



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

Εφαρμογή τεχνικών μεταφοράς μάθησης για τη  
βελτίωση της προβλεπτικής ακρίβειας  
νευρωνικών δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

των

ΚΡΙΤΣΩΤΑΚΗ Μ. ΕΛΕΥΘΕΡΙΟΥ  
ΣΤΑΘΑ Ι. ΕΥΣΤΑΘΙΑ

Επιβλέπων: Ασημακόπουλος Βασίλειος  
Καθηγητής Ε.Μ.Π.

Υπεύθυνος : Ευάγγελος Σπηλιώτης, Αρτέμιος Ανάργυρος Σεμένογλου  
Διδάκτορας Ε.Μ.Π. Υ.Δ. Ε.Μ.Π.

Αθήνα, Ιούλιος 2021





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων

# Εφαρμογή τεχνικών μεταφοράς μάθησης για τη βελτίωση της προβλεπτικής ακρίβειας νευρωνικών δικτύων

## ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

των

**ΚΡΙΤΣΩΤΑΚΗ Μ. ΕΛΕΥΘΕΡΙΟΥ  
ΣΤΑΘΑ Ι. ΕΥΣΤΑΘΙΑ**

**Επιβλέπων:** Ασημακόπουλος Βασίλειος  
Καθηγητής Ε.Μ.Π.

**Υπεύθυνος :** Ευάγγελος Σπηλιώτης, Αρτέμιος Ανάργυρος Σεμένογλου  
Διδάκτορας Ε.Μ.Π. Υ.Δ. Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16 Ιουλίου 2021.

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

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

.....  
Ασκούνης Δημήτριος  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2021





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων

Copyright © Ελευθέριος Μ. Κριτσωτάκης, Ευσταθία Ι. Σταθά, 2021.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

.....  
**Ελευθέριος Μ. Κριτσωτάκης**

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

.....  
**Ευσταθία Ι. Σταθά**

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



# Περίληψη

Στην εποχή μας, οι προβλέψεις συνιστούν, πιο πολύ από ποτέ, ένα αναπόσπαστο κομμάτι της καθημερινότητας του ανθρώπου. Αποφάσεις μικρής κλίμακας, αλλά και μεγάλες επιχειρηματικές δράσεις, απαιτούν την παραγωγή ποιοτικών προβλέψεων, με αποτέλεσμα η επιστήμη των προβλέψεων να γνωρίζει διαρκή ανάπτυξη. Προσφάτως, συντελέστηκε ένα από τα βασικότερα βήματα στην πορεία εξέλιξης του κλάδου, καθώς ενσωματώθηκε σε αυτόν η χρήση τεχνικών μηχανικής μάθησης. Η ίδια παρέχει μια πληθώρα νέων προσεγγίσεων, δίνοντας τη δυνατότητα για καθολική χρήση των δεδομένων, αξιοποιώντας τη γνώση μεταξύ χρονοσειρών.

Στα πλαίσια της αξιοποίησης γνώσης μεταξύ χρονοσειρών, μια τεχνική που χρήζει έρευνας είναι η χρήση της μεταφοράς μάθησης μεταξύ διαφορετικών συνόλων δεδομένων. Η παρούσα εργασία, εστίασε στη χρήση τέτοιων μεθόδων με την εκπαίδευση ενός υπερμοντέλου επί ενός συνόλου δεδομένων με ποικιλία χρονοσειρών, με σκοπό την παραγωγή ποιοτικών προβλέψεων για τις ετήσιες χρονοσειρές του συνόλου δεδομένων του M4. Σκοπός μας ήταν να ερευνήσουμε αν η ιδέα της μεταφοράς μάθησης θα μπορούσε να λειτουργήσει καλά στον τομέα των προβλέψεων και, επιπρόσθετα, ποιες μέθοδοι επανεκπαίδευσης κρίνονται περισσότερο κατάλληλες. Επιπλέον, ερευνήσαμε το αν η συγκεκριμένη μέθοδος προσφέρεται και για μη καθολική χρήση των δεδομένων του περιβάλλοντος στόχου, δοκιμάζοντας να επανεκπαιδεύσουμε το υπερμοντέλο αρχικά χωρίζοντας το σύνολο δεδομένων σε υποομάδες κι, έπειτα, χρησιμοποιώντας τις χρονοσειρές του με τοπικό τρόπο.

Με βάση τα πειράματά μας, εξήχθησαν ορισμένα συμπεράσματα ως προς την αξία της μεταφοράς μάθησης και τις δυνατότητες χρήσης της. Η ίδια αποδείχτηκε αρκετά χρήσιμη, καθώς οδήγησε σε μείωση του σφάλματος των προβλέψεων, ενώ οι απαιτήσεις της σε χρόνο και υπολογιστικούς πόρους δεν την κατέστησαν ανεπίτρεπτη. Μάλιστα, οι δύο μέθοδοι που απέδωσαν τα μέγιστα οφέλη, ήταν και αυτές με το μικρότερο κόστος. Τέλος, όλα τα αποτελέσματα που λάβαμε αποδεικνύουν πως η καθολική χρήση των δεδομένων που έχουμε στη διάθεσή μας, συνιστά την καλύτερη δυνατή επιλογή προς σκοπούς επανεκπαίδευσης, κατά τη χρήση τεχνικών μεταφοράς μάθησης.

## Λέξεις Κλειδιά

Πρόβλεψη Χρονοσειρών, Μηχανική Μάθηση, Μεταφορά Μάθησης, Νευρωνικά Δίκτυα, Πολυστρωματικά Δίκτυα Perceptron





# Abstract

In our times, forecasting constitutes, more than ever, an indispensable part of humans. Small-scale decisions and large corporate operations alike require the production of quality forecasts, leading to the continuous development of forecasting science. Following the trend of science, the forecasting domain realized one of its major steps by incorporating the use of machine learning. The latter offers a variety of new approaches giving, among others, the possibility of global use of data utilizing cross-series knowledge.

In the scope of utilizing cross-series knowledge, a technique to be researched is transfer learning between different datasets. The present work focuses on the use of such methods by training a supermodel on a dataset with a variety of time series aiming to produce quality forecasts for the M4 dataset. Our goal was to examine whether the idea of transfer learning could perform adequately on the task of forecasting and, in addition, which retraining methods are deemed more appropriate. On top of that, we investigated whether transfer learning is suitable for non global use of the target dataset by retraining the supermodel firstly, on a split subgroup of the target dataset and secondly, on its time series locally.

Based on our experiments, some conclusions were deduced relevant to the value of transfer learning and its possible uses. The latter was proven to be rather helpful, as it decreased the forecasting error while its requirements in time and computing resources didn't render it inapplicable. As a matter of fact, the two methods that yielded the best results were those bearing the least computational cost. Lastly, all our results signify that a global use of the data in our disposal is the best possible choice with regards to retraining when using transfer learning techniques.

## Keywords

Time series Forecasting, Transfer Learning, Machine Learning, Neural Networks, Multi-layer Perceptron



# Ευχαριστίες

Με την ολοκλήρωση της παρούσας εργασίας και, παράλληλα, τη λήψη των διπλωμάτων μας ως Ηλεκτρολόγοι Μηχανικοί και Μηχανικοί Υπολογιστών, από το Εθνικό Μετσόβιο Πολυτεχνείο, θα θέλαμε να απευθύνουμε τις ευχαριστίες μας σε όσους συνέβαλαν στην επιτυχία μας.

Αρχικά, θα θέλαμε να ευχαριστήσουμε τον καθηγητή κ. Βασίλειο Ασημακόπουλο που μας εμπιστεύτηκε και μας έδωσε την ευκαιρία να συνεργαστούμε ερευνητικά με τη Μονάδα Προβλέψεων και Στρατηγικής, αναθέτοντας μας αυτή την εργασία. Παράλληλα, θα θέλαμε να ευχαριστήσουμε τον καθηγητή κ. Ιωάννη Ψαρρά και τον καθηγητή κ. Δημήτριο Ασκούνη για τη συμμετοχή τους στην επιτροπή εξέτασης της εργασίας.

Επιπρόσθετα, θα θέλαμε να ευχαριστήσουμε θερμά τον Υ.Δ. Αρτέμη Σεμένογλου, καθώς και τον Δρ. Βαγγέλη Σπηλιώτη για την πολύτιμη καθοδήγηση και βοήθεια που μας προσέφεραν στα πλαίσια εκπόνησης της διπλωματικής μας εργασίας.

Στο σημείο αυτό, εγώ, ο Λευτέρης, θα ήθελα να εκφράσω την ευγνωμοσύνη μου για την αμέριστη φροντίδα της οικογένειάς μου όλα αυτά τα χρόνια. Αφιερώνω λοιπόν το πόνημα αυτό στην μητέρα μου Βάσω, στον πατέρα μου Μιχάλη και στην αδερφή μου Άννα-Μαρία. Ακόμη, ευχαριστώ την σύντροφό μου Αναστασία για την αδιάκοπη αγάπη της και τους φίλους μου Σωκράτη, Άμρ, Φίλιππο, Δημήτρη, Λευτέρη και Κλεοπάτρα για την πολύτιμη παρουσία τους στην καθημερινότητά μου.

Κλείνοντας, προσωπικά ως Ευσταθία, θα ήθελα να ευχαριστήσω τους γονείς μου Γιάννη και Ελένη για την οικονομική και ψυχολογική τους στήριξη στην πορεία των σπουδών μου, τις αδερφές μου Μάρω και Βίκη που στάθηκαν δίπλα μου σε όλη την πορεία της φοίτησης μου, καθώς και τους φίλους μου Αριάδνη Κ., Γιώργο Β., Μαλβίνα Σ., Μαρία Α., Χαράλαμπο Μ., Χρήστο Σ., που έδωσαν όμορφες νότες στα χρόνια της φοίτησης μου χαρίζοντας μου πληθώρα αναμνήσεων.



# Περιεχόμενα

Περίληψη	i
Abstract	iii
Ευχαριστίες	v
Περιεχόμενα	ix
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Σύγχρονη Κοινωνία και Προβλέψεις	1
1.2 Αντικείμενο της Εργασίας	2
1.3 Δομή της Εργασίας	2
<b>2 Τεχνικές Προβλέψεων και Χρονοσειρές</b>	<b>5</b>
2.1 Εισαγωγή	5
2.2 Χρονοσειρές	6
2.2.1 Συχνότητα Χρονοσειρών	6
2.2.2 Ποιοτικά Χαρακτηριστικά Χρονοσειρών	7
2.2.3 Αποσύνθεση Χρονοσειρών	8
2.2.4 Προεπεξεργασία Χρονοσειρών	10
2.3 Προβλέψεις	11
2.3.1 Είδη Προβλέψεων	11
2.3.2 Ορίζοντας Πρόβλεψης	12
2.3.3 Διαστήματα Εμπιστοσύνης	13
2.3.4 Συμβατικές Μέθοδοι Στατιστικής Πρόβλεψης	13
2.4 Στατιστικοί Δείκτες Ακρίβειας Προβλέψεων	15
<b>3 Μηχανική Μάθηση και Νευρωνικά Δίκτυα</b>	<b>19</b>
3.1 Εισαγωγή	19
3.2 Ιστορική Αναδρομή	20
3.3 Μηχανική Μάθηση	20
3.3.1 Κατηγορίες Αλγορίθμων Μηχανικής Μάθησης	20
3.3.2 Μοντέλα Μηχανικής Μάθησης	21

3.4	Νευρωνικά Δίκτυα . . . . .	22
3.4.1	Τεχνητός Νευρώνας Perceptron . . . . .	22
3.4.2	Συναρτήσεις Ενεργοποίησης . . . . .	23
3.4.3	Πολυστρωματικά Δίκτυα Perceptron . . . . .	24
3.4.4	Ο Αλγόριθμος Backpropagation . . . . .	26
3.4.5	Συναρτήσεις Απώλειας . . . . .	27
3.4.6	Αλγόριθμοι Βελτιστοποίησης . . . . .	27
<b>4</b>	<b>Μεταφορά Μάθησης</b>	<b>31</b>
4.1	Εισαγωγή . . . . .	31
4.2	Βασικοί Ορισμοί και Είδη Μεταφοράς Μάθησης . . . . .	33
4.3	Περαιτέρω Διάκριση Ειδών Μεταφοράς Μάθησης . . . . .	35
4.3.1	Με βάση τον Χώρο Χαρακτηριστικών $\mathcal{X}$ . . . . .	35
4.3.2	Με βάση την Οριακή Κατανομή Πιθανότητας $P(X)$ . . . . .	35
4.3.3	Με βάση την Δεσμευμένη Κατανομή Πιθανότητας $P(Y X)$ . . . . .	36
4.3.4	Με βάση τον Χώρο Ετικετών $\mathcal{Y}$ . . . . .	36
4.4	Ενδεικτικές Μέθοδοι Μεταφοράς Μάθησης . . . . .	36
4.4.1	Ευνοϊκό Ξεκίνημα . . . . .	36
4.4.2	Φραγμένες Εποχές . . . . .	37
4.4.3	Πάγωμα Επιπέδων . . . . .	37
4.4.4	Αντικατάσταση της Κεφαλής . . . . .	38
<b>5</b>	<b>Πειραματική Διαδικασία</b>	<b>39</b>
5.1	Εισαγωγή . . . . .	39
5.2	Σύνολα Δεδομένων: M4 και ForeDeCk . . . . .	39
5.2.1	Δεδομένα . . . . .	39
5.3	Προετοιμασία Πειραμάτων . . . . .	41
5.3.1	Επεξεργασία και Παραθυροποίηση Δεδομένων . . . . .	41
5.3.2	Οπτικοποίηση Δεδομένων με Βάση τα Χαρακτηριστικά . . . . .	43
5.4	Μοντέλο Βάσης . . . . .	44
5.5	Benchmarks . . . . .	46
<b>6</b>	<b>Αποτελέσματα Πειραμάτων</b>	<b>49</b>
6.1	Εισαγωγικές Παρατηρήσεις . . . . .	49
6.2	Πειράματα με χρήση του ForeDeCk . . . . .	50
6.2.1	Ευνοϊκό Ξεκίνημα . . . . .	50
6.2.2	Φραγμένες Εποχές . . . . .	52
6.2.3	Συγκέντρωση Αποτελεσμάτων . . . . .	53
6.3	Δημιουργία του Συνόλου Δεδομένων Left SubForeDeCk . . . . .	54
6.4	Πειράματα σε επίπεδο εκπαίδευσης με χρήση του Left SubForeDeCk . . . . .	55
6.4.1	Ευνοϊκό Ξεκίνημα . . . . .	55
6.4.2	Φραγμένες Εποχές . . . . .	56

---

6.4.3	Πάγωμα Επιπέδων . . . . .	58
6.4.4	Επαναρχικοποίηση του Επιπέδου Εξόδου . . . . .	59
6.4.5	Συγκέντρωση των αποτελεσμάτων . . . . .	60
6.5	Πειράματα σε επίπεδο δεδομένων με χρήση του Left SubForeDeCk . . . . .	61
6.5.1	Ανά Χρονοσειρά . . . . .	62
6.5.2	Κόψιμο σε τεταρτημόρια . . . . .	62
6.5.3	Κοντινότερα Κεντροειδή . . . . .	64
6.5.4	Συγκέντρωση αποτελεσμάτων . . . . .	65
<b>7</b>	<b>Συμπεράσματα και Προεκτάσεις</b>	<b>67</b>
7.1	Συμπεράσματα . . . . .	67
7.2	Περιορισμοί και Προεκτάσεις . . . . .	69
	<b>Bibliography</b>	<b>71</b>





# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Σύγχρονη Κοινωνία και Προβλέψεις

Στην σύγχρονη εποχή, η παραγωγή αξιόπιστων και άμεσα διαθέσιμων προβλέψεων έχει αποκτήσει εξαιρετικά μεγάλη σπουδαιότητα. Τόσο οι μικρές επιχειρήσεις, οι μεγάλες πολυεθνικές εταιρείες όσο και οι διακρατικές οργανώσεις βασίζονται στον κλάδο των προβλέψεων για μπορέσουν να ανταποκριθούν με τρόπο βέλτιστο στις απαιτήσεις του μέλλοντος. Η ενέργεια, η βιομηχανία, τα καταναλωτικά αγαθά, οι ενδοεπιχειρησιακές διαδικασίες και ο καιρός αποτελούν ορισμένους από τους τομείς εφαρμογής του κλάδου των προβλέψεων, ο οποίος βρίσκει εφαρμογή, πέραν των παραπάνω, σε μια πληθώρα άλλων διαδικασιών.

Για τον λόγο αυτό, το ενδιαφέρον της ερευνητικής κοινότητας για τον κλάδο των προβλέψεων συνεχώς αυξάνεται ενώ νέες καινοτόμες μέθοδοι προτείνονται συνεχώς. Ακολουθώντας την γενικότερη τάση του κλάδου της επιστήμης των υπολογιστών και την ραγδαία ανάπτυξη της τεχνητής νοημοσύνης και της μηχανικής μάθησης, τα τελευταία χρόνια παρατηρείται μία έμφαση στην εξερεύνηση των μεθόδων μηχανικής μάθησης πάνω στον κλάδο των προβλέψεων.

Ιστορικά, η πρόβλεψη χρονοσειρών έκανε χρήση στατιστικών μεθόδων για την εξαγωγή συμπερασμάτων. Ωστόσο, στις αρχές του 21ου αιώνα φαίνεται πως η μηχανική μάθηση αποκτά πρωτεύοντα ρόλο στο πεδίο των προβλέψεων [1]. Αρχικά, χρησιμοποιήθηκαν θεμελιώδεις αρχιτεκτονικές όπως τα δίκτυα MLP ενώ στην πορεία η έρευνα φαίνεται να έχει προσανατολιστεί σε μοντέλα βαθύτερων αρχιτεκτονικών (deep learning) όπως για παράδειγμα είναι τα βαθιά δίκτυα MLP (DMLP), τα RNN, τα LSTM οι περιορισμένες μηχανές Boltzmann (restricted Boltzmann machines), τα βαθιά δίκτυα πίστης (deep belief networks) και οι αυτοκωδικοποιητές (autoencoders) [2–4]. Τέλος, αξίζει να αναφερθεί πως έχουν προταθεί αρκετές υβριδικές μέθοδοι που συνδυάζουν τόσο την στατιστική πρόβλεψη όσο και την μηχανική μάθηση. Παράδειγμα τέτοιας μεθόδου αποτελεί η νικητρία υποβολή στον διαγωνισμό προβλέψεων M4 [5].

## 1.2 Αντικείμενο της Εργασίας

Η παρούσα εργασία επικεντρώνεται στην χρήση νευρωνικών δικτύων, μίας από τις κατηγορίες μοντέλων μηχανικής μάθησης, τα οποία ενδείκνυνται για προβλήματα παλινδρόμησης. Κατά συνέπεια, οι προβλέψεις χρονοσειρών είναι ένα από τα πεδία στα οποία βρίσκουν εξαιρετική εφαρμογή τα νευρωνικά δίκτυα.

Για την πρόβλεψη χρονοσειρών είναι απαραίτητος ο καθαρισμός των δεδομένων, η συμπλήρωση τυχόν απουσιαζόντων τιμών καθώς και η τυποποίησή τους, δηλαδή ο σχηματισμός χρονικών παραθύρων εισόδου και εξόδου με συγκεκριμένες διαστάσεις. Επομένως, ως είσοδος του νευρωνικού δικτύου λαμβάνεται ένας αριθμός από κάποιες παρατηρήσεις που αντιπροσωπεύουν παρελθοντικές τιμές και ως έξοδος ένας αριθμός από παρατηρήσεις που έπονται οι οποίες αντιπροσωπεύουν την χρονική εξέλιξη του φαινομένου. Αξίζει να αναφερθεί πως για την εξαγωγή τέτοιων χρονικών παραθύρων μπορούμε να χρησιμοποιήσουμε πληροφορία αποκλειστικά προερχόμενη από την χρονοσειρά που επιθυμούμε να προβλέψουμε ή να εκμεταλλευτούμε και πληροφορία από άλλες χρονοσειρές. Στην πρώτη περίπτωση αναφερόμαστε σε τοπική εκπαίδευση (local) ενώ στην δεύτερη σε δια-μαθησιακή εκπαίδευση (cross-learning). Στην παρούσα διπλωματική εργασία εξετάζουμε μεθόδους εκπαίδευσης που βασίζονται τόσο στις αρχές καθολικής εκμετάλλευσης της γνώσης όσο και στην αξιοποίηση της τοπικής πληροφορίας.

Πιο συγκεκριμένα, η μέθοδος που εξετάζουμε αφορά στην αξιοποίηση γνώσης που έχει αποκτηθεί από την εκμάθηση κάποιας διαφορετικής αλλά συναφής δραστηριότητας και την μεταφορά της γνώσης αυτής στο εν λόγω πρόβλημα πρόβλεψης. Η τεχνική αυτή είναι γνωστή στον κλάδο της μηχανικής μάθησης ως μεταφορά μάθησης.

## 1.3 Δομή της Εργασίας

Τα κεφάλαια δύο, τρία και τέσσερα αποτελούν το θεωρητικό έρεισμα της παρούσας διπλωματικής εργασίας. Στο δεύτερο κεφάλαιο παραθέτουμε τις βασικές αρχές της θεωρίας των προβλέψεων, όπως είναι λόγου χάριν τα ποιοτικά χαρακτηριστικά μιας χρονοσειράς και αναφερόμαστε συνοπτικά στις θεμελιώδεις στατιστικές μεθόδους πρόβλεψων. Ακόμη, ορίζουμε έννοιες στις οποίες θα βασιστούμε στην συνέχεια όπως είναι για παράδειγμα οι στατιστικοί δείκτες σφάλματος. Στο τρίτο κεφάλαιο εισάγουμε την θεωρία της μηχανικής μάθησης και των νευρωνικών δικτύων, δίνοντας βάση τόσο στις αρχές λειτουργίας τους όσο και στα δομικά συστατικά τους στοιχεία ενώ παράλληλα αναφερόμαστε στους σημαντικότερους αλγόριθμους που διέπουν την λειτουργία τους. Στο τέταρτο κεφάλαιο, θεμελιώνουμε θεωρητικά την τεχνική της μεταφοράς μάθησης αναφέροντας αναλυτικά τα είδη και τις στρατηγικές με τις οποίες αυτή επιτυγχάνεται ενώ ταυτόχρονα εισάγονται μερικές από τις μεθόδους μεταφοράς μάθησης που έχουν προταθεί στην βιβλιογραφία.

Τα κεφάλαια πέντε, έξι και επτά απαρτίζουν το πειραματικό μέρος της παρούσας διπλωματικής εργασίας. Στο κεφάλαιο πέντε αναλύουμε εκτενώς την πειραματική διαδικασία που

ακολουθήθηκε, τα σύνολα δεδομένων που χρησιμοποιήθηκαν, τις τεχνικές προετοιμασίας, επεξεργασίας και μετασχηματισμού των δεδομένων, την αρχιτεκτονική του νευρωνικού δικτύου που εφαρμόστηκε καθώς και την δομή των πειραμάτων που εκτελέστηκαν. Τέλος εισάγονται και δύο ακόμα μέθοδοι πρόβλεψης ως σημεία αναφοράς και σύγκρισης με τα δικά μας μοντέλα. Στο κεφάλαιο έξι παραθέτονται τα αποτελέσματα καθενός από τα εκτελεσθέντα πειράματα έχοντας σαν βασικό άξονα το σύνολο δεδομένων που χρησιμοποιήθηκε καθώς και την τεχνική μεταφοράς μάθησης που εφαρμόστηκε. Στο κεφάλαιο επτά γίνεται μελέτη και ερμηνεία των πειραματικών αποτελεσμάτων και εισάγονται τα βασικά συμπεράσματα που εξήχθησαν. Τέλος, γίνεται αναφορά στους εγγενείς περιορισμούς που συναντήθηκαν κατά την κατάστροση και εκτέλεση των πειραμάτων ενώ προτείνονται πιθανές προεκτάσεις προς διερεύνηση και πιθανές εφαρμογές των μέχρι τώρα συμπερασμάτων.



## Κεφάλαιο 2

# Τεχνικές Προβλέψεων και Χρονοσειρές

### 2.1 Εισαγωγή

Οι προβλέψεις είναι ένα αναπόσπαστο κομμάτι της καθημερινότητας των ανθρώπων. Είτε πρόκειται για προβλέψεις μικρής κλίμακας, όπως η πρόβλεψη του χρόνου που θα μεσολαβήσει από τη στιγμή εκκίνησης του δρομολογίου με το λεωφορείο έως τη στιγμή άφιξης στον τελικό προορισμό, είτε για προβλέψεις μεγάλης κλίμακας και επιχειρηματικού ενδιαφέροντος, όπως η πορεία κάποιου δείκτη στο χρηματιστήριο, οι άνθρωποι ερχόμαστε καθημερινά σε επαφή με την έννοια της πρόβλεψης. Παρά την προηγούμενη παρατήρηση, η επιστήμη των προβλέψεων είναι μια από τις πλέον πρόσφατα ανεπτυγμένες και πηγάζει από την, εξίσου πρόσφατη, επιστήμη των πιθανοτήτων και της στατιστικής.

Λόγω της μεγάλης της σημασίας και εφαρμοσιμότητας, η επιστήμη των προβλέψεων γνωρίζει ραγδαία ανάπτυξη, με ένα μεγάλο πλήθος ερευνητών να παρουσιάζουν νέες προτάσεις και μεθόδους ανά τακτά χρονικά διαστήματα. Αν και η πρόοδος στον τομέα είναι ήδη αξιοσημείωτη, υπάρχει ακόμα μεγάλο περιθώριο εξέλιξης.

Η ίδια συνιστά κομμάτι της επιχειρησιακής έρευνας και χρησιμοποιείται ευρέως για την υποστήριξη σημαντικών αποφάσεων, τόσο σε μικρές όσο και σε μεγάλες επιχειρήσεις. Ωστόσο, πάντα υπάρχει η πιθανότητα εσφαλμένων προβλέψεων και αυτές οφείλονται στην αναμφισβήτητη ύπαρξη της αβεβαιότητας. Παρά την ύπαρξη προτύπων στα περισσότερα μεγέθη των οποίων οι τιμές τίθενται προς πρόβλεψη, το παρελθόν μπορεί να μας βοηθήσει μόνο έως ένα βαθμό στην παραγωγή προβλέψεων. Δεν πρέπει ποτέ να ξεχνάμε τους πολυποίκιλους παράγοντες που επηρεάζουν τα περισσότερα μεγέθη και, επίσης, δεν πρέπει να αγνοούμε την αδυναμία μας να τους συνυπολογίσουμε κατά την παραγωγή προβλέψεων.

Μερικοί από τους βασικότερους κλάδους εφαρμογής της επιστήμης των προβλέψεων, επί του επιχειρηματικού κόσμου, είναι σύμφωνα με τη βιβλιογραφία [6], [7] οι ακόλουθοι:

- **Οικονομία και Χρηματοοικονομία**

- Ηλεκτρική Ενέργεια
- Περιβάλλον και Κλίμα
- Κοινωνικό Περιβάλλον
- Τουρισμός
- Μεταφορές και Μετακινήσεις
- Ακίνητα και Κτηματικές Περιουσίες

## 2.2 Χρονοσειρές

Με τον όρο χρονοσειρά (time series) αναφερόμαστε στην συγκεντρωτική καταγραφή της τιμής ενός μεγέθους προς παρατήρηση, ανά αυστηρά καθορισμένα χρονικά διαστήματα. Παραδείγματα χρονοσειρών είναι η καταγραφή των πωλήσεων ενός καταστήματος στο τέλος κάθε ημέρας, ή το πλήθος των αφίξεων πελατών σε ένα κατάστημα ανά ώρα λειτουργίας. Ένα από τα βασικότερα χαρακτηριστικά των χρονοσειρών είναι πως κάθε τιμή τους δεν είναι ανεξάρτητη των υπολοίπων, καθώς τα προς παρακολούθηση μεγέθη τείνουν να έχουν κάποια συγκεκριμένα χαρακτηριστικά. Έτσι, σε αρκετές περιπτώσεις οι παρελθοντικές τιμές μιας χρονοσειράς μπορούν να βοηθήσουν στην πρόβλεψη της μελλοντικής της πορείας.

### 2.2.1 Συχνότητα Χρονοσειρών

Χάρη στη φύση των χρονοσειρών να αφορούν φυσικά ή άλλα μεγέθη ένα βασικό χαρακτηριστικό τους είναι η ποικιλομορφία τους ως προς τη συχνότητα παρατήρησης. Μάλιστα, με βάση αυτό το χαρακτηριστικό μπορεί να γίνει διαχωρισμός τους, όπως φαίνεται ακολούθως:

- **Ετήσιες Χρονοσειρές:** Υπάρχει μια καταγεγραμμένη παρατήρηση για κάθε χρόνο.
- **Μηνιαίες Χρονοσειρές:** Υπάρχει μια καταγεγραμμένη παρατήρηση για κάθε μήνα.
- **Εβδομαδιαίες Χρονοσειρές:** Υπάρχει μια καταγεγραμμένη παρατήρηση για κάθε εβδομάδα.
- **Ημερήσιες Χρονοσειρές:** Υπάρχει μια καταγεγραμμένη παρατήρηση για κάθε μέρα.
- **Ωριαίες Χρονοσειρές:** Υπάρχει μια καταγεγραμμένη παρατήρηση για κάθε ώρα.

με το διαχωρισμό να μπορεί να προχωρήσει περαιτέρω. Ωστόσο, οι προαναφερθείσες συνιστούν τις πλέον συνηθεις χρονοσειρές.

Ο διαχωρισμός αυτός είναι αρκετά σημαντικός, καθώς πέραν της διαφοράς ως προς τη συχνότητα παρατήρησης, οι χρονοσειρές που ανήκουν σε διαφορετικές κατηγορίες τείνουν να

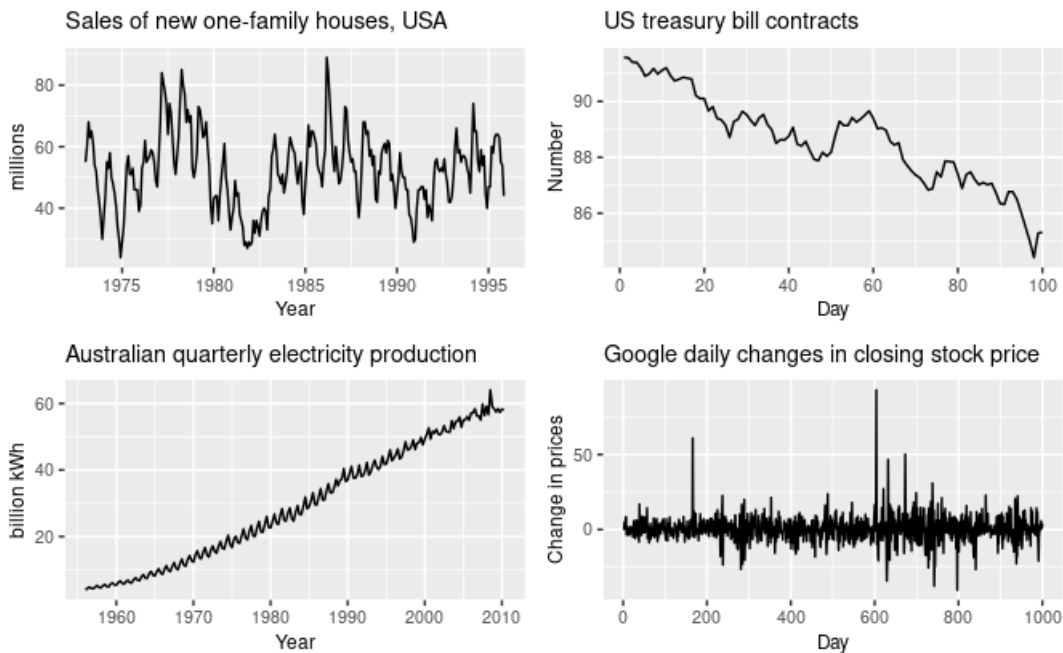
εμφανίζουν διαφορές και ως προς άλλα σημαντικά χαρακτηριστικά, όπως την ύπαρξη εποχιακής συμπεριφοράς για τις μηνιαίες χρονοσειρές και τη μη ύπαρξη για της χρονιαίες. Οι διαφορές αυτές, οφείλονται στη φύση των χρονοσειρών, καθώς η εξάρτηση μεταξύ των παρατηρήσεων αποκτά διαφορετική έννοια ανάλογα με τη συχνότητα καταγραφής των παρατηρήσεων. Επομένως, κρίνεται απαραίτητη η αναγνώριση της κατηγορίας στην οποία ανήκει κάθε χρονοσειρά πριν επιχειρηθεί η ανάλυση της.

### 2.2.2 Ποιοτικά Χαρακτηριστικά Χρονοσειρών

Οι χρονοσειρές μπορούν να αναλυθούν σε κάποια βασικά ποιοτικά χαρακτηριστικά, τα οποία είναι η τάση, η εποχικότητα, η κυκλικότητα και η τυχαιότητα, η οποία μπορεί να εμφανιστεί είτε με τη μορφή ασυνεχειών, είτε με οποιαδήποτε άλλη μορφή.

- **Τάση:** Η τάση είναι το ποιοτικό χαρακτηριστικό εκείνο που εκφράζει την μακροπρόθεσμη πορεία των τιμών του μεγέθους που μελετά η χρονοσειρά. Η ίδια μπορεί να είναι είτε ανοδική, είτε καθοδική, είτε ακόμα και σταθερή. Έχουν παρατηρηθεί χρονοσειρές στις οποίες δεν μπορούμε να μιλήσουμε καθαρά για ύπαρξη τάσης, καθώς σε διαφορετικά τους υποδιαστήματα έχουν διαφορετική συμπεριφορά ως προς αυτό το χαρακτηριστικό. Η τάση μπορεί να είναι γραμμική, εκθετική ή φθίνουσα. Για να μπορέσει να αναγνωριστεί η τάση μιας χρονοσειράς είναι απαραίτητη η ύπαρξη πληθώρας παρατηρήσεων για ένα επαρκώς μεγάλο χρονικό διάστημα ώστε να χαρακτηρίζεται “μακροπρόθεσμο”.
- **Εποχικότητα:** Το χαρακτηριστικό αυτό δεν υφίσταται σε όλες τις χρονοσειρές, αλλά μόνο σε αυτές που έχουν συχνότητα παρατήρησης τέτοια η οποία να επιτρέπει την εποχιακή συμπεριφορά. Στις χρονοσειρές που παρουσιάζουν το χαρακτηριστικό αυτό υπάρχει μια συσχέτιση ανάμεσα σε παρατηρήσεις που ανήκουν σε αντίστοιχες χρονικές περιόδους που απέχουν μεταξύ τους σταθερό διάστημα, μικρότερο του ενός έτους. Ένα παράδειγμα που μπορεί να διευκολύνει την κατανόηση του χαρακτηριστικού αυτού είναι μια χρονοσειρά που καταγράφει τις ενοικιάσεις σκαφών με μηνιαία συχνότητα. Αναμένουμε, οι πωλήσεις να είναι αυξημένες κατά τους θερινούς μήνες και μειωμένες κατά τους χειμερινούς.
- **Κυκλικότητα:** Η κυκλικότητα αφορά την εμφάνιση διακυμάνσεων στις τιμές της χρονοσειράς, χάρη σε διάφορους εξωγενείς παράγοντες. Οι διακυμάνσεις αυτές δεν ισαπέχουν μεταξύ τους και η εμφάνιση τους δεν μπορεί να προβλεφθεί. Παράλληλα, η κυκλική συμπεριφορά έχει νόημα να μελετάται σε διαστήματα ετών και, ως εκ τούτου, απαιτείται η ύπαρξη πληθώρας παρατηρήσεων για να μπορεί κανείς με βεβαιότητα να αναγνωρίσει αυτό το χαρακτηριστικό. Ένα σύνηθες λάθος είναι να συγχέεται η κυκλικότητα με την εποχικότητα, ωστόσο τα δύο αυτά χαρακτηριστικά διαφέρουν σε όσα προαναφέρθηκαν και έχουν ως μόνο κοινό τους το γεγονός πως οδηγούν σε επαναλαμβανόμενα μοτίβα στις τιμές μιας χρονοσειράς.

- Τυχαιότητα:** Το συγκεκριμένο χαρακτηριστικό αφορά τις διακυμάνσεις εκείνες που απομένουν όταν από μια χρονοσειρά αφαιρεθούν όλα τα υπόλοιπα χαρακτηριστικά που προαναφέρθηκαν. Η ίδια μπορεί να λαμβάνει τη μορφή είτε μιας εντελώς τυχαίας μεταβλητής, με τη στατιστική έννοια, είτε να εκφράζεται με τη μορφή ασυνεχειών και αλλοίωσης της κανονικότητας του προτύπου που υπάρχει στις τιμές του προς παρατήρηση μεγέθους. Στην δεύτερη περίπτωση, οι ασυνέχειες αυτές μπορούν να εμφανιστούν είτε παροδικά, με αποτέλεσμα την επαναφορά της χρονοσειράς μετά το πέρας τους στην “κανονικότητα” της, και οφείλονται σε απρόβλεπτα γεγονότα χωρίς διάρκεια, είτε με τη μορφή μόνιμων αλλαγών στο μοτίβο της. Οι πρώτες, ονομάζονται ειδικά γεγονότα (special events) ή outliers, ενώ οι δεύτερες ονομάζονται αλλαγές επιπέδου (level-shifts), καθώς πολλές φορές εμφανίζονται σαν απότομες αυξήσεις ή μειώσεις της μέσης τιμής της χρονοσειράς. Τόσο τα ειδικά γεγονότα, όσο και οι αλλαγές επιπέδου είναι μεν απρόβλεπτες, μπορούν όμως να εξηγηθούν με βάση κάποια γεγονότα. Το ίδιο δεν συμβαίνει με τις υπόλοιπες εκφάνσεις της τυχαιότητας σε μια χρονοσειρά.



Σχήμα 2.1: Παραδείγματα χρονοσειρών με έντονα τα ποιοτικά τους χαρακτηριστικά. Πάνω αριστερά παρατηρούμε μια χρονοσειρά με έντονη κυκλικότητα. Πάνω δεξιά έχουμε μια χρονοσειρά με έντονη τάση και εποχιακή συμπεριφορά, με αντίστοιχο μοτίβο να παρατηρείται και στην κάτω αριστερά χρονοσειρά. Κάτω δεξιά έχουμε μια χρονοσειρά με πολύ έντονη την ύπαρξη της τυχαιότητας. [8]

### 2.2.3 Αποσύνθεση Χρονοσειρών

Όπως προαναφέρθηκε, στις χρονοσειρές παρατηρούνται κάποιες βασικές συνιστώσες. Τα ίδια είναι τόσο συνήθη και σημαντικά, ώστε αντιμετωπίζονται ως συνιστώσες των χρονοσειρών και συχνά είναι αναγκαίος ο ποσοτικός προσδιορισμός τους. Η διαδικασία κατά την οποία αυτό επιτυγχάνεται ονομάζεται αποσύνθεση.



Η αποσύνθεση έχει εξέχουσα σημασία στην επιστήμη των προβλέψεων, καθώς ο υπολογισμός των συνιστωσών των χρονοσειρών μπορεί να μας βοηθήσει σε εναλλακτικές απεικονίσεις τους στο χώρο. Επιπρόσθετα, ο υπολογισμός αυτός δύναται να βοηθήσει και στην αναγνώριση ομοιοτήτων μεταξύ χρονοσειρών με αυτοματοποιημένο τρόπο με σκοπό την ομαδοποίηση τους.

Η ίδια διατυπώνεται μαθηματικά ως ακολούθως:

$$Y_t = f(S_t, T_t, C_t, R_t)$$

$Y_t$ : παρατήρηση την χρονική στιγμή  $t$

$S_t$ : συνιστώσα της εποχικότητας τη χρονική στιγμή  $t$

$T_t$ : συνιστώσα της τάσης τη χρονική στιγμή  $t$

$C_t$ : συνιστώσα της κυκλικότητας τη χρονική στιγμή  $t$

$R_t$ : συνιστώσα της τυχαιότητας τη χρονική στιγμή  $t$

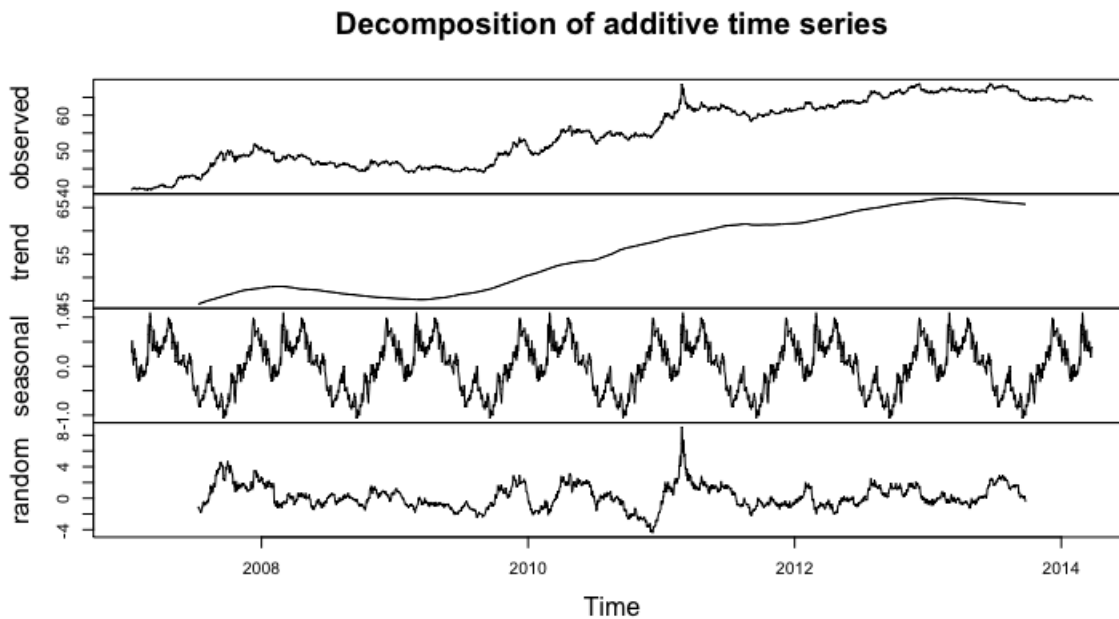
$f$ : συναρτησιακή σχέση μεταξύ της παρατήρησης και των συνιστωσών

Οι συνηθέστερες μορφές που έχει η συναρτησιακή σχέση  $f$  είναι η πολλαπλασιαστική και η προσθετική, ωστόσο η κλασσική μέθοδος αποσύνθεσης που αναφέρεται ακολούθως θεωρεί πολλαπλασιαστική σχέση των συνιστωσών.

Η διαδικασία της αποσύνθεσης έχει κάποια διακριτά αυστηρά καθορισμένα βήματα με βάση τα οποία επιτυγχάνεται ο ποσοτικός προσδιορισμός όλων των συνιστωσών σε μια χρονοσειρά. Τα βήματα αυτά αναφέρονται και επεξηγούνται ακολούθως:

1. Υπολογίζεται η χρονοσειρά τάσης-κύκλου, δηλαδή οι δύο αυτές συνιστώσες απομονώνονται από κοινού από τη χρονοσειρά. Αυτό επιτυγχάνεται με τη χρήση κάποιου τεχνητού μέσου όρου (ΚΜΟ) επί των παρατηρήσεων της χρονοσειράς. Ο ΚΜΟ αυτός έχει παράθυρο αντίστοιχο του μήκους της εποχικότητας της χρονοσειράς (στην περίπτωση εποχικών χρονοσειρών).
2. Διαιρείται η χρονοσειρά με τη χρονοσειρά τάσης-κύκλου προκειμένου να παραχθεί η χρονοσειρά εποχικότητας-τυχειότητας, η οποία αναφέρεται συνήθως με τον όρο “λόγοι εποχικότητας”.
3. Γίνεται εύρεση της συνιστώσας της εποχικότητας, γνωστή και με τον όρο “δείκτες εποχικότητας” από τους λόγους εποχικότητας. Αυτό επιτυγχάνεται με τον υπολογισμό του μέσου όρου των αντίστοιχων λόγων εποχικότητας
4. Απαλείφεται η εποχικότητα από τη χρονοσειρά και προσδιορίζεται η χρονοσειρά τάσης-κύκλου-τυχειότητας. Για την επίτευξη του προσδιορισμού αυτού, γίνεται διαίρεση της χρονοσειράς με τους δείκτες εποχικότητας που έχουν υπολογιστεί.

5. Υπολογίζεται η τάση της χρονοσειράς. Αυτό το βήμα απαιτεί την επιλογή του μοντέλου τάσης που αντιπροσωπεύει περισσότερο τη χρονοσειρά και την εφαρμογή του μαθηματικού του φορμαλισμού επί της χρονοσειράς για τον υπολογισμό της συνιστώσας της τάσης.
6. Υπολογίζεται η συνιστώσα της κυκλικότητας. Για την επίτευξη του υπολογισμού η χρονοσειρά τάσης-κύκλου διαιρείται με τη συνιστώσα της τάσης.



Σχήμα 2.2: Αποσύνθεση χρονοσειράς στα βασικά της χαρακτηριστικά. Στο πρώτο διάγραμμα βλέπουμε τη χρονοσειρά πριν την εφαρμογή της μεθόδου της αποσύνθεσης, με βάση τις τιμές παρατήρησης. Το δεύτερο διάγραμμα απεικονίζει την τάση της χρονοσειράς, ενώ το τρίτο της εποχικότητα της. Τέλος, βλέπουμε την συνιστώσα της τυχαιότητας. [9]

#### 2.2.4 Προεπεξεργασία Χρονοσειρών

Έχοντας εξηγήσει τι είναι μια χρονοσειρά, χρειάζεται να αναφερθούμε στην ύπαρξη πιθανών μηδενικών τιμών σε αυτή και να αναλύσουμε τη σημασία και τη διαχείριση τους. Συγκεκριμένα, καθώς κάθε μια τιμή μιας χρονοσειράς αναπαριστά την τιμή που είχε το προς παρατήρηση μέγεθος κατά τη στιγμή της παρατήρησης, μια μηδενική τιμή θα μπορούσε να μην θεωρείται ως ειδική περίπτωση. Ωστόσο, είναι αρκετά σύνηθες η μηδενική τιμή μιας παρατήρησης να σημαίνει έλλειψη καταγραφής της παρατήρησης, δηλαδή κενό. Στην περίπτωση αυτή είναι απαραίτητη η ειδική διαχείριση αυτής της μηδενικής τιμής και η αντικατάστασή της με μια τιμή που θα είχε περισσότερο νόημα. Υπάρχουν αρκετές προσεγγίσεις ως προς το πως μπορεί κανείς να αντικαταστήσει μια μηδενική τιμή, ενδεικτικά αναφέρουμε τις επικρατέστερες:

- Προσπάθεια προσδιορισμού της τιμής με κριτικό τρόπο, στην περίπτωση που είναι σαφές το επίπεδο στο οποίο η ίδια κινήθηκε.

- Χρήση του ημιαθροίσματος της προηγούμενης κι επόμενης τιμής από τη μηδενική, προς αντικατάσταση της.
- Χρήση μέσου όρου των τιμών για τις αντίστοιχες τιμές με την μηδενική από τις υπόλοιπες περιόδους, σε περίπτωση χρονοσειράς με εποχιακό χαρακτήρα.

Είναι σημαντικό, βέβαια, πριν την αντικατάσταση μιας μηδενικής τιμής να προσδιοριστεί αν η ίδια αντιστοιχεί πράγματι σε έλλειψη παρατήρησης ή είναι μια πραγματική και σωστή παρατήρηση.

Για τον προαναφερθέν προσδιορισμό μας βοηθάει η συνολική εικόνα και συμπεριφορά της χρονοσειράς. Συγκεκριμένα, υπάρχει ένα ειδικό είδος χρονοσειρών, οι οποίες έχουν διακοπτόμενη φύση και τις οποίες διαχειριζόμαστε με ειδικό τρόπο. Οι χρονοσειρές αυτές χαρακτηρίζονται από την ιδιότητα τους να έχουν αρκετές μηδενικές τιμές που εμφανίζονται με τυχαίο τρόπο και ποικίλες μη μηδενικές που δεν χαρακτηρίζονται τόσο από τα προαναφερθέντα ποιοτικά χαρακτηριστικά. Ένα προφανές παράδειγμα τέτοιων χρονοσειρών είναι η ημερήσια καταγραφή παραγγελιών ανταλλακτικών από κάποια εταιρεία.

## 2.3 Προβλέψεις

Η πρόβλεψη (forecasting) συνιστά την διαδικασία κατά την οποία γίνεται προσπάθεια προσδιορισμού της μελλοντικής πορείας μιας χρονοσειράς με βάση τις υπάρχουσες παρατηρήσεις γι' αυτή. Η ίδια είναι αρκετά πολύπλοκη, καθώς σε πολλές περιπτώσεις το παρελθόν δεν εκφράζει απόλυτα το μέλλον και υπάρχουν διάφορα λανθάνοντα μοτίβα που χρειάζεται να ληφθούν υπόψιν.

### 2.3.1 Είδη Προβλέψεων

Υπάρχουν τρία βασικά είδη προβλέψεων τα οποία εφαρμόζονται και η επιλογή τους γίνεται ανάλογα με την εκάστοτε εφαρμογή.

- **Στατιστική Πρόβλεψη (Statistical Forecasting):** Το είδος αυτό συνιστά το πιο διαδεδομένο είδος προβλέψεων, καθώς προσφέρει το πλεονέκτημα της εύκολης χρήσης ακόμα και από άτομα που δεν έχουν πλήρη γνώση σχετικά με την επιστήμη των προβλέψεων. Κατά τη στατιστική πρόβλεψη γίνεται χρήση ενός στατιστικού μοντέλου επί μιας χρονοσειράς με σκοπό την αυτοματοποιημένη παραγωγή προβλέψεων. Ο εκάστοτε χρήσης του μοντέλου μπορεί να το χρησιμοποιεί ως “μαύρο κουτί”, προσφέροντας του τις πληροφορίες της χρονοσειράς και λαμβάνοντας στην έξοδο του την πρόβλεψη για τον επιλεχθέν ορίζοντα πρόβλεψης. Ένα βασικό μειονέκτημα του συγκεκριμένου είδους είναι πως αγνοεί τους πιθανούς εξωγενείς παράγοντες που θα μπορούσαν να επηρεάσουν την τιμή της χρονοσειράς και απλώς θεωρεί πως το πρότυπο του παρελθόντος της χρονοσειράς θα εφαρμοστεί στο μέλλον.

- **Κριτική Πρόβλεψη (Judgmental Forecasting):** Το συγκεκριμένο είδος δεν συνιστά έναν αυτοματοποιημένο τρόπο παραγωγής προβλέψεων, αλλά απαιτεί την συνεισφορά ανθρώπινου δυναμικού με γνώση και εμπειρία στον τομέα των προβλέψεων. Οι κριτικές προβλέψεις μπορούν να γίνονται ατομικά από έναν άνθρωπο, όμως για την αποφυγή πιθανής προκατάληψης είναι καλύτερο να γίνονται από ομάδες ατόμων. Το μεγαλύτερο πλεονέκτημα της συγκεκριμένης μεθόδου είναι πως λαμβάνει υπόψιν όλη την πληροφορία που υπάρχει τόσο σχετικά με την ίδια την χρονοσειρά, όσο και με το περιβάλλον του προς παρατήρηση μεγέθους. Έτσι, κατά τις κριτικές προβλέψεις μπορούν να ληφθούν υπόψιν παράγοντες που επηρεάζουν σημαντικά την πορεία μιας χρονοσειράς και αγνοούνται πλήρως κατά τη στατιστική πρόβλεψη. Ωστόσο, η συγκεκριμένη μέθοδος κοστίζει κι, έτσι, συνήθως επιλέγεται μόνο για λίγες κρίσιμες χρονοσειρές.
- **Πρόβλεψη Προϋπολογισμού (Target/Budget Forecasting):** Το συγκεκριμένο είδος προβλέψεων μοιάζει περισσότερο με την τοποθέτηση ενός στόχου και την προσπάθεια επίτευξης του και λιγότερο με την αυθεντική παραγωγή προβλέψεων. Υπάρχει, δηλαδή, εξ' αρχής προκατάληψη και αισιοδοξία κατά την παραγωγή προβλέψεων προϋπολογισμού. Τέτοιες προβλέψεις συναντάμε συχνά σε εταιρείες με τη μορφή στόχου (για παράδειγμα στόχος πωλήσεων). Συνήθως, οι προβλέψεις αυτές γίνονται με τρόπο τέτοιο ώστε να είναι μεν εφικτές αλλά παράλληλα και αισιόδοξες και έχουν ως σκοπό να δώσουν κίνητρο στους εργαζόμενους να αυξήσουν την παραγωγικότητά τους.

Κατά την παραγωγή προβλέψεων μπορεί κανείς να επιλέξει ένα είδος από τα προαναφερθέντα ή να κάνει έναν συνδυασμό τους. Ο τρόπος που επιλέγεται τελικά για την παραγωγή των προβλέψεων ονομάζεται τελική πρόβλεψη.

### 2.3.2 Ορίζοντας Πρόβλεψης

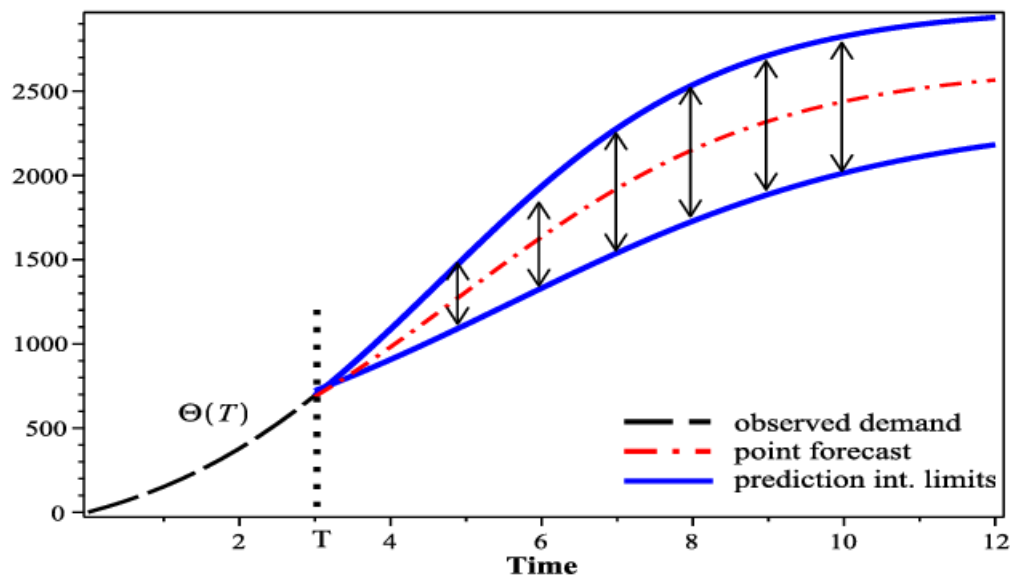
Ένα από τα σημαντικότερα στοιχεία που χρειάζεται να γνωρίζουμε κατά την παραγωγή προβλέψεων είναι ο ορίζοντας πρόβλεψης (forecast horizon). Ο ίδιος συνιστά το πλήθος των μελλοντικών χρονικών στιγμών για τις οποίες είναι επιθυμητή η παραγωγή προβλέψεων για την εκάστοτε χρονοσειρά. Ο ορίζοντας πρόβλεψης μπορεί σε μεγάλο βαθμό να καθορίσει τη μέθοδο που θα επιλέξουμε για την παραγωγή προβλέψεων επί μιας χρονοσειράς, καθώς διαφορετικές μέθοδοι κρίνονται κατάλληλες ανά περιπτώσεις. Συγκεκριμένα, ανάλογα με τον ορίζοντα πρόβλεψης διακρίνουμε τρεις κατηγορίες, για κάθε μια εκ των οποίων ενδείκνυνται διαφορετικές μέθοδοι πρόβλεψης:

- **Βραχυπρόθεσμη Πρόβλεψη (Short-Term Forecast):** Αναφέρεται σε παραγωγή προβλέψεων για χρονικό διάστημα έως και τριών μηνών.
- **Μεσοπρόθεσμη Πρόβλεψη (Mid-Term Forecasting):** Είναι η συνηθέστερη εκ των τριών κατηγοριών και αναφέρεται σε έναν ορίζοντα πρόβλεψης μεγέθους από τρεις έως και δεκαπέντε μήνες. Ωστόσο, η συγκεκριμένη μέθοδος κοστίζει κι, έτσι, συνήθως επιλέγεται μόνο για λίγες κρίσιμες χρονοσειρές.

- **Μακροπρόθεσμη Πρόβλεψη (Long-Term Forecasting):** Ο ορίζοντας πρόβλεψης είναι συνήθως μεγαλύτερος από τρία έτη. Πρόκειται για τη δυσκολότερη κατηγορία εκ των τριών, ειδικά στην περίπτωση των κριτικών προβλέψεων, αφού το βαθύτερο μέλλον χαρακτηρίζεται από μεγαλύτερη αβεβαιότητα και το πρότυπο της χρονοσειράς τείνει να αλλοιώνεται σε μεγάλο βάθος χρόνου.

### 2.3.3 Διαστήματα Εμπιστοσύνης

Ένα εξίσου σημαντικό χαρακτηριστικό με τον ορίζοντα πρόβλεψης και, μάλιστα, αρκετά αλληλένδετο με αυτόν, είναι το διάστημα εμπιστοσύνης (prediction interval). Το ίδιο καθορίζει ένα διάστημα που περικλείει συμμετρικά τη σημειακή πρόβλεψη και καθορίζει ένα σύνολο τιμών που τελικά παρέχονται ως πιθανές με βάση την πρόβλεψη μας. Το μέγεθος του καθορίζεται από το επίπεδο εμπιστοσύνης (confidence level), το οποίο σε περίπτωση ανεξάρτητων πειραμάτων συνιστά το πλήθος των φορών που η πραγματική τιμή για το μέγεθος που παρατηρείται θα ανήκει πράγματι στο διάστημα εμπιστοσύνης. Όσο μεγαλύτερο είναι το επίπεδο εμπιστοσύνης, δηλαδή όσο εγχυρότερο επιθυμούμε να είναι το διάστημα εμπιστοσύνης, τόσο μεγαλώνει και το τελευταίο. Τελικά, σε βάθος χρόνου και για περίπτωση μεγάλου ορίζοντα πρόβλεψης, το διάστημα εμπιστοσύνης μοιάζει σαν μια βεντάλια που ανοίγει συμμετρικά σε σχέση με τη σημειακή πρόβλεψη.



Σχήμα 2.3: Απεικόνιση του διαστήματος εμπιστοσύνης -με μπλε-, σε σχέση με τη σημειακή πρόβλεψη που παρέχεται -με κόκκινο-. Το ίδιο ανοίγει με το χρόνο συμμετρικά πάνω και κάτω από την εξαχθείσα σημειακή πρόβλεψη. [10]

### 2.3.4 Συμβατικές Μέθοδοι Στατιστικής Πρόβλεψης

Παραδοσιακά η παραγωγή προβλέψεων γινόταν με μεθόδους πρόβλεψης οι οποίες αποτελούνταν από κάποιο μαθηματικό τύπο που εφαρμοζόταν επί των προηγούμενων παρατηρήσεων μιας χρονοσειράς με σκοπό να παραχθούν οι προβλέψεις για τις επερχόμενες χρονικές στιγμές.

Θα αναφέρουμε μια πληθώρα τέτοιων μεθόδων σε αυτή την ενότητα, μαζί με κάποιες βασικές πληροφορίες για κάθε μια εξ' αυτών.

Σύμφωνα με τη βιβλιογραφία [6], [11], [7] οι βασικότερες συμβατικές μέθοδοι πρόβλεψης είναι οι ακόλουθες:

- **Απλοϊκή μέθοδος (Naive):** Η συγκεκριμένη μέθοδος χρησιμοποιείται συνήθως μόνο ως σημείο αναφοράς (benchmark), καθώς είναι εξαιρετικά απλή στη λογική της. Η πρόβλεψη για κάθε χρονική στιγμή ταυτίζεται με την τιμή της χρονοσειράς για την προηγούμενη χρονική στιγμή.
- **Μέθοδοι Εκθετικής Εξομάλυνσης (Exponential Smoothing):** Οι μέθοδοι αυτές είναι αρκετά διαδεδομένες και ακριβείς, ενώ εφαρμόζονται κυρίως για βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις, ενώ αριθμούν περισσότερα από 50 χρόνια από όταν ξεκίνησαν να εδραιώνονται [12]. Ανάλογα με τη μορφή της τάσης διακρίνουμε την απλή εκθετική εξομάλυνση (simple exponential smoothing) για χρονοσειρές χωρίς τάση, την εκθετική εξομάλυνση γραμμικής τάσης (holt exponential smoothing) για χρονοσειρές με γραμμική τάση και την εκθετική εξομάλυνση μη γραμμικής ή φθίνουσας τάσης (damped exponential smoothing).
- **Μοντέλα Παλινδρόμησης (Regression Models):** Τα μοντέλα αυτά σε μεγάλο βαθμό κάνουν μια απεικόνιση της τάσης και, ως εκ τούτου, κρίνονται κυρίως κατάλληλα για μακροπρόθεσμες προβλέψεις. Τα ίδια διαχωρίζονται σε απλά ή πολλαπλά ανάλογα με το πλήθος των ανεξάρτητων μεταβλητών που χρησιμοποιούνται για την παλινδρόμηση.
- **Μέθοδος Theta (Theta Model):** Η μέθοδος αυτή προτάθηκε [13] από τους Ασημακόπουλο και Νικολόπουλο το 2000. Η ιδέα της μεθόδου είναι η αποσύνθεση της χρονοσειράς σε δύο ή περισσότερες γραμμές Θητα (theta lines), η ξεχωριστή επέκταση κάθε μιας εξ' αυτών και ο συνδυασμός των προβλέψεων τους για την παραγωγή της τελικής πρόβλεψης. Οι γραμμές Theta παράγονται με τρόπο τέτοιο, ώστε να τροποποιούν τις τοπικές καμπυλότητες της χρονοσειράς.
- **Ολοκληρωμένα Αυτοπαλινδρομικά Μοντέλα Κινητών Μέσων Όρων (ARIMA Models):** Πρόκειται για στοχαστικά μοντέλα μαθηματικής φύσης τα οποία χρησιμοποιούνται για την περιγραφή της διαχρονικής εξέλιξης του προς παρατήρηση μεγέθους. Στα μοντέλα αυτά εμπεριέχεται ο τυχαίος παράγοντας, οι προηγούμενες τιμές της χρονοσειράς και, ενδεχομένως, κάποιοι ακόμα στοχαστικοί παράγοντες. Το τελικό μοντέλο προκύπτει ως γραμμικός συνδυασμός των προαναφερθέντων ποσοτήτων.

Σύμφωνα με τη βιβλιογραφία [14], [5], [7] ο συνδυασμός δύο ή περισσότερων στατιστικών μεθόδων πρόβλεψης έχει αποδειχθεί αρκετά ανταγωνιστικός. Κατά το συνδυασμό μπορούν να χρησιμοποιηθούν ίσα ή άνισα βάρη για τις προβλέψεις που προκύπτουν από κάθε μέθοδο. Καθώς οι διαφορετικές μέθοδοι μπορούν να δώσουν διαφορετικά πλεονεκτήματα, αν γίνει προσεκτική επιλογή των προς συνδυασμό μεθόδων είναι δυνατή η επίτευξη σημαντικά καλύτερων αποτελεσμάτων.

## 2.4 Στατιστικοί Δείκτες Ακρίβειας Προβλέψεων

Στην επιστήμη των προβλέψεων κρίνεται εξέχουσας σημασίας η στατιστική ανάλυση της ακρίβειας των προβλέψεων, προκειμένου να είναι δυνατή η βελτίωση τους. Ο υπολογισμός του σφάλματος μπορεί να μας βοηθήσει να κατανοήσουμε βασικά χαρακτηριστικά γι' αυτό, όπως την πηγή του, τη συστηματικότητα ή μή και την προκατάληψη ή μη. Παράλληλα, κάθε δείκτης εκφράζει διαφορετικά χαρακτηριστικά του σφάλματος και επιλέγεται σε διαφορετικές περιπτώσεις.

Προκειμένου να μπορέσουμε να υπολογίσουμε τον εκάστοτε στατιστικό δείκτη ακρίβειας χρειαζόμαστε τόσο την χρονοσειρά που αποτελείται από τις πραγματικές παρατηρήσεις, όσο και την χρονοσειρά που αποτελεί την πρόβλεψη για το προς παρατήρηση μέγεθος. Η πραγματική χρονοσειρά συμβολίζεται με  $Y$ , ενώ η χρονοσειρά που συνιστά την πρόβλεψη με  $F$ .

Θα ξεκινήσουμε ορίζοντας το σφάλμα, το οποίο συνιστά τη διαφορά μεταξύ της πραγματικής τιμής και της τιμής πρόβλεψης για μια χρονική στιγμή. Θετικό σφάλμα συνεπάγεται απαισιοδοξία, αφού η τιμή της πρόβλεψης ήταν μικρότερη από την πραγματική, ενώ αρνητικό σφάλμα συνεπάγεται αισιοδοξία.

$$e_t = Y_t - F_t$$

$Y_t$ : παρατήρηση την χρονική στιγμή  $t$

$F_t$ : πρόβλεψη τη χρονική στιγμή  $t$

$e_t$ : σφάλμα πρόβλεψης για τη χρονική στιγμή  $t$

Στη συνέχεια παραθέτουμε τους βασικότερους στατιστικούς δείκτες ακρίβειας προβλέψεων, όπως αυτοί παρατίθενται στη βιβλιογραφία [7], με αναφορά στη μαθηματική τους έκφραση και στα χαρακτηριστικά τους. Τα σφάλματα υπολογίζονται για όλες τις χρονικές στιγμές για τις οποίες έχουμε διαθέσιμη και την πραγματική παρατήρηση αλλά και την πρόβλεψη. Το πλήθος των χρονικών αυτών στιγμών είναι  $n$ .

- **Μέσο Σφάλμα (Mean Error):** Βασικό χαρακτηριστικό του συγκεκριμένου δείκτη είναι πως εκφράζει την ύπαρξη ή μη συστηματικότητας στο σφάλμα. Τιμές που βρίσκονται κοντά στο μηδέν δηλώνουν τυχαιότητα του σφάλματος. Οι θετικές τιμές του δείκτη δηλώνουν απαισιοδοξία, ενώ οι αρνητικές τιμές εκφράζουν αισιοδοξία.

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_t - F_t)$$

- **Μέσο απόλυτο σφάλμα (Mean Absolute Error):** Ο συγκεκριμένος δείκτης εκφράζει ένα μέτρο σχετικά με την αστοχία, όμως δεν δίνει κάποια πληροφορία για την

κατεύθυνση του σφάλματος. Μεγαλύτερες τιμές του δείκτη συνεπάγονται μικρότερη ακρίβεια.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_t - F_t|$$

- **Μέσο τετραγωνικό σφάλμα (Mean Squared Error):** Όπως και ο προηγούμενος δείκτης, ο συγκεκριμένος μας δίνει πληροφορίες σχετικά με το μέτρο της αστοχίας, όχι όμως σχετικά με την κατεύθυνση του. Βασικό του χαρακτηριστικό είναι πως, λόγω του τετραγωνισμού, δίνει μεγαλύτερη βαρύτητα στα μεγάλα σφάλματα και μικρότερη στα μικρά.

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

- **Ρίζα μέσου τετραγωνικού σφάλματος (Root Mean Squared Error):** Ο συγκεκριμένος δείκτης σφάλματος έχει ακριβώς τις ίδιες ιδιότητες με τον προηγούμενο. Η μόνη του διαφορά είναι πως λόγω της τετραγωνικής ρίζας, ο ίδιος είναι εκφρασμένος στις μονάδες της αρχικής χρονοσειράς.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_t - F_t)^2}$$

- **Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error):** Ο συγκεκριμένος δείκτης ενδείκνυται για περιπτώσεις όπου η ίδια μέθοδος πρόβλεψης εφαρμόζεται σε περισσότερες από μια χρονοσειρές (οι οποίες έχουν διαφορετικά επίπεδα), προκειμένου να είναι δυνατή η σύγκριση των αποτελεσμάτων. Όπως φαίνεται και από το όνομα του, ο ίδιος είναι εκφρασμένος ως ποσοστό. Οι μικρότερες τιμές του εκφράζουν μεγαλύτερη ακρίβεια στην πρόβλεψη. Είναι σημαντικό να αναφέρουμε πως ο συγκεκριμένος δείκτης δεν ενδείκνυται για χρονοσειρές διακοπτόμενης ζήτησης, καθώς λόγω της τοποθέτησης της πραγματικής τιμής της χρονοσειράς στον παρονομαστή υπάρχει ο κίνδυνος απειρισμού του.

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

- **Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error):** Ο συγκεκριμένος δείκτης είναι μια παραλλαγή του προηγούμενου. Η βασική του διαφορά είναι η αλλαγή του όρου του παρονομαστή που οδηγεί στην ύπαρξη πάνω ορίου, με τις τιμές του δείκτη να κυμαίνονται στο διάστημα [0%, 200%]. Παρά την ονομασία, ο ίδιος δεν είναι ιδιαίτερα συμμετρικός, καθώς διαχειρίζεται τις απασιόδοξες και αισιόδοξες προβλέψεις με διαφορετικό τρόπο. Στον ίδιο, όπως και στον προηγούμενο, μικρότερη τιμή συνεπάγεται μεγαλύτερη ακρίβεια. Στα πλαίσια της παρούσας διπλωματικής όλα τα αποτελέσματα μας εκφράζονται με χρήση αυτού του



δείκτη, ο οποίος είναι από τους πλέον διαδεδομένους στην επιστήμη των προβλέψεων.

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

- **Μέσο απόλυτο κανονικοποιημένο σφάλμα (Mean Absolute Scaled Error):** Ο συγκεκριμένος δείκτης προτάθηκε [15] από τους Hyndman και Koehler το 2006 με σκοπό να αποφευχθεί η απροσδιοριστία των δεικτών MAPE και sMAPE, η οποία οφείλεται στον πιθανό μηδενισμό του παρονομαστή αυτών των σφαλμάτων. Παράλληλα, βασικός τους στόχος ήταν να αποδοθεί η ίδια βαρύτητα τόσο στα μικρά όσο και στα μεγάλα σφάλματα, σε αντίθεση με τους δείκτες MSE και RMSE. Η κανονικοποίηση γίνεται με βάση το σφάλμα που θα προέκυπτε με εφαρμογή της μεθόδου πρόβλεψης Naïve επί της χρονοσειράς. Αν ο δείκτης λάβει τιμή μικρότερη της μονάδας, τότε η μέθοδος που εφαρμόστηκε απέδωσε καλύτερα συγκριτικά με τη Naïve, ενώ αν λάβει τιμή μεγαλύτερη της μονάδας, τότε συνέβη το αντίθετο.

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |Y_t - F_t|}{\frac{1}{n-1} \sum_{i=2}^n |Y_t - Y_{t-1}|}$$

Είναι αρκετά σημαντικό να γίνει ένας διαχωρισμός ανάμεσα στο σφάλμα που υπολογίζεται από την εφαρμογή της μεθόδου σε χρονικές στιγμές που έχουν παρέλθει, κι άρα, κατά την παραγωγή των προβλέψεων έχουμε πληροφορία για την πραγματική τιμή του προς παρατήρηση μεγέθους και στο σφάλμα που υπολογίζεται για τις προβλέψεις που κάνουμε αφού παρέλθει το χρονικό διάστημα του ορίζοντα και πρόβλεψης και γίνουν σε εμάς διαθέσιμες οι πραγματικές τιμές του προς παρατήρηση μεγέθους. Το πρώτο σφάλμα ονομάζεται in-sample error, και στα πλαίσια της παρούσας διπλωματικής, όπου εργαζόμαστε με νευρωνικά δίκτυα, το ίδιο μπορεί να ταυτιστεί με το validation error. Το δεύτερο, ονομάζεται out-of-sample error και για την παρούσα εργασία μπορεί να θεωρηθεί ισοδύναμο με το error που παίρνουμε στο τεστ σετ. Ένα χαμηλό in-sample error δεν συνεπάγεται πάντα και χαμηλό out-of-sample error, ωστόσο κατά την επιλογή του μοντέλου πρόβλεψης έχουμε διαθέσιμο μόνο αυτό.



## Κεφάλαιο 3

# Μηχανική Μάθηση και Νευρωνικά Δίκτυα

### 3.1 Εισαγωγή

Η μηχανική μάθηση (machine learning) αποτελεί κλάδο της τεχνητής νοημοσύνης και αντλεί το θεωρητικό της υπόβαθρο από την υπολογιστική στατιστική και την επιστήμη των υπολογιστών. Έχει πολλαπλές εφαρμογές σε ένα ευρύτατο φάσμα ερευνητικών και εμπορικών προβλημάτων και η χρήση της συστήνεται κυρίως σε δύο περιπτώσεις: Πρώτον, όταν η μορφή του προβλήματος καθιστά αδύνατη την οργάνωση της λύσης σε μία σειρά από κανόνες και δεύτερον, όταν ο ακριβής υπολογισμός της βέλτιστης λύσης είναι τόσο πολύπλοκος υπολογιστικά που θα συνέφερε μια προσέγγιση της λύσης, κάνοντας χρήση αισθητά λιγότερων υπολογιστικών πόρων. Έτσι, ένα σύστημα μηχανικής μάθησης μαθαίνει, μέσω ενός συνόλου δεδομένων, να επιτελεί μία εργασία η οποία έπειτα γενικεύεται σε άλλα δεδομένα με τα οποία το μοντέλο δεν έχει έρθει σε επαφή [16].

Σε επιχειρησιακό επίπεδο η μηχανική μάθηση βρίσκει εφαρμογή σε μία πληθώρα περιστάσεων όπως είναι τα συστήματα συστάσεων, η εξατομικευμένη προώθηση και το marketing, ο εντοπισμός απάτης στις πιστωτικές κάρτες καθώς και σε προβλήματα βελτιστοποίησης και υποστήριξης αποφάσεων στον κλάδο της επιχειρησιακής έρευνας. Τέτοια παραδείγματα αποτελούν η διαχείριση των αποθηκών και της εφοδιαστικής αλυσίδας [17]. Ακόμη, στον κλάδο της υγείας έχουν αναπτυχθεί μία πληθώρα από εφαρμογές όπως η πρόγνωση και η διάγνωση ασθενειών, η εξατομικευση της φαρμακευτικής αγωγής και η ανακάλυψη νέων φαρμάκων [18]. Στον κλάδο της επεξεργασίας της φυσικής γλώσσας εξέχοντα παραδείγματα εφαρμογών αποτελούν η αυτόματη μετάφραση, η εξαγωγή πληροφοριών από κείμενο, η αυτόματη δημιουργία περιλήψεων, η κατηγοριοποίηση κειμένων και η αναγνώριση φωνής και συναισθήματος [19]. Επιπρόσθετα, ποικίλες είναι οι εφαρμογές στον κλάδο της όρασης υπολογιστών όπως είναι η αναγνώριση δραστηριότητας, η εκτίμηση της στάσης του σώματος, η αναγνώριση προσώπων και αντικειμένων [20,21]. Μάλιστα, αξιολογα είναι και τα επιτεύγματα της μηχανικής μάθησης στην τέχνη με παραδείγματα όπως την δημιουργία μουσικής σύνθεσης βασισμένης σε κάποια

συναισθηματική κατάσταση ή την μεταφορά του ζωγραφικού στυλ σε εικόνες [22].

## 3.2 Ιστορική Αναδρομή

Ο Alan Turing παρατηρεί το 1950 την ανάγκη δημιουργίας μηχανών που “να μαθαίνουν από την εμπειρία” ανοίγοντας τον δρόμο για αυτό που σήμερα θεωρούμε ως μηχανική μάθηση [23]. Οι πρώτες εξελίξεις στον τομέα της μηχανικής μάθησης χρονολογούνται στην διάρκεια της δεκαετίας του 1960 ενώ ο όρος μηχανική μάθηση συναντάται πρώτη φορά στην εργασία [24] του Arthur Samuel, ερευνητή της IBM. Ωστόσο, ένας πιο αυστηρός ορισμός για την μηχανική μάθηση θα έρθει αρκετά χρόνια αργότερα, το 1997, από τον Tom Mitchell [16]:

**Λήμμα 3.1.** *Λέγεται ότι ένα πρόγραμμα μαθαίνει από μία εμπειρία  $E$  να επιτελεί μια εργασία  $T$ , βάση μιας μετρικής επίδοσης  $P$ , εάν η μετρική επίδοσης  $P$  για την εργασία  $T$  βελτιώνεται με την εμπειρία  $E$ .*

Τα νευρωνικά δίκτυα, υποπεδίο της μηχανικής μάθησης με πολλά state-of-the-art αποτελέσματα, έχουν την θεωρητική τους βάση στην εφεύρεση του perceptron από τον Rosenblatt [25]. Ωστόσο, η περιορισμένη ικανότητα του perceptron για την αναγνώριση μη γραμμικών προτύπων θα κάμψει το ενδιαφέρον για τα νευρωνικά δίκτυα για αρκετά χρόνια μέχρι την εμφάνιση των πολυστρωματικών δικτύων perceptron, με σαφώς μεγαλύτερη υπολογιστική ισχύ.

Μερικές εφευρέσεις-σταθμοί στην πορεία εξέλιξης της μηχανικής μάθησης κατά το δεύτερο μισό του 20ού αιώνα αποτέλεσαν το αντίληπτρο (perceptron), ιδέα η οποία συλλήφθηκε το 1943 [26] και εισήχθη το 1957 [25], ο ταξινομητής κοντινότερων γειτόνων (nearest neighbor classifier) το 1967 [27], η πρώτη σύλληψη της ιδέας του αλγορίθμου της οπισθοδιάδοσης (backpropagation) το 1970 [28,29], τα δίκτυα Hopfield, ένα είδος επαναληπτικών νευρωνικών δικτύων (recurrent neural networks) το 1982 [30], η εφεύρεση του αλγορίθμου Q-learning το 1989 [31] που ανοίγει τον δρόμο για την ανάπτυξη του κλάδου της ενισχυτικής μάθησης, η ανακάλυψη των τυχαίων δασών απόφασης (random decision forests) το 1995 [32] όπως και των μηχανών διανυσμάτων υποστήριξης (support vector machines) την ίδια χρονιά [33], τα δίκτυα μακράς βραχείας μνήμης LSTM (long-short-term memory) το 1997 [34] και τα συνελκτικά νευρωνικά δίκτυα (convolutional neural networks) κατά την διάρκεια των δεκαετιών ογδόντα και ενενήντα [35].

## 3.3 Μηχανική Μάθηση

### 3.3.1 Κατηγορίες Αλγορίθμων Μηχανικής Μάθησης

Με βάση την βιβλιογραφία [36] προσδιορίζονται τρεις βασικές κατηγορίες αλγορίθμων μηχανικής μάθησης, κάθε μία από τις οποίες αφορά σε διαφορετικά είδη προβλημάτων. Οι κατηγορίες αυτές παρουσιάζονται εδώ συνοπτικά:

- **Επιβλεπόμενη Μάθηση (Supervised Learning):** Η επιβλεπόμενη μάθηση συνίσταται στην εκμάθηση ενός μοντέλου να επιτελεί μία συγκεκριμένη εργασία μέσα από δεδομένα των οποίων η επιθυμητή έξοδος είναι γνωστή (labeled data). Τέτοιες μέθοδοι προτιμώνται όταν τα δεδομένα είναι ευρέως διαθέσιμα ή όταν η απόκτησή τους είναι σχετικά οικονομική.

Βασικότερες κατηγορίες προβλημάτων επιβλεπόμενης μάθησης είναι η κατηγοριοποίηση (classification) και η παλινδρόμηση (regression). Στην πρώτη περίπτωση προσπαθούμε να κατατάξουμε μία είσοδο σε μία επιθυμητή έξοδο με διακριτές καταστάσεις. Τέτοιο παράδειγμα αποτελεί η κατηγοριοποίηση κυττάρων σε μία ιατρική εικόνα ως καρκινικά ή μη. Στην δεύτερη προσπαθούμε να συσχετίσουμε τις μεταβλητές εξόδου με μία ή περισσότερες μεταβλητές εισόδου. Για παράδειγμα τέτοιο πρόβλημα αποτελεί η πρόβλεψη χρονοσειρών. Στο πρόβλημα αυτό προσπαθούμε να συσχετίσουμε τον αριθμό των  $m$  επόμενων παρατηρήσεων χρησιμοποιώντας ως είσοδο τις  $n$  προηγούμενες παρατηρήσεις. Στην γενική περίπτωση, η παλινδρόμηση είναι μη-γραμμική.

- **Μη επιβλεπόμενη Μάθηση (Unsupervised Learning):** Η μη επιβλεπόμενη μάθηση αναγνωρίζει ένα υποκείμενο μοτίβο στα δεδομένα χωρίς να είναι εκ των προτέρων επισημασμένη η έξοδός τους. Αυτές οι τεχνικές είναι ιδιαίτερα χρήσιμες καθώς η απόκτηση επισημασμένων δεδομένων είναι σχετικά ακριβή και απαιτεί την χρονοβόρα εργασία ανθρώπων, οι οποίοι μάλιστα σε αρκετές περιπτώσεις πρέπει να είναι εξαιρετικά καταρτισμένοι. Διασημότερες χρήσεις της προσέγγισης αυτής είναι η συσταδοποίηση (clustering) και η μείωση της διαστατικότητας (dimensionality reduction).
- **Ενισχυτική Μάθηση (Reinforcement Learning):** Η ενισχυτική μάθηση αποτελεί μία μέθοδο επίλυσης προβλημάτων κατά την οποία το μοντέλο μαθαίνει να εκτελεί την ζητούμενη εργασία μέσω της εξερεύνησης του περιβάλλοντός του, χωρίς να του δοθούν ξεκάθαρες οδηγίες. Σε περίπτωση θετικής έκβασης, η συμπεριφορά ενισχύεται μέσω μιας ανταμοιβής και η εκπαίδευση συνεχίζεται. Έτσι, το μοντέλο μπορεί να επιλύει όλο και πιο περίπλοκα προβλήματα, όπως για παράδειγμα η μετακίνηση ενός αντικειμένου από μια θέση A σε μία θέση B από ένα αυτόνομο ρομποτικό χέρι.

### 3.3.2 Μοντέλα Μηχανικής Μάθησης

Η επίλυση ενός προβλήματος με τεχνικές μηχανικής μάθησης περιλαμβάνει την εκπαίδευση ενός μοντέλου. Εδώ αναφερόμαστε στις βασικότερες κατηγορίες μοντέλων μηχανικής μάθησης [37]:

- **Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks):** Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) αποτελούν συλλογές από συνδεδεμένους κόμβους που μιμούνται ως ένα βαθμό την λειτουργία των βιολογικών νευρώνων. Υπάρχουν πολλά διαφορετικά είδη κόμβων κάθε ένα από τα οποία επιτελεί μια διαφορετική λειτουργία

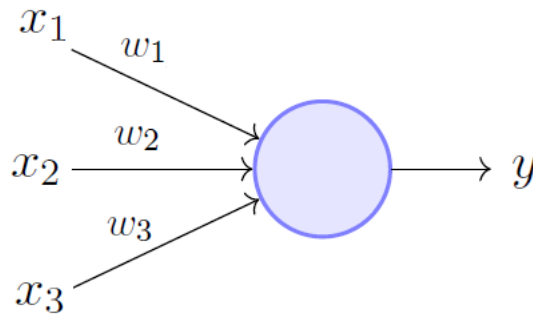
και συνθέτει μια πληθώρα από ΤΝΔ. Τέτοια παραδείγματα κόμβων ΤΝΔ αποτελούν ο τεχνητός νευρώνας perceptron ή οι κόμβοι LSTM. Στα νευρωνικά δίκτυα γίνεται εκτενέστερη αναφορά στην επόμενη υποενότητα.

- **Δέντρα Απόφασης (Decision Trees):** Τα δέντρα απόφασης αποτελούν δενδρικές δομές στις οποίες κάθε εσωτερικός κόμβος αναπαριστά ένα από τα χαρακτηριστικά του δείγματος ενώ κάθε φύλλο αναπαριστά την προβλεπόμενη έξοδο. Για παράδειγμα, ένα δέντρο απόφασης για την τιμολόγηση ενός ακινήτου προς πώληση θα μπορούσε να έχει ως χαρακτηριστικά στους εσωτερικούς κόμβους την δεκαετία κατασκευής, ή τα τετραγωνικά μέτρα της κατοικίας ενώ στα φύλλα θα είχε τις διαφορετικές προκύπτουσες τιμές τιμολόγησης.
- **Μηχανές Διανυσμάτων Υποστήριξης (Support-Vector Machines):** Οι μηχανές διανυσμάτων υποστήριξης είναι μία μη-πιθανοτική, δυαδική, κατά βάση γραμμική μέθοδος ταξινόμησης και απαιτεί επισημασμένα δεδομένα για τα λειτουργήσει. Επεκτάσεις αυτής της μεθόδου, μπορούν να την καταστήσουν κατάλληλη και για μη γραμμικά δεδομένα, όπως είναι για παράδειγμα το τέχνασμα του πυρήνα (kernel trick).
- **Δίκτυα Bayes (Bayes Networks):** Τα δίκτυα Bayes ή δίκτυα πίστης (belief networks) όπως αλλιώς ονομάζονται, αποτελούν ένα πιθανοτικό μοντέλο αναπαράστασης της συσχέτισης τυχαίων μεταβλητών μέσω κατευθυνόμενων ακυκλικών γράφων (directed acyclic graphs). Με ένα τέτοιο μοντέλο μπορούμε να αναπαραστήσουμε για παράδειγμα την εξάρτηση μεταξύ των συμπτωμάτων και των ασθενειών.
- **Γενετικοί Αλγόριθμοι (Generative Algorithms):** Οι γενετικοί αλγόριθμοι είναι μία κατηγορία αλγορίθμων εμπνευσμένοι από την Δαρβινική φυσική επιλογή. Ο αλγόριθμος αρχικοποιεί τυχαία μία σειρά από μοντέλα, που αποτελούν την πρώτη γενιά. Οι καλύτερες λύσεις επιλέγονται ενώ οι υπόλοιπες απορρίπτονται. Έτσι συγκροτείται ένα πληθυσμός λύσεων που δημιουργεί την επόμενη γενιά. Κάθε γενιά λύσεων παράγει την επόμενη με μηχανισμούς μίξης και μετάλλαξης. Τέλος, προκύπτει το εκπαιδευμένο μοντέλο ως η βέλτιστη λύση της τελευταίας γενεάς.

## 3.4 Νευρωνικά Δίκτυα

### 3.4.1 Τεχνητός Νευρώνας Perceptron

Ο τεχνητός νευρώνας (artificial neuron) γνωστός με την ονομασία αντίληπτρο (perceptron) αποτελεί την βασική δομική μονάδα της μεγάλης οικογένειας των προσθιοδρομικών νευρωνικών δικτύων (feed-forward neural networks). Το perceptron προτάθηκε ως ένας δυαδικός ταξινομητής [25, 38]. Στην γενικότερη περίπτωση, ανάλογα με την συνάρτηση ενεργοποίησής του, μπορεί να χρησιμοποιηθεί για δυαδική ταξινόμηση, γραμμική ή και μη γραμμική παλινδρόμηση.



Σχήμα 3.1: Σχηματική απεικόνιση νευρώνα perceptron [39]

Στην γενική περίπτωση ένας perceptron έχει  $n$  εισόδους και μία έξοδο. Κάθε μία από αυτές τις εισόδους πολλαπλασιάζεται με την συνιστώσα του βάρους στο οποίο αντιστοιχεί και το άθροισμά τους, το οποίο ουσιαστικά αποτελεί το εσωτερικό γινόμενο του διανύσματος εισόδου με το διάνυσμα των βαρών, μετασχηματίζεται μέσω της βηματικής συνάρτησης ενεργοποίησης αφού προστεθεί πρώτα σε αυτό μία σταθερά γνωστή ως προκατάληψη (bias). Έτσι προκύπτει η έξοδος του perceptron:

$$y = g\left(\sum_{i=0}^n x_i w_i + b\right)$$

- $y$ : έξοδος νευρώνα perceptron
- $g$ : συνάρτηση ενεργοποίησης
- $x_i$ : συνιστώσα διανύσματος εισόδου
- $w_i$ : συνιστώσα διανύσματος βαρών
- $b$ : προκατάληψη (bias)

Ο νευρώνας perceptron, αν και αποτελεί το θεωρητικό έναυσμα της ανάδειξης του πεδίου των νευρωνικών δικτύων έχει πολύ περιορισμένες εφαρμογές. Εκ γεγονότος, μπορεί να ταξινομήσει ορθά μόνο γραμμικώς διαχωρίσιμα δεδομένα και κρίνεται ακατάλληλος ακόμα και για απλά μοτίβα όπως για παράδειγμα το διάσημο πρόβλημα ταξινόμησης XOR.

### 3.4.2 Συναρτήσεις Ενεργοποίησης

Στην γενική περίπτωση, το εσωτερικό γινόμενο των διανυσμάτων της εισόδου και των βαρών ενός perceptron (έστω input) μετασχηματίζεται μέσω μιας συνάρτησης πριν αυτό φτάσει στην έξοδο και τροφοδοτήσει τα επόμενα επίπεδα νευρώνων. Η συνάρτηση αυτή καλείται συνάρτηση ενεργοποίησης (activation function). Όπως έχει φανεί από άλλες μελέτες [40] οι συναρτήσεις αυτές διαδραματίζουν πρωτεύοντα ρόλο στην συνολική επίδοση ενός ΤΝΔ. Οι βασικότερες συναρτήσεις με βάση την βιβλιογραφία παρατίθενται στη συνέχεια ακολουθούμενες από την μαθηματική τους έκφραση:

- **Ταυτοτική Συνάρτηση (Identity Function):** Η ταυτοτική συνάρτηση απεικονίζει την είσοδο στην έξοδο χωρίς κανένα ενδιάμεσο μετασχηματισμό. Αποτελεί ουσιαστικά την απλούστερη μορφή συνάρτησης ενεργοποίησης και χρησιμοποιείται κυρίως όταν το αποτέλεσμα στην έξοδο ενός νευρώνα έχει κάποιο φυσικό νόημα και άρα δεν θέλουμε να το μεταβάλλουμε.

$$f(x) = x$$

- **Βηματική Συνάρτηση (Step Function):** Η βηματική συνάρτηση προσομοιάζει περισσότερο στην δομή και λειτουργία των φυσικών νευρώνων του εγκεφάλου οι οποίοι είτε πυροδοτούνται, είτε όχι. Για εισόδους μεγαλύτερες από μία τιμή κατωφλίου, έστω  $b$ , η έξοδος του νευρώνα είναι μονάδα και μηδέν διαφορετικά. Αν και αποτελεί μία κομψή επιλογή, χρησιμοποιείται σε λίγες κατηγορίες εφαρμογών, κυρίως λόγω της ασυνέχειας της, πράγμα που δυσχεραίνει την παραγωγή και την διάδοση του σφάλματος.

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

- **Ανορθωμένη Γραμμική Συνάρτηση (Rectified Linear Unit):** Η ανορθωμένη γραμμική συνάρτηση ή ReLU όπως είναι συνήθως γνωστή, αποτελεί μία δημοφιλή επιλογή συνάρτησης ενεργοποίησης. Συνήθως χρησιμοποιείται όταν θέλουμε να έχουμε αποτελέσματα παραπλήσια με αυτά της βηματικής συνάρτησης, αλλά παράλληλα επιθυμούμε να μεταδίδεται το μέγεθος των χαρακτηριστικών μεταξύ των στρώσεων του νευρωνικού δικτύου [40].

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

- **Κανονικοποιημένη Εκθετική Συνάρτηση (Normalized Exponential Function):** Η κανονικοποιημένη εκθετική συνάρτηση γνωστή ως softmax, αποτελεί την συνήθη επιλογή όταν στην έξοδο ενός νευρωνικού δικτύου θέλουμε να εκφράσουμε πιθανότητες. Δηλαδή, όταν θέλουμε όλες οι έξοδοι να αθροίζονται στην μονάδα, πράγμα ιδιαίτερα χρήσιμο σε προβλήματα ταξινόμησης.

$$f(x) = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}}$$

### 3.4.3 Πολυστρωματικά Δίκτυα Perceptron

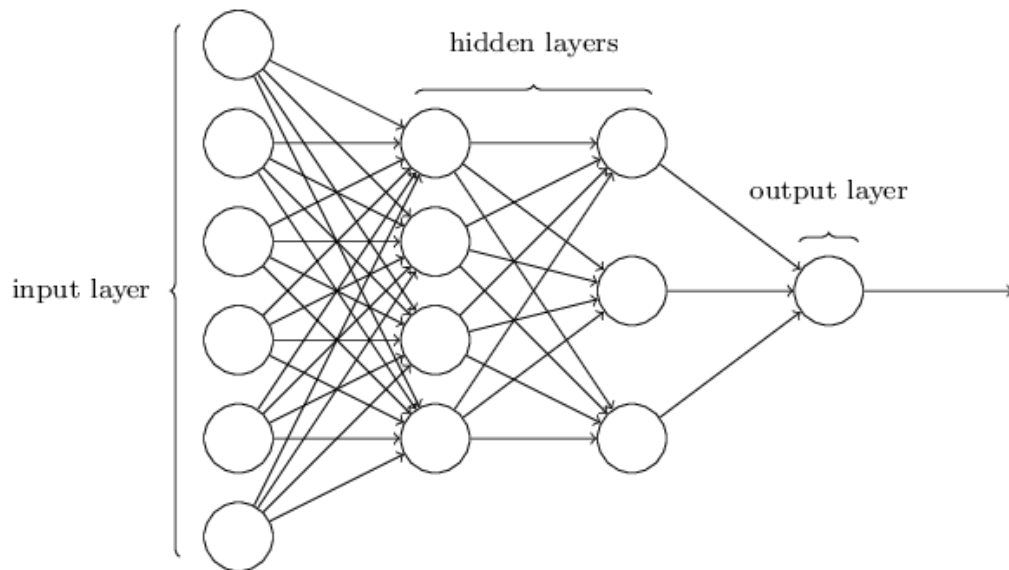
Όπως αναφέραμε σε προηγούμενη υποενότητα, η υπολογιστική ισχύ των νευρώνων perceptron είναι σαφώς περιορισμένη. Η ανάγκη για την δημιουργία τεχνητών νευρωνικών δικτύων που να μπορούν να μάθουν οποιαδήποτε κατανομή δεδομένων οδήγησε στην δημιουργία των πολυστρωματικών δικτύων perceptron, τα οποία αποτελούν την βασική δομική αρχιτεκτονική των νευρωνικών δικτύων.

Ένα πολυστρωματικό δίκτυο perceptron (multi-layer perceptron) ή MLP όπως είναι ευρέως γνωστό, αποτελεί ένα είδος προσθιοδρομικού νευρωνικού δικτύου που χρησιμοποιεί



σαν δομική του μονάδα τους νευρώνες perceptron [41]. Ένα τέτοιο δίκτυο αποτελείται από στρώσεις νευρώνων perceptron, κάθε μία από τις οποίες είναι πλήρως συνδεδεμένη (fully connected) με τους νευρώνες της επόμενης στρώσης. Πιο αναλυτικά, η έξοδος κάθε νευρώνα μιας στρώσης τροφοδοτεί την είσοδο καθενός από τους νευρώνες της επόμενης στρώσης. Μια στρώση που συνδέεται με τέτοιο τρόπο με εκείνη που ακολουθεί ονομάζεται πυκνή (dense).

Ένα MLP δίκτυο αποτελείται από τουλάχιστον τρεις πυκνές στρώσεις νευρώνων. Η πρώτη αποτελεί την στρώση εισόδου, η τελευταία την στρώση εξόδου ενώ στο ενδιάμεσο μπορούν να παρεμβάλλονται οσοσδήποτε πυκνές τέτοιες στρώσεις οι οποίες καλούνται κρυμμένες στρώσεις (hidden layers).



Σχήμα 3.2: Σχηματική αναπαράσταση ενός MLP δικτύου με 2 κρυφές στρώσεις [42]

Τα πολυστρωματικά perceptron (MLP) αποτελούν μία από τις αρχιτεκτονικές μοντέλων που έχουν φανεί να έχουν καλή επίδοση σε προβλήματα πρόβλεψης χρονοσειρών. Συνήθως αποτελούν μία καλή αρχική εκτίμηση της λύσης και λόγω της μικρής υπολογιστικής πολυπλοκότητας που τα χαρακτηρίζει προσφέρονται ως εναρκτήρια σημεία σύγκρισης (benchmarks) για πιο σύνθετες αρχιτεκτονικές. Μία άλλη αρχιτεκτονική ΤΝΔ κατάλληλη για την πρόβλεψη χρονοσειρών αποτελούν τα επαναληπτικά δίκτυα. Εκείνα χρησιμοποιούνται στην πρόβλεψη ακολουθιακών δεδομένων όπως για παράδειγμα είναι ο ήχος, οι λέξεις μιας πρότασης ή μία ακολουθία DNA. Ως εκ τούτου, βρίσκουν εξαιρετική εφαρμογή στην πρόβλεψη χρονοσειρών [43]. Τέλος, τα συνελκτικά δίκτυα (CNN) μπορούν να αναγνωρίζουν την χωρική συσχέτιση μεταξύ των δεδομένων, όπως για παράδειγμα συμβαίνει με τις διάφορες εφαρμογές όρασης υπολογιστών. Αντίστοιχα, όπως έχει προταθεί από την βιβλιογραφία [44], ένα συνελκτικό δίκτυο μπορεί να αναγνωρίσει τα υποκείμενα μοτίβα και τις συσχετίσεις μεταξύ μελλοντικών και παρελθοντικών χρονικών παρατηρήσεων και να παράξει ποιοτικές προβλέψεις.

### 3.4.4 Ο Αλγόριθμος Backpropagation

Ο αλγόριθμος οπισθοδιάδοσης (backpropagation) αποτελεί μια οικογένεια μεθόδων για την εξυπηρέτηση της μάθησης ενός MLP μέσω της διάδοσης του σφάλματος [45]. Στην πιο απλοϊκή εκδοχή ενός τέτοιου δικτύου, δηλαδή στην περίπτωση ενός μόνο νευρώνα perceptron, η διάδοση του σφάλματος από την έξοδο στην είσοδο και η ανανέωση των βαρών είναι άμεση: στην περίπτωση λανθασμένης ταξινόμησης, τα βάρη του perceptron αναπροσαρμόζονται έτσι ώστε να μειώνεται η απόσταση μεταξύ της προβλεπόμενης και της πραγματικής τιμής. Η ανανέωση των βαρών του perceptron είναι αυτή που συντελεί στην μάθησή του και καθορίζεται από την κάθοδο κλίσης της συνάρτησης του σφάλματος (gradient descent). Στην γενική της περίπτωση είναι η εξής:

$$w_{i+1} = w_i + a \times Err \times g'(in) \times x_i$$

$w_{i+1}$ : ανανεωμένη συνιστώσα διανύσματος βαρών

$w_i$ : συνιστώσα διανύσματος βαρών

$x_i$ : συνιστώσα διανύσματος εισόδου

$a$ : ρυθμός εκμάθησης, υπερπαράμετρος στο διάστημα (0,1)

$Err$ : σφάλμα προβλεπόμενης από πραγματική τιμή

$in$ : εσωτερικό γινόμενο διανύσματος εισόδου με το διάνυσμα βαρών

Ωστόσο στην περίπτωση ενός MLP δικτύου η διαδικασία αυτή είναι μη τετριμμένη. Η οπισθοδιάδοση επομένως αποτελεί μία αποδοτική επαναληπτική μέθοδο για την ανανέωση των βαρών ενός MLP δικτύου έτσι ώστε αυτό να μπορεί να απεικονίσει κάθε είσοδο στην σωστή έξοδο.

Ξεκινώντας από μία εσφαλμένη πρόβλεψη εισόδου, ανανεώνεται το βάρος των νευρώνων της τελευταίας στρώσης χρησιμοποιώντας τον αλγόριθμο καθόδου κλίσης ή κάποια άλλη βελτιστοποιημένη παραλλαγή του, όπως θα περιγράψουμε παρακάτω. Έπειτα με βάση τον κανόνα της αλυσίδας του Leibniz, το σφάλμα διαδίδεται χρησιμοποιώντας τις μερικές παραγώγους των βαρών της επόμενης στοιβάδας. Η διαδικασία αυτή επαναλαμβάνεται για κάθε έναν από τα perceptrons κάθε στρώσης από το τέλος προς την αρχή και με αυτό τον τρόπο επιτυγχάνεται η μάθηση.

Καθώς η διαδικασία της οπισθοδιάδοσης για κάθε ένα από τα δεδομένα μπορεί να αποβεί αρκετά κοστοβόρα υπολογιστικά, τα δεδομένα συνενώνονται συνήθως σε παρτίδες (batches). Με αυτό τον τρόπο αποτιμάται πρώτα η έξοδος όλων των δειγμάτων της παρτίδας κι έπειτα πραγματοποιείται ένα μόνο τρέξιμο της οπισθοδιάδοσης του σφάλματος για όλα τα δείγματα της παρτίδας αυτής.

### 3.4.5 Συναρτήσεις Απώλειας

Με βάση την βιβλιογραφία [46], παραθέτουμε εδώ μερικές από τις πιο συνήθεις μετρικές για την παρατήρηση του σφάλματος (error) κατά την διαδικασία της εκπαίδευσης που αφορούν προβλήματα παλινδρόμησης. Οι μετρικές αυτές αποτελούν τον μέσο όρο της συνάρτησης απώλειας (loss function) όλων των δειγμάτων.

- **Μέσο τετραγωνικό σφάλμα (Mean Squared Error):** Το μέσο τετραγωνικό σφάλμα, γνωστό και ως απώλεια L2, επιβάλλει αναλογικά μεγαλύτερη ποινή στα μεγαλύτερα σφάλματα του μοντέλου και ως εκ τούτου συνεισφέρει σε γρηγορότερη σύγκλιση του μοντέλου. Υπολογίζεται από τον τύπο που ακολουθεί:

$$L2 = |y - f(x)|^2$$

- **Μέσο απόλυτο σφάλμα (Mean Absolute Error):** Το μέσο απόλυτο σφάλμα, γνωστό και ως απώλεια L1, παρότι είναι πιο ακλόνητο από ασυνήθιστα μεγάλες τιμές (outliers) χαρακτηρίζεται από την δυσκολία διαχείρισης της απόλυτης τιμής σε διάφορες μαθηματικές πράξεις. Υπολογίζεται από τον τύπο που ακολουθεί:

$$L1 = |y - f(x)|$$

- **Απώλεια Huber (Huber Loss):** Η απώλεια Huber είναι μια μέθοδος που συνδυάζει τα θετικά των δύο παραπάνω προσεγγίσεων. Με βάση μια σταθερά  $d$  ακολουθείται η πρώτη ή η δεύτερη προσέγγιση. Στην περίπτωση σφαλμάτων μικρότερων του  $d$  ακολουθείται η μετρική L2 ενώ σε άλλη περίπτωση μία παραλλαγή της μετρικής L1

$$L_d(a) = \begin{cases} \frac{1}{2}a^2 & |a| \leq d \\ d(|a| - \frac{1}{2}d) & |a| > d \end{cases}$$

όπου  $a = y - f(x)$  και  $d$  επιλεγόμενη σταθερά.

### 3.4.6 Αλγόριθμοι Βελτιστοποίησης

Η πλειονότητα των αλγορίθμων βελτιστοποίησης των ΤΝΔ αποτελούν προεκτάσεις και παραλλαγές του αλγορίθμου καθόδου κλίσης (gradient descent). Ο αλγόριθμος καθόδου κλίσης, όπως είχε προταθεί αρχικά από τον γάλλο μαθηματικό Cauchy, είναι μια επαναληπτική αριθμητική μέθοδος πρώτης τάξης για την επίλυση εξισώσεων πολλών μεταβλητών [47]. Στον κλάδο των ΤΝΔ χρησιμοποιείται για την εύρεση ελαχίστων της συνάρτησης απώλειας και την διαδικασία βελτιστοποίησης των βαρών του. Εδώ θα αναφερθούμε σύντομα σε μερικές από τις πλέον διαδεδομένες παραλλαγές της μεθόδου αυτής με πολλές πρακτικές εφαρμογές, σύμφωνα με την βιβλιογραφία [48]:

- **Κάθοδος Κλίσης (Batch Gradient Descent):** Ο αλγόριθμος καθόδου κλίσης περιλαμβάνει την εύρεση της κλίσης μιας συνάρτησης πολλών μεταβλητών κατά μήκος

της οποίας κινούμαστε για να βρεθούμε σε κάποιο ελάχιστο. Ο βαθμός στον οποίο κινούμαστε πάνω στην καμπύλη αυτή, αποτελεί μία από τις βασικές υπερπαραμέτρους του αλγορίθμου, γνωστή ως ρυθμός εκμάθησης (learning rate). Στην συμβατική του εκδοχή, ο αλγόριθμος καθόδου κλίσης υπολογίζει την συνάρτηση απώλειας ολόκληρου του συνόλου δεδομένων πριν προβεί στην ανανέωση των βαρών. Επομένως χρειάζεται να πραγματοποιήσει πολλαπλούς υπολογισμούς για να διεξάγει μία και μόνη επανάληψη, κάτι που επιβραδύνει κατα πολύ την επίδοσή του ή ακόμα χειρότερα καθιστά την εφαρμογή του αλγορίθμου ανέφικτη σε περίπτωση που το σύνολο δεδομένων δεν χωράει στην μνήμη. Ως εκ τούτου, η σημασία του αλγορίθμου περιορίζεται κυρίως στην χρήση του ως θεωρητικό υπόβαθρο για πιο εξελιγμένες μεθόδους. Στο Σχ. 3.3 παρουσιάζεται εποπτικά η κάθοδος κλίσης. Οι ελλειψοειδείς γραμμές αντιστοιχούν σε ισομετρικές καμπύλες ενώ οι ευθείες αναπαριστούν διαδοχικές επαναλήψεις του αλγορίθμου.



Σχήμα 3.3: Διαδοχικές επαναλήψεις του αλγορίθμου καθόδου κλίσης [48]

- **Στοχαστική Κάθοδος Κλίσης (Stochastic Gradient Descent):** Ο αλγόριθμος στοχαστικής καθόδου κλίσης αποτελεί μία από τις βελτιωμένες επεκτάσεις του κλασικού αλγορίθμου. Η ουσιαστική του διαφορά έγκειται στο γεγονός ότι ο υπολογισμός της συνάρτησης απώλειας πραγματοποιείται για κάθε ένα από τα δείγματα του συνόλου ξεχωριστά αντί για το σύνολο των δεδομένων. Έτσι, κάθε ανανέωση των βαρών ελαχιστοποιεί το σφάλμα για την δική της μία και μόνη είσοδο κι όχι καθολικά. Αυτή η παραλλαγή οδηγεί σε βελτιωμένη ταχύτητα σύγκλισης αλλά και μεγαλύτερες διακυμάνσεις κατά την διαδικασία εκμάθησης.
- **Κάθοδος Κλίσης Βασισμένη σε Παρτίδες (Mini-Batch Gradient Descent):** Ο αλγόριθμος καθόδου κλίσης βασισμένης σε παρτίδες συνδυάζει επιτυχώς τα θετικά χαρακτηριστικά των δύο παραπάνω προσεγγίσεων. Αντί να προσπαθεί να ελαχιστοποιήσει την συνάρτηση σφάλματος κάθε δείγματος μεμονωμένα σε κάθε επανάληψη, συνενώνει περισσότερα από ένα δείγματα σε παρτίδες (batches) και πραγματοποιεί την διαδικασία που προαναφέραμε. Έτσι, έχει αυξημένη ταχύτητα σύγκλισης αλλά και μεγαλύτερη ευστάθεια. Σε αυτή την μέθοδο βασίζονται όλες οι επιτυχημένες προεκτάσεις του αλγορίθμου, οι οποίες εφαρμόζονται ευρύτατα στην πράξη. Τέτοιο παράδειγμα αποτελεί ο αλγόριθμος Adam [49] (Adaptive Moment Estimation) ο οποίος κάνει χρήση της έννοιας της ορμής (momentum), δηλαδή ανανεώνει τα βάρη έτσι ώστε να λαμ-

βάνεται υπόψη η γενικότερη κατεύθυνση των προηγούμενων ανανεώσεων. Επιπλέον, ο ίδιος αλγόριθμος χρησιμοποιεί μη σταθερούς, προσαρμοζόμενους (adaptive) ρυθμούς εκμάθησης για κάθε μία από τις παραμέτρους του. Άλλα παραδείγματα επιτυχημένων αλγορίθμων βασισμένων σε αυτή την τεχνική είναι οι AdaMax [49], RMSprop [50] και Nadam [51]



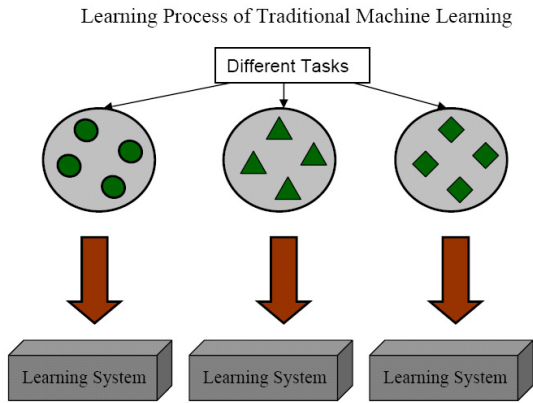
## Κεφάλαιο 4

# Μεταφορά Μάθησης

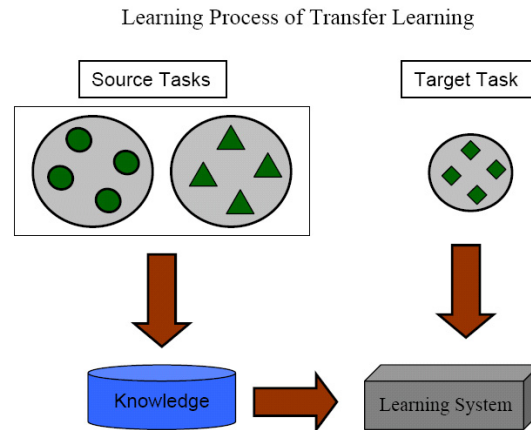
### 4.1 Εισαγωγή

Ένα από τα σημαντικότερα προβλήματα στην κλασική μηχανική μάθηση είναι η ανάγκη ύπαρξης ενός ικανοποιητικά μεγάλου συνόλου δεδομένων εκπαίδευσης, το οποίο χρειάζεται να είναι αντιπροσωπευτικό του συνόλου δεδομένων πάνω στο οποίο θα κληθεί το μοντέλο να παράξει προβλέψεις. Στη δική μας περίπτωση, αυτό μεταφράζεται σε ανάγκη συλλογής ενός μεγάλου πλήθους χρονοσειρών της ίδιας χρονικής συχνότητας και με αντίστοιχη μορφή ως προς τα βασικά ποιοτικά χαρακτηριστικά, αλλά και με αντιστοιχία ως προς τον κλάδο από τον οποίο προέρχονται, μια συνθήκη η οποία σε αρκετές περιπτώσεις είναι δύσκολο να ικανοποιηθεί. Έτσι, καλούμαστε να αναζητήσουμε εναλλακτικές επιλογές είτε με την παραγωγή τεχνητών δεδομένων (data augmentation), όπου ήδη προτείνονται αρκετές μέθοδοι [52], είτε με χρήση της μεταφοράς μάθησης (transfer learning), στην οποία εστιάζει η παρούσα διπλωματική εργασία.

Η μεταφορά μάθησης συνιστά τη μέθοδο μάθησης, κατά την οποία ένα μοντέλο στήνεται και εκπαιδεύεται με δεδομένα ενός συγκεκριμένου περιβάλλοντος, το λεγόμενο περιβάλλον προέλευσης (source domain), με σκοπό να παράξει προβλέψεις σε ένα άλλο περιβάλλον, το λεγόμενο περιβάλλον στόχο (target domain). Ακόμη, το μοντέλο εκπαιδεύεται στο περιβάλλον προέλευσης με τρόπο τέτοιο ώστε να επιτελεί ορθά την εργασία προέλευσης (source task) και, καλείται, έπειτα να επιτελέσει ορθά μια εργασία στόχο (target task). Αντί, λοιπόν, να εκπαιδεύσουμε ένα μοντέλο εκ του μηδενός (from scratch) για να επιτελέσει την εργασία στόχο, το εκπαιδεύουμε αρχικά με βάση την εργασία προέλευσης και μεταφέρουμε τη γνώση αυτή εμπλουτίζοντας την, προκειμένου να βελτιστοποιήσουμε την απόδοση του μοντέλου στην εργασία στόχο. Στο υποκεφάλαιο που ακολουθεί ορίζονται πληρέστερα οι προαναφερθείσες έννοιες, με χρήση μαθηματικού φορμαλισμού.

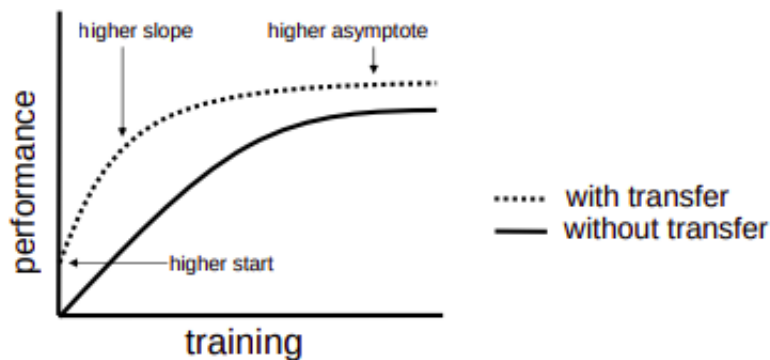


Σχήμα 4.1: Παραδοσιακή Μηχανική Μάθηση [53]



Σχήμα 4.2: Μεταφορά Μάθησης [53]

Η ίδια η μεταφορά μάθησης, συνιστά ένα είδος μάθησης με πολλαπλά οφέλη σε σχέση με την παραδοσιακή μηχανική μάθηση. Πρώτον, η αρχική επίδοση του μοντέλου προέλευσης στην εργασία στόχο πριν συμβεί οποιαδήποτε επανεκπαίδευση είναι υψηλότερη από την επίδοση ενός τυχαία αρχικοποιημένου μοντέλου για την ίδια εργασία (*higher start*). Δεύτερον, το εν λόγω μοντέλο συγκλίνει γρηγορότερα, άρα απαιτείται λιγότερος χρόνος μέχρι να ολοκληρωθεί η εκπαίδευσή του (*higher slope*). Τρίτον, το οποίο ίσως αποτελεί και το σπουδαιότερο πλεονέκτημα, το μοντέλο συγκλίνει σε υψηλότερα επίπεδα επίδοσης (*higher asymptote*). Τα παραπάνω παρουσιάζονται με εποπτικό τρόπο στο Σχ. 4.3.



Σχήμα 4.3: Πλεονεκτήματα μεταφοράς μάθησης. [54]

Εξέχουσα είναι η σημασία της μεταφοράς μάθησης στον τομέα των προβλέψεων και ως προς την συνεισφορά που η ίδια μπορεί να έχει από οικονομικής σκοπιάς, ειδικά σε επίπεδο μικρομεσαίων οργανισμών και επιχειρήσεων. Συγκεκριμένα, ένα από τα βασικότερα προβλήματα που αντιμετωπίζουν οι υπεύθυνοι προβλέψεων κατά τον σχεδιασμό συστημάτων που θα παρέχουν ποιοτικές προβλέψεις είναι η έλλειψη πληθώρας δεδομένων, γεγονός που δεν συνιστά μεγάλο πρόβλημα στις τοπικές προσεγγίσεις και τις παραδοσιακές μεθόδους πρόβλεψης, αλλά



δημιουργεί κωλύματα στο σχεδιασμό συστημάτων πρόβλεψης με χρήση μηχανικής μάθησης και νευρωνικών δικτύων. Καθώς στις μέρες μας η καταγραφή δεδομένων γίνεται όλο και πιο διαδεδομένη και καθολική και υπάρχουν ποικίλες μεγάλες ανοιχτές βιβλιοθήκες δεδομένων, υπάρχει η ευκαιρία για δημιουργία υπερμοντέλων, που θα προεκπαιδευτούν σε μεγάλα σύνολα δεδομένων και θα παρέχονται ανοιχτά προς πάσα χρήση, ακριβώς όπως συμβαίνει στον κλάδο των εικόνων, με μοντέλα όπως το ResNet [55] ή το ImageNet [56]. Μια τέτοια προσέγγιση μπορεί να οφελήσει τους οργανισμούς, καθώς πλέον θα μπορούν να αναπτύξουν μοντέλα που θα επιτελούν επιτυχώς την εργασία στόχο, χωρίς να χρειάζεται να τα δομήσουν εκ του μηδενός, αλλά κάνοντας κάποιες αλλαγές επί των παρεχόμενων υπερμοντέλων. Άλλωστε, η επανεκπαίδευση δεν έχει τις ίδιες απαιτήσεις ως προς τον όγκο των δειγμάτων, αλλά ούτε και το ίδιο κόστος. Εν κατακλείδι, μέσω της μεταφοράς μάθησης μπορεί να δοθεί η δυνατότητα σε μια πληθώρα οργανισμών και επιχειρήσεων να παράγουν ποιοτικότερες προβλέψεις με χαμηλότερο κόστος, κάνοντας χρήση της μηχανικής μάθησης.

Αν και τα οφέλη της μεταφοράς μάθησης είναι πολλαπλά, είναι σημαντικό να μην αγνοούμε την πιθανότητα της αρνητικής μεταφοράς (negative transfer), δηλαδή της αρνητικής επιρροής της μεταφοράς μάθησης στις επιδόσεις του μοντέλου. Κάτι τέτοιο θα μπορούσε να οφείλεται, κυρίως, σε μια κακή συσχέτιση μεταξύ του περιβάλλοντος συσχέτισης και στόχου, με αποτέλεσμα η γνώση που παίρνει το μοντέλο εκμεταλλευόμενο το περιβάλλον και την εργασία προέλευσης να παρακωλύει την εκπαίδευση του στο περιβάλλον στόχου και την ορθή επίτευξη της εργασίας στόχου. Στη βιβλιογραφία [57] αναφέρεται, πως στην περίπτωση ύπαρξης πολλαπλών διαθέσιμων συνόλων δεδομένων, θα πρέπει να επιλέγεται κάθε φορά το πλέον σχετικό με το περιβάλλον στόχου, ενώ, στην περίπτωση ύπαρξης μόνο ενός συνόλου δεδομένων και μερικής ομοιότητας των δύο περιβαλλόντων ανά σημεία, θα πρέπει να γίνεται διαχωρισμός του περιβάλλοντος προέλευσης σε τμήματα και χρήση μόνο των τμημάτων εκείνων που σχετίζονται επαρκώς με το περιβάλλον στόχου.

## 4.2 Βασικοί Ορισμοί και Είδη Μεταφοράς Μάθησης

Για να αποσαφηνιστούν καλύτερα οι έννοιες που πραγματεύεται η παρούσα εργασία και να διευκολυνθεί η κατανόηση όσων θα αναφερθούν στην συνέχεια αναφορικά με τη μεταφορά μάθησης, είναι σημαντικό να τοποθετηθούν οι αυστηροί ορισμοί τους, με χρήση μαθηματικού φορμαλισμού.

**Ορισμός 4.1** ([57]). Ένα περιβάλλον (domain)  $\mathcal{D}$  ορίζεται από δύο βασικά στοιχεία, τον χώρο χαρακτηριστικών (feature space)  $\mathcal{X}$  και την οριακή κατανομή πιθανότητας (marginal probability distribution)  $P(X)$ , όπου  $X = x_1, \dots, x_n \in \mathcal{X}$ . Επομένως, ισχύει  $\mathcal{D} = \{\mathcal{X}, P(X)\}$ .

**Ορισμός 4.2** ([57]). Δοθέντος ενός περιβάλλοντος  $\mathcal{D}$ , μια εργασία (task)  $\mathcal{T}$  ορίζεται από δύο βασικά στοιχεία, το χώρο ετικετών (label space)  $\mathcal{Y}$  και μια συνάρτηση πρόβλεψης  $f(\cdot)$ , την οποία μαθαίνουμε από τα ζεύγη διανυσμάτων χαρακτηριστικών (feature vector) και ετικετών (labels)  $\{x_i, y_i\}$ , όπου  $x_i \in \mathcal{X}$  και  $y_i \in \mathcal{Y}$ . Επομένως, ισχύει  $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$  ή ισοδύναμα  $\mathcal{T} = \{\mathcal{Y}, P(Y|X)\}$ .

**Ορισμός 4.3** ([57]). Ορίζουμε, επιπρόσθετα, ως δεδομένα του περιβάλλοντος προέλευσης (*domain source data*) και συμβολίζουμε με  $D_S$ , το περιβάλλον  $D_S = \{(x_{S1}, y_{S1}), \dots, (x_{Sn}, y_{Sn})\}$ , όπου  $x_{Si} \in \mathcal{X}_S$  είναι το  $i$ -οστό στοιχείο του  $D_S$  και  $y_{Si} \in \mathcal{Y}_S$  είναι η αντίστοιχη ετικέτα για το  $x_{Si}$ .

**Ορισμός 4.4** ([57]). Ορίζουμε, αντίστοιχα, ως δεδομένα του περιβάλλοντος στόχου (*target domain data*) και συμβολίζουμε με  $D_T$ , το περιβάλλον  $D_T = \{(x_{T1}, y_{T1}), \dots, (x_{Tn}, y_{Tn})\}$ , όπου  $x_{Ti} \in \mathcal{X}_T$  είναι το  $i$ -οστό στοιχείο του  $D_T$  και  $y_{Ti} \in \mathcal{Y}_T$  είναι η αντίστοιχη ετικέτα για το  $x_{Ti}$ .

Οι εργασίες προέλευσης και στόχου συμβολίζονται με  $\mathcal{T}_S$  και  $\mathcal{T}_T$  αντίστοιχα. Η συνάρτηση πρόβλεψης της προέλευσης συμβολίζεται με  $f_S(\cdot)$ , ενώ η συνάρτηση πρόβλεψης στόχου με  $f_T(\cdot)$ .

**Ορισμός 4.5** ([57]). Δοθέντος ενός περιβάλλοντος προέλευσης  $D_S$  μαζί με τη αντίστοιχη εργασία προέλευσης  $\mathcal{T}_S$  και ενός περιβάλλοντος στόχου  $D_T$  με την αντίστοιχη εργασία στόχου  $\mathcal{T}_T$ , η μεταφορά μάθησης είναι η διαδικασία κατά την οποία προκαλείται βελτίωση στην συνάρτηση πρόβλεψης  $f_T(\cdot)$ , χρησιμοποιώντας τη γνώση που προέρχεται από τα  $D_S$  και  $\mathcal{T}_S$ , όπου  $D_S \neq D_T$  και  $\mathcal{T}_S \neq \mathcal{T}_T$ .

Έχοντας δώσει όλους τους προηγούμενους εκτενείς ορισμούς, μπορούμε να διαχωρίσουμε τη μεταφορά μάθησης στα είδη της, με βάση το πόσο διαφέρουν τα περιβάλλοντα προέλευσης και στόχου και οι εργασίες προέλευσης και στόχου [53]:

- **Επαγωγική μεταφορά μάθησης (Inductive Transfer Learning):** Τα περιβάλλοντα προέλευσης και στόχου είναι κοινά, όμως οι εργασίες προέλευσης και στόχου διαφέρουν αλλά σχετίζονται.
- **Μη επιβλεπόμενη μεταφορά μάθησης (Unsupervised Transfer Learning):** Τα περιβάλλοντα προέλευσης και στόχου διαφέρουν αλλά σχετίζονται, ενώ το ίδιο ισχύει και για τις εργασίες προέλευσης και στόχου.
- **Μεταγωγική μεταφορά μάθησης (Transductive Transfer Learning):** Τα περιβάλλοντα προέλευσης και στόχου διαφέρουν αλλά σχετίζονται, ενώ οι εργασίες προέλευσης και στόχου είναι κοινές.

Ασφαλώς, η περίπτωση που προέλευση και στόχος ταυτίζονται αναφορικά τόσο με την εργασία, όσο και με το περιβάλλον, είναι η περίπτωση της παραδοσιακής μηχανικής μάθησης. Φυσικά, όσο περισσότερο διαφέρουν η προέλευση και ο στόχος, ως προς τους δύο βασικούς άξονες, τόσο δυσκολότερο γίνεται το έργο της μεταφοράς μάθησης. Στην περίπτωση αυτή, κρίνεται απαραίτητη η εύρεση μεθόδων επανεκπαίδευσης (*retraining*) του μοντέλου, με σκοπό την βελτίωση της απόδοσης του στην εργασία στόχο. Ωστόσο, ακόμη και σε αυτές τις περιπτώσεις, τα πλεονεκτήματα της χρήσης της μεταφοράς μάθησης είναι μεγάλα, καθώς η επανεκπαίδευση κοστίζει λιγότερο σε χρόνο σε σχέση με την δημιουργία ενός νέου μοντέλου εκ του μηδενός (*from scratch*) και, παράλληλα, παρέχει μεγαλύτερη ακρίβεια, καθώς τα αρχικά βάρη στην περίπτωση αυτή δεν είναι τυχαία αρχικοποιημένα.

### 4.3 Περαιτέρω Διάκριση Ειδών Μεταφοράς Μάθησης

Όπως προκύπτει από τους ορισμούς του περιβάλλοντος  $\mathcal{D} = \{\mathcal{X}, P(X)\}$  και της εργασίας  $\mathcal{T} = \{\mathcal{Y}, P(Y|X)\}$  σε ένα πρόβλημα μεταφοράς μάθησης, είναι φανερό ότι τα περιβάλλοντα ή οι εργασίες προέλευσης και στόχου μπορούν να διαφέρουν μεταξύ τους σε ένα ή και περισσότερα εκ των  $\mathcal{X}, P(X), \mathcal{Y}, P(Y|X)$ . Με βάση αυτή την διαφορά, διακρίνουμε τις τέσσερις παρακάτω κατηγορίες.

#### 4.3.1 Με βάση τον Χώρο Χαρακτηριστικών $\mathcal{X}$

Μια ουσιώδης διάκριση των προβλημάτων μεταφοράς μάθησης αφορά στους χώρους χαρακτηριστικών του περιβάλλοντος προέλευσης  $\mathcal{X}_S$  και του περιβάλλοντος στόχου  $\mathcal{X}_T$ . Όταν οι δύο χώροι χαρακτηριστικών ταυτίζονται, δηλαδή όταν ισχύει  $\mathcal{X}_S = \mathcal{X}_T$ , τότε μιλάμε για ομογενή μεταφορά μάθησης (homogeneous transfer learning) ενώ σε αντίθετη περίπτωση, δηλαδή όταν  $\mathcal{X}_S \neq \mathcal{X}_T$ , αναφερόμαστε σε ετερογενή μεταφορά μάθησης (heterogeneous transfer learning). Αξίζει να σημειώσουμε ότι στην περίπτωση της ετερογενούς μεταφοράς μάθησης, η οριακή κατανομή πιθανότητας  $P(X)$  θεωρείται ίδια και στα δύο περιβάλλοντα.

Για να γίνει η διάκριση αυτή περαιτέρω κατανοητή αναφέρουμε δύο παραδείγματα. Ας σκεφτούμε, για παράδειγμα, την εργασία πρόβλεψης της αξίας ενός πελάτη (customer lifetime value). Εάν στο περιβάλλον προέλευσης ο χώρος χαρακτηριστικών είναι  $\mathcal{X}_S = (\text{ηλικία}, \text{ετήσιο εισόδημα}, \text{μέσο ύψος αγορών})$  και στο περιβάλλον στόχου ο χώρος χαρακτηριστικών είναι  $\mathcal{X}_T = (\text{αριθμός επισκέψεων ετησίως}, \text{ύψος αγορών μέχρι τώρα})$ , τότε προφανώς αντιλαμβανόμαστε ότι πρόκειται για περίπτωση ετερογενούς μεταφοράς μάθησης. Αντίστοιχα, θεωρώντας πως το περιβάλλον προέλευσης και το περιβάλλον στόχου είναι κοινά κι έστω, ακόμη, ότι η εργασία στόχου είναι η πρόβλεψη της επόμενης επίσκεψης του πελάτη, ενώ η εργασία προέλευσης είναι η αξία του πελάτη όπως περιγράφηκε παραπάνω, τότε αυτό αποτελεί ένα παράδειγμα ομογενούς μεταφοράς μάθησης. Η παρούσα εργασία επικεντρώνεται στην ομογενή μεταφορά μάθησης.

#### 4.3.2 Με βάση την Οριακή Κατανομή Πιθανότητας $P(X)$

Εκτός από τον χώρο χαρακτηριστικών  $\mathcal{X}$ , το περιβάλλον προέλευσης  $D_S$  και το περιβάλλον στόχου  $D_T$  μπορούν να διαφέρουν και ως προς την οριακή κατανομή πιθανότητας  $P(X)$ . Ας θεωρήσουμε για παράδειγμα το πρόβλημα ομογενούς μεταφοράς μάθησης που περιγράψαμε προηγουμένως κατά το οποίο τόσο η εργασία προέλευσης όσο και η εργασία στόχου αφορούν στην πρόβλεψη της αξίας ενός πελάτη. Αν το περιβάλλον προέλευσης αφορά σε δεδομένα λιανικής πώλησης τροφίμων ενώ το περιβάλλον στόχου αφορά σε δεδομένα πώλησης αυτοκινήτων πολυτελείας, είναι λογικό να αναμένουμε ότι  $P_S(X) \neq P_T(X)$ . Αυτό το είδος προκατάληψης (bias) είναι γνωστό στην βιβλιογραφία ως προκατάληψη συχνότητας των χαρακτηριστικών (frequency feature bias).

### 4.3.3 Με βάση την Δεσμευμένη Κατανομή Πιθανότητας $P(Y|X)$

Σε αντιστοιχία, οι εργασίες προέλευσης  $\mathcal{T}_S$  και οι εργασίες στόχου  $\mathcal{T}_T$  μπορεί να διαφέρουν είτε ως προς την δεσμευμένη κατανομή πιθανότητας  $P(Y|X)$  (conditional probability distribution), είτε ως προς τον χώρο ετικετών  $\mathcal{Y}$ . Στην περίπτωση όπου  $P(Y_S|X_S) \neq P(Y_T|X_T)$ , μπορούμε να ανατρέξουμε σε ένα παράδειγμα ανάλυσης κειμένου για την κατηγοριοποίηση κριτικών προϊόντων. Εύκολα μπορούμε να φανταστούμε ότι εάν, για παράδειγμα, το περιβάλλον προέλευσης αποτελείται από κριτικές σε κινητά τηλέφωνα ενώ το περιβάλλον στόχου από κριτικές δωματίων ξενοδοχείου, τότε στην πρώτη περίπτωση η λέξη “μικρό” υποδηλώνει ένα μάλλον θετικό χαρακτηριστικό, ενώ στην δεύτερη ένα μάλλον αρνητικό. Αυτό το είδος προκατάληψης είναι γνωστό στην βιβλιογραφία ως προκατάληψη εκ των συμφραζομένων των χαρακτηριστικών (context feature bias).

### 4.3.4 Με βάση τον Χώρο Ετικετών $\mathcal{Y}$

Η περίπτωση  $\mathcal{Y}_S \neq \mathcal{Y}_T$  λαμβάνει χώρα όταν το αποτέλεσμα μιας εργασίας στόχου εκφράζεται με διαφορετικό τρόπο από το αποτέλεσμα της εργασίας προέλευσης. Επιστρέφουμε στο προηγούμενο παράδειγμά μας για τον υπολογισμό της αξίας ενός πελάτη. Υποθέτουμε ότι η εργασία προέλευσης αποφαινεται με δυαδικές τιμές: *true* για πελάτες μεγάλης αξίας και *false* για τους υπόλοιπους. Εάν εμείς στην εργασία στόχου θέλουμε να επιτύχουμε παλινδρόμηση της αξίας του πελάτη, δηλαδή στην έξοδο του μοντέλου μας αναμένουμε αριθμητική τιμή, τότε το συγκεκριμένο πρόβλημα μεταφοράς μάθησης ανήκει στην κατηγορία αυτή.

## 4.4 Ενδεικτικές Μέθοδοι Μεταφοράς Μάθησης

Στη βιβλιογραφία υπάρχει πληθώρα προτεινόμενων μεθόδων μεταφοράς μάθησης με άλλες εξ αυτών να προσφέρονται για γενική εφαρμογή, άλλες να είναι εξειδικευμένες σε συγκεκριμένες αρχιτεκτονικές και άλλες να ενδείκνυνται για συγκεκριμένους κλάδους. Μια από τις βασικότερες και γενικότερες μεθόδους αφορά την προεκπαίδευση ενός υπερ-μοντέλου με χρήση ενός τμήματος ή ολόκληρου του συνόλου δεδομένων που έρχεται από το περιβάλλον προέλευσης, προκειμένου το ίδιο να εκτελέσει την εργασία στόχου, με ή χωρίς επανεκπαίδευση. Η προσέγγιση αυτή ενδείκνυται ανεξαρτήτως αρχιτεκτονικής και έχει εφαρμοστεί σε προβλήματα με εικόνες [56, 58], επεξεργασία φυσικής γλώσσας [59], άλλα και ποικίλους άλλους κλάδους. Στα πλαίσια της στρατηγικής αυτής, έχουν μελετηθεί διαφορετικές τεχνικές ως προς την επανεκπαίδευση του υπερ-μοντέλου, προκειμένου να βελτιωθεί η προβλεπτική του ακρίβεια, με τις βασικότερες εξ αυτών να αναφέρονται στη συνέχεια.

### 4.4.1 Ευνοϊκό Ξεκίνημα

Στη μέθοδο του ευνοϊκού ξεκινήματος (warm start) τα βάρη του υπερ-μοντέλου, όπως αυτά έχουν καθοριστεί από την προεκπαίδευση του, χρησιμοποιούνται ως αρχικά βάρη για την

εκπαίδευση με το σύνολο δεδομένων του περιβάλλοντος στόχου. Αυτή η μέθοδος αφοραμάται από το γεγονός ότι όταν ένα πρόβλημα μάθησης μοιράζεται μία κοινή εργασία ή ένα σύνολο δεδομένων με κάποιο άλλο, τότε ένα μέρος της συνολικής γνώσης έχει ήδη αποκτηθεί. Με άλλα λόγια, τα βάρη του δικτύου που επανεκπαιδεύεται δεν θα είναι τυχαία αρχικοποιημένα και, για τον λόγο αυτό, θα είναι πιο κοντά στην τιμή στην οποία θα συγκλίνουν τελικώς μετά το τέλος της επανεκπαίδευσης. Ως εκ τούτου, παρατηρείται πως ο ρυθμός με τον οποίο αυξάνεται η προβλεπτική ακρίβεια του μοντέλου ανά μονάδα χρόνου εκπαίδευσης είναι πολύ μεγαλύτερος. Μάλιστα, μπορούμε να πούμε πως ένα ευνοϊκό ξεκίνημα αποτελεί πολύ σημαντικό σημείο υπεροχής, ιδιαίτερα στην περίπτωση που τα δεδομένα του περιβάλλοντος στόχου είναι λίγα σε πλήθος. Ένα ακόμη πλεονέκτημα της μεθόδου αυτής είναι ότι η τυχαία αρχικοποίηση καθορίζει σε μεγάλο βαθμό την πορεία προς την οποία θα κατευθυνθούν τα βάρη του μοντέλου, με αποτέλεσμα αν ο ρυθμός εκμάθησης είναι ακατάλληλος, να είναι δυνατό τα βάρη να οδηγηθούν σε τοπικά ελάχιστα του σφάλματος και το μοντέλο να μην συγκλίνει τελικά σε ένα καλό αποτέλεσμα. Αυτό στην περίπτωση του ευνοϊκού ξεκινήματος αποφεύγεται.

#### 4.4.2 Φραγμένες Εποχές

Η μέθοδος των φραγμένων εποχών (bounded epochs) αξιοποιεί τα βάρη του υπερ-μοντέλου ως αρχικοποίηση, ωστόσο δεν επιτρέπει μια πλήρη επανεκπαίδευση στο περιβάλλον στόχου. Αντιθέτως, φράσει την επανεκπαίδευση σε ένα μικρό πλήθος εποχών. Έναυσμα της μεθόδου αυτής είναι η ανάγκη για παραγωγή ποιοτικών προβλέψεων σε σύντομο χρονικό διάστημα. Μέσα σε έναν περιορισμένο αριθμό εποχών, συνήθως 1, 2 ή 5, το μοντέλο θα έχει καταφέρει να αναγνωρίσει με ικανοποιητική ακρίβεια, αλλά όχι πλήρως, το σύνολο δεδομένων στόχου. Στην διαφορετική περίπτωση που ξεκινούσαμε από ένα μοντέλο με τυχαία αρχικοποιημένα βάρη, η προβλεπτική ακρίβεια του μοντέλου μετά από εκπαίδευση ελάχιστων εποχών θα ήταν ασήμαντη. Για τον λόγο αυτό, και ιδιαίτερα στην περίπτωση της ανάγκης παραγωγής προβλέψεων ταχύτατα με ένα μικρό μόνο κόστος στην ποιότητα, η μέθοδος των φραγμένων εποχών αποκτά εξέχουσα σημασία.

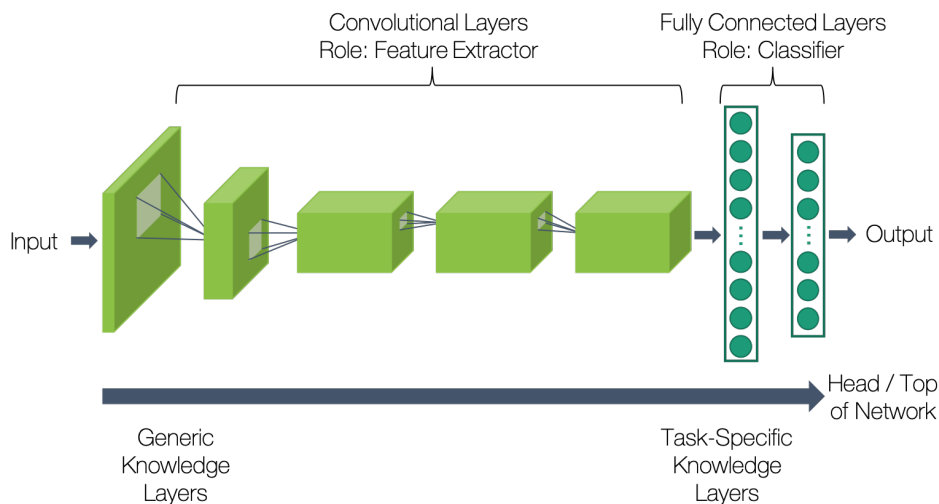
#### 4.4.3 Πάγωμα Επιπέδων

Η μέθοδος του παγώματος επιπέδων (layers freezing) αξιοποιεί ακόμα περισσότερο τα βάρη του υπερ-μοντέλου, παγώνοντας την εκπαίδευση για ένα υποσύνολο των επιπέδων του μοντέλου, διατηρώντας τα ακριβώς όπως προέκυψαν από την προεκπαίδευση. Στα υπόλοιπα επίπεδα, τα ήδη υπάρχοντα βάρη χρησιμοποιούνται και πάλι ως αρχικοποίηση. Το πάγωμα αφορά τα κρυμμένα επίπεδα και ξεκινάει από το επίπεδο που ακολουθεί το επίπεδο εισόδου προς το επίπεδο εξόδου. Η κατεύθυνση αυτή έχει αποδειχτεί βέλτιστη, διότι τα πρώτα επίπεδα ενός μοντέλου τείνουν να απεικονίζουν τα γενικά χαρακτηριστικά των δεδομένων εκπαίδευσης, ενώ αυτά που βρίσκονται πιο κοντά στην έξοδο -τα οποία χαρακτηρίζονται και ως κεφαλή- είναι υπεύθυνα για την παραγωγή των προβλέψεων -ή την κατηγοριοποίηση σε περίπτωση προβλημάτων ταξινόμησης-. Ένα από τα πλεονεκτήματα της μεθόδου αυτής είναι το γεγονός

πως το πάγωμα ορισμένων επιπέδων μειώνει τις παραμέτρους προς εκπαίδευση του δικτύου και συνεπώς τον όγκο των μαθηματικών υπολογισμών που απαιτούνται για την διενέργεια ενός βήματος εκπαίδευσης. Έτσι, ο χρόνος της επανεκπαίδευσης επιταχύνεται αισθητά.

#### 4.4.4 Αντικατάσταση της Κεφαλής

Η μέθοδος αυτή προτείνει την αντικατάσταση της κεφαλής (head replacement) του μοντέλου με ταυτόχρονο πάγωμα ή μη των προηγούμενων επιπέδων. Χρησιμοποιείται εκτενώς σε περιπτώσεις όπου η εργασία στόχου διαφέρει από την εργασία προέλευσης, προκειμένου η νέα κεφαλή να είναι προσαρμοσμένη στην εργασία στόχου. Η ίδια έχει καλύτερα αποτελέσματα από τις προηγούμενες μεθόδους όταν το περιβάλλον στόχου απέχει σημαντικά από το περιβάλλον προέλευσης. Η μέθοδος αυτή έχει ευρύτερη εφαρμογή στην κατηγορία των βαθιών συνελικτικών δικτύων (CNN) βλ. Σχ. 4.4, όπου τα πρώτα συνελικτικά επίπεδα αναπαριστούν πληροφορία χαμηλού επιπέδου, όπως για παράδειγμα η αναγνώριση ακμών ή γωνιών, η οποία μπορεί να χρησιμοποιηθεί από τα επόμενα συνελικτικά επίπεδα για την κωδικοποίηση χαρακτηριστικών υψηλού επιπέδου, λόγω χάριν τροχούς οχημάτων, με σκοπό την αναγνώριση μοτίβων διαφορετικών από εκείνα της αρχικής εργασίας προέλευσης, όπως είναι για παράδειγμα η αναγνώριση ενός άλλου είδους τροχοφόρου οχήματος.



Σχήμα 4.4: Αναπαράσταση της δομής ενός συνελικτικό νευρωνικό δίκτυο (CNN). Τα πρώτα επίπεδα κωδικοποιούν γενική πληροφορία ενώ τα τελευταία αναπαριστούν χαρακτηριστικά προσαρμοσμένα στο εκάστοτε πρόβλημα. Από εδώ προκύπτουν οι μέθοδοι μεταφοράς μάθησης όπως η αντικατάσταση της κεφαλής και το πάγωμα επιπέδων. [60]

## Κεφάλαιο 5

# Πειραματική Διαδικασία

### 5.1 Εισαγωγή

Στο παρόν κεφάλαιο περιγράφεται εκτενώς η πειραματική διαδικασία που ακολουθήσαμε για τις ανάγκες της παρούσας εργασίας. Αρχικά, αναφερόμαστε στην μορφή και το είδος των δεδομένων με τα οποία εργαστήκαμε καθώς και στην μεθοδολογία επεξεργασίας, τυποποίησης και μετασχηματισμού τους. Έπειτα, εισάγεται το μοντέλο βάσης μας, δηλαδή η αρχιτεκτονική του νευρωνικού δικτύου με το οποίο διεξήχθησαν τα πειράματά μας. Στην συνέχεια, περιγράφεται λεπτομερώς η διάταξη και η εκτέλεση των πειραμάτων και, τέλος, εισάγονται δύο άλλα μοντέλα πρόβλεψης ως benchmarks της επίδοσης του μοντέλου μας.

### 5.2 Σύνολα Δεδομένων: M4 και ForeDeCk

#### 5.2.1 Δεδομένα

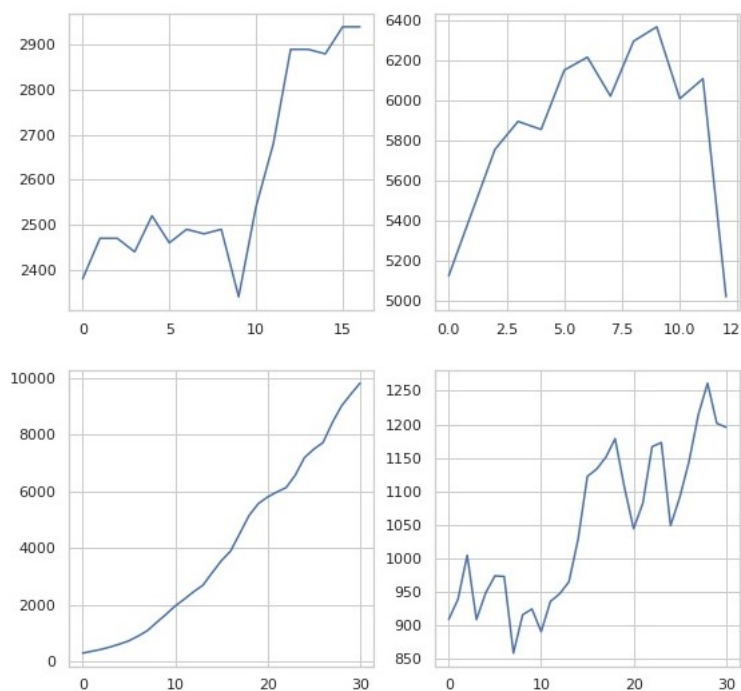
Το σύνολο δεδομένων M4 [61] αποτελεί μία συλλογή από 100.000 συνεχείς χρονοσειρές διαφόρων συχνοτήτων, με διαφορετικό πλήθος παρατηρήσεων η καθεμία, προερχόμενων από ένα μεγάλο εύρος τομέων όπως είναι η βιομηχανία, οι επιχειρήσεις, η διακυβέρνηση, τα δημογραφικά φαινόμενα και το χρηματιστήριο, ανάμεσα σε άλλα. Το σύνολο αυτό δημιουργήθηκε για να εξυπηρετήσει τις ανάγκες του ομώνυμου διαγωνισμού προβλέψεων M4 που διεξήχθη το 2018.

Το σύνολο δεδομένων M4 επιλέχθηκε ως σύνολο δεδομένων στόχου για την μέτρηση της επίδοσης των στρατηγικών μεταφοράς μάθησης που πραγματεύεται η παρούσα εργασία. Η επιλογή του M4 δικαιολογείται από το γεγονός ότι δίνει την δυνατότητα να μετρηθεί η αποτελεσματικότητα των μεθόδων μεταφοράς μάθησης συγκριτικά με τις πιο επιτυχημένες μεθόδους του διαγωνισμού.

Το σύνολο δεδομένων ForeDeCk αποτελεί μια μεγαλύτερη συλλογή δεδομένων που ανθολογήθηκε από το Εθνικό Μετσόβιο Πολυτεχνείο και περιλαμβάνει περί τις 900.000 χρονοσειρές, όμοιες με αυτές που περιγράφονται παραπάνω. Αξίζει να αναφερθεί μάλιστα, πως το

σύνολο δεδομένων M4 δημιουργήθηκε δειγματοληπτώντας τυχαία 100.000 χρονοσειρές από το ForeDeCk, οι οποίες αφαιρέθηκαν στην συνέχεια από εκείνο για να μην υπάρχει επικάλυψη μεταξύ των δύο συνόλων δεδομένων. Επομένως, υπό την μαθηματική έννοια, το σύνολο δεδομένων M4 δεν είναι υποσύνολο του ForeDeCk, όμως τα δύο σύνολα αυτά μοιράζονται μια αρκετά κοντινή οριακή κατανομή πιθανότητας των δεδομένων τους, γεγονός στο οποίο θα αναφερθούμε πιο αναλυτικά σε επόμενη υποενότητα του κεφαλαίου. Ένα υποσύνολο του ForeDeCk που αριθμεί 463.977 χρονοσειρές, αποτέλεσε το σύνολο δεδομένων προέλευσης. Στο πλαίσιο της παρούσας εργασίας θα αναφερόμαστε στο σύνολο αυτό απλά ως ForeDeCk.

Η παρούσα διπλωματική εστιάζεται στην πρόβλεψη χρονοσειρών χαμηλής συχνότητας, πιο συγκεκριμένα στις ετήσιες χρονοσειρές. Οι ετήσιες χρονοσειρές του M4 έχουν πλήθος 23.000 και κάθε μία από αυτές αριθμεί τουλάχιστον 13 παρατηρήσεις, ενώ η μεγαλύτερη αριθμεί 835. Ο ορίζοντας πρόβλεψης για τις ετήσιες χρονοσειρές του διαγωνισμού M4 ήταν 6 έτη, σύμβαση η οποία ακολουθήθηκε και στην παρούσα διπλωματική εργασία. Στην συνέχεια της παρούσας εργασίας, θα αναφερόμαστε σε αυτές τις 23.000 ετήσιες χρονοσειρές ως M4.



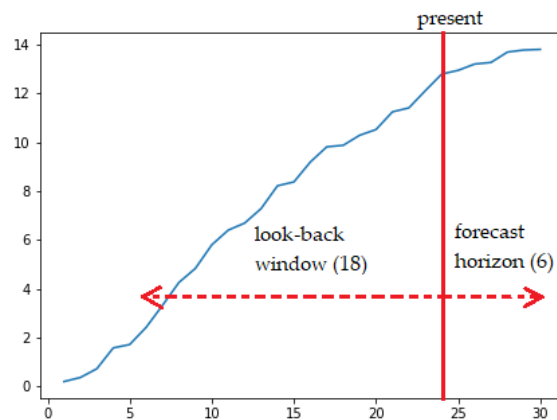
Σχήμα 5.1: Γραφική παράσταση τεσσάρων διαφορετικών χρονοσειρών του M4. Αξίζει να σημειωθεί το διαφορετικό μήκος κάθε χρονοσειράς. Πάνω αριστερά: έντονη αλλαγή επιπέδου. Πάνω δεξιά: αυξητική τάση και απότομη μείωση στην συνέχεια. Κάτω αριστερά: αυξητική γραμμική τάση. Κάτω δεξιά: αυξητική τάση με διακυμάνσεις



## 5.3 Προετοιμασία Πειραμάτων

### 5.3.1 Επεξεργασία και Παραθυροποίηση Δεδομένων

Για την πρόβλεψη μιας χρονοσειράς απαιτείται πρώτα ο καθαρισμός και η επεξεργασία των δεδομένων (data cleansing) όπως ακριβώς και στις άλλες μεθόδους πρόβλεψης. Ένα τέτοιο παράδειγμα αποτελεί η συμπλήρωση απουσιαζόντων τιμών (missing values). Η πρώτη ενέργεια που πρέπει να γίνει είναι η επεξεργασία των ακατέργαστων δεδομένων (raw data) έτσι ώστε αυτά να έρθουν σε τυποποιημένη μορφή. Αυτό είναι απαραίτητο καθώς οι περισσότερες κατηγορίες νευρωνικών δικτύων μαθαίνουν από δεδομένα με σταθερές διαστάσεις εισόδου και εξόδου. Με τον τρόπο αυτό, τα αποτελέσματα στην έξοδο ενός νευρωνικού δικτύου είναι ερμηνεύσιμα κι έχουν φυσικό νόημα. Στον κλάδο πρόβλεψης χρονοσειρών αυτό σημαίνει πρακτικά να παραθυροποιήσουμε την χρονοσειρά, κατακερματίζοντας τη σε όμοια χρονικά παράθυρα εισόδου και εξόδου κατασκευάζοντας έτσι τα δείγματα μας για εκπαίδευση. Για ετήσιες χρονοσειρές, τις οποίες πραγματεύεται η παρούσα εργασία, οι συνήθεις διαστάσεις εισόδου και εξόδου είναι 18 και 6 αντίστοιχα [11]. Αυτό πρακτικά σημαίνει ότι στο μοντέλο εισάγονται χρονικές παρατηρήσεις των 18 τελευταίων ετών κι αυτό καλείται να αποφανθεί σχετικά με την εξέλιξη του φαινομένου κατα τα επόμενα 6 έτη. Δηλαδή το χρονικό παράθυρο ενδιαφέροντος (look-back window) έχει μήκος 18, ενώ ο ορίζοντας πρόβλεψης 6. Ένα ζεύγος ενός χρονικού παραθύρου εισόδου μαζί με την αντίστοιχη έξοδό του αποτελεί ένα δείγμα εκπαίδευσης (training sample).



Σχήμα 5.2: Σχηματική απεικόνιση του χρονικού παραθύρου ενδιαφέροντος και του ορίζοντα πρόβλεψης.

Για να κατασκευάσουμε το σύνολο εκπαίδευσης του μοντέλου που περιέχει τα δείγματα πάνω στα οποία θα εκπαιδεύσουμε το μοντέλο μας μπορούμε να ακολουθήσουμε δύο προσεγγίσεις:

- **Τελευταία Παράθυρα (Last Windows):** Ο πρώτος και πιο απλός τρόπος κάνει χρήση μόνο του τελευταίου χρονικού παραθύρου μιας χρονοσειράς. Παραδείγματος χάριν, για τις διαστάσεις εισόδου-εξόδου που αναφέραμε παραπάνω, εξάγονται από κάθε χρονοσειρά τα 24 τελευταία δείγματα, όπου τα 18 πρώτα αποτελούν την είσοδο ενώ τα 6 τελευταία την έξοδο. Όπως είναι προφανές, με την μέθοδο των τελευταίων χρονικών

παραθύρων (last windows) από κάθε χρονοσειρά εξάγεται ένα μόνο ζεύγος για την εκπαίδευση του μοντέλου.

- **Όλα τα Παράθυρα (All Windows):** Ο δεύτερος τρόπος είναι αυτός που χρησιμοποιείται ευρύτατα στην πράξη μιας και κάνει χρήση όλης της προσφερόμενης πληροφορίας μιας χρονοσειράς. Η μέθοδος αυτή εξάγει ζεύγη εισόδου-εξόδου μετακυλώντας το χρονικό παράθυρο που αναφέραμε παραπάνω σε όλο το μήκος της χρονοσειράς, αποχρυσταλλώνοντας έτσι κάθε ένα από τα διαφορετικά στιγμιότυπά της σε όλα τα δυνατά παράθυρα (all windows). Η μέθοδος αυτή προσφέρει ένα πιο πλούσιο σύνολο δεδομένων εκπαίδευσης, όπως είναι αναμενόμενο, κι έχει προταθεί ως μέθοδος επαύξησης δεδομένων (data augmentation).

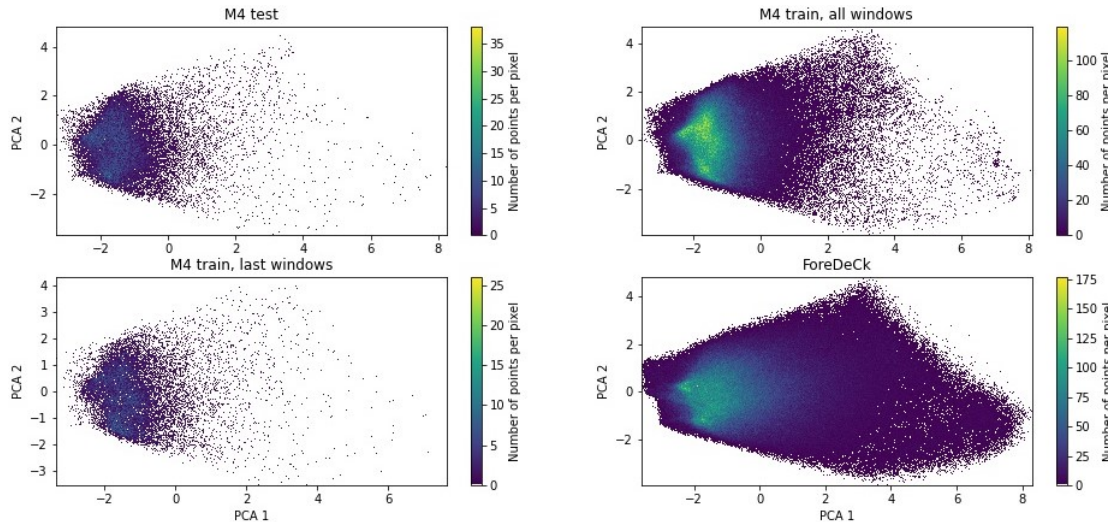
Έτσι κατασκευάζουμε τα σύνολα δεδομένων εκπαίδευσης M4 last windows και M4 all windows. Με όμοιο τρόπο κατασκευάζεται και το σύνολο δεδομένων ForeDeCk last windows, το οποίο χρησιμοποιείται εξ' ολοκλήρου στην παρούσα εργασία. Αξίζει να σημειωθεί ότι η πληθικότητα των συνόλων M4 last windows και M4 all windows είναι 14.242 και 235.460 αντίστοιχα ενώ του ForeDeCk είναι 463.977 παράθυρα εκπαίδευσης, όπως φαίνεται στο Σχ. 5.3. Αξίζει να αναφέρουμε πως το σύνολο δεδομένων M4 all windows είναι αυτό που χρησιμοποιείται τόσο για επανεκπαίδευση όσο και για εκπαίδευση εκ του μηδενός του μοντέλου μας.

Έπειτα σε κάθε ένα από τα χρονικά παράθυρα τα οποία προορίζονται για εκπαίδευση εφαρμόζουμε κλιμάκωση ελαχίστου-μεγίστου (min-max scaling): Αρχικά, η κλιμάκωση των δεδομένων μεγίστου-ελαχίστου κρίνεται απαραίτητη για την μάθηση ενός νευρωνικού δικτύου. Πραγματοποιώντας την κλιμάκωση φέρνουμε όλα τα δεδομένα στην ίδια κλίμακα από 0 έως 1 διατηρώντας παράλληλα την εσωτερικά συνοχή μεταξύ των παρατηρήσεων κάθε χρονοσειράς. Έτσι το νευρωνικό δίκτυο μπορεί να αντιληφθεί το υποκείμενο μοτίβο χωρίς να εστιάζεται στην φυσική τιμή του μεγέθους υπό παρατήρηση. Έπειτα από την παραγωγή των προβλέψεων, εκείνες μετατρέπονται με ανάλογη διαδικασία στην κανονική τους κλίμακα για να αποκτήσουν και πάλι φυσικό νόημα.

Τέλος, αξίζει να σημειωθεί πως για την δημιουργία του συνόλου δεδομένων ελέγχου (test dataset) χρησιμοποιούμε ως παράθυρο εισόδου τις τελευταίες 18 παρατηρήσεις από κάθε μία από τις 23.000 χρονοσειρές στο σύνολο του M4 (σε περίπτωση που η χρονοσειρά έχει λιγότερες από 18 παρατηρήσει αντικαθιστούμε τις απουσιάζουσες τιμές με την παλαιότερη μη απουσιάζουσα τιμή ώστε να φτάσει να έχει μήκος 18) κι έτσι προκύπτει το σύνολο ελέγχου εισόδου. Το σύνολο ελέγχου εξόδου μας δίνεται έτοιμο από τους διοργανωτές του διαγωνισμού σε ξεχωριστό αρχείο. Με τον τρόπο αυτό ελέγχεται η ικανότητα γενίκευσής (generalization) του μοντέλου μας, αφού καλείται να προβλέψει δεδομένα με τα οποία δεν έχει έρθει ποτέ σε επαφή κατά την διαδικασία της εκπαίδευσης.

### 5.3.2 Οπτικοποίηση Δεδομένων με Βάση τα Χαρακτηριστικά

Στην υποενότητα αυτή εισάγουμε την βασική μέθοδο μετασχηματισμού και οπτικοποίησης των δεδομένων που αξιοποιούμε στην πειραματική μας διαδικασία.



Σχήμα 5.3: Γραφικές παραστάσεις των χαρακτηριστικών των συνόλων δεδομένων προβλεπόμενα σε δύο κύριες συνιστώσες μέσω της τεχνικής PCA. Πάνω αριστερά: Το σύνολο δεδομένων στόχου το οποίο επιθυμούμε να προβλέψουμε, αριθμεί 23.000 χρονικά παράθυρα των οποίων είναι γνωστές μόνοι οι εισόδοι. Πάνω δεξιά: Το σύνολο δεδομένων στόχου στο να επανεκπαιδύσουμε το μοντέλο μας, αριθμεί 235.460 χρονικά παράθυρα εισόδων και γνωστών εξόδων. Αποτελεί την επαυξημένη εκδοχή του συνόλου κάτω αριστερά. Κάτω αριστερά: Το σύνολο δεδομένων επανεκπαίδευσης στόχου, που προκύπτει εάν εκμεταλλευτούμε μόνο τα τελευταία χρονικά παράθυρα εισόδων-εξόδων, αριθμεί 14.242 δείγματα. Κάτω δεξιά: Το σύνολο δεδομένων πρόελευσης στο οποίο προεκπαιδύουμε το μοντέλο μας, αριθμεί 463.977 δείγματα.

Εκτός από τα σύνολα δεδομένων που περιέχουν την ποσοτική πληροφορία για την τιμή του μεγέθους κάθε χρονοσειράς, δηλαδή των χρονικών παρατηρήσεων, έχουμε στην διάθεσή μας ένα ακόμη σύνολο δεδομένων, το σύνολο δεδομένων των χαρακτηριστικών (features dataset). Κάθε εγγραφή στο σύνολο αυτό αφορά σε μετα-πληροφορία για τα καθένα από τα χρονικά παράθυρα εισόδου. Η μεταπληροφορία αυτή που εξάγεται για κάθε ένα από τα χρονικά παράθυρα εισόδου αφορά σε ένα σύνολο από οκτώ χαρακτηριστικά κάθε τέτοιου παραθύρου: επίπεδο τυχαιότητας (randomness level), επίπεδο τάσης (trend level), αυτοσυσχέτιση (autocorrelation), το λάμβδα του μετασχηματισμού Box-Cox (Box-Cox lambda), ασυμμετρία (skewness), κύρτωση (kurtosis), μη γραμμικότητα (nonlinearity), εκθέτης Hurst (Hurst exponent). Όπως θα φανεί στην συνέχεια, από αυτήν την πλούσια πηγή μεταπληροφορίας για τα χαρακτηριστικά κάθε χρονοσειράς θα μπορέσουμε να εξάγουμε πολύτιμα συμπεράσματα.

Με βάση τα χαρακτηριστικά αυτά του κάθε χρονικού παραθύρου καταφέρνουμε να προβάλουμε το σύνολο των οκτώ χαρακτηριστικών αυτών σε ένα δύο διαστάσεις χρησιμοποιώντας την τεχνική της ανάλυσης κύριων συνιστωσών (principal component analysis) ή αλλιώς PCA [62]. Η τεχνική αυτή προβάλλει έναν χώρο διάστασης  $m$  σε έναν άλλον διάστασης

$n$ , όπου  $n < m$ , διατηρώντας την μέγιστη δυνατή διακύμανση των δεδομένων. Οι κύριες συνιστώσες εξάγονται ως τα ιδιοδιανύσματα (eigenvectors) του πίνακα της συνδιακύμανσης (covariance matrix). Στην παρούσα εργασία λοιπόν, γίνεται χρήση αυτής της τεχνικής για να προβάλλουμε τα σύνολα δεδομένων των χαρακτηριστικών (features dataset) σε δύο διαστάσεις έτσι ώστε να αποκτήσουμε εποπτεία πάνω σε αυτά. Χαρακτηριστικά αναφέρουμε πως το ποσοστό της συνολικής διακύμανσης των δεδομένων που επεξηγείται από τις δύο κύριες συνιστώσες φτάνει, στην περίπτωση μας, στο 68,4%. Όπως φαίνεται από την εξίσωση του μετασχηματισμού, η PC1 επεξηγεί κυρίως τα χαρακτηριστικά της τυχαιότητας, της τάσης και του συντελεστή Hurst, η PC2 επεξηγεί κυρίως τα χαρακτηριστικά του λάμβδα Box-Cox και της μη γραμμικότητας ενώ τα υπόλοιπα χαρακτηριστικά, δηλαδή η αυτοσυσχέτιση, η ασυμμετρία και η κύρτωση εξηγούνται σε μεγάλο ποσοστό και από τις δύο κύριες συνιστώσες.

$$\begin{bmatrix} PC1 \\ PC2 \end{bmatrix} = \begin{bmatrix} 0.42 & -0.42 & -0.36 & -0.18 & 0.38 & 0.4 & 0.04 & -0.43 \\ 0.14 & -0.3 & -0.28 & 0.51 & -0.45 & -0.34 & 0.42 & -0.22 \end{bmatrix} \mathbf{F}$$

$$\mathbf{F} = \left[ \text{rand} \quad \text{trend} \quad AC \quad \text{Box} - \text{Cox} \quad \text{skew} \quad \text{kurt} \quad \text{nonlin} \quad \text{Hurst} \right]^T$$

όπου  $PC1, PC2$  οι δύο συνιστώσες του κάθε δείγματος και  $\mathbf{F}$ , το διάνυσμα των χαρακτηριστικών κάθε δείγματος.

Συμπερασματικά, μπορούμε να πούμε πως όσο μεταβαίνουμε προς τα δεξιά στο διδιάστατο επίπεδο τόσο αυξάνεται η τυχαιότητα ενώ φθίνει η τάση και ο συντελεστής Hurst. Αντίστοιχα όσο μεταβαίνουμε προς τα πάνω μέρος του επιπέδου αυξάνεται η μη γραμμικότητα και ο συντελεστής Box-Cox ενώ φθίνει η ασυμμετρία, η κύρτωση και η αυτοσυσχέτιση.

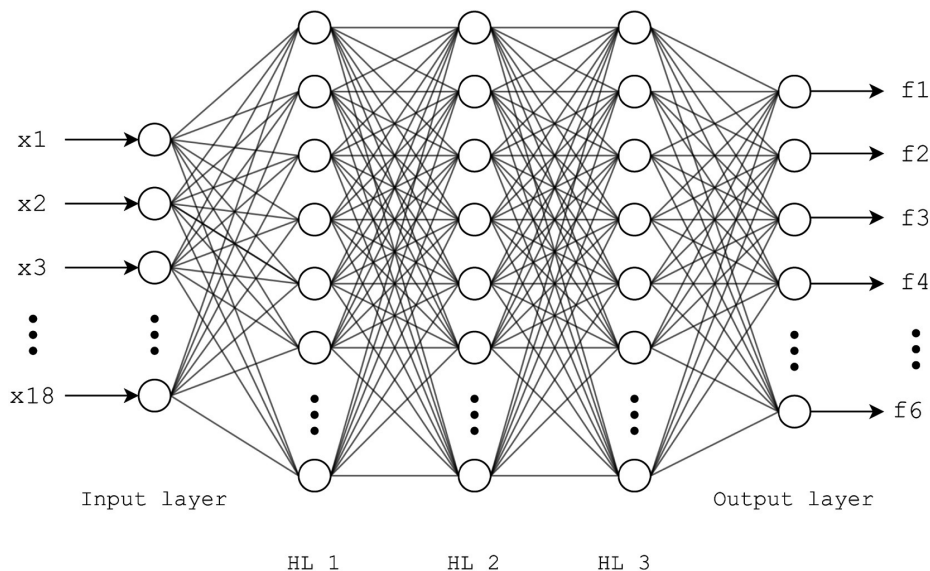
Στο Σχ. 5.3 βλέπουμε την οπτικοποίηση διαφορετικών συνόλων δεδομένων ενδιαφέροντος. Με βάση αυτό και τον πίνακα μετασχηματισμού σε δύο κύριες συνιστώσες αντιλαμβανόμαστε πως το σύνολο δεδομένων, με μεγάλο όγκο χρονοσειρών στο αριστερό ημιεπίπεδο, απαρτίζεται κατά κύριο λόγο από χρονοσειρές με μεγάλη τάση, χαμηλή τυχαιότητα και αυτοσυσχέτιση μεταξύ των παρατηρήσεων. Στην συνέχεια χρησιμοποιούμε τις δύο κυριότερες συνιστώσες για να διαχωρίσουμε το σύνολο δεδομένων σε υποσύνολα.

## 5.4 Μοντέλο Βάσης

Το βασικό μας μοντέλο αποτελείται ένα νευρωνικό δίκτυο MLP. Οι λόγοι για την επιλογή αυτής της αρχιτεκτονικής είναι πολλαπλοί: Από την μία πλευρά, η απλότητα των δικτύων MLP δίνει την δυνατότητα για γρήγορη κατάστρωση της πειραματικής διαδικασίας, μιας και αυτά αποδίδουν αρκετά καλά για ένα μεγάλο πλήθος εφαρμογών χωρίς να είναι απαραίτητη η χρονοβόρα αναζήτηση της βέλτιστης αρχιτεκτονικής δικτύου (network architecture search), διαδικασία η οποία είναι απαραίτητη σε πιο βαθιές και πολύπλοκες state-of-the-art αρχιτεκτονικές [63]. Ακόμη, η ύπαρξη ενός σχετικά μικρού πλήθους εκπαιδευσιμων βαρών σε ένα δίκτυο MLP μειώνει καταλυτικά το χρόνο εκπαίδευσης, επιτρέποντάς μας να δοκιμάσουμε ένα μεγαλύτερο εύρος πειραμάτων από αυτό που θα είμασταν σε θέση να διεξάγουμε με μια πιο

ακριβή υπολογιστικά δομή. Από την άλλη πλευρά, στα μειονεκτήματα αυτής της επιλογής συγκαταλέγεται το γεγονός πως τα δίκτυα αυτά αποτελούν απο την φύση τους εργαλεία γενικού σκοπού, δεν προσφέρονται συγκεκριμένα για την πρόβλεψη χρονοσειρών και ως εκ τούτου αναμένουμε η απόδοσή τους να είναι λιγότερο ικανοποιητική τα δίκτυα LSTM κι από άλλες πιο ελπιδοφόρες αρχιτεκτονικές όπως αυτές που αναφέραμε σε προηγούμενο υποκεφάλαιο.

Για την επιλογή των βασικών υπερ-παραμέτρων του δικτύου βασιστήκαμε στα συμπεράσματα της βιβλιογραφίας σχετικά με την βέλτιστη τοπολογία ενός MLP δικτύου για πρόβλεψη χρονοσειρών [11]. Με βάση τα συμπεράσματα αυτά, το δίκτυο μας αποτελείται από το επίπεδο εισόδου μήκους 18, το επίπεδο εξόδου μήκους 6, ενδιάμεσα από τα οποία παρεμβάλλονταν 3 κρυμμένα επίπεδα με 27 νευρώνες στο καθένα, δηλαδή 1,5 φορά το μήκος της εισόδου. Για συνάρτηση ενεργοποίησης επιλέξαμε την ReLU, ενώ για βελτιστοποιητή τον αλγόριθμο Adam με ρυθμό εκμάθησης 0,001 και μέγεθος παρτίδας 64. Τέλος, επιλέξαμε ως συνάρτηση απώλειας το μέσο απόλυτο σφάλμα MAE. Το παραπάνω μοντέλο συνιστά το μοντέλο πρό-έλευσης και αποτέλεσε αφετηρία για όλα μας τα πειράματα, εκπαιδευμένο στο ForeDeCk ή σε ένα υποσύνολο αυτού, όπως αναλύεται σε επόμενη ενότητα.



Σχήμα 5.4: Το MLP νευρωνικό δίκτυο που χρησιμοποιήσαμε: τρία κρυφά επίπεδα με 27 τεχνητούς νευρώνες το καθένα, μήκος επιπέδου εισόδου 18 και μήκος επιπέδου εξόδου 6 [11].

Ακόμη, αξίζει να αναφερθεί ότι για κάθε πείραμα χρησιμοποιήθηκε ένα πλήθος από 50 όμοια μοντέλα, ενώ ως αποτέλεσμα του πειράματος θεωρήθηκε η τελική πρόβλεψη του συνόλου των μοντέλων μας (ensemble model) που προέκυψε από τη διάμεσο (median) των προβλέψεων όλων των μοντέλων. Σημειώνεται πως η παραπάνω συνήθης πρακτική στον κλάδο των προβλέψεων έχει στόχο την απομάκρυνση της τυχαιότητας των αποτελεσμάτων και την μείωση του σφάλματος. Σε ότι αφορά την αξιολόγηση των αποτελεσμάτων του κάθε πειράματος, χρησιμοποιήθηκε το συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα sMAPE.

Επιπλέον, είναι σκόπιμο να αναφέρουμε εδώ, πως όταν αναφερόμαστε σε πλήρη επανεκπα-

ίδευση του δικτύου, γίνεται χρήση της τεχνικής του πρόωρου τερματισμού της εκπαίδευσης (early stopping), με περίοδο ανοχής (patience) τις δέκα εποχές.

Τέλος, για την εκτέλεση των πειραμάτων αναπτύχθηκε ο αντίστοιχος κώδικας στην γλώσσα προγραμματισμού Python κάνοντας χρήση βιβλιοθηκών ανοιχτού κώδικα (open source). Πιο συγκεκριμένα, για την επεξεργασία και τον μετασχηματισμό των δεδομένων έγινε χρήση των NumPy, Pandas και Scikit Learn [64–66]. Για την δημιουργία και εκπαίδευση των μοντέλων μας βασιστήκαμε στην βιβλιοθήκη Tensorflow και στο Keras API αυτής [67], ενώ για την οπτικοποίηση των δεδομένων κάναμε χρήση των πακέτων Matplotlib και Seaborn [68,69].

## 5.5 Benchmarks

Για να μετρήσουμε την επίδοση των τεχνικών μεταφοράς μάθησης που προτείνουμε, εισάγουμε εδώ δύο άλλες μεθόδους ως μέτρο σύγκρισης (benchmark). Οι μέθοδοι αυτές είναι οι εξής:

- Μέθοδος Theta (Theta Model):** Ο μηχανισμός λειτουργίας του μοντέλου Theta έχει αναλυθεί σε προηγούμενο κεφάλαιο. Εδώ αναφέρουμε συνοπτικά ότι η μέθοδος αυτή αποσυνθέτει την σειρά προς πρόβλεψη σε δύο ή περισσότερες γραμμές Theta, κάθε μία από τις οποίες εκμεταλλεύεται με διαφορετικό τρόπο την καμπυλότητα της χρονοσειράς για να προβλέψει την εξέλιξη της στο εγγές ή μακρινό μέλλον. Έπειτα οι γραμμές Theta συνενώνονται και παράγουν την πρόβλεψη της χρονοσειράς [13]. Ο λόγος για τον οποίον η μέθοδος αυτή επιλέγεται ως benchmark στην παρούσα εργασία είναι επειδή αποτελεί την καλύτερη αμιγώς στατιστική μέθοδο πρόβλεψης του διαγωνισμού M4 (2018) καθώς και την νικήτρια μέθοδο του διαγωνισμού M3 (2000), όταν και προτάθηκε για πρώτη φορά [70].
- Μέθοδος Smyl (Smyl Method):** Η μέθοδος αυτή αποτελεί μία υβριδική μέθοδο πρόβλεψης που συνδυάζει στοιχεία μηχανικής μάθησης και στατιστικής πρόβλεψης σε ένα ενιαίο ιεραρχικό πλαίσιο [5]. Η προσέγγιση αυτή συνδυάζει ένα μοντέλο εκθετικής εξομάλυνσης που εκπαιδεύεται τοπικά για κάθε χρονοσειρά χρησιμοποιώντας πληροφορία αποκλειστικά από την χρονοσειρά αυτή (local), καθώς κι ένα νευρωνικό δίκτυο LSTM εκπαιδευμένο στον σύνολο των χρονοσειρών συνολικά (global). Προτείνουμε την μέθοδο αυτή σαν benchmark στην παρούσα εργασία καθώς αποτέλεσε την νικήτρια μέθοδο του διαγωνισμού M4 [61].
- ForeDeCk:** Σε αυτή την μέθοδο εκπαιδεύουμε το μοντέλο βάσης, που προαναφέρθηκε, στο σύνολο του ForeDeCk και προβλέπουμε το σύνολο του M4 χωρίς καμία επανεκπαίδευση. Αξίζει να τονίσουμε εδώ πως κάτι τέτοιο είναι αρκετά σημαντικό για να διαπιστώσουμε την προβλεπτική ακρίβεια που έχει το μοντέλο μας την στιγμή που ξεκινάει να επανεκπαίδευεται. Αν η ακρίβεια αυτή πέσει κατά την επανεκπαίδευση, πράγμα που δεν είναι απίθανο, τότε η διαδικασία της επανεκπαίδευσης καθίσταται μη παραγωγική.

- **M4 All Windows:** Σε αυτή την μέθοδο εφαρμόζουμε την συνήθη προσέγγιση μηχανικής μάθησης. Έτσι, εκπαιδεύουμε το μοντέλο βάσης στο σύνολο δεδομένων εκπαίδευσης M4 All Windows και εν συνεχεία προβλέπουμε το σύνολο M4. Αυτή είναι η μέθοδος που θα εφαρμόζαμε εάν δεν επιθυμούσαμε να κάνουμε χρήση κάποιας τεχνικής μεταφοράς μάθησης.





## Κεφάλαιο 6

# Αποτελέσματα Πειραμάτων

### 6.1 Εισαγωγικές Παρατηρήσεις

Αναφορικά με τα πειράματα που υλοποιήθηκαν, τα ίδια είχαν ως σύνολο δεδομένων προέλευσης είτε το πλήρες ForeDeCk είτε την αριστερή τομή του, το Left SubForeDeCk. Επιλέξαμε το δεύτερο σύνολο καθώς θέλαμε να εκτελέσουμε τα πειράματα με ένα σύνολο προέλευσης που να είναι μεν αντιπροσωπευτικό του συνόλου στόχου, αλλά ταυτόχρονα να μην μοιράζεται μία τόσο κοντινή κατανομή πιθανότητας μαζί του. Δηλαδή, αφητηρία του κάθε πειράματος αποτέλεσε ένα προεκπαιδευμένο μοντέλο σε ένα από τα δύο σύνολα αυτά. Ως εκ τούτου, τα πειράματά μας εστιάζουν στο κομμάτι της μεταφοράς της αποκτηθείσας γνώσης στο πεδίο του M4, δηλαδή στο πως να προσαρμόσουμε το μοντέλο μας στο πεδίο στόχου κατά τρόπο βέλτιστο. Αντίστοιχα, ως σύνολο δεδομένων στόχου χρησιμοποιήθηκε τόσο το επαυξημένο σύνολο εκπαίδευσης M4 (all windows) όσο και το απλό (last windows). Τα αποτελέσματα που παρουσιάζονται, ωστόσο, αναφέρονται μόνο στο επαυξημένο σύνολο δεδομένων εκπαίδευσης, μιας κι εκείνο δίνει πάντα πιο ποιοτικά αποτελέσματα, όπως άλλωστε είναι αναμενόμενο.

Εκ πρώτης όψευς, κινηθήκαμε σε δύο βασικούς πειραματικούς άξονες: το επίπεδο της επανεκπαίδευσης, και το επίπεδο δεδομένων. Σε επίπεδο επανεκπαίδευσης προσπαθήσαμε να προσαρμόσουμε το μοντέλο μας στο M4, χρησιμοποιώντας όλο το πλήθος των δειγμάτων του και ρυθμίζοντας μία σειρά από υπερ-παραμέτρους. Από την άλλη μεριά, σε επίπεδο δεδομένων προσπαθήσαμε να προσαρμόσουμε το μοντέλο μας σε υποομάδες του M4, εξειδικεύοντας ένα διαφορετικό μοντέλο σε κάθε τέτοια υποομάδα δεδομένων. Η λεπτομερής περιγραφή των πειραμάτων ακολουθεί στις επόμενες ενότητες.

Είναι σημαντικό να αναφέρουμε πως σε κάθε ένα από τα πειράματα που εκτελέστηκαν έγινε μία ενδελεχής αναζήτηση του ρυθμού εκμάθησης του αλγορίθμου βελτιστοποίησης με σκοπό την εύρεση του συνδυασμού των υπερ-παραμέτρων που οδηγούν στο βέλτιστο αποτέλεσμα. Η αναζήτηση των υπερ-παραμέτρων που συνδυαστικά οδηγούν στην βέλτιστη λύση είναι γνωστή στη βιβλιογραφία ως εξερεύνηση πλέγματος (grid search) [71]. Οι τιμές του ρυθμού εκμάθησης που δοκιμάστηκαν κυμαίνονται σε ένα πολύ μεγάλο εύρος από  $10^{-2}$  έως  $10^{-8}$ . Ο

πρώτος αποτελεί έναν σχετικά μεγάλο ρυθμό εκμάθησης που επιφέρει, κατά συνέπεια, γρήγορη σύγκλιση αλλά και μεγάλες διακυμάνσεις κατά την εκμάθηση, ενώ ο δεύτερος αποτελεί έναν πολύ μικρό ρυθμό εκμάθησης που προκαλεί ανεπαίσθητες αλλαγές στα βάρη.

Στο κεφάλαιο αυτό θα περιγράψουμε τις πειραματικές διαδικασίες που προαναφέρθηκαν, μαζί με τα αποτελέσματα που λάβαμε γι' αυτές. Θα δούμε, αρχικά, τα πρώτα πειράματα με χρήση του ForeDeCk, ως περιβάλλοντος στόχου. Έπειτα, θα δούμε τα πειράματα του Left SubForeDeCk, χωρισμένα ως πειράματα σε επίπεδο εκπαίδευσης και πειράματα σε επίπεδο δεδομένων. Όλα μας τα αποτελέσματα συγκρίνονται με τα αποτελέσματα των benchmarks της παρούσας εργασίας, τα οποία φαίνονται στον Πίνακα 6.1

Learning Rate	sMAPE
Theta	13.593
Smyl Method	13.176
ForeDeCk	12.916
Left SubForeDeCk	14.026
M4 All Windows	13.098

Πίνακας 6.1: sMAPE των μεθόδων που τοποθετήθηκαν ως Benchmarks.

## 6.2 Πειράματα με χρήση του ForeDeCk

Στο πρώτο στάδιο των πειραμάτων μας εξερευνήσαμε το κατά πόσο η επανεκπαίδευση σε όλο το πλήθος των δεδομένων του M4 μπορεί να βελτιώσει την απόδοση του δικτύου, όταν αυτό έχει προεκπαιδευτεί με το ForeDeCk ως σύνολο δεδομένων περιβάλλοντος προέλευσης. Για τον λόγο αυτό διεξήχθησαν τα πειράματα που ακολουθούν, όπου δοκιμάστηκαν ποικίλες μέθοδοι επανεκπαίδευσης.

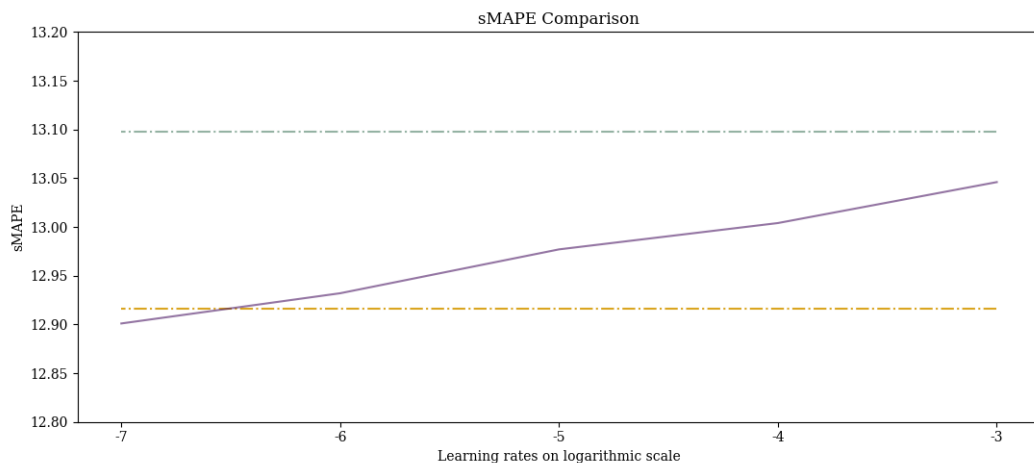
### 6.2.1 Ευνοϊκό Ξεκίνημα

Στο πείραμα αυτό εκτελούμε πλήρη επανεκπαίδευση του μοντέλου μας έχοντας ως αφετηρία τα ήδη προσαρμοσμένα βάρη του. Η τεχνική αυτή αναφέρεται συχνά ως warm start μιας και θυμίζει την παραδοσιακή μηχανική μάθηση με την μόνη διαφορά ότι τα βάρη δεν είναι τυχαία αρχικοποιημένα (cold start) αλλά προέρχονται από ένα μοντέλο που έχει προεκπαιδευτεί σε ένα συναφές σύνολο δεδομένων προέλευσης για να εκτελεί μία παρόμοια εργασία προέλευσης.

Στον Πίνακα 6.2 και το Σχήμα 6.1 παρουσιάζονται τα αποτελέσματα που λάβαμε για τη δοκιμή μιας πληθώρας ρυθμών εκμάθησης στο διάστημα μεταξύ  $10^{-3}$  και  $10^{-7}$ . Όπως παρατηρούμε, σε κάθε περίπτωση, η επανεκπαίδευση των μοντέλων με χρήση του M4 All Windows έχει επιφέρει μείωση του σφάλματος των προβλέψεων σε σύγκριση με την πρόβλεψη χωρίς

καμία επανεκπαίδευση (12.916) μόνο για ορισμένους ρυθμούς εκμάθησης και μόνο κατά 0.1% στην καλύτερη περίπτωση. Όμως, το σφάλμα είναι μικρότερο και από το αντίστοιχο για απλή εκπαίδευση των μοντέλων με χρήση του M4 All Windows, χωρίς εφαρμογή μεθόδων μεταφοράς μάθησης (13.098), κατά 1.5% στην καλύτερη περίπτωση. Επομένως, η συγκεκριμένη μέθοδος κρίνεται, ανά περίπτωση, επιτυχημένη.

Παράλληλα, βλέπουμε ένα μοτίβο μείωσης του sMAPE με μείωση του ρυθμού εκμάθησης. Η συγκεκριμένη συμπεριφορά μπορεί να εξηγηθεί αν αναλογιστούμε το εξαιρετικά μικρό σφάλμα που παίρνουμε με απλή εκπαίδευση των μοντέλων με το ForeDeCk και, συνεπώς, την πολύ καλή αρχικοποίηση των βαρών των μοντέλων, παράλληλα με την ιδιότητα των μεγαλύτερων ρυθμών εκμάθησης να προκαλούν μεγάλες μεταβολές στα βάρη. Είναι προφανές, πως δεν επιθυμούμε μεγάλες μεταβολές στα βάρη, αλλά μόνο κάποιες μικρές διορθώσεις με βάση τη γνώση που θα λάβουν τα μοντέλα από το περιβάλλον στόχου, προκειμένου να αυξηθεί η προβλεπτική τους ακρίβεια.



Σχήμα 6.1: Διάγραμμα όπου φαίνεται η εξέλιξη του δείκτη sMAPE για τις διαφορετικές τιμές του ρυθμού εκμάθησης, με χρήση λογαριθμικής κλίμακας για τον δεύτερο, συγκριτικά και με τα δύο βασικότερα benchmarks του πειράματος. Με μωβ γραμμή απεικονίζεται η εξέλιξη αυτή, με πράσινο βλέπουμε την τιμή του δείκτη για την περίπτωση εκπαίδευσης με το All Windows, ενώ η πορτοκαλί γραμμή απεικονίζει την τιμή του δείκτη στην περίπτωση απλής προεκπαίδευσης με το ForeDeCk.

Learning Rate	sMAPE
$10^{-3}$	13.046
$10^{-4}$	13.004
$10^{-5}$	12.977
$10^{-6}$	12.932
$10^{-7}$	<b>12.901</b>

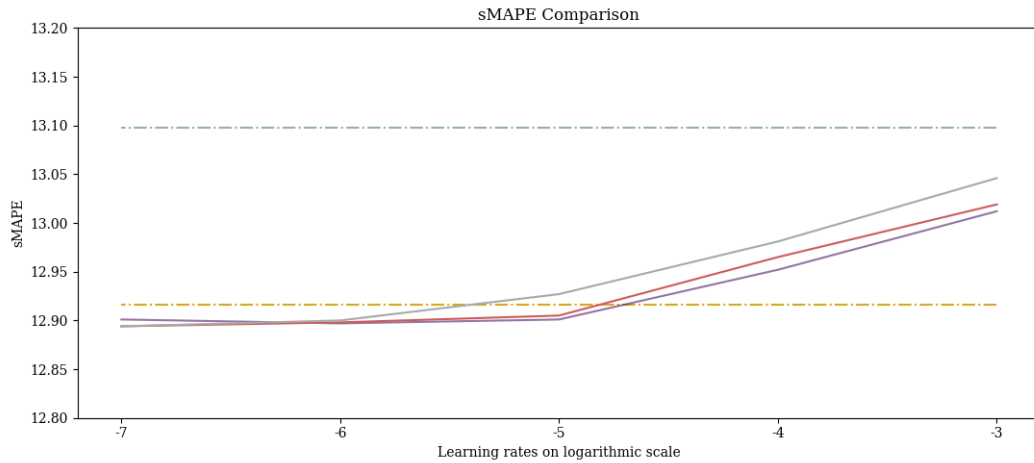
Πίνακας 6.2: sMAPE της μεθόδου του ευνοϊκού ξεκινήματος για διαφορετικούς ρυθμούς εκμάθησης με χρήση του ForeDeCk ως περιβάλλοντος προέλευσης.

### 6.2.2 Φραγμένες Εποχές

Στο πείραμα αυτό επανεκπαιδύουμε το μοντέλο μας για μερικές μόνο εποχές. Θεωρώντας πως το μοντέλο έχει εκπαιδευτεί σε ένα όμοιο σύνολο δεδομένων προέλευσης, διατυπώνουμε την υπόθεση ότι ένας μικρός αριθμός εποχών επανεκπαίδευσης στο σύνολο στόχου θα αυξήσει την προβλεπτική του ικανότητα. Για αυτό εξετάζουμε την επανεκπαίδευση του για 1, 2 και 5 εποχές, σε αντιδιαστολή με την πλήρη εκπαίδευση του προηγούμενου πειράματος.

Στον Πίνακα 6.3 και το Σχήμα 6.2 παρουσιάζονται τα αποτελέσματα που λάβαμε για τη δοκιμή μιας πληθώρας ρυθμών εκμάθησης στο διάστημα μεταξύ  $10^{-3}$  και  $10^{-7}$ , αλλά και για περιορισμένο πλήθος εποχών επανεκπαίδευσης, με 1, 2 ή 5 εποχές. Όπως παρατηρούμε, η επανεκπαίδευση των μοντέλων με χρήση του M4 All Windows έχει επιφέρει μείωση του σφάλματος των προβλέψεων σε σύγκριση με την πρόβλεψη χωρίς καμία επανεκπαίδευση (12.916) μόνο σε ορισμένες περιπτώσεις και μόνο κατά 0.1% στην καλύτερη περίπτωση, όπως ακριβώς συνέβη και με την προηγούμενη μέθοδο. Ωστόσο, το σφάλμα είναι σε όλες τις περιπτώσεις μικρότερο από το αντίστοιχο για απλή εκπαίδευση των μοντέλων με χρήση του M4 All Windows, χωρίς εφαρμογή μεθόδων μεταφοράς μάθησης (13.098), φτάνοντας μια βελτίωση κατά 1.5% στην καλύτερη περίπτωση. Ως εκ τούτου, τα αποτελέσματα της μεθόδου την καθιστούν κατάλληλη μόνο υπό περιπτώσεις.

Διακρίνουμε και πάλι το ίδιο μοτίβο μείωσης του sMAPE με μείωση του ρυθμού εκμάθησης, ενώ όσο περισσότερες οι εποχές επανεκπαίδευσης τόσο μεγαλύτερη η διαφορά που δημιουργείται με μείωση του ρυθμού εκμάθησης. Παράλληλα, τα αποτελέσματα για τα διαφορετικά πλήθη εποχών είναι συγκρίσιμα, με την αύξηση του πλήθους των εποχών να δουλεύει χειρότερα για μεγαλύτερους ρυθμούς εκμάθησης και καλύτερα για μικρότερους. Το αποτέλεσμα εξηγείται σε μεγάλο βαθμό χάρη στις εντονότερες μεταβολές που προκαλούνται ανά εποχή με μεγαλύτερο ρυθμό εκμάθησης και την πολύ καλή αρχικοποίηση που δίνει στα βάρη η προεκπαίδευση με το περιβάλλον προέλευσης. Προφανώς, σε περίπτωση μεγάλων μεταβολών των βαρών, επιθυμούμε λιγότερες εποχές για ένα καλό fine-tuning, ενώ σε αντίθετη περίπτωση τα μοντέλα χρειάζονται περισσότερες εποχές επανεκπαίδευσης για να οδηγηθούν σε μεγαλύτερη προβλεπτική ακρίβεια, μαθαίνοντας περισσότερα από το περιβάλλον στόχου.



Σχήμα 6.2: Διάγραμμα όπου φαίνεται η εξέλιξη του δείκτη sMAPE για τις διαφορετικές τιμές του ρυθμού εκμάθησης, με χρήση λογαριθμικής κλίμακας για τον δεύτερο, συγκριτικά και με τα δύο βασικότερα benchmarks του πειράματος. Με μωβ γραμμή απεικονίζεται η εξέλιξη αυτή για μια εποχή επανεκπαίδευσης, με κόκκινο η εξέλιξη για δύο εποχές, με γκρι για πέντε εποχές, με πράσινο βλέπουμε την τιμή του δείκτη για την περίπτωση εκπαίδευσης με το All Windows, ενώ η πορτοκαλί γραμμή απεικονίζει την τιμή του δείκτη στην περίπτωση απλής προεκπαίδευσης με το ForeDeCk.

Learning Rate	1 Epoch	2 Epochs	5 Epochs
$10^{-3}$	13.012	13.019	13.046
$10^{-4}$	12.952	12.965	12.981
$10^{-5}$	12.901	12.905	12.927
$10^{-6}$	12.897	12.898	12.9
$10^{-7}$	12.901	<b>12.894</b>	<b>12.894</b>

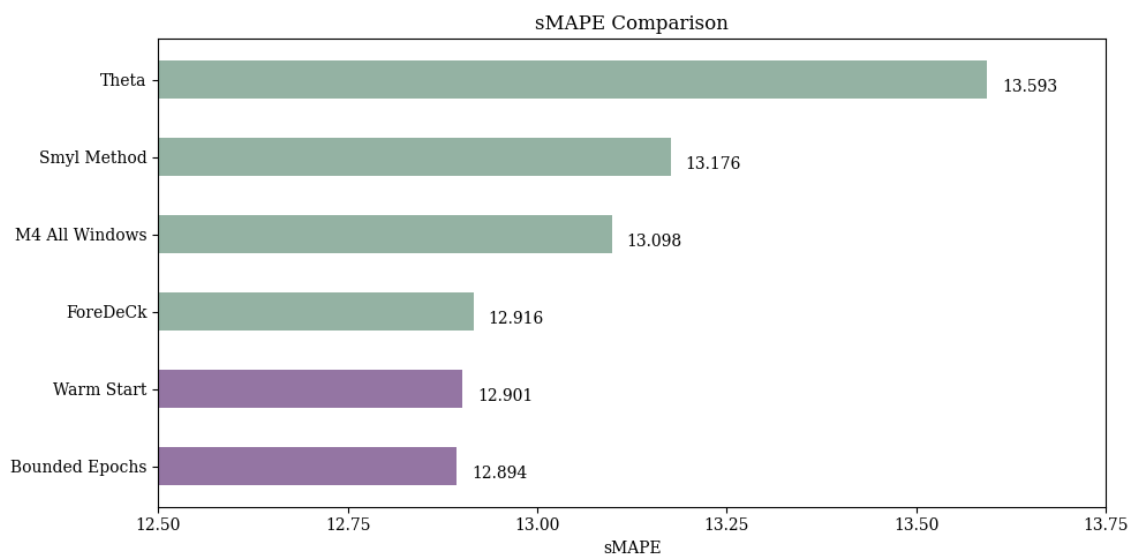
Πίνακας 6.3: sMAPE της μεθόδου των φραγμένων εποχών για διαφορετικούς ρυθμούς εκμάθησης και πλήθος εποχών εκπαίδευσης με χρήση του ForeDeCk ως περιβάλλοντος προέλευσης.

### 6.2.3 Συγκέντρωση Αποτελεσμάτων

Στο Σχήμα 6.3 παρατηρούμε μια συγκεντρωτική καταγραφή των καλύτερων αποτελεσμάτων που απέδωσε κάθε μια από τις μεθόδους που δοκιμάσαμε, μαζί με τα αποτελέσματα που είχαμε από το benchmarking αυτών των μεθόδων. Είναι εμφανές, πως και οι δύο μέθοδοι μεταφοράς μάθησης που δοκιμάσαμε έχουν καταφέρει να πετύχουν μικρότερο σφάλμα πρόβλεψης από τα δύο προσφερόμενα benchmarks, με μικρές διαφορές απόδοσης να παρατηρούνται μεταξύ τους. Το καλύτερο αποτέλεσμα έρχεται από τη μέθοδο των φραγμένων εποχών και, συγκεκριμένα, με επανεκπαίδευση για 2 ή 5 εποχές.

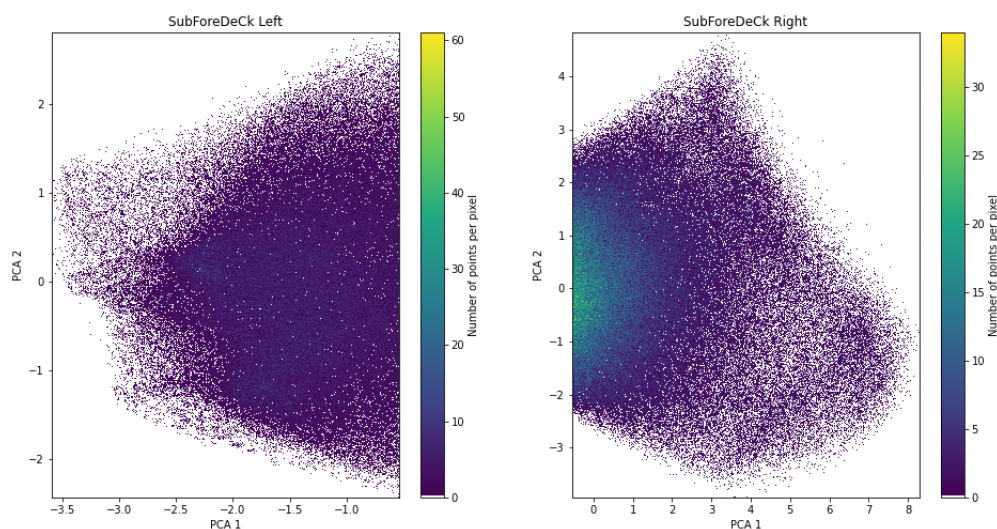
Είναι, φυσικά, άξια λόγου η πολύ μικρή διαφορά του σφάλματος χωρίς και με επανεκπαίδευση, ανεξαρτήτως της μεθόδου που θα επιλέξουμε, με μια ποσοστιαία βελτίωση κοντά στο

0.1% για κάθε μέθοδο. Η ίδια οφείλεται σε μεγάλο βαθμό στο γεγονός πως το ForeDeCk είναι ένα πολύ αντιπροσωπευτικό σύνολο δεδομένων για το M4.



Σχήμα 6.3: Ραβδόγραμμα των δύο μεθόδων μεταφοράς μάθησης (μωβ μπάρες) με τα τέσσερα benchmarks που χρησιμοποιήθηκαν γι' αυτές τις μεθόδους (πράσινες μπάρες), με χρήση του ForeDeCk ως περιβάλλοντος πρόελευσης.

### 6.3 Δημιουργία του Συνόλου Δεδομένων Left SubForeDeCk



Σχήμα 6.4: Γραφικές παραστάσεις των χαρακτηριστικών των συνόλων δεδομένων προβλεβημένα σε δύο κύριες συνιστώσες μέσω της τεχνικής PCA. Αριστερά: Left SubForeDeck, Η πάνω αριστερή του ForeDeCk, αριθμεί 231.989 δείγματα. Δεξιά: Right SubForeDeck, Η δεξιά τομή του ForeDeCk, αριθμεί 231.988 δείγματα.

Στα πειράματα που τοποθετήθηκαν και σχολιάστηκαν προηγουμένως το ForeDeCk χρησιμοποιήθηκε σαν σύνολο δεδομένων προέλευσης. Όπως ήδη αναφέρθηκε, το ίδιο συνιστά ένα σύνολο που σε μεγάλο βαθμό καλύπτει πλήρως το χώρο των χαρακτηριστικών, έχοντας ως αποτέλεσμα ένα μικρό περιθώριο βελτίωσης της απόδοσης του μοντέλου με χρήση μεθόδων μεταφοράς μάθησης. Εμείς, ωστόσο, θέλοντας να διερευνήσουμε το πόσο καλά μπορεί να αποδώσει η λογική της μεταφοράς μάθησης, αποφασίσαμε να παράξουμε ένα σύνολο δεδομένων προέλευσης με λιγότερα δείγματα και με πιο περιορισμένη κάλυψη του χώρου των χαρακτηριστικών. Μια τέτοια κίνηση θα έδινε μεγαλύτερα περιθώρια βελτίωσης της απόδοσης, σε περίπτωση που οι μέθοδοι ήταν πράγματι πετυχημένες, ενώ παράλληλα θα επέτρεπε την σύγκριση της απόδοσης των διαφορετικών μεθόδων. Άλλωστε, η ύπαρξη ενός συνόλου δεδομένων προέλευσης τόσο αντιπροσωπευτικού σε σχέση με το σύνολο δεδομένων στόχου είναι ένα αρκετά σπάνιο ενδεχόμενο κι εμείς επιθυμούσαμε να δουλέψουμε σε ένα πιο ρεαλιστικό σενάριο για να παράξουμε συμπεράσματα που θα είναι χρήσιμα στην γενική περίπτωση. Δημιουργήσαμε, λοιπόν, τα ακόλουθα δύο υποσύνολα του ForeDeCk, στα οποία θα αναφερόμαστε ως τομές του ForeDeCk.

**Left/Right SubForeDeCk:** Αυτά τα δύο σύνολα δεδομένων δημιουργούνται από το σύνολο χαρακτηριστικών των δεδομένων (features dataset) του ForeDeCk προβεβλημένο στις δύο κύριες διαστάσεις με βάση την τεχνική PCA. Ακολούθως, βρίσκεται η διάμεσος κατά μήκος της πρώτης κύριας συνιστώσας PCA1. Έτσι, τα δείγματα χρονικών παραθύρων που βρίσκονται αριστερά της διαμέσου συγκαταλέγονται στο Left SubForeDeck και αντίστοιχα για το Right ForeDeCk. Αξίζει να σημειώσουμε πως τα δύο σύνολα δεδομένων είναι σχεδόν ισοπληθή: Το Left SubForeDeck αριθμεί 231.989 δείγματα ενώ το Right SubForeDeck 231.988. Από τα δύο παραπάνω σύνολα επιλέξαμε το Left SubForeDeck ως πλέον κατάλληλο για να διεξάγουμε τα λοιπά πειράματά μας.

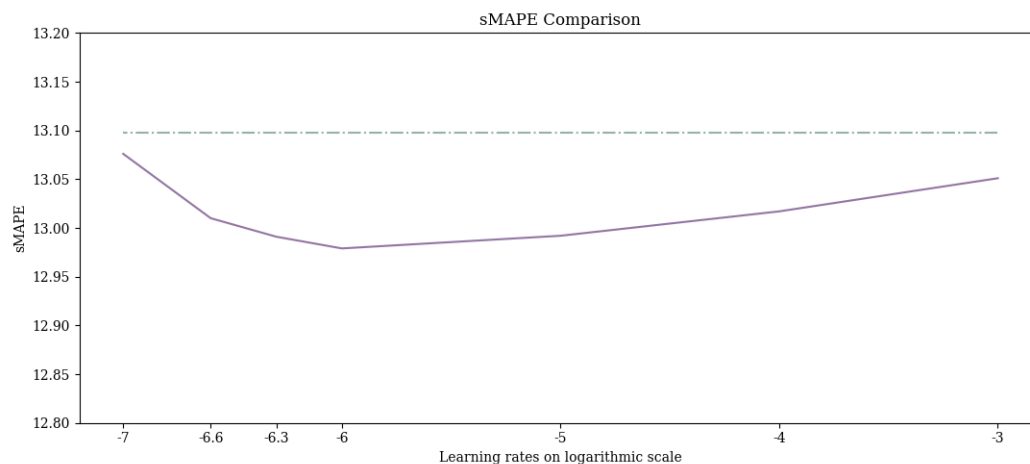
## 6.4 Πειράματα σε επίπεδο εκπαίδευσης με χρήση του Left SubForeDeCk

### 6.4.1 Ευνοϊκό Ξεκίνημα

Στον Πίνακα 6.4 και το Σχήμα 6.5 παρουσιάζονται τα αποτελέσματα που λάβαμε για τη δοκιμή μιας πληθώρας ρυθμών εκμάθησης στο διάστημα μεταξύ  $10^{-3}$  και  $10^{-7}$ . Όπως παρατηρούμε, σε κάθε περίπτωση, η επανεκπαίδευση των μοντέλων με χρήση του M4 All Windows έχει επιφέρει μείωση του σφάλματος των προβλέψεων σε σύγκριση με την πρόβλεψη χωρίς καμία επανεκπαίδευση (14.026). Μάλιστα, στην καλύτερη περίπτωση έχουμε μια βελτίωση κατά 7.5%. Επίσης, το σφάλμα είναι μικρότερο και από το αντίστοιχο για απλή εκπαίδευση των μοντέλων με χρήση του M4 All Windows, χωρίς εφαρμογή μεθόδων μεταφοράς μάθησης (13.098), μέχρι και κατά 0.9% στην καλύτερη περίπτωση. Επομένως, η συγκεκριμένη μέθοδος κρίνεται επιτυχημένη.

Παράλληλα, βλέπουμε ένα μοτίβο μείωσης του sMAPE με μείωση του ρυθμού εκμάθη-

σης μέχρι μια τιμή κι έπειτα αύξηση του στα αρχικά επίπεδα. Η συγκεκριμένη συμπεριφορά μπορεί να εξηγηθεί αν αναλογιστούμε πως ένας μικρότερος ρυθμός εκμάθησης οδηγεί και σε μικρότερες μεταβολές των βαρών των μοντέλων και, ως εκ τούτου, σε μικρές βελτιώσεις τους (fine-tuning). Ωστόσο, μια πολύ μικρή τιμή του ρυθμού εκμάθησης μπορεί να οδηγήσει σε ελαχιστοποίηση της βελτίωσης, αφού οι μεταβολές των βαρών δεν είναι επαρκείς ώστε να προκαλέσουν την καλύτερη προβλεπτική ακρίβεια των μοντέλων.



Σχήμα 6.5: Διάγραμμα όπου φαίνεται η εξέλιξη του δείκτη sMAPE για τις διαφορετικές τιμές του ρυθμού εκμάθησης, με χρήση λογαριθμικής κλίμακας για τον δεύτερο, συγκριτικά και με το βασικότερο benchmark του πειράματος. Με μωβ γραμμή απεικονίζεται η εξέλιξη αυτή, ενώ με πράσινο βλέπουμε την τιμή του δείκτη για την περίπτωση εκπαίδευσης με το All Windows.

Learning Rate	sMAPE
$10^{-3}$	13.051
$10^{-4}$	13.017
$10^{-5}$	12.992
$10^{-6}$	<b>12.979</b>
$5 \cdot 10^{-7}$	12.991
$2.5 \cdot 10^{-7}$	13.010
$10^{-7}$	13.076

Πίνακας 6.4: sMAPE της μεθόδου του ευνοϊκού ξεκινήματος για διαφορετικούς ρυθμούς εκμάθησης με χρήση του Left SubForeDeCk ως περιβάλλοντος προέλευσης

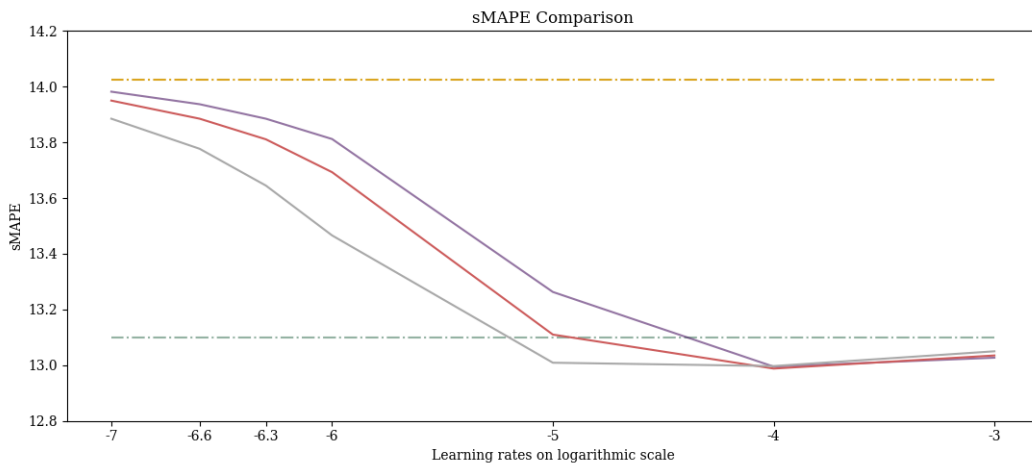
#### 6.4.2 Φραγμένες Εποχές

Στον Πίνακα 6.5 και στο Σχήμα 6.6 παρουσιάζονται τα αποτελέσματα που λάβαμε για τη δοκιμή μιας πληθώρας ρυθμών εκμάθησης στο διάστημα μεταξύ  $10^{-3}$  και  $10^{-7}$ , αλλά και για περιορισμένο πλήθος εποχών επανεκπαίδευσης, με 1, 2 ή 5 εποχές. Όπως παρατηρούμε, σε κάθε περίπτωση, η επανεκπαίδευση των μοντέλων με χρήση του M4 All Windows έχει



επιφέρει μείωση του σφάλματος των προβλέψεων σε σύγκριση με την πρόβλεψη χωρίς καμία επανεκπαίδευση (14.026) έως και κατά 7.4%, όπως συνέβη και με την προηγούμενη μέθοδο. Ωστόσο, το σφάλμα δεν είναι σε όλες τις περιπτώσεις μικρότερο από το αντίστοιχο για απλή εκπαίδευση των μοντέλων με χρήση του M4 All Windows, χωρίς εφαρμογή μεθόδων μεταφοράς μάθησης (13.098) και, ως εκ τούτου, τα αποτελέσματα της μεθόδου την καθιστούν κατάλληλη μόνο υπό περιπτώσεις.

Διακρίνουμε και πάλι το ίδιο μοτίβο αρχικής μείωσης και στη συνέχεια αύξησης του sMAPE με μείωση του ρυθμού εκμάθησης, ωστόσο σε αυτή τη μέθοδο το ελάχιστο έρχεται για ρυθμό εκμάθησης μεγαλύτερο σε σχέση με προηγουμένως. Παράλληλα, τα αποτελέσματα για τα διαφορετικά πλήθη εποχών είναι συγκρίσιμα, με την αύξηση του πλήθους των εποχών να δουλεύει χειρότερα για μεγαλύτερους ρυθμούς εκμάθησης και καλύτερα για μικρότερους. Το αποτέλεσμα εξηγείται σε μεγάλο βαθμό χάρη στις εντονότερες μεταβολές που προκαλούνται ανά εποχή με μεγαλύτερο ρυθμό εκμάθησης και την αρκετά καλή αρχικοποίηση που δίνει στα βάρη η προεκπαίδευση με το περιβάλλον προέλευσης. Προφανώς, σε περίπτωση μεγάλων μεταβολών των βαρών, επιθυμούμε λιγότερες εποχές για ένα καλό fine-tuning, ενώ σε αντίθετη περίπτωση τα μοντέλα χρειάζονται περισσότερες εποχές επανεκπαίδευσης για να οδηγηθούν σε μεγαλύτερη προβλεπτική ακρίβεια, μαθαίνοντας περισσότερα από το περιβάλλον στόχου.



Σχήμα 6.6: Διάγραμμα όπου φαίνεται η εξέλιξη του δείκτη sMAPE για τις διαφορετικές τιμές του ρυθμού εκμάθησης, με χρήση λογαριθμικής κλίμακας για τον δεύτερο, συγκριτικά και με τα δύο βασικότερα benchmarks του πειράματος. Με μωβ γραμμή απεικονίζεται η εξέλιξη αυτή για μια εποχή επανεκπαίδευσης, με κόκκινη η εξέλιξη για δύο εποχές, με γκρι για πέντε εποχές, με πράσινο βλέπουμε την τιμή του δείκτη για την περίπτωση εκπαίδευσης με το All Windows, ενώ η πορτοκαλί γραμμή απεικονίζει την τιμή του δείκτη στην περίπτωση απλής προεκπαίδευσης με το Left SubForeDeCk.

Learning Rate	1 Epoch	2 Epochs	5 Epochs
$10^{-3}$	13.027	13.035	13.050
$10^{-4}$	12.995	<b>12.988</b>	12.997
$10^{-5}$	13.263	13.110	13.009
$10^{-6}$	13.812	13.693	13.466
$5 \cdot 10^{-7}$	13.885	13.811	13.645
$2.5 \cdot 10^{-7}$	13.937	13.885	13.777
$10^{-7}$	13.982	13.950	13.885

Πίνακας 6.5: sMAPE της μεθόδου των φραγμένων εποχών για διαφορετικούς ρυθμούς εκμάθησης και πλήθος εποχών εκπαίδευσης με χρήση του Left SubForeDeCk ως περιβάλλοντος προέλευσης

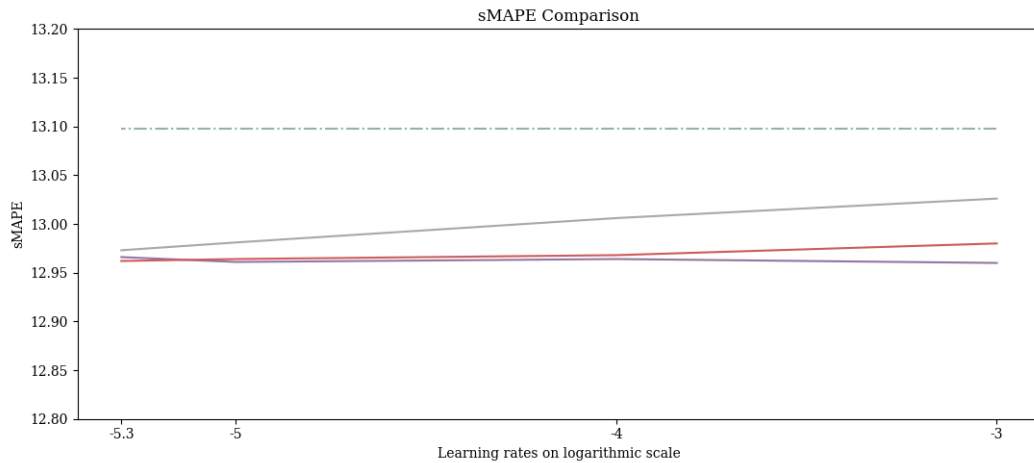
### 6.4.3 Πάγωμα Επιπέδων

Στο πείραμα αυτό επιλέγουμε να εκπαιδύσουμε μερικώς το δίκτυό μας, πακτώνοντας κάποια από τα βάρη του στις αρχικές τους τιμές πριν την επανεκπαίδευση, μία τεχνική γνωστή ως “πάγωμα” επιπέδων του δικτύου (layers freezing). Αρχικά “παγώνουμε” όλα τα βάρη του δικτύου έτσι ώστε να μην επιδέχονται αλλαγές από τον αλγόριθμο βελτιστοποίησης. Στην συνέχεια, ξεκινώντας από το επίπεδο εξόδου “ξεπαγώνουμε” προοδευτικά όλα τα επίπεδα μέχρι και το πρώτο κρυμμένο επίπεδο, και κάθε ένα από τα μοντέλα που προκύπτουν, τα οποία έχουν διαφορετικό πλήθος εκπαιδευσιμων βαρών υποβάλλεται σε πλήρη επανεκπαίδευση.

Στον Πίνακα 6.6 και στο Σχήμα 6.7 παρουσιάζονται τα αποτελέσματα που λάβαμε για τη δοκιμή μιας πληθώρας ρυθμών εκμάθησης στο διάστημα μεταξύ  $10^{-3}$  και  $5 \cdot 10^{-6}$ , αλλά και για περιορισμένο πλήθος επιπέδων προς εκπαίδευση, με τον περιορισμό να επιτρέπει την επανεκπαίδευση 1, 2 ή 3 επιπέδων. Όπως παρατηρούμε, σε κάθε περίπτωση, η επανεκπαίδευση των μοντέλων με χρήση του M4 All Windows έχει επιφέρει μείωση του σφάλματος των προβλέψεων σε σύγκριση με την πρόβλεψη χωρίς καμία επανεκπαίδευση (14.026) έως και κατά 7.6%, όπως συνέβη και με τις προηγούμενες μεθόδους. Το ίδιο παρατηρείται και συγκριτικά με το σφάλμα για απλή εκπαίδευση των μοντέλων με χρήση του M4 All Windows, χωρίς εφαρμογή μεθόδων μεταφοράς μάθησης (13.098) και, μάλιστα, τα αποτελέσματα της μεθόδου είναι τα καλύτερα που έχουμε λάβει ως τώρα με τις μεθόδους που δοκιμάστηκαν, αφού η ποσοστιαία βελτίωση αγγίζει το 1%. Πρόκειται, λοιπόν, για μια εξαιρετικά επιτυχημένη στρατηγική.

Σε αντίθεση με τις προηγούμενες μεθόδους, δεν παρατηρείται κάποιο μοτίβο αναφορικά με το ρυθμό εκμάθησης και το προκύπτον sMAPE. Βλέπουμε πως διαφορετικοί ρυθμοί εκμάθησης προκαλούν ανεπαίσθητες διαφορές, οι οποίες καθίστανται λίγο περισσότερο εμφανείς με την αύξηση του πλήθους των επιπέδων προς επανεκπαίδευση. Αναφορικά με τα τελικά αποτελέσματα, φαίνεται πως η επιλογή της επανεκπαίδευσης μόνο του επιπέδου εξόδου δίνει και το μικρότερο σφάλμα, ωστόσο με μικρή διαφορά από τις άλλες δύο επιλογές. Το αποτέλεσμα

εξηγείται με βάση το γεγονός πως τα πρώτα επίπεδα τείνουν να αποτυπώνουν γενικά τα χαρακτηριστικά του συνόλου εκπαίδευσης, σε αντίθεση με τα τελικά που επιφορτίζονται με την παραγωγή των προβλέψεων, αλλά και την καλή αρχικοποίηση που δίνει το περιβάλλον προέλευσης. Συγκεκριμένα, καθώς το τελευταίο είναι αρκετά αντιπροσωπευτικό για το περιβάλλον στόχου, η αποτύπωση των χαρακτηριστικών του, όπως αυτή έχει γίνει στα πρώτα επίπεδα, κρίνεται αρκετά επιτυχημένη και, ως εκ τούτου, δεν χρειάζεται να γίνει μεταβολή της. Μια μικρή μεταβολή των βαρών του επιπέδου εξόδου είναι αρκετή, προκειμένου να βελτιστοποιηθεί η προβλεπτική ακρίβεια των μοντέλων.



Σχήμα 6.7: Διάγραμμα όπου φαίνεται η εξέλιξη του δείκτη sMAPE για τις διαφορετικές τιμές του ρυθμού εκμάθησης, με χρήση λογαριθμικής κλίμακας για τον δεύτερο, συγκριτικά και με τα δύο βασικότερα benchmarks του πειράματος. Με μωβ γραμμή απεικονίζεται η εξέλιξη αυτή για ένα επίπεδο προς επανεκπαίδευση, με κόκκινο η εξέλιξη για δύο επίπεδα, με γκρι για τρία επίπεδα, ενώ με πράσινο βλέπουμε την τιμή του δείκτη για την περίπτωση εκπαίδευσης με το All Windows.

Learning Rate	1 Trainable Layer	2 Trainable Layers	3 Trainable Layers
$10^{-3}$	<b>12.960</b>	12.980	13.026
$10^{-4}$	12.964	12.968	13.006
$10^{-5}$	12.961	12.964	12.981
$5 \cdot 10^{-6}$	12.966	12.962	12.973

Πίνακας 6.6: sMAPE της μεθόδου του παγώματος επιπέδων για διαφορετικό πλήθος επιπέδων που επιδέχονται εκπαίδευση με χρήση του Left SubForeDeCk ως περιβάλλοντος προέλευσης

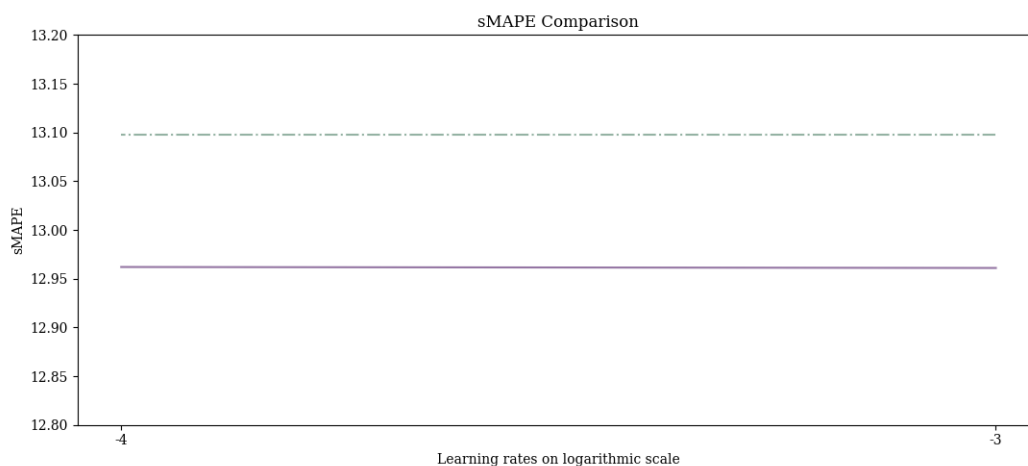
#### 6.4.4 Επαναρχικοποίηση του Επιπέδου Εξόδου

Η συγκεκριμένη μέθοδος δοκιμάστηκε συμπληρωματικά της προηγούμενης και σε μεγάλο βαθμό αντιστοιχεί στην επανεκπαίδευση ενός και μόνο επιπέδου, με πάγωμα των υπολοίπων. Αυτό που την χαρακτηρίζει, είναι η τοποθέτηση ενός νέου επιπέδου εξόδου στην θέση του προηγούμενου εκπαιδευμένου -η οποία αντιστοιχεί σε επαναρχικοποίηση του- και η επανεκπα-

ίδευση μόνο αυτού.

Στον Πίνακα 6.7 και το Σχήμα 6.8 βλέπουμε τα αποτελέσματα για τη δοκιμή δύο και μόνο ρυθμών εκμάθησης,  $10^{-3}$  και  $10^{-4}$ . Η ουσιαστική της διαφοροποίηση σε σχέση με την προαναφερθείσα μέθοδο είναι πως το επίπεδο εξόδου αλλάζει και, επομένως, δεν διατηρεί καμία προηγούμενη γνώση από το περιβάλλον προέλευσης, αλλά αρχικοποιείται με τυχαίο τρόπο. Όλη η αποκτηθείσα γνώση συγκεντρώνεται, λοιπόν, στα υπόλοιπα επίπεδα.

Τα αποτελέσματα είναι αρκετά αναμενόμενα, αφού ταυτίζονται σε μεγάλο βαθμό με τις τιμές που είδαμε για την επανεκπαίδευση μόνο του επιπέδου εξόδου, στα πλαίσια της προηγούμενης μεθόδου. Η διαφορετική αρχικοποίηση δεν φαίνεται να επηρέασε την τελική προβλεπτική ακρίβεια, γεγονός που εξηγείται, αφού αφήσαμε τα μοντέλα να εκπαιδευτούν για ένα μεγάλο πλήθος εποχών και, άρα, συνέκλιναν στο ίδιο αποτέλεσμα.



Σχήμα 6.8: Διάγραμμα όπου φαίνεται η εξέλιξη του δείκτη sMAPE για τις διαφορετικές τιμές του ρυθμού εκμάθησης, με χρήση λογαριθμικής κλίμακας για τον δεύτερο, συγκριτικά και με το βασικότερο benchmark του πειράματος. Με μωβ γραμμή απεικονίζεται η εξέλιξη αυτή, ενώ με πράσινο βλέπουμε την τιμή του δείκτη για την περίπτωση εκπαίδευσης με το All Windows.

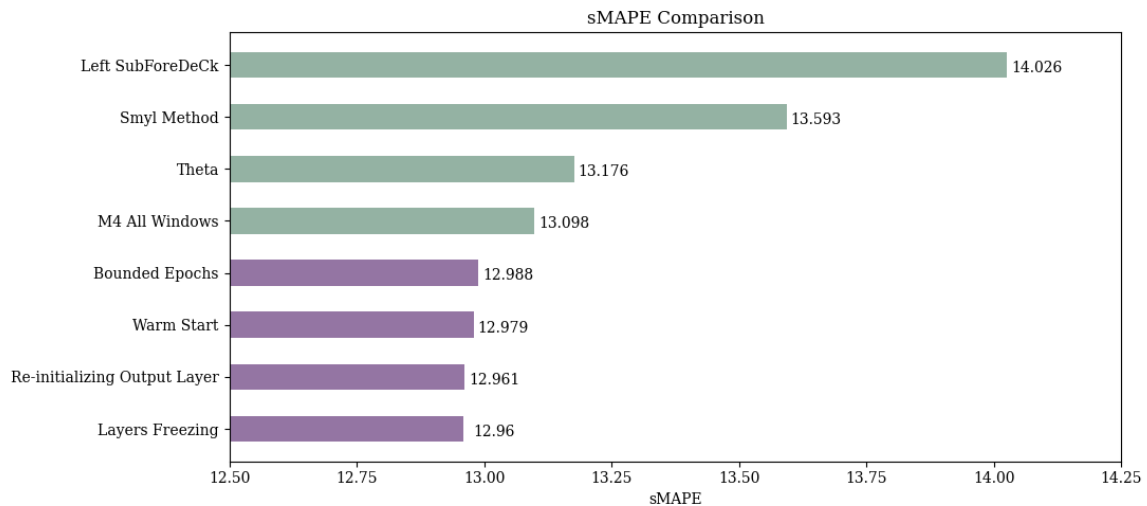
Learning Rate	sMAPE
$10^{-3}$	<b>12.961</b>
$10^{-4}$	12.962

Πίνακας 6.7: sMAPE της μεθόδου αλλαγής του επιπέδου εξόδου για δύο διαφορετικές τιμές του ρυθμού εκμάθησης με χρήση του Left SubForeDeCk ως περιβάλλοντος προέλευσης

#### 6.4.5 Συγκέντρωση των αποτελεσμάτων

Στο Σχήμα 6.9 παρατηρούμε μια συγκεντρωτική καταγραφή των καλύτερων αποτελεσμάτων που απέδωσε κάθε μια από τις μεθόδους που δοκιμάσαμε, μαζί με τα αποτελέσματα

που είχαμε από το benchmarking αυτών των μεθόδων. Είναι εμφανές, πως όλες οι μέθοδοι μεταφοράς γνώσης που δοκιμάσαμε έχουν καταφέρει να πετύχουν μικρότερο σφάλμα πρόβλεψης από τα δύο προσφερόμενα benchmarks, με μικρές διαφορές απόδοσης να παρατηρούνται μεταξύ τους. Το καλύτερο αποτέλεσμα έρχεται από τη μέθοδο παγώματος επιπέδων με επιλογή επανεκπαίδευσης μόνο του επιπέδου εξόδου ή αντικατάστασης του από ένα νέο, όπου η ποσοστιαία βελτίωση σε σχέση με την απλή εκπαίδευση με το All Windows είναι της τάξης του 1%, ενώ η αντίστοιχη ποσοστιαία βελτίωση σε σχέση με την πρόβλεψη χωρίς επανεκπαίδευση αγγίζει το 7.6%.



Σχήμα 6.9: Ραβδόγραμμα των τεσσάρων μεθόδων μεταφοράς μάθησης (μωβ μπάρες) με τα τέσσερα benchmarks που χρησιμοποιήθηκαν γι' αυτές τις μεθόδους (πράσινες μπάρες).

## 6.5 Πειράματα σε επίπεδο δεδομένων με χρήση του Left SubForeDeCk

Στο δεύτερο στάδιο των πειραμάτων μας επιλέγοντας να εφαρμόσουμε την μέθοδο του ευνοϊκού ξεκινήματος για την επανεκπαίδευση των μοντέλων μας, δοκιμάσαμε να κάνουμε χρήση των δεδομένων του M4 με ένα λιγότερο καθολικό τρόπο, σε σχέση με όλα τα προηγούμενα πειράματα. Αντί να κάνουμε επανεκπαίδευση στο σύνολο των δεδομένων καθολικά, χωρίσαμε τον χώρο σε γειτονιές χρονοσειρών, εκπαιδεύοντας ένα μοντέλο για κάθε μία από τις γειτονιές αυτές. Έτσι, εκμεταλλευόμαστε την τοπικότητα των χρονοσειρών, δηλαδή επανεκπαιδεύουμε το μοντέλο μας σε χρονοσειρές που έχουν περισσότερα κοινά χαρακτηριστικά με τις χρονοσειρές τις οποίες στην συνέχεια θα προβλέψουμε. Για αυτό τον λόγο αναμένουμε να δούμε μία βελτίωση στην απόδοση της προβλεπτικής ικανότητας.

Αρχικά, δοκιμάσαμε μια απλή τοπική επανεκπαίδευση των μοντέλων, προκειμένου να εκπαιδεύονται για την πρόβλεψη κάθε χρονοσειράς μόνο με τα δικά της παράθυρα. Ακολούθως, χωρίσαμε το σύνολο στόχου, δηλαδή το M4, σε μικρότερα υποσύνολα και προσπαθήσαμε να προσαρμόσουμε τα μοντέλα μας σε κάθε ένα από αυτά τα υποσύνολα ξεχωριστά, σαν να ήταν

διαφορετικά υποπροβλήματα, έτσι ώστε για την πρόβλεψη των χρονοσειρών κάθε υποσυνόλου να γίνεται επανεκπαίδευση με χρήση των παραθύρων, ως δείγματα (samples), που ανήκουν σε αυτό. Ο διαχωρισμός των δεδομένων έγινε με βάση τα δύο κύρια χαρακτηριστικά τους όπως αυτά προέκυψαν μετά από την κλιμάκωση και τον μετασχηματισμό τους στις δύο κυριότερες συνιστώσες τους, βάσει της τεχνικής PCA. Το συνολικό σφάλμα της προσέγγισης αυτής προέκυψε ως ο σταθμισμένος μέσος όρος των σφαλμάτων κάθε υποσυνόλου.

### 6.5.1 Ανά Χρονοσειρά

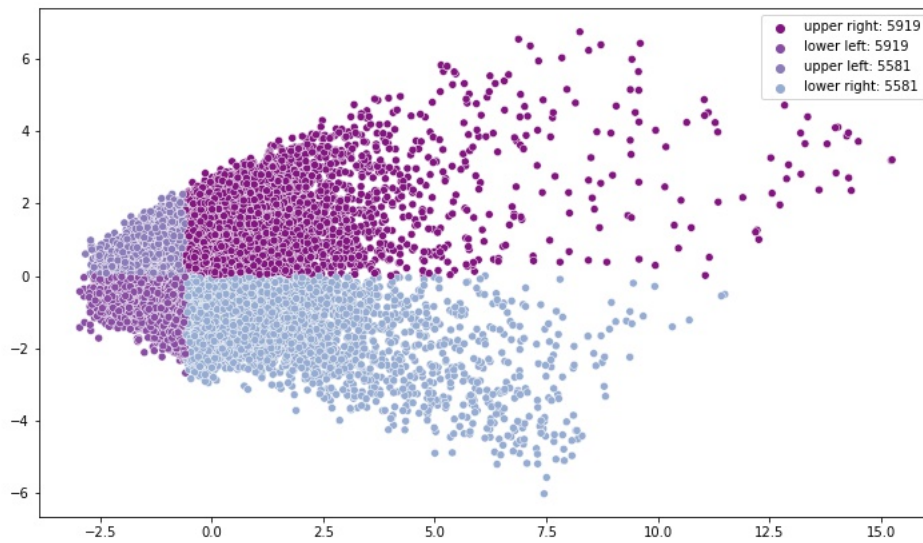
Σε αυτό το πείραμα διαχωρίζουμε τον χώρο επανεκπαίδευσης έτσι ώστε κάθε χρονοσειρά προς πρόβλεψη να επανεκπαιδύεται μόνο στα δικά της προηγούμενα χρονικά παράθυρα. Αν η χρονοσειρά προς πρόβλεψη έχει μικρό πλήθος παρατηρήσεων και δεν διαθέτει προηγούμενα χρονικά παράθυρα τότε θα προβλεφθεί χωρίς καμία επανεκπαίδευση του μοντέλου προέλευσης. Επομένως, με αυτό το πείραμα πετυχαίνουμε τον μεγαλύτερο δυνατό βαθμό λεπτομέρειας (granularity) όσον αφορά στην επανεκπαίδευση κάθε χρονοσειράς.

Συμπεριλάβαμε στην έρευνα μας το παρόν πείραμα, διότι η ίδια η μεταφορά μάθησης προσφέρεται, όπως προαναφέρθηκε, και σε μικρούς οργανισμούς. Είναι, λοιπόν, πολύ πιθανό κάποιος οργανισμός να επιθυμεί να παράξει προβλέψεις για μια και μόνο χρονοσειρά του και, ως εκ τούτου, θέλαμε να εξάγουμε συμπεράσματα σχετικά με το πόσο αποδοτική θα ήταν μια τέτοια χρήση του υπερμοντέλου. Παράλληλα, συμβατικά η πρόβλεψη χρονοσειρών στηριζόταν σε τοπική χρήση των δεδομένων τους, οπότε κρίθηκε αρκετά σημαντικό να ελέγξουμε πόσο καλά μπορούσε να συνυπάρξει μια μέθοδος διασταυρούμενης μάθησης με τη συμβατική χρήση των δεδομένων προς πρόβλεψη.

Στο πείραμα αυτό, το μοντέλο επανεκπαιδεύτηκε για κάθε χρονοσειρά μόνο με χρήση των δικών της παραθύρων, για 20 εποχές και με ρυθμό εκμάθησης  $10^{-3}$ . Επομένως, το ίδιο χρησιμοποίησε τη γνώση που είχε από την εκπαίδευση στο περιβάλλον προέλευσης και βελτιστοποιήθηκε για να προβλέπει κάθε φορά μια μόνο χρονοσειρά. Η τελική τιμή του sMAPE της συγκεκριμένης μεθόδου ήταν 14.367, νούμερο μεγαλύτερο και από τα δύο βασικά μας benchmarks και, φυσικά, χειρότερο από το αποτέλεσμα που λάβαμε με καθολική χρήση των δεδομένων.

### 6.5.2 Κόψιμο σε τεταρτημόρια

Σε αυτό το πείραμα διαχωρίζουμε τον χώρο επανεκπαίδευσης σε τέσσερα υποσύνολα βασισμένοι στα δύο κύρια χαρακτηριστικά (features) των χρονοσειρών επανεκπαίδευσης, έτσι ώστε κάθε κόψιμο να μοιράζει τον χώρο σε τέσσερα ίσα μέρη ή τεταρτημόρια (quadrants), με ίσο πλήθος δειγμάτων (cardinality). Ακολούθως, επαναλαμβάνουμε την διαδικασία αυτή αναδρομικά για κάθε τεταρτημόριο διαχωρίζοντας έτσι περαιτέρω τον χώρο επανεκπαίδευσης μας σε 16, 64 κοκ. υποομάδες (bins) βλ. Σχ. 6.10.



Σχήμα 6.10: Το σύνολο δεδομένων M4 test όπως αυτό διαχωρίζεται από τον αλγόριθμο κοψίματος σε τεταρτημόρια.

Το συγκεκριμένο πείραμα διεξήχθη προκειμένου να ερευνησουμε κατά πόσο η ομαδοποίηση χρονοσειρών με κοινά χαρακτηριστικά θα μπορούσε να παράξει περισσότερο ακριβή αποτελέσματα. Προφανώς, ο διαχωρισμός σε ισοπληθής ομάδες με έναν ορθοκανονικό τρόπο επί της απεικόνισης των χρονοσειρών στο διδιάστατο χώρο χαρακτηριστικών ήταν μια προσέγγιση που μπορούσε να αποδώσει μια πιο απλοϊκή ομαδοποίηση, προσφέροντας τα θετικά της ομαδοποίησης χωρίς να εισάγει μεγάλη πολυπλοκότητα επί του διαχωρισμού σε ομάδες.

Στον Πίνακα 6.8 βλέπουμε τα αποτελέσματα της συγκεκριμένης μεθόδου για διαχωρισμό του περιβάλλοντος στόχου σε ομάδες με χρήση αναδρομικού αλγορίθμου που χωρίζει πάντα το σύνολο σε δυνάμεις του 4, για 4, 16 και 64 ομάδες δεδομένων. Επανεκπαιδύουμε τα 50 μοντέλα με ρυθμό εκμάθησης  $10^{-5}$  με χρήση της μεθόδου της πρόωρης διακοπής early stopping για λήξη της εκπαίδευσης.

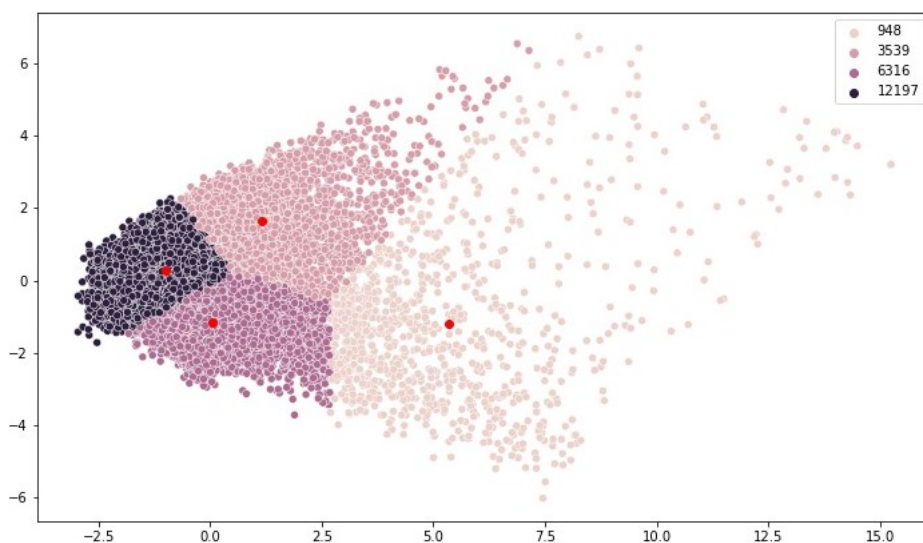
Παρατηρούμε πως αύξηση του πλήθους των δημιουργούμενων υποομάδων χρονοσειρών οδηγεί σε αύξηση του sMAPE, ενώ ήδη ο διαχωρισμός των δεδομένων του περιβάλλοντος στόχου σε τέσσερις ομάδες, έναντι της χρήσης του ως ενιαίο σύνολο, οδηγεί σε ποσοστιαία αύξηση του δείκτη σφάλματος κατά 1.3%. Βέβαια, η αύξηση από 4 σε 16 κι έπειτα σε 64 ομάδες είναι σχεδόν ανεπαίσθητη. Επομένως, κρίνεται προτιμότερη η χρήση του συνόλου με έναν καθολικό global τρόπο, συγκριτικά με το διαχωρισμό του με χρήση της μεθόδου αυτής.

Number of bins	sMAPE
4	<b>13.153</b>
16	13.167
64	13.209

Πίνακας 6.8: sMAPE της μεθόδου του κοψίματος σε τεταρτημόρια με χρήση του Left Sub-ForeDeCk ως περιβάλλοντος προέλευσης

### 6.5.3 Κοντινότερα Κεντροειδή

Σε αυτό το πείραμα διαχωρίζουμε τον χώρο επανεκπαίδευσης κάνοντας χρήση του αλγορίθμου συσταδοποίησης  $k$ -μέσων ( $k$ -means clustering). Ακολούθως, χρησιμοποιώντας τον ταξινομητή κοντινότερων γειτόνων ( $k$ -nearest neighbors classifier) με έναν κοντινότερο γείτονα, αναθέτουμε κάθε χρονοσειρά του M4 προς πρόβλεψη στην συστάδα με την οποία μοιράζεται τα περισσότερα κοινά χαρακτηριστικά, δηλαδή αυτή με το κοντινότερο κεντροειδές, όπως φαίνεται στο Σχ. 4.5. Ο συνδυασμός των αλγορίθμων  $k$ -means και  $k$ -nearest neighbors για ομαδοποίηση και ταξινόμηση (classification) είναι γνωστός στην βιβλιογραφία ως ταξινομητής κοντινότερου κεντροειδούς (nearest centroid classifier).



Σχήμα 6.11: Το σύνολο δεδομένων M4 test όπως αυτό διαχωρίζεται από τον αλγόριθμο  $k$ -μέσων για  $n = 4$  συστάδες. Το κέντρο κάθε συστάδας σημειώνονται με κόκκινη κουκίδα ενώ η πληθικότητα αυτών αναφέρεται στο υπόμνημα.

Το παρόν πείραμα συμπεριλήφθηκε στην ερευνητική μας εργασία ως συμπλήρωση του προηγούμενου. Επιθυμούσαμε να διερευνήσουμε αν η ομαδοποίηση με χρήση περισσότερων προηγμένων μεθόδων θα μπορούσε να αποδώσει καλύτερα από την απλοϊκή μέθοδο που χρησιμοποιήθηκε προηγουμένως. Καθώς ο αλγόριθμος των  $k$ -μέσων ενδείκνυται για διαχωρισμό ενός συνόλου σε  $k$  ομάδες με καλό τρόπο, τον επιλέξαμε προκειμένου να αναλάβει τον διαχωρισμό εστιάζοντας μόνο στο πόσο κοινά είναι τα χαρακτηριστικά των χρονοσειρών ανά ομάδα.



Άλλωστε, ο συγκεκριμένος αλγόριθμος φροντίζει να βρει την τοποθέτηση των κεντροειδών κάθε ομάδας κι έπειτα διαχωρίζει το χώρο με βάση τη μεσοκάθετο που έχουν τα σημεία αυτά, ανά δύο.

Στον Πίνακα 6.9 βλέπουμε τα αποτελέσματα της συγκεκριμένης μεθόδου για διαχωρισμό του περιβάλλοντος στόχου σε ομάδες με χρήση του αλγορίθμου k-Means, σε 2, 4 και 32 συστάδες (clusters). Επανεκπαιδύουμε τα 50 μοντέλα με ρυθμό εκμάθησης  $10^{-4}$  με χρήση της μεθόδου της πρόωρης διακοπής early stopping για λήξη της εκπαίδευσης.

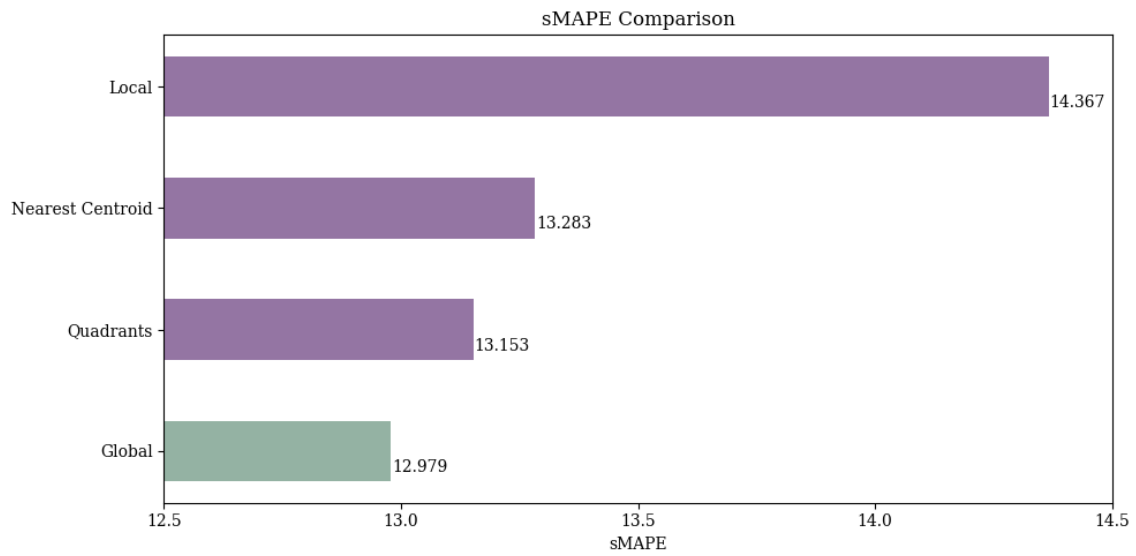
Παρατηρούμε πως αύξηση του πλήθους των δημιουργούμενων συστάδων οδηγεί σε αύξηση του sMAPE, ενώ ήδη ο διαχωρισμός των δεδομένων του περιβάλλοντος στόχου σε δύο συστάδες, έναντι της χρήσης του ως ενιαίο σύνολο, οδηγεί σε ποσοστιαία αύξηση του δείκτη σφάλματος κατά 2.3%. Επομένως, κρίνεται προτιμότερη η χρήση του συνόλου με έναν καθολικό global τρόπο, συγκριτικά με το διαχωρισμό του με χρήση της μεθόδου αυτής.

Number of clusters	sMAPE
2	<b>13.283</b>
4	13.797
32	14.085

Πίνακας 6.9: sMAPE της μεθόδου των κοντινότερων κεντροειδών με χρήση του Left SubForeDeCk ως περιβάλλοντος προέλευσης

#### 6.5.4 Συγκέντρωση αποτελεσμάτων

Στο Σχήμα 6.12 παρατηρούμε μια συγκεντρωτική καταγραφή των καλύτερων αποτελεσμάτων που απέδωσε κάθε μια από τις μεθόδους που δοκιμάσαμε, μαζί με τα αποτελέσματα που είχαμε από το benchmarking τους, το οποίο συνιστά τη χρήση της ίδιας μεθόδου επανεκπαίδευσης με καθολική χρήση των δεδομένων. Το αποτέλεσμα δείχνει μια υπεροχή της καθολικής χρήσης των δεδομένων, έναντι της τοπικής, καθώς ανεξαρτήτως μεθόδου διαχωρισμού παρατηρούμε αύξηση του σφάλματος πρόβλεψης με αύξηση του πλήθους των υποομάδων στις οποίες χωρίζουμε τα δεδομένα. Μάλιστα, η ποσοστιαία αύξηση είναι της τάξης του 1% στην καλύτερη περίπτωση και φτάνει έως και το 10% στη χειρότερη περίπτωση. Η τελευταία εμφανίζεται για την τοπική χρήση των δεδομένων, όπου πλέον δουλεύουμε σε επίπεδο χρονοσειράς για να κάνουμε την επανεκπαίδευση.



Σχήμα 6.12: Ραβδόγραμμα των τεσσάρων μεθόδων μεταφοράς μάθησης (πράσινες μπάρες) με τα δύο benchmarks που χρησιμοποιήθηκαν γι' αυτές τις μεθόδους (μωβ μπάρες).

## Κεφάλαιο 7

# Συμπεράσματα και Προεκτάσεις

### 7.1 Συμπεράσματα

Η παρούσα διπλωματική εστίασε στη μελέτη της εφαρμογής μεθόδων μεταφοράς μάθησης με σκοπό την βελτίωση της προβλεπτικής ακρίβειας νευρωνικών δικτύων στις ετήσιες χρονοσειρές του M4. Η αρχιτεκτονική που χρησιμοποιήθηκε ήταν ένα πολυστρωματικό δίκτυο perceptron με 3 κρυμμένα επίπεδα, καθένα εκ των οποίων αποτελούνταν από 27 νευρώνες. Εκτελέσαμε πειράματα με δύο διαφορετικά περιβάλλοντα προέλευσης, το ForeDeCk, το οποίο μοιράζεται μια αρκετά κοινή οριακή κατανομή πιθανότητας με το περιβάλλον στόχου -το M4-, και το Left SubForeDeCk, το οποίο διαφοροποιείται σημαντικά από το περιβάλλον στόχου ως προς την οριακή κατανομή πιθανότητας. Ακόμη, τα πειράματα που διεξήχθησαν με χρήση του δεύτερου περιβάλλοντος στόχου χωρίστηκαν σε δύο βασικές κατηγορίες. Η πρώτη αφορούσε την μέθοδο με την οποία γινόταν η επανεκπαίδευση χρησιμοποιώντας πάντα προς επανεκπαίδευση τα δεδομένα του περιβάλλοντος στόχου με καθολικό τρόπο. Η δεύτερη εστίασε στον τρόπο με τον οποίο κάνουμε χρήση των δεδομένων προς επανεκπαίδευση, δοκιμάζοντας διάφορες λιγότερο καθολικές μεθόδους, ενώ η επανεκπαίδευση γινόταν πάντα με την μέθοδο του ευνοϊκού ξεκινήματος.

Αναφορικά με την περίπτωση χρήσης του ForeDeCk ως περιβάλλοντος προέλευσης συμπεραίνουμε πως καθώς το ίδιο συνιστά ένα πολύ αντιπροσωπευτικό σύνολο δεδομένων του M4, η επανεκπαίδευση στον M4 δεν μπορεί να προσφέρει σημαντική βελτίωση της προβλεπτικής ακρίβειας του μοντέλου. Συγκεκριμένα, λαμβάνουμε μια ποσοστιαία βελτίωση της τάξης του 0.1% στην καλύτερη περίπτωση. Αν αναλογιστούμε το γεγονός πως η επανεκπαίδευση του μοντέλου έχει απαιτήσεις σε χρόνο και υπολογιστικούς πόρους η ίδια κρίνεται ακατάλληλη, καθώς δεν αποδίδει σε βαθμό τέτοιο ώστε να αντισταθμίσει το κόστος που έχει.

Στο σημείο αυτό, αξίζει να τονιστεί ξανά πως ο λόγος για τον οποίον το σύνολο δεδομένων του M4 μοιράζεται μία τόσο κοντινή κατανομή πιθανότητας με το ForeDeCk είναι επειδή το πρώτο προέκυψε από δειγματοληψία επί του δεύτερου, τα οποία δείγματα στην συνέχεια αφαιρέθηκαν από το αρχικό σύνολο δεδομένων. Ωστόσο, η πλειονότητα των προβλημάτων που καλούμαστε να επιλύσουμε με μεθόδους μεταφοράς μάθησης δεν μοιράζεται μία τόσο κοντινή

κατανομή με το σύνολο δεδομένων προέλευσης. Για τον λόγο αυτό, τα συμπεράσματα που προκύπτουν από την μελέτη που διεξήχθη με το Left SubForeDeCk ίσως να είναι και αυτά με την μεγαλύτερη αξία.

Στην περίπτωση χρήσης του Left SubForeDeCk ως περιβάλλοντος προέλευσης έχουμε μια πληθώρα συμπερασμάτων τόσο ως προς την χρήση των δεδομένων όσο και ως προς τις μεθόδους με τις οποίες γίνεται η επανεκπαίδευση.

- Η επανεκπαίδευση κρίνεται σε κάθε περίπτωση χρήσιμη και αποδοτική, καθώς καταφέρνει να μειώνει σημαντικά το σφάλμα πρόβλεψης του μοντέλου, από 14.026 σε 12.988 ή και λιγότερο, δίνοντας μια ποσοστιαία βελτίωση της τάξης του 7.6% στην καλύτερη περίπτωση.
- Η χρήση της μεταφοράς μάθησης κρίνεται επιτυχημένη, αφού το μοντέλο καταφέρνει μια ποσοστιαία μείωση του σφάλματος sMAPE κατά 1% στην καλύτερη περίπτωση.
- Καθώς το περιβάλλον προέλευσης είναι αρκετά αντιπροσωπευτικό του περιβάλλοντος στόχου είναι αποδοτικότερη η προσέγγιση των μικρών μεταβολών (fine-tuning) σε κάθε μέθοδο. Αυτό μεταφράζεται σε ένα σχετικά μικρό ρυθμό εκμάθησης και στην επανεκπαίδευση λιγότερων επιπέδων -με πάγωμα όσων βρίσκονται πιο κοντά στην είσοδο-.
- Η μέθοδος των φραγμένων εποχών αποδίδει συγκρίσιμα καλά με τη μέθοδο του ευνοϊκού ξεκινήματος. Οι δύο τους μοιράζονται την ίδια λογική, όμως η πρώτη μειώνει το χρόνο και την πολυπλοκότητα της εκπαίδευσης και, ως εκ τούτου, αν και η δεύτερη καταφέρνει να μειώσει περισσότερο το σφάλμα πρόβλεψης δεν αντισταθμίζει το κόστος της και, έτσι, κρίνεται λιγότερο προτιμητέα σε σχέση με τη μέθοδο των φραγμένων εποχών.
- Οι μέθοδοι του παγώματος επιπέδων και της επαναρχικοποίησης του επιπέδου εξόδου καταφέρνουν τη μεγαλύτερη μείωση του σφάλματος πρόβλεψης που επετεύχθη στην παρούσα εργασία. Οι ίδιες συνιστούν, μάλιστα, και τις μεθόδους με το μικρότερο κόστος. Ως εκ τούτου, προτείνονται από εμάς ως οι καλύτερες μέθοδοι επανεκπαίδευσης για το συγκεκριμένο πρόβλημα.
- Η καθολική χρήση των δεδομένων κατά την επανεκπαίδευση κρίνεται η βέλτιστη, καθώς η τοπική τους χρήση χειροτερεύει το sMAPE κατά 10%, ενώ και ο διαχωρισμός τους σε υποομάδες προκαλεί την χειρότερη απόδοση του μοντέλου με μια ποσοστιαία αύξηση κατά 1.3% του δείκτη sMAPE στην καλύτερη περίπτωση.

Σαν γενικότερο συμπέρασμα μπορούμε να πούμε πως η χρήση της μεταφοράς μάθησης είναι ωφέλιμη τόσο ως προς την βελτίωση της προβλεπτικής ακρίβειας του μοντέλου όσο και ως προς τη μείωση του χρόνου εκπαίδευσης. Σε σύνολα δεδομένων που απέχουν αρκετά μεταξύ τους η μεταφορά μάθησης είναι πιθανόν να μην οδηγήσει σε αξιόλογη ή και καθόλου βελτίωση της προβλεπτικής ακρίβειας του μοντέλου. Ωστόσο, μπορεί να πετύχει το ίδιο σφάλμα σε ένα κλάσμα του χρόνου εκπαίδευσης που θα απαιτούνταν για μια εκ του μηδενός εκπαίδευση. Η

χρήση της τεχνικής επαναρχικοποίησης του επιπέδου εξόδου, ή φραγμένων εποχών, οδηγούν σε επιτυχημένη μεταφορά μάθησης με πολύ μικρό χρόνο επανεκπαίδευσης, με τον μικρότερο να αντιστοιχεί στην μέθοδο φραγμένων εποχών για μερικές εποχές εκπαίδευσης.

Σε σύνολα δεδομένων που δεν διαφέρουν μεταξύ τους η μεταφορά μάθησης προσφέρει βελτίωση της προβλεπτικής ακρίβειας, λόγω της μεγάλης συνάφειας της αποκτηθείσας γνώσης. Μπορούμε να θεωρήσουμε πως η περίπτωση κατά την οποία οι κατανομές είναι εξαιρετικά κοντινές προσομοιάζει στην τεχνική επαύξησης των δεδομένων, μιας και το μοντέλο μας έχει στην διάθεσή του περισσότερα δεδομένα εκπαίδευσης. Στην ίδια περίπτωση, η επανεκπαίδευση καθίσταται μη παραγωγική και συστήνεται η πρόβλεψη απευθείας από το μοντέλο προέλευσης, ή η μέθοδος φραγμένων εποχών με έναν πιο μικρό ρυθμό εκμάθησης για μία ή δύο εποχές. Μια τέτοια κίνηση μπορεί να προσφέρει μικρές αλλαγές στα βάρη του μοντέλου, χάρη στη σύντομη επαφή του ίδιου με τα δεδομένα του περιβάλλοντος στόχου, με αποτέλεσμα να διατηρηθούν οι ήδη πολύ καλές ισορροπίες. Υπο αυτό το πρίσμα διαφαίνεται η μεγάλη σπουδαιότητα της μεταφοράς μάθησης.

## 7.2 Περιορισμοί και Προεκτάσεις

Όπως προαναφέρθηκε, στην παρούσα διπλωματική εργαστήκαμε με μια αρκετά απλή αρχιτεκτονική. Το ίδιο το μοντέλο επέβαλε περιορισμούς στη βελτίωση της προβλεπτικής ακρίβειας, ωστόσο κατάφερε να μας προσφέρει κάποια πολύ σημαντικά αρχικά συμπεράσματα αναφορικά με το πόσο αξιόλογη είναι η χρήση της μεταφοράς μάθησης στον τομέα των προβλέψεων. Υπάρχει, ωστόσο, αρκετό περιθώριο επιπλέον έρευνας με χρήση πιο σύνθετων αρχιτεκτονικών, όπως επαναληπτικά ή συνελικτικά δίκτυα ή, ακόμα, και βαθύτερες αρχιτεκτονικές πολυστρωματικών δικτύων perceptron. Η έρευνα τέτοιων αρχιτεκτονικών καθίσταται μεγάλης χρησιμότητας για να αποσαφηνιστεί ο τρόπος με τον οποίον ανταποκρίνονται τα δίκτυα αυτά στην εκπαίδευση σε συνδυασμό και με το μέγεθος του συνόλου δεδομένων στόχου. Για παράδειγμα, σε ένα πρόβλημα μεταφοράς μάθησης με λιγιστά δεδομένα συνόλου στόχου μπορεί να γίνει η υπόθεση ότι ένα λιτό και υπολογιστικά ελαφρύ δίκτυο MLP θα προσαρμόζεται καλύτερα στο πρόβλημα στόχου σε σχέση με ένα βαρύ υπολογιστικά δίκτυο LSTM.

Επιπρόσθετα, η έρευνα μας στηρίχθηκε αποκλειστικά και μόνο στη χρήση ετήσιων χρονοσειρών. Οι ίδιες καθιστούν μια περίπτωση χρονοσειρών χωρίς πολύ μεγάλες διακυμάνσεις, με πιο σταθερά μοτίβο. Ως εκ τούτου, η μελέτη χρονοσειρών με μεγαλύτερη συχνότητα, όπως μηνιαίες χρονοσειρές, συνιστά ένα πεδίο ανοιχτό προς περαιτέρω έρευνα, η οποία θα μπορούσε να αποδώσει αρκετά διαφορετικά αποτελέσματα. Μία τέτοια μελέτη θα είχε ιδιαίτερο ερευνητικό ενδιαφέρον καθώς η βιβλιογραφία δεν έχει αποφανθεί σχετικά με το εάν τα δίκτυα MLP μπορούν ή όχι να μάθουν την εποχικότητα μιας χρονοσειράς [72].

Ακόμα, όλη μας η εργασία στηρίχθηκε στην μεταφορά μάθησης μεταξύ δύο αρκετά όμοιων περιβαλλόντων, ακόμα και στην περίπτωση χρήσης του Left SubForeDeCk, το οποίο δημιουργήθηκε με τρόπο τέτοιο ώστε να διαφέρει αρκετά από το περιβάλλον στόχου μας. Επιπλέον,

δουλέψαμε με ίδιες εργασίες προέλευσης και στόχου. Επομένως, υπάρχει χώρος για περαιτέρω έρευνα με χρήση περιβάλλοντων που να απέχουν περισσότερο μεταξύ τους ή με διαφορετικές εργασίες στόχου, όπως η περίπτωση μεταφοράς μάθησης μεταξύ χρονοσειρών διαφορετικής συχνότητας. Ιδιαίτερο ερευνητικό ενδιαφέρον παρουσιάζει η τελευταία περίπτωση καθώς θα μπορούσε να καταδείξει το κατά πόσο είναι εφικτή ή μη η δημιουργία ενός υπερμοντέλου προεκπαιδευμένου σε ένα τεράστιο όγκο δεδομένων το οποίο στην συνέχεια θα προσαρμόζεται στις χρονοσειρές που επιθυμούμε να προβλέψουμε ανεξάρτητα από το εάν αυτές είναι μηνιαίες, ετήσιες ή άλλης συχνότητας.

Κλείνοντας, με βάση τα συμπεράσματα της παρούσας εργασίας ενισχύουμε έντονα τη δημιουργία μιας πλατφόρμας συγκέντρωσης δεδομένων και χρονοσειρών από όλη τη βιομηχανία για την χρήση τους με σκοπό τη μεταφορά μάθησης. Συν τοις άλλοις, ενισχύουμε έντονα την δημιουργία υπερμοντέλων που θα παρέχονται προς χρήση ή επανεκπαίδευση, ακριβώς όπως συμβαίνει στον κλάδο των εικόνων [56]. Το κέρδος αυτών των προσεγγίσεων για τις επιχειρήσεις και τους οργανισμούς μπορεί να είναι σημαντικά μεγάλο, καθώς καταφέρνει να διαχωρίσει τον όγκο δεδομένων που υπάρχουν προς εκπαίδευση από την ποιότητα των προβλέψεων που παράγονται. Κάθε εταιρεία, ακόμα και με λίγα δεδομένα στη διάθεση της, θα μπορούσε να εργάζεται με επανεκπαίδευση ενός εκ των προσφερόμενων υπερμοντέλων, προκειμένου να δημιουργήσει μοντέλα που θα παράγουν επιτυχείς προβλέψεις με αρκετά χαμηλό κόστος. Κατά την άποψη των συγγραφέων, σε ένα μέλλον όπου τα ανοιχτά δεδομένα και τα μοντέλα ανοιχτού κώδικα είναι όλο και πιο διαδεδομένα, η μεταφορά μάθησης θα αποτελέσει μία εξαιρετικά αποδοτική προσέγγιση για την εκμετάλλευση και ενσωμάτωση της διαθέσιμης πληροφορίας από επιχειρήσεις και οργανισμούς.

# Bibliography

- [1] A. Tealab, “Time series forecasting using artificial neural networks methodologies: A systematic review,” *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 334–340, 2018.
- [2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [3] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural networks*, vol. 61, pp. 85–117, 2015.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [5] S. Smyl, “A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting,” *International Journal of Forecasting*, vol. 36, no. 1, pp. 75–85, 2020. M4 Competition.
- [6] G. Mahalakshmi, S. Sridevi, and S. Rajaram, “A survey on forecasting of time series data,” in *2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE’16)*, pp. 1–8, 2016.
- [7] Πετρόπουλος, Φ. and Ασημακόπουλος, Β., *Επιχειρησιακές Προβλέψεις*. Εκδόσεις Συμμετρία, 2013.
- [8] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice (2nd ed)*. OTexts, 04 2018.
- [9] V. Mehra, “Forecasting used to inr foreign exchange rate using time series analysis techniques like holtwinters simple exponential smoothing, arima and neural networks,” 12 2017.
- [10] S. Wu, K. Kempf, M. Atan, B. Aytac, S. Shirodkar, and A. Mishra, “Extending bass for improved new product forecasting,” 06 2021.
- [11] A.-A. Semenoglou, E. Spiliotis, S. Makridakis, and V. Assimakopoulos, “Investigating the accuracy of cross-learning time series forecasting methods,” *International Journal of Forecasting*, vol. 37, no. 3, pp. 1072–1084, 2021.

- 
- [12] R. G. Brown, R. F. Meyer, and D. A. D'Esopo, "The fundamental theorem of exponential smoothing," *Operations Research*, vol. 9, no. 5, pp. 673–687, 1961.
- [13] V. Assimakopoulos and K. Nikolopoulos, "The theta model: a decomposition approach to forecasting," *International Journal of Forecasting*, vol. 16, no. 4, pp. 521–530, 2000. The M3- Competition.
- [14] D. Shaub, "Fast and accurate yearly time series forecasting with forecast combinations," *International Journal of Forecasting*, vol. 36, no. 1, pp. 116–120, 2020. M4 Competition.
- [15] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
- [16] T. M. Mitchell, *Machine Learning*. New York City, NY: McGraw-Hill, 1997.
- [17] M. Ivanović and M. Radovanović, "Modern machine learning techniques and their applications," in *International Conference on Electronic, Communication, and Network*, 2014.
- [18] A. Panesar, *Machine learning and AI for healthcare*. Springer, 2019.
- [19] N. Indurkha and F. J. Damerau, *Handbook of natural language processing*, vol. 2. CRC Press, 2010.
- [20] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, 2018.
- [21] T. M. Mitchell, *The Discipline of Machine Learning*. Pittsburgh, PA: Carnegie Mellon University, School of Computer Science, Machine Learning Dept., 2006.
- [22] R. Fiebrink, "Machine learning education for artists, musicians, and other creative practitioners," *ACM Transactions on Computing Education (TOCE)*, vol. 19, no. 4, pp. 1–32, 2019.
- [23] S. Harnad, "The annotation game: On turing (1950) on computing, machinery, and intelligence (published version bowdlerized)," in *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (R. Epstein, G. Roberts, and G. Beber, eds.), pp. 23–66, Springer, 2008. Chapter: 3 Commentary On: Turing, A.M. (1950) Computing Machinery and Intelligence. Mind 49 433-460 Address: Amsterdam.
- [24] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.
- [25] F. Rosenblatt, "The perceptron—a perceiving and recognizing automaton," 1957. Report 85-460-1.



- [26] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [27] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [28] S. Linnainmaa, "Taylor expansion of the accumulated rounding error," *BIT Numerical Mathematics*, vol. 16, no. 2, pp. 146–160, 1976.
- [29] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [30] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.
- [31] C. J. C. H. Watkins, "Learning from delayed rewards," 1989.
- [32] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1, pp. 278–282, IEEE, 1995.
- [33] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [35] B. Sahiner, H.-P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images," *IEEE transactions on Medical Imaging*, vol. 15, no. 5, pp. 598–610, 1996.
- [36] S. J. Russell, *Artificial intelligence: a modern approach*. Harlow: Pearson, 2016.
- [37] J. E. T. Akinsola, "Supervised machine learning algorithms: Classification and comparison," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48, pp. 128 – 138, 06 2017.
- [38] C. M. Bishop, *Pattern recognition and machine learning*. Boston: Academic Press, 1992.
- [39] G. Saporito, "What is a perceptron?," 2019. <https://towardsdatascience.com/what-is-a-perceptron-210a50190c3b> [Accessed: 12-05-2021].
- [40] Y. Wang, Y. Li, Y. Song, and X. Rong, "The influence of the activation function in a convolution neural network model of facial expression recognition," *Applied Sciences*, vol. 10, p. 1897, 03 2020.

- [41] S. S. Haykin, *Neural networks and learning machines*. Delhi: Pearson, 2016.
- [42] D-R-E, “multilayer-perceptron.” <https://github.com/d-r-e/multilayer-perceptron>. [Accessed: 14-May-2021].
- [43] R. Madan and P. S. Mangipudi, “Predicting computer network traffic: A time series forecasting approach using dwt, arima and rnn,” in *2018 Eleventh International Conference on Contemporary Computing (IC3)*, pp. 1–5, 2018.
- [44] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman, “Stock price prediction using lstm, rnn and cnn-sliding window model,” in *2017 international conference on advances in computing, communications and informatics (icacci)*, pp. 1643–1647, IEEE, 2017.
- [45] D. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, pp. 533–536, 1986.
- [46] L. Rosasco, E. De Vito, A. Caponnetto, M. Piana, and A. Verri, “Are loss functions all the same?,” *Neural computation*, vol. 16, pp. 1063–76, 06 2004.
- [47] C. Lemaréchal, “Cauchy and the gradient method,” *Doc Math Extra*, vol. 251, no. 254, p. 10, 2012.
- [48] S. Ruder, “An overview of gradient descent optimization algorithms,” *arXiv preprint arXiv:1609.04747*, 2016.
- [49] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [50] G. Hinton, N. Srivastava, and K. Swersky, “Neural networks for machine learning lecture 6a overview of mini-batch gradient descent,” *Cited on*, vol. 14, no. 8, 2012.
- [51] T. Dozat, “Incorporating nesterov momentum into adam,” 2016.
- [52] Q. Wen, L. Sun, X. Song, J. Gao, X. Wang, and H. Xu, “Time series data augmentation for deep learning: A survey,” *ArXiv*, vol. abs/2002.12478, 2020.
- [53] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [54] L. Torrey and J. Shavlik, “Transfer learning,” in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pp. 242–264, IGI global, 2010.
- [55] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

- [56] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.
- [57] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *Journal of Big Data*, vol. 3, 2016.
- [58] M. Hussain, J. J. Bird, and D. R. Faria, “A study on cnn transfer learning for image classification,” in *Advances in Computational Intelligence Systems* (A. Lotfi, H. Bouchachia, A. Gegov, C. Langensiepen, and M. McGinnity, eds.), (Cham), pp. 191–202, Springer International Publishing, 2019.
- [59] N. Durrani, H. Sajjad, and F. Dalvi, “How transfer learning impacts linguistic knowledge in deep nlp models?,” 05 2021.
- [60] S. Müller, “Car model classification i: Transfer learning with resnet.” <https://www.rcraft.org/r-news/car-model-classification-i-transfer-learning-with-resnet/> [Accessed: 13-Jun-2021].
- [61] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The m4 competition: 100,000 time series and 61 forecasting methods,” *International Journal of Forecasting*, vol. 36, no. 1, pp. 54–74, 2020. M4 Competition.
- [62] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [63] T. Elsken, J. H. Metzen, and F. Hutter, “Neural architecture search: A survey,” *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 1997–2017, 2019.
- [64] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy,” *Nature*, vol. 585, pp. 357–362, Sept. 2020.
- [65] T. pandas development team, “pandas-dev/pandas: Pandas,” Feb. 2020.
- [66] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [67] F. Chollet *et al.*, “Keras,” 2015.
- [68] J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing in science & engineering*, vol. 9, no. 3, pp. 90–95, 2007.

- 
- [69] M. Waskom, O. Botvinnik, D. O’Kane, P. Hobson, S. Lukauskas, D. C. Gemperline, T. Augspurger, Y. Halchenko, J. B. Cole, J. Warmenhoven, J. de Ruiter, C. Pye, S. Hoyer, J. Vanderplas, S. Villalba, G. Kunter, E. Quintero, P. Bachant, M. Martin, K. Meyer, A. Miles, Y. Ram, T. Yarkoni, M. L. Williams, C. Evans, C. Fitzgerald, Brian, C. Fonnesbeck, A. Lee, and A. Qalieh, “mwaskom/seaborn: v0.8.1 (september 2017),” Sept. 2017.
- [70] S. Makridakis and M. Hibon, “The m3-competition: results, conclusions and implications,” *International Journal of Forecasting*, vol. 16, no. 4, pp. 451–476, 2000. The M3- Competition.
- [71] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, “Hyperparameter optimization for machine learning models based on bayesian optimizationb,” *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, 2019.
- [72] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and machine learning forecasting methods: Concerns and ways forward,” *PloS one*, vol. 13, no. 3, p. e0194889, 2018.