περιληπτικό» χάρτη. Υπάρχουν αρκετοί διαφορετικοί τρόποι για να πραγματοποιηθεί μία τέτοια συμπύκνωση αλλά ένας πολύ συνηθισμένος τρόπος είναι να επιλεγούν οι μέγιστες τιμές από όλα τα μη επικαλυπτόμενα τμήματα ενός χάρτη που έχουν μία προκαθορισμένη διάσταση, π.χ. 2×2 .

Ένα παράδειγμα της διαδικασίας φαίνεται ακολούθως.

$$\text{Επιλογή μεγίστου} \begin{bmatrix} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 6 \\ 7 & 8 & 8 & 3 \\ 9 & 10 & 6 & 12 \end{bmatrix} \rightarrow \begin{bmatrix} 4 & 6 \\ 10 & 12 \end{bmatrix}$$

Το παραγόμενο, είναι ένας ελαττωμένος χάρτης χαρακτηριστικών κατά έναν παράγοντα ίσο με 2 σε κάθε διάσταση. Παρέχεται με αυτό τον τρόπο η ιδιότητα της μη χωρικής αλλοίωσης (location invariance), ήτοι, εφ' όσον υπάρχει μία μεγάλη τιμή στο μπλοκ που εξετάζουμε, τότε το σύνολο του μπλοκ ανάγεται σε αυτήν την μεγάλη τιμή στον τελικό «περιληπτικό» χάρτη [42].

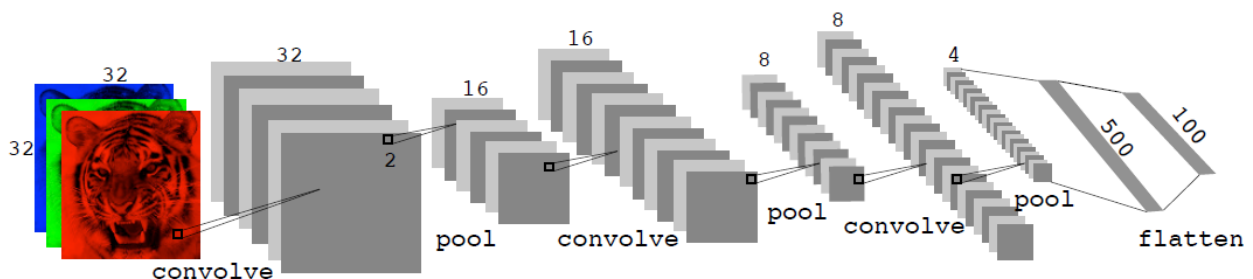
4.3.6 Επιπέδωση (flattening)

Μέχρι στιγμής οι χάρτες χαρακτηριστικών που συναντήσαμε είναι δύο διαστάσεων, ενώ σε άλλες περιπτώσεις μπορούν να είναι τριών διαστάσεων κ.λ.π. Ωστόσο, στις αρχιτεκτονικές των συνελκτικών δικτύων, οι χάρτες αυτοί, πριν μας δώσουν την τελική ζητούμενη έξοδο, εισάγονται σε ένα ή και περισσότερα πυκνά επίπεδα, τα οποία θα παρέχουν ένα ακόμα στάδιο εκπαίδευσης σύμφωνα με αυτά που περιγράφηκαν στην ενότητα των Multi Layer Perceptrons.

Για να επιτευχθεί αυτό, οι χάρτες πρέπει να μετατραπούν σε μονοδιάστατους χάρτες, και το ρόλο αυτό αναλαμβάνουν τα λεγόμενα επίπεδα επιπέδωσης (flattening layers).

4.3.7 Πανόραμα της αρχιτεκτονικής ενός συνελκτικού νευρωνικού δικτύου.

Όλα τα βήματα και οι διαδικασίες που περιγράψαμε παραπάνω, συντίθεται σε μία συνολική αρχιτεκτονική η οποία στην περίπτωση κατηγοριοποίησης εικόνας παρουσιάζεται εποπτικά ακολούθως.

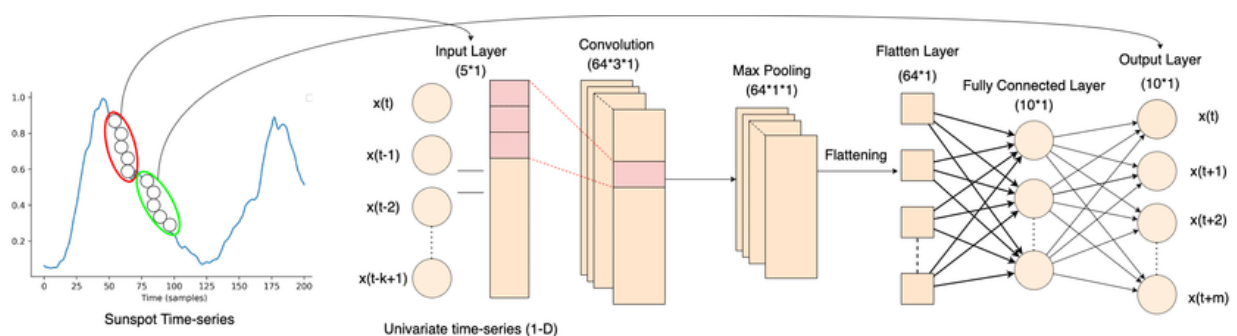


Εικόνα 29. Κατηγοριοποίηση εικόνας με τη χρήση συνελκτικών νευρωνικών δικτύων [42]

Όπως βλέπουμε, το επίπεδο εισόδου αποτελείται από έναν χάρτη χαρακτηριστικών τριών διαστάσεων (Έχουμε 3 κανάλια RGB μίας εικόνας διαστάσεων 32 x 32 pixels και συνεπώς έναν χάρτη χαρακτηριστικών διαστάσεων (32 x 32 x 3). Ο χάρτης αυτός εισάγεται διαδοχικά σε 3 συνελκτικά επίπεδα, το καθένα εκ των οποίων ακολουθείται από ένα επίπεδο επιλογής (max pooling) που συμπυκνώνει την έξοδο των συνελκτικών επιπέδων κατά έναν παράγοντα ίσο με 2.

Ακολουθεί η επιπέδωση, μέσω της οποίας τα pixels πλέον αντιμετωπίζονται ως μεμονωμένες τιμές και εισάγονται σε πλήρως συνδεδεμένα επίπεδα (fully connected layers) προτού φτάσουν στο επίπεδο εξόδου το οποίο κατηγοριοποιεί την εικόνα μας σε μία από τις 100 πιθανές κατηγορίες.

Θα παραθέσουμε ένα ακόμα παράδειγμα της αρχιτεκτονικής, προκειμένου να γίνει αντιληπτό πως τα παραπάνω εφαρμόζονται σε ένα μονοδιάστατο χώρο και συνεπώς και στις περιπτώσεις των χρονολογικών σειρών που αποτελούν το κατ' εξοχήν αντικείμενο της παρούσας εργασίας.



Εικόνα 30. Μονοδιάστατο συνελκτικό νευρωνικό δίκτυο για πρόβλεψη πολλαπλών βημάτων χρονοσειράς [63]

Όπως βλέπουμε από το σχήμα, ζητούμενο του προβλήματος είναι η πρόβλεψη 5 χρονικών στιγμών μετά την τρέχουσα χρονική στιγμή, αξιοποιώντας τις τιμές της χρονοσειράς την τρέχουσα χρονική στιγμή καθώς και των τεσσάρων προηγούμενων τιμών της σειράς.

Η είσοδος, που έγκειται σε έναν χάρτη χαρακτηριστικών διαστάσεων 5 x 1 τροφοδοτείται στο επίπεδο συνέλιξης και με χρήση 64 φίλτρων και διάσταση φίλτρου 3, έχουμε μία έξοδο διαστάσεων (64 x 3 x 1). Αυτή η έξοδος τροφοδοτείται σε ένα επίπεδο επιλογής μεγίστου στο οποίο «συμπυκνώνεται» σε ένα χώρο διαστάσεων (64 x 1 x 1), δηλαδή από κάθε

συνελισσόμενο παράθυρο διαστάσεων (3 x 1) επιλέγεται η μέγιστη τιμή του. Κατόπιν έχουμε το επίπεδο flattening, απ' όπου παίρνουμε μία έξοδο διαστάσεων (64 X 1) και τέλος, τροφοδοτούμε αυτές τις μεμονωμένες πλέον τιμές σε ένα πλήρως συνδεδεμένο επίπεδο (fully connected layer) λαμβάνοντας ένα επίπεδο εξόδου που μας δίνει 5 τιμές (5 νευρώνες εξόδου) κάθε μία εκ των οποίων αντιστοιχεί σε μία από τις τιμές του ορίζοντα που θέλουμε να προβλέψουμε και που αντιστοιχούν στις επισημασμένες με πράσινο χρώμα στην εικόνα.

4.4 Χρονικά (temporal) συνελικτικά δίκτυα (TCN)

Κλείνοντας την παρουσίαση των νευρωνικών δικτύων, θεωρούμε σκόπιμο να παρουσιάσουμε την ειδική διάταξη συνελικτικών δικτύων που έχει καθιερωθεί να αποκαλούνται «χρονικά» (temporal) εξαιτίας του γεγονότος ότι η χρήση τους ενδείκνυται στην μελέτη χρονολογικών φαινομένων (συμπεριλαμβανομένων των χρονοσειρών).

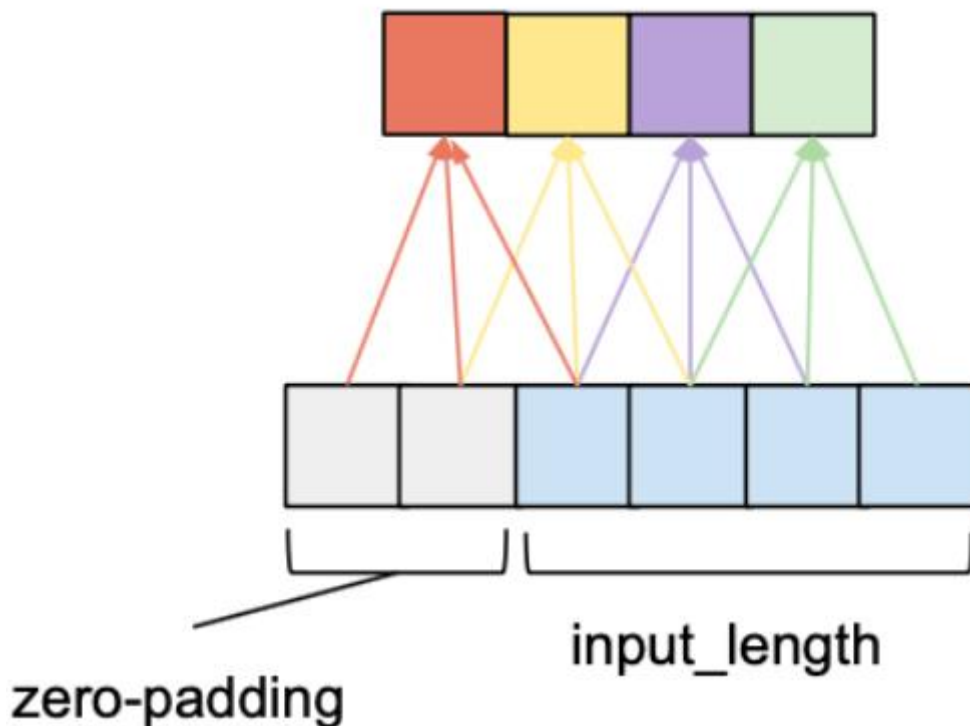
Μέχρι πρόσφατα, δύο είδη αρχιτεκτονικών αναδρομικών νευρωνικών δικτύων, τα LSTM και τα GRUs είχαν συσχετιστεί σε μεγάλο βαθμό με την μοντελοποίηση ακολουθιακών δεδομένων όπως είναι οι χρονοσειρές. Ωστόσο, πρόσφατες μελέτες θεωρούν αυτόν τον τρόπο σκέψης απαρχαιωμένο και ότι πλέον τα συνελικτικά δίκτυα θα έπρεπε να θεωρούνται από τους κύριους υποψήφιους, όταν έχουμε να κάνουμε με τέτοιου είδους δεδομένα. Αποδείχθηκε στην πορεία, πως τα συνελικτικά δίκτυα, μπορούν να έχουν καλύτερα αποτελέσματα σε σχέση με τα αναδρομικά, σε πολλές περιπτώσεις, αποφεύγοντας ταυτόχρονα πολλές από τις «παγίδες» στις οποίες υποπίπτουν συνήθως τα τελευταία. Επιπλέον, η χρήση των συνελικτικών δικτύων μπορεί να οδηγήσει σε καλύτερη χρονική απόδοση, αφού επιτρέπει την παραλληλοποίηση στον υπολογισμό των εξόδων [64]. Η προτεινόμενη αρχιτεκτονική, ονομάζεται Χρονικό Συνελικτικό Δίκτυο (Temporal Convolutional Network).

4.4.1 Αιτιατή συνέλιξη (Causal Convolution)

Προκειμένου ένα συνελικτικό επίπεδο να είναι αιτιατό, θα πρέπει κάθε στοιχείο της εξόδου, να εξαρτάται μόνο από στοιχεία που βρίσκονται πριν από αυτό στην ακολουθία εισόδου. Όπως αναφέραμε, προκειμένου να διασφαλιστεί ότι η έξοδος θα έχει τις ίδιες διαστάσεις με την είσοδο, συνήθως εφαρμόζουμε την τεχνική του padding, που συνήθως γίνεται με συμμετρικό τρόπο και στην αρχή και στο τέλος της υπό μελέτη χρονοσειράς. Αντιθέτως, για να επιτευχθεί η αιτιατότητα του δικτύου, χρειάζεται απλώς να εφαρμόσουμε padding μόνο στην αριστερή πλευρά της χρονοσειράς [64]. Δεδομένου λοιπόν, ότι δεν υπάρχει γέμισμα

στην δεξιά πλευρά, το τελευταίο στοιχείο από το οποίο εξαρτάται η έξοδος του επιπέδου, είναι και το τελευταίο στοιχείο της εισόδου. Αντίστοιχα, το προτελευταίο στοιχείο της εξόδου, εφόσον προκύπτει από συνέλιξη τμήματος της χρονοσειράς μετατοπισμένης κατά μία παρατήρηση προς τα αριστερά σε σχέση με αυτό που χρησιμοποιήθηκε για το τελευταίο στοιχείο της εξόδου, θα εξαρτάται από στοιχεία το τελευταίο εκ των οποίων μπορεί να είναι το προτελευταίο στοιχείο της εισόδου, κ.ο.κ.

Αυτό παρουσιάζεται σχηματικά στην ακόλουθη εικόνα :

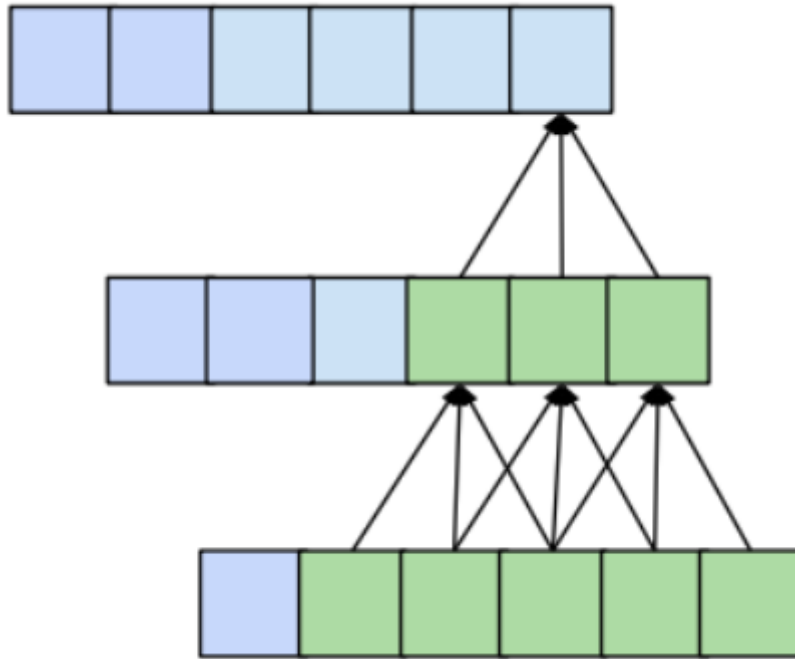


Εικόνα 31. Διασφάλιση της αιτιατότητας με padding μόνο στην αριστερή πλευρά της χρονοσειράς [64].

4.4.2 Εκμετάλλευση του dilation rate

Μία εκ των επιθυμητών ιδιοτήτων του μοντέλου πρόβλεψης, είναι ότι μία συγκεκριμένη τιμή της εξόδου, πρέπει να εξαρτάται από όλες τις προηγούμενες τιμές τις εισόδου (δηλαδή όλες τις τιμές που έχουν δείκτη μικρότερο ή ίσο από αυτήν) [64]. Αυτό μπορεί να επιτευχθεί, όταν το πεδίο λήψης (receptive field), ήτοι το σύνολο τιμών της εισόδου που επηρεάζουν μία συγκεκριμένη τιμή της εξόδου, έχει μέγεθος όσο και το μέγεθος της εισόδου. Αυτή η ιδιότητα αποκαλείται «πλήρης κάλυψη της ιστορικότητας» [64]. Όπως είδαμε στα προηγούμενα, στις συμβατικές αρχιτεκτονικές συνελκτικών δικτύων, κάθε τιμή της εξόδου, επηρεάζεται από

πλήθος τιμών εισόδου ίσο με το μέγεθος του φίλτρου (kernel size). Έτσι για παράδειγμα, στην παραπάνω εικόνα, το στοιχείο 3 της εξόδου εξαρτάται μόνο από τα στοιχεία 1, 2 και 3 της εισόδου κ.ο.κ. Σε περίπτωση που αλληλεπιθέτουμε συνελκτικά επίπεδα, το μέγεθος αυτό του πεδίου λήψης μπορεί να αυξηθεί όπως φαίνεται ακολούθως.



Εικόνα 32. Αύξηση μεγέθους του πεδίου λήψης με χρήση δύο συνελκτικών επιπέδων [64].

Στην γενική περίπτωση, ένα συνελκτικό δίκτυο μονοδιάστατης εισόδου που αποτελείται από n επίπεδα κι έχει μέγεθος πυρήνα ίσο με k , τότε το πεδίο λήψης της εξόδου του θα δίνεται από την εξίσωση [64] :

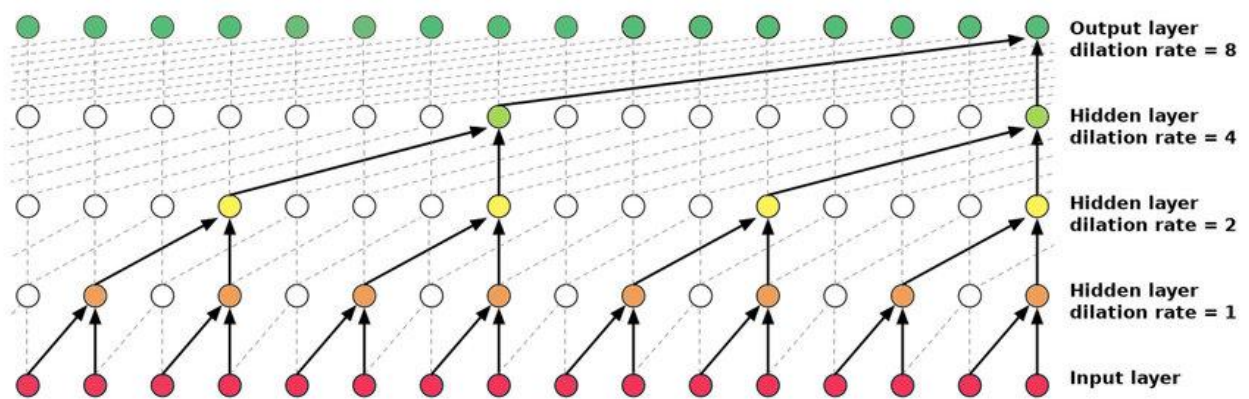
$$r = 1 + n * (k - 1)$$

Κι επομένως για να γνωρίζουμε πόσα συνελκτικά επίπεδα χρειαζόμαστε για να επιτύχουμε την πλήρη κάλυψη, λύνουμε την εξίσωση ως προς n και θα έχουμε:

$$n = (l - 1)/(k - 1)$$

Η παραπάνω σχέση όμως, σημαίνει πως ο αριθμός των επιπέδων που απαιτούνται για την πλήρη κάλυψη, αυξάνεται πολύ γρήγορα και απαιτείται ένας μεγάλος αριθμός παραμέτρων

καθώς επίσης και πολύ μεγάλος χρόνος εκπαίδευσης. Τη λύση στο πρόβλημα αυτό έρχεται να δώσει η χρήση του dilation rate, αυξάνοντας με εκθετικό ρυθμό την τιμή του, καθώς μετακινούμαστε από επίπεδο σε επίπεδο. Ξεκινάμε επιλέγοντας έναν ακέραιο b , ο οποίος θα υψώνεται κάθε φορά στη δύναμη που αντιστοιχεί στον αριθμό του επιπέδου που βρισκόμαστε. Δηλαδή, για το πρώτο επίπεδο θα υψωθεί στην μηδενική δύναμη και θα έχουμε dilation rate ίσο με 1. Ακολούθως αν το b που επιλέχθηκε είναι ίσο με 2, το δεύτερο επίπεδο θα έχει dilation rate ίσο με 2 και το τρίτο επίπεδο 4 κ.ο.κ, όπως φαίνεται στην ακόλουθη εικόνα όπου παρουσιάζεται η αρχιτεκτονική του wavenet, ενός χρονικού συνελκτικού δικτύου το οποίο αναπτύχθηκε από την εταιρεία deep mind, αρχικά με σκοπό να εξυπηρετήσει τις ανάγκες για σύνθεση φωνής.



Εικόνα 33. Η αρχιτεκτονική του wavenet που δημιουργεί ένα χρονικό συνελκτικό δίκτυο [65].

Όπως μπορούμε να αντιληφθούμε από την εικόνα, η τιμή του επάνω δεξιά νευρώνα εξόδου, επηρεάζεται από το σύνολο των τιμών της εισόδου σύμφωνα με τον μηχανισμό που περιγράφηκε. Με τη μεθοδολογία αυτή, προκειμένου να πετύχουμε την πλήρη κάλυψη, χρειαζόμαστε αριθμό επιπέδων που αυξάνονται λογαριθμικά σε σχέση με τον αριθμό των δεδομένων και όχι γραμμικά όπως θα συνέβαινε στην απλή περίπτωση.

Η χρήση του εν λόγω δικτύου (wavenet) έχει δοκιμαστεί στον χώρο των προβλέψεων χρονοσειρών, έχει αποδειχθεί ότι έχει πολύ καλά αποτελέσματα σε προβλήματα που προσομοιάζουν παλινδρόμηση και είναι ικανό να «μαθαίνει» αποτελεσματικά τις εξαρτήσεις που υπάρχουν εντός και ενδιάμεσως μίας χρονοσειράς, χωρίς να υπάρχει ανάγκη για μεγάλο όγκο ιστορικών δεδομένων. Είναι επίσης αποδοτικό από πλευράς υπολογιστικού χρόνου, εύκολο στην υλοποίηση και υπερτερεί σε σχέση με γραμμικά και αναδρομικά μοντέλα [66].

5 Πειραματική διαδικασία

5.1 Εισαγωγή

Σκοπός της παρούσας εργασίας ήταν αφενός η διερεύνηση των δυνατοτήτων των συνελκτικών νευρωνικών δικτύων (Convolutional Neural Networks) στον κλάδο των προβλέψεων και αφετέρου η αξιολόγηση του χρόνου εκπαίδευσης και των υπολογιστικών πόρων που απαιτούν.

Η σύγκριση των επιδόσεών τους, έγινε σε σχέση με καθιερωμένες μεθόδους (MLP, Theta) καθώς και με μεθόδους που διακρίθηκαν στα πλαίσια του διαγωνισμού M4 (ES-RNN που υποβλήθηκε από τον Slawlek Smyl).

Η υλοποίηση των δικτύων, η πραγματοποίηση των πειραμάτων αλλά και η δημιουργία γραφημάτων απεικόνισης των αποτελεσμάτων έγιναν με τη χρήση της γλώσσας python καθώς και βιβλιοθηκών που ανήκουν στο οικοσύστημά της (Pandas, Keras, Matplotlib κ.λ.π).

Τα βασικά βήματα της πειραματικής διαδικασίας που ακολουθήθηκε παρουσιάζονται συνοπτικά:

- *Περιγραφή και ανάλυση των δεδομένων.* Στην ενότητα αυτή παρουσιάζουμε τα δεδομένα που χρησιμοποιήθηκαν, τα οποία αξιολογούνται βάσει 8 κύριων χαρακτηριστικών. Εφαρμόσαμε τη μέθοδο της ανάλυσης σε κύριες συνιστώσες προκειμένου να ελαττώσουμε τον χώρο των διαστάσεων σε 2 και να διευκολυνθεί η οπτική αναπαράστασή τους.
- *Προεπεξεργασία των δεδομένων:* Στο στάδιο αυτό, το σύνολο δεδομένων εκπαίδευσης μετασχηματίστηκε σε μορφή κατάλληλη για να εισαχθεί στα δίκτυα προς εκπαίδευση.
- *Δημιουργία των συνελκτικών δικτύων:* Στο στάδιο αυτό έγιναν οι βασικές επιλογές της δομής και της εκπαίδευσης των δικτύων και δημιουργήθηκε προγραμματιστικά ένα framework προκειμένου να μπορούν να διεξάγονται με αυτοματοποιημένο τρόπο πολλαπλά πειράματα.

Ορίστηκε επίσης ένας χώρος υπερπαραμέτρων (δηλαδή παράμετροι των οποίων οι τιμές δεν μεταβάλλονται κατά την μάθηση), των οποίων ο βέλτιστος καθορισμός διαδραματίζει σημαντικό ρόλο στην τελική ακρίβεια των μοντέλων. Με χρήση συγκεκριμένης μεθοδολογίας (Bayesian optimization) έγινε επανάληψη των

πειραμάτων με διαφορετικές τιμές παραμέτρων, προκειμένου να προσδιοριστούν οι τιμές που επιφέρουν τα βέλτιστα αποτελέσματα.

- *Επιλογή μέτρων αξιολόγησης:* Στο στάδιο αυτό, παράλληλα με την εκπαίδευση των δικτύων, έγινε και η αξιολόγησή τους με την επιλεγθείσα μετρική (sMAPE), ενώ ταυτόχρονα καταγράφηκε και ο χρόνος εκπαίδευσής τους.
- *Επιλογή μεθόδων αναφοράς:* Στο στάδιο αυτό επιλέγονται οι μέθοδοι με τις οποίες θα συγκριθεί το μοντέλο μας και συγκεκριμένα μία καθιερωμένη στατιστική μέθοδος Theta, η νικήτρια μέθοδος του διαγωνισμού M4 (ES-RNN του Slawlek Smyl⁴), καθώς και ένα Multi-Layer Perceptron που εκπαιδεύτηκε ειδικά για τις ανάγκες της παρούσας εργασίας.

5.2 Περιγραφή και ανάλυση του συνόλου δεδομένων

Το σύνολο δεδομένων μελέτης της διπλωματικής αποτελείται από τις 23.000 ετήσιες χρονοσειρές του διαγωνισμού M4. Το μέγεθος των σειρών ποικίλλει με μερικές να έχουν λιγότερες από 20 παρατηρήσεις, ενώ κάποιες άλλες υπερβαίνουν τις 400. Ο ορίζοντας πρόβλεψης (και συνεπώς το σύνολο ελέγχου) όπως προαναφέρθηκε αποτελείται από 6 παρατηρήσεις.

Οι σειρές, έχουν αναλυθεί με βάση 10 βασικά χαρακτηριστικά (που αφορούν την εποχικότητα, την τάση, τη συχνότητα, την αυτοσυσχέτιση κ.λπ.) και αποδόθηκαν τιμές σε καθεμία για τα χαρακτηριστικά αυτά, των οποίων ακολουθεί μια συνοπτική αναφορά:

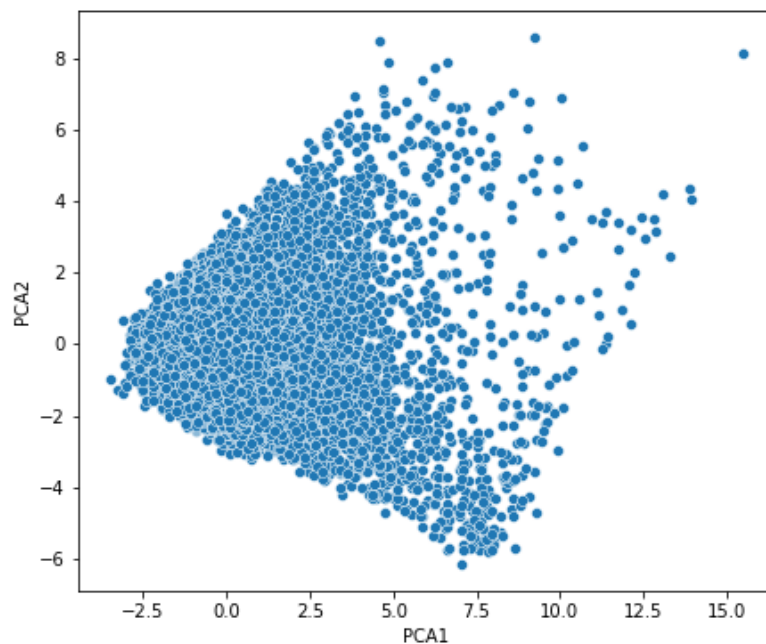
- Εντροπία φάσματος ή Spectral Entropy (**Rlevel**)
- Ένταση τάσης ή Strength of Trend (**Tlevel**)
- Ένταση εποχικότητας ή Strength of seasonality (**Slevel**)
- Συχνότητα ή Frequency (**Flevel**)
- Αυτοσυσχέτιση πρώτης τάξης ή First order auto correlation(**Alevel**)
- Βέλτιστη παράμετρος λάμδα του μετασχηματισμού BoxCox ή Optimal lambda parameter of BoxCox transformation (**Blevel**)
- Ασυμμετρία ή skewness (**SKlevel**)
- Κυρτότητα ή kurtosis (**Klevel**)

⁴ Η μέθοδος αυτή είναι υβριδική, και συνδυάζει αναδρομικά νευρωνικά δίκτυα, με την μέθοδο της εκθετικής εξομάλυνσης.

- Teraesvirta τεστ νευρωνικών δικτύων περί μη γραμμικότητας ή Teraesvirta neural network test for non-linearity (**Nlevel**)
- Παράμετρος κλασματικής διαφοροποίησης ή Fractional differencing parameter (**SSlevel**)

Στις ετήσιες χρονοσειρές, τα χαρακτηριστικά Slevel και Frequency (Εποχικότητα και Συχνότητα) είναι σταθερά και συνεπώς δεν χρησιμοποιήθηκαν στα πλαίσια της παρούσας μελέτης.

Προκειμένου να ελαττώσουμε τις διαστάσεις και ως εκ τούτου να διευκολυνθεί η μελέτη και ανάλυση των αποτελεσμάτων, εφαρμόστηκε ανάλυση σε κύριες συνιστώσες, κι έτσι τα 8 χαρακτηριστικά που απομένουν ανάχθηκαν σε δύο βασικές συνιστώσες (PC1 και PC2). Έτσι δίνεται η δυνατότητα να απεικονίσουμε τις χρονοσειρές στο επίπεδο, όπως φαίνεται στο ακόλουθο διάγραμμα.



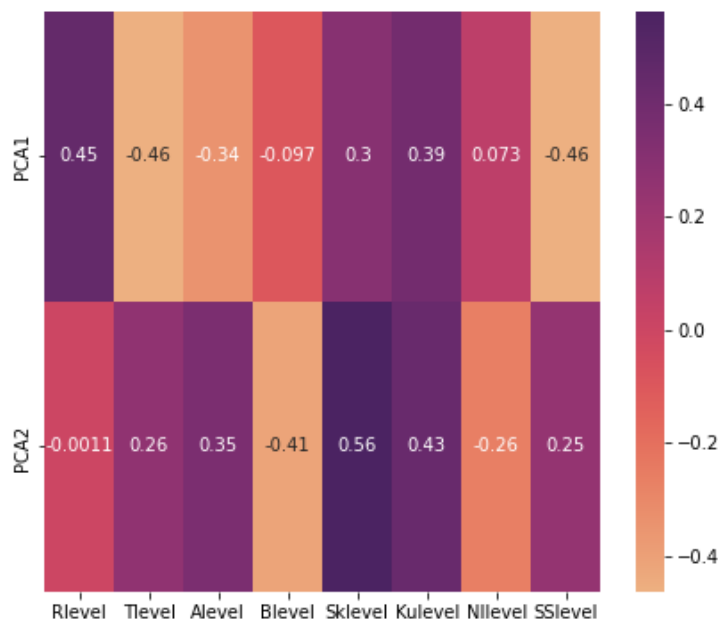
Διάγραμμα 1. Αναγωγή των 8 βασικών χαρακτηριστικών σε χώρο 2 διαστάσεων, με ανάλυση PCA

Με αυτή τη διαδικασία, ουσιαστικά ορίζονται δύο γραμμικές σχέσεις που συνδέουν τα 8 βασικά χαρακτηριστικά με τις κύριες συνιστώσες, ως ακολούθως:

$$PC1 = 0.448 \cdot Rlevel - 0.462 \cdot Tlevel - 0.340 \cdot Alevel - 0.097 \cdot Blevel + 0.301 \cdot Sklevel + 0.388 \cdot Kulevel + 0.073 \cdot Nllevel - 0.464 \cdot SSlevel \quad (1)$$

$$PC2 = -0.001 \cdot Rlevel + 0.264 \cdot Tlevel + 0.353 \cdot Alevel - 0.414 \cdot Blevel + 0.563 \cdot Sklevel + 0.433 \cdot Kulevel - 0.261 \cdot Nllevel - 0.249 \cdot SSlevel \quad (2)$$

Κατόπιν, και προκειμένου να αποκτηθεί μία οπτική αντίληψη των παραπάνω σχέσεων, δημιουργήθηκε ο χάρτης θερμότητας (heatmap), που οπτικοποιεί το βαθμό συσχέτισης του καθενός εκ των 8 χαρακτηριστικών, με τις δύο κύριες συνιστώσες.



Διάγραμμα 2. Χάρτης θερμότητας των κυρίων συνιστωσών σε σχέση με τα 8 βασικά χαρακτηριστικά.

Όπως φαίνεται στον χάρτη, η 1^η κύρια συνιστώσα εξαρτάται έντονα με αναλογική σχέση από τα χαρακτηριστικά Rlevel και Kulevel και αντιστρόφως ανάλογα από τα Tlevel, Alevel και SSlevel. Αντίστοιχα η 2^η κύρια συνιστώσα, εξαρτάται με αναλογική σχέση από τα Sklevel και Kulevel και σε μικρότερο βαθμό από τα Alevel, Tlevel και SSlevel. Αντιστρόφως ανάλογη σχέση υπάρχει με το Blevel και σε μικρότερο βαθμό με το Nllevel.

5.3 Προεπεξεργασία δεδομένων

Ως γνωστόν τα νευρωνικά δίκτυα εν αντιθέσει με τις στατιστικές μεθόδους οι οποίες μπορούν να δώσουν προβλέψεις για μεταβλητά μήκη χρονοσειρών, πρέπει να λάβουν στο επίπεδο

εισόδου τους ένα σταθερό αριθμό παρατηρήσεων. Έτσι λοιπόν, προκειμένου να δημιουργηθεί κατάλληλο σύνολο εκπαίδευσης, το αρχικό dataset χωρίστηκε σε «παράθυρα» παρατηρήσεων τα οποία είχαν ένα σταθερό αριθμό παρατηρήσεων εισόδου και 6 παρατηρήσεις εξόδου.

Για παράδειγμα, σε περίπτωση που η αρχιτεκτονική ενός δικτύου είναι τέτοια, ώστε να λαμβάνει 18 τιμές έως είσοδο και με τη χρήση αυτών να προβλέπεται ο ζητούμενος ορίζοντας των 6 παρατηρήσεων, είναι αναγκαία η ύπαρξη ενός παραθύρου $18 + 6 = 24$ παρατηρήσεων. Εάν στο αρχικό σύνολο δεδομένων υπάρχει μία χρονοσειρά με 200 παρατηρήσεις, μπορούμε απ' αυτήν να κατασκευάσουμε συνολικά $200 - 24 + 1 = 177$ παράθυρα, μήκους 24 παρατηρήσεων, λαμβάνοντας διαδοχικά ως παράθυρα τις παρατηρήσεις 1 - 24, 2 - 25 κ.ο.κ. Με τον τρόπο αυτό επιτυγχάνεται η σταθερότητα των παρατηρήσεων εισόδου, αλλά ταυτόχρονα έχουμε και μία αξιοσημείωτη αύξηση του συνόλου δεδομένων εκπαίδευσης, το οποίο στην περίπτωση των 18 παρατηρήσεων εισόδου, ανέρχεται σε 235.000 δείγματα (samples).

Επόμενο βήμα της προεπεξεργασίας ήταν η κανονικοποίηση των δεδομένων, δηλαδή η αναγωγή όλων των τιμών στην κλίμακα 0 - 1. Για την κανονικοποίηση μίας οποιασδήποτε τιμής μιας χρονοσειράς χρησιμοποιήθηκε ο τύπος:

$$z_i = \frac{x_i - \min(x_{in})}{\max(x_{in}) - \min(x_{in})}$$

Όπου z_i η κανονικοποιημένη τιμή, x_i η αρχική τιμή και τα $\max(x_{in})$ και $\min(x_{in})$ αναφέρονται στη μέγιστη και ελάχιστη τιμή της εισόδου. Ο λόγος που χρησιμοποιούνται αυτές και όχι το σύνολο των τιμών του παραθύρου, είναι ο εξής: Με τον τρόπο αυτό, στην περίπτωση που κάποια από τις 6 τιμές ελέγχου, υπερβαίνει την μέγιστη ή υπολείπεται της ελάχιστης τιμής της εισόδου, τότε η κανονικοποίηση της τιμής αυτής θα δώσει μία τιμή μεγαλύτερη του 1 ή μικρότερη του 0 αντίστοιχα. Σε περίπτωση που η κανονικοποίηση γινόταν λαμβάνοντας ως μέγιστη και ελάχιστη τιμή την μέγιστη τιμή του παραθύρου, τότε όλες οι κανονικοποιημένες τιμές θα βρίσκονταν ανάμεσα στο 0 και στο 1. Σε αυτή την περίπτωση όμως, το εκπαιδευμένο πλέον δίκτυο, δεν θα είχε τη δυνατότητα να προβλέψει τιμές εκτός του εύρους των δύο ακραίων τιμών του παραθύρου, πράγμα που επιτυγχάνεται με τον τρόπο ο οποίος περιγράφηκε.

Σαν ένα τελευταίο βήμα προεπεξεργασίας, αφαιρέθηκαν όλες οι χρονοσειρές που είχαν τιμές μικρότερες του 10, καθώς και όλες οι στάσιμες χρονοσειρές οι οποίες δεν είναι δυνατόν να κανονικοποιηθούν, αφού η μέγιστη και η ελάχιστη τιμή τους είναι ίσες και θα προέκυπτε διαίρεση με το μηδέν.

Τέλος αναφέρουμε, πως οι τιμές εξόδου που προέκυψαν από τα δίκτυα που εκπαιδεύτηκαν, ακολουθούσαν την συγκεκριμένη κανονικοποίηση και για να βρεθεί η τελική προβλεπόμενη τιμή, εφαρμόστηκε η αντίστροφη διαδικασία (δηλαδή ο υπολογισμός του x_i έχοντας ως δεδομένο το z_i).

5.4 Δημιουργία συνελκτικών δικτύων (CNN)

5.4.1 Τεχνική τερματισμού εκπαίδευσης και βελτιστοποιητής.

Προκειμένου να αποφύγουμε το φαινόμενο του overfitting, και να εξοικονομηθεί χρόνος και υπολογιστικοί πόροι, επιλέχθηκε για τις ανάγκες της εκπαίδευσης η μεθοδολογία early stopping, όπως αυτή περιγράφηκε σε προηγούμενη ενότητα, ενώ για την βελτιστοποίηση του αλγορίθμου κατάβασης δυναμικού, χρησιμοποιήθηκε ο βελτιστοποιητής adam, που όπως είδαμε στο αντίστοιχο κεφάλαιο χρησιμοποιεί τους εκθετικά σταθμισμένους μέσους όρους για να επιτύχει ταχύτερη σύγκλιση του αλγορίθμου κατάβασης δυναμικού και είναι πολύ χρήσιμος σε προβλήματα με μεγάλο αριθμό παραμέτρων.

5.4.2 Εκπαίδευση πολλαπλών μοντέλων (ensembling) και διάμεσες (median) προβλέψεις.

Είναι γνωστό ότι η εκπαίδευση των νευρωνικών δικτύων καθώς και οι επιδόσεις τους διέπονται από μεγάλη στοχαστικότητα, και σε κάθε διαδικασία εκπαίδευσης, τα τελικά βάρη που θα προκύψουν μπορεί να είναι διαφορετικά από μια άλλη με ίδιες τιμές υπερπαραμέτρων. Προκειμένου να ελαχιστοποιηθεί αυτή η τυχαιότητα και να ενισχυθεί η ακρίβεια της πρόβλεψης, κατά την διεξαγωγή των πειραμάτων, οι προβλέψεις δεν προέκυψαν από ένα μοναδικό μοντέλο. Αντιθέτως, ένα σταθερό πλήθος (10) μοντέλων, εκπαιδεύτηκαν με τις ίδιες παραμέτρους. Σαν τελική πρόβλεψη, χρησιμοποιήθηκαν οι διάμεσες (median) προβλέψεις των μοντέλων αυτών. Όπως παρατηρήθηκε, η μέθοδος αυτή επιφέρει μεγαλύτερη ακρίβεια σε σχέση με αυτή του βέλτιστου των μεμονωμένων μοντέλων.

5.4.3 Βελτιστοποίηση υπερπαραμέτρων

Για την εκτέλεση της διαδικασίας βελτιστοποίησης, χρησιμοποιήθηκε η `rython` βιβλιοθήκη `hyperopt` η οποία προσφέρει τη δυνατότητα για τυχαία αναζήτηση, εκτιμητές `Tree Parzen (TPE)` καθώς και προσαρμοζόμενοι (`Adaptive`) εκτιμητές `Tree Parzen`. Η βιβλιοθήκη παρέχει μία συνάρτηση, η οποία δέχεται ως ορίσματα την συνάρτηση, της οποίας την τιμή επιδιώκουμε να ελαχιστοποιήσουμε, ένα λεξικό (`dictionary`) της `rython`, το οποίο περιέχει τον χώρο των υπερπαραμέτρων, τον αλγόριθμο που επιλέγεται καθώς και τον αριθμό των επαναλήψεων του πειράματος.

Επιλέχθηκε ο χώρος παραμέτρων που απεικονίζεται στους πίνακες 1, 2 και διενεργήθηκαν 100 διαδικασίες εκπαίδευσης, κάθε μία από τις οποίες περιλάμβανε 10 μοντέλα με τις ίδιες παραμέτρους και υπολογισμό του `sMAPE` των διάμεσων τιμών των προβλέψεων.

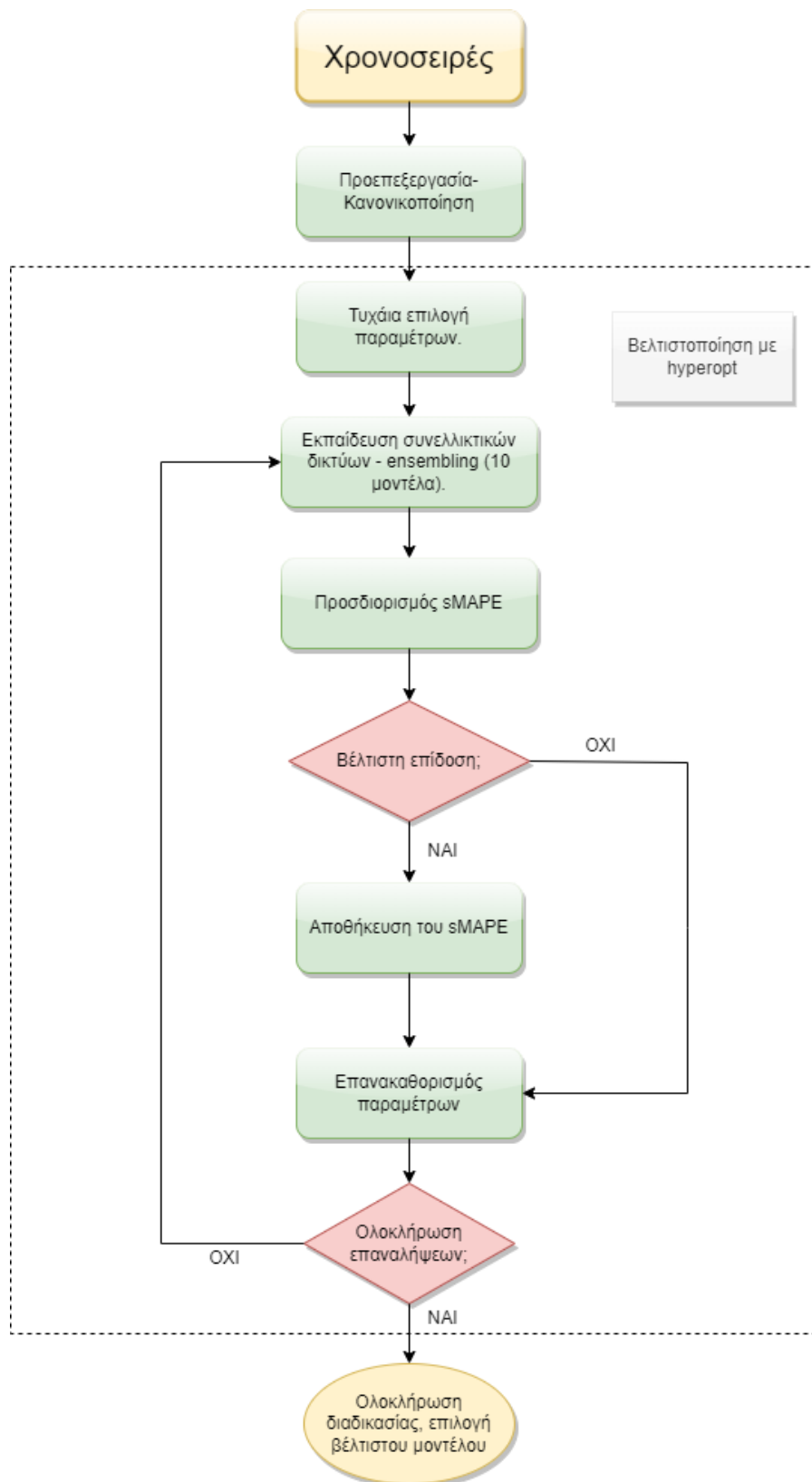
Όνομασία παραμέτρου	Περιγραφή	Χώρος τιμών	Βέλτιστη
Filters	Αριθμός φίλτρων	[16, 32, 64, 128]	16
Kernel_size	Μέγεθος πυρήνα	[3, 5]	3
Dense_neurons	Αριθμός νευρώνων στα «πυκνά επίπεδα»	[30, 40, 50, 60]	50
Activation	Συνάρτηση ενεργοποίησης	ReLU	ReLU
Strides	Βήμα συνέλιξης επιπέδου	1	1
Dilation_rate	Βαθμός διαστολής επιπέδου	[1,2]	2
padding	Στρατηγική padding	[same, valid, causal]	valid
Conv_layers_num	Πλήθος συνελκτικών επιπέδων	[1, 2, 3]	3
Dense_layers_num	Πλήθος «πυκνών» επιπέδων	[1, 2]	1

Πίνακας 1. Χώρος παραμέτρων αρχιτεκτονικής, πρώτης σειράς πειραμάτων.

Όνομασία παραμέτρου	Περιγραφή	Χώρος τιμών	Βέλτιστη
Bs	Μέγεθος δεσμίδας (batch)	[32, 64, 128, 256, 512]	512
Learning_rate	Ρυθμός μεταβολής βαρών	[0.0005, 0.001, 0.005]	0.005
Loss_f	Συνάρτηση απώλειας	mae	mae
Test_size	Ποσοστό συνόλου ελέγχου στο σύνολο των δεδομένων.	0.2	0.2
n_inputs	Πλήθος παρατηρήσεων εισόδου	[12, 15, 18, 21, 24]	18

Πίνακας 2. Χώρος παραμέτρων εκπαίδευσης, πρώτης σειράς πειραμάτων.

Η διαδικασία εκπαίδευσης και βελτιστοποίησης των υπερπαραμέτρων, παρουσιάζεται σχηματικά στο ακόλουθο BPMN διάγραμμα.



Διάγραμμα 3. Σχηματική αναπαράσταση της πειραματικής διαδικασίας.

5.4.4 Τεχνολογίες υλοποίησης.

Όπως αναφέραμε και προηγουμένως, προκειμένου να δομήσουμε τα μοντέλα και να διεξάγουμε τα πειράματα, χρησιμοποιήθηκε η γλώσσα Python και το οικοσύστημα τεχνολογιών της.

Η Python είναι γλώσσα προγραμματισμού γενικού σκοπού, η οποία χρησιμοποιείται ευρέως και σε άλλους τομείς, όπως ο γενικός προγραμματισμός, η δημιουργία desktop εφαρμογών, τα γεωγραφικά πληροφοριακά συστήματα (GIS), οι εφαρμογές διαδικτύου (με κυρίαρχα frameworks τα Django και flask), η επεξεργασία φυσικής γλώσσας, η μηχανική μάθηση κ.α.

Έχει σχεδιαστεί προκειμένου να είναι εύκολα επεκτάσιμη (με modules) και επιδιώκει μία απλούστερη σύνταξη και γραμματική, ενώ ταυτόχρονα προσφέρει στον προγραμματιστή την επιλογή της μεθοδολογίας συγγραφής κώδικα. Μεγάλο πλεονέκτημα της είναι ότι υποστηρίζει πλήρως το παράδειγμα του αντικειμενοστραφούς καθώς και του δομημένου προγραμματισμού, ενώ έχει υποστήριξη και για συναρτησιακό προγραμματισμό παρόμοιας μορφής με τη Lisp.

Η χρήση της στον τομέα της βαθιάς μηχανικής μάθησης έχει εκτοξευθεί, λόγω της ύπαρξης βιβλιοθηκών όπως είναι οι TensorFlow, Keras, Pytorch και Scikit-learn. Ως εκ τούτου, επιλέχθηκε μαζί με το οικοσύστημά τεχνολογιών της, για την εκπόνηση της παρούσας διπλωματικής.

Η TensorFlow είναι δωρεάν και ανοιχτού κώδικα βιβλιοθήκη για χρήση στη μηχανική μάθηση και την τεχνητή νοημοσύνη. Μπορεί να χρησιμοποιηθεί για μία πλειάδα εργασιών, όμως εστιάζει ειδικότερα στην εκπαίδευση και εξαγωγή συμπερασμάτων από βαθιά νευρωνικά δίκτυα (deep neural networks). Αναπτύχθηκε από την ομάδα Google Brain για εσωτερική χρήση στα πλαίσια της Google στην έρευνα και την παραγωγή και μπορεί να χρησιμοποιηθεί από γλώσσες προγραμματισμού, όπως είναι οι Python, Javascript, C++ και Java.

Το Keras αντίστοιχα, είναι μία βιβλιοθήκη ανοιχτού κώδικα, η οποία παρέχει μία Python διεπαφή για τεχνητά νευρωνικά δίκτυα (ANN). Λειτουργεί ως διεπαφή για την βιβλιοθήκη Tensorflow. Έχει σχεδιαστεί για να παρέχει τη δυνατότητα γρήγορου πειραματισμού με βαθιές αρχιτεκτονικές νευρωνικών δικτύων (deep neural networks) και εστιάζει στην

φιλικότητα προς τον χρήστη, την modular αρχιτεκτονική και την επεκτασιμότητα. Ο βασικός της δημιουργός ονομάζεται Francois Chollet ο οποίος είναι μηχανικός στην google.

Το Keras περιλαμβάνει πολλές υλοποιήσεις των δομικών συστατικών των συνηθέστερων νευρωνικών δικτύων, όπως είναι τα επίπεδα, οι συναρτήσεις ενεργοποίησης και οι βελτιστοποιητές (optimizers). Εκτός των συνηθισμένων νευρωνικών δικτύων, προσφέρει υποστήριξη για συνελκτικά δίκτυα (CNN) και ανατροφοδοτούμενα (RNN).

5.5 Δείκτες αξιολόγησης.

Για την σύγκριση των σημειακών προβλέψεων, αξιοποιήθηκε ο δείκτης σφάλματος sMAPE. Η επιλογή του εν λόγω δείκτη έγινε για τους εξής λόγους:

- Είναι ανεξάρτητος της κλίμακας της εκάστοτε χρονοσειράς και επιτρέπει τη σύγκριση μεταξύ χρονοσειρών με διαφορετικό μέσο επίπεδο
- Δεν λαμβάνει απροσδιόριστες τιμές όπως το MAPE
- Είναι φραγμένος μεταξύ των ορίων [0%, 200%]

Ταυτόχρονα, επειδή ζητούμενο εκτός από τη μελέτη της ποιότητας των προβλέψεων ήταν και το να αξιολογηθούν τα μοντέλα ως προς το ζήτημα του υπολογιστικού χρόνου και του ενεργειακού αποτυπώματός τους, μετρήθηκε ο χρόνος εκπαίδευσης των μοντέλων σε λεπτά (min).

5.6 Επιλογή μεθόδων αναφοράς (benchmarks)

Επιλέχθηκαν τρεις μέθοδοι που χρησιμοποιήθηκαν ως μέθοδοι αναφοράς, για να εξακριβωθεί αν εν τέλει τα συνελκτικά δίκτυα είναι σε θέση να παράγουν καλύτερες προβλέψεις σε σχέση με αναγνωρισμένες και παγιωμένες μεθόδους.

- **Multi-Layer Perceptron.** Την πιο αναγνωρισμένη αρχιτεκτονική νευρωνικού δικτύου αποτελούν τα γνωστά σε όλους έχουν μία επαφή με τον τομέα, Multi-Layer Perceptrons (MLP). Εκπαιδεύτηκε λοιπόν ένα τέτοιο δίκτυο, δεχόμενο ως σύνολο εκπαίδευσης το σύνολο που δημιουργήθηκε μέσω του σταδίου της προεπεξεργασίας. Το δίκτυο περιλάμβανε ένα επίπεδο με 18 νευρώνες εισόδου, τρία κρυφά επίπεδα με 27 νευρώνες το καθένα, καθώς και ένα επίπεδο εξόδου με 6 νευρώνες που αντιστοιχούν στον ζητούμενο ορίζοντα πρόβλεψης.

- **Theta.** Η μέθοδος Theta είναι μία αναγνωρισμένη και ώριμη στατιστική μέθοδος πρόβλεψης που βασίζεται στην αποσύνθεση της αρχικής χρονοσειράς. Καθιερώθηκε και έγινε γνωστή μετά από τις εξαιρετικές επιδόσεις που πέτυχε στα πλαίσια του διαγωνισμού M3.
- **ES-RNN.** Στον διαγωνισμό M4 διαγωνίστηκαν πολλές ομάδες με ποικίλλες μεθοδολογίες (στατιστικές και μηχανικής μάθησης). Ωστόσο την έκπληξη του διαγωνισμού αποτέλεσε η μέθοδος που υπέβαλλε ο Slawlek Smyl, ο οποίος δημιούργησε μία πραγματικά υβριδική μέθοδο στην οποία με τη χρήση ενός αναδρομικού νευρωνικού δικτύου (RNN) βελτιστοποιούνται οι παράμετροι εκθετικής εξομάλυνσης για την πρόβλεψη της χρονοσειράς. Ως εκ τούτου, δεν ήταν δυνατόν να μη συμπεριληφθεί στις μεθόδους αξιολόγησης της διπλωματικής, αφού το ζητούμενο ήταν οι μέθοδοι που δοκιμάζουμε να «αναμετρηθούν» με τους πιο σκληρούς όρους ανταγωνισμού.

6 Μελέτη και ανάλυση των Αποτελεσμάτων.

Στο κεφάλαιο αυτό, μελετήθηκαν τα αποτελέσματα της πειραματικής διαδικασίας από πολλές διαφορετικές γωνίες.

Αρχικά συγκρίνεται το sMAPE του βέλτιστου μοντέλου με αυτά των μεθόδων αναφοράς και διαπιστώνεται ότι υπερτερεί έναντι όλων. Κατόπιν, και δεδομένης της τυχαιότητας που είναι εγγενής στα νευρωνικά δίκτυα, λόγω της τυχαίας αρχικοποίησης των βαρών τους, επιχειρούμε να διαπιστώσουμε, αν η βελτίωση που επέφερε το μοντέλο μας σε σχέση με τις μεθόδους αναφοράς, βρίσκεται ή όχι στα όρια του στατιστικού λάθους. Επίσης διερευνήθηκε το ποιο είναι το βέλτιστο πλήθος μοντέλων που συμμετέχουν στο ensembling όσον αφορά την ακρίβεια.

Σε επόμενο στάδιο διερευνάται το κατά πόσον οι βασικές υπερπαραμέτροι αρχιτεκτονικής και εκπαίδευσης επιδρούν στην ακρίβεια του μοντέλου. Αυτό γίνεται μεταβάλλοντας ανά μία τις παραμέτρους και εξετάζοντας την επίδραση που έχει αυτή η μεταβολή στην ακρίβεια. Διερευνάται και ο χρόνος εκπαίδευσης των μοντέλων προκειμένου να εντοπιστούν τα πιο ακριβή, αλλά και αυτά που παρουσιάζουν την βέλτιστη συμπεριφορά από πλευράς ενεργειακού αποτυπώματος.

Κατόπιν, και προκειμένου να αποκτηθεί μια διαίσθηση του τρόπου με τον οποίο τα συνελκτικά δίκτυα μαθαίνουν και αποτυπώνουν στην εσωτερική τους δομή τα χαρακτηριστικά των χρονοσειρών, γίνεται διαγραμματική απεικόνιση των ενδιάμεσων εξόδων τους που αφορούν το πρώτο συνελκτικό επίπεδο και σχολιάζονται τα ευρήματα.

Ακολούθως, κάνοντας χρήση της ανάλυσης κύριων συνιστωσών που έγινε για το σύνολο δεδομένων στο προηγούμενο κεφάλαιο, χωρίστηκαν οι χρονοσειρές σε συστάδες (clusters) στον νέο χώρο δύο διαστάσεων και αποτιμώνται συγκριτικά οι επιδόσεις του CNN και των άλλων μεθόδων αναφοράς για κάθε cluster.

Τέλος με χρήση απλής γραμμικής παλινδρόμησης, επιχειρήθηκε να συσχετιστεί το σφάλμα (sMAPE) με τα 8 βασικά χαρακτηριστικά των χρονοσειρών και δημιουργήθηκε μία γραμμική σχέση που τα συνδέει, ενώ εξετάστηκε με διάφορα στατιστικά μέτρα κατά πόσον τελικά υπάρχει σημαντική συσχέτιση μεταξύ της εξαρτημένης και των ανεξάρτητων μεταβλητών.

6.1 Επιλογή βέλτιστου μοντέλου και σύγκριση επιδόσεων με μεθόδους αναφοράς.

Στην ενότητα αυτή, παρουσιάζουμε το μοντέλο που επιλέχθηκε από την διαδικασία βελτιστοποίησης υπερπαραμέτρων, συγκριτικά με τις επιλεγμένες μεθόδους benchmarking.

Όπως αναφέρεται και στον πίνακα με τις βέλτιστες υπερπαραμέτρους, το τελικό μοντέλο (στο οποίο εφεξής θα γίνεται αναφορά ως CNN), περιλαμβάνει 3 συνελκτικά επίπεδα με 16 φίλτρα και μέγεθος πυρήνα 3, καθώς και 1 πυκνό επίπεδο με 50 νευρώνες. Δεν εφαρμόστηκε η τεχνική του padding (επιλογή valid) ενώ χρησιμοποιήθηκε βαθμός διαστολής 2. Το πλήθος παρατηρήσεων εισόδου που δέχεται είναι 18 και έχει επίπεδο εξόδου με 6 νευρώνες που αντιστοιχούν στον ορίζοντα 6 ετών πρόβλεψης.

Όσον αφορά τις παραμέτρους εκπαίδευσής του, χρησιμοποιήθηκε μέγεθος δεσμίδας 512, ρυθμός μάθησης 0.005 και ποσοστό του συνόλου ελέγχου επί του συνολικού ίσο με 20%.

Μέθοδος	sMAPE	Βελτίωση %
CNN	13.055	-
MLP	13.087	0.245
THETA	14.593	10.539
ES-RNN	13.175	0.911

Πίνακας 3. Σύγκριση του CNN με τις μεθόδους αναφοράς

Στον πίνακα 3 παρουσιάζονται οι επιδόσεις του μοντέλου σε σχέση με τις μεθόδους αναφοράς που επιλέχθηκαν. Παρατηρούμε ότι το CNN εμφάνισε μία πολύ μεγάλη βελτίωση σε σχέση με τη μέθοδο Theta, της τάξεως του 10.5%, και 0.9% σε σχέση με τη νικήτρια μέθοδο του διαγωνισμού M4 (ESRNN). Τέλος μικρότερη αλλά υπαρκτή βελτίωση (0.245%) εμφανίζεται ως προς την μέθοδο MLP.

6.2 Στατιστική ανάλυση και διαστήματα εμπιστοσύνης

Επειδή όπως προαναφέρθηκε τα νευρωνικά δίκτυα, ακόμα και με τη χρήση ensembling που έγινε, διέπονται από μεγάλη τυχαιότητα, γίνεται εμφανές ότι αν επανεκπαιδευτεί η ίδια διάταξη που παρουσιάσαμε στην προηγούμενη ενότητα, δεν θα λάβουμε το ίδιο αποτέλεσμα. Αυτό που ωστόσο μας ενδιαφέρει να διερευνηθεί, είναι κατά πόσον η βελτίωση που επιφέρει σε σχέση με τις μεθόδους αναφοράς είναι στατιστικά σημαντική. Προκειμένου να διερευνηθεί αυτό, δημιουργήθηκε ένα δείγμα 50 επαναλήψεων της εκπαίδευσης του

μοντέλου, δηλαδή κάνοντας χρήση των βέλτιστων παραμέτρων που προσδιορίστηκαν, εκπαιδεύτηκαν 50 φορές από 10 μοντέλα και υπολογίστηκε το sMAPE των διάμεσων προβλέψεών τους. Κατόπιν προσδιορίστηκαν τα στατιστικά στοιχεία του δείγματος. Η μέση τιμή ήταν $\bar{x} = 13.068$ ενώ οι μέγιστες και ελάχιστες τιμές ήταν αντίστοιχα: $Min = 13.0283$ και $Max = 13.1177$.

Η τυπική απόκλιση ήταν $s = 0.0156^5$ και το τυπικό σφάλμα $s_{\bar{x}} = 0.00221^6$.

Όπως προκύπτει από την παραπάνω ανάλυση, το 95% διάστημα εμπιστοσύνης για τον μέσο όρο της κατανομής, βρίσκεται μεταξύ των ορίων 13.064 και 13.072 όπως φαίνεται στην ακόλουθη σχέση.

$$95\%, 1.960 \cdot s_{\bar{x}} = 13.068 \pm 0.004 (\pm 0.03\%)$$

Έτσι λοιπόν το άνω όριο του διαστήματος προκύπτει 13.072 και παρατηρούμε, ότι η μέθοδος, για διάστημα εμπιστοσύνης 95% εμφανίζει βελτίωση και ως προς τις 3 μεθόδους αναφοράς⁷.

6.3 Βέλτιστο πλήθος εκπαιδευόμενων μοντέλων (ensembling)

Κατόπιν, επιχειρήθηκε να εξεταστεί, κατά πόσον το πλήθος των μοντέλων που εκπαιδεύονται και των οποίων λαμβάνεται η διάμεση πρόβλεψη (ensembling), επηρεάζει την ακρίβεια των προβλέψεων. Για το σκοπό αυτό, διεξήχθησαν επιπλέον εκπαιδεύσεις με τις ίδιες παραμέτρους, στις οποίες εκπαιδεύτηκαν 20, 30, 40 και 50 μοντέλα αντίστοιχα, και υπολογίστηκε η ακρίβεια των διάμεσων προβλέψεών τους όπως φαίνεται στον ακόλουθο πίνακα.

⁵Η τυπική απόκλιση του δείγματος που δίνεται από τη σχέση:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

⁶Το τυπικό σφάλμα που δίνεται από τη σχέση:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

⁷ Όσον αφορά την μέθοδο Theta επειδή τα αποτελέσματά της είναι ντετερμινιστικά, είμαστε σίγουροι ότι ο μέσος όρος της κατανομής είναι αυτός που προσδιορίζεται από το SMAPE. Αντιθέτως για τις άλλες δύο μεθόδους το δεχόμαστε σαν υπόθεση ότι γνωρίζουμε τον μέσο όρο του πληθυσμού, καθώς δεν ήταν εφικτό να επαναληφθεί το πείραμα και να δημιουργηθούν δείγματα για τις δύο αυτές μεθόδους.

Πλήθος μοντέλων	Σφάλμα (sMAPE)	Χρόνος (min)
10 (μέση τιμή δείγματος)	13.068	12.5
20	13.060	25.8
30	13.044	37.5
40	13.051	48.8
50	13.056	58.5

Πίνακας 4. sMAPE με βάση τον αριθμό των εκπαιδευμένων μοντέλων (χρήση διάμεσων προβλέψεων)

Όπως προκύπτει από τον πίνακα, η ακρίβεια βελτιώνεται (ελάττωση σφάλματος) μέχρι και την τιμή των 30 μοντέλων, και κατόπιν επιδεινώνεται για τις τιμές των 40 και 50 μοντέλων. Ως εκ τούτου, για τα επόμενα θεωρήθηκε ότι ο αριθμός 30 είναι ο βέλτιστος και οι όποιες επιπλέον εκπαιδεύσεις έγιναν με τον αριθμό αυτό. Η ακρίβεια 13.044 που επιτεύχθηκε, θεωρείται πλέον αυτή ως ακρίβεια του CNN.

Για λόγους υπολογιστικού χρόνου, δεν επαναλήφθηκε η διαδικασία προσδιορισμού στατιστικής σημαντικότητας του προηγούμενου κεφαλαίου, ωστόσο είναι μάλλον εμφανές και αποτελεί αίσθηση του μελετητή, ότι η απόσταση ασφαλείας από τις υπόλοιπες μεθόδους αυξάνει σε σχέση με το ensembling των 10 μοντέλων.

6.4 Επίδραση των βασικών παραμέτρων στην ακρίβεια του μοντέλου.

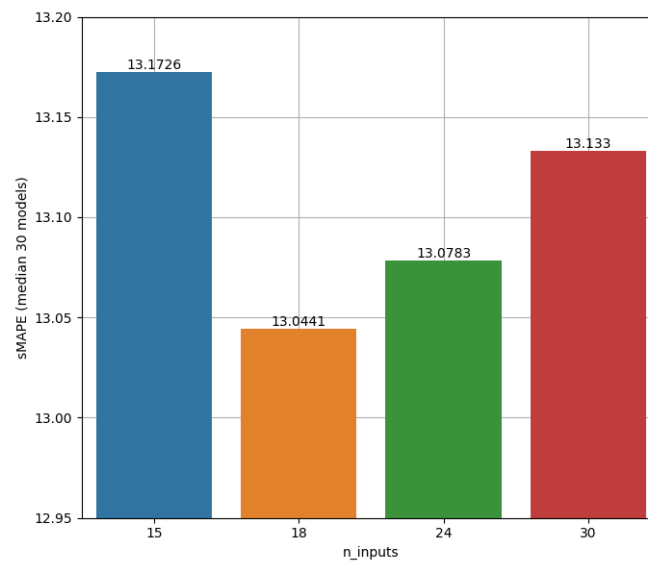
Στην ενότητα αυτή, εν αντιθέσει με αυτήν της πειραματικής διαδικασίας όπου έγινε αναζήτηση στο χώρο των παραμέτρων με τη βοήθεια Bayesian optimization, πραγματοποιήθηκε αναζήτηση, μεταβάλλοντας κάθε φορά μία, από τις παραμέτρους αρχιτεκτονικής ή εκπαίδευσης, και εξετάζοντας την επίδραση που είχε η αλλαγή αυτή στην ακρίβεια. Τα αποτελέσματα παρουσιάζονται αναλυτικά στους ακόλουθους πίνακες και διαγράμματα.

6.4.1 Παρατηρήσεις εισόδου

Παρατηρήσεις εισόδου (n_inputs)	Διάρκεια (min)	Σφάλμα (sMAPE)
15	26.12	13.173
18 ⁸	39.92	13.044
24	20.37	13.078
30	18.17	13.133

Πίνακας 5. Αποτελέσματα πειραμάτων με μεταβολή των παρατηρήσεων εισόδου

Τα αποτελέσματα από τη μεταβολή των παρατηρήσεων εισόδου δείχνουν ότι η ακρίβεια πρόβλεψης βελτιώνεται όσο κινούμαστε προς τις 18 παρατηρήσεις (αρχική επιλογή) όπου και έχουμε την καλύτερη επίδοση και επιδεινώνεται όσο αυξάνουμε τον αριθμό της εισόδου πέραν των 18 παρατηρήσεων. Πρέπει να επισημανθεί εδώ, ότι η αύξηση ή ελάττωση των παρατηρήσεων εισόδου επιδρά και στο μέγεθος του συνόλου εκπαίδευσης, αφού όσο περισσότερες παρατηρήσεις χρειάζεται το παράθυρο εισόδου, τόσο λιγότερα είναι τα παράθυρα που μπορούν να δημιουργηθούν από τα αρχικά δεδομένα.



Διάγραμμα 4. Ραβδόγραμμα του sMAPE ως προς τις παρατηρήσεις εισόδου (διάμεσος 30 μοντέλων)

Αξιοσημείωτη είναι εδώ η παρατήρηση, ότι με τη χρήση 24 παρατηρήσεων εισόδου αντί για 18, έχουμε ελάττωση της ακρίβειας της τάξεως του 0.26% (από 13.044 σε 13.078), με ελάττωση του χρόνου εκπαίδευσης στο 50% περίπου.

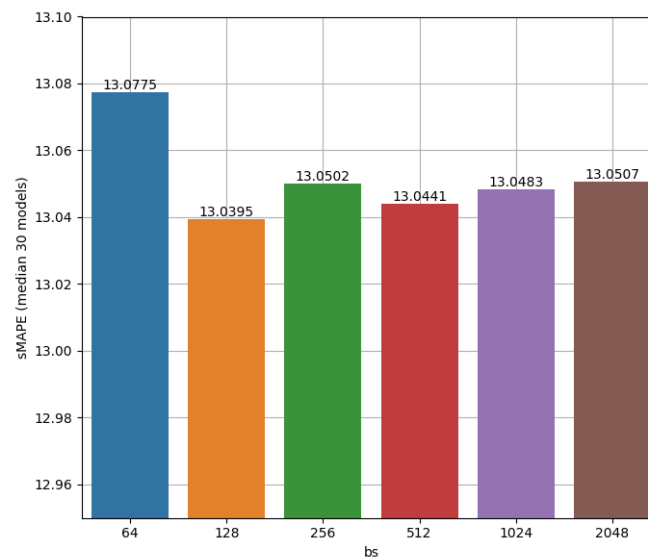
⁸Η τιμή της υπερπαραμέτρου που προσδιορίστηκε στο στάδιο του Bayesian optimization.

6.4.2 μέγεθος δεσμίδας (batch size)

Μέγεθος δεσμίδας (batch size)	Διάρκεια (min)	Ακρίβεια (sMAPE)
64	67.03	13.078
128	72.7	13.039
256	55.92	13.05
512 ⁹	39.92	13.044
1024	41.8	13.048
2048	58.3	13.050

Πίνακας 6. Αποτελέσματα πειραμάτων με μεταβολή του μεγέθους δεσμίδας

Παρατηρούμε πως η βέλτιστη επίδοση επιτεύχθηκε για μέγεθος δεσμίδας ίσο με 128 (βελτίωση από 13.044 σε 13.039) ωστόσο αυτή η επιλογή ήταν και αρκετά πιο χρονοβόρα όσον αφορά την εκπαίδευση με αύξηση κατά 28% του χρόνου της.



Διάγραμμα 5. Ραβδόγραμμα του sMAPE ως προς το μέγεθος δεσμίδας (διάμεσος 30 μοντέλων)

Αντίθετα, το μέγεθος δεσμίδας 64 επέφερε χειρότερο αποτέλεσμα (0.26% από 13.044 σε 13.078) με ταυτόχρονη αύξηση του χρόνου εκπαίδευσης κατά 20% περίπου. Από το ραβδόγραμμα παρατηρούμε επίσης πως δεν υπάρχει κάποια εμφανής τάση της ακρίβειας σε σχέση με την αύξηση ή μείωση του μεγέθους δεσμίδας, με εξαίρεση την τιμή 64 η οποία είχε εμφανώς χειρότερη επίδοση από τις υπόλοιπες, παρατηρούμε μικρές αυξομειώσεις.

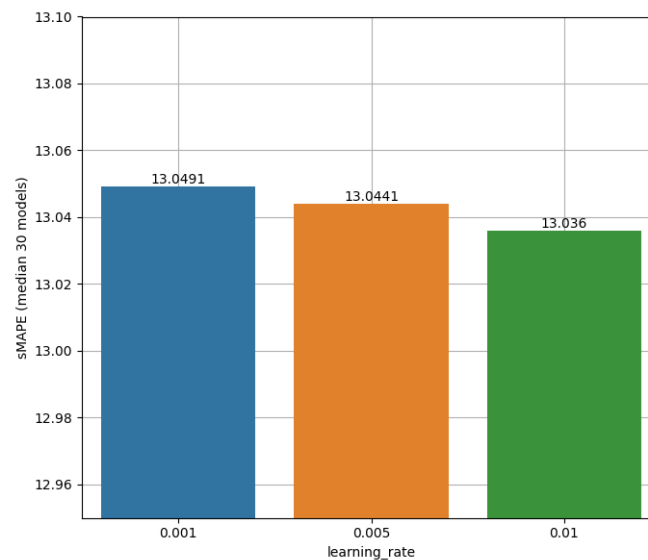
⁹Η τιμή της υπερπαραμέτρου που προσδιορίστηκε στο στάδιο του Bayesian optimization.

6.4.3 Ρυθμός μάθησης (learning rate)

Ρυθμός μάθησης (learning rate)	Διάρκεια (min)	Σφάλμα (sMAPE)
0.001	57.55	13.049
0.005 ¹⁰	39.92	13.044
0.01	31.48	13.036

Πίνακας 7. Αποτελέσματα πειραμάτων με μεταβολή του ρυθμού μάθησης

Η αύξηση του ρυθμού μάθησης στο 0.01 οδήγησε σε βελτίωση της ακρίβειας πρόβλεψης (0.06% από 13.044 σε 13.036), και είναι και πολύ σημαντικό το γεγονός ότι η επιλογή αυτή απαίτησε και κατά 22% λιγότερο χρόνο εκπαίδευσης.



Διάγραμμα 6. Ραβδόγραμμα του sMAPE ως προς τον ρυθμό μάθησης (διάμεσος 30 μοντέλων)

Στο διάγραμμα παρατηρούμε ότι με την αύξηση του learning rate βελτιώνεται η ακρίβεια πρόβλεψης και ταυτόχρονα ελαττώνεται ο χρόνος εκπαίδευσης. Η παρατήρηση αυτή σηματοδοτεί κατά τη γνώμη του μελετητή, την ανάγκη να διερευνηθούν ακόμα μεγαλύτερες τιμές του learning rate (π.χ. 0.02 - 0.05) οι οποίες θα μπορούσαν να εμφανίσουν περαιτέρω βελτίωση της ακρίβειας ή/και ελάττωση του χρόνου.

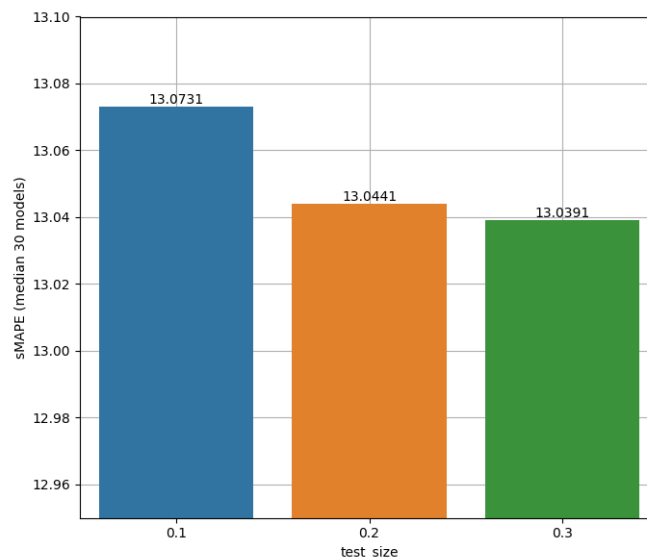
¹⁰Η τιμή της υπερπαραμέτρου που προσδιορίστηκε στο στάδιο του Bayesian optimization.

6.4.4 Ποσοστό συνόλου ελέγχου (test size)

Σύνολο ελέγχου (test size)	Διάρκεια (min)	Σφάλμα (sMAPE)
0.1	51.53	13.073
0.2 ¹¹	39.92	13.044
0.3	42.15	13.039

Πίνακας 8. Αποτελέσματα πειραμάτων με μεταβολή του ποσοστού του συνόλου ελέγχου

Η αύξηση του ποσοστού που καταλαμβάνει το σύνολο ελέγχου σε 0.3 οδήγησε σε βελτίωση της ακρίβειας πρόβλεψης (0.038% από 13.044 σε 13.039) με μία μικρή αύξηση του χρόνου εκπαίδευσης της τάξεως του 5%. Αντιθέτως η μείωσή του σε 10% επί του συνόλου των δεδομένων, οδήγησε σε σημαντική πτώση της ακρίβειας πρόβλεψης με ταυτόχρονη αύξηση του χρόνου εκπαίδευσης κατά 29%.



Διάγραμμα 7. Ραβδόγραμμα του sMAPE ως προς το ποσοστό του συνόλου ελέγχου (διάμεσος 30 μοντέλων)

Κι εδώ όπως και με την περίπτωση του learning rate, το διάγραμμα δίνει την αίσθηση αύξησης της ακρίβειας με την ταυτόχρονη αύξηση του ποσοστού. Συνεπώς θα είχε ενδιαφέρον ο πειραματισμός με ακόμα μεγαλύτερες τιμές (π.χ. 0.35, 0.4).

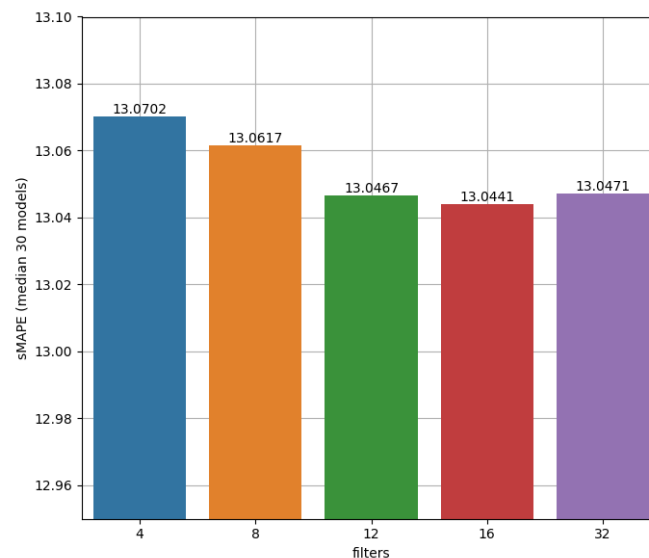
¹¹Η τιμή της υπερπαραμέτρου που προσδιορίστηκε στο στάδιο του Bayesian optimization.

6.4.5 Αριθμός φίλτρων

Αρ. φίλτρων (filters)	Διάρκεια (min)	Σφάλμα (sMAPE)
4	22.03	13.07
8	25.07	13.061
12	32.35	13.046
16 ¹²	39.92	13.044
32	61.35	13.047

Πίνακας 9. Αποτελέσματα πειραμάτων με μεταβολή του αριθμού φίλτρων

Καμία από τις μεταβολές στον αριθμό φίλτρων δεν επέφερε βελτίωση στην ακρίβεια της πρόβλεψης, ωστόσο είναι αξιοσημείωτο το γεγονός ότι η ελάττωση των φίλτρων σε 12, οδήγησε και σε ελάττωση του χρόνου εκπαίδευσης κατά 20% έχοντας ως τίμημα μία πολύ μικρή ελάττωση της ακρίβειας κατά 0.015% (από 13.0441 σε 13.0466).



Διάγραμμα 8. Ραβδόγραμμα του sMAPE ως προς τον αριθμό φίλτρων (διάμεσος 30 μοντέλων)

Στο διάγραμμα παρατηρούμε πως η ακρίβεια παρουσιάζει βελτίωση όσο κινούμαστε από αριστερά και δεξιά, προς την τιμή 16. Ενδιαφέρον παρουσιάζει η μελέτη της περιοχής των 10 – 20 φίλτρων (δηλαδή μια περιοχή εκατέρωθεν των 16, όπου ενδεχομένως να εμφανιστεί βελτιωμένη ακρίβεια με παρόμοιους ή και καλύτερους χρόνους).

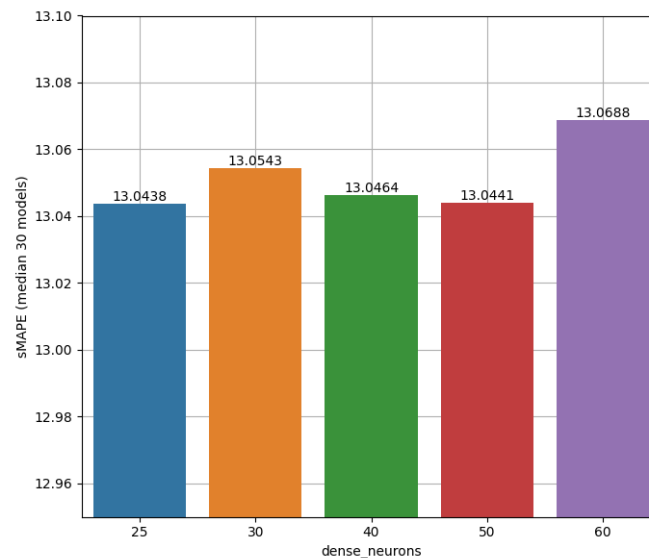
¹²Η τιμή της υπερπαραμέτρου που προσδιορίστηκε στο στάδιο του Bayesian optimization.

6.4.6 Αριθμός νευρώνων «πυκνού» επιπέδου (dense neurons)

Νευρώνες πυκνού επιπέδου	Διάρκεια (min)	Σφάλμα (sMAPE)
25	43.13	13.043
30	38.57	13.054
40	41.07	13.046
50 ¹³	39.92	13.044
60	43.2	13.068

Πίνακας 10. Αποτελέσματα πειραμάτων με μεταβολή των νευρώνων πυκνού επιπέδου

Η ελάττωση των νευρώνων πυκνού επιπέδου κατά το ήμισυ (25 από 50) οδήγησε σε μια ανεπαίσθητη βελτίωση της ακρίβειας κατά 0.0003 με μία μικρή αύξηση του χρόνου εκπαίδευσης.



Διάγραμμα 9. Ραβδόγραμμα του sMAPE ως προς τον αριθμό νευρώνων πυκνού επιπέδου (διάμεσος 30 μοντέλων)

Από το διάγραμμα συμπεραίνουμε πως η μεταβολή της ακρίβειας σε συνάρτηση με τον αριθμό των νευρώνων δεν εμφανίζει κάποια συγκεκριμένη τάση, αλλά αντίθετα εμφανίζονται αυξομειώσεις με βέλτιστη ωστόσο περιοχή τους 40-50 νευρώνες.

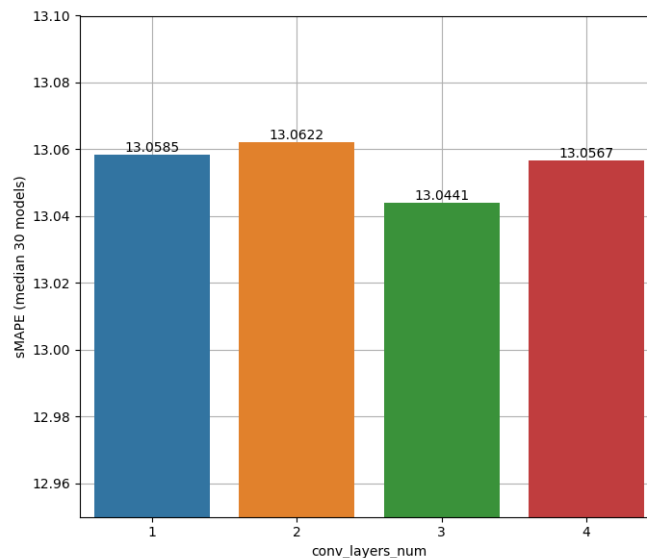
¹³Η τιμή της υπερπαραμέτρου που προσδιορίστηκε στο στάδιο του Bayesian optimization.

6.4.7 Αριθμός συνελκτικών επιπέδων (convolutional layers).

Αρ. συνελκτικών επιπέδων	Διάρκεια (min)	Σφάλμα (sMAPE)
1	15.37	13.059
2	27.67	13.062
3 ¹⁴	39.92	13.044
4	47.02	13.057

Πίνακας 11. Αποτελέσματα πειραμάτων με μεταβολή του αριθμού συνελκτικών επιπέδων.

Η μεταβολή του αριθμού των συνελκτικών επιπέδων (convolutional layers) δεν επέφερε κάποια βελτίωση στην ακρίβεια, ωστόσο, η χρήση μόνο ενός επιπέδου απαίτησε σχεδόν 50% λιγότερο χρόνο εκπαίδευσης, ενώ η ζημιά στην ακρίβεια ήταν 0.1% (από 13.044 σε 13.059).



Διάγραμμα 10. Ραβδόγραμμα του sMAPE ως προς τον αριθμό συνελκτικών επιπέδων (διάμεσος 30 μοντέλων)

Στο διάγραμμα παρατηρούμε ότι δεν υπάρχει κάποια γνησίως μονότονη σχέση του αριθμού της ακρίβειας με τον αριθμό συνελκτικών επιπέδων, ενώ εμφανίζονται κι εδώ αυξομειώσεις, με βέλτιστη τιμή αυτή των τριών επιπέδων, όμως η παράμετρος αυτή θα έπρεπε να εξεταστεί και σε σχέση με τον αριθμό των φίλτρων σε κάθε επίπεδο, τα οποία σε ενδεχόμενη μελλοντική πειραματική διαδικασία θα μπορούσαν να διαφοροποιούνται σε κάθε επίπεδο και να μελετηθεί κατ' αυτό τον τρόπο η επίδρασή τους.

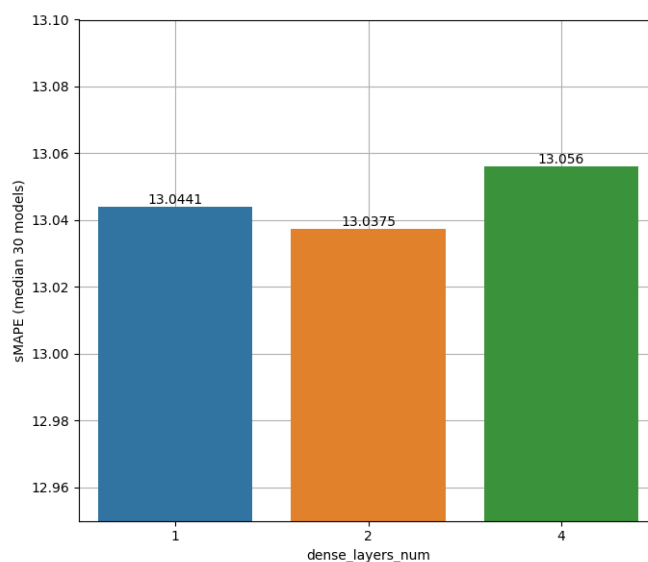
¹⁴Η τιμή της υπερπαραμέτρου που προσδιορίστηκε στο στάδιο του Bayesian optimization.

6.4.8 Αριθμός πυκνών επιπέδων (dense layers)

Αριθμός πυκνών επιπέδων (dense layers)	Διάρκεια (min)	Σφάλμα (sMAPE)
1 ¹⁵	39.92	13.044
2	39.37	13.038
4	46.88	13.056

Πίνακας 12. Αποτελέσματα πειραμάτων με μεταβολή του αριθμού πυκνών επιπέδων.

Τα δύο πυκνά επίπεδα, πέτυχαν καλύτερη επίδοση κατά 0.046% (από 13.044 σε 13.038) έχοντας παρόμοιο χρόνο εκπαίδευσης, ενώ η χρήση 4 επιπέδων ελάττωσε την ακρίβεια αυξάνοντας τον χρόνο εκπαίδευσης κατά 17% περίπου.



Διάγραμμα 11. Ραβδόγραμμα του sMAPE ως προς τον αριθμό πυκνών επιπέδων (διάμεσος 30 μοντέλων)

Από το γεγονός ότι η αύξηση των πυκνών επιπέδων από 1 σε 2 δεν επέφερε αύξηση του χρόνου εκπαίδευσης, μπορούμε να αντιληφθούμε την πολυπλοκότητα και την συνεπαγόμενη δυσκολία στην κατανόηση της εσωτερικής δομής των νευρωνικών δικτύων. Θα περίμενε κανείς, δεδομένης της ύπαρξης αρκετά περισσότερων παραμέτρων εκπαίδευσης που επέφερε το ένα επιπλέον επίπεδο, να απαιτείται μεγαλύτερος χρόνος εκπαίδευσης, ωστόσο απαιτήθηκε οριακά λιγότερος χρόνος και μάλιστα με βελτίωση της ακρίβειας, κάτι που κατά πάσα πιθανότητα οφείλεται στο γεγονός ότι τα δύο πυκνά επίπεδα μπόρεσαν να αποτυπώσουν με καλύτερο τρόπο κάποια χαρακτηριστικά (features) της χρονοσειράς, με

¹⁵Η τιμή της υπερπαραμέτρου που προσδιορίστηκε στο στάδιο του Bayesian optimization.

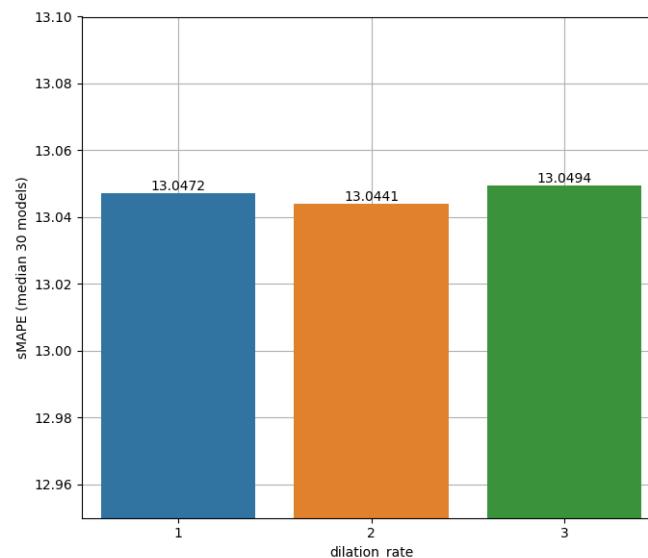
αποτέλεσμα να είναι σε θέση να ολοκληρώσουν την εκπαίδευση σε λιγότερες εποχές (epochs) σε σχέση με την αρχιτεκτονική του ενός επιπέδου και με αυτό τον τρόπο να αντισταθμίζεται από χρονικής πλευράς ο μεγαλύτερος αριθμός παραμέτρων.

6.4.9 Βαθμός διαστολής (dilation rate)

Βαθμός διαστολής (dilation rate)	Διάρκεια (min)	Σφάλμα (sMAPE)
1	38.42	13.047
2 ¹⁶	39.92	13.044
3	39.65	13.049

Πίνακας 13. Αποτελέσματα πειραμάτων με μεταβολή του βαθμού διαστολής.

Η μεταβολή του βαθμού διαστολής επέφερε παρόμοια αποτελέσματα όσον αφορά την ακρίβεια αλλά και όσον αφορά το χρόνο εκπαίδευσης.



Διάγραμμα 12. Ραβδόγραμμα του sMAPE ως προς τον βαθμό διαστολής (διάμεσος 30 μοντέλων)

Στο διάγραμμα 11 παρατηρούμε πως η συνάρτηση της ακρίβειας σε σχέση με το χρόνο εμφανίζει τη βέλτιστη τιμή (ελάχιστο) για την τιμή 2, ενώ εκατέρωθεν της εμφανίζεται μια μικρή επιδείνωση. Η χρήση του dilation rate παίζει έναν ιδιαίτερο ρόλο στην αρχιτεκτονική των temporal convolutional networks, στα οποία η τιμή του αυξάνεται εκθετικά ανάμεσα στα επίπεδα, προκειμένου τελικά η έξοδος να εξαρτάται από το σύνολο των παρατηρήσεων εισόδου, όπως αναφέρθηκε σε προηγούμενο κεφάλαιο.

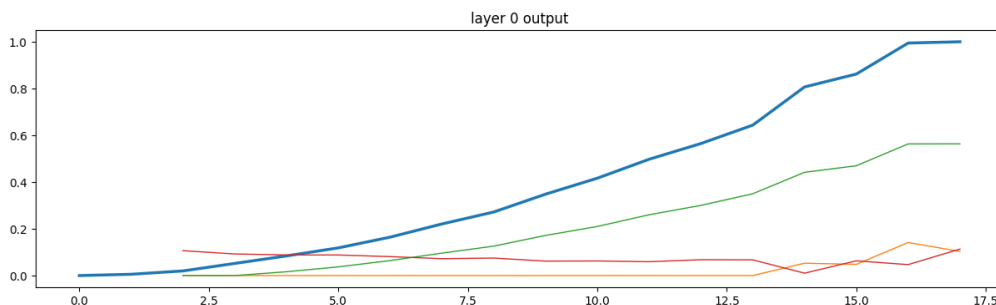
¹⁶Η τιμή της υπερπαραμέτρου που προσδιορίστηκε στο στάδιο του Bayesian optimization.

Μέσω της παραπάνω διαδικασίας, προέκυψαν αρχιτεκτονικές που επέφεραν μικρές βελτιώσεις σε σχέση με το βέλτιστο μοντέλο που προσδιορίστηκε στο πειραματικό στάδιο. Για τις ανάγκες των επόμενων ενοτήτων έγινε χρήση της βέλτιστης από αυτές τις αρχιτεκτονικές, η οποία επιτυγχάνεται με την αλλαγή του learning rate από 0.005 σε 0.01, και η οποία, είχε βελτίωση του sMAPE από **13.044** σε **13.036**, ενώ ταυτόχρονα ο χρόνος εκπαίδευσης ελαττώθηκε από 40 λεπτά περίπου σε 31.5 λεπτά, συνδυάζοντας έτσι το τερπνόν μετά του ωφελίμου.

6.4.10 Απεικόνιση ενδιάμεσων εξόδων των συνελικτικών επιπέδων

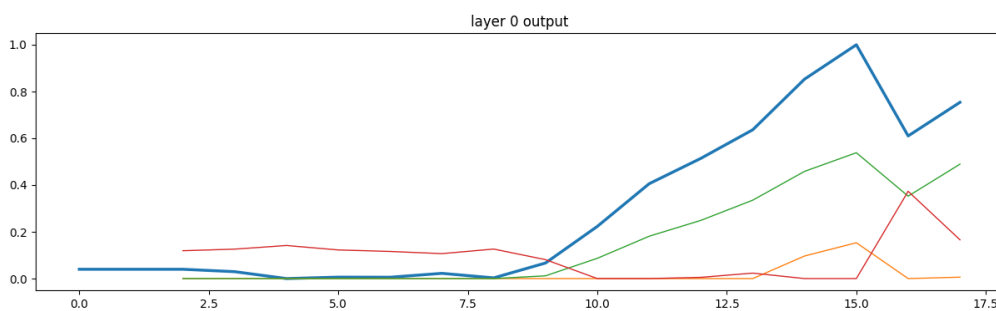
Σε μία προσπάθεια να κατανοηθεί καλύτερα ο τρόπος μάθησης των συνελικτικών επιπέδων, και ο τρόπος με τον οποίο αποτυπώνουν βασικές έννοιες των χρονοσειρών, όπως είναι η τάση, η εποχικότητα, η κυκλικότητα και η τυχαιότητα, επιλέχθηκαν συγκεκριμένες χρονοσειρές που εμφανίζουν έντονα τα παραπάνω χαρακτηριστικά, και παραστάθηκαν γραφικά οι εξοδοί των φίλτρων τους, όσον αφορά το πρώτο συνελικτικό επίπεδο. Το μοντέλο που χρησιμοποιήθηκε ήταν αυτό που προέκυψε από το στάδιο της πειραματικής διαδικασίας, κατόπιν όλων των σταδίων βελτιστοποίησης. Τα αποτελέσματα φαίνονται στα παρακάτω διαγράμματα.

Όπως αναφέρθηκε στο κεφάλαιο της πειραματικής διαδικασίας, η βέλτιστη αρχιτεκτονική περιείχε 16 φίλτρα. Ωστόσο από την μελέτη των εξόδων του πρώτου επιπέδου, ένας σημαντικός αριθμός φίλτρων είχε μηδενική έξοδο, ενώ κάποια άλλα δεν εμφάνιζαν κάποιο μοτίβο τέτοιο ώστε να θεωρηθεί σκόπιμη η αποτύπωσή τους. Έτσι ο μελετητής κατέληξε σε 3 φίλτρα, την έξοδο των οποίων αποτύπωσε γραφικά, προκειμένου ο αναγνώστης να αποκτήσει μία αίσθηση του τι συμβαίνει στο εσωτερικό ενός συνελικτικού δικτύου.



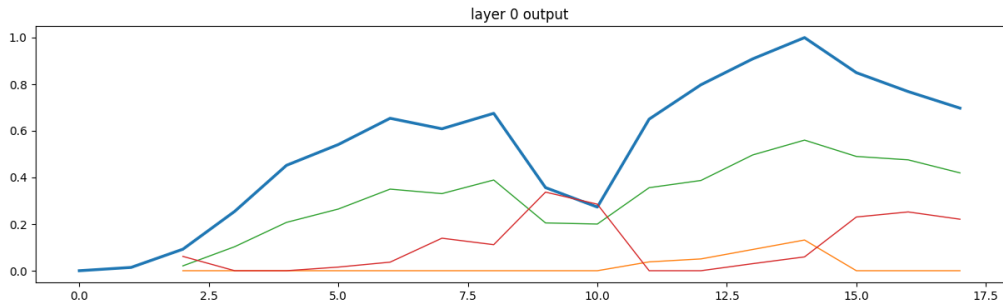
Διάγραμμα 13. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά με τάση, χωρίς εποχικότητα ή τυχαιότητα.

Από τη μελέτη του πρώτου διαγράμματος όπου υπάρχει έντονο το φαινόμενο της τάσης σε όλο το εύρος της χρονοσειράς, παρατηρούμε ότι το ένα από τα φίλτρα (πράσινο χρώμα) διατηρεί τη μορφή της αρχικής χρονοσειράς (έντονη γραμμή), ωστόσο φαίνεται πιο εξομαλυμένη και σε χαμηλότερα επίπεδα τιμών σε σχέση με αυτήν. Ένα δεύτερο φίλτρο (πορτοκαλί γραμμή) αποτυπώνει τη μορφή της σε μία μικρή περιοχή προς το τέλος, ενώ στα αρχικά στάδια η τιμή του έχει μηδενιστεί. Την πιο αξιοσημείωτη συμπεριφορά εμφανίζει το φίλτρο που απεικονίζεται με κόκκινη γραμμή, το οποίο φαίνεται ότι στα σημεία στα οποία αλλάζει η κλίση της αρχικής χρονοσειράς, αλλάζει και αυτό κλίση, με αντίθετη όμως κατεύθυνση από αυτή της αρχικής.



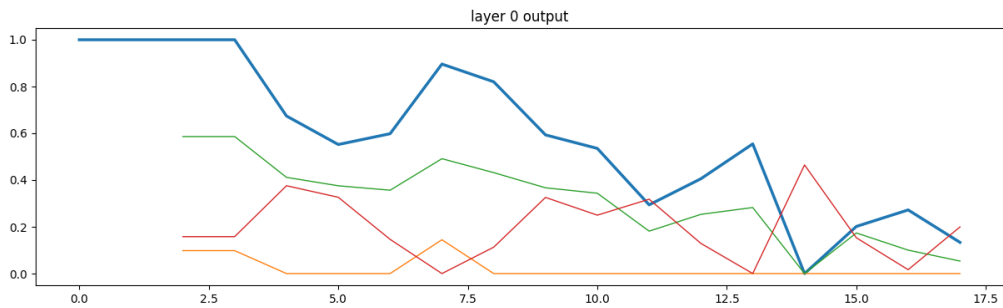
Διάγραμμα 14. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά , με αλλαγή μοτίβου (από στασιμότητα αποκτάει τάση)

Στο δεύτερο διάγραμμα, απεικονίζεται χρονοσειρά (μπλε έντονο χρώμα) η οποία στη μέση της εμφανίζει αλλαγή μοτίβου δηλαδή από στάσιμη αποκτά μια έντονη ανοδική τάση. Και στην περίπτωση αυτή, παρατηρούμε ένα φίλτρο που διατηρεί τη μορφή της, πιο εξομαλυμένο και σε πιο χαμηλά επίπεδα τιμών. Το δεύτερο φίλτρο επίσης όπως και στο προηγούμενο διάγραμμα, αποτυπώνει τη μορφή της χρονοσειράς σε ακόμα πιο εξομαλυμένη μορφή και χαμηλά επίπεδα, σε βαθμό στον οποίο σε ένα μεγάλο τμήμα του να έχει μηδενιστεί εντελώς η επίδρασή του. Τέλος το φίλτρο που απεικονίζεται με την κόκκινη γραμμή, διατηρεί την ιδιότητά να εμφανίζει κλίσεις αντίθετες από αυτές της χρονοσειράς, στα σημεία όπου αυτή μεταβάλλεται.



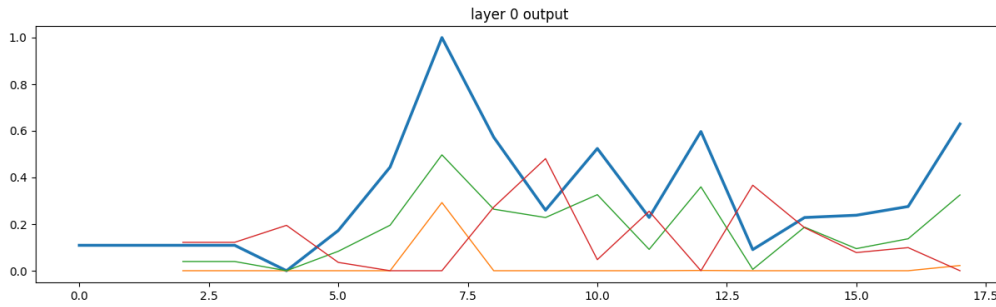
Διάγραμμα 15. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά με κυκλικότητα.

Στο τρίτο διάγραμμα, η χρονοσειρά εμφανίζει δύο κύκλους μακράς διάρκειας ο καθένας εκ των οποίων χαρακτηρίζεται από μία περίοδο ανοδικής τάσης ακολουθούμενης από μία περίοδο καθοδικής. Τα χαρακτηριστικά των τριών φίλτρων που απεικονίζονται παραμένουν ίδια. Το φίλτρο με την πορτοκαλί γραμμή έχει μη μηδενικές τιμές μόνο στο τμήμα όπου η αρχική χρονοσειρά εμφανίζει τον δεύτερο κύκλο της, ενώ στις υπόλοιπες περιοχές έχει μηδενική επίδραση.



Διάγραμμα 16. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά με πτωτική τάση και κυκλικότητα

Η χρονοσειρά του διαγράμματος 14 παρουσιάζει μία πτωτική τάση και κάποια κυκλικότητα (διαστήματα με πτώση ακολουθούμενα από ανοδικά διαστήματα). Το πράσινο φίλτρο και σε αυτήν την περίπτωση αποδίδει τη μορφή της χρονοσειράς, σε εξομαλυμένη μορφή και σε πιο χαμηλά επίπεδα τιμών, ενώ το πορτοκαλί είναι στα περισσότερα τμήματά του μηδενικό, ακολουθώντας μόνο σε δύο σημεία την μορφή της χρονοσειράς.



Διάγραμμα 17. Απεικόνιση εξόδων πρώτου επιπέδου σε χρονοσειρά με τυχαιότητα χωρίς τάση.

Η χρονοσειρά του τελευταίου διαγράμματος χαρακτηρίζεται από μεγάλες διακυμάνσεις και τυχαιότητα, κάτι που παρατηρείται και στα διαγράμματα των φίλτρων. Τα φίλτρα του διαγράμματος εμφανίζουν παρόμοια συμπεριφορά με τις προηγούμενες περιπτώσεις και αξιοσημείωτη είναι η μορφή του φίλτρου που απεικονίζεται με κόκκινη γραμμή, το οποίο θα μπορούσε να πει κανείς πως αποτελεί «είδωλο» της αρχικής χρονοσειράς στα περισσότερα σημεία.

Μία γενική παρατήρηση από τη μελέτη των παραπάνω περιπτώσεων, είναι ότι ένας σημαντικός αριθμός φίλτρων δεν αξιοποιήθηκε από το μοντέλο, είχαν μηδενική επίδραση και ως εκ τούτου δεν απεικονίστηκαν και στα διαγράμματα. Η παρατήρηση αυτή είχε ως αποτέλεσμα και την δοκιμή των αρχιτεκτονικών με μικρότερο αριθμό φίλτρων που παρουσιάστηκε στην προηγούμενη ενότητα. Η αρχιτεκτονική με 12 φίλτρα όπως είδαμε, είχε σχεδόν εφάμιλλα αποτελέσματα με αυτήν των 16 φίλτρων, ενώ είχε σημαντική ελάττωση στον χρόνο εκπαίδευσης.

Ωστόσο, τα παραπάνω διαγράμματα αναφέρονται στις εξόδους του πρώτου από τα τρία συνελκτικά επίπεδα που αποτέλεσαν την βέλτιστη αρχιτεκτονική. Στην αρχιτεκτονική αυτή, όλα τα επίπεδα αποτελούνταν από τον ίδιο αριθμό φίλτρων. Τα ευρήματα της ενότητας αυτής θα μπορούσαν να σημαίνουν ότι ενδεχομένως μία αρχιτεκτονική με λιγότερα φίλτρα μόνο στο πρώτο επίπεδο αλλά διατήρησή τους στα υπόλοιπα, θα μπορούσε να επιφέρει βελτιωμένα ή παρόμοια αποτελέσματα με μικρότερο χρόνο εκπαίδευσης.

6.5 Συσχέτιση του σφάλματος των προβλέψεων με τα χαρακτηριστικά της χρονοσειράς

Στην ενότητα αυτή, έγινε ανάλυση γραμμικής παλινδρόμησης, προκειμένου να διαπιστωθεί κατά πόσον υπάρχει συσχέτιση των επιδόσεων του CNN με τα 8 βασικά χαρακτηριστικά των χρονοσειρών. Με τη διαδικασία αυτή, προσδιορίζονται και οι συντελεστές που συσχετίζουν την εξαρτημένη μεταβλητή (sMAPE) με τις ανεξάρτητες (τιμές των βασικών χαρακτηριστικών).

Το συμπέρασμα που προκύπτει είναι ότι υπάρχει μία ασθενής συσχέτιση μεταξύ των τιμών των 8 χαρακτηριστικών και της επίδοσης του CNN δικτύου με τιμή R^2 ίση με 0.138. Η τιμή του F-STATISTIC για την παλινδρόμηση είναι 461.1, και η αντίστοιχη τιμή του p-value που προσδιορίζεται είναι <0.001 . Λόγω του ότι η τελευταία είναι μικρότερη του 0.05 συμπεραίνουμε, ότι τουλάχιστον μία από τις 8 εξαρτημένες μεταβλητές είναι σημαντικά συσχετισμένη με το sMAPE.

Χαρακτηριστικό	Συντελεστής (coefficient)	$P > t $
Rlevel	-10.015	< 0.001
Tlevel	-37.334	< 0.001
Alevel	-13.974	< 0.001
Blevel	-3.936	< 0.001
Sklevel	7.116	< 0.001
Kulevel	0.198	0.331
Nllevel	0.445	0.217
SSlevel	32.050	< 0.001
Const	29.936	< 0.001

Πίνακας 14. Αποτελέσματα γραμμικής παλινδρόμησης για τη συσχέτιση της επίδοσης των συνελκτικών δικτύων με τα 8 βασικά χαρακτηριστικά.

Στον πίνακα 14, φαίνονται οι συντελεστές β της γραμμικής παλινδρόμησης για το καθένα από τα χαρακτηριστικά, καθώς και η τιμή $P > |t|$ η οποία για κάθε μεμονωμένη μεταβλητή, μας υποδηλώνει αν υπάρχει σημαντική συσχέτιση με την εξαρτημένη. Στην περίπτωσή μας, όλες οι μεταβλητές εκτός των Nllevel και Kulevel έχουν στο πεδίο αυτό τιμή < 0.05 που σημαίνει ότι υπάρχει η ζητούμενη σημαντικότητα στη συσχέτιση. Όσον αφορά τον τρόπο με τον οποίο επηρεάζουν οι μεταβλητές την ακρίβεια, είναι προφανές ότι επιδιώκουμε την ελάττωση της εξαρτημένης μεταβλητής. Συνεπώς η αύξηση της τιμής των χαρακτηριστικών που έχουν

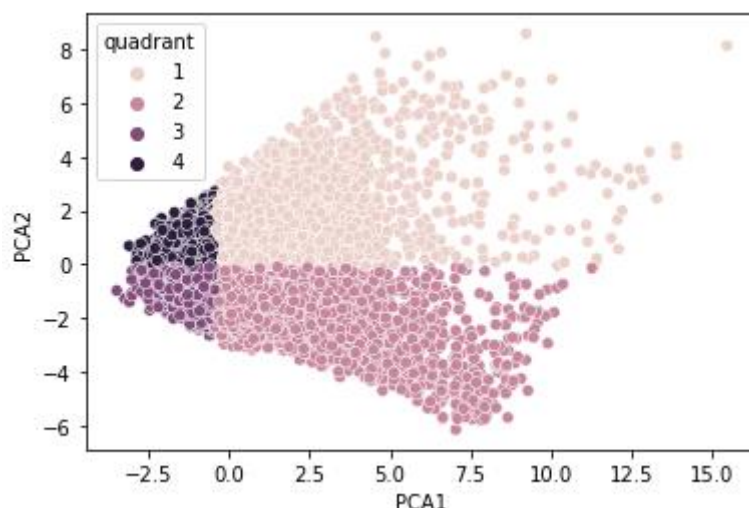
αρνητικό συντελεστή, οδηγεί σε βελτίωση της ακρίβειας, ενώ αντίστοιχα αυτών που έχουν θετικό συντελεστή σε ελάττωσή της.

Παρατηρούμε πως τους μεγαλύτερους αρνητικούς συντελεστές έχουν τα χαρακτηριστικά Tlevel και Alevel με -37.334 και -13.974 αντίστοιχα, ενώ τους μεγαλύτερους θετικούς συντελεστές έχουν τα Sslevel και Sklevel με 32.050 και 7.116 αντίστοιχα. Πρακτικά αυτό σημαίνει πως το CNN δίκτυο προβλέπει καλύτερα τις χρονοσειρές που χαρακτηρίζονται από έντονη τάση και ισχυρή αυτοσυσχέτιση πρώτης τάξης, ενώ η απόδοσή του ελαττώνεται με την ύπαρξη ασυμμετρίας (skewness) καθώς και για μεγάλες τιμές του χαρακτηριστικού SSlevel (Fractional differencing parameter).

6.6 Συσταδοποίηση των χρονοσειρών (clustering) και σύγκριση επιδόσεων των μεθόδων αναφοράς ανά συστάδα.

Σε προηγούμενη ενότητα παρουσιάστηκε η ανάλυση του συνόλου των δεδομένων σε κύριες συνιστώσες. Στην ενότητα αυτή οι χρονοσειρές διαχωρίζονται σε συστάδες (clusters) στον χώρο των δύο διαστάσεων που ορίζουν οι δύο αυτές συνιστώσες. Κατόπιν γίνεται συγκριτική ανάλυση των επιδόσεων του CNN καθώς και των μεθόδων αναφοράς για καθεμία από αυτές. Με τον τρόπο αυτό, μπορούμε να αποκτήσουμε πληροφορία σχετικά με το ποια ή ποιες από τις μεθόδους εμφανίζουν καλύτερη ακρίβεια και σε ποιο είδος χρονοσειρών.

Προκειμένου, να καταλήξουμε σε όσο το δυνατόν ισοπληθείς συστάδες, το επίπεδο των κυρίων συνιστωσών χωρίστηκε σε τεταρτημόρια, που είχαν ως αρχή των αξόνων τις διάμεσες τιμές των χρονοσειρών όσον αφορά τις κύριες συνιστώσες.



Διάγραμμα 18. Διαχωρισμός των χρονοσειρών σε συστάδες (clusters)

Η διάμεσος της πρώτης συνιστώσας είναι -0.430 ενώ της δεύτερης -0.074 περίπου. Στο σημείο αυτό ορίζεται ένα νοητό σύστημα κάθετων αξόνων, που χωρίζουν τις χρονοσειρές όπως φαίνεται στο διάγραμμα (scatterplot).

Επόμενο βήμα ήταν η συγκριτική ανάλυση των επιδόσεων των συνελκτικών δικτύων, με το MLP benchmark που δημιουργήθηκε για τις ανάγκες της παρούσας μελέτης, με την μέθοδο ES-RNN, η οποία κατετάγη πρώτη στον διαγωνισμό M4, και την μέθοδο Theta η οποία είναι μία καθιερωμένη στατιστική μέθοδος πρόβλεψης με ιδιαίτερα καλές επιδόσεις στον διαγωνισμό M3. Συγκεκριμένα υπολογίστηκε το sMAPE της κάθε μίας από τις 4 μεθόδους, για κάθε τεταρτημόριο. Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα:

Τεταρτημόριο	CNN	MLP	THETA	ESRNN
1	18.007	18.148	20.122	18.196
2	15.970	16.005	17.221	16.213
3	8.922	8.924	9.725	9.005
4	9.183	9.207	11.257	9.227

Πίνακας 15. sMAPE των μεθόδων ανά τεταρτημόριο

Ζητούμενο του διαχωρισμού αυτού, είναι να εντοπιστεί, αν η μέθοδος των συνελκτικών δικτύων, εμφανίζει καλύτερες ή χειρότερες επιδόσεις σε σχέση με τις υπόλοιπες μεθόδους σε συγκεκριμένες κατηγορίες χρονοσειρών.

Το πρώτο συμπέρασμα από τη μελέτη του πίνακα είναι ότι και οι 4 μέθοδοι εμφανίζουν διαφορετικές επιδόσεις ανά τεταρτημόριο. Συγκεκριμένα οι καλύτερες επιδόσεις

εμφανίζονται στα τεταρτημόρια 3 και 4, σημαντικά χειρότερες επιδόσεις συναντάμε στο τεταρτημόριο 2, και ακόμα χειρότερες στο τεταρτημόριο 1, όπου το sMAPE του CNN έχει διπλάσια τιμή σε σχέση με αυτή των τεταρτημορίων 3 και 4.

Προκύπτει επίσης, ότι τα συνελκτικά δίκτυα υπερτερούν σε σχέση με τις υπόλοιπες μεθόδους σε όλα τα τεταρτημόρια. Μάλιστα, σε όλα τα τεταρτημόρια, διατηρείται και η ίδια κατάταξη όσον αφορά τις επιδόσεις των μεθόδων. Δηλαδή πρώτο έρχεται το CNN, δεύτερο το MLP, τρίτη η μέθοδος ESRNN και τελευταία η μέθοδος Theta.

Τέλος δημιουργήσαμε τον ακόλουθο πίνακα, που δείχνει την ποσοστιαία βελτίωση που εμφανίζει το CNN σε σχέση με τις άλλες τρεις μεθόδους ανά τεταρτημόριο¹⁷.

Τεταρτημόριο	CNN-MLP	CNN-THETA	CNN-ESRNN
1	0.78%	10.51%	1.04%
2	0.22%	7.26%	1.50%
3	0.02%	8.26%	0.92%
4	0.26%	18.42%	0.48%

Πίνακας 16. Ποσοστιαία βελτίωση του sMAPE από το CNN ως προς τις μεθόδους αναφοράς.

Όπως είχε επισημανθεί και από τη μελέτη του προηγούμενου πίνακα το CNN παρουσιάζει βελτίωση σε όλα τα τεταρτημόρια και ως προς όλες τις μεθόδους. Ως προς το MLP που αποτελεί και την πιο κοντινή του μέθοδο αναφορικά με τις επιδόσεις, η μεγαλύτερη βελτίωση της τάξεως του 0.777% εμφανίζεται στο πρώτο τεταρτημόριο, ενώ σχεδόν αμελητέα βελτίωση (0.022%) στο τρίτο τεταρτημόριο. Ως προς την μέθοδο THETA οι βελτιώσεις είναι αξιόλογες σε όλα τα τεταρτημόρια, με την μεγαλύτερη να σημειώνεται στο τεταρτημόριο 4 της τάξεως 18.424% και τη μικρότερη στο τεταρτημόριο 2 (7.264%). Τέλος, ως προς τη μέθοδο ESRNN η μεγαλύτερη βελτίωση (1.499%) επιτυγχάνεται στο τεταρτημόριο 1 και η μικρότερη στο τεταρτημόριο 4 με 0.477%.

¹⁷ Η ποσοστιαία βελτίωση υπολογίζεται με βάση τον τύπο $\frac{sMAPE_{Bench} - sMAPE_{CNN}}{sMAPE_{Bench}} \cdot 100\%$. Έτσι όταν το CNN έχει μικρότερο SMAPE δηλαδή καλύτερη ακρίβεια, το ποσοστό παίρνει θετικές τιμές ενώ σε αντίθετη περίπτωση αρνητικές.

7 Συμπεράσματα – Μελλοντική έρευνα

Στην παρούσα διπλωματική εργασία, μελετήθηκαν και υλοποιήθηκαν ποικίλλες αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων, σε μια προσπάθεια να προβλεφθούν βέλτιστα τα δεδομένα του διαγωνισμού M4. Συγκεκριμένα από το σύνολο των δεδομένων του διαγωνισμού αυτού, χρησιμοποιήθηκαν οι ετήσιες χρονοσειρές και έγιναν σημειακές προβλέψεις. Τα αποτελέσματα συγκρίθηκαν με ένα MLP που υλοποιήθηκε επί τούτου, καθώς και με τη νικήτρια μέθοδο του διαγωνισμού ES-RNN (μέθοδος Smyl), και την μέθοδο Theta που αποτελεί μία καθιερωμένη στατιστική μέθοδο πρόβλεψης.

Μέσα από σειρές πειραμάτων που είχαν ως σκοπό την βελτιστοποίηση των υπερπαραμέτρων της αρχιτεκτονικής, προσδιορίστηκε μία περιοχή τους, στην οποία, το συνελκτικό δίκτυο δίνει αποτελέσματα αισθητά βελτιωμένα σε σχέση με τις υπόλοιπες μεθόδους.

Τα βασικά συμπεράσματα από την όλη διαδικασία μπορούν να συνοψιστούν στα εξής:

- Η χρήση των συνελκτικών δικτύων στον τομέα των προβλέψεων, δεν έχει αρκετό χρονικό διάστημα που έχει αρχίσει να ερευνάται, ενώ το παραδοσιακό προνομιακό πεδίο των δικτύων αυτών είναι η αναγνώριση εικόνας. Με την παρούσα έρευνα, αποδεικνύεται ότι τα δίκτυα αυτά, όχι απλώς επιβιώνουν, αλλά έχουν και καλύτερα αποτελέσματα στις προβλέψεις χρονολογικών σειρών σε σχέση με τα πιο παραδοσιακά feed forward δίκτυα όπως είναι τα MLP.
- Η οπτική αναπαράσταση που έγινε στις ενδιάμεσες εξόδους των συνελκτικών επιπέδων καταδεικνύει ένα ακόμα πλεονέκτημα από τη χρήση τους, το οποίο έγκειται στο ότι είναι πλέον δυνατό να οπτικοποιηθεί ο τρόπος με τον οποίο αντιλαμβάνονται την έννοια της χρονοσειράς και να αποκτηθεί μία αίσθηση του τρόπου μάθησής τους. Αντίθετα, στα MLP, είναι από πολύ δύσκολο έως και αδύνατο να εντοπίσει ο μελετητής μοτίβα στην εσωτερική δομή τους, γεγονός που τα καθιστά μαύρα κουτιά και δυσχεραίνει τις προσπάθειες βελτιστοποίησης των επιδόσεών τους.
- Οι αρχιτεκτονικές που επέφεραν τα καλύτερα αποτελέσματα δεν ήταν κατ' ανάγκην και αυτές που απαιτούσαν τον μεγαλύτερο χρόνο εκπαίδευσης. Αντιθέτως, σε πολλές περιπτώσεις και ιδιαίτερα στην πρώτη φάση των πειραμάτων, εντοπίστηκε μία αρνητική συσχέτιση του χρόνου εκπαίδευσης με την ακρίβεια των προβλέψεων.

- Μετά τον προσδιορισμό των βέλτιστων παραμέτρων μέσω της πειραματικής διαδικασίας, ακολούθησε ένα δεύτερο στάδιο στο οποίο οι παράμετροι μεταβάλλονταν κατά μία κάθε φορά. Στο στάδιο αυτό δεν εντοπίστηκαν μεγάλες διακυμάνσεις στην ακρίβεια των προβλέψεων, κάτι που μας οδηγεί στο συμπέρασμα ότι το εκπαιδευμένο μοντέλο είναι αρκετά ευσταθές όσον αφορά μικρές αλλαγές στις τιμές των παραμέτρων.
- Τα δεδομένα εκπαίδευσης, διαχωρίστηκαν σε 4 τεταρτημόρια με βάση τις δύο κύριες συνιστώσες που προέκυψαν από την αντίστοιχη ανάλυση. Η μέθοδος των συνελικτικών δικτύων εμφάνισε βελτίωση στην ακρίβεια πρόβλεψης και στα 4 τεταρτημόρια σε σχέση με τις υπόλοιπες 3 μεθόδους αναφοράς. Αυτό σημαίνει ότι χαρακτηρίζεται από μία εν γένει ποιοτικότερη προσέγγιση ως προς την μάθηση-πρόβλεψη και η βελτίωση δεν εξαρτάται από κάποιο συγκεκριμένο χαρακτηριστικό της χρονοσειράς.
- Το παραπάνω συμπέρασμα, σε συνδυασμό με το γεγονός ότι οι χρονοσειρές του διαγωνισμού δεν είναι εξειδικευμένες αλλά αντλούνται από πολλούς τομείς του επιχειρηματικού κόσμου καθιστά κατά τη γνώμη του μελετητή τα ήδη εκπαιδευμένα μοντέλα μία αξιόπιστη λύση πρόβλεψης και για ad hoc χρονοσειρές που μπορεί να έχει ανάγκη ο οποιοσδήποτε ενδιαφερόμενος (επιχείρηση, ακαδημαϊκό ίδρυμα, ιδιώτης).

Η εργασία αυτή θα μπορούσε να αποτελέσει το έναυσμα για μία σειρά βελτιώσεων και επεκτάσεων εκ των οποίων οι κυριότερες είναι οι εξής:

- Εκπαίδευση μοντέλων με αντίστοιχες αρχιτεκτονικές και για τα υπόλοιπα είδη χρονοσειρών του διαγωνισμού (τριμηνιαίες εβδομαδιαίες, ετήσιες κ.λπ). Λόγω του γεγονότος ότι οι σειρές αυτές εμφανίζουν και το στοιχείο της εποχικότητας που δεν εμφανίζεται στις ετήσιες, αναμένουμε ακόμα καλύτερες επιδόσεις.
- Εκπαίδευση μοντέλων για την παραγωγή πιθανοτικών προβλέψεων.
- Ενίσχυση των αρχιτεκτονικών με επιπλέον γνώση. Συγκεκριμένα προτείνεται η τροφοδότηση των πυκνών επιπέδων που έπονται των συνελικτικών, με στοιχεία όπως είναι οι τιμές των κύριων χαρακτηριστικών των χρονοσειρών που έχουν προσδιοριστεί και ενδεχομένως με τις προβλέψεις άλλων μεθόδων (π.χ Theta). Υπάρχει εκ μέρους του μελετητή η αίσθηση και η αισιοδοξία ότι αυτές οι παρεμβάσεις θα μπορούσαν να επιφέρουν σημαντική βελτίωση στην ακρίβεια των προβλέψεων.

Διαδίδεται από πολλούς ειδικούς του χώρου (π.χ. Andrew Ng) η άποψη, ότι το επόμενο στάδιο της εξέλιξης στην μηχανική μάθηση, βρίσκεται στο λεγόμενο Machine learning engineering for production (MLOps). Ο τομέας αυτός αναφέρεται στα εργαλεία, στις τεχνικές και στην πρακτική εμπειρία, που μπορούν να μεταμορφώσουν την θεωρητική γνώση της μηχανικής μάθησης σε ένα εργαλείο έτοιμο προς παραγωγική πλέον χρήση. Μία θεμελιώδης διαφορά που υπάρχει στον τομέα αυτό σε σχέση με την παραδοσιακή μηχανική μάθηση, είναι ότι τα συστήματα που βρίσκονται σε παραγωγική λειτουργία οφείλουν να διαχειρίζονται δυναμικά δεδομένα, τα οποία εξελίσσονται με φρενήρεις ρυθμούς. Ταυτόχρονα, ένα σύστημα παραγωγικής λειτουργίας οφείλει να λειτουργεί ανελλιπώς με το μικρότερο δυνατό κόστος και έχοντας την μέγιστη δυνατή απόδοση. Ο μηχανικός που θα ασχοληθεί με τον τομέα αυτό, οφείλει να διαθέτει στην φαρέτρα του όχι μόνο τη θεωρητική γνώση της μηχανικής μάθησης, αλλά και εμπειρία στις σύγχρονες μεθόδους ανάπτυξης λογισμικού προκειμένου να ανταποκριθεί στις απαιτήσεις.

Στα πλαίσια των παραπάνω, και προκειμένου τα αποτελέσματα της παρούσας εργασίας να τύχουν χρήσης από ευρύτερο κοινό, μία ιδέα του μελετητή είναι η δημιουργία διαδικτυακής εφαρμογής, δομημένης με σύγχρονες αρχιτεκτονικές (Service Oriented Architecture ή Microservices) και τεχνολογίες ανάπτυξης (π.χ React για το frontend κομμάτι και Django ή παρεμφερείς python τεχνολογίες για το backend) η οποία θα καθιστά προσβάσιμα τα απότοκα της παρούσας έρευνας στο online κοινό.

Σε μία τέτοια εφαρμογή θα μπορούσε ο ενδιαφερόμενος, μέσω γραφικού περιβάλλοντος αλλά και μέσω προγραμματιστικής διεπαφής (restful api) να εισάγει τις τιμές μίας ή περισσοτέρων χρονοσειρών, καθώς και κάποιες επιλογές (όπως το αν η χρονοσειρά είναι ετήσια κ.λπ.) και να λαμβάνει τις προβλέψεις των ήδη υπαρχόντων μοντέλων. Σε μία ακόμα πιο προχωρημένη εκδοχή της εφαρμογής, ο χρήστης θα μπορούσε να εκπαιδεύει το δικό του μοντέλο, ορίζοντας μέσω του γραφικού περιβάλλοντος ή του api, τις υπερπαραμέτρους της προτίμησής του και η εφαρμογή να διενεργεί την εκπαίδευση στο backend κομμάτι της. Το παραγόμενο μοντέλο θα είναι διαθέσιμο στον χρήστη για να το χρησιμοποιήσει κατά το δοκούν. Οι επιλογές είναι πραγματικά ατελείωτες, και θα μπορούσαν να συμπεριλάβουν ακόμα και την παροχή συνόλου δεδομένων εκπαίδευσης από τον ίδιο το χρήστη.

Ο μελετητής πιστεύει ότι μία τέτοια εφαρμογή, θα προσέλκυε το ενδιαφέρον αρκετών ερευνητών οι οποίοι θα έχουν τη δυνατότητα αξιοποίησης της γνώσης και εμπειρίας που αποκτήθηκε από την παρούσα εργασία στα πλαίσια των δικών τους ερευνών, και γιατί όχι, το ενδιαφέρον επιχειρήσεων και ιδιωτών που θα θέλουν να εκπονήσουν ή και να συγκρίνουν προβλέψεις στον τομέα ενδιαφέροντός τους.

Κλείνοντας, μιας και είναι και ιδιαίτερος επίκαιρο λόγω της ενεργειακής κρίσης που βιώνουμε, θα κάνουμε μια σύντομη αναφορά στο ενεργειακό αποτύπωμα (carbon footprint) της μηχανικής μάθησης και για το ποια είναι η θέση της παρούσας εργασίας στην προβληματική που δημιουργεί το ζήτημα αυτό. Τα τελευταία χρόνια η μηχανική μάθηση έχει αποκτήσει εξέχουσα θέση στον χώρο των ΤΠΕ και το γεγονός αυτό, οδήγησε σε ανησυχίες που σχετίζονται με τα ολοένα αυξανόμενα κόστη που αφορούν την υπολογιστική ισχύ που απαιτείται, καθώς και τις εκπομπές αερίων θερμοκηπίου που σχετίζονται με αυτήν. Υπάρχουν απαισιόδοξες απόψεις που θεωρούν ότι οι ανάγκες για κατανάλωση ενέργειας θα συνεχίσουν να αυξάνονται εκθετικά σε σημείο ώστε πλέον η μηχανική μάθηση να καταστεί μη βιώσιμη, αλλά υπάρχουν και οι αισιόδοξες που θεωρούν ότι με τις κατάλληλες πρακτικές η κατανάλωση θα σταθεροποιηθεί τελικά σε ένα σημείο (plateau). Οι πρακτικές αυτές συνοψίζονται στις εξής [67]:

- Επιλογή πιο αποδοτικών αρχιτεκτονικών, όπως τα αραιά μοντέλα (sparse models), τα οποία μπορούν να βελτιώσουν την ποιότητα των αποτελεσμάτων ενώ παράλληλα ελαττώνουν τις υπολογιστικές ανάγκες κατά 3 έως 10 φορές.
- Κατάλληλη επιλογή hardware. Χρησιμοποιώντας συστήματα και επεξεργαστές που έχουν βελτιστοποιηθεί για την εκπαίδευση μοντέλων μηχανικής μάθησης, σε αντίθεση με τους επεξεργαστές γενικού σκοπού, οι ανάγκες σε ενέργεια μπορούν να ελαττωθούν από 2 έως 5 φορές.
- Χρήση του υπολογιστικού νέφους (cloud). Οι υπολογισμοί στο cloud αντί για τοπικά data centers ελαττώνουν τις ενεργειακές απαιτήσεις κατά 1 έως 4 φορές.
- Χωρική βελτιστοποίηση. Στο υπολογιστικό νέφος επιτρέπεται συνήθως στους χρήστες να επιλέγουν την τοποθεσία που χρησιμοποιεί την «καθαρότερη» ενέργεια, ελαττώνοντας έτσι το αποτύπωμα του άνθρακα κατά 5 έως 10 φορές.

Οι τέσσερις αυτές πρακτικές, μπορούν συνδυαστικά να ελαττώσουν την απαιτούμενη ενέργεια μέχρι και 100 φορές και τις εκπομπές άνθρακα κατά 1000 φορές.

Στην παρούσα διπλωματική, η οποία θεωρούμε ότι απεικονίζει σε ένα βαθμό τη σύγχρονη «αλυσίδα παραγωγής» της μηχανικής μάθησης σε μικρογραφία, έγινε αντιληπτό και αποδείχθηκε και στην πράξη, πως ένα πιο ελαφρύ και αποδοτικό μοντέλο (CNN) μπορεί να αποφέρει καλύτερα αποτελέσματα αλλά και βελτιωμένη ενεργειακή συμπεριφορά σε σχέση με ένα βαρύ μοντέλο όπως είναι το MLP.

Ωστόσο, η βασική κατανάλωση ενέργειας στην παρούσα εργασία, δεν προήλθε από την εκπαίδευση του τελικού μοντέλου καθαυτού, αλλά από την διεξαγωγή αλληπαλλήλων πειραμάτων προκειμένου να προσδιοριστούν οι βέλτιστες υπερπαραμέτροι. Ο μελετητής θεωρεί πως το γεγονός αυτό, σε συνδυασμό με το ότι πολλοί ερευνητές στον πλανήτη αναγκάζονται να «επανεφευρίσκουν τον τροχό» (ήτοι να εκπαιδεύουν και να δοκιμάζουν μοντέλα που έχουν ήδη εκπαιδευτεί από συναδέλφους τους), αποτελεί και ένα από τα σημαντικότερα προβλήματα που πρέπει να αντιμετωπιστούν προκειμένου να επιτευχθεί η πολυπόθητη ενεργειακή βιωσιμότητα. Με την ευχή και ελπίδα πως τα πειράματα που διενεργήθηκαν στην παρούσα εργασία, θα αποτελέσουν μία μικρή έστω καθοδήγηση στους συναδέλφους που θα εκπονήσουν παρεμφερή έρευνα στο τόσο συναρπαστικό αυτό πεδίο, την εναποθέτει στην διάθεση και την κριτική τους σκέψη.

8 Βιβλιογραφικές αναφορές

- [1] S. Makridakis, R. Hogarth και A. Gaba, *Dance with Chance: Making luck work for you*, Oneworld Publications, 2010.
- [2] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed., Australia: OTexts, 2018.
- [3] H. R. Max Roser και E. Ortiz-Ospina, «World Population Growth,» *Our World in Data*, 2013.
- [4] R. A. Neher, R. Dyrdak, V. Druelle, E. B. Hodcroft και J. Albert, «Impact of seasonal forcing on a potential SARS-CoV-2 pandemic,» *medRxiv*, 2020.
- [5] R. U. Ayres, «Barriers and breakthroughs: an “expanding frontiers” model of the technology-industry life cycle,» *Technovation*, τόμ. 7, pp. 87-115, 1988.
- [6] A. G. Investors, «“The Sixth Kondratieff”-long waves of prosperity,» 2010. [Ηλεκτρονικό]. Available: https://www.allianz.com/content/dam/onemarketing/azcom/Allianz_com/migration/media/press/document/other/kondratieff_en.pdf.
- [7] ESRI, «Understanding outliers in time series analysis,» 2020. [Ηλεκτρονικό]. Available: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/understanding-outliers-in-time-series-analysis.htm>.
- [8] T. Bex, «How to Detect Random Walk and White Noise in Time Series Forecasting. Find out if the target is worth forecasting,» 2021. [Ηλεκτρονικό]. Available: <https://towardsdatascience.com/how-to-detect-random-walk-and-white-noise-in-time-series-forecasting-bdb5bbd4ef81>.
- [9] Φ. Ασημακόπουλος Β. Πετρόπουλος, *Επιχειρησιακές προβλέψεις*, Εκδόσεις Συμμετρία,

2015.

- [10] C. Zaiontz, «Basic Concepts of Correlation,» 2015. [Ηλεκτρονικό]. Available: <http://www.real-statistics.com/correlation/basic-concepts-correlation/>.
- [11] Η. Κατσούφης, Α. Κώνστα, Κ. Χριστοδουλίδης, Δ. Νταουκάκη, Β. Πέογλος, Ρ. Βλαστού, Κ. Παπαδόπουλος, Γ. Ράπτης, Π. Πίσσης, Ε. Αναστασάκης, Κ. Στεφανής, Λ. Απέκης, Μ. Μακροπούλου και Δ. Πίτλιγγερ, Εργαστηριακές Ασκήσεις Φυσικής, Εκδόσεις Συμμετρία, 2010.
- [12] P. R. Winters, «Forecasting Sales by Exponentially Weighted Moving Averages,» *Management Science*, τόμ. 6, pp. 324-342, 1960.
- [13] V. Assimakopoulos και Κ. Nikolopoulos, «The theta model: a decomposition approach to forecasting,» *International Journal of Forecasting*, τόμ. 16, pp. 521-530, 2000.
- [14] S. Makridakis και M. Hibon, «The M3-Competition: results, conclusions and implications,» *International Journal of Forecasting*, τόμ. 16, pp. 451-476, 2000.
- [15] R. Nau., «ARIMA models for time series forecasting,» [Ηλεκτρονικό]. Available: <https://people.duke.edu/rnau/411arim.htm>.
- [16] J. Scott Armstrong και S. P. Schnaars, «Long-range forecasting: From crystal ball to computer, 2nd ed. (Wiley, New York, 1985) [UK pound]22.95 (paper), pp. 689,» *International Journal of Forecasting*, τόμ. 2, pp. 387-390, 1986.
- [17] P. Goodwin και R. Lawton, «On the asymmetry of the symmetric MAPE,» *International Journal of Forecasting*, τόμ. 15, pp. 405-408, 1999.
- [18] R. J. Hyndman και A. B. Koehler, «Another look at measures of forecast accuracy,» *International Journal of Forecasting*, τόμ. 22, pp. 679-688, 2006.
- [19] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen και R. Winkler, «The accuracy of extrapolation (time series)

- methods: Results of a forecasting competition,» *Journal of Forecasting*, τόμ. 1, pp. 111-153, 1982.
- [20] M. D. Geurts και J. P. Kelly, «Forecasting retail sales using alternative models,» *International Journal of Forecasting*, τόμ. 2, pp. 261-272, 1986.
- [21] C. Chatfield, «A personal view of the M2-competition,» *International Journal of Forecasting*, τόμ. 9, pp. 23-24, 1993.
- [22] S. Makridakis, «Accuracy measures: theoretical and practical concerns,» *International Journal of Forecasting*, τόμ. 9, pp. 527-529, 1993.
- [23] S. Makridakis, E. Spiliotis και V. Assimakopoulos, «The M4 Competition: Results, findings, conclusion and way forward,» *International Journal of Forecasting*, τόμ. 34, pp. 802-808, 2018.
- [24] S. Makridakis, E. Spiliotis και V. Assimakopoulos, «The M4 Competition: 100,000 time series and 61 forecasting methods,» *International Journal of Forecasting*, τόμ. 36, pp. 54-74, 2020.
- [25] S. F. Crone, M. Hibon και K. Nikolopoulos, «Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction,» *International Journal of Forecasting*, τόμ. 27, pp. 635-660, 2011.
- [26] K. Nikolopoulos και F. Petropoulos, «Forecasting for big data: Does suboptimality matter?,» *Computers & Operations Research*, τόμ. 98, pp. 322-329, 2018.
- [27] S. Makridakis, E. Spiliotis και V. Assimakopoulos, «The M5 competition: Background, organization, and implementation,» *International Journal of Forecasting*, 2021.
- [28] Kaggle, «M5 Forecasting-Accuracy. Estimate the unit sales of Walmart retail goods,» 2020. [Ηλεκτρονικό]. Available: <https://www.kaggle.com/c/m5-forecasting-accuracy>.
- [29] S. Makridakis, E. Spiliotis και V. Assimakopoulos, «The M5 Accuracy competition:

Results, findings and conclusions,» October 2020.

- [30] «The M6 financial forecasting competition,» 2021. [Ηλεκτρονικό]. Available: <https://mofc.unic.ac.cy/the-m6-competition>.
- [31] S. Russell και P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th US ed. επιμ., Prentice Hall, 2010.
- [32] F. Chollet, *Deep Learning with Python*, Manning, 2017.
- [33] P. McCorduck, *Machines Who Think*, (2nd ed.) επιμ., Natick, MA: A. K. Peters, Ltd., 2004.
- [34] T. Mitchell, *Machine Learning*, New York: McGraw Hill., 1997.
- [35] J. Hu, H. Niu, J. Carrasco, B. Lennox και F. Arvin, «Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning,» *IEEE Transactions on Vehicular Technology*, τόμ. 69, pp. 14413-14423, 2020.
- [36] E. García-Gonzalo, Z. Fernández-Muñiz, P. J. Garcia Nieto, A. Sánchez και M. Menéndez, «Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Classifiers,» *Materials*, τόμ. 9, p. 531, June 2016.
- [37] T. Joachims, «Text categorization with Support Vector Machines: Learning with many relevant features,» σε *Machine Learning: ECML-98*, Berlin, 1998.
- [38] J. Kaur και J. R. Saini, «A study of text classification natural language processing algorithms for Indian languages,» *VNSGU J Sci Technol*, τόμ. 4, p. 162–167, 2015.
- [39] K. R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen και V. Vapnik, «Predicting time series with support vector machines,» σε *Artificial Neural Networks — ICANN'97*, Berlin, 1997.

- [40] I. Karsznia και K. Wereszczyńska, «When Traditional Selection Fails: How to Improve Settlement Selection for Small-Scale Maps Using Machine Learning,» *ISPRS International Journal of Geo-Information*, τόμ. 9, p. 230, April 2020.
- [41] A. Nikhil και B. Techstudent, «A Predictive Analytic study on stock market trend by supervised machine learning algorithms,» 2020.
- [42] G. James, D. Witten, T. Hastie και R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, Springer, 2013.
- [43] J. Dunstan, M. Aguirre, M. Bastías, C. Nau, T. Glass και F. Tobar, «Predicting nationwide obesity from food sales using machine learning,» *Health Informatics Journal*, τόμ. 26, pp. 653-655, May 2019.
- [44] L. Rokach και O. Maimon, «Clustering Methods.,» σε *The Data Mining and Knowledge Discovery Handbook*, O. Maimon και L. Rokach, Επιμ., Springer, 2005, pp. 321-352.
- [45] Simplilearn, «What is Perceptron: A Beginners Guide for Perceptron,» 2022. [Ηλεκτρονικό]. Available: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/perceptron>.
- [46] D. Long, S. Wuest, J. Williams, R. Rauwendaal και M. Bailey, «Contour Planting: A Strategy to Reduce Soil Erosion on Steep Slopes,» 2010.
- [47] F. Rosenblatt, «The perceptron - A perceiving and recognizing automaton,» Ithaca, 1957.
- [48] N. Ceneda, «Quantile Regression of High-Frequency Data Tail Dynamics via a Recurrent Neural Network,» 2020.
- [49] Z. Knok, K. Pap και M. Hrcic, «Implementation of intelligent model for pneumonia detection,» *Tehnički glasnik*, τόμ. 13, pp. 315-322, December 2019.
- [50] R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit,"

Nature, vol. 408, 21 December 2000.

- [51] R. Hahnloser και Η. S. Seung, «Permitted and Forbidden Sets in Symmetric Threshold-Linear Networks,» σε *Advances in Neural Information Processing Systems*, 2000.
- [52] X. Glorot, A. Bordes και Y. Bengio, «Deep Sparse Rectifier Neural Networks,» σε *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, USA, 2011.
- [53] Y. Yucesan, A. Von Zuben, F. Viana και J. Mahfoud, «Estimating Parameters and Discrepancy of Computer Models with Graphs and Neural Networks,» 2020.
- [54] R. Gençay και M. Qi, «Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging,» *Neural Networks, IEEE Transactions on*, τόμ. 12, pp. 726-734, August 2001.
- [55] K. E. Pilario, Y. Cao και M. Shafiee, «A Kernel Design Approach to Improve Kernel Subspace Identification,» *IEEE Transactions on Industrial Electronics*, τόμ. PP, pp. 1-1, May 2020.
- [56] M. Parsa, C. Schuman, P. Date, D. Rose, B. Kay, J. Mitchell, S. Young, R. Dellana, W. Severa, T. Potok και K. Roy, *Hyperparameter Optimization in Binary Communication Networks for Neuromorphic Deployment*, 2020.
- [57] S. Mishra, «Nonlinear System Identification Using Functional Link Multilayer Perceptron Artificial Neural Networks,» *Nonlinear System Identification Using Functional Link Multilayer Perceptron Artificial Neural Networks*, τόμ. 10, pp. 31542-31546, January 2015.
- [58] T. Hastie, R. Tibshirani και J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2 επιμ., Springer, 2009.
- [59] G. Cybenko, «Approximation by superpositions of a sigmoidal function,» *Mathematics of Control, Signals, and Systems (MCSS)*, τόμ. 2, p. 303–314, 1 December 1989.

- [60] F. Thiesing και O. Vornberger, «Time Series Prediction by Neural Networks,» June 1995.
- [61] I. Goodfellow, Y. Bengio και A. Courville, Deep Learning, MIT Press, 2016.
- [62] P.-L. Pröve, *An introduction to different Types of Convolutions in Deep Learning*, 2017.
- [63] R. Chandra, S. Goyal και R. Gupta, «Evaluation of Deep Learning Models for Multi-Step Ahead Time Series Prediction,» *IEEE Access*, τόμ. PP, pp. 1-1, May 2021.
- [64] F. Lässig, *Temporal Convolutional Networks and Forecasting*, 2021.
- [65] D. Impedovo, V. Dentamaro, G. Pirlo και L. Sarcinella, «TrafficWave: Generative Deep Learning Architecture for Vehicular Traffic Flow Prediction,» *Applied Sciences*, τόμ. 9, December 2019.
- [66] A. Borovykh, S. Bohte και C. W. Oosterlee, *Conditional Time Series Forecasting with Convolutional Neural Networks*, 2018.
- [67] D. Patterson, J. Gonzalez, U. Hölzle, Q. H. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier και J. Dean, «The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink,» April 2022.
- [68] S. Wang, C. Li και A. Lim, *Why Are the ARIMA and SARIMA not Sufficient*, arXiv, 2019.
- [69] E. Parzen, *Stochastic Processes*, Dover Publications, 2015.
- [70] keras.io, *Why choose Keras?*, 2021.
- [71] R. J. Hyndman και A. B. Koehler, «Another look at measures of forecast accuracy,» *International Journal of Forecasting*, τόμ. 22, pp. 679-688, 2006.
- [72] J. Everette S. Gardner και E. Mckenzie, «Forecasting Trends in Time Series,»

Management Science, τόμ. 31, pp. 1237-1246, 1985.

[73] T. J. B. community, *Jupyter Meets the Earth*, 2020.

[74] G. E. P. Box και G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day and Wisconsin Madison Uni. Dept. of Statistics., 1970.

[75] J. M. Bland και D. G. Altman, «Statistics notes: measurement error,» *BMG*, τόμ. 312 (7047), p. 1654, 1996.