



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΒΙΟΙΑΤΡΙΚΗ

Πρόγνωση καιρού με χρήση τεχνικών Μη- χανικής Μάθησης



Αθανάσιος Μπριάνης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Επιβλέποντες

Κ. Δελήμπασης και Β. Πλαγιανάκος

Λαμία, 2019

Υπεύθυνη Δήλωση μη λογοκλοπής και ανάληψης προσωπικής ευθύνης

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, και γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα και ενυπογράφως ότι η παρούσα εργασία με τίτλο [«τίτλος εργασίας»] αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές από τις οποίες χρησιμοποίησα δεδομένα, ιδέες, φράσεις, προτάσεις ή λέξεις, είτε επακριβώς (όπως υπάρχουν στο πρωτότυπο ή μεταφρασμένες) είτε με παράφραση, έχουν δηλωθεί κατάλληλα και ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο ΔΗΛΩΝ

12/05/2019

Αθανάσιος Μπριάνης



Περιεχόμενα

Εισαγωγή.....	9
Τεχνητή Νοημοσύνη	10
Μηχανική Μάθηση	10
Νευρωνικά Δίκτυα	11
1. Τεχνητά Νευρωνικά Δίκτυα	11
1.1.1 Συνάρτηση ενεργοποίησης.....	13
1.1.2 Γραμμική Συνάρτηση (Linear)	13
1.1.3 ReLU (Rectified Linear Unit)	14
1.2 Εκπαίδευση Δικτύου	15
1.2.1 Συνάρτηση Κόστους	15
1.2.2 Μέσο τετραγωνικό Σφάλμα	15
1.2.3 Αλγόριθμος απότομης καθόδου	16
1.2.4 Αλγόριθμος Οπισθοδιάδοσης του λάθους	17
1.2.5 Μέτρηση ακρίβειας	18
1.3 Ανατροφοδοτούμενα Νευρωνικά Δίκτυα.....	18
1.3.1 Long Short-Term Memory Νευρωνικά Δίκτυα.....	19
1.3.2 Gated Recurrent Unit Νευρωνικά Δίκτυα	21
2. Πειραματική Διαδικασία	22
2.1 Τεχνικά Χαρακτηριστικά	22
2.2 Δεδομένα	22
2.2.1 Στοιχεία Καιρού και Μονάδες Μέτρησης.....	24
2.2.2 Ποιοτικός Έλεγχος.....	25
3. Εισαγωγή και Γραφική Απεικόνιση Δεδομένων	26
3.1 Καθαρισμός Δεδομένων και Στατιστική Περιγραφή	26
3.2 Μορφοποίηση Δεδομένων	32
3.3 Κανονικοποίηση Δεδομένων	34

4. Αρχιτεκτονική Νευρωνικών Δικτύων	35
4.1 Πλήρως Διασυνδεδεμένο Απλό Νευρωνικό Δίκτυο	35
4.1.1 Δεδομένα Εκπαίδευσης και Ανάπτυξη μοντέλου σε κώδικα	36
4.2 Ανατροφοδοτούμενο Νευρωνικό Δίκτυο LSTM	38
4.2.1 Δεδομένα Εκπαίδευσης και Ανάπτυξη μοντέλου σε κώδικα	39
4.3 Ανατροφοδοτούμενο Νευρωνικό Δίκτυο GRU	41
4.3.1 Δεδομένα Εκπαίδευσης και Ανάπτυξη μοντέλου σε κώδικα	42
5. Αποτελέσματα.....	44
5.1 Εκπαίδευση	44
5.2 Επιβεβαίωση (Validation)	46
5.3 Πρόβλεψη	49
5.4 Σύγκριση Αποτελεσμάτων.....	50
6. Συμπεράσματα - Συζήτηση	51
Βιβλιογραφία	55

Ευρετήριο Σχημάτων

Σχήμα 1 • Νευρωνικό Δίκτυο ενός Στρώματος Εισόδου	12
Σχήμα 2 • Νευρωνικό δίκτυο με στρώμα κρυφού επιπέδου	13
Σχήμα 3 • Γραμμική συνάρτηση ενεργοποίησης νευρώνων.....	14
Σχήμα 4 • Συνάρτηση διορθωμένης γραμμικής μονάδας ReLU.....	15
Σχήμα 5 • Ανατροφοδοτούμενο Νευρωνικό Δίκτυο με επαναληπτικό βρόχο.....	19
Σχήμα 6 • Ανεπτυγμένο Ανατροφοδοτούμενο Νευρωνικό Δίκτυο	19
Σχήμα 7 • Ανεπτυγμένο Επαναληπτικό Μοντέλο, σε ένα τυπικό RNN με ένα tanh στρώμα	20
Σχήμα 8 • Ανεπτυγμένο Επαναληπτικό Μοντέλο σε LSTM που εμπεριέχει τέσσερις πύλες.....	21
Σχήμα 9 • Ανατροφοδοτούμενο GRU δίκτυο με τρεις πύλες.....	21
Σχήμα 10 • Απλό Νευρωνικό Δίκτυο Τεσσάρων Στρωμάτων	36
Σχήμα 11 • Ανατροφοδοτούμενο LSTM Νευρωνικό Δίκτυο Τεσσάρων Στρωμάτων	39
Σχήμα 12 • Ανατροφοδοτούμενο GRU Νευρωνικό Δίκτυο Τεσσάρων Στρωμάτων.....	42

Ευρετήριο Πινάκων

Πίνακας 1 • Μη Επεξεργασμένα Δεδομένα.....	24
Πίνακας 2 • Συνοπτικά Στοιχεία Καιρού	25
Πίνακας 3 • Επεξεργασία Δεδομένων – Πλήρης πίνακας δεδομένων	28
Πίνακας 4 • Τύποι Δεδομένων.....	28
Πίνακας 5 • Στατιστική Απεικόνιση Δεδομένων	29
Πίνακας 6 • Συσχετίσεις Δεδομένων	29
Πίνακας 7 • Επεξεργασία Δεδομένων – Μορφοποίηση δεδομένων με ανάπτυξη 32 χαρακτηριστικών.....	32
Πίνακας 8 • Επεξεργασία Δεδομένων – Δεδομένα εκπαίδευσης X (X_test data).....	33
Πίνακας 9 • Επεξεργασία Δεδομένων – Δεδομένα εκπαίδευσης y (y_test data)	33
Πίνακας 10 • Επεξεργασία Δεδομένων – Δεδομένα εισόδου πρόβλεψης (target_X)	34
Πίνακας 11 • Επεξεργασία Δεδομένων – Δεδομένα στόχου πρόβλεψης (target_y)	34
Πίνακας 12 • Παράμετροι Αρχιτεκτονικής Απλού Νευρωνικού Δικτύου	37
Πίνακας 13 • Παράμετροι Αρχιτεκτονικής LSTM Νευρωνικού Δικτύου.....	40
Πίνακας 14 • Παράμετροι Αρχιτεκτονικής GRU Νευρωνικού Δικτύου	43
Πίνακας 15 • Συγκεντρωτικό Σφάλμα Μοντέλων	50
Πίνακας 16 • Συγκεντρωτικό Σφάλμα Μοντέλων στα Δεδομένα Επιβεβαίωσης.....	51
Πίνακας 17 • Συγκεντρωτική Ακρίβεια Μοντέλων	51
Πίνακας 18 • Συγκεντρωτική Ακρίβεια Μοντέλων στα Δεδομένα Επιβεβαίωσης.....	51
Πίνακας 19 • Μέσο Απόλυτο Σφάλμα Προβλέψεων	51

Ευρετήριο Εικόνων

Εικόνα 1 • Γραφικό Παράδειγμα Εξίσωσης Υπολογισμού του Λάθους των Τετραγώνων	16
Εικόνα 2 • Πολιτική Ορθής Χρήσης Δεδομένων	23

Ευρετήριο Κώδικα

Απόσπασμα Κώδικα 1 • Εισαγωγή Δεδομένων	24
Απόσπασμα Κώδικα 2 • Επεξεργασία Δεδομένων	27
Απόσπασμα Κώδικα 3 • Κανονικοποίηση δεδομένων με χρήση μεθόδου MinMaxScaler()	35
Απόσπασμα Κώδικα 4 • Διαχωρισμός δεδομένων με χρήση μεθόδου train_test_split()	36
Απόσπασμα Κώδικα 5 • Δημιουργία Απλού Νευρωνικού Δικτύου.....	37

Απόσπασμα Κώδικα 6 • Εκπαίδευση Απλού Νευρωνικού Δικτύου	37
Απόσπασμα Κώδικα 7 • Κανονικοποίηση Δεδομένων	39
Απόσπασμα Κώδικα 8 • Διαχωρισμός δεδομένων με χρήση μεθόδου train_test_split()	39
Απόσπασμα Κώδικα 9 • Δημιουργία LSTM Νευρωνικού Δικτύου	40
Απόσπασμα Κώδικα 10 • Εκπαίδευση LSTM Νευρωνικού Δικτύου	40
Απόσπασμα Κώδικα 11 • Δημιουργία GRU Νευρωνικού Δικτύου	42
Απόσπασμα Κώδικα 12 • Εκπαίδευση GRU Νευρωνικού Δικτύου	43

Ευρετήριο Διαγραμμάτων

Διάγραμμα 1 • Ετήσια Κατανομή Θερμοκρασίας	30
Διάγραμμα 2 • Ετήσια Κατανομή Ατμοσφαιρικής Πίεσης	31
Διάγραμμα 3 • Γράφημα Ετήσιας Κατανομής Ταχύτητας Ανέμου	31
Διάγραμμα 4 • Εξέλιξη Σφάλματος στις 100 Εποχές	44
Διάγραμμα 5 • Εξέλιξη Σφάλματος στις 250 Εποχές	45
Διάγραμμα 6 • Εξέλιξη Σφάλματος στις 500 Εποχές	45
Διάγραμμα 7 • Εξέλιξη Ακρίβειας στις 100 Εποχές	45
Διάγραμμα 8 • Εξέλιξη Ακρίβειας στις 250 Εποχές	46
Διάγραμμα 9 • Εξέλιξη Ακρίβειας στις 500 Εποχές	46
Διάγραμμα 10 • Εξέλιξη Σφάλματος Δεδομένων Επιβεβαίωσης στις 100 Εποχές	47
Διάγραμμα 11 • Εξέλιξη Σφάλματος Δεδομένων Επιβεβαίωσης στις 250 Εποχές	47
Διάγραμμα 12 • Εξέλιξη Σφάλματος Δεδομένων Επιβεβαίωσης στις 500 Εποχές	48
Διάγραμμα 13 • Εξέλιξη Ακρίβειας Δεδομένων Επιβεβαίωσης στις 100 Εποχές	48
Διάγραμμα 14 • Εξέλιξη Ακρίβειας Δεδομένων Επιβεβαίωσης στις 250 Εποχές	48
Διάγραμμα 15 • Εξέλιξη Ακρίβειας Δεδομένων Επιβεβαίωσης στις 500 Εποχές	49
Διάγραμμα 16 • Πρόβλεψη δεδομένων στις 100 Εποχές	49
Διάγραμμα 17 • Πρόβλεψη δεδομένων στις 250 Εποχές	50
Διάγραμμα 18 • Πρόβλεψη δεδομένων στις 500 Εποχές	50

Περίληψη

Η πρόγνωση του καιρού είναι η επιστήμη της εκτίμησης των μελλοντικών μετεωρολογικών φαινομένων που συμβαίνουν στην ατμόσφαιρα. Είναι άκρως σημαντική παράμετρος από την αρχαιότητα μέχρι τις σύγχρονες μέρες τόσο στην αγροτικές περιοχές όσο και στα αστικά κέντρα. Η χρήση τεχνητών νευρωνικών δικτύων παρουσιάζεται ως κατάλληλη μέθοδος για να περιγράψει την πολυπλοκότητα της ατμόσφαιρας. Τα νευρωνικά δίκτυα καθώς και τα ανατροφοδοτούμενα νευρωνικά δίκτυα είναι δύο υλοποιήσεις μηχανικής μάθησης που θα συγκριθούν σε αυτή την εργασία για την προβλεπτική τους ικανότητα στα στοιχεία καιρού. Πιο συγκεκριμένα με βάση τις μετρήσεις δεδομένων καιρού από μετεωρολογικό σταθμό της Λαμίας, τα μοντέλα θα προγνώσουν τις μέσες ημερήσιες τιμές θερμοκρασίας για την περιοχή της Λαμίας.

Αρχικά θα παρουσιάσουμε κάποια θεωρητικά στοιχεία για την ανάπτυξη και τον τρόπο λειτουργίας των νευρωνικών δικτύων. Ύστερα θα αναπτύξουμε τον τρόπο μορφοποίησης των δεδομένων που θα εισάγουμε στα δίκτυα. Στην συνέχεια θα αναλύσουμε την αρχιτεκτονική των μοντέλων που δημιουργήθηκαν για την πρόγνωση. Και τέλος θα παρουσιαστούν τα αποτελέσματα της έρευνας, τα συμπεράσματα για την απόδοση των δικτύων και προτάσεις για μελλοντική έρευνα στο αντικείμενο.

Abstract

Weather forecasting is the science of predicting the future weather conditions and climatic events. Since the ancient times till nowadays it is examined to be a crucial parameter for people's everyday lives both in agricultural and industrial sectors. Artificial Neural Network (ANN) modeling it is considered to be suitable for weather forecasting due to the complexity of atmosphere. In this thesis we will be studying the implementation of two famous architectures for machine learning with ANN, the Simple ANN model and the Recurrent ANN model. We will look into their ability to predict meteorological elements accurately. The object of this research is to predict the daily mean surface temperature based on various measured weather elements gained from meteorological station located in Lamia, Greece.

At first we will present some theory about the development and functionality of ANN's. Later we will format the prediction data to be suitable for Neural Network modeling and we will propose a suitable architecture for weather forecasting. In the end we will present and discuss the results and make suggestions for future research in the field.

Εισαγωγή

Η αποφασιστική παράμετρος της καθημερινότητάς μας, καθορίζεται από τις ενέργειες και τις πράξεις μας κατά την διάρκεια της ημέρας. Ένας από τους κυριότερους παράγοντες που μπορεί να επηρεαστεί η δράση μας είναι ο καιρός και τα φαινόμενα που τον συνοδεύουν. Η πρόγνωση του καιρού είναι ένα ζήτημα που απασχολεί την ανθρωπότητα από την αρχαιότητα. Οι προσπάθειες για έγκαιρη και έγκυρη εκτίμηση του είναι αντικείμενο των επιστημών για εκατοντάδες χρόνια και στην σύγχρονη εποχή με μεγαλύτερη ένταση.

Παρότι τον τελευταίο αιώνα έχουν καθοριστεί πλήρως οι μαθηματικές εξισώσεις που περιγράφουν την κίνηση της γήινης ατμόσφαιρας, όπου και λαμβάνουν χώρα τα καιρικά φαινόμενα, πρόκειται για ένα πρόβλημα χασοτικό. Τα δυναμικά συστήματα που το περιγράφουν είναι μη γραμμικά και εξαιρετικά ευαίσθητα στις αρχικές συνθήκες τους κατά την ολοκλήρωση τους στο χρόνο. Μία καλή πρακτική πέραν των μαθηματικών μοντέλων θα ήταν να χρησιμοποιήσουμε την δυνατότητα των υπολογιστών και της τεχνικής νοημοσύνης να διακρίνουν πρότυπα και να κάνουν προγνώσεις μαθαίνοντας από δεδομένα.

Ο κλάδος αυτός των επιστημών ταιριάζει στο πρόβλημα της πρόγνωσης του καιρού και των μετεωρολογικών στοιχείων του καθώς με τις σύγχρονες συσκευές που διαθέτουμε έχουμε πρόσβαση σε δεδομένα καιρού κάθε χρονική στιγμή της ημέρας. Μερικά από αυτά τα δεδομένα θα μπορούσαν να είναι η θερμοκρασία, η πίεση της ατμόσφαιρας, το σημείο δρόσου, ο υετός, η συννεφοκάλυψη, η ορατότητα κ.α.

Στην έρευνά μας θα χρησιμοποιήσουμε την μηχανική μάθηση και ιδιαίτερα τις τεχνικές των τεχνητών νευρωνικών δικτύων για να κάνουμε προγνώσεις για τον καιρό στην πόλη της Λαμίας. Πιο συγκεκριμένα θα εκπαιδεύσουμε τρία τεχνητά νευρωνικά δίκτυα για να πραγματοποιήσουν προγνώσεις για τις μέσες θερμοκρασίες ημέρας στην περιοχή της Λαμίας, εισάγοντας ως δεδομένα μάθησης μετεωρολογικά στοιχεία της κατώτερης ατμόσφαιρας. Τα δεδομένα θα εισάγονται ανά τέσσερις ημέρες και τα μοντέλα μας θα καλούνται προγνώσουν την πέμπτη ημέρα. Το πρώτο μοντέλο θα αποτελείται από ένα απλό πολυστρωματικό πλήρως διασυνδεδεμένο νευρωνικό δίκτυο, το δεύτερο από ένα ανατροφοδοτούμενο πολυστρωματικό πλήρως διασυνδεδεμένο νευρωνικό δίκτυο LSTM και το τρίτο από ένα ανατροφοδοτούμενο πολυστρωματικό πλήρως διασυνδεδεμένο νευρωνικό δίκτυο GRU.

Αρχικά θα αναφερθούν κάποιες θεωρητικές έννοιες βασικών στοιχείων που θα χρησιμοποιηθούν στην εργασία μας. Στην συνέχεια θα αναλύσουμε τον τρόπο επεξεργασίας των δεδομένων μας για

να καθορίσουμε ποια κρίνονται χρήσιμα για το πρόβλημα μας. Ύστερα θα αναλύσουμε την μοντελοποίηση των δικτύων που επιλέχθηκαν. Και τέλος θα παρουσιαστούν τα αποτελέσματα μας και θα συζητηθούν για προβληματισμό και περαιτέρω επιστημονική έρευνα.

Τεχνητή Νοημοσύνη

Τεχνητή νοημοσύνη ορίζεται σύμφωνα με τον Τζον Μακάρθι ως η *«επιστήμη και η μεθοδολογία της δημιουργίας νοούντων μηχανών»* (McCarthy, 2007). Πρόκειται για τον κλάδο της πληροφορικής που ασχολείται με την σχεδίαση και την υλοποίηση υπολογιστικών συστημάτων που μιμούνται στοιχεία ανθρώπινης συμπεριφοράς τα οποία υπονοούν έστω και στοιχειώδη ευφυΐα, όπως μάθηση προσαρμοστικότητα, εξαγωγή συμπερασμάτων από συμφραζόμενες έννοιες.

Παρότι εννοιολογικά παραπέμπει σε πληροφορικό κομμάτι της επιστήμης η τεχνητή νοημοσύνη είναι σημείο επαφής πολλαπλών επιστημών όπως η ψυχολογία, η φιλοσοφία, η νευρολογία, η επιστήμη των μηχανών κ.ο.κ., με στόχο την ανάπτυξη συλλογιστικής μάθησης και την σύνθεσης ευφυούς συμπεριφοράς. Παραδοσιακά προβλήματα τεχνητής νοημοσύνης είναι η Λογική, η εκμάθηση, ο σχεδιασμός, η επεξεργασία φυσικής γλώσσας και η μίμηση (Russell & Norvig, 2003).

Μερικά από τα εργαλεία που έχουν επιστρατευτεί για την υλοποίηση της Τεχνητής Νοημοσύνης είναι οι ευριστικοί αλγόριθμοι, οι αλγόριθμοι βελτιστοποίησης, οι μέθοδοι στατιστικής ανάλυσης πιθανοτήτων καθώς και τεχνητά νευρωνικά δίκτυα.

Μηχανική Μάθηση

Η μηχανική μάθηση αποτελεί πεδίο της επιστήμης των υπολογιστών. Αναπτύχθηκε στα πλαίσια της μελέτης μεθόδων αναγνώρισης προτύπων και της θεωρίας της μάθησης της τεχνητής νοημοσύνης (Hosch, 2009). Ο Άρθουρ Σάμιουελ ορίζει την μηχανική μάθηση ως το *«Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί»* (Simon, 2013). Πρόκειται για την μελέτη αλγορίθμων που σκοπό έχουν να εκπαιδεύουν και να κάνουν προβλέψεις ή επιλογές βάσει πειραματικών δεδομένων που τροφοδοτούνται σε αυτά.

Όταν τα δεδομένα ενός προβλήματος είναι εξαιρετικά μεγάλου μεγέθους δεν είναι εύκολο να γίνουν απευθείας αντιληπτά από τον άνθρωπο, ή να αναλυθούν με συμβατικές μεθόδους επεξεργασίας. Έτσι δημιουργείται η ανάγκη για ανακάλυψη προτύπων κατανοητών στην ανθρώπινη νόηση. Ο τομέας αυτός των επιστημών αναφέρεται ως εξόρυξη δεδομένων (data

mining). Ένα υπολογιστικό σύστημα καλείται να εκπαιδευτεί και να αναγνωρίσει πολύτιμη για τον άνθρωπο πληροφορία χρησιμοποιώντας τεχνικές μάθησης. Υπάρχει ποικιλία τεχνικών που έχουν αναπτυχθεί για την βαθιά μάθηση των μηχανών. Μερικές από αυτές είναι η εκμάθηση με δέντρο απόφασης (Maimon & Rokach, 2008), οι κανόνες συσχέτισης (Frawley, Piatetsky-Shapiro, & Matheus, 1991), ο επαγωγικός λογικός προγραμματισμός (Muggleton & Lucde, 1994), οι μηχανές υποστήριξης διανυσμάτων (Cortes & Varnik, 1995), τα δίκτυα Bayes (Pearl, 2000), οι γενετικοί αλγόριθμοι (Mitchell, 1996) και τα τεχνητά νευρωνικά δίκτυα, με τα οποία και θα ασχοληθούμε στην παρούσα εργασία.

Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα βρίσκουν διαρκώς αυξανόμενες εφαρμογές στις θετικές και όχι μόνον επιστήμες καθώς έχουν γίνει γνωστά τα τελευταία σαράντα χρόνια περίπου. Το κύριο χαρακτηριστικό τους είναι ότι οι αρχές λειτουργίας τους είναι βασισμένες στον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου. Όπως η ανθρώπινη σκέψη χρησιμοποιεί τεχνικές μάθησης, εκπαίδευσης, μνήμης δεδομένων για την λήψη μιας απόφασης, έτσι και ένα δίκτυο νευρώνων. Η χρήση ηλεκτρονικών υπολογιστών για την προσομοίωση τέτοιων δομών δικτύων είναι ιδανική γιατί με την υπολογιστική τους δύναμη και σε συνάρτηση με μαθηματικά εργαλεία που έχουν αναπτυχθεί, υπάρχει η δυνατότητα επεξεργασίας μεγάλης κλίμακας δεδομένων σε μικρό χρονικό διάστημα.

Η δομή των νευρωνικών δικτύων παρομοιάζεται με την δομή των νευρικών συστημάτων που εμφανίζονται σε βιολογικούς οργανισμούς. Κάθε νευρωνικό δίκτυο αποτελείται από ένα μεγάλο αριθμό μονάδων που λέγονται νευρώνες. Ο νευρώνας είναι η μικρότερη ανεξάρτητη μονάδα ενός δικτύου. Έχει την δυνατότητα συνεχόμενα να παίρνει και να δίνει ηλεκτρικά σήματα από και προς άλλους νευρώνες. Η μελέτη τέτοιων δομών του φυσικού κόσμου οδήγησε στην ιδέα ανάπτυξης σε ηλεκτρονικό υπολογιστή αλγορίθμων και μοντέλων που θα προσομοιάζουν την λειτουργία των δικτύων νευρώνων και ονομάζονται τεχνητά νευρωνικά δίκτυα (Argyarakis, 2019).

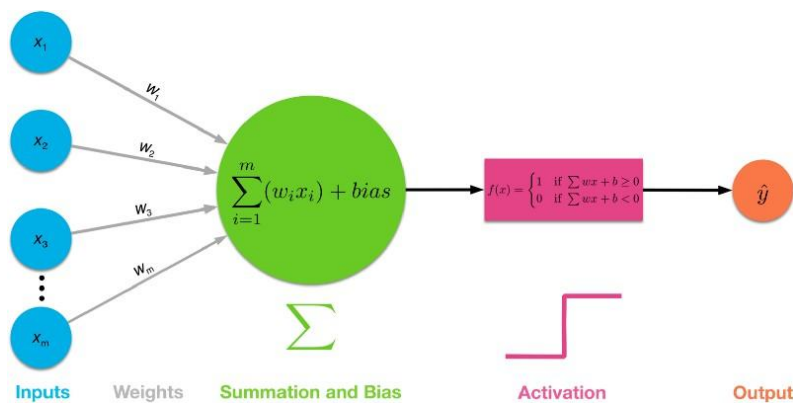
1. Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) είναι μοντέλα από μεγάλο πλήθος διασυνδεδεμένων νευρώνων. Κάθε μονάδα νευρώνα δέχεται ένα σήμα εισόδου από το περιβάλλον ή από άλλους νευρώνες και διαμέσου μιας συνάρτησης ενεργοποίησης παράγει ένα νέο σήμα εξόδου προς

άλλη μονάδα νευρώνα ή έξοδο ολόκληρου συστήματος νευρώνων. Οι αλληλεπιδράσεις οποιουδήποτε ζεύγους νευρώνων χαρακτηρίζεται από μια τιμή βάρους. Αυτή η τιμή μεταβάλλεται καθώς ο ΤΝΔ μαθαίνει και είναι ουσιαστικά η πληροφορία που αποθηκεύεται στο δίκτυο. Τα σήματα εισόδου ενός νευρώνα πολλαπλασιάζονται από τα αντίστοιχα βάρη και το άθροισμα αυτών αποτελεί την είσοδο της συνάρτησης ενεργοποίησης του νευρώνα όπως παρουσιάζεται στο σχήμα 1. Εάν x_i είναι η i -οστή είσοδος του j νευρώνα, W_{ji} είναι i -οστό βάρος του j νευρώνα, όπου αποτελείται από N συνδέσεις και Φ η συνάρτηση ενεργοποίησης του νευρώνα τότε η έξοδος του νευρώνα δίδεται από την εξίσωση:

$$Y_j = \Phi \left(\sum_{i=0}^{N-1} (x_i W_{ji}) \right) \quad (1)$$

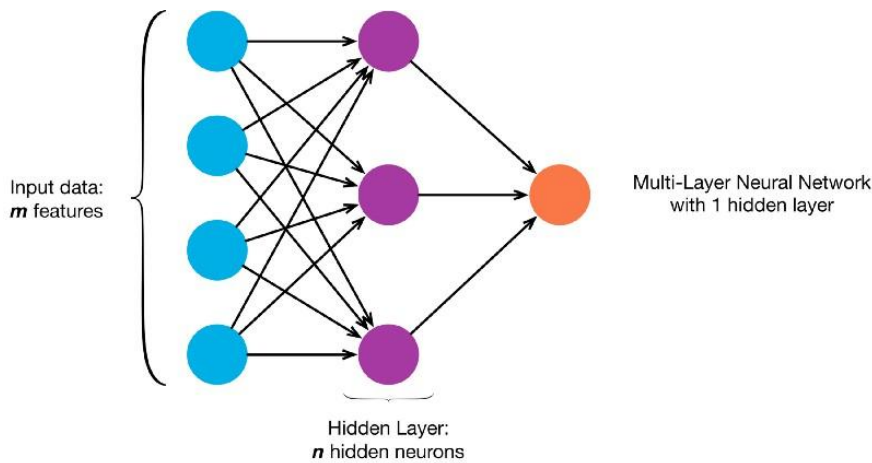
Συνήθεις συναρτήσεις ενεργοποίησης είναι η βηματική συνάρτηση, η συνάρτηση προσήμου, η ταυτοτική συνάρτηση, η σιγμοειδής συνάρτηση, η συνάρτηση γραμμικής ανόρθωσης κ.α.



Σχήμα 1 • Νευρωνικό Δίκτυο ενός Στρώματος Εισόδου

Ένα νευρωνικό δίκτυο μπορεί να είναι πιο πολύπλοκο από απλώς μια είσοδο και μία έξοδο όπως περιγράφηκε, αρκεί να προσθέσουμε και άλλα επίπεδα ενδιάμεσα, τα λεγόμενα κρυφά επίπεδα. Ονομάζονται κρυφά επειδή δεν έχουν άμεση αλληλεπίδραση με τον έξω κόσμο του συστήματός μας. Επομένως όπως φαίνεται στο σχήμα 2, για X το επίπεδο εισόδου, S την συνάρτηση ενεργοποίησης του κρυφού επιπέδου, σ την συνάρτηση ενεργοποίησης της εξόδου, W_1 τον πίνακα βαρών των νευρώνων του κρυφού επιπέδου και W_2 τον πίνακα βαρών των νευρώνων του επιπέδου της εξόδου τότε η έξοδος αυτού του δικτύου δίδεται από την εξίσωση:

$$Y = \Phi(W_2 * S(W_1 * X)) \quad (2)$$



Σχήμα 2 • Νευρωνικό δίκτυο με στρώμα κρυφού επιπέδου

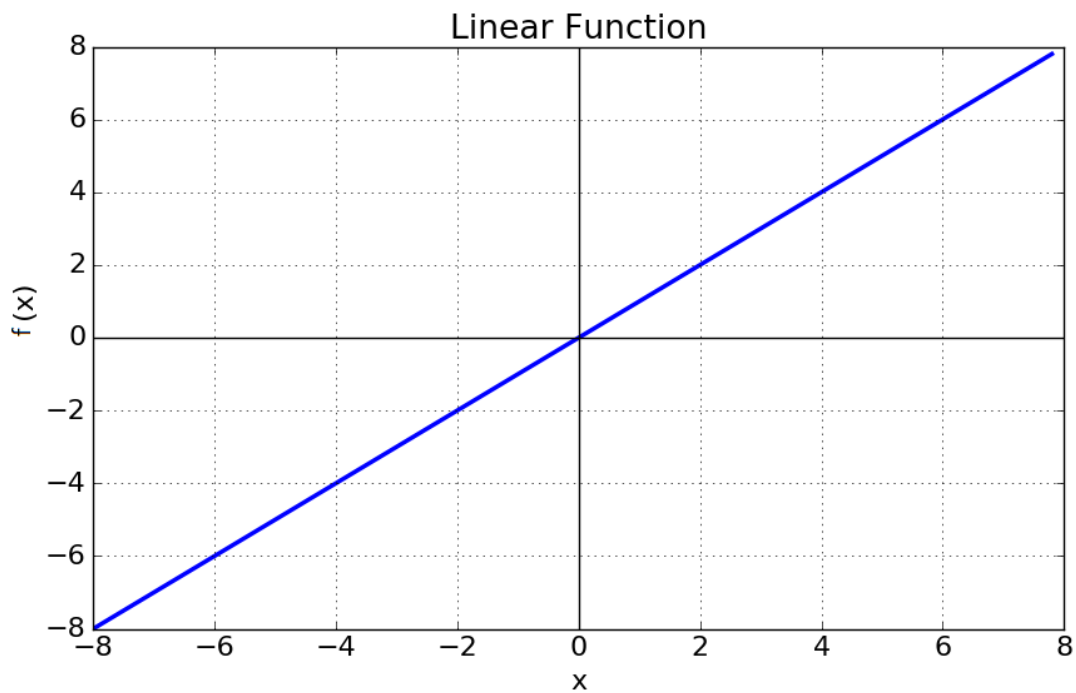
1.1.1 Συνάρτηση ενεργοποίησης

Συχνά στα προβλήματα γραμμικής παλινδρόμησης χρησιμοποιείται η συνάρτηση ενεργοποίησης κατά την εξαγωγή δεδομένων από ένα δίκτυο νευρώνων. Η συνάρτηση ενεργοποίησης μας επιτρέπει να καθορίζουμε τα όρια για την έξοδο των νευρώνων. Η σωστή επιλογή της, θα επηρεάσει και την μορφοποίηση των δεδομένων εισόδου. Είναι συχνά μη-γραμμικός τελεστής στις περιπτώσεις που εξετάζονται σύνθετα θέματα με ποικιλομορφία στα δεδομένα εισόδου. Παρακάτω θα παρουσιάσουμε τις πιο διαδεδομένες.

1.1.2 Γραμμική Συνάρτηση (Linear)

Η πιο απλή περίπτωση είναι η γραμμική συνάρτηση ενεργοποίησης. Όπου ορίζεται σε όλο διάστημα πραγματικών αριθμών.

$$f(x) = x \quad (3)$$



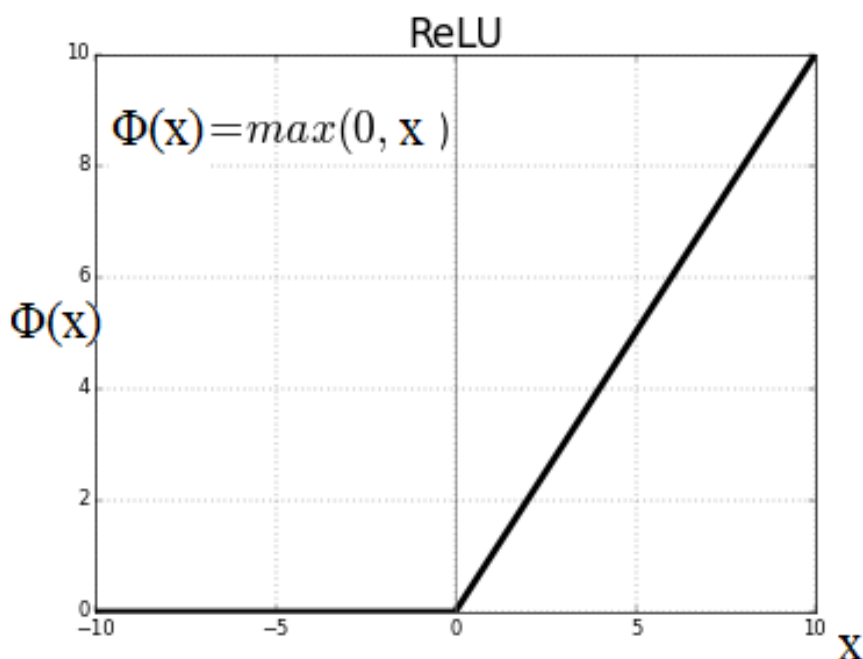
Σχήμα 3 • Γραμμική συνάρτηση ενεργοποίησης νευρώνων

1.1.3 ReLU (Rectified Linear Unit)

Η συνάρτηση διορθωμένης γραμμικής μονάδας (ReLU) (Agarap, 2018) είναι από τις πιο σύγχρονες και πιο διαδεδομένες καθώς αποδεικνύεται να λειτουργεί και να αποδίδει καλύτερα από άλλες. Χρησιμοποιείται κυρίως στην ενεργοποίηση κρυφών στρωμάτων νευρώνων και ορίζεται ως:

$$\Phi(x) = \max(0, x) \quad (4) \text{ όπου:}$$

$$\frac{dy}{dx} \Phi(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases}$$



Σχήμα 4 • Συνάρτηση διορθωμένης γραμμικής μονάδας ReLU

1.2 Εκπαίδευση Δικτύου

1.2.1 Συνάρτηση Κόστους

Ένα ΤΝΔ για να εκπαιδευτεί, χρησιμοποιεί μία διαφορίσιμη συνάρτηση κόστους. Η λειτουργία της είναι η βαθμολόγηση με μία τιμή του πόσο λάθος είναι το αποτέλεσμα του ΤΝΔ. Και αντίστοιχα προβαίνουμε στην βελτιστοποίηση των παραμέτρων του μοντέλου για να επιτύχουμε καλύτερη τιμή.

Έστω ότι Y είναι η πρόβλεψη του μοντέλου και f η συνάρτηση που τρέχει το μοντέλο μας για W βάρη (bias weights) και X δεδομένα εισόδου.

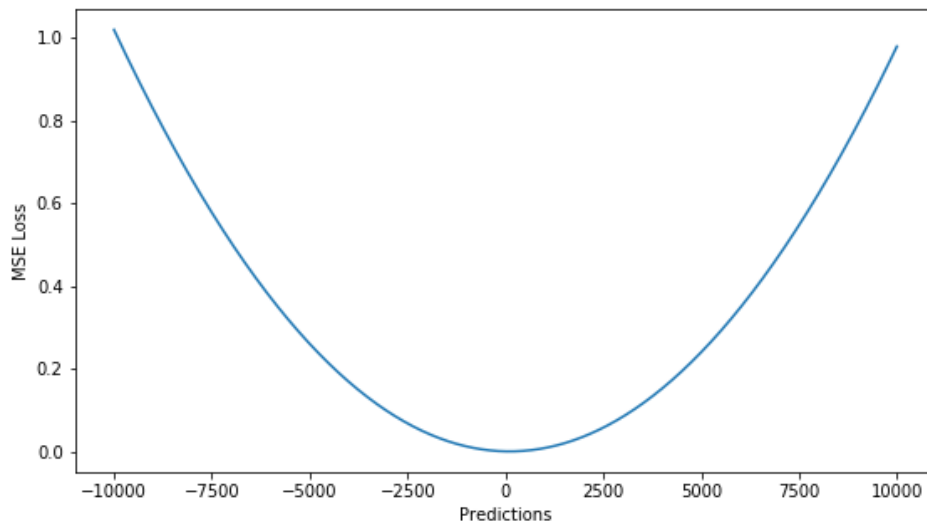
$$f(W, X_i) = Y_i \quad (5)$$

1.2.2 Μέσο τετραγωνικό Σφάλμα

Μία συνήθης συνάρτηση υπολογισμού του κόστους ενός μοντέλου είναι το μέσο τετραγωνικό σφάλμα (Mean Square Error - MSE):

$$Cost_f(W, Y_n, X_n) = \frac{1}{n} \sum_0^{n-1} |f(W, X_i) - Y_i| \quad (6)$$

Το MSE πρόκειται για την μέση τιμή των διαφορών που προκύπτουν από τις προβλέψεις και τα πραγματικά δεδομένα που θέλουμε να εκπαιδύσουμε το μοντέλο. Τα τετράγωνα των διαφορών τιμωρούν περισσότερο τις παραμέτρους των μοντέλων που έδωσαν μεγαλύτερο σφάλμα.



Εικόνα 1 • Γραφικό Παράδειγμα Εξίσωσης Υπολογισμού του Λάθους των Τετραγώνων

1.2.3 Αλγόριθμος απότομης καθόδου

Ο αλγόριθμος σύγκλισης με ελάττωση της παραγώγου (Gradient Descent) χρησιμοποιείται για να προσαρμοστούν οι παράμετροι ενός μοντέλου και να μειωθεί το κόστος. Η παράγωγος μας δίνει την κατεύθυνση στην οποία πρέπει να κινηθούν οι παράμετροι των βαρών του μοντέλου, ώστε να μειωθεί το κόστος. Το μέγεθος της μεταβολής προσδιορίζεται ευριστικά από τον ρυθμό μάθησης λ . Η εξίσωση για ένα βήμα μάθησης δίδεται παρακάτω:

$$W_{i+1} = W_i - \lambda \frac{\partial \text{Costf}(W_i, X_n, Y_n)}{\partial W_i} \quad (7)$$

όπου i ορίζεται το βήμα στο χρόνο.

Για την ομαλότερη και αποδοτικότερη βελτιστοποίηση των παραμέτρων ενός δικτύου με την χρήση του αλγορίθμου απότομης καθόδου έχει προταθεί μία λύση που χρησιμοποιείται ευρέως: η εκπαίδευση σε παρτίδες (batches).

Η εκπαίδευση σε μικρό αριθμό δεδομένων επί του συνόλου ωφελεί στην ταχύτητα επεξεργασίας καθώς βοηθά και στον σκοπό της μάθησης του δικτύου. Κάθε πακέτο δεδομένων είναι διαφορετικό σε κάθε πέρασμα και το μοντέλο μας βλέπει διαφορετικά δεδομένα.

Η μέθοδος αυτή ανήκει στους πρώτης τάξης αλγορίθμους βελτιστοποίησης. Αυτό σημαίνει πως συχνά δεν έχουμε απόλυτη σιγουριά πως θα βρεθεί το ολικό ελάχιστο της συνάρτησης καθώς από πειράματα που έχουν γίνει τα νευρωνικά δίκτυα έχουν πολλά τοπικά ελάχιστα και ενδέχεται να εγκλωβιστεί ο αλγόριθμος σε κάποιο από αυτά.

1.2.4 Αλγόριθμος Οπισθοδιάδοσης του λάθους

Όσο τα δίκτυα των νευρώνων γίνονται ολοένα και πιο πολύπλοκα με πολλαπλά επίπεδα γίνονται δύσκολο να υπολογιστεί το λάθος σε κάθε αποτέλεσμα ενός μοντέλου. Έτσι αναπτύχθηκε ο αλγόριθμος οπισθοδιάδοσης.

Η ιδέα στην ανάπτυξη του αλγορίθμου βρίσκεται στον υπολογισμό των μερικών παραγώγων της συνάρτησης του κόστους $Cost_f$ ως προς τα βάρη (weights) W και το κατώφλι (bias) b . Θα χρησιμοποιήσουμε το άρθιοσμα του τετραγώνου μεταξύ των πραγματικών τιμών και των τιμών που προβλέπει ένα ΤΝΔ.

$$Cost_f(W) = \frac{1}{2} \sum_j |f(W, x_j) - \hat{y}_j|^2 \quad (8)$$

Όπου $f_{out}(W, x_j)$ η επιθυμητή συνάρτηση εξόδου με W τα βάρη και x_j τις τιμές εισόδου, \hat{y}_j το διάνυσμα των συναρτήσεων ενεργοποίησης των εξόδων του ΤΝΔ και ο δείκτης j είναι ο εκάστοτε νευρώνας. Χρησιμοποιούμε το κλάσμα $\frac{1}{2}$ για να απλοποιήσουμε την παράγωγο του κόστους. Με τον κανόνα την αλυσίδας υπολογίζουμε την παράγωγο του κόστους ως προς τα βάρη εξόδου.

$$\frac{\partial Cost_f}{\partial W_{out}} = \frac{\partial Cost_f}{\partial f_{out}} \frac{\partial f_{out}}{\partial W_{out}} = - \sum_{j=1}^N (y_j - f(x_j)) \frac{\partial f_{out}}{\partial W_{out_j}} \quad (9)$$

Όπου η f_{out} συμβολίζει την ενεργοποίηση εξόδου που δίδεται από την σχέση:

$$f_{out} = \Phi \left(\sum_{k=1}^K w_k x_j \right) \quad (10)$$

με Φ την συνάρτηση ενεργοποίησης. Η παράγωγος f_{out} ως προς τα βάρη εξόδου w_j που ανήκουν στον χώρο W_{out} είναι x_j . Συνεπώς:

$$\frac{\partial Cost_f}{\partial W_{out}} = - \sum_{j=1}^N (y_j - f(x_j)) x_j \quad (11)$$

Με αυτό τον τρόπο έχουμε την κατεύθυνση που πρέπει να αλλάξουν τα βάρη που συνδέουν το τελευταίο κρυφό επίπεδο με τους νευρώνες της εξόδου. Ο τρόπος για να υπολογίσουμε τις αλλαγές που πρέπει να γίνουν και στα βάρη των προηγούμενων επιπέδων είναι να

χρησιμοποιήσουμε και πάλι τον κανόνα της αλυσίδας. Αυτή την φορά όμως με όλες τις ενδιάμεσες παραγώγους από την έξοδο έως και τα βάρη της εισόδου και των ενεργοποιήσεων.

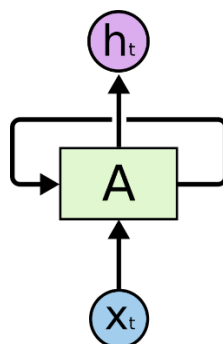
1.2.5 Μέτρηση ακρίβειας

Στα προβλήματα πρόβλεψης συχνά είναι απαραίτητο να αξιολογούμε τις εκτιμήσεις μας σε κάθε εποχή εκτέλεσης του μοντέλου μας. Συνήθως χρησιμοποιούμε ποιοτικές μεταβλητές που συγκρίνουν πόσο κοντά βρισκόμαστε στα επιθυμητά προς εκτίμηση αποτελέσματα. Στις περιπτώσεις παλινδρόμησης όπου οι μεταβλητές πρόβλεψης είναι συνεχείς, η μέτρηση της ακρίβειας είναι πιο απαιτητική. Στο πλαίσιο αυτό θα χρειαστεί να κατηγοριοποιήσουμε τις προβλέψεις μας. Θα επιλέξουμε την μέθοδο της ελαφριάς ακρίβειας (soft accuracy) όπου θα επιλέγουμε για κάθε αριθμητική τιμή πρόβλεψης την στρογγυλοποίηση της στον κοντινότερο ακέραιο αριθμό. Εάν η τιμή βρίσκεται σε ίση απόσταση από δύο ακεραίους, τότε θα επιλέγεται ο ακέραιος που είναι ζυγός αριθμός.

1.3 Ανατροφοδοτούμενα Νευρωνικά Δίκτυα

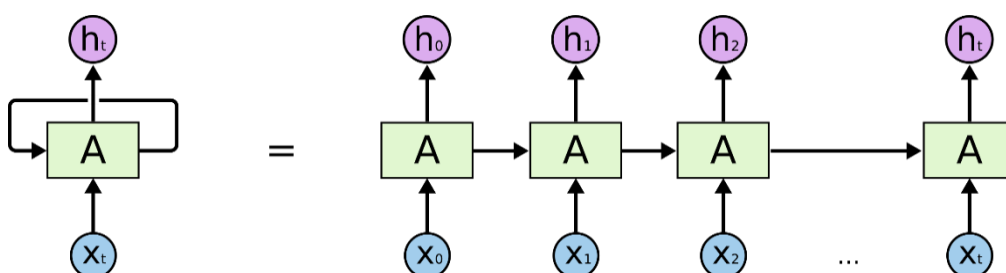
Όπως οι άνθρωποι δεν σκέφτονται από το μηδέν σε κάθε τους απόφαση έτσι δημιουργήθηκε και η ανάγκη για την ανάπτυξη ενός τρόπου επεξεργασίας των δεδομένων από μοντέλα νευρωνικών δικτύων τα οποία θα κρατάνε κάποιου είδους πληροφορία από προηγούμενες εκτελέσεις του μοντέλου. Τα συνηθισμένα νευρωνικά δίκτυα δεν έχουν τέτοιο μηχανισμό. Για παράδειγμα όταν παρακολουθούμε μία ταινία κάθε στιγμιότυπο της ταινίας δεν έχει μεμονωμένα νόημα, αλλά όλα τα στιγμιότυπα μαζί δημιουργούν μνήμη των προηγούμενων σκηνών και μπορούμε να αντιληφθούμε μια στατική σκηνή τι περιγράφει. Δεν είναι σαφές πως ένα απλό δίκτυο νευρώνων θα μπορούσε να αξιοποιήσει αυτές τις πληροφορίες.

Τα ανατροφοδοτούμενα δίκτυα (Recurrent Neural Network - RNN) έχουν την δυνατότητα να διατηρήσουν πληροφορίες μέσω συνεχών επαναλήψιμων βρόχων.



Σχήμα 5 • Ανατροφοδοτούμενο Νευρωνικό Δίκτυο με επαναληπτικό βρόχο

Στο σχήμα 3, ένα νευρωνικό δίκτυο A , δέχεται ως είσοδο x_t και δίνει έξοδο h_t . Ο βρόχος του επιτρέπει να περνάει την πληροφορία από το ένα δίκτυο στο επόμενο. Αυτές οι επαναλήψεις δεν είναι τίποτα άλλο από ένα απλό νευρωνικό δίκτυο το οποίο συνδέεται με όμοια του και περνά την πληροφορία του από το προηγούμενο, στο τωρινό και ύστερα στο επόμενο κ.ο.κ. Το σχήμα 6 δείχνει γραφικά πως θα ήταν η διάταξη εάν αναπτύσσαμε ένα τέτοιο δίκτυο σε απλούστερη μορφή.



Σχήμα 6 • Ανεπτυγμένο Ανατροφοδοτούμενο Νευρωνικό Δίκτυο

Η αλυσιδωτή φύση του δικτύου που παρουσιάζεται στο σχήμα 4, μας δίνει την αίσθηση πως υπάρχει η δυνατότητα να συμπεριφέρεται σαν χρονολογική σειρά. Και πως μια τέτοια αρχιτεκτονική θα ήταν ιδανική για την επεξεργασία δεδομένων που εξαρτώνται από μνήμη προηγούμενων καταστάσεων.

Πράγματι η χρήση RNNs τα τελευταία χρόνια δείχνει να αποδίδει εξαιρετικά σε προβλήματα όπως η αναγνώριση φυσικής γλώσσας, η μετάφραση και η μοντελοποίηση γλώσσας, η αναγνώριση προτύπων εικόνων κ.α.

1.3.1 Long Short-Term Memory Νευρωνικά Δίκτυα

Όπως είδαμε προηγουμένως η δυνατότητα της διατήρησης πληροφορίας κατά την εκπαίδευση των δικτύων είναι μείζονος σημασίας. Υπάρχουν περιπτώσεις που τα RNN έχουν αποδώσει εξαιρετικά καλά σε τέτοια προβλήματα. Όμως δεν είναι πάντα η διατήρηση της πληροφορίας σε όλη την διάρκεια ενός γεγονότος το ζητούμενο.

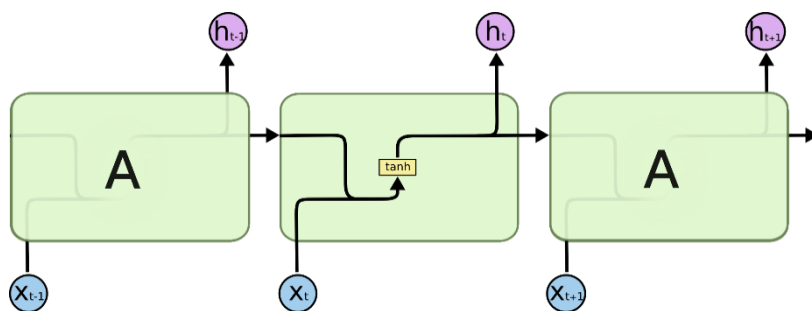
Πολλές φορές αυτό που μας ενδιαφέρει είναι η πρόσφατη πληροφορία. Για παράδειγμα για να προβλέψουμε την τελευταία λέξη στην έκφραση “τα σύννεφα είναι στον ουρανό” δεν χρειαζόμαστε περισσότερη πληροφορία, είναι εμφανές ότι η επόμενη λέξη θα είναι ουρανός. Σε τέτοιες περιπτώσεις που η σχετική με το πρόβλημα πληροφορία και η τοποθεσία της βρίσκονται

κοντά, τα RNN μπορούν να εκπαιδευτούν και να μεταδώσουν την πληροφορία.

Αντίθετα υπάρχουν περιπτώσεις όπου χρειαζόμαστε περισσότερο περιεχόμενο στην μνήμη. Παράδειγμα εάν θέλουμε να προβλέψουμε την τελευταία λέξη στην έκφραση “Μεγάλωσα στη Γαλλία ... ομιλώ με ευφράδεια Γαλλικά”. Εάν θέλουμε να εντοπίσουμε την λέξη “Γαλλικά”, το περιεχόμενο της λέξης “Γαλλία” θα μας έδινε την σωστή απάντηση. Αυτό όμως θα αποτελεί μεγάλο κενό μεταξύ της λέξης που σχετίζεται με το περιεχόμενο που αναζητούμε και το σημείο πρόβλεψης.

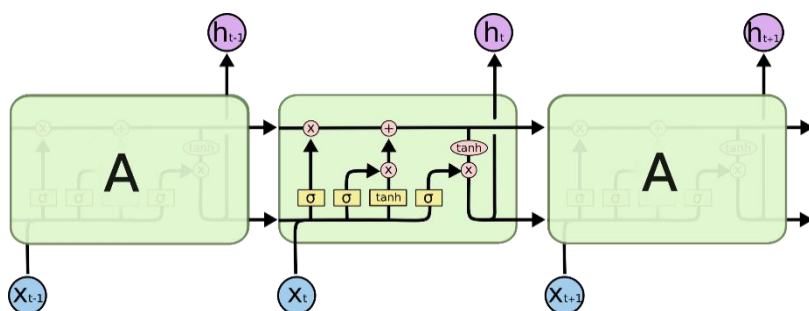
Τα ανατροφοδοτούμενα δίκτυα νευρώνων αποδεικνύεται ότι δεν αποδίδουν καλά στις περιπτώσεις όπου η μάθησή τους χρειάζεται μεγάλο διάστημα για να βρουν την σωστή πληροφορία.

Τα νευρωνικά δίκτυα Long Short-Term Memory ή απλά LSTMs είναι η λύση στο παραπάνω πρόβλημα. Αναπτύχθηκαν από τους Hochreiter και Schmidhuber το 1997 (Hochreiter & Schmidhuber, 1997) με σκοπό να αντιμετωπίσουν προβλήματα όπου η εξάρτηση της χρήσιμης πληροφορίας από μια μεγάλη περίοδο εισαγωγής δεδομένων σε ένα μοντέλο δεν θα αποτελούσε πρόβλημα.



Σχήμα 7 • Ανεπτυγμένο Επαναληπτικό Μοντέλο, σε ένα τυπικό RNN με ένα tanh στρώμα

Τα LSTMs διατηρούν την αλυσιδωτή δομή των RNN, αλλά το επαναλαμβανόμενο δομικό στοιχείο μπορεί να έχει απλά νευρωνικά δίκτυα που αλληλεπιδρούν μεταξύ τους με ένα ιδιαίτερο τρόπο.



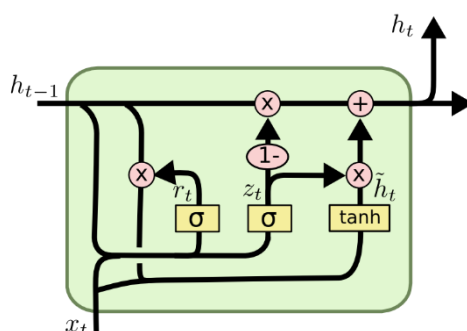
Σχήμα 8 • Ανεπτυγμένο Επαναληπτικό Μοντέλο σε LSTM που εμπεριέχει τέσσερις πύλες

Στο παραπάνω διάγραμμα κάθε γραμμή συμβολίζει το διάνυσμα από την έξοδο ενός κόμβου, στην είσοδο του επόμενου. Οι ροζ κύκλοι αναπαριστούν πράξεις που συμβαίνουν στο σημείο, όπως για παράδειγμα διανυσματική άθροιση, ενώ τα κίτρινα ορθογώνια είναι στρώματα νευρωνικών δικτύων. Οι γραμμές που ενώνονται συμβολίζουν συνένωση, ενώ αυτές που ανοίγουν συμβολίζουν την αντιγραφή όπου το αντιγραμμένα στοιχεία αλλάζουν κατεύθυνση.

Η κύρια ιδέα πίσω από τα LSTMs και το κύτταρο κατάστασης (cell state), είναι η οριζόντια γραμμή που διατρέχει τον ανεπτυγμένο γράφο και περνά την πληροφορία από από τον ένα κόμβο στον επόμενο. Αυτή η γραμμή πληροφορίας διακόπτεται για να αφαιρεθεί ή να προστεθεί πληροφορία με βάση τις διεργασίες που γίνονται μέσα στο κύτταρο. Παράδειγμα φαίνεται στο παραπάνω σχήμα, όπου στην πρώτη πύλη “απόρριψης” η πληροφορία που έρχεται από τον προηγούμενο κόμβο φιλτράρεται από μια σιγμοειδή συνάρτηση που δίνει αριθμητικές τιμές στο πεδίο $[0,1]$ και διατηρεί τα κριτήρια του πόσο χρήσιμη είναι η πληροφορία για να την περάσει στον επόμενο. Ακόμη το κύτταρο αποτελείται από δύο επιπλέον πύλες (gates). Την πύλη εισόδου και την πύλη εξόδου (Olah, 2015).

1.3.2 Gated Recurrent Unit Νευρωνικά Δίκτυα

Μια σύγχρονη απλοποιημένη εκδοχή των LSTM's είναι τα δίκτυα GRU. Συστήθηκαν από τον Cho (Chung, Gulcehre, Cho, & Bengio, 2014). Η κεντρική ιδέα πίσω από την αρχιτεκτονική τους είναι πως ενώνονται οι πύλες απόρριψης και εισόδου και δημιουργούν την πύλη ενημέρωσης. Όπως παρατηρούμε στο παρακάτω σχήμα διατηρούνται η κατάσταση του κυττάρου, η πύλη εξόδου και η εν γέννη φιλοσοφία των ανατροφοδοτούμενων δικτύων όσο αναφορά την επανάληψη και την μετάδοση πληροφορίας.



Σχήμα 9 • Ανατροφοδοτούμενο GRU δίκτυο με τρεις πύλες

2. Πειραματική Διαδικασία

Σε αυτό το κεφάλαιο θα πραγματοποιηθεί μια σύντομη αναφορά στο τεχνικό μέρος της εργασίας και θα παρουσιαστούν τα τεχνικά χαρακτηριστικά του υπολογιστή, καθώς και οι πλατφόρμες που εκτελέστηκε ο πηγαίος κώδικας. Θα παρουσιαστούν επιπλέον οι τύποι και οι πηγές των δεδομένων καθώς και τα δικαιώματα χρήσης τους.

2.1 Τεχνικά Χαρακτηριστικά

Λειτουργικό σύστημα και Προγράμματα

Η ανάπτυξη των νευρωνικών μοντέλων θα πραγματοποιηθεί σε λειτουργικό Ubuntu OS 18.04.2 LTS (Ubuntu, 2019). Ο υπολογιστής διαθέτει επεξεργαστή Intel® Core™ i5-8250U CPU @ 1.60GHz × 8, κάρτα γραφικών Intel® UHD Graphics 620 (Kabylake GT2) και 8GB RAM. Η λήψη των δεδομένων μας πραγματοποιήθηκε από το διαδίκτυο σε αρχείο CSV και για την ανάγνωση τους χρησιμοποιήθηκε το LibreOffice Calc v.6.0.7.3 (Λήψη Libre Office, 2019). Ο κώδικας αναπτύχθηκε στην πλατφόρμα προγραμματισμού Pycharm-Community v.2018.3 (Download PyCharm , 2019).

Γλώσσα προγραμματισμού και Βιβλιοθήκες

Για την συγγραφή του κώδικα χρησιμοποιήσαμε την γλώσσα προγραμματισμού Python v.3.6.6 (Python 3.6.8, 2019). Η γλώσσα Python επιλέχθηκε λόγω της ευκολίας και της δημοφιλίας της σε θέματα προγραμματισμού μηχανικής μάθησης. Εκτός των βασικών βιβλιοθηκών που έχει η αρχική εγκατάσταση της γλώσσας θα εγκαταστήσουμε και τις ακόλουθες. Για την φόρτωση και την μαθηματική επεξεργασία των δεδομένων: Pandas (Get pandas!, 2019), Numpy (NumPy, 2019) και Scipy (Obtaining SciPy libraries, 2019). Για την γραφική απεικόνιση Matplotlib (Matplotlib , 2019). Τέλος για την ανάπτυξη των μοντέλων Scikit-learn (scikit-learn 0.20.3, 2019) και Keras (Keras 2.2.4, 2019).

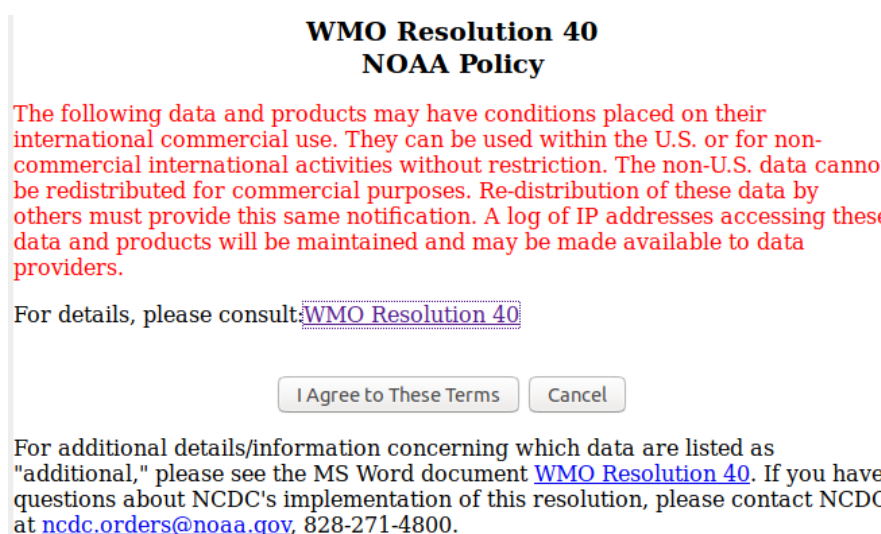
2.2 Δεδομένα

Για την έρευνα μας τα δεδομένα ανακτήθηκαν από το Κέντρο Κλιματικών Ερευνών Εθνικής Υπηρεσίας Ωκεανών και Ατμόσφαιρας των Η.Π.Α. (National Oceanic Atmospheric Administration U.S.A.) (NOAA Official Website, 2019). Το N.O.A.A. υπάγεται στο υπουργείο εμπορίου των Ηνωμένων πολιτειών (U.S. Department of Commerce) (U.S. Department of Commerce Official Website, 2019) και μας παρέχει μέσω της υπηρεσίας NESDIS (National Environmental Satellite,

Data and Information Service) (NESDIS NOAA Official Website, 2019) μετεωρολογικά στοιχεία καιρού σε παγκόσμιο επίπεδο καθώς συλλέγει και επεξεργάζεται δεδομένα καθημερινά από δορυφόρους, μετεωρολογικούς σταθμούς και άλλες πηγές. Η ανάγκη για υψηλής ποιότητας περιβαλλοντικά δεδομένα και πληροφόρηση οδήγησε στην δημιουργία του Εθνικού Κέντρου για Περιβαλλοντική Πληροφόρηση NCEI (National Center for Environmental Information). Η αρμοδιότητα του NCEI είναι η συλλογή, αποθήκευση και παροχή πρόσβασης σε ένα από τα μεγαλύτερα αρχεία στον κόσμο με Ωκεανογραφικά, Ατμοσφαιρικά και γεωφυσικά δεδομένα. Η προηγούμενη ονομασία της υπηρεσίας ήταν NCDC (National Climatic Data Center) (NCDC NOAA Official Website, 2019).

Τα δεδομένα που θα χρησιμοποιήσουμε βασίζονται στην ανταλλαγή στοιχείων σύμφωνα με τον κανονισμό που έχει θεσπίσει η Παγκόσμια Μετεωρολογική Οργάνωση (World Meteorological Organization) (WMO Official Website, 2019) με το Παγκόσμιο Πρόγραμμα Παρακολούθησης Καιρού και την συνθήκη WMO Resolution 40 (Cg-XII) (WMO, 2019). Σύμφωνα με αυτήν επιτρέπεται σε χώρες που είναι μέλη του WMO να περιορίζουν την πρόσβαση στα δεδομένα που παράγονται σε αυτήν και εξάγονται προς άλλες χώρες όταν πρόκειται να υπάρξει εμπορική χρήση τους.

Η πρόσβαση στα δεδομένα της εργασίας και η λήψη τους γίνεται από την ηλεκτρονική διεύθυνση <https://www7.ncdc.noaa.gov/CDO/cdoselect.cmd>. Με την είσοδο στην ιστοσελίδα, ζητείται η αποδοχή των όρων χρήσης των δεδομένων για περιοχές εκτός Ηνωμένων Πολιτειών (εικόνα 2).



Εικόνα 2 • Πολιτική Ορθής Χρήσης Δεδομένων

Στη συνέχεια αφού επιλέξουμε την γεωγραφική περιοχή που μας ενδιαφέρει έχουμε την

δυνατότητα να επιλέξουμε χρονικό εύρος των δεδομένων και τον μετεωρολογικό σταθμό που θέλουμε για να τα ανακτήσουμε σε αρχείο μορφοποίησης CSV. Στην περίπτωση μας επιλέγουμε για τον μετεωρολογικό σταθμό της πόλης της Λαμίας τα δεδομένα που διατίθενται για το έτος 2018.

Τυπώνουμε στην κονσόλα μας τις πρώτες και τις τελευταίες πέντε σειρές δεδομένων για να ελέγξουμε τι στοιχεία ανακτήσαμε και την μορφή τους (πίνακας 1).

```
import pandas as pd

filename = 'lamia_2018.csv'
df=pd.read_csv(filename)
print(df.head(), '\n')
print(df.tail())
```

Απόσπασμα Κώδικα 1 • Εισαγωγή Δεδομένων

STN---	WBAN	YEARMODA	TEMP	...	MIN	PRCP	SNDP	FRSHTT	
0	166750	99999	20180126	37.7	...	27.7	0.00I	999.9	0
1	166750	99999	20180127	40.4	...	30	0.00I	999.9	0
2	166750	99999	20180128	41.6	...	30.2	0.00I	999.9	0
3	166750	99999	20180129	43.1	...	29.5	0.00I	999.9	0
4	166750	99999	20180130	45.6	...	31.1	0.00I	999.9	0

[5 rows x 22 columns]

STN---	WBAN	YEARMODA	TEMP	...	MIN	PRCP	SNDP	FRSHTT	
321	166750	99999	20181227	37.7	...	30.9	0.00I	999.9	0
322	166750	99999	20181228	38.4	...	28.9	0.00I	999.9	0
323	166750	99999	20181229	42.5	...	29.5	0.00I	999.9	0
324	166750	99999	20181230	47.1	...	39.4	0.00I	999.9	0
325	166750	99999	20181231	41.1	...	37.6*	0.56F	999.9	10000

[5 rows x 22 columns]

Πίνακας 1 • Μη Επεξεργασμένα Δεδομένα

Παρατηρούμε στην πρώτη στήλη το πλήθος των σειρών που μας δίνει το πλήθος των ημερών του έτους που έχουμε μετεωρολογικά δεδομένα δηλαδή 326. Καθώς και ότι υπάρχουν 22 στήλες συνεπώς 22 διαφορετικού τύπου δεδομένα. Στη συνέχεια θα παρουσιάσουμε αναλυτικά τις στήλες και τις τιμές των δεδομένων.

2.2.1 Στοιχεία Καιρού και Μονάδες Μέτρησης

Τα παρακάτω στοιχεία επεξεργάζονται από το NCDC και αναφέρονται ως ημερήσια παγκόσμια δεδομένα επιφάνειας και παρέχονται από το USAF Climatology Center. Αφορούν τις μέσες τιμές που συλλέγονται καθημερινά. Η τελική απεικόνιση των παρατηρήσεων είναι διαθέσιμη ύστερα

από μία με δύο εργάσιμες ημέρες.

Δεδομένα	Ακρίβεια - Μονάδα μέτρησης
Μέση θερμοκρασία	(.1 Fahrenheit)
Μέσο σημείο δρόσου	(.1 Fahrenheit)
Μέση πίεση στην επιφάνεια της θάλασσας	(.1 mb)
Μέση πίεση στην περιοχή του σταθμού	(.1 mb)
Μέση ορατότητα	(.1 miles)
Μέση ταχύτητα ανέμου	(.1 knots)
Μέση μέγιστη διατηρούμενη ταχύτητα ανέμου	(.1 knots)
Μέση μέγιστη ταχύτητα ριπών ανέμου	(.1 knots)
Μέση μέγιστη θερμοκρασία	(.1 Fahrenheit)
Μέση ελάχιστη θερμοκρασία	(.1 Fahrenheit)
Μέσο μεγεθος υετού	(.01 inches)
Μέση χιονόπτωση	(.1 inches)

Πίνακας 2 • Συνοπτικά Στοιχεία Καιρού

2.2.2 Ποιοτικός Έλεγχος

Το USAF δέχεται καθημερινά δεκαοκτώ διαφορετικού τύπου μετεωρολογικά στοιχεία καιρού. Για κάποιες χρονικές περιόδους μιας ή περισσότερων χωρών τα δεδομένα ίσως δεν είναι διαθέσιμα είτε λόγω περιορισμών που έχουν τεθεί είτε λόγω επικοινωνιακών ζητημάτων. Για αυτό τον λόγο δεν παρουσιάζονται καθόλου δεδομένα για τα οποία οι ημερήσιες παρατηρήσεις που διαθέτουμε είναι λιγότερες από τέσσερις. Στα δεδομένα εμπεριέχεται σφάλμα στρογγυλοποίησης για να αποδοθούν στην εκάστοτε μονάδα μέτρησης.

Ακόμη οι μέσες ημερήσιες τιμές που συλλέγουμε από τους σταθμούς βασίζονται σε ωριαίες μετρήσεις λειτουργίας του σταθμού. Παραδείγματος χάριν κάποιιοι σταθμοί μπορεί να δίδουν την μόνιμα την ορατότητα τους κοντά στα δέκα μίλια σε όλες τις μετρήσεις τους και αυτό να συμβαίνει όχι επειδή δεν υπάρχει μεγαλύτερη ορατότητα στην περιοχή απλώς επειδή ο σταθμός είναι ρυθμισμένος να συγκεντρώνει τιμές μέχρι τα δέκα μίλια και όχι παραπάνω.

Τέλος τα δεδομένα υπόκεινται σε αυστηρούς ποιοτικούς ελέγχους έτσι ώστε να αποκωδικοποιηθεί το δυνατόν περισσότερη πληροφορία των συνοπτικών δεδομένων που συλλέγονται και να ελαχιστοποιηθούν τα τυχαία σφάλματα που υπάρχουν στα δεδομένα.

Σε αυτή την εργασία θα ασχοληθούμε κυρίως με τα δεδομένα που δίδονται σε ακεραίους ή δεκαδικούς αριθμούς όπως τα δεδομένα που παρουσιάζονται στον πίνακα 1. Όπου μία τιμή εκλείπει η καταχώρηση από την βάση των δεδομένων θα έχει τιμή 99.9 ή 999.9 ή 9999.9. Σε κάποιες τιμές κυρίως των στηλών ημερήσια μέση μέγιστη και ελάχιστη θερμοκρασία παρουσιάζεται αστερίσκος οποίος καταδεικνύει ότι η τιμή εισήχθη από το report του σταθμού αυτούσια και δεν είναι προϊόν επεξεργασίας των ωριαίων μετρήσεων.

Για παρατηρήσεις και διευκρινίσεις σχετικά με την διαδικασία λήψης, επεξεργασίας και διανομής των δεδομένων στο κοινό, αποκλειστικά υπεύθυνοι είναι η υπηρεσία NCDC, Με στοιχεία επικοινωνίας: ncdc.orders@noaa.gov.

3. Εισαγωγή και Γραφική Απεικόνιση Δεδομένων

Σε αυτό το κεφάλαιο θα μελετήσουμε τα δεδομένα και θα τα επεξεργαστούμε με τέτοιο τρόπο, ώστε οι τιμές και τα στοιχεία που θα προκύψουν να είναι κατάλληλα για την γραμμική παλινδρόμηση με νευρωνικά δίκτυα που θα υλοποιήσουμε.

3.1 Καθαρισμός Δεδομένων και Στατιστική Περιγραφή

Όπως αναφέραμε στο προηγούμενο κεφάλαιο κατά την λήψη των δεδομένων μας κάποιες τιμές περιέχουν αστερίσκο. Άλλες περιέχουν κενά και δεν διαβάζονται ως αριθμητικοί χαρακτήρες αλλά ως αλφαριθμητικοί. Άλλες έχουν τιμή 999,9. Επιπλέον κάποια ονόματα στηλών, χαρακτηριστικών δηλαδή του πίνακα δεδομένων μας περιέχουν χαρακτήρες ακατανόητους. Με τις παρακάτω εντολές θα επιλύσουμε αυτά τα ζητήματα.

Ακόμη θα αφαιρέσουμε τις στήλες που μετρούν μόνον το πλήθος των ημερήσιων τιμών που λάβαμε από τον σταθμό και από τις οποίες προέκυψε ο μέσος όρος της τιμής που χρησιμοποιούμε. Και θα αφαιρέσουμε τις στήλες που τα δεδομένα τους είτε είναι σε μεγάλο βαθμό ελλειπή, είτε δεν υπάρχει μεταβλητότητα ικανή να επηρεάσει ένα πρόβλημα γραμμικής παλινδρόμησης όπως το δικό μας.

```

df = df.drop(columns=[' ', ' .1', ' .2', ' .3', ' .4', ' .5'])
df.columns = ['STN', 'WBAN', 'YEARMODA', 'TEMP', 'DEWP', 'SLP', 'STP', 'VISIB', 'WDSP',
'MXSPD', 'GUST', 'MAX', 'MIN', 'PRCP', 'SNDP', 'FRSHTT']
df['YEARMODA'] = pd.to_datetime(df['YEARMODA'], format='%Y%m%d')
df = df.set_index('YEARMODA')
df = df.drop(columns=['STN', 'WBAN', 'VISIB', 'GUST', 'PRCP', 'SNDP', 'FRSHTT'])
df['MAX'] = df['MAX'].str.replace('*', '')
df['MAX'] = df['MAX'].str.replace(' ', '')
df['MIN'] = df['MIN'].str.replace('*', '')
df['MIN'] = df['MIN'].str.replace(' ', '')
df['MAX'] = df.MAX.astype(float)
df['MIN'] = df.MIN.astype(float)

```

Απόσπασμα Κώδικα 2 • Επεξεργασία Δεδομένων

Στο παραπάνω κομμάτι κώδικα αφαιρούμε από τον πίνακα δεδομένων μας την ταυτότητα του σταθμού συλλογής δεδομένων (STN = 166750), την ταυτότητα του σταθμού σύμφωνα με την Υπηρεσία καιρού της Πολεμικής Αεροπορίας (WBAN = 9999, η τιμή αυτή ουσιαστικά καταδεικνύει την μη ύπαρξη WBAN για τον σταθμό της Λαμίας), την ορατότητα (VISIB = η τιμή αυτή ουσιαστικά καταδεικνύει την μη ύπαρξη VISIB για τον σταθμό της Λαμίας), τις ριπές του ανέμου (GUST = 999,9) και τη χιονόπτωση (SNDP = 999,9). Τέλος αφαιρούμε τον υετό (PRCP) και την μεταβλητή (FRSHTT, πρόκειται για ενδείξεις ύπαρξης η μη ακραίων καιρικών φαινομένων) καθώς είναι κωδικοποιημένες μη αριθμητικές τιμές.

Θέτοντας ως δείκτη των σειρών την ημερομηνία (YEARMODA) ο πίνακας των δεδομένων έχει ως εξής:

YEARMODA	TEMP	DEWP	SLP	STP	WDSP	MXSPD	MAX	MIN
2018-01-26	37.7	29.0	1028.0	1026.6	3.5	5.1	54.0	27.7
2018-01-27	40.4	33.7	1031.7	1030.3	4.4	5.1	55.6	30.0
2018-01-28	41.6	37.2	1032.8	1031.4	3.5	4.1	58.1	30.2
2018-01-29	43.1	35.1	1029.7	1028.3	2.8	6.0	61.5	29.5
2018-01-30	45.6	34.2	1024.6	1023.2	2.4	2.9	62.6	31.1
2018-01-31	44.2	35.4	1024.1	1022.7	2.7	5.1	63.1	32.4
2018-02-01	47.0	37.6	1019.2	1017.8	2.7	6.0	62.1	30.7
2018-02-02	49.5	42.1	1017.8	1016.4	1.7	2.9	63.1	39.6
2018-02-03	57.9	42.5	1012.8	1011.4	7.1	14.0	73.0	43.5
2018-02-04	53.6	44.5	1011.2	1009.8	6.8	13.0	60.1	44.8
...
2018-12-22	43.5	37.7	1023.4	1022.0	1.8	2.9	61.7	30.9
2018-12-23	48.4	39.0	1022.4	1021.0	2.6	7.0	69.4	34.9
2018-12-24	47.8	40.8	1017.1	1015.7	2.3	5.1	63.7	34.7
2018-12-25	45.7	37.5	1017.9	1016.6	8.8	14.0	49.3	43.3
2018-12-26	38.0	22.0	1026.1	1024.7	7.6	11.1	48.0	32.5
2018-12-27	37.7	26.1	1025.4	1024.0	6.3	8.9	53.8	30.9
2018-12-28	38.4	28.5	1023.8	1022.4	3.5	6.0	55.2	28.9
2018-12-29	42.5	33.3	1021.3	1019.9	2.2	5.1	60.1	29.5
2018-12-30	47.1	34.4	1022.6	1021.3	3.2	5.1	57.2	39.4
2018-12-31	41.1	35.8	1020.7	1019.3	6.0	9.9	44.8	37.6

[326 rows x 8 columns]

Πίνακας 3 • Επεξεργασία Δεδομένων – Πλήρης πίνακας δεδομένων

Στα δεδομένα μας ενδέχεται να υπάρχουν μετρήσεις με τιμή 999,9. Οι τιμές αυτές θα αντικατασταθούν με το μέσο όρο της προηγούμενης και της επόμενης μέτρησης. Σε δεδομένα καιρού αυτό είναι μια αποδεκτή πρακτική διότι είναι γραμμική η μεταβολή των τιμών μας.

Με την βοήθεια συναρτήσεων της βιβλιοθήκης pandas έχουμε μια σύντομη περιγραφή των δεδομένων και κάποιες πληροφορίες για αυτά.

```
DatetimeIndex: 326 entries,
2018-01-26 to 2018-12-31
Data columns (total 8 columns):
TEMP    326 non-null float64
DEWP    326 non-null float64
SLP     326 non-null float64
STP     326 non-null float64
WDSP    326 non-null float64
MXSPD   326 non-null float64
MAX     326 non-null float64
MIN     326 non-null float64
dtypes: float64(8)
memory usage: 22.9 KB
```

Πίνακας 4 • Τύποι Δεδομένων

Από τον πίνακα 4 παρατηρούμε ότι τα δεδομένα μας αποτελούνται από 326 διανύσματα χαρακτηριστικών, με διάσταση (αριθμό χαρακτηριστικών -features) 8. Τα δεδομένα εκτίνονται από τις 26/1/2018 έως 31/12/2018. Το διάστημα των εγγραφών που παρατίθεται, υπολείπεται δέκα ημερών για τις οποίες δεν είχαμε καθόλου τιμές από την Αμερικανική μετεωρολογική υπηρεσία.

	count	mean	std	min	25%	50%	75%	max
TEMP	326.0	63.963497	12.521135	37.7	54.150	63.90	75.650	87.4
DEWP	326.0	50.041718	9.206565	22.0	42.425	52.00	57.200	68.3
SLP	326.0	1014.735890	6.184619	997.8	1010.400	1014.45	1018.875	1032.8
STP	326.0	1013.406442	6.165791	996.5	1009.025	1013.10	1017.475	1031.4
WDSP	326.0	4.603067	2.102787	1.5	3.200	4.20	5.500	14.8
MXSPD	326.0	8.256442	3.673322	2.9	6.000	7.00	9.900	26.0
MAX	326.0	76.949693	14.702933	42.4	65.150	77.80	90.700	101.8
MIN	326.0	53.685276	11.349528	27.7	45.100	54.95	64.000	74.1

Πίνακας 5 • Στατιστική Απεικόνιση Δεδομένων

Μία στατιστική σύνοψη των δεδομένων παρουσιάζεται στον πίνακα 5. Με την μέση τιμή τους, την τυπική απόκλιση, τα εκατοστημόρια των 25%, 50% (διάμεσος), 75% και την μέγιστη και ελάχιστη τιμή κάθε χαρακτηριστικού.

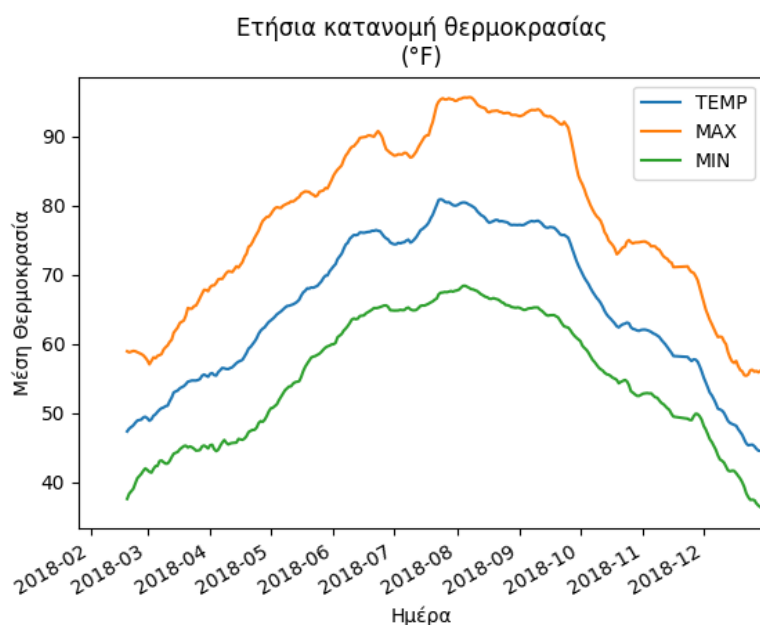
	TEMP	DEWP	SLP	STP	WDSP	MXSPD	MAX	MIN
TEMP	1.000000	0.838773	-0.392883	-0.388909	0.116211	0.091811	0.954985	0.946983
DEWP	0.838773	1.000000	-0.314976	-0.311566	-0.132116	-0.132728	0.766030	0.888187
SLP	-0.392883	-0.314976	1.000000	0.999968	-0.307021	-0.373645	-0.329647	-0.410215
STP	-0.388909	-0.311566	0.999968	1.000000	-0.306672	-0.373641	-0.325946	-0.406290
WDSP	0.116211	-0.132116	-0.307021	-0.306672	1.000000	0.880497	0.014669	0.151051
MXSPD	0.091811	-0.132728	-0.373645	-0.373641	0.880497	1.000000	0.000986	0.123359
MAX	0.954985	0.766030	-0.329647	-0.325946	0.014669	0.000986	1.000000	0.850418
MIN	0.946983	0.888187	-0.410215	-0.406290	0.151051	0.123359	0.850418	1.000000

Πίνακας 6 • Συσχετίσεις Δεδομένων

Από τις συσχετίσεις μεταξύ των μεταβλητών στον πίνακα 6, με χρήση του συντελεστή pearson παρατηρούμε τις συσχετίσεις με την μεταβλητή TEMP μέση θερμοκρασία ημέρας. Φαίνεται η μεγάλη συσχέτιση με τις μέγιστες και ελάχιστες θερμοκρασίες που παρατηρούνται μέσα στην ημέρα. Μεγάλη συσχέτιση επίσης με το μέσο σημείο δρόσου (DEWP). Αρνητική συσχέτιση εμφανίζεται με τις μέσες πιέσεις στο επίπεδο της θάλασσας (SLP) και στο επίπεδο του σταθμού (STP) το οποίο είναι καθόλα φυσιολογικό αφού οι θερμές μάζες αέρα ανεβαίνουν στην ατμόσφαιρα και δημιουργούν υποπίεση στο επίπεδο της γήινης επιφάνειας. Τέλος οι μέσες ταχύτητες του αέρα (WDSP, MXSPD) δεν παρουσιάζουν συσχέτιση με τις μετρήσεις μας για την μέση θερμοκρασία. Αλλά θα κρατήσουμε τις μετρήσεις τους, για τον έλεγχο της παλινδρόμησης

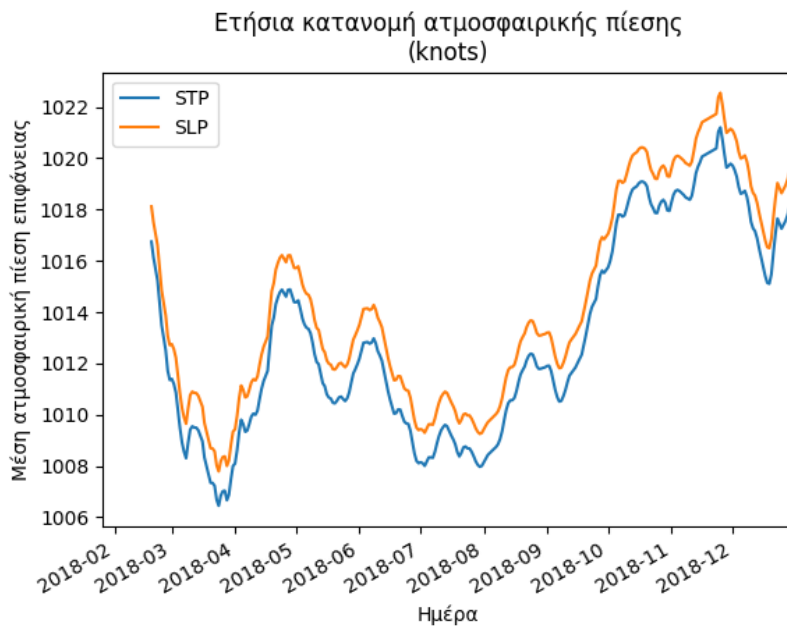
που θα πραγματοποιήσουμε διότι θα χρησιμοποιήσουμε νευρωνικά δίκτυα και η πληθώρα πληροφορίας προτείνεται.

Ενδιαφέρον θα ήταν να παρακολουθήσουμε πως μεταβάλλονται τα δεδομένα μας γραφικά για να αποκτήσουμε μια διαίσθηση της ποιότητας τους.



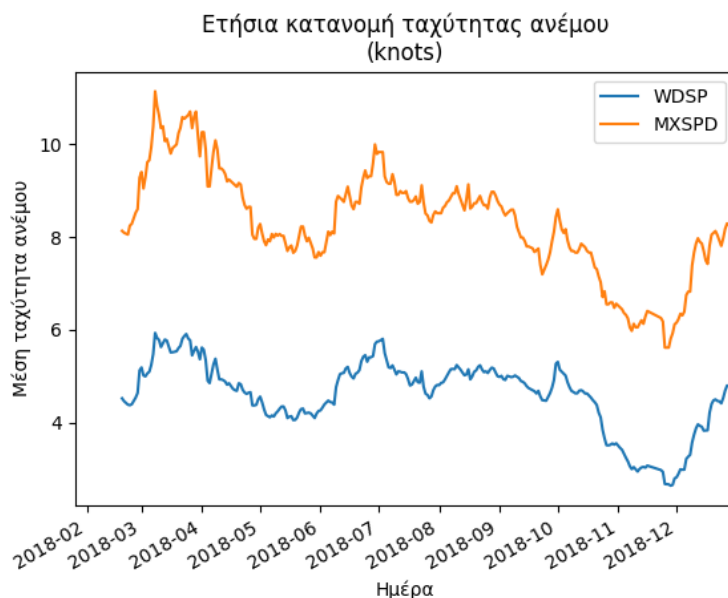
Διάγραμμα 1 • Ετήσια Κατανομή Θερμοκρασίας

Στο διάγραμμα 1 βλέπουμε πως κατανέμεται η μηνιαία μέση θερμοκρασία σε ετήσια βάση. Τα δεδομένα μας παρουσιάζουν υψηλές θερμοκρασίες τους καλοκαιρινούς μήνες και χαμηλές τον χειμώνα (TEMP). Ενώ η μέση μέγιστη θερμοκρασία ημέρας (MAX) και η μέση ελάχιστη (MIN) είναι σταθερές στις μεταβολές του TEMP και ακολουθούν την καμπυλότητα του.



Διάγραμμα 2 • Ετήσια Κατανομή Ατμοσφαιρικής Πίεσης

Στο διάγραμμα 2 παρατηρούμε την κατανομή της ετήσιας μεταβολής των μετρούμενων πιέσεων στην επιφάνεια του σταθμού (STP) και στην επιφάνεια της θάλασσας (SLP). Πράγματι η πίεση στην επιφάνεια της θάλασσας πρέπει να είναι μεγαλύτερη καθώς όσο αυξάνει το ύψος μέτρησης της ατμοσφαιρικής πίεσης, αυτή μειώνεται.



Διάγραμμα 3 • Γράφημα Ετήσιας Κατανομής Ταχύτητας Ανέμου

Τέλος από την εικόνα 5 βλέπουμε την συνοχή στις μετρήσεις της ταχύτητας του ανέμου. Τόσο στην μέση ταχύτητα ημέρας (WDSP) όσο και στην μέγιστη διατηρούμενη ταχύτητα (MXSPD) όσο

αναφορά τις μεταβολές που παρατηρούνται.

3.2 Μορφοποίηση Δεδομένων

Στη παρούσα εργασία θα χρησιμοποιήσουμε μοντελοποίηση μηχανικής μάθησης για να εκπαιδευτεί ένα δίκτυο νευρώνων και να πραγματοποιεί πρόγνωση καιρού με βάσει τις μετρήσεις προηγούμενων ημερών. Από αναφορές προτείνεται αυτή η μέθοδος (Holmstrom, Liu, & Vo, 2016), (McQuistan, 2017) και ενδεχομένως όσο μεγαλύτερο χρονικό διάστημα μετρήσεων χρησιμοποιείται για είσοδος στο δίκτυο τόσο καλύτερη η πρόγνωση που μπορεί να πραγματοποιηθεί.

Πιο συγκεκριμένα, θα διερευνήσουμε την πρόγνωση της μέσης θερμοκρασίας της ατμόσφαιρας στην περιοχή της Λαμίας κάθε πέμπτη μέρα, ενώ το δίκτυο θα εκπαιδεύεται με τις μετρήσεις μετεωρολογικών στοιχείων που περιγράψαμε παραπάνω για τις προηγούμενες τέσσερις.

Θα αναπτύξουμε επομένως τα δεδομένα μας ώστε κάθε σειρά δεδομένων να περιλαμβάνει μετρήσεις τεσσάρων ημερών και θα χρησιμοποιήσουμε αυτό το πακέτο δεδομένων για εκπαίδευση. Ενώ θα φτιάξουμε και μία στήλη για τα δεδομένα που θα στοχεύει το μοντέλο να επιτύχει.

	TEMP	DEWP	SLP	STP	WDSP	...	MAX3	MAX4	MIN2	MIN3	MIN4
0	37.7	29.0	1028.0	1026.6	3.5	...	58.1	61.5	30.0	30.2	29.5
1	40.4	33.7	1031.7	1030.3	4.4	...	61.5	62.6	30.2	29.5	31.1
2	41.6	37.2	1032.8	1031.4	3.5	...	62.6	63.1	29.5	31.1	32.4
3	43.1	35.1	1029.7	1028.3	2.8	...	63.1	62.1	31.1	32.4	30.7
4	45.6	34.2	1024.6	1023.2	2.4	...	62.1	63.1	32.4	30.7	39.6
..
321	37.7	26.1	1025.4	1024.0	6.3	...	60.1	57.2	28.9	29.5	39.4
322	38.4	28.5	1023.8	1022.4	3.5	...	57.2	44.8	29.5	39.4	37.6
323	42.5	33.3	1021.3	1019.9	2.2	...	44.8	1.0	39.4	37.6	1.0
324	47.1	34.4	1022.6	1021.3	3.2	...	1.0	1.0	37.6	1.0	1.0
325	41.1	35.8	1020.7	1019.3	6.0	...	1.0	1.0	1.0	1.0	1.0

[326 rows x 32 columns]

Πίνακας 7 • Επεξεργασία Δεδομένων – Μορφοποίηση δεδομένων με ανάπτυξη 32 χαρακτηριστικών

Γίνεται προφανές ότι ο νέος πίνακας δεδομένων δεν θα εμπεριέχει τις τελευταίες τέσσερις σειρές του, καθότι θα χρησιμοποιηθούν για την πρόγνωση της τελευταίας ημέρας. Πρόκειται για τις σειρές που περιέχουν dummy τιμές 1.0 (πίνακας 7). Επιπλέον θα αφαιρέσουμε και τις νέες 20 τελευταίες μετρήσεις που θα τις χρησιμοποιήσουμε για πρόγνωση και έλεγχο εγκυρότητας της εκπαίδευσής μας. Τελικά θα έχουμε για τις τιμές εισόδου X τον πίνακα 8 και τις τιμές

εκπαίδευσης γ τον πίνακα 9 αντίστοιχα:

```
      TEMP  DEWP  SLP      STP      WDSP  ...  MAX3  MAX4  MIN2  MIN3  MIN4
0      37.7  29.0  1028.0  1026.6  3.5  ...  58.1  61.5  30.0  30.2  29.5
1      40.4  33.7  1031.7  1030.3  4.4  ...  61.5  62.6  30.2  29.5  31.1
2      41.6  37.2  1032.8  1031.4  3.5  ...  62.6  63.1  29.5  31.1  32.4
3      43.1  35.1  1029.7  1028.3  2.8  ...  63.1  62.1  31.1  32.4  30.7
4      45.6  34.2  1024.6  1023.2  2.4  ...  62.1  63.1  32.4  30.7  39.6
..      ...   ...   ...     ...     ...   ...   ...   ...   ...   ...   ...
297    45.1  36.7  1018.3  1016.9  4.0  ...  60.3  54.7  35.1  41.4  36.7
298    43.1  39.3  1013.8  1012.4  2.4  ...  54.7  58.1  41.4  36.7  33.3
299    50.1  41.6  1016.9  1015.5  3.1  ...  58.1  57.0  36.7  33.3  30.7
300    44.6  28.2  1022.7  1021.3  7.6  ...  57.0  65.5  33.3  30.7  41.0
301    41.6  30.8  1023.3  1021.9  4.2  ...  65.5  60.4  30.7  41.0  44.1
[302 rows x 32 columns]
```

Πίνακας 8 • Επεξεργασία Δεδομένων – Δεδομένα εκπαίδευσης X (X_test data)

```
      TEMP
4      45.6
5      44.2
6      47.0
7      49.5
8      57.9
...
301    41.6
302    42.0
303    54.7
304    51.8
305    44.3
Length: 302
dtype: float64
```

Πίνακας 9 • Επεξεργασία Δεδομένων – Δεδομένα εκπαίδευσης γ (y_test data)

Τα δεδομένα πρόβλεψης (ελέγχου) που θα χρησιμοποιηθούν για να υπολογίσουν την τελική απόδοση και αξιολόγηση των μοντέλων ΤΝΔ παρουσιάζονται πίνακα 10 και τον πίνακα 11 για τις μεταβλητές target_x και target_y αντίστοιχα.

	TEMP	DEWP	SLP	STP	WDSP	...	MAX3	MAX4	MIN2	MIN3	MIN4
302	42.0	35.2	1016.0	1014.6	3.1	...	60.4	52.5	41.0	44.1	38.7
303	54.7	41.1	1010.0	1008.6	8.1	...	52.5	57.0	44.1	38.7	32.5
304	51.8	39.0	1006.8	1005.5	7.7	...	57.0	42.8	38.7	32.5	32.4
305	44.3	32.0	1013.8	1012.5	5.9	...	42.8	49.8	32.5	32.4	41.5
306	39.7	28.8	1020.2	1018.8	4.4	...	49.8	54.7	32.4	41.5	47.1
..
317	48.4	39.0	1022.4	1021.0	2.6	...	49.3	48.0	34.7	43.3	32.5
318	47.8	40.8	1017.1	1015.7	2.3	...	48.0	53.8	43.3	32.5	30.9
319	45.7	37.5	1017.9	1016.6	8.8	...	53.8	55.2	32.5	30.9	28.9
320	38.0	22.0	1026.1	1024.7	7.6	...	55.2	60.1	30.9	28.9	29.5
321	37.7	26.1	1025.4	1024.0	6.3	...	60.1	57.2	28.9	29.5	39.4

[20 rows x 32 columns]

Πίνακας 10 • Επεξεργασία Δεδομένων – Δεδομένα εισόδου πρόβλεψης (target_x)

```

TEMP
306    39.7
307    39.9
308    45.8
309    49.4
310    52.1
...
321    37.7
322    38.4
323    42.5
324    47.1
325    41.1
Length: 20
dtype: float64

```

Πίνακας 11 • Επεξεργασία Δεδομένων – Δεδομένα στόχου πρόβλεψης (target_y)

3.3 Κανονικοποίηση Δεδομένων

Τα δεδομένα που θα χρησιμοποιήσουμε ως είσοδο στα μοντέλα πρόγνωσης είναι βαθμονομημένα σε διαφορετικές κλίμακες μονάδων μέτρησης. Παράδειγμα οι πιέσεις μετριοούνται στην κλίμακα των χιλιάδων μονάδων ενώ οι θερμοκρασίες των δεκάδων μονάδων. Εάν τροφοδοτήσουμε το δίκτυο με τις τιμές των δεδομένων ως έχουν, ενδέχεται να δημιουργήσουμε πρόβλημα στην αναγνώριση των τιμών και των προτύπων που θα θέλαμε το δίκτυο να αντιληφθεί και να αναγνωρίσει.

Μία καλή πρακτική που εφαρμόζεται είναι η κανονικοποίηση των δεδομένων. Μπορούμε να μετασχηματίσουμε τα δεδομένα ώστε να έχουν ακτίνα τιμών [0, 1].

Αυτό επιτυγχάνεται με την μέθοδο `MinMaxScaler` της βιβλιοθήκης `Scikit-learn` αφαιρώντας από κάθε στοιχείο των δεδομένων ξεχωριστά την ελάχιστη τιμή της στήλης του και διαιρώντας με την διαφορά την μέγιστης τιμής και της ελάχιστης τιμής της στήλης.

```
scalerMm = MinMaxScaler(feature_range=(0,1))
scalerMm_X = scalerMm.fit(X)
scalerMm_y = scalerMm.fit(y)
X = scalerMm_X.transform(X)
y = scalerMm_y.transform(y)
```

Απόσπασμα Κώδικα 3 • Κανονικοποίηση δεδομένων με χρήση μεθόδου MinMaxScaler()

4. Αρχιτεκτονική Νευρωνικών Δικτύων

Σε αυτό το κεφάλαιο θα μελετήσουμε την αρχιτεκτονική των μοντέλων που θα αναπτυχθούν για την πρόγνωση των θερμοκρασιών επιφανείας στην περιοχή της Λαμίας. Θα αναπτυχθούν δύο (3) μοντέλα. Το πρώτο μοντέλο θα είναι ένα πολυστρωματικό πλήρως διασυνδεδεμένο δίκτυο νευρώνων. Το δεύτερο μοντέλο ένα πολυστρωματικό πλήρως διασυνδεδεμένο ανατροφοδοτούμενο δίκτυο νευρώνων με χρήση της τεχνικής LSTM. Και το τρίτο μοντέλο θα είναι πολυστρωματικό πλήρως διασυνδεδεμένο ανατροφοδοτούμενο δίκτυο νευρώνων και θα χρησιμοποιηθεί η τεχνική GRU. Τα μοντέλα διαμορφώθηκαν με real-life τρεξίματα στην υπολογιστική πλατφόρμα. Επιλέχθηκαν οι αρχιτεκτονικές που απέδιδαν καλύτερα. Επιπλέον τα βασικά χαρακτηριστικά των αρχιτεκτονικών των τριών (3) μοντέλων είναι τα ίδια ώστε να υπάρχουν άξονες σύγκρισης των μεθόδων για την αποτελεσματικότητά τους.

4.1 Πλήρως Διασυνδεδεμένο Απλό Νευρωνικό Δίκτυο

Η απλούστερη μορφή ενός τεχνητού νευρωνικού δικτύου είναι ένα δίκτυο που όλοι οι νευρώνες ενός στρώματος συνδέονται με όλους τους νευρώνες του προηγούμενου και όλους τους νευρώνες του επόμενου στρώματος. Το μοντέλο που θα κατασκευάσουμε για την υλοποίηση της πρώτης πρόγνωσης είναι ένα δίκτυο τεσσάρων (4) στρωμάτων. Το πρώτο στρώμα είναι το στρώμα εισόδου των δεδομένων μας. Το δεύτερο και το τρίτο είναι τα κρυφά στρώματα. Τέλος το τέταρτο είναι το στρώμα εξόδου.

Το πρώτο στρώμα θα δέχεται ως είσοδο μετεωρολογικά στοιχεία σύμφωνα με τα χαρακτηριστικά του πίνακα των δεδομένων μας. Εφόσον έχουμε οκτώ (8) διαφορετικά μετεωρολογικά στοιχεία και έχουμε υπολογίσει τις τιμές τους για τέσσερις (4) συνεχόμενες ημέρες, θα έχουμε 32 χαρακτηριστικά. Συνεπώς οι νευρώνες του στρώματος εισόδου θα είναι ίσου πλήθους (32).

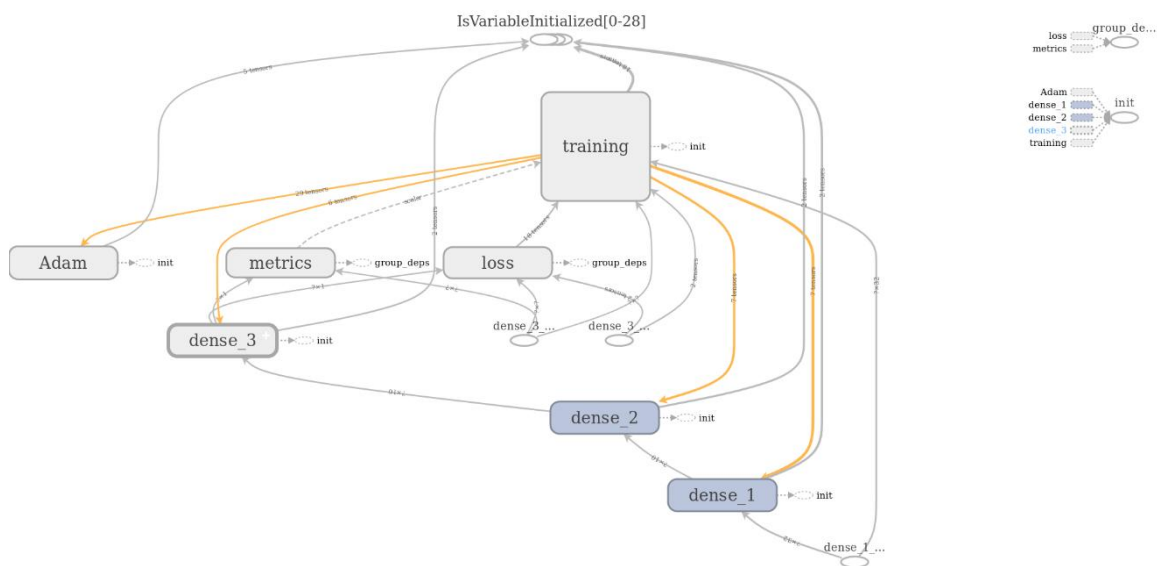
Το δεύτερο στρώμα αποτελεί το πρώτο κρυφό στρώμα νευρώνων. Αποτελείται από δέκα (10) νευρώνες. Η αρχικοποίηση των βαρών που ενώνουν το στρώμα εισόδου, με το πρώτο κρυφό στρώμα υλοποιείται με ομοιόμορφη κατανομή, ενώ ως συνάρτηση ενεργοποίησης επιλέγεται η

συνάρτηση διορθωμένης γραμμικής μονάδας (ReLU).

Το τρίτο στρώμα αποτελεί το δεύτερο κρυφό στρώμα νευρώνων. Συνδέεται πλήρως με το προηγούμενο στρώμα και αποτελείται και αυτό από δέκα (10) νευρώνες. Η συνάρτηση διορθωμένης γραμμικής μονάδας (ReLU) επιλέγεται και σε αυτό το στρώμα για την ενεργοποίηση.

Το τέταρτο στρώμα, το στρώμα εξόδου αποτελείται από ένα (1) νευρώνα. Η έξοδος του μοντέλου μας, παράγει μία αριθμητική τιμή σε κάθε εκτέλεση. Ως συνάρτηση ενεργοποίησης επιλέγεται η γραμμική συνάρτηση (linear).

Με την βοήθεια του Tensorboard (TensorFlow Core Official Website, 2019) και την χρήση callback στην μέθοδο του Keras παρουσιάζουμε γραφικά την αρχιτεκτονική του δικτύου μας.



Σχήμα 10 • Απλό Νευρωνικό Δίκτυο Τεσσάρων Στρωμάτων

4.1.1 Δεδομένα Εκπαίδευσης και Ανάπτυξη μοντέλου σε κώδικα

Τα δεδομένα που σε κάθε εποχή που τρέχει το μοντέλο και εκπαιδεύεται παράγονται από την βιβλιοθήκη sklearn και την μέθοδο `train_test_split()`. Επομένως αφού κανονικοποιήσουμε τα δεδομένα σύμφωνα με την μέθοδο που περιγράψαμε στην παράγραφο 3.4, θα χωρίσουμε τα δεδομένα μας σε δεδομένα εκπαίδευσης και δεδομένα εξέτασης. Τα δεδομένα εξέτασης θα είναι σε κάθε εποχή το είκοσι τοις εκατό (20%) των δεδομένων εισόδου (πίνακας 8).

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

Απόσπασμα Κώδικα 4 • Διαχωρισμός δεδομένων με χρήση μεθόδου `train_test_split()`

Στη συνέχεια θα δημιουργήσουμε το δίκτυο με την χρήση της βιβλιοθήκης Keras.

```
model = Sequential()
model.add(Dense(10, input_dim=32, kernel_initializer='uniform', activation='relu'))
model.add(Dense(10, activation='relu'))
model.add(Dense(1, activation='linear'))
```

Απόσπασμα Κώδικα 5 • Δημιουργία Απλού Νευρωνικού Δικτύου

Έπειτα θα τρέξουμε το μοντέλο όπου η συνάρτηση λάθους των δεδομένων εισαγωγής και των δεδομένων εξέτασης σε κάθε εποχή θα δίδονται από το μέσο τετραγωνικό σφάλμα. Σαν αλγόριθμος βελτιστοποίησης επιλέχθηκε ο Adam. Και μετρική απόδοσης του δικτύου το `soft_accuracy`.

```
model.compile(loss='mean_squared_error',
              optimizer='adam',
              metrics=[soft_acc])
model.fit(X_train, y_train,
        batch_size=nb_batches,
        epochs=nb_epochs,
        validation_data=(X_test, y_test),
        callbacks=callbacks)
```

Απόσπασμα Κώδικα 6 • Εκπαίδευση Απλού Νευρωνικού Δικτύου

Με την μέθοδο `.summary()` της βιβλιοθήκης Keras αποκτούμε μία σύντομη περιγραφή των παραμέτρων που υπολογίστηκαν για την διαμόρφωση του δικτύου.

```
>>> model.summary()
```

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 10)	330
dense_2 (Dense)	(None, 10)	110
dense_3 (Dense)	(None, 1)	11

=====
Total params: 451
Trainable params: 451
Non-trainable params: 0

Πίνακας 12 • Παράμετροι Αρχιτεκτονικής Απλού Νευρωνικού Δικτύου

Για το πρώτο κρυφό στρώμα εισάγονται τριάντα δύο (32) χαρακτηριστικά δεδομένων σε δέκα (10) νευρώνες και δέκα (10) τα βάρη μεροληψίας (bias weights), κατά συνέπεια έχουμε τριακόσιες

τριάντα (330) παραμέτρους συνολικά. Στο δεύτερο κρυφό στρώμα εισάγονται δέκα (10) τιμές δεδομένων από πρώτο κρυφό στρώμα που αθροίζουν στους δέκα (10) νευρώνες και στα δέκα (10) βάρη μεροληψίας και προκύπτουν εκατόν δέκα (110) παράμετροι. Τέλος το στρώμα εξόδου που έχει ένα (1) νευρώνα δέχεται από δεύτερο κρυφό στρώμα δέκα (10) βάρη, αθροιστικά με το ένα (1) βάρος μεροληψίας έχουμε έντεκα (11) παραμέτρους. Συνολικά το δίκτυο μας υπολογίζει 451 παραμέτρους.

4.2 Ανατροφοδοτούμενο Νευρωνικό Δίκτυο LSTM

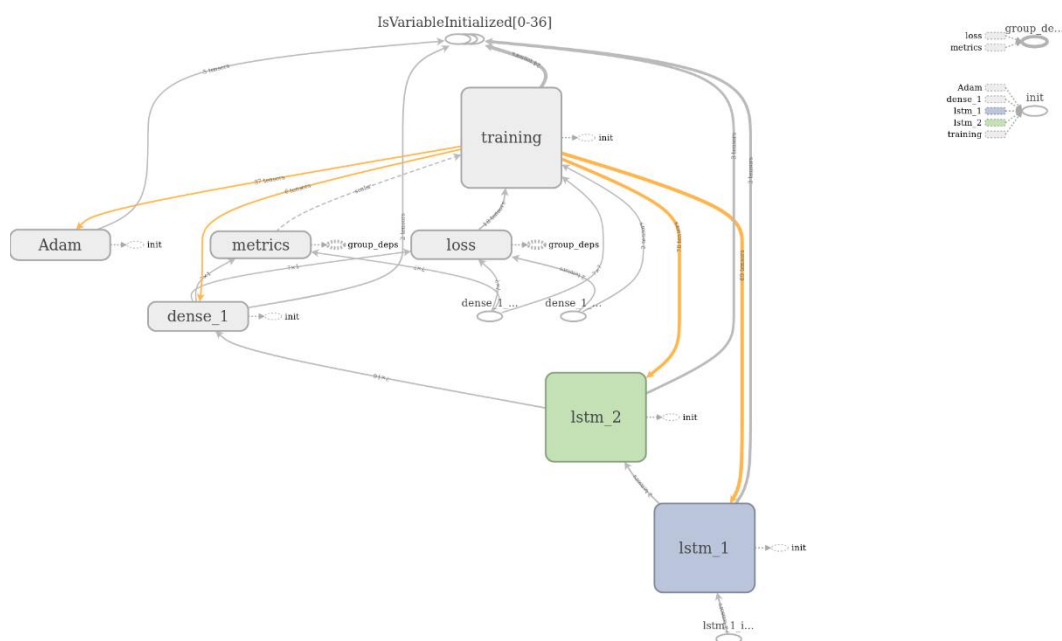
Η δεύτερη μοντελοποίηση θα περιλαμβάνει την χρήση των νευρώνων LSTM. Το μοντέλο αποτελείται από τέσσερα (4) στρώματα. Το στρώμα εισόδου, δύο (2) κρυφά στρώματα και ένα (1) στρώμα εξόδου.

Το πρώτο στρώμα δέχεται ως είσοδο μετεωρολογικά στοιχεία σύμφωνα με τα χαρακτηριστικά του πίνακα των δεδομένων μας. Εφόσον έχουμε οκτώ (8) διαφορετικά μετεωρολογικά στοιχεία και έχουμε υπολογίσει τις τιμές τους για τέσσερις (4) συνεχόμενες ημέρες, θα έχουμε τριανταδύο (32) στήλες. Συνεπώς οι νευρώνες του στρώματος εισόδου θα είναι τριανταδύο (32).

Το δεύτερο και τρίτο στρώμα αποτελούν κρυφά στρώματα νευρώνων. Αποτελούνται από δέκα (10) LSTM νευρώνες. Η αρχικοποίηση των βαρών που ενώνουν το στρώμα εισόδου, με το πρώτο κρυφό στρώμα υλοποιείται με ομοιόμορφη κατανομή. Ενώ ως συνάρτηση ενεργοποίησης επιλέγεται η συνάρτηση διορθωμένης γραμμικής μονάδας (ReLU).

Το τέταρτο στρώμα, το στρώμα εξόδου αποτελείται από ένα (1) νευρώνα. Η έξοδος του μοντέλου μας, παράγει μία αριθμητική τιμή σε κάθε εκτέλεση. Ως συνάρτηση ενεργοποίησης επιλέγεται η γραμμική συνάρτηση (linear).

Με την βοήθεια του Tensorboard παρουσιάζουμε γραφικά την αρχιτεκτονική του δικτύου μας.



Σχήμα 11 • Ανατροφοδοτούμενο LSTM Νευρωνικό Δίκτυο Τεσσάρων Στρωμάτων

4.2.1 Δεδομένα Εκπαίδευσης και Ανάπτυξη μοντέλου σε κώδικα

Τα δεδομένα που σε κάθε εποχή που τρέχει το μοντέλο και εκπαιδεύεται παράγονται από την μέθοδο `train_test_split()` της βιβλιοθήκης `sklearn`. Αρχικά θα κανονικοποιήσουμε τα δεδομένα.

```
scalerMm = MinMaxScaler(feature_range=(0,1))
scalerMm_X = scalerMm.fit(X)
scalerMm_y = scalerMm.fit(y)
X = scalerMm_X.transform(X).reshape(302,1,32)
y = scalerMm_y.transform(y)
target_X = scalerMm_X.transform(target_X).reshape(20,1,32)
target_y = scalerMm_y.transform(target_y)
```

Απόσπασμα Κώδικα 7 • Κανονικοποίηση Δεδομένων

Στη συνέχεια θα χωρίσουμε τα δεδομένα σε υποσύνολο εκπαίδευσης και ελέγχου. Τα δεδομένα εξέτασης θα είναι σε κάθε εποχή το είκοσι τοις εκατό (20%) των δεδομένων εισόδου.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

Απόσπασμα Κώδικα 8 • Διαχωρισμός δεδομένων με χρήση μεθόδου `train_test_split()`

Στη συνέχεια δημιουργήσαμε το δίκτυο με την χρήση της βιβλιοθήκης `Keras`.

```

model = Sequential()
model.add(LSTM(10, input_shape=(1,32),return_sequences=True,
kernel_initializer='uniform', activation='relu'))
model.add(LSTM(10, input_shape=(1,32), activation='relu'))
model.add(Dense(1, activation='linear'))

```

Απόσπασμα Κώδικα 9 • Δημιουργία LSTM Νευρωνικού Δικτύου

Έπειτα εκτελούμε το μοντέλο ορίζοντας τη συνάρτηση σφάλματος των δεδομένων εισαγωγής και των δεδομένων εξέτασης σε κάθε εποχή ως το μέσο τετραγωνικό σφάλμα. Σαν αλγόριθμος βελτιστοποίησης επιλέχθηκε ο Adam και μετρική απόδοσης του δικτύου το `soft_accuracy`.

```

model.compile(loss='mean_squared_error',
optimizer='adam',
metrics=[soft_acc])
model.fit(X_train, y_train,
batch_size=nb_batches,
epochs=nb_epochs,
validation_data=(X_test, y_test),
callbacks=callbacks)

```

Απόσπασμα Κώδικα 10 • Εκπαίδευση LSTM Νευρωνικού Δικτύου

Με την μέθοδο `.summary()` της βιβλιοθήκης Keras υπολογίζουμε τις παραμέτρους που διαμορφώνουν το νευρωνικό δίκτυο που κατασκευάστηκε.

```
>>> model.summary()
```

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 1, 10)	1720
lstm_2 (LSTM)	(None, 10)	840
dense_1 (Dense)	(None, 1)	11

=====
 Total params: 2,571
 Trainable params: 2,571
 Non-trainable params: 0

Πίνακας 13 • Παράμετροι Αρχιτεκτονικής LSTM Νευρωνικού Δικτύου

Στους LSTM νευρώνες υπάρχουν τρεις (3) λογικές πύλες και η κατάσταση του κυττάρου. Συνεπώς οι παράμετροι σε κάθε στρώμα για κάθε νευρώνα θα είναι επί τέσσερα και θα δίδονται από τον τύπο $4(nm + n^2 + n)$ όπου n οι νευρώνες και m οι τιμές εισόδου κάθε στρώματος (Dey & Salem, 2017)]. Στο πρώτο κρυφό στρώμα έχουμε δέκα (10) νευρώνες, που δέχονται τριάντα δύο (32) τιμές εισόδου και δέκα (10) τιμές μεροληψίας (bias weights) με σύνολο παραμέτρων 1720. Το

δεύτερο κρυφό στρώμα διαθέτει δέκα (10) νευρώνες, δέκα (10) τιμές εισόδου και δέκα (10) τιμές μεροληψίας, σύνολο 840 παράμετροι. Και τέλος το στρώμα εξόδου είναι απλός νευρώνας με μία (1) τιμή εξόδου, μία (1) τιμή μεροληψίας και δέκα (10) τιμές εισόδου από το προηγούμενο στρώμα σύνολο έντεκα (11). Τελικά για να δομηθεί ένα LSTM δίκτυο της παραπάνω αρχιτεκτονικής χρειάζεται να υπολογιστούν 2.571 παράμετροι.

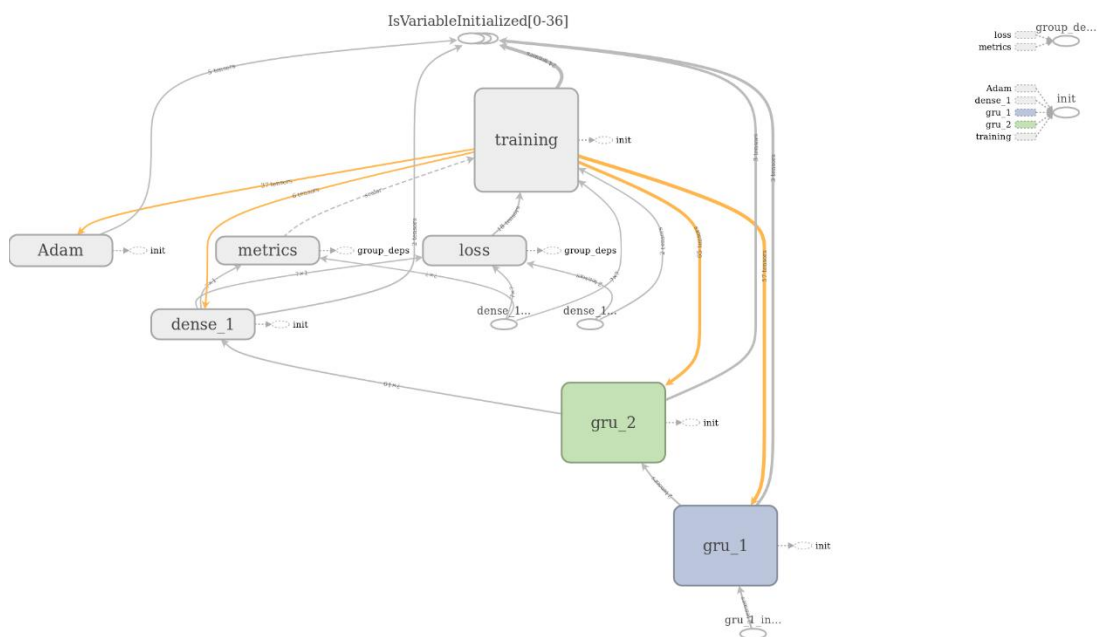
4.3 Ανατροφοδοτούμενο Νευρωνικό Δίκτυο GRU

Η τρίτη υλοποίηση περιλαμβάνει την χρήση των νευρώνων GRU. Το μοντέλο θα αποτελείται από τέσσερα (4) στρώματα. Το στρώμα εισόδου, δύο (2) κρυφά στρώματα και ένα (1) στρώμα εξόδου.

Το πρώτο στρώμα θα δέχεται ως είσοδο μετεωρολογικά στοιχεία σύμφωνα με τα χαρακτηριστικά του πίνακα των δεδομένων μας. Εφόσον έχουμε οκτώ (8) διαφορετικά μετεωρολογικά στοιχεία και έχουμε υπολογίσει τις τιμές τους για τέσσερις (4) συνεχόμενες ημέρες, θα έχουμε τριανταδύο (32) στήλες. Συνεπώς οι νευρώνες του στρώματος εισόδου θα είναι τριάντα δύο (32).

Το δεύτερο και τρίτο στρώμα αποτελούν κρυφά στρώματα νευρώνων. Αποτελούνται από δέκα (10) GRU νευρώνες. Η αρχικοποίηση των βαρών που ενώνουν το στρώμα εισόδου, με το πρώτο κρυφό στρώμα υλοποιείται με ομοιόμορφη κατανομή, ενώ ως συνάρτηση ενεργοποίησης επιλέγεται η συνάρτηση διορθωμένης γραμμικής μονάδας (ReLU).

Το τέταρτο στρώμα, το στρώμα εξόδου αποτελείται από ένα (1) νευρώνα. Η έξοδος του ΤΝΔ παράγει μία αριθμητική τιμή σε κάθε εκτέλεση. Ως συνάρτηση ενεργοποίησης επιλέγεται η γραμμική συνάρτηση (linear). Με την βοήθεια του Tensorboard παρουσιάζουμε γραφικά την αρχιτεκτονική του δικτύου μας.



Σχήμα 12 • Ανατροφοδοτούμενο GRU Νευρωνικό Δίκτυο Τεσσάρων Στρωμάτων

4.3.1 Δεδομένα Εκπαίδευσης και Ανάπτυξη μοντέλου σε κώδικα

Όμοια με την παράγραφο 4.2.1, κανονικοποιούμε τα δεδομένα και τα χωρίζουμε σε δεδομένα εκπαίδευσης και δεδομένα ελέγχου. Με την χρήση του Keras δημιουργούμε τον δίκτυο και το εκπαιδεύουμε.

```

model = Sequential()
model.add(GRU(10, input_shape=(1,32),return_sequences=True,
kernel_initializer='uniform', activation='relu'))
model.add(GRU(10, input_shape=(1,32), activation='relu'))
model.add(Dense(1, activation='linear'))

```

Απόσπασμα Κώδικα 11 • Δημιουργία GRU Νευρωνικού Δικτύου

Έπειτα εκτελούμε το μοντέλο με συνάρτηση λάθους των δεδομένων εισαγωγής και των δεδομένων εξέτασης σε κάθε εποχή ίση με το μέσο τετραγωνικό σφάλμα. Αλγόριθμος βελτιστοποίησης και σε αυτή την περίπτωση επιλέγεται ο Adam και μετρική απόδοσης του δικτύου το `soft_accuracy`.

```

model.compile(loss='mean_squared_error',
              optimizer='adam',
              metrics=[soft_acc])
model.fit(X_train, y_train,
        batch_size=nb_batches,
        epochs=nb_epochs,
        validation_data=(X_test, y_test),
        callbacks=callbacks)

```

Απόσπασμα Κώδικα 12 • Εκπαίδευση GRU Νευρωνικού Δικτύου

Και πάλι με την μέθοδο `.summary()` της βιβλιοθήκης Keras υπολογίζουμε τις παραμέτρους του δικτύου.

```
>>> model.summary()
```

Layer (type)	Output Shape	Param #
gru_1 (GRU)	(None, 1, 10)	1290
gru_2 (GRU)	(None, 10)	630
dense_1 (Dense)	(None, 1)	11

=====
 Total params: 1,931
 Trainable params: 1,931
 Non-trainable params: 0

Πίνακας 14 • Παράμετροι Αρχιτεκτονικής GRU Νευρωνικού Δικτύου

Οι GRU νευρώνες διαθέτουν δύο (2) λογικές πύλες, επομένως ο συνολικός αριθμός των παραμέτρων του μοντέλου να βρίσκεται από τον τύπο $3(n^2 + nm + n)$ όπου n οι νευρώνες και m οι τιμές εισόδου κάθε στρώματος (Dey & Salem, 2017). Το πρώτο κρυφό στρώμα διαθέτει δέκα (10) νευρώνες, που δέχονται τριάντα δύο (32) τιμές εισόδου και δέκα (10) τιμές μεροληψίας (bias weights) με σύνολο παραμέτρων 1290. Το δεύτερο κρυφό στρώμα διαθέτει δέκα (10) νευρώνες, δέκα (10) τιμές εισόδου και δέκα (10) τιμές μεροληψίας, σύνολο 630 παράμετροι. Και τέλος το στρώμα εξόδου είναι απλός νευρώνας με μία (1) τιμή εξόδου, μία (1) τιμή μεροληψίας και δέκα (10) τιμές εισόδου από το προηγούμενο στρώμα σύνολο έντεκα (11). Τελικώς για την δομή του GRU μοντέλου υπολογίζουμε 1.931 παραμέτρους.

5. Αποτελέσματα

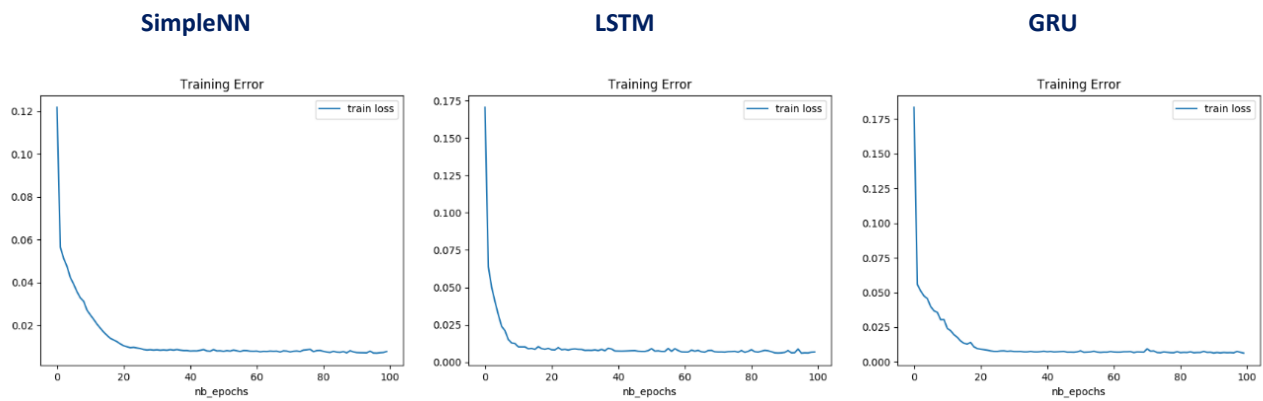
Σε αυτό το κεφάλαιο θα παρουσιάσουμε τα αποτελέσματα των τριών μοντέλων που κατασκευάσαμε γραφικά. Κάθε μοντέλο εκτελέστηκε 5 φορές και οι τιμές που παρουσιάζονται είναι ο μέσος όρος των αποτελεσμάτων.

5.1 Εκπαίδευση

Στα διαγράμματα που ακολουθούν θα παρατηρήσουμε την σταδιακή μείωση του σφάλματος κατά την εκπαίδευση των μοντέλων στον αντίστοιχο αριθμό εποχών. Το σφάλμα υπολογίζεται όπως ορίστηκε στην συνάρτηση (6).

LOSS

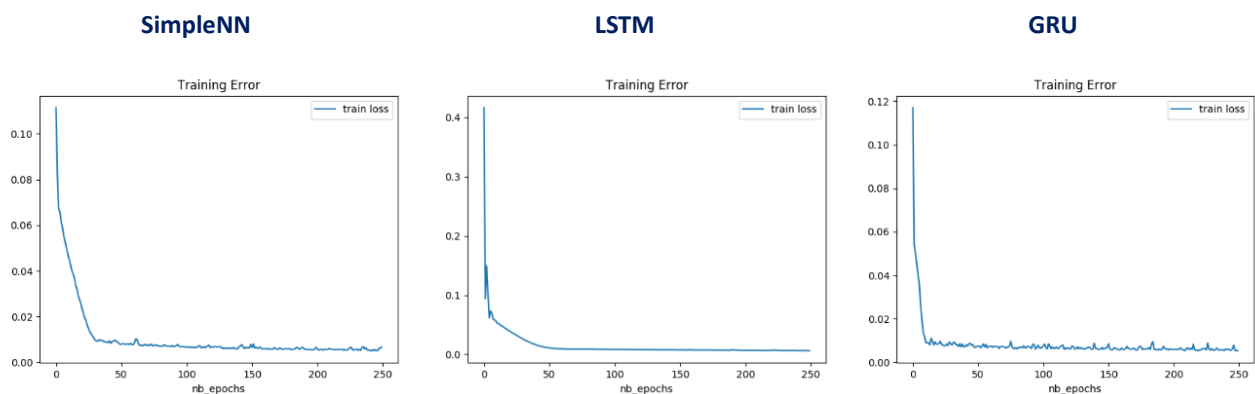
`nb_epochs=100, nb_batches=50, train_set`



Διάγραμμα 4 • Εξέλιξη Σφάλματος στις 100 Εποχές

LOSS

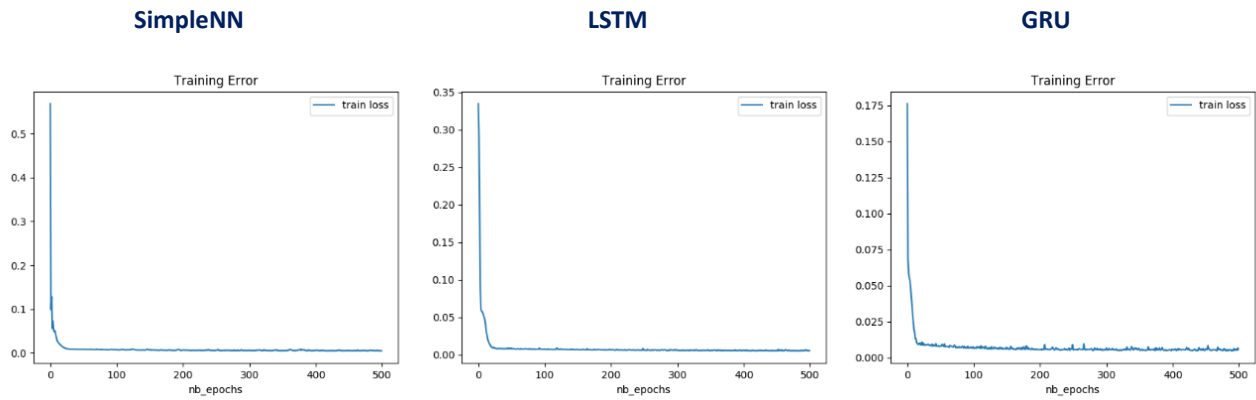
`nb_epochs=250, nb_batches=50, train_set`



Διάγραμμα 5 • Εξέλιξη Σφάλματος στις 250 Εποχές

LOSS

nb_epochs=500, nb_batches=50, train_set

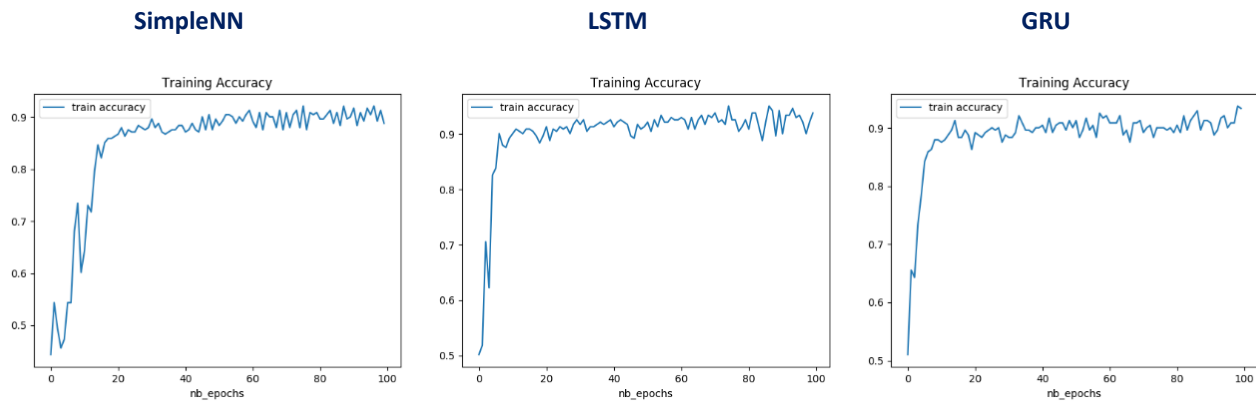


Διάγραμμα 6 • Εξέλιξη Σφάλματος στις 500 Εποχές

Στη συνέχεια παρατηρούμε την αύξηση της μετρικής soft accuracy των μοντέλων για τις αντίστοιχες εποχές. Ο τρόπος λειτουργίας της παρουσιάστηκε στην παράγραφο 1.2.5 .

SOFT ACCURACY

nb_epochs=100, nb_batches=50, train_set



Διάγραμμα 7 • Εξέλιξη Ακρίβειας στις 100 Εποχές

SOFT ACCURACY

nb_epochs=250, nb_batches=50, train_set

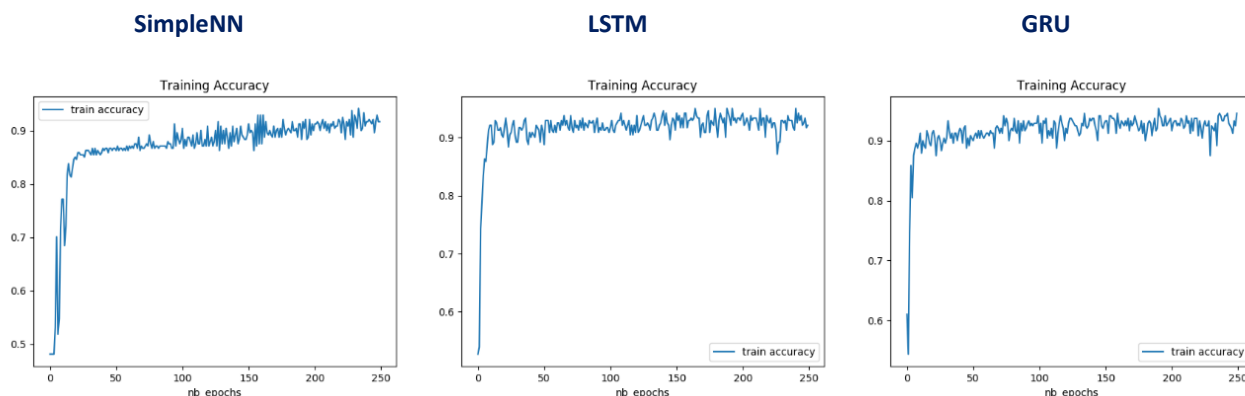
SimpleNN

LSTM

GRU

SOFT ACCURACY

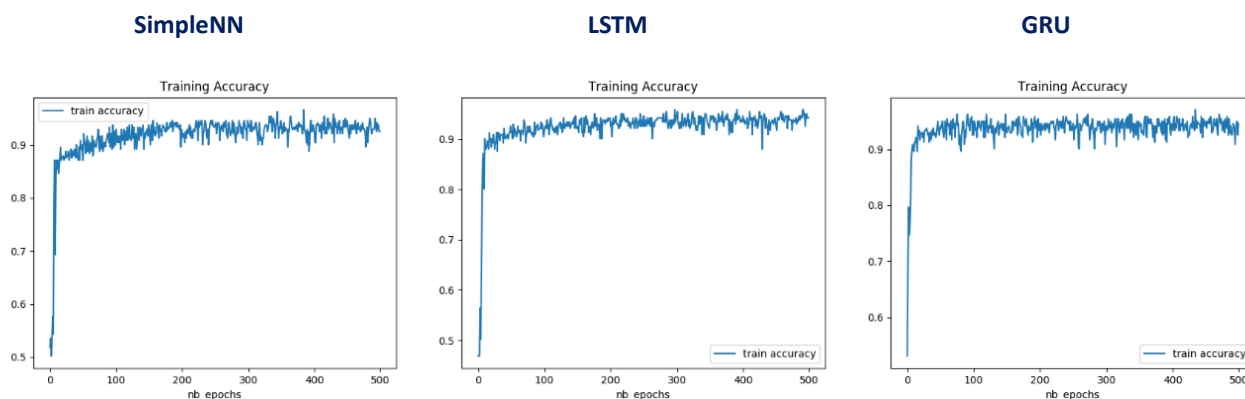
nb_epochs=250, nb_batches=50, train_set



Διάγραμμα 8 • Εξέλιξη Ακρίβειας στις 250 Εποχές

SOFT ACCURACY

nb_epochs=500, nb_batches=50, train_set



Διάγραμμα 9 • Εξέλιξη Ακρίβειας στις 500 Εποχές

5.2 Επιβεβαίωση (Validation)

Κατά την διάρκεια της εκπαίδευσης των μοντέλων, το 20% των δεδομένων ανά εποχή χρησιμοποιούνταν για επιβεβαίωση (validation) βελτίωσης της απόδοσης τους. Στα παρακάτω γραφήματα παρατηρούμε την βελτίωση.

VALIDATION LOSS

nb_epochs=100, nb_batches=50, test_set

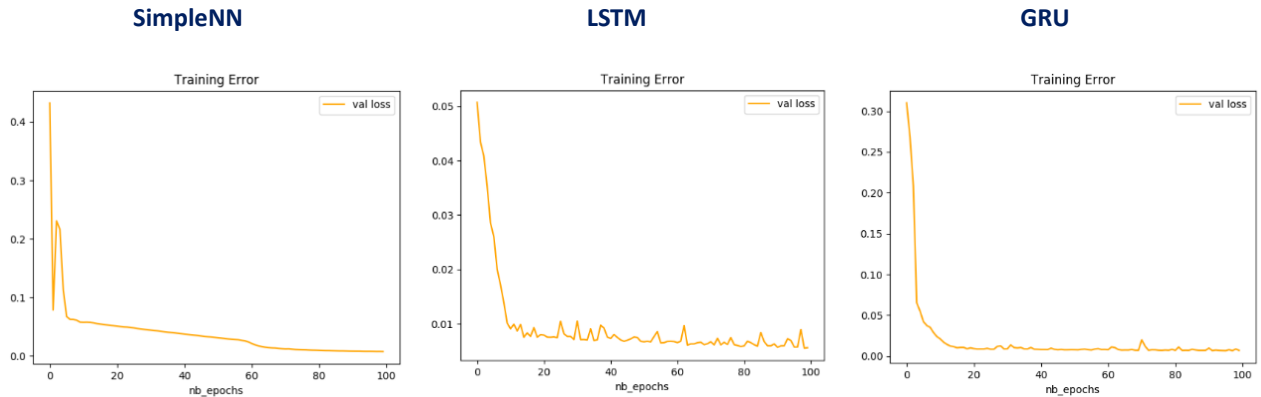
SimpleNN

LSTM

GRU

VALIDATION LOSS

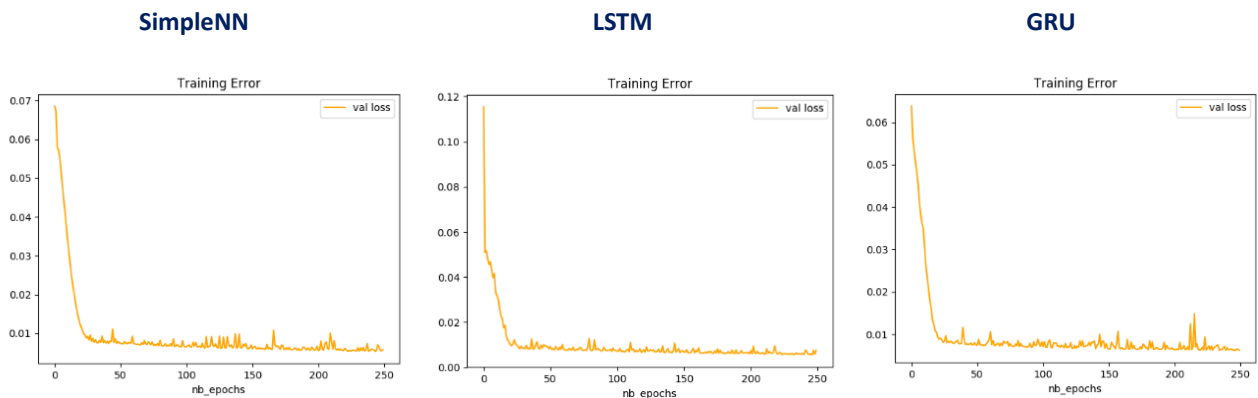
nb_epochs=100, nb_batches=50, test_set



Διάγραμμα 10 • Εξέλιξη Σφάλματος Δεδομένων Επιβεβαίωσης στις 100 Εποχές

VALIDATION LOSS

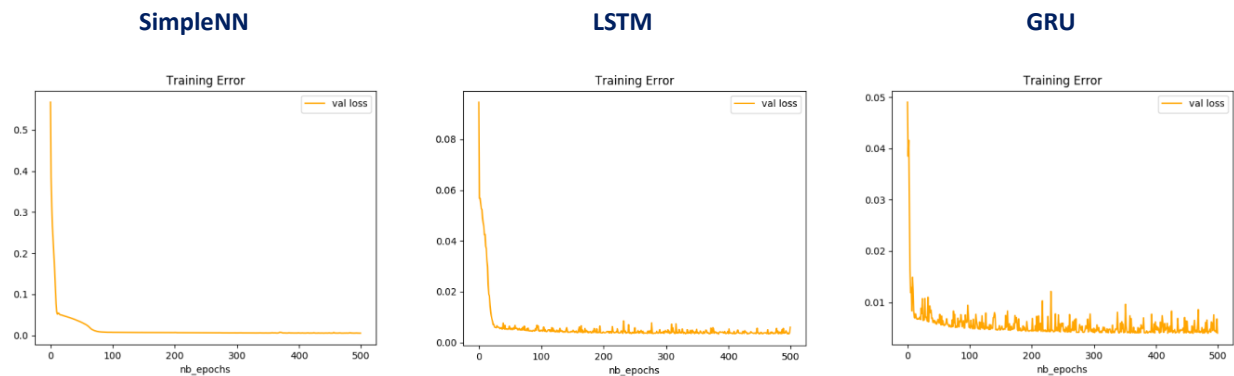
nb_epochs=250, nb_batches=50, test_set



Διάγραμμα 11 • Εξέλιξη Σφάλματος Δεδομένων Επιβεβαίωσης στις 250 Εποχές

VALIDATION LOSS

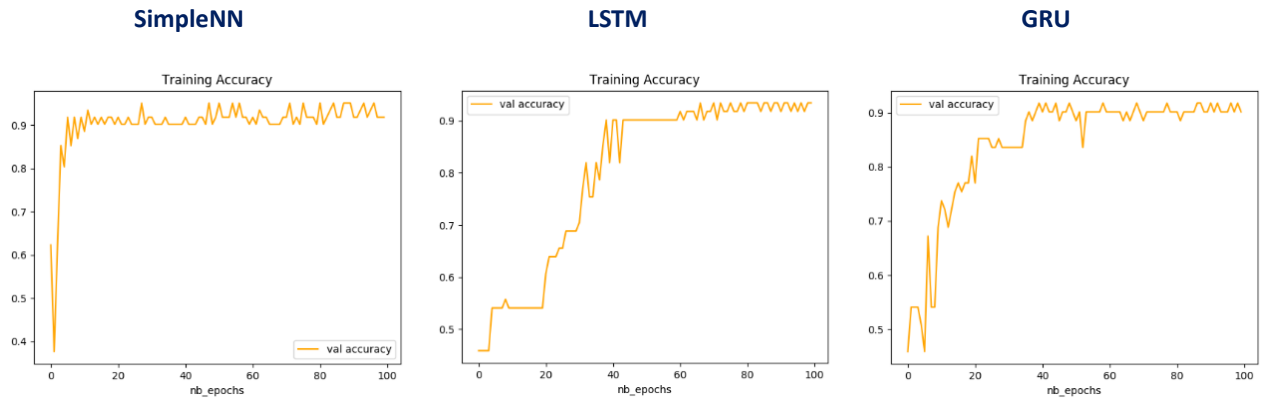
nb_epochs=500, nb_batches=50, test_set



Διάγραμμα 12 • Εξέλιξη Σφάλματος Δεδομένων Επιβεβαίωσης στις 500 Εποχές

VALIDATION SOFT ACCURACY

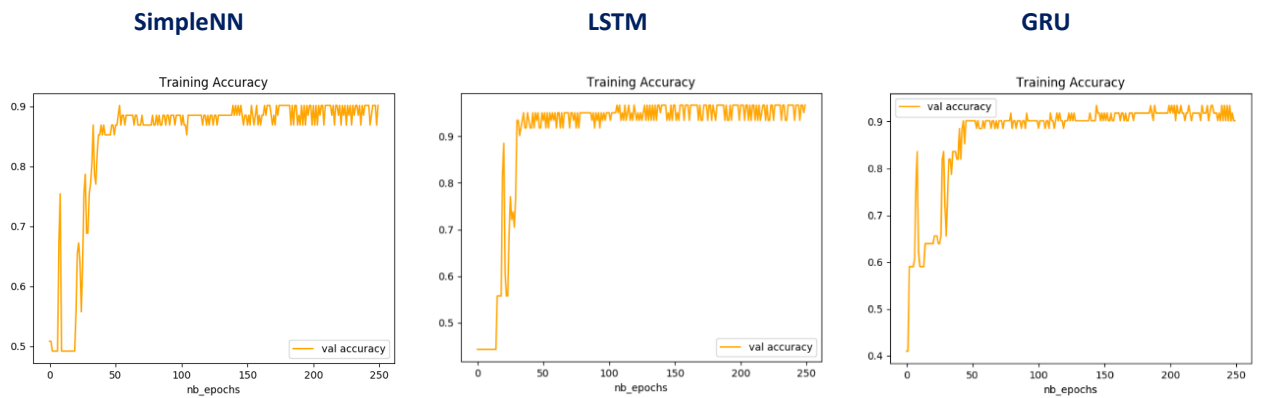
nb_epochs=100, nb_batches=50, test_set



Διάγραμμα 13 • Εξέλιξη Ακρίβειας Δεδομένων Επιβεβαίωσης στις 100 Εποχές

VALIDATION SOFT ACCURACY

nb_epochs=250, nb_batches=50, test_set



Διάγραμμα 14 • Εξέλιξη Ακρίβειας Δεδομένων Επιβεβαίωσης στις 250 Εποχές

VALIDATION SOFT ACCURACY

nb_epochs=500, nb_batches=50, test_set

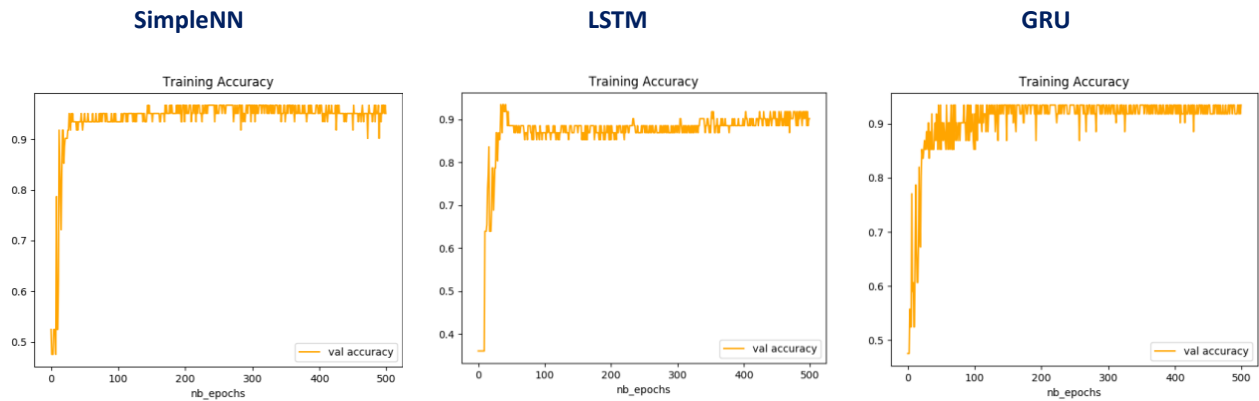
SimpleNN

LSTM

GRU

VALIDATION SOFT ACCURACY

nb_epochs=500, nb_batches=50, test_set



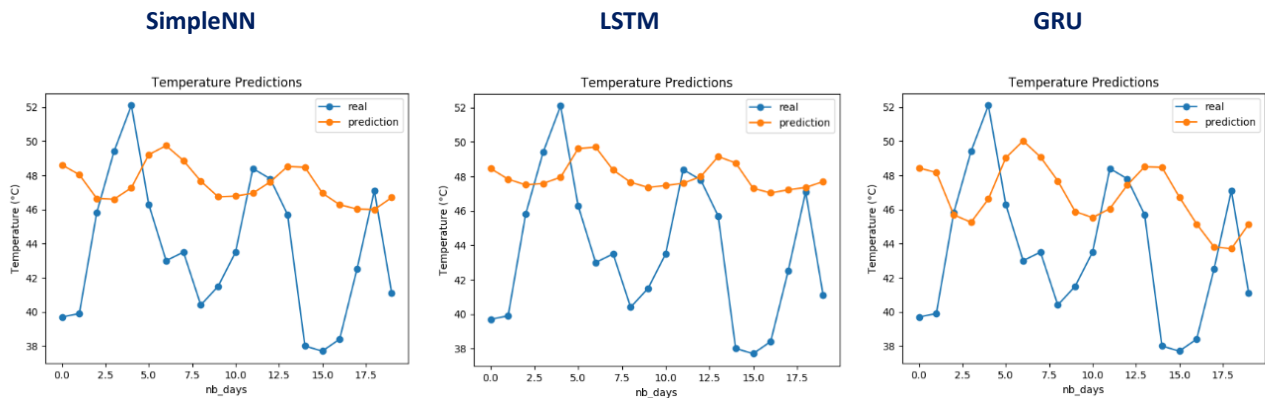
Διάγραμμα 15 • Εξέλιξη Ακρίβειας Δεδομένων Επιβεβαίωσης στις 500 Εποχές

5.3 Πρόβλεψη

Στα διαγράμματα που ακολουθούν θα παρατηρήσουμε την προβλεπτική ικανότητα που έχουν τα μοντέλα για είκοσι ημέρες μετά την εκπαίδευσή τους σε δεδομένα που εισάγονται πρώτη φορά.

PREDICTION

nb_epochs=100, nb_batches=50, target_set



Διάγραμμα 16 • Πρόβλεψη δεδομένων στις 100 Εποχές

PREDICTION

nb_epochs=250, nb_batches=50, target_set

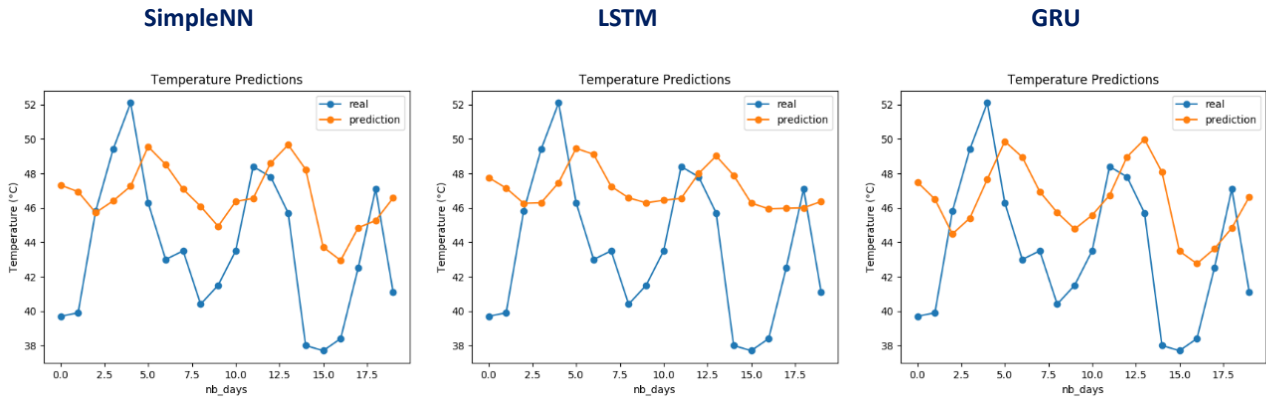
SimpleNN

LSTM

GRU

PREDICTION

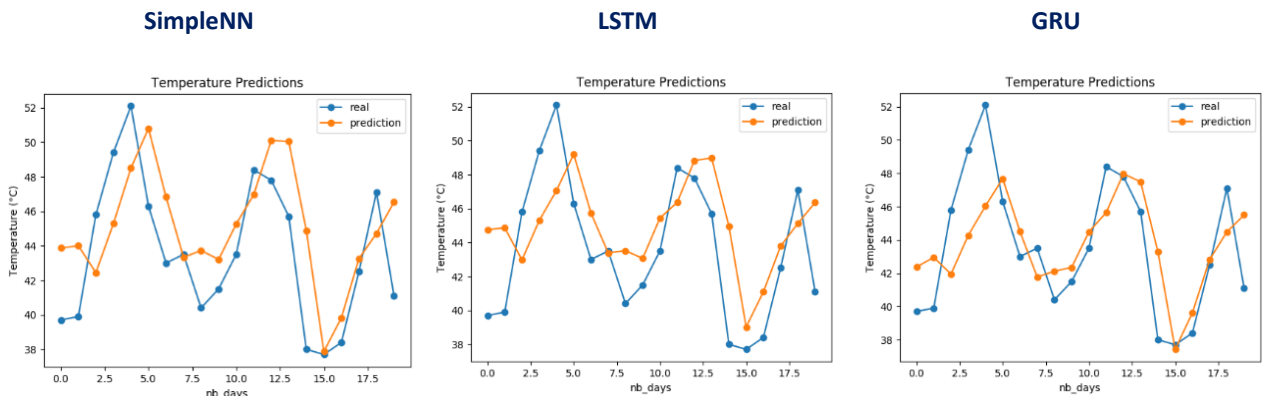
nb_epochs=250, nb_batches=50, target_set



Διάγραμμα 17 • Πρόβλεψη δεδομένων στις 250 Εποχές

PREDICTION

nb_epochs=500, nb_batches=50, target_set



Διάγραμμα 18 • Πρόβλεψη δεδομένων στις 500 Εποχές

5.4 Σύγκριση Αποτελεσμάτων

Σε αυτή την παράγραφο παραθέτουμε τους πίνακες των ευρημάτων μας συγκεντρωτικά.

LOSS			
	100	250	500
SimpleNN	0,0091	0,0068	0,0052
LSTM	0,0063	0,0056	0,0051
GRU	0,0070	0,0055	0,0046

Πίνακας 15 • Συγκεντρωτικό Σφάλμα Μοντέλων

VALIDATION LOSS			
	100	250	500
SimpleNN	0,0088	0,0090	0,0058
LSTM	0,0065	0,0065	0,0064
GRU	0,0057	0,0060	0,0077

Πίνακας 16 • Συγκεντρωτικό Σφάλμα Μοντέλων στα Δεδομένα Επιβεβαίωσης

SOFT ACCURACY (%)			
	100	250	500
SimpleNN	90,82	93,36	93,78
LSTM	92,12	92,95	92,95
GRU	91,33	94,19	93,36

Πίνακας 17 • Συγκεντρωτική Ακρίβεια Μοντέλων

VALIDATION SOFT ACCURACY (%)			
	100	250	500
SimpleNN	93,91	91,80	91,80
LSTM	94,52	88,52	95,08
GRU	92,15	90,16	93,44

Πίνακας 18 • Συγκεντρωτική Ακρίβεια Μοντέλων στα Δεδομένα Επιβεβαίωσης

PREDICTION MAE (°F)			
	100	250	500
SimpleNN	4,165	3,282	2,830
LSTM	4,560	3,825	3,622
GRU	4,100	3,525	3,149

Πίνακας 19 • Μέσο Απόλυτο Σφάλμα Προβλέψεων

6. Συμπεράσματα - Συζήτηση

Στόχος της εργασίας ήταν να προβλέψουμε για τις τελευταίες είκοσι ημέρες του έτους 2018 στην περιοχή της Λαμίας τις μέσες θερμοκρασίες ημέρας με χρήση νευρωνικών μοντέλων. Τα μοντέλα που επιστρατεύτηκαν ήταν ένα απλό νευρωνικό δίκτυο δύο κρυφών στρωμάτων Simple Neural Net, ένα ανατροφοδοτούμενο LSTM δίκτυο δύο κρυφών στρωμάτων και μία εξέλιξη του LSTM, το GRU ανατροφοδοτούμενο δίκτυο με όμοια αρχιτεκτονική.

Παρατηρήσαμε την εξέλιξη των μοντέλων για 100, 250 και 500 εποχές. Αρχικά για την εκπαίδευσή τους υπολογίσαμε την μείωση του σφάλματος στην εκτίμηση των τιμών-στόχων. Οι αλγόριθμοι LSTM και GRU δείχνουν να μαθαίνουν με ταχύτερο ρυθμό καθώς περίπου στις 20 εποχές έχουν φθάσει τις χαμηλότερες τιμές του σφάλματος σε αντίθεση με το SimpleNN που χρειάζεται περισσότερο χρόνο.

Για τον υπολογισμό της επιτυχούς πρόβλεψης τιμής της μέσης θερμοκρασίας κατά την εκπαίδευση χρησιμοποιήσαμε την μετρική soft accuracy. Σε αυτή την “ελαφριά” εκτίμηση ακρίβειας βλέπουμε ότι όλα τα μοντέλα φθάνουν σε σχετικά υψηλές τιμές πρόβλεψης αρκετά γρήγορα. Τις υψηλότερες τιμές, μας τις δίνει το GRU μοντέλο για τις 250 εποχές. Τις χαμηλότερες το SimpleNN για τις 100 εποχές. Ενώ όλα τα μοντέλα έχουν ένα εύρος τιμών που ταλαντώνονται από όταν επιτευχθεί η μέγιστη τιμή τους και έπειτα.

Σε κάθε εποχή εκπαίδευσης κρατούσαμε το 20% των δεδομένων για επιβεβαίωση (validation) της μάθησης σε δεδομένα που δεν έχουν δει τα μοντέλα μας. Με αυτόν τον τρόπο καταγράφουμε μία πιο ρεαλιστική εξέλιξη της μάθησης. Πράγματι αυτό αποτυπώνεται στον υπολογισμό του σφάλματος από τα γραφήματα καθώς τα μοντέλα έχουν μεγαλύτερες αποκλίσεις στις πρώτες εποχές και στη συνέχεια συγκλίνουν σε χαμηλές τιμές. Οι οποίες ωστόσο δεν είναι τόσο μικρές όσο στον υπολογισμό του σφάλματος των δεδομένων εκπαίδευσης γεγονός που σημαίνει πως αποφεύγεται η υπερπροσαρμογή του δικτύου στο συγκεκριμένο τύπο δεδομένων.

Όμοια αποτελέσματα καταγράφονται και στον υπολογισμό της ακρίβειας των δεδομένων επιβεβαίωσης. Παρατηρείται η μείωση στην συνολική απόδοση των δεδομένων σε σχέση με τα δεδομένα εκπαίδευσης. Ενώ σε αντίθεση με τις μέχρι στιγμής επισημάνσεις μας, όπου τα μοντέλα LSTM και GRU έδειχναν μία ομαλότερη μετάβαση στις τιμές σύγκλισης, στο συγκριμένο κριτήριο σε όλες τις εποχές το SimpleNN παρουσιάζει μία σταθερότητα ως προς την απόδοση.

Δίδεται η αίσθηση ότι το εκάστοτε μοντέλο γενικεύει αρκετά καλά τα δεδομένα που καλείται να προβλέψει. Οπότε καλούμαστε να προβλέψουμε τις τιμές της μέσης θερμοκρασίας και να αξιολογήσουμε την εκπαίδευση.

Από τον πίνακα υπολογισμού του μέσου απόλυτου σφάλματος αρχικά παρατηρούμε ότι υπάρχει απόκλιση στις τιμές που προβλέφθηκαν από 4,6 έως 2,8 βαθμούς την κλίμακας Fahrenheit. Στη συνέχεια γίνεται αντιληπτό πως όσο περισσότερες εποχές μάθησης επιτρέψουμε στα μοντέλα τόσο καλύτερη πρόβλεψη πραγματοποιείται. Στις 500 εποχές επομένως καταγράφεται η καλύτερη απόδοση.

Ακόμη, στις προβλεφθείσες τιμές εμφανίζεται μία συνέπεια όσον αφορά την μεταβολή των πραγματικών τιμών. Εφόσον καταφέρνουν να ακολουθήσουν τις διακυμάνσεις από την μία ημέρα στην άλλη αποδίδοντας την αλλαγή μικροκλίματος που ενδεχομένως να συμβαίνει.

Τέλος ένας σημαντικός παράγοντας για την απόδοση ενός μοντέλου είναι η υπολογιστική δύναμη που χρειάζεται σε συνδυασμό με την καλύτερη ποιότητα προβλέψεων. Όπως αναφέραμε στο κεφάλαιο 4 κάθε υλοποίηση έχει διαφορετικό αριθμό παραμέτρων για την εκτέλεση και εξαγωγή αποτελεσμάτων. Στην υλοποίηση SimpleNN υπολογίστηκαν 451 παράμετροι, στον LSTM και στον GRU 2.571 και 1.931 αντίστοιχα. Επομένως η απλούστερη υλοποίηση του SimpleNN χρειάζεται και την λιγότερη υπολογιστική δύναμη.

Συνολικά λοιπόν καταλήγουμε πως η υλοποίηση SimpleNN είναι η πιο αποδοτική από άποψη καλύτερης πρόβλεψης θερμοκρασιών, καλύτερης απεικόνισης της ημερήσιας διακύμανσης των θερμοκρασιών και της μικρότερης χρήσης υπολογιστικών πόρων.

Παρότι θα περιμέναμε οι αλγόριθμοι LSTM και GRU να έχουν σημαντικά καλύτερη απόδοση καθότι διαμορφώνονται με σκοπό της επεξεργασία χρονικά μεταβαλλόμενων δεδομένων η πιο απλή υλοποίηση νευρωνικών δικτύων αποδίδει καλύτερα στην υπόθεση εργασίας μας.

6.1 Μελλοντική Έρευνα

Η παρούσα εργασία αποτελεί μία αρχική προσέγγιση στην προσπάθεια πρόγνωσης μετεωρολογικών στοιχείων με την χρήση μηχανικής μάθησης. Οι τεχνικές υλοποίησης δικτύων νευρώνων που αναπτύσσονται μέχρι και σήμερα αποτελούν εργαλεία χρήσιμα για πειραματισμό και εξερεύνηση των απειριορίστων δυνατοτήτων τους, στην πρόβλεψη και στην διαχείριση μεγάλου όγκου δεδομένων.

Παρουσιάσαμε αλγόριθμους που αξιοποιούν δεδομένα χρονικά μεταβαλλόμενα και τους παραμετροποιήσαμε σε βαθμό που να μπορεί να εξαχθεί χρήσιμη πληροφορία από την σύγκριση μεταξύ τους. Ο μελλοντικός ερευνητής μπορεί να χρησιμοποιήσει τα αποτελέσματα αυτής της εργασίας για την περαιτέρω διερεύνηση προβλημάτων πρόγνωσης καιρού.

Η επιστημονική κοινότητα ανακαλύπτει νέες μεθόδους που εστιάζουν στην διατήρηση πληροφορίας εντός της επεξεργαστικής ικανότητας των αλγορίθμων, όμως εξίσου σημαντική αποδεικνύεται η μορφοποίηση των δεδομένων που θα εισαχθούν σε αυτούς.

Τέλος, η επεξεργαστική δύναμη των αλγορίθμων επιτρέπει την αναγνώριση προτύπων σε ποικίλου τύπου δεδομένα. Επομένως μπορεί να εξεταστεί η εισαγωγή και άλλων τύπων

δεδομένων που παρατηρούνται στην ατμόσφαιρα, όπως και το μεγαλύτερο χρονικό εύρος τους.

Βιβλιογραφία

- Agarap, A. F. (2018). *Deep Learning using Rectified Linear Units (ReLU)*. Manila, Metro Manila, Philippines: De La Salle University.
- Argyris, P. (2019). *Computational Physics Group A.U.Th.* . Ανάκτηση από <http://kelifos.physics.auth.gr/COURSES/neural/K1.pdf>
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Gated Feedback Recurrent Neural Networks*. University of Montreal.
- Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks*. *Machine Learning*, 273-297.
- Dey, R., & Salem, F. M. (2017). *Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks* . *Circuits, Systems, and Neural Networks (CSANN) LAB Department of Electrical and Computer Engineering Michigan State University* .
- Download PyCharm . (2019). Ανάκτηση από <https://www.jetbrains.com/pycharm/download/#section=linux>
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1991). *Knowledge Discovery in Databases: An Overview*. *AI Magazine*.
- Get pandas! (2019). Ανάκτηση από <https://pandas.pydata.org/getpandas.html>
- Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-term Memory*. *Neural Computation*, 1735-1780.
- Holmstrom, M., Liu, D., & Vo, C. (2016). *Machine Learning Applied to Weather Forecasting*. Stanford University.
- Hosch, W. L. (2009). *Machine learning*. *Encyclopedia Britannica*.
- Keras 2.2.4. (2019). Ανάκτηση από <https://pypi.org/project/Keras/>
- Maimon, O. Z., & Rokach, L. (2008). *Data Mining with Decision Trees: Theory and Applications (Machine Perception and Artificial Intelligence)* . World Scientific Pub Co Inc.
- Matplotlib . (2019). Ανάκτηση από <https://matplotlib.org/downloads.html>
- McCarthy, J. (2007). *WHAT IS ARTIFICIAL INTELLIGENCE?* Computer Science Department, Stanford University.
- McQuistan, A. (2017, November 14). *Using Machine Learning to Predict the Weather: Part 1*. Ανάκτηση από <https://stackabuse.com/using-machine-learning-to-predict-the-weather-part-1/>
- Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. MIT Press.
- Muggleton, S., & Lucde, R. (1994). *Inductive Logic Programming: Theory and Methods*. *The Journal of Logic Programming*, 629-679.
- NCDC NOAA Official Website. (2019). Ανάκτηση από <https://www.ncdc.noaa.gov/>
- NESDIS NOAA Official Website. (2019). Ανάκτηση από <https://www.nesdis.noaa.gov/>
- NumPy. (2019). Ανάκτηση από <http://www.numpy.org/>
- Obtaining SciPy libraries. (2019). Ανάκτηση από <https://www.scipy.org/scipylib/download.html>
- Olah, C. (2015, August 27). *Understanding LSTM Networks*. Ανάκτηση από colah's blog: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Pearl, J. (2000). *Causality: Models, Reasoning and Inference* . Cambridge University Press.
- Python 3.6.8. (2019). Ανάκτηση από <https://www.python.org/downloads/release/python-368/>
- Russell, S. J., & Norvig, P. (2003). *Artificial Intelligence : A Modern Approach*. Malaysia; Pearson Education Limited.
- scikit-learn 0.20.3. (2019). Ανάκτηση από <https://pypi.org/project/scikit-learn/>
- Simon, P. (2013). *Too Big to Ignore: The Business Case for Big Data*. Wiley and SAS Business Series.
- TensorFlow Core Official Website. (2019). Ανάκτηση από https://www.tensorflow.org/guide/summaries_and_tensorboard
- U.S. Department of Commerce Official Website. (2019). Ανάκτηση από

<https://www.commerce.gov/>
Ubuntu. (2019). Download Ubuntu Desktop. Ανάκτηση από
<https://www.ubuntu.com/download/desktop>
USAF. (n.d.). Climatology Center. Ανάκτηση από
https://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=nasa_ancillary&KeywordPath=DataCenters%7CGOVERNMENT+AGENCIES-U.S.+FEDERAL+AGENCIES%7CDOD%7CDOD%2FUSAF%2FAFCCC&OrigMetadataNode=GCMD&EntryId=USAFETAC_SATMO&MetadataView=Full&MetadataType=0&lbnode=mdlb5
WMO. (2019). Resolution 40. Ανάκτηση από
http://www.wmo.int/pages/prog/hwrrp/documents/wmo_827_enCG-XII-Res40.pdf
WMO Official Website. (2019). Ανάκτηση από <https://public.wmo.int/en>
Λήψη Libre Office. (2019). Ανάκτηση από <https://el.libreoffice.org/download/>
NOAA Official Website. (2019). Ανάκτηση από <https://www.noaa.gov/>