ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Αναγνώριση φωνητικών εντολών με νευρωνικά δίκτυα

Διπλωματική Εργασία

Κεράστας Ιωάννης

Επιβλέπουσα: Τσαλαπάτα Χαρίκλεια

Βόλος 2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Αναγνώριση φωνητικών εντολών με νευρωνικά δίκτυα

Διπλωματική Εργασία

Κεράστας Ιωάννης

Επιβλέπουσα: Τσαλαπάτα Χαρίκλεια

Βόλος 2021

UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Voice command recognition using neural networks

Diploma Thesis

Kerastas Ioannis

Supervisor: Tsalapata Hariklia

Volos 2021

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπουσα **Τσαλαπάτα Χαρίκλεια**
Ε.ΔΙ.Π, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος **Δασκαλοπούλου Ασπασία**
Επίκουρος Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος **Σταμούλης Γεώργιος**
Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Ημερομηνία έγκρισης: 25-2-2021

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά την καθηγήτρια κυρία Χαρίκλεια Τσαλαπάτα καθώς και τους συνεπιβλέποντες καθηγητές Ασπασία Δασκαλοπούλου και Γεώργιο Σταμούλη.Τέλος,θα ήθελα να ευχαριστήσω την οικογένεια μου για την στήριξη τους όλα αυτά τα χρόνια.

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ

«Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων, αναφέρονται 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 Δηλών

Κεράστας Ιωάννης
15-2-2021

Περίληψη

Στην σημερινή εποχή, και ειδικότερα με την αύξηση των έξυπνων συσκευών, απαιτούνται συστήματά αναγνώρισης φωνητικών εντολών. Η διπλωματική εργασία μελετάει συστήματα αναγνώρισης φωνητικών εντολών που βασίζονται στην μηχανική μάθηση. Πιο συγκεκριμένα δοκιμάζουμε μοντέλα που χρησιμοποιούν αναδρομικά τεχνητά νευρωνικά δίκτυα ή συνελικτικά τεχνητά νευρωνικά δίκτυα και επιπλέον μελετάμε τρόπους προ-επεξεργασίας δεδομένων ήχου, ώστε να εξάγουμε χαρακτηριστικά που θα δοθούν στα μοντέλα μας. Στο τέλος συνοψίζουμε τα αποτελέσματά μας και παρουσιάζουμε το καλύτερο μοντέλο.

Abstract

Nowadays, especially with the rise of smart devices, systems that require accurate voice recognition are needed. This thesis studies voice recognition systems based on machine learning. More specifically, we experiment with models that make use of recurrent neural networks and convolutional neural networks and in addition we study way of preprocessing audio data, in order to extract features for our models. In the end, we summarize our results and present the best model.

Πίνακας περιεχομένων

Κατάλογος σχημάτων

Κατάλογος πινάκων

Συντομογραφίες

ANN	Artificial Neural Network
LSTM	Long Short-Term Memory
MLP	multilayer Perceptron
GRU	Gated Recurrent Unit
CNN	Convolutional Neural Network
DNN	Deep Neural Network
εικ	εικόνα
πιν	πίνακας
εξ	εξίσωση

Κεφάλαιο 1

Εισαγωγή

1.1 Αντικείμενο της διπλωματικής

Στην παρούσα διπλωματική θα αναφερθούν τρόποι και μέθοδοι αναγνώρισης φωνητικών εντολών με την χρήση μηχανικής μάθησης. Στην καθημερινότητα μας συναναστρεφόμαστε με όλο και περισσότερες ευφυείς συσκευές και εφαρμογές που έχουν σκοπό να μας διευκολύνουν την ζωή. Συνεπώς απαιτείται ένας εύκολος τρόπος ώστε να πραγματοποιούν τις ενέργειες που επιθυμούμε. Ο άνθρωπος από την φύση του επικοινωνεί κυρίως με την φωνή του επομένως πρέπει να βρούμε ένα τρόπο να δίνουμε εντολές στις συσκευές μέσω της φωνής μας.

Στο παρελθόν δοκιμάστηκαν πολλές μέθοδοι και τεχνικές για αναγνώριση φωνητικών εντολών, όπως η χρήση κρυπτομαρκοβιανών μοντέλων (Hidden Markov Models) [14], όμως τα τελευταία χρονιά γίνεται εκτεταμένη μελέτη και έρευνα στο πεδίο της μηχανικής μάθησης και λόγω τις αύξησης της υπολογιστικής ισχύος αλλά και τις ικανότητας της να δίνει λύση σε πολλά προβλήματα, μπορεί να εφαρμοστεί και στο πρόβλημα αναγνώρισης φωνητικών εντολών. Η αναγνώριση φωνητικών εντολών με την χρήση μηχανικής μάθησης είναι δύσκολο πρόβλημα εξαιτίας της μορφής των δεδομένων μας. Τα δεδομένα ήχου πρέπει να περάσουν μερικά στάδια προεπεξεργασίας, προκειμένου να έχουν κατάλληλη μορφή για είσοδο σε συστήματα μηχανικής μάθησης. Στην συνέχεια πρέπει να δημιουργηθεί το κατάλληλο μοντέλο το οποίο θα δέχεται τα δεδομένα λαμβάνοντας υπόψη και την χρονική συσχέτιση που υπάρχει στα δεδομένα ήχου.

Σχετικές εργασίες

Έχουν γίνει παρόμοιες εργασίες για την αναγνώριση φωνητικών εντολών οι οποίες χρησιμοποιούσαν πολλές και διαφορετικές μεθόδους. Μια από αυτές τις μεθόδους είναι με την χρήση κρυπτομαρκοβιανών μοντέλων [15] ή με την χρήση ασαφής λογικής [16]. Επιπλέον έχουν γίνει πολλές απόπειρες για να λυθεί αυτό το πρόβλημα με την χρήση νευρωνικών δικτύων, με το οποίο θα ασχοληθούμε και εμείς. Μερικές από αυτές τις προσπάθειες έγιναν χωρίς τα δεδομένα του ήχου να υποστούν κάποια προ-επεξεργασία. [17] και είχαν ικανοποιητικά αποτελέσματά. Κάποιες άλλες όμως ακολούθησαν διαφορετική κατεύθυνση και εξήγαγαν χαρακτηριστικά από τον ήχο μέσω διάφορων μετασχηματισμών [18], [19] με την απόδοση τους να είναι καλύτερη σε σχέση με τις προσπάθειες που δεν έγινε προ-επεξεργασία στα δεδομένα ήχου. Και οι δυο διαφορετικές κατευθύνσεις, δηλαδή της προ-επεξεργασίας ή όχι των δεδομένων, χρησιμοποίησαν αναδρομικά ή συνελικτικά τεχνητά δίκτυα ή και κάποιον συνδυασμό τους.

1.1.1 Συνεισφορά

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Στην αρχή έγινε εκτενής επισκόπηση του πεδίου της μηχανικής μάθησης και πιο συγκεκριμένα μελετήθηκαν τα τεχνητά νευρωνικά δίκτυα.

2. Έπειτα έγινε εστίαση στην εξαγωγή χαρακτηριστικών από δεδομένα ήχου και συγκεκριμένα αναλύθηκε ο μετασχηματισμός Fourier σύντομου χρόνου (STFT) για την παραγωγή φασμοτογραφήματος από ήχο.

3. Υλοποιήθηκαν δύο βασικά μοντέλα που βασίζονται σε τεχνητά νευρωνικά δίκτυα. Το ένα από αυτά βασίζεται σε αναδρομικά νευρωνικά δίκτυα ενώ το άλλο σε συνελικτικά νευρωνικά δίκτυα.

4. Πραγματοποιήθηκαν δοκιμές, αλλάζοντας διάφορες παραμέτρους στα μοντέλα μας, ώστε να βρεθεί το καταλληλότερο μοντέλο για τα δεδομένα μας.

5. Τέλος συνοψίσαμε τα αποτελέσματα μας καταλήγοντας στα αντίστοιχα συμπεράσματα.

1.2 Οργάνωση του τόμου

Η παρούσα διπλωματική χωρίζεται σε πέντε κεφάλαια. Στο Κεφάλαιο 2 γίνεται ανάλυση του τομέα της μηχανικής μάθησης όπου αναλύονται απαραίτητες έννοιες για την κατανόηση των μοντέλων που παρουσιάζονται. Στην συνέχεια στο Κεφάλαιο 3 γίνεται παρουσίαση των δεδομένων που θα χρησιμοποιήσουμε και αναλύεται ο μετασχηματισμός Fourier και παραλλαγές του όπου θα χρησιμοποιηθούν για εξαγωγή χαρακτηριστικών από τα δεδομένα μας. Στο Κεφάλαιο 4 παρουσιάζεται η βασική αρχιτεκτονική των μοντέλων που αναπτύξαμε και τα αποτελέσματα τους. Τέλος στο Κεφάλαιο 5 παρατίθενται συγκεντρωτικά τα αποτελέσματα από τα μοντέλα μας και τα συμπεράσματά μας.

Κεφάλαιο 2

Εισαγωγή στην μηχανική μάθηση

Η επιστήμη των υπολογιστών είναι πολυδιάστατη και χωρίζεται σε διάφορα υποπεδία. Ένα υποπεδίο είναι η μηχανική μάθηση που αναπτύχθηκε από την μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. Ορίζεται ως η ικανότητα ενός υπολογιστικού συστήματος να δημιουργεί μοντέλα ή πρότυπα από ένα σύνολο δεδομένων. Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Οι αλγόριθμοι αυτοί κατασκευάζουν μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις ή να εξάγουν αποφάσεις.

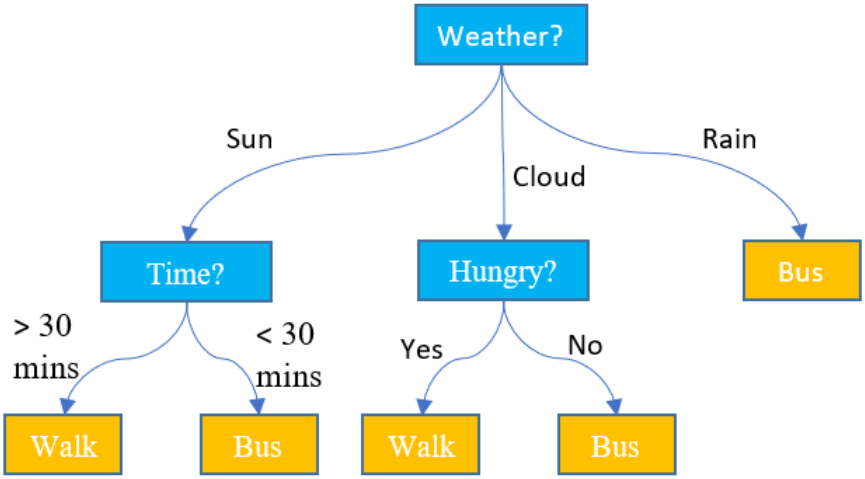
2.1 Μοντέλα μηχανικής μάθησης

Υπάρχουν πολλά μοντέλα και αλγόριθμοι μηχανικής μάθησης οπού έχουν εφαρμογή σε διάφορα προβλήματα και μερικοί παρουσιάζονται παρακάτω.

2.1.1 Δέντρα απόφασης

Τα δέντρα αποφάσεων έχουν επηρεάσει μια ευρεία περιοχή της μηχανικής μάθησης καλύπτοντας τόσο την ταξινόμηση όσο και την παλινδρόμιση [20]. Στην ανάλυση αποφάσεων ένα δέντρο μπορεί να χρησιμοποιηθεί για την οπτική απεικόνιση των αποφάσεων και της λήψης αποφάσεων. Με τα δέντρα αποφάσεων προσπαθούμε να απλοποιήσουμε ένα πρόβλημα και να παρουσιάσουμε τις σημαντικότερες ενέργειες και γεγονότα που περιλαμβάνει. Οι ενέργειες για τις οποίες είμαστε σίγουροι ότι δεν πρέπει να γίνουν, δεν θα περιληφθούν στο δέντρο καθώς θα περιπλέκουν την ανάλυση του προβλήματος. Τα βασικά πλεονεκτήματα

της ανάλυσης των δέντρων αποφάσεων είναι: Παρουσιάζουν με απλότητα και σαφήνεια το πρόβλημα και τις πιθανές του εκβάσεις και μπορούν να εφαρμοστούν σε πολλά διαφορετικά είδη προβλημάτων. Διευκολύνει την προσθήκη άλλων πιθανών εκβάσεων, καθώς μπορεί να γίνει περαιτέρω ανάπτυξη ορισμένων κλάδων εκ των υστέρων.



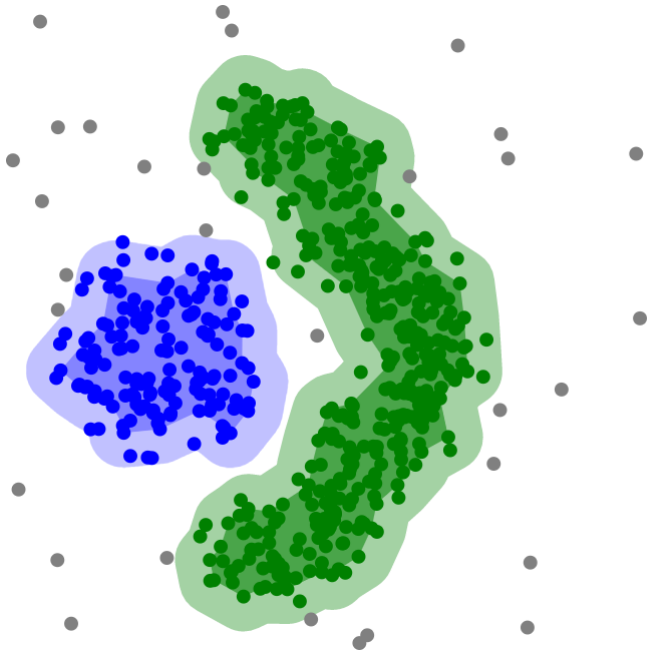
Σχήμα 2.1: Ένα δέντρο απόφασης. [1]

2.1.2 Ομαδοποίηση

Ομαδοποίηση (clustering) είναι μια κατηγορία αλγορίθμων όπου διαχωρίζουν δεδομένα σε υποσύνολα που ονομάζονται ομάδες (clusters). Ο διαχωρισμός γίνεται σύμφωνα με ορισμένα κριτήρια, οπού χρησιμοποιούνται για τον υπολογισμό μιας απόστασής, με σκοπό να ομαδοποιήσουμε δεδομένα της ίδιας ομάδας κοντά ενώ από τις υπόλοιπές ομάδες μακριά. Είναι μέθοδος μη επιβλεπόμενης μάθησης η οποία χρησιμοποιείται επίσης και στην στατιστική ανάλυση δεδομένων.

2.1.3 Μηχανές διανυσμάτων υποστήριξης

Είναι μια ομάδα αλγορίθμων επιβλεπόμενης μάθησης που χρησιμοποιούνται κυρίως για την επίλυση προβλημάτων ταξινόμησης αλλά και σπανιότερα παλινδρόμησης [21], [22]. Η λειτουργία των μηχανών διανυσμάτων υποστήριξης βασίζεται στην εύρεση ενός υπερεπιπέδου που διαχωρίζει τα δεδομένα μας βρίσκοντας το μέγιστο περιθώριο από αυτά, δηλαδή

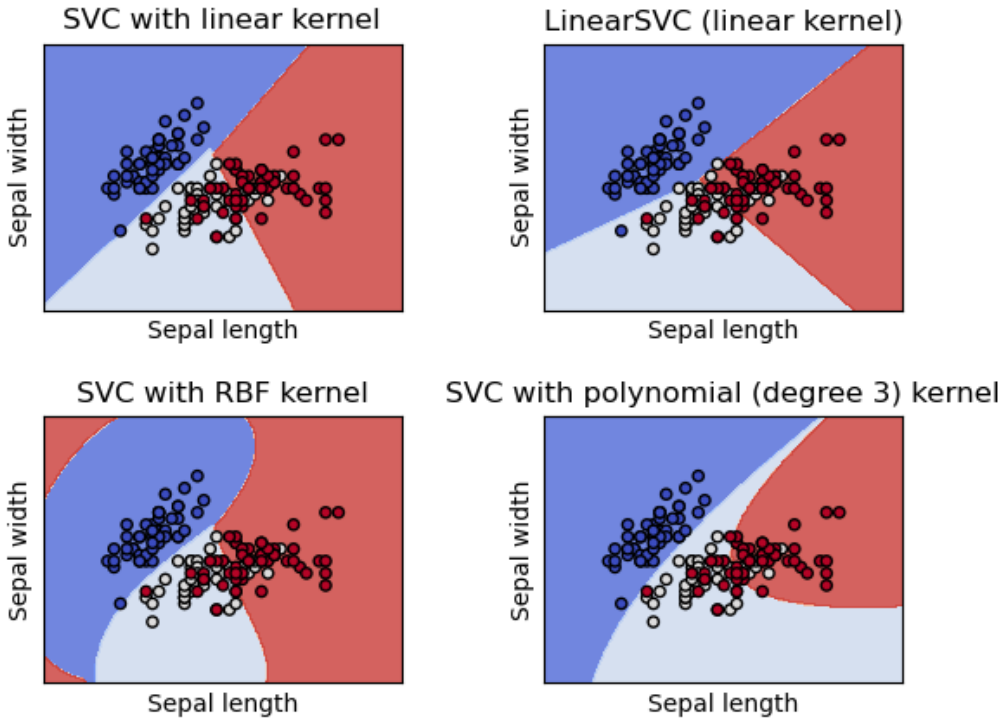


Σχήμα 2.2: Ομαδοποίηση με βάση την πυκνότητα. [2]

ένα υπερεπίπεδο όπου έχει την μέγιστη απόσταση από όλες τις κλάσεις των δεδομένων μας. Κάθε μηχανή διανυσμάτων υποστήριξης έχει την δυνατότητα κατηγοριοποίησης σε δύο κλάσεις, αλλά υπάρχουν παραλλαγές στον αλγόριθμο όπου μας επιτρέπουν να τον εφαρμόσουμε σε παραπάνω αριθμό κλάσεων. Ένα αρνητικό χαρακτηριστικό των μηχανών διανυσμάτων υποστήριξης είναι ότι τα δεδομένα μας θα πρέπει να είναι γραμμικά διαχωρίσιμα ώστε να λειτουργεί ο αλγόριθμος, όμως με την χρήση κατάλληλων απεικονίσεων μπορούμε να μεταφέρουμε και μη γραμμικά διαχωρίσιμα δεδομένα σε κάποια μεγαλύτερη διάσταση όπου εκεί θα είναι γραμμικά διαχωρίσιμα.

2.1.4 Τεχνητά νευρωνικά δίκτυά

Ένα τεχνητό νευρωνικό δίκτυο (Artificial Neural Network) είναι ένα υπολογιστικό σύστημα υλικού και λογισμικού του οποίου η δομή και η λειτουργία είναι εμπνευσμένη από τον τρόπο λειτουργίας των βιολογικών νευρικών δικτύων, τα οποία αποτελούν δομικά συστατικά των εγκεφάλων των ζώων και των ανθρώπων (εικ. 2.4, 2.5). Περιλαμβάνεται από απλούς υπολογιστικούς κόμβους (νευρώνες) διασυνδεδεμένους μεταξύ τους. Τα δομικά στοιχεία του νευρωνικού δικτύου είναι οι νευρώνες, κάθε ένας από τους οποίους δέχεται ένα σύνολο αριθμητικών εισόδων από διαφορετικές πηγές(περιβάλλον,άλλους νευρώνες), πραγματοποιεί έναν υπολογισμό με βάση τις εισόδους που έλαβε και παράγει μια έξοδο. Η έξο-



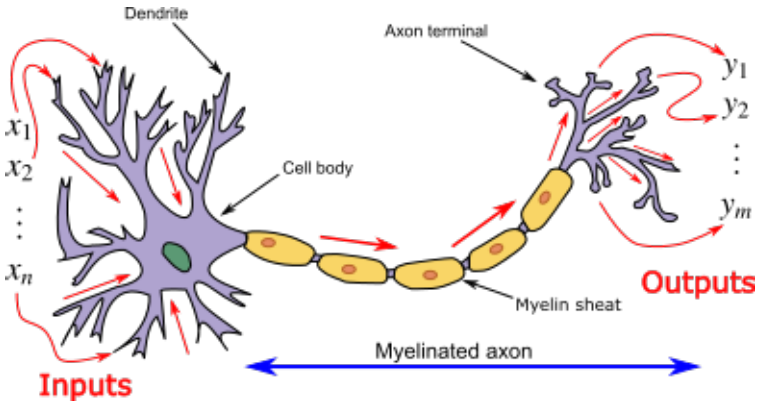
Σχήμα 2.3: Παράδειγμα Μηχανών διανυσμάτων υποστήριξης. [3]

δος αυτή τροφοδοτείται ως είσοδος σε άλλους νευρώνες, είτε κατευθύνεται στο περιβάλλον. Υπάρχουν διάφορα είδη νευρώνων. Οι νευρώνες εισόδου δεν επιτελούν κανέναν υπολογισμό και απλώς συνδέουν τις περιβαλλοντικές εισόδους του δικτύου με τους υπολογιστικούς νευρώνες. Οι νευρώνες εξόδου οι οποίοι προωθούν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου. Οι κρυμμένοι νευρώνες πολλαπλασιάζουν την είσοδό τους με το αντίστοιχο βάρος και υπολογίζουν το ολικό άθροισμα αυτών γινομένων. Το άθροισμα αυτό στην συνέχεια τροφοδοτείται σε μια συνάρτηση ενεργοποίησης, η οποία υλοποιείται εσωτερικά σε κάθε κόμβο. Η τελική τιμή του νευρώνα είναι η έξοδος της συνάρτησης ενεργοποίησης.

2.2 Είδη τεχνητών νευρωνικών δικτύων

2.2.1 Ο perceptron

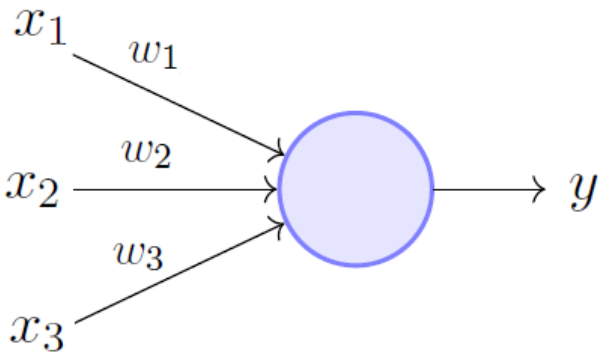
Είναι το αρχαιότερο και απλούστερο νευρωνικό δίκτυο (εικ. 2.5). Πρόκειται για έναν μόνο νευρώνα με βηματική συνάρτηση ενεργοποίησης (εξ. 2.1) και χρησιμοποιείται για δυαδική κατηγοριοποίηση δεδομένων που είναι γραμμικά διαχωρίσιμα. Είναι ένα είδος τεχνη-



Σχήμα 2.4: Το μοντέλο του βιολογικού νευρώνα. [4]

τού νευρωνικού δικτύου που εφευρέθηκε το 1958 στο αεροναυτικό εργαστήριο του Κορνέλλ (Cornell Aeronautical Laboratory) από τον Φρανκ Ρόζενμπλαττ (Frank Rosenblatt) [23].

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > threshold, \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

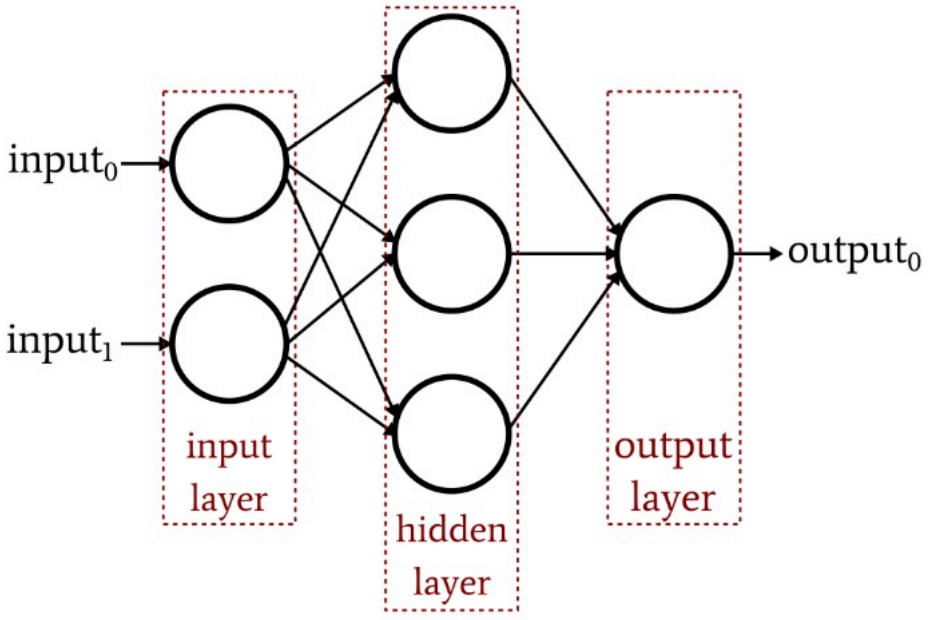


Σχήμα 2.5: Παράδειγμα perceptron με τρεις εισόδους.

Τα βάρη $w_1, w_2, ..., w_n$ είναι πραγματικοί αριθμοί που εκφράζουν την σημασία των αντίστοιχων εισροών στην έξοδο. Η έξοδος, η οποία θα είναι είτε 0 είτε 1, καθορίζεται από την τιμή που θα δώσει το άθροισμα $\sum_i w_i x_i$ και εάν αυτή είναι μεγαλύτερη η μικρότερη από κάποια τιμή κατωφλίου (threshold). Με αυτόν τον τρόπο το perceptron λαμβάνει αποφάσεις, δηλαδή με την τιμή του κατωφλίου και της βαρύτητας των στοιχείων που είναι διαθέσιμα. Επίσης πολλοί νευρώνες perceptron μπορούν να χρησιμοποιηθούν για να δημιουργήσουμε ένα επίπεδο το οποίο ονομάζεται πλήρως συνδεδεμένο επίπεδο.

2.2.2 Πολυεπίπεδα τεχνητά νευρωνικά δίκτυά

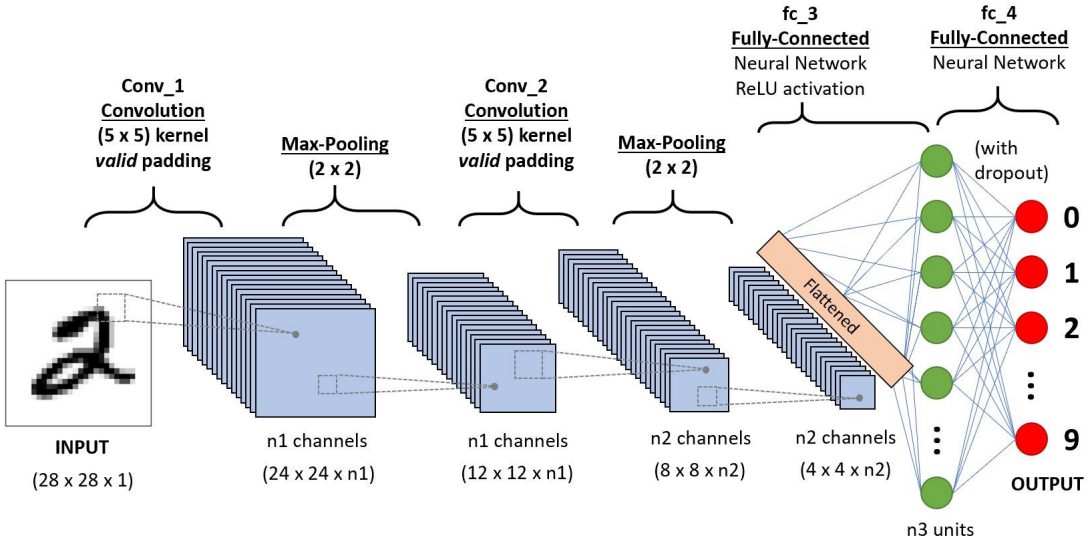
Το πολυεπίπεδο perceptron (MultiLayer Perceptron (MLP)) είναι ένα δίκτυο με πολλά πλήρως συνδεδεμένα επίπεδα πρόσθιας τροφοδότησης (feedforward) (εικ. 2.6) [24]. Οι νευρώνες είναι οργανωμένοι σε επίπεδα (layers) και δεν υπάρχουν συνδέσεις μεταξύ νευρώνων του ίδιου επιπέδου. Με αυτόν τον τρόπο διευκολύνεται η μαθηματική ανάλυση και υπάρχει η δυνατότητα παράλληλης επεξεργασίας. Ένα MLP αποτελείται από τουλάχιστον τρία επίπεδα κόμβων: ένα επίπεδο εισόδου, ένα κρυφό επίπεδο και ένα επίπεδο εξόδου. Η συνάρτηση ενεργοποίησης των κρυμμένων νευρώνων του κρυφού επιπέδου είναι μη γραμμική (συνήθως λογιστική). Στο επίπεδο εξόδου η συνάρτηση ενεργοποίησης είναι συνήθως γραμμική η λογιστική, ανάλογα με το πρόβλημα προς επίλυση. Για προβλήματα ταξινόμησης προτιμάται η λογιστική και για προβλήματα συναρτησιακής προσέγγισης η γραμμική..Υπάρχει πλήρης διασύνδεση μεταξύ των νευρώνων δύο διαδοχικών επιπέδων και συνήθως δεν επιτρέπονται συνδέσεις μεταξύ νευρώνων που ανήκουν σε επίπεδα που δεν είναι διαδοχικά. Το MLP χρησιμοποιεί μια εποπτευόμενη (supervised) τεχνική μάθησης που ονομάζεται backpropagation για εκπαίδευση.



Σχήμα 2.6: Παράδειγμα πολυεπίπεδου perceptron με δύο εισόδους. [5]

2.2.3 Συνελικτικά νευρωνικά δίκτυα

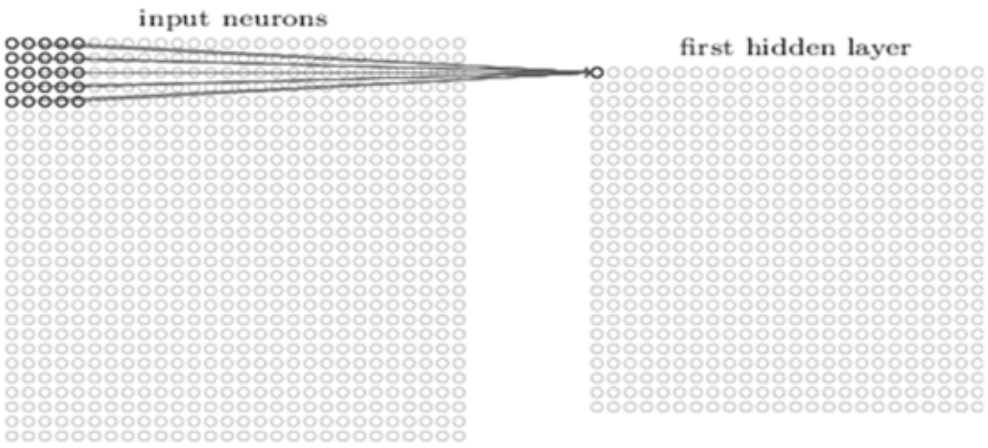
Τα συνελικτικά νευρωνικά δίκτυα (CNN) είναι μια κατηγορία των βαθέων νευρωνικών δικτύων (deep neural networks), τα οποία χρησιμοποιούνται συνήθως για την ανάλυση οπτικών εικόνων [25]. Αποτελούνται από ένα ή πολλά επίπεδα συνέλιξης (convolutional layers) τα οποία συνήθως ακολουθούνται απο αντίστοιχα επίπεδα υπερ-δειγματοληψίας (pooling layers) και πιθανόν κοντά στην έξοδο, ένα η περισσότερα πλήρως συνδεδεμένα επίπεδα (εικ. 2.7). Τα CNN είναι σχεδιασμένα ώστε να εκμεταλλεύονται την τοπικότητα στον χώρο, αφού γειτονικά-κοντίνα pixels έχουν μεγαλύτερη σχέση μεταξύ τους από ότι με μακρινά. Ένα συνελικτικό νευρωνικό δίκτυο αποτελείται από ένα επίπεδο εισόδου και ένα εξόδου, καθώς και από πολλαπλά κρυμμένα επίπεδα.



Σχήμα 2.7: Δομή συνελικτικου νευρωνικού δικτύου. [6]

Οι παράμετροι ενός συνελικτικού επιπέδου αποτελούνται από ένα σύνολο φίλτρων (εικ. 2.8), οπού κάθε φίλτρο είναι αρκετά μικρότερο από την είσοδο, και μεταβάλουν τα βάρη τους προκειμένου να εκπαιδευτούν. Ένα τυπικό φίλτρο σε ένα στρώμα CNN μπορεί να έχει μέγεθος $4x4x3$ και η τελευταία διάσταση μας δείχνει το βάθος του πίνακά, για παράδειγμα σε μια έγχρωμη εικόνα θα είναι 3 ενώ σε μια ασπρόμαυρη 1. Κατά την εμπρός διάδοση, μετατοπίζουμε κάθε φίλτρο κατά ολόκληρο το πλάτος και μήκος της εισόδου και υπολογίζουμε το γινόμενο μεταξύ της εισόδου σε οποιαδήποτε θέση και των αντίστοιχων φίλτρων.

Μια σημαντική παράμετρος για την μείωση του επιπέδου εξόδου του (νευρωνικού δικτύου/επιπέδου) είναι βηματισμός (stride). Πρόκειται για μια παράμετρο που επιλέγει ο δη-



Σχήμα 2.8: Η δράση ενός φίλτρου 5x5. [7]

μιουργός του νευρωνικού επίπεδο για κάθε ένα από τα επίπεδα συνέλιξης. Αυτή η παράμετρος ορίζει το πόσο θα μετακινείται κάθε φορά ο πυρήνας της συνέλιξης.

Μια άλλη σημαντική παράμετρος των νευρωνικών επιπέδων είναι το padding, το οποίο το χρησιμοποιούμε για να ελέγξουμε το μέγεθος της εξόδου για το κάθε επίπεδο. Είναι ένα πλήθος στοιχείων στα οποία δίνεται μια αυθαίρετη τιμή (συνήθως 0 ή την τιμή του πιο κοντινού στοιχείου) και τοποθετούνται στα άκρα της εικόνας, στις κατευθύνσεις του πλάτους και του ύψους.

Το βασικό πλεονέκτημα των συνελικτικών νευρωνικών δικτύων σε σχέση με τα πλήρως συνδεδεμένα επίπεδα είναι η μείωση των συνδέσεων από επίπεδο σε επίπεδο, γεγονός που προκαλεί μείωση των συνολικών βαρών του δικτύου. Αυτή η μείωση κάνει ταχύτερη την εκπαίδευση και μειώνει την πολυπλοκότητά του δικτύου.

2.2.4 Επίπεδο υπερ-δειγματοληψίας

Το επίπεδο υπερ-δειγματοληψίας (pooling layer) είναι ένα στρώμα δειγματοληψίας μεταξύ διαδοχικών επιπέδων συνέλιξης. Με το pooling layer επιτυγχάνουμε μείωση των διαστάσεων της εξόδου ενός συνελικτικού επιπέδου και συνεπώς μειώνουμε τον αριθμό των παραμέτρων και των υπολογισμών στο δίκτυο. Αυτή η μείωση γίνεται με την χρήση ενός πυρήνα οπού εφαρμόζεται μια συνάρτησή στα στοιχεία του πυρήνα. Υπάρχουν διαφορετικές συναρτήσεις που χρησιμοποιούνται και ορισμένες από αυτές είναι οι παρακάτω.

Δειγματοληψία μεγίστου

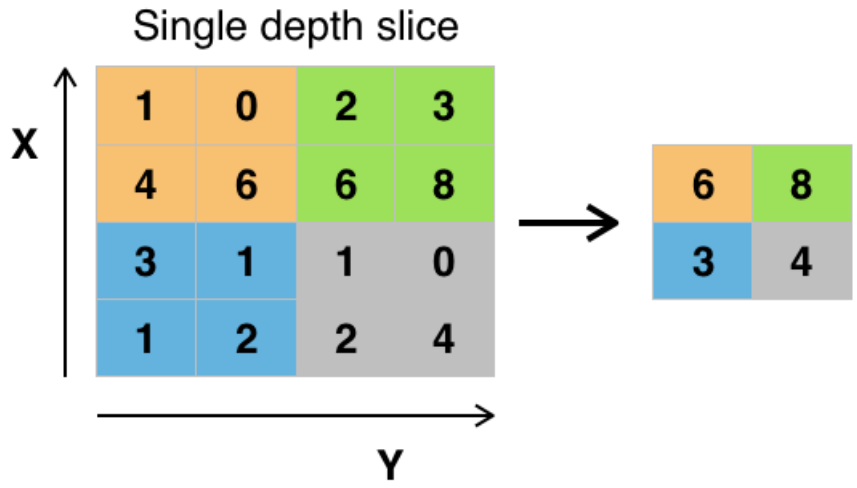
Είναι η μέθοδος που χρησιμοποιείται συνήθως, καθώς διατηρεί μόνο την μέγιστη τιμή της απόκρισης του φίλτρου. Δεν επηρεάζεται από αδιάφορα χαρακτηριστικά στην περιοχή και βοηθάει στην ταχύτερη εκμάθηση του δικτύου.

Δειγματοληψία μέσου όρου

Χρησιμοποιείται σε περιπτώσεις όπου τα χαρακτηριστικά που θέλουμε να υπολογίσουμε προκύπτουν από ολόκληρη την εικόνα. Επιλέγεται ο μέσος όρος των κελιών του πυρήνα.

Δειγματοληψία μέσου

Σε αυτή την μέθοδο επιλέγεται η μεσαία τιμή. Έχει το πλεονέκτημα ότι δεν επηρεάζεται από ακραίες απόκρισης του φίλτρου, αλλά χρειάζεται περισσότερο χρόνο για να υπολογιστεί αφού πρώτα πρέπει να γίνει η ταξινόμηση ώστε να βρεθεί η μεσαία τιμή.

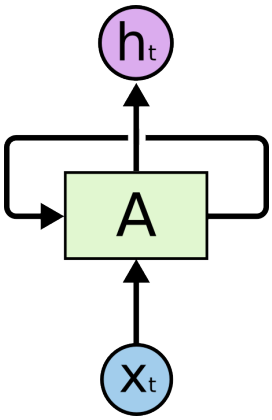


Σχήμα 2.9: Δειγματοληψία μεγίστου. [8]

2.2.5 Αναδρομικά τεχνητά νευρωνικά Δίκτυά

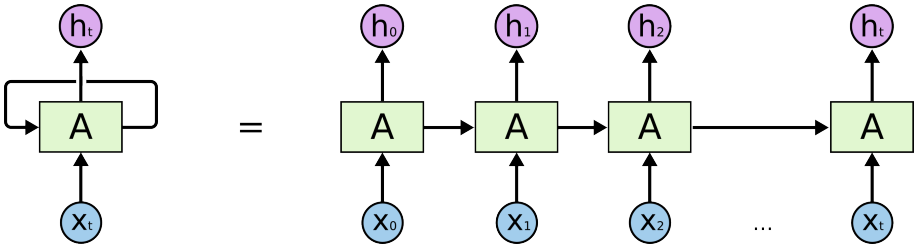
Τα αναδρομικά νευρωνικά δίκτυα (ANN) μοιάζουν με την λειτουργία του ανθρώπινου εγκεφάλου, ο οποίος δεν ξεκινάει την σκέψη του από το μηδέν, όταν για παράδειγμα διαβάζει ένα κείμενο, αλλά κατανοεί την κάθε λέξη με βάση την κατανόηση των προηγούμενων λέξεων. Το κύριο χαρακτηριστικό τους είναι ότι περιέχουν τουλάχιστον μια ανάδραση ανά-

μεσα στους κόμβους του ίδιου επιπέδου ή ανάμεσα σε κόμβους διαφορετικών επιπέδων (εικ. 2.10), γεγονός που τα διαφοροποιεί από τα feedforward δίκτυα .



Σχήμα 2.10: Τα Αναδρομικά Νευρωνικά Δίκτυα αποτελούνται από βρόχους. [9]

Η ιδιότητα αυτή των αναδρομικών νευρωνικών δικτύων τους προσδίδει δύο πολύ σημαντικά χαρακτηριστικά. Το πρώτο είναι η αντίληψη του χρόνου, δηλαδή σε δυναμικά συστήματα μπορούν να χρησιμοποιηθούν για να προβλέψουν την μελλοντική τιμή $t + 1$ έχοντας δεδομένα της στιγμής t ή και προηγουμένων. Το δεύτερο είναι η εισαγωγή της μνήμης στα νευρωνικά δίκτυα.



Σχήμα 2.11: ”Ξετυλιγμένο” ANN. [9]

Τα αναδρομικά νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν σε διάφορες εφαρμογές όπως στην κατάταξη δεδομένων σε κλάσεις, την μοντελοποίηση στοχαστικών ακολουθιών, την συμπίεση δεδομένων κ.α.

2.2.6 Εξαφάνιση/Εκτόξευση της κλίσης

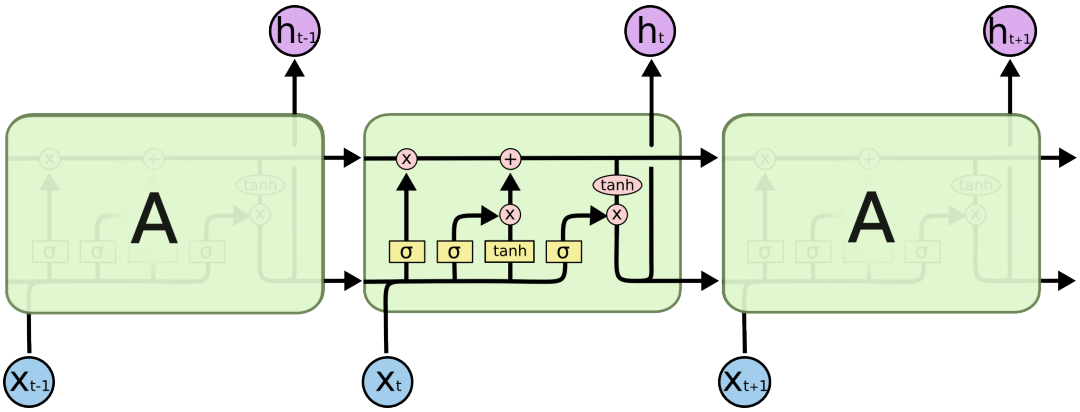
Για τον υπολογισμό της κλίσης του κόστους (gradient of the loss function) για κάποια παρελθοντική είσοδο, παραγωγίζεται μια σύνθεση συναρτήσεων η οποία οδηγεί σε ολοένα και αυξανόμενο αριθμό παραγόντων γινομένου, για παρελθοντικές τιμές. Υπάρχει πιθανότητα κάποιοι από τους (παράγοντες) αυτού του γινομένου να οδηγήσουν σε μεγάλη αύξηση

ή μείωση της κλίσης του κόστους ως προς τις παραμέτρους, και να δημιουργήσουν αστάθεια στο αποτέλεσμα [26]. Οι ανανεώσεις των παραμέτρων βασίζονται στον υπολογισμό της κλίσης, και αν ο υπολογισμός αυτός μας δώσει υπερβολικά μεγάλες ή μικρές τιμές η μάθηση θα αποτύχει. Το πρόβλημα αυτό ονομάζεται ως πρόβλημα μακροπρόθεσμων εξαρτήσεων ή εκτόξευση της κλίσης (Vanishing or exploding gradient problem) και έχει σαν αποτέλεσμα η διαδικασία εκπαίδευσης να αποτύχει καθώς δεν μπορεί να εντοπίσει ικανοποιητικές τιμές για τα βάρη του μοντέλου. Για αυτόν τον λόγο αναπτύχθηκαν νέες τεχνικές για να αποφύγουμε αυτήν την αστάθεια.

2.2.7 Long short-term memory (LSTM)

Για την αντιμετώπιση του προβλήματος Εκτόξευσης της κλίσης δημιουργήθηκαν διάφορες νέες αρχιτεκτονικές και μια πρώτη προσπάθεια ήταν τα νευρωνικά δίκτυα LSTM [27], [9], [28]. Τα LSTM είναι ένα ειδικό είδος αναδρομικού νευρωνικού δικτύου, που αποτελείται από έναν αριθμό μονάδων συνδεδεμένων μεταξύ τους σε κάθε επίπεδο. Κάθε μονάδα των LSTM αποτελείται από ένα ή περισσότερα κελιά μνήμης συνδεδεμένα μεταξύ τους, καθώς και τις πύλες εισόδου, εξόδου και επιλεκτικής συγκράτησης. Οι πύλες εισόδου είναι υπεύθυνες για τις λειτουργίες εγγραφής, οι πύλες εξόδου για τις λειτουργίες ανάγνωσης και οι πύλες επιλεκτικής συγκράτησης για την επαναφορά των κελίων. Με την χρήση αυτών των πυλών εξασφαλίζεται η αποθήκευση και η πρόσβαση στις πληροφορίες ακόμα και μετά την πάροδο μεγάλων χρονικών περιόδων. Το σχήμα (εικ. 2.12) απεικονίζει τα δομικά στοιχεία των μονάδων μνήμης των LSTM και τις συνδέσεις μεταξύ τους. Κάθε γραμμή φέρει ένα διάνυσμα από την έξοδο ενός κόμβου έως τις εισόδους άλλων. Οι ροζ κύκλοι αφορούν πράξεις μεταξύ των διανυσμάτων και τα κίτρινα πλαίσια είναι διακριτά επίπεδα αναδρομικών δικτύων που περιλαμβάνουν συναρτήσεις όπως η σιγμοειδής και η υπερβολική εφαπτομένη και χρησιμοποιούνται στην εκπαίδευση των LSTM. Οι γραμμές που ενώνονται υποδηλώνουν συγχωνεύσεις και οι γραμμές που διακλαδώνονται περιέχουν αντίγραφα της ίδιας πληροφορίας. Το κλειδί για τα LSTM είναι η οριζόντια γραμμή που διατρέχει το πάνω μέρος του διαγράμματος και αντιπροσωπεύει την κατάσταση των μονάδων της μνήμης τους.

Το πρώτο βήμα στο LSTM είναι να αποφασίσουμε για το ποίες πληροφορίες πρόκειται να πετάξουμε από την κατάστασή μονάδας μνήμης (cell state), το οποιο το αποφασίζει η πύλη επιλεκτικής συγκράτησης του συστήματος (εξ. 2.2). Το επόμενο βήμα είναι να αποφασίσουμε ποίες νέες πληροφορίες πρόκειται να αποθηκεύσουμε στο cell state για το οποίο



Σχήμα 2.12: Επίπεδο LSTM. [10]

είναι υπεύθυνη η πύλη εισόδου που ξεχωρίζει ποίες από τις υπάρχουσες πληροφορίες θα παραμείνουν στην μνήμη (εξ. 2.3). Τέλος, θα πρέπει να αποφασιστεί για το αποτέλεσμα h_t που θα εξάγουμε το οποίο θα γίνει είσοδος στο επόμενο επίπεδο. Αρχικά, γίνεται πέρασμα της πληροφορίας εισόδου με σκοπό να αποφασιστεί ποίο ποσοστό αυτής θα προωθηθεί ως στην έξοδο (εξ. 2.5) και στην συνέχεια μαζί με το περιεχόμενο της μνήμης παίρνουμε την επιθυμητή έξοδο.(εξ. 2.6).

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{2.2}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2.3}$$

$$\hat{C}_t = tanh(W_C[h_{t-1}, x_t] + b_C) \tag{2.4}$$

$$o_t = tanh(W_o[h_{t-1}, x_t] + b_o) \tag{2.5}$$

$$h_t = o_t * tanh(C_t) \tag{2.6}$$

Το LSTM μπορεί να χρησιμοποιηθεί σε διάφορες εφαρμογές, όπως αναγνώριση ομιλίας, αναγνώριση χειρόγραφου και αναγνώριση εισβολών ή επιθέσεων σε δίκτυο (intrusion detection systems).

2.2.8 Gated recurrent units (GRU)

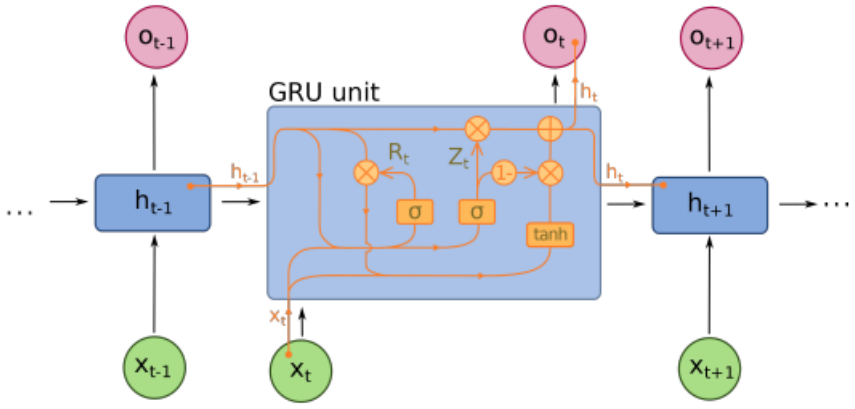
Μια άλλη αρχιτεκτονική είναι τα Gated Recurrent Units (GRUs) (εικ. 2.13) τα οποία είναι μια επέκταση των αναδρομικών νευρωνικών δικτύων [29], [30]. Συνδυάζει τις πύλες εισόδου και forget σε μια, η οποία ονομάζεται πύλη ενημέρωσης(update), ενώ η πύλη εξόδου ονομάζεται πύλη επαναφοράς (reset). Σε κάθε βήμα η κατάσταση στην οποία βρίσκεται τώρα το δίκτυο μπορεί είτε να προστεθεί καινούρια μνήμη, αν θεωρείται σημαντική τόσο αυτή όσο και η προηγούμενη μνήμη, είτε να δημιουργηθεί εκ νέου στην περίπτωση που η προηγούμενη μνήμη δεν θεωρείται πλέον σημαντική. Επίσης, επιλέγει σε τι βαθμό θα μεταφερθεί η τρέχουσα κατάσταση στο επόμενο βήμα, καθώς μπορεί να μεταφερθεί ολόκληρη, να μεταφερθεί ένα μέρος της ή να μην μεταφερθεί καθόλου. Οι εξισώσεις που περιγράφουν τα GRUs είναι:

$$z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1}) \tag{2.7}$$

$$r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1}) \tag{2.8}$$

$$\bar{h_t} = tanh(r_t \circ Uh_{t-1} + Wx_t) \tag{2.9}$$

$$h_t = (1 - z_t) \circ \bar{h_t} + z_t \circ h_{t-1} \tag{2.10}$$



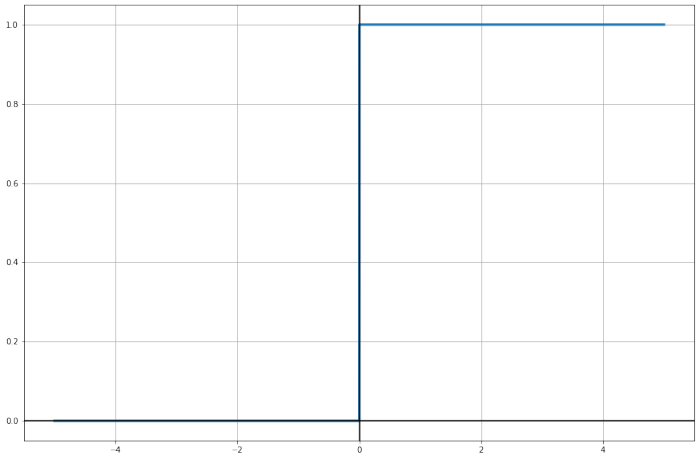
Σχήμα 2.13: Επίπεδο GRU. [11]

2.3 Συνάρτησεις ενεργοποίησης

2.3.1 Βηματική συνάρτηση

Η βηματική συνάρτηση είναι η απλούστερη συνάρτηση ενεργοποίησης και χρησιμοποιείται στον Perceptron.

$$f(x) = \begin{cases} 0 & \text{αν } x < 0 \\ 1 & \text{αν } x \geq 0 \end{cases} \tag{2.11}$$

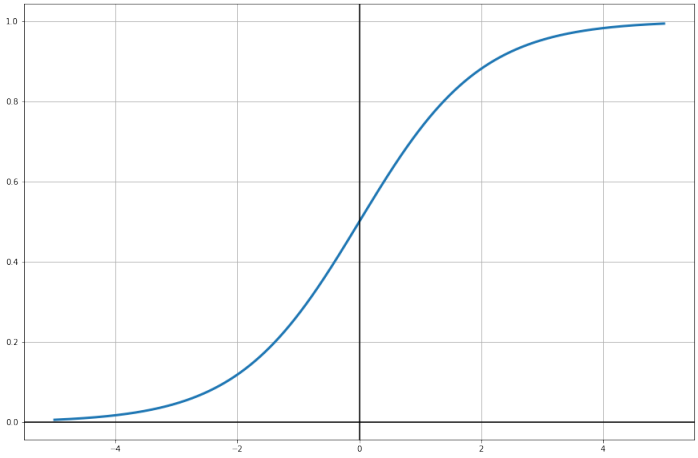


Σχήμα 2.14: Βηματική συνάρτηση.

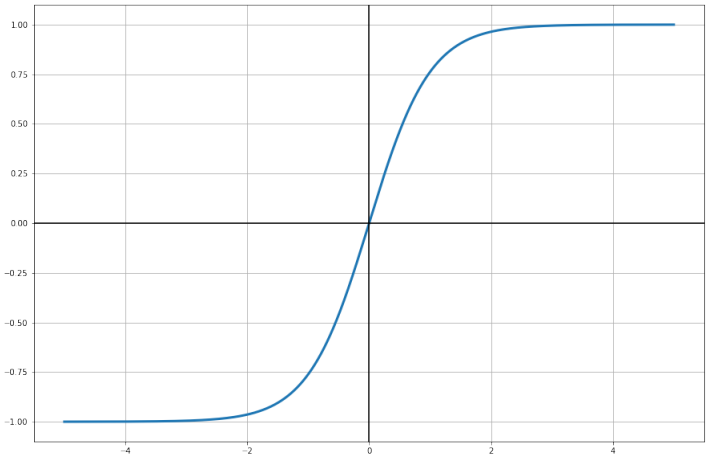
2.3.2 Σιγμοειδής συνάρτηση

Μια σιγμοειδής συνάρτηση είναι μια μαθηματική συνάρτηση, που έχει χαρακτηριστική καμπύλη σχήματος ”S” (εικ. 2.15), (εικ. 2.16). Οι σιγμοειδής συναρτήσεις έχουν πεδίο τιμών όλο το σύνολο των πραγματικών αριθμών, με την τιμή επιστροφής να αυξάνεται συνήθως μονοτονικά, αλλά μπορεί και να μειώνεται. Εμφανίζουν συχνότερα τιμή επιστροφής (άξονας y) στην περιοχή $(0, 1)$ ή συχνά από $(-1, 1)$. Αν δίδεται έντονη αρνητική τιμή εισόδου, η σιγμοειδής συνάρτηση (εξ. 2.12) εκπέμπει τιμές πολύ κοντά στο μηδέν και έτσι η ενημέρωση των παραμέτρων κατά την διάρκεια της εκπαίδευσης δεν γίνεται όσο τακτικά θέλουμε.Για την λύση του προβλήματος χρησιμοποιείται η συνάρτηση υπερβολικής εφαπτομένης που περιγράφεται παρακάτω (εξ. 2.13).

$$s(x) = \frac{1}{1 + e^{-x}} \tag{2.12}$$



Σχήμα 2.15: Σιγμοειδής συνάρτηση.

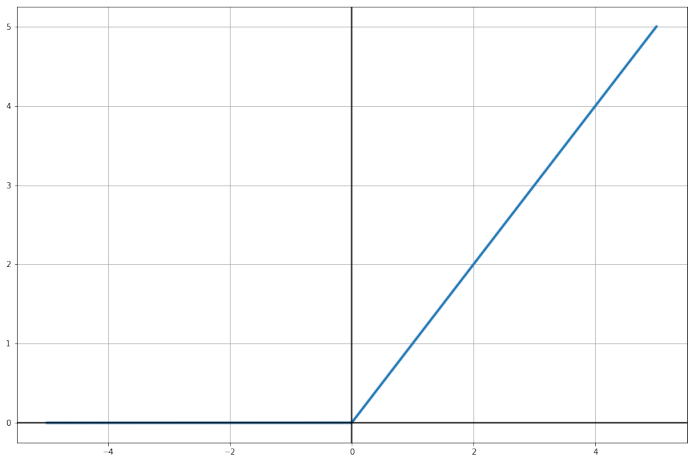


Σχήμα 2.16: Συνάρτηση Υπερβολικής Εφαπτομένης.

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \tag{2.13}$$

2.3.3 Ανορθωμένη γραμμική συνάρτηση ράμπας

Στα τεχνητά νευρωνικά δίκτυα η ανορθωμένη γραμμική συνάρτηση ράμπας (Rectified Linear Unit) είναι μια συνάρτηση ενεργοποίησης που περιγράφεται απο την εξής σχέση: $f(x) = max(0, x)$ όπου x είναι η είσοδος του νευρώνα. Η συνάρτηση αυτή είναι από το 2017 η πιο δημοφιλής συνάρτηση ενεργοποίησης για βαθιά νευρωνικά δίκτυα (DNN). Έχει την δυνατότητα να εκπαιδεύσει ένα δίκτυο γρηγορότερα από τις υπόλοιπες συναρτήσεις, δίνοντας παράλληλα ακριβή αποτελέσματα. Το σημαντικότερο μειονέκτημα της είναι ότι ορισμένες φορές μπορεί να οδηγήσει κάποιους νευρώνες του δικτύου σε τιμές βαρών, που τους αποτρέπει να ενεργοποιηθούν, συνεπώς σταματάνε να εκπαιδεύονται.

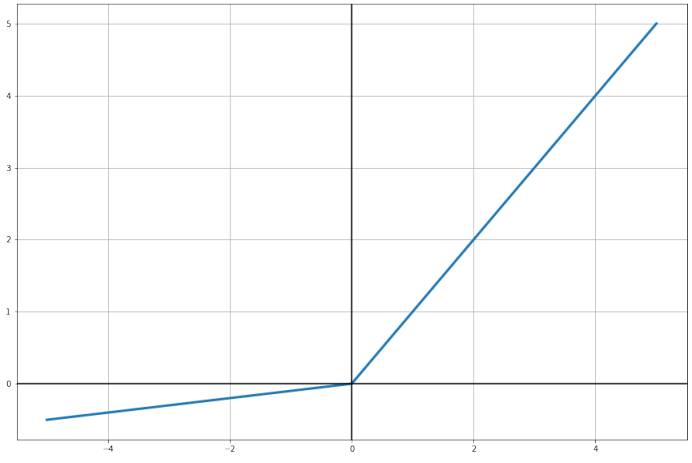


Σχήμα 2.17: Ανορθωμένη γραμμική συνάρτηση ράμπας.

2.3.4 Παραμετροποιημένη συνάρτηση ράμπας

Στην περίπτωση που η τιμή εισόδου είναι αρνητική ,η παραμετροποιημένη συνάρτηση ράμπας (PReLU) την πολλαπλασιάζει με μια μικρή τιμή α. Η τιμή της παραμέτρου α δεν είναι σταθερή αλλά κατά την διάρκεια της εκπαίδευσης ”μαθαίνετε” από το δίκτυο. Αν το $\alpha = 0$ η συνάρτηση μετατρέπεται σε ReLU, ενώ αν πάρει κάποια σταθερή τιμή ονομάζεται Leaky ReLU.

$$f(x) = \begin{cases} \alpha x & \text{αν } x < 0 \\ x & \text{αν } x \geq 0 \end{cases} \tag{2.14}$$



Σχήμα 2.18: Leaky ReLU με $\alpha = 0.1$.

2.3.5 Συνάρτηση Softmax

Η συνάρτηση Softmax γνωστή ως Softargmax ή ομαλοποιημένη εκθετική συνάρτηση είναι μια γενίκευση της λογιστικής συνάρτησης σε πολλαπλές διαστάσεις και εφαρμόζεται στην έξοδο ενός ολοκλήρου επίπεδου. Χρησιμοποιείται συχνά ως η τελευταία λειτουργία ενεργοποίησης του νευρωνικού δικτύου για την ομαλοποίηση της εξόδου, έτσι ώστε οι τιμές εξόδου να είναι μεταξύ του 0 και 1, αλλά και το άθροισμά τους να ισούται με την μονάδα. Χρησιμοποιείται ευρέως στα βαθιά νευρωνικά δίκτυα όταν θέλουμε να αντιμετωπίσουμε προβλήματα κατηγοριοποίησης.Μαθηματικά εκφράζεται απο τον τύπο (εξ. 2.15).

$$\frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}} \tag{2.15}$$

2.4 Εκπαίδευση νευρωνικού δικτύου

2.4.1 Εισαγωγή

Εκπαίδευση νευρωνικού δικτύου 1) Εκπαίδευση με επίβλεψη: Το δίκτυο τροφοδοτείται με μια σειρά από δεδομένα που έχουν αντιστοίχηση με κάποιο στόχο. Το δίκτυο προσαρμόζει τις παραμέτρους του ώστε να μπορεί να προβλέψει την «σωστή» έξοδο σε άγνωστα δεδομένα. 2)Εκπαίδευση χωρίς επίβλεψη: Το δίκτυο τροφοδοτείται με μια σειρά από δεδομένα, τα οποία προσπαθεί να ομαδοποιήσει βάσει κοντινών χαρακτηριστικών [].

2.4.2 Overfitting

Η αιτία της κακής απόδοσης στην μηχανική μάθηση είναι είτε το overfitting, είτε το underfitting των δεδομένων. Η εποπτευόμενη μηχανική μάθηση είναι το πρόβλημα της προσέγγισης μια συνάρτησης στόχου f έχοντάς τις μεταβλητές εισόδου X σε μια μεταβλητή εξόδου Y $Y = f(X)$.Αυτός ο χαρακτηρισμός περιγράφει το εύρος των προβλημάτων ταξινόμησης και πρόβλεψης και τους αλγορίθμους που μπορούν να χρησιμοποιηθούν για την αντιμετώπιση τους. Το σημαντικότερο πρόβλημά είναι η συνάρτηση που προσεγγίζουμε να μπορεί να γενικεύεται σε νέα δεδομένα, διότι τα δεδομένα εκπαίδευσής είναι μονό ένα μικρο δείγμα. Συνεπώς ένα μοντέλο μηχανικής μάθησης θα πρέπει να μπορεί να κάνει γενικεύσεις απο τα δεδομένα που χρησιμοποίησε κατά την εκπαίδευσή ώστε να μπορεί να κάνει σωστές προβλέψεις σε νέα δεδομένα όπου του δίνονται πρώτη φορά Η ορολογία που χρησιμοποιούμε στην

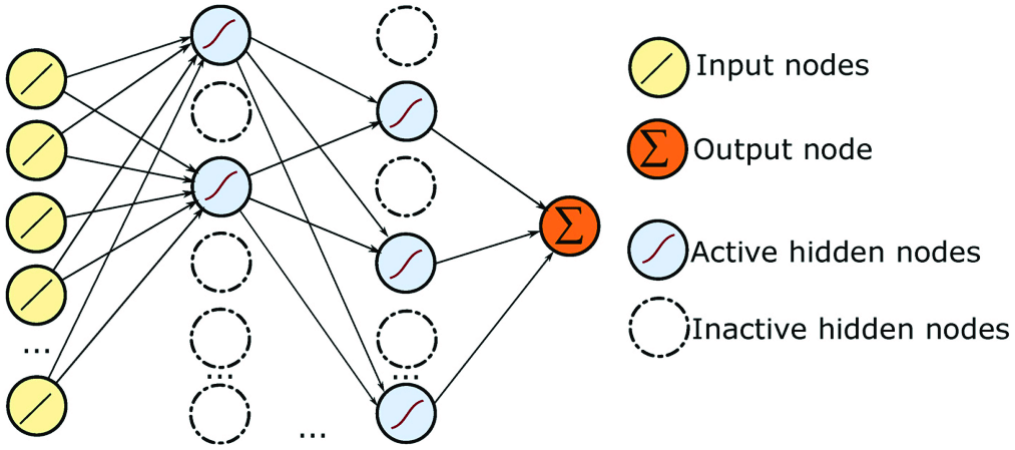
μηχανική μάθηση όταν μιλάμε για το πόσο καλά ένα μοντέλο μηχανικής μάθησης μαθαίνει και γενικεύεται σε νέα δεδομένα είναι το overfitting και το underfitting, οι οποίες είναι οι δύο μεγαλύτερες αιτίες για κακή απόδοση αλγορίθμων μηχανικής μάθησης. Το overfitting συμβαίνει όταν ένα μοντέλο μαθαίνει τον θόρυβο από τα δεδομένα εκπαίδευσης ως έννοιες σε σημείο οπού τα αποτελέσματα επηρεάζονται αρνητικά, διότι αυτές οι έννοιες δεν ισχύουν σε νέα δεδομένα. Στόχος είναι να βρούμε το σημείο μεταξύ του overfittng και underfitting. Για την κατανόηση του στόχου μπορούμε να δούμε την απόδοση ενός αλγορίθμου μηχανικής μάθησης, με την πάροδο του χρόνου, καθώς μαθαίνει τα δεδομένα εκπαίδευσης. Με την πάροδο του χρόνου, καθώς μαθαίνει ο αλγόριθμος, το σφάλμα για το μοντέλο στα δεδομένα εκπαίδευσης μειώνεται όπως και το σφάλμα στο σύνολο δεδομένων δοκιμής. Εάν η εκπαίδευση συνεχιστεί για μεγάλο χρονικό διάστημα, η απόδοση στο σύνολο δεδομένων εκπαίδευσης μπορεί να συνεχίσει να μειώνεται επειδή το μοντέλο έχει overfitting και μαθαίνει τον θόρυβο στο σύνολο δεδομένων εκπαίδευσης.

2.4.3 Dropout

Μια προσέγγιση στην αντιμετώπιση του προβλήματος του overfitting περιλαμβάνει το dropout [31], [28]. Σε κάθε επανάληψη της εκπαίδευσης, απενεργοποιούνται τυχαία κάποιοι νευρώνες, κρυφοί ή μη, του δικτύου μαζί με όλες τις εισερχόμενες και εξερχόμενες συνδέσεις. Οι νευρώνες θα απενεργοποιηθούν σε κάθε επανάληψη γίνεται με τυχαίο τρόπο και ο κάθε νευρώνας έχει μια συγκεκριμένη πιθανότητα παραμονής. Μια συνηθισμένη τιμή για τους εσωτερικούς νευρώνες είναι το 0,5, άλλα εξαρτάται και από την εφαρμογή του δικτύου, ενώ για τους νευρώνες εισόδου δεν χρησιμοποιείτε dropout. Με αυτήν την μέθοδο οι νευρώνες μαθαίνουν να μην προσαρμόζονται πάρα πολύ στα δεδομένα μας και να έχουν λιγότερο εξάρτησή από γειτονικούς νευρώνες.

2.4.4 Κατάβαση δυναμικού

Σκοπός του αλγορίθμου Gradient descent είναι να βρει όλα τα βάρη και τις πολώσεις έτσι ώστε η συνάρτηση που μελετάμε να πλησιάζει το 0 [32]. Στόχος μας είναι να βρούμε το σημείο στο οποίο η συνάρτηση κόστους $C(\theta)$ έχει ολικό ελάχιστο, οπού θ του δικτύου μας. Οι παράμετροι αυτοί ανανεώνονται προς την αντίθετη κατεύθυνση της κλίσης της $C(\theta)$. Επιπλέον, η κλίση αυτή πολλαπλασιάζεται με έναν συντελεστή n ο οποίος ονομάζεται ρυθμός εκμάθησης (learning rate), ο οποίος καθορίζει το μέγεθος του άλματος που κάνουμε για



Σχήμα 2.19: Παράδειγμα Dropout. [12]

να φτάσουμε στο ελάχιστο. Στα προβλήματα που μελετάμε, συνήθως έχουμε μεγάλο αριθμό μεταβλητών οπότε ο αναλυτικός τρόπος επίλυσης με τον υπολογισμό των μερικών παραγώγων είναι δύσκολος και θα πρέπει να χρησιμοποιηθεί αλγόριθμος. Ο αλγόριθμος Gradient descent, ξεκινάει από ένα τυχαίο σημείο και προχωράει με διαδοχικές επισκέψεις σε άλλα σημεία έτσι ώστε η συνάρτηση κόστους C να μειώνεται σε κάθε επανάληψη μέχρι να φτάσουμε το ελάχιστο.

Κατάβαση δυναμικού κατά παρτίδες

Η κατάβαση δυναμικού κατά παρτίδες (Batch Gradient descent) πρώτα υπολογίζει το κόστος και τις κλίσεις για όλα τα δεδομένα που υπάρχουν στο σύνολο εκπαίδευσης και στην συνέχεια ανανεώνει κατάλληλα τα βάρη του δικτύου. Για τα σύνολα εκπαίδευσης που είναι μεγάλα σε όγκο η κατάβαση δυναμικού κατά παρτίδες μπορεί να γίνει πολύ αργή, λόγω του ότι υπολογίζει και κρατάει τις κλίσεις για ολόκληρο το σύνολο των δεδομένων έως το τέλος της εποχής εκπαίδευσης όπου ανανεώνει τις παραμέτρους. Επιπλέον ένα ακόμα αρνητικό χαρακτηριστικό του Batch Gradient descent είναι το ότι υπολογίζει κλίσεις για παρόμοια παραδείγματα, πριν γίνει η τελική ενημέρωση, το οποίο είναι κοστοβόρο [33].

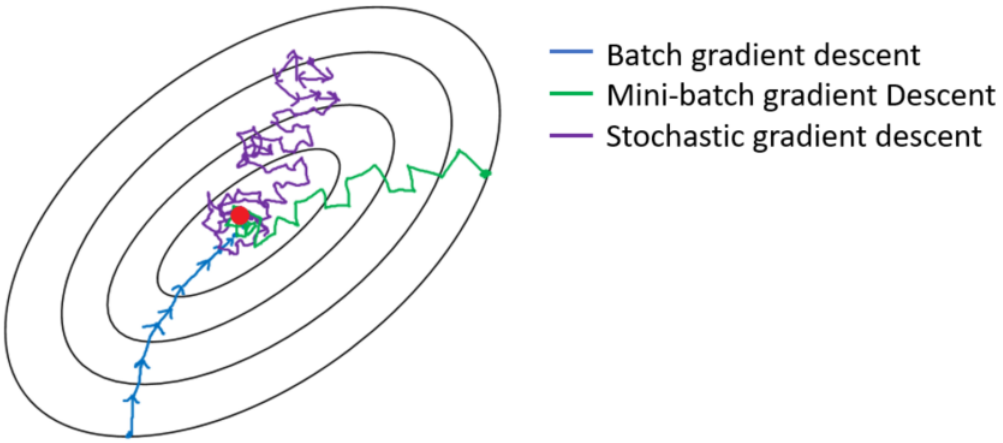
Στοχαστική κατάβαση δυναμικού

Η Στοχαστική κατάβαση δυναμικού δεν ανανεώνει τις παραμέτρους μετά το τέλος της εποχής, αλλά αντιθέτως τις ανανεώνει για κάθε παράδειγμα ξεχωριστά, επιλέγοντάς τα τυχαία από το σύνολο δεδομένων μας. Η στοχαστική κατάβαση δυναμικού σε αντίθεση με το Batch Gradient descent δεν επαναπροσδιορίζει κλίσεις για παρόμοια παραδείγματα διότι η

ενημέρωση γίνεται μετά από κάθε παράδειγμα και έτσι η εκπαίδευση είναι ταχύτερη. Ένα αρνητικό της στοχαστικής κατάβασης δυναμικού είναι το ότι η συνάρτηση κόστους, όπου προσπαθούμε να ελαχιστοποιήσουμε, έχει λιγότερο ομαλή πορεία προς το ελάχιστο [33].

Κατάβαση δυναμικού κατά μικρές παρτίδες

Η κατάβαση δυναμικού κατά μικρές παρτίδες είναι ένας συνδυασμός των παραπάνω δύο μεθόδων, όπου αντί να επιλέγουμε ένα παράδειγμα κάθε φορά, επιλέγουμε μια παρτίδα παραδειγμάτων μεγέθους n ώστε να ανανεώσουμε τις παραμέτρους του δικτύου μας. Αυτό έχει αποτέλεσμα να αποφύγουμε το αρνητικό της στοχαστικής κατάβασης δυναμικού ενώ συγχρόνως διατηρείται η ταχύτητα του αλγορίθμου [33].



Σχήμα 2.20: Σύγκριση μεθόδων Gradient descent. [13]

2.4.5 Οπισθοδιάδοση

Στη μηχανική μάθηση, ο αλγόριθμος οπισθοδιάδοσης (backpropagation) [34], [32] είναι ένας ευρέως χρησιμοποιούμενος αλγόριθμος στην εκπαίδευση ενός δικτύου που αποτελείται από πολλά επίπεδα. Τα βάρη του δικτύου αρχικοποιούνται σε τυχαίες τιμές και κατά την διάρκεια της εκπαίδευσής αν η έξοδος του δικτύου είναι λάθος τα βάρη αλλάζουν ώστε να μειωθεί το λάθος. Μέτα από πολλές επαναλήψεις το λάθος θα γίνει πολύ μικρό και η εκπαίδευση θα έχει τελειώσει. Ο αλγόριθμος backpropagation περιλαμβάνει 2 στάδια. Στο πρώτο στάδιο τα βάρη του δικτύου παραμένουν σταθερά και το σήμα εισόδου μεταδίδεται στο δίκτυο προς τα μπροστά μέχρι να φτάσει στο τελευταίο επίπεδο εξόδου. Στο δεύτερο στάδιο

τα βάρη του δικτύου αλλάζουν τιμές και παράγεται ένα σήμα σφάλματος απο την διαφορά της πραγματικής εξόδου του δικτύου με μια επιθυμητή έξοδο, το σήμα αυτό διαδίδεται προς τα πίσω. Ο αλγόριθμος backpropagation απαιτεί ότι οι παράγωγοι ενεργοποίησης κατά τον σχεδιασμό του δικτύου είναι γνωστοί. Στο πλαίσιο της μάθησης, η οπισθοδιάδοση χρησιμοποιείται από τον αλγόριθμο βελτιστοποίησης μείωσης κλίσης ώστε να ρυθμίσει το βάρος των νευρώνων υπολογίζοντας την κλίση της συνάρτησης απώλειας. Η οπισθοδιάδοση υπολογίζει τις κλίσεις, ενώ η gradient descent (στοχαστική μείωση της κλίσης) χρησιμοποιεί τις κλίσεις για την εκπαίδευση του μοντέλου. Η διαδικασία εκπαίδευσης λαμβάνει χώρα με την παρουσίαση και εφαρμογή στο νευρωνικό δίκτυο ενός συνόλου παραδειγμάτων εκπαίδευσης. Η παρουσίαση όλων των προτύπων στο δίκτυο ονομάζεται εποχή. Εκτελούνται επαναλήψεις των εποχών, ώσπου τα βάρη του δικτύου να σταθεροποιηθούν σε συγκεκριμένες τιμές που θα προκαλούν σύγκλιση της μέσης τιμής των σφαλμάτων στην ελάχιστη δυνατή τιμή της. Ο backpropagation αλλάζει τα βάρη έτσι ώστε να ελαχιστοποιήσει αυτό το σφάλμα.

Κεφάλαιο 3

Δεδομένα και Προεπεξεργασία

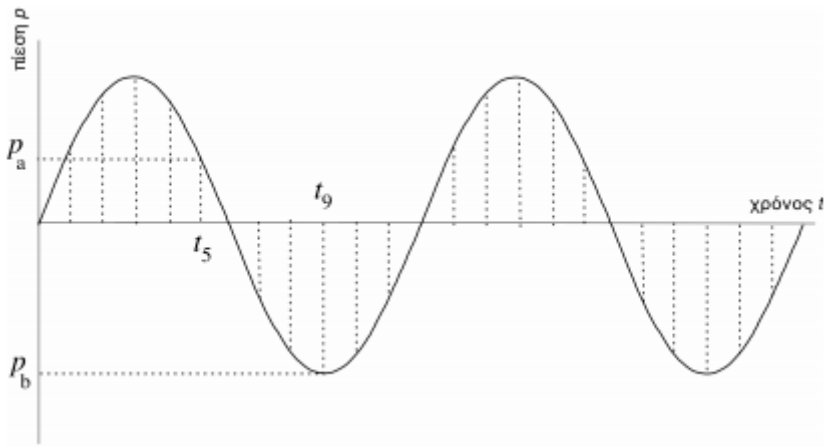
3.1 Προεπεξεργασία

3.1.1 Ο ήχος σαν κυματομορφή

Ο ήχος αποτελεί ένα κύμα που ταξιδεύει σε ένα υλικό μέσο, στερεό, υγρό η αέριο [35]. Η αναπαράσταση ενός κύματος ήχου μπορεί να γίνει στις δύο διαστάσεις, σε ένα σύστημα δύο κάθετων αξόνων x και y. Ο άξονας y αφορά την μεταβολή που προκαλεί το κύμα στο μέσο που μεταδίδεται και αφορά το μέγεθος της έντασης, ενώ ο άξονας x τον χρόνο που συντελείται η αντίστοιχη μεταβολή. Επομένως ο ήχος αποτελεί μια χρονική συνάρτηση μεταβολής της έντασης. Σε αλγορίθμους μηχανικής μάθησης όμως, δεν θα έχουμε τα καλύτερα αποτελέσματά αν δώσουμε σαν είσοδο τον ήχο στην αρχική του μορφή, για αυτόν τον λόγο συνηθίζεται να περνάει από κάποια στάδια προεπεξεργασίας για την εξαγωγή χαρακτηριστικών. Σε σήματα ήχου συνηθίζεται να μετατρέπονται από το πεδίο του ήχου στο πεδίο της συχνότητάς με την χρήση του μετασχηματισμού Fourier.

3.1.2 Ανάλυση Fourier

Η ανάλυση Fourier είναι ένα πεδίο των εφαρμοσμένων μαθηματικών [36]. Προέκυψε από την προσπάθεια αναπαράστασης μια συνάρτησης ως άθροισμα απλούστερων περιοδικών τριγωνομετρικών συναρτήσεων. Η κεντρική ιδέα στην ανάλυση Fourier είναι η προσπάθεια για κατανόηση μια συνάρτησης μέσω της διάσπασης της σε στοιχειώδη μέρη (αποσύνθεση) (εξ. 3.1). Η αντίστροφη διαδικασία, δηλαδή η κατασκευή μια συνάρτησης από γνωστές συναρτήσεις ονομάζεται σύνθεση (εξ. 3.2). Η ανάλυση Fourier περιλαμβάνει και



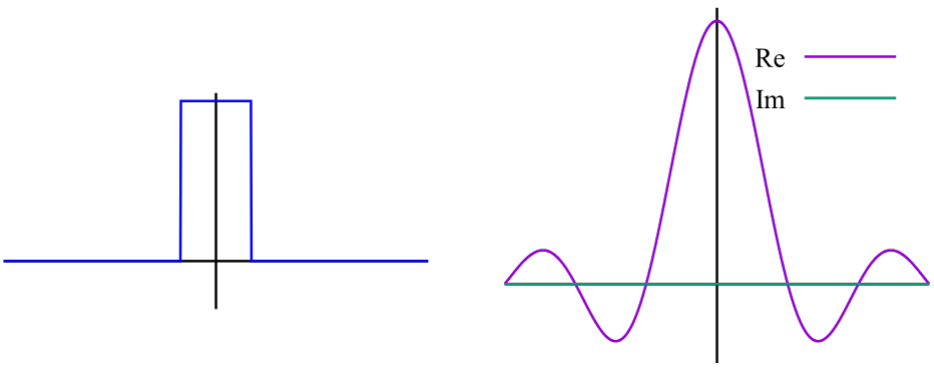
Σχήμα 3.1: Ο ήχος σαν κυματική συνάρτηση

τις δύο διαδικασίες.

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(j\Omega)\, e^{j\Omega t}\, d\Omega \tag{3.1}$$

$$X(j\Omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\Omega t}\, dt \tag{3.2}$$

Ο μετασχηματισμός Fourier αποσυνθέτει μια συνάρτηση σε άθροισμα περιοδικών ημιτονοειδών και συνημιτονοειδών συναρτήσεων. Το αποτέλεσμα είναι μια συνάρτηση γνωστή ως φάσμα, που περιγράφει κατά πόσο συμμετέχει κάθε στοιχειώδες ημίτονο στον σχηματισμό της αρχικής συνάρτησης. Η συνάρτηση που δημιουργείται έχει διαφορετικό πεδίου ορισμού από την αρχική. Σε ένα σήμα χρόνου που εφαρμόσαμε μετασχηματισμό Fourier μεταφέρετε από το πεδίο του χρόνου στο πεδίο της συχνότητας.



Σχήμα 3.2: Μετασχηματισμός Fourier μιας συνάρτησης

3.1.3 Μετασχηματισμός Fourier μικρού χρόνου

Με τον απλό μετασχηματισμός Fourier χάνουμε εντελώς πληροφορία σχετικά με τον χρόνο η οποία μπορεί να φανεί χρήσιμη στο μοντέλο μας [37]. Ο μετασχηματισμός Fourier μικρού χρόνου (Short-time Fourier transform(STFT)) χρησιμοποιείται για τον προσδιορισμό της ημιτονοειδούς συχνότητας και της φάσης μικρών κομματιών του σήματος. Στην πράξη η διαδικασία υπολογισμού του STFT είναι να χωρίσουμε ένα μεγάλο σήμα σε μικρότερα σήματα, και στην συνέχεια να υπολογίσουμε τον μετασχηματισμό Fourier για καθένα από τα επιμέρους σήματα ξεχωριστά. Αυτό έχει σαν αποτέλεσμά να πάρουμε το φάσμα του κάθε τμήματος, διατηρώντας παράλληλα πληροφορία σχετικά με τον τον χρόνο. Στον συνεχή χρόνο, η συνάρτηση που πρόκειται να μετασχηματιστεί, πολλαπλασιάζεται με μια άλλη συνάρτηση, με αποτέλεσμά η αρχική συνάρτηση να μηδενίζεται παντού με εξαίρεση ένα διάστημα. Από τον μετασχηματισμό Fourier του προκύπτοντος σήματος, λαμβάνεται μια δισδιάστατη αναπαράσταση του σήματος. Η μαθηματική συνάρτηση που περιγράφει την παραπάνω διαδικασία είναι:

$$\mathbf{STFT}\{x(t)\}(\tau,\omega) \equiv X(\tau,\omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-i\omega t}\,dt \tag{3.3}$$

Όπου το $w(t)$ είναι η "συνάρτηση παραθύρου" και είναι κεντραρισμένη στο μηδέν. Το $x(t)$ είναι το σήμα που πρόκειται να μετασχηματιστεί, $(,)$ είναι ουσιαστικά ο μετασχηματισμός Fourier του $x(t)w(\ -t)$, μια σύνθετη συνάρτηση που αντιπροσωπεύει την φάση και το μέγεθος του σήματος με την πάροδο του χρόνου και της συχνότητας. Ο μετασχηματισμός STFT είναι αντιστρέψιμος (αναστρέψιμος), δηλαδή το αρχικό σήμα μπορεί να ανακτηθεί με τον αντίστροφο του STFT. Ο πιο συνηθισμένος και ευρέως διαδεδομένος τρόπος αντιστροφής του STFT είναι με την χρήση της μεθόδου overlap-add, η οποία επιτρέπει τροποποιήσεις στο φάσμα του STFT. Με αυτόν τον τρόπο δημιουργείται μια ευέλικτη μέθοδος επεξεργασίας σήματος.

Όταν το παράθυρο (χρονικό) είναι μεγάλο, μπορούμε να διακρίνουμε με ακρίβεια το ποία ακριβώς συχνότητα είναι παρούσα στο σήμα κάθε στιγμή. Τα χρονικά όρια στα οποία γίνονται οι εναλλαγές των συχνοτήτων δεν είναι πολύ σαφή. Υπάρχει επικάλυψη στον άξονα του χρόνου, η οποία δημιουργεί την λανθασμένη εντύπωση ότι υπάρχουν ταυτόχρονα δύο συχνότητες στο σήμα αυτό. Αυτό συμβαίνει επειδή κάποια από τα τμήματα του σήματος που αποκόβουμε, περιέχουν και τις δύο συχνότητες, όχι ταυτόχρονα, αλλά διαδοχικά. Αν το μήκος του παραθύρου μικρύνει, μπορούμε να παρατηρήσουμε καλύτερα τις μεταβολές συχνο-

τήτων στον χρόνο, αλλά δεν έχουμε καλή ανάλυση στον άξονα των συχνοτήτων. Επομένως όταν έχουμε μεγάλο παράθυρο τότε η ανάλυση στην συχνότητα είναι καλή ενώ η ανάλυση στον χρόνο είναι κακή. Αντιθέτως με ένα μικρό παράθυρο η ανάλυση στον χρόνο είναι καλή είναι χειρότερη η ανάλυση στην συχνότητα. Αυτό ονομάζεται νόμος απροσδιοριστίας του Heisenberg, και αναφέρεται στο ότι δεν μπορούμε να ξέρουμε ποιά συχνότητα υπάρχει σε μία χρονική στιγμή, αλλά μπορούμε να ξέρουμε ποίο είναι το εύρος των συχνοτήτων που υπάρχει σε ένα χρονικό διάστημα.

3.2 Τα δεδομένα που χρησιμοποιήθηκαν

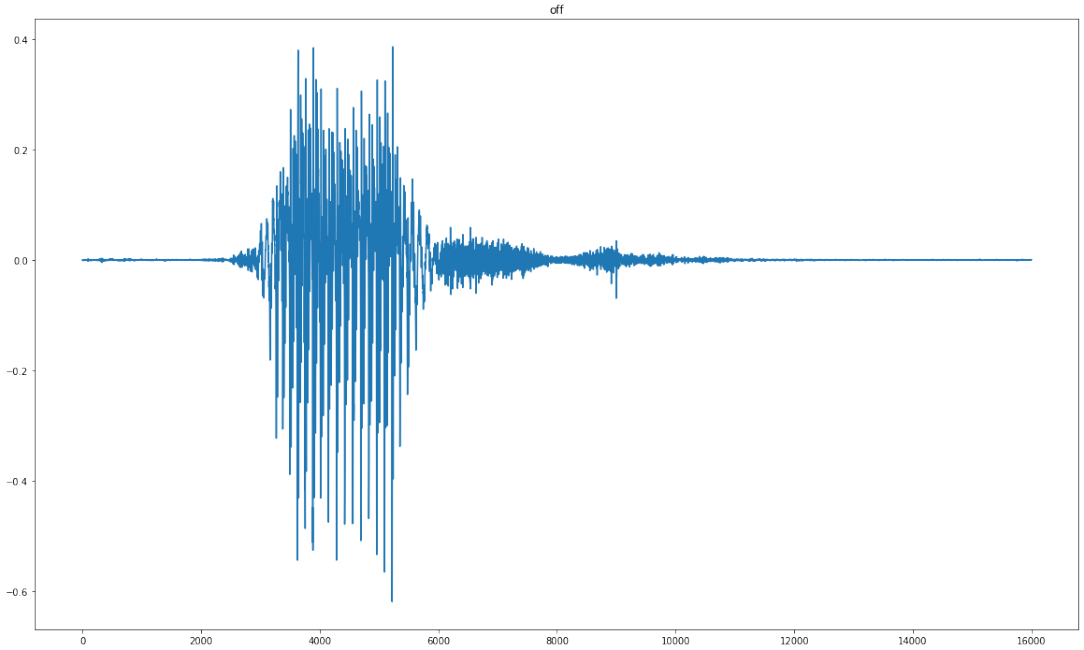
Για αυτήν την εργασία χρησιμοποιήσαμε το dataset 'SpeechCommands' [38]. Το dataset αυτό αποτελείται από διάφορα αρχεία ήχου οπού το καθένα περιέχει μια από τις παρακάτω κατηγορίες:

Category	Number of Files
silence	1197
unknown	60760
down	3917
go	3880
left	3801
no	3941
off	3745
on	3845
right	3778
stop	3872
up	3723
yes	4044

Παρατηρούμε ότι στο dataset μας, σχεδόν όλες οι κατηγορίες έχουν ίδιο αριθμό αρχείων (≈ 4000) και επιπλέον υπάρχει και η κατηγορία 'unknown' ώστε να είναι δυνατή η αναγνώρισή κάποια λέξης που δεν ανήκει σε κάποια γνωστή κατηγορία. Στο σύνολο υπάρχουν 100503 αρχεία και τα 60760 ανήκουν στην κατηγορία 'unknown'.

Τα δεδομένα μας αρχικά βρίσκονται σε μορφή κυματομορφής όπως φαίνεται στην εικόνα (εικ. 3.3). Όπως αναφέρθηκε παραπάνω όμως αυτή η μορφή δεν είναι η κατάλληλη για είσοδο σε νευρωνικό δίκτυο. Κάθε σήμα ήχου είναι ένα πολύπλοκο σήμα που αποτελείται από

πολλαπλά σήματα μιας συγκεκριμένης συχνότητας που ταξιδεύουν μαζί στον χώρο και δημιουργούν το τελικό σήμα ήχου που ακούμε. Συνεπώς μπορούμε να χρησιμοποιήσουμε αυτά τα σήματά σαν είσοδο στο μοντέλο μας δημιουργώντας το φασματογράφημα του αρχικού μας σήματος.

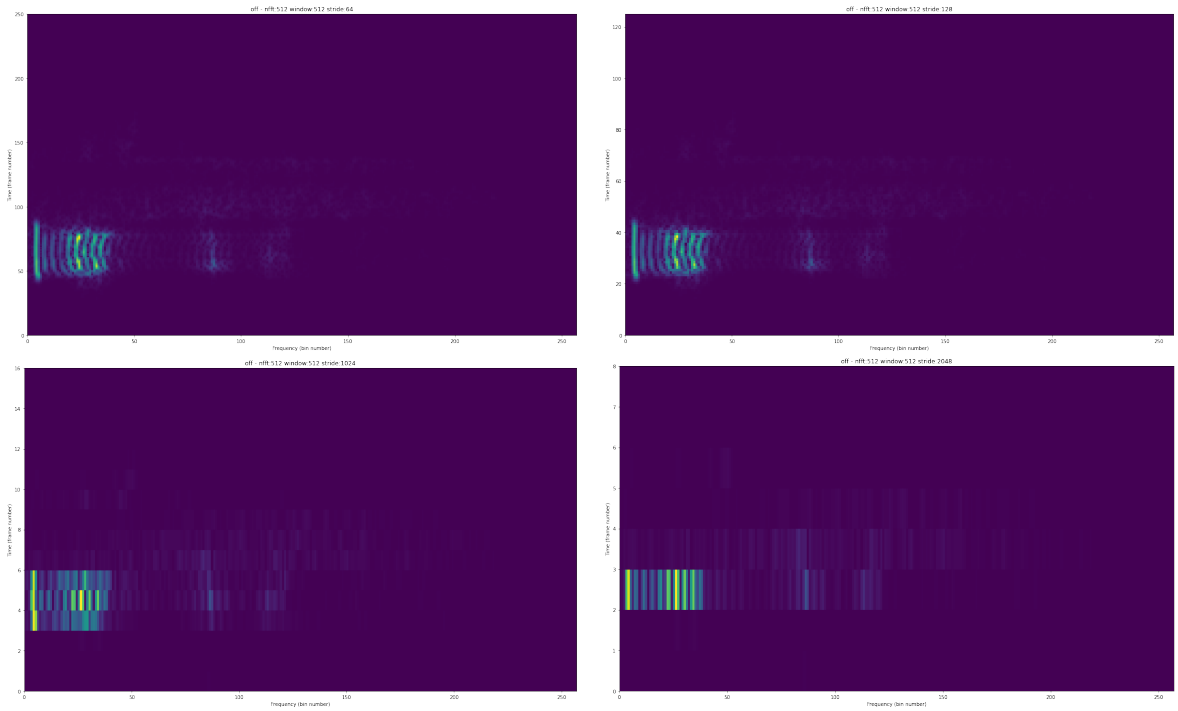


Σχήμα 3.3: Κυματομορφή της λέξης ”right”

Για την δημιουργία του φασματογραφήματος έχουμε διάφορες παραμέτρους που μπορούμε να αλλάξουμε. Μια παράμετρος είναι ο αριθμός των δειγμάτων (samples) από το σήμα θα εφαρμόσουμε μετασχηματισμό Fourier (window), ο αριθμός των δειγμάτων οπού θα προσπεράσουμε πριν εφαρμόσουμε πάλι μετασχηματισμό Fourier (stride) και τέλος το μέγεθος του γρήγορου μετασχηματισμού Fourier (nfft), δηλαδή το εύρος συχνότητάς που θα καταλαμβάνει κάθε τμήμα στον άξονα της συχνότητάς.

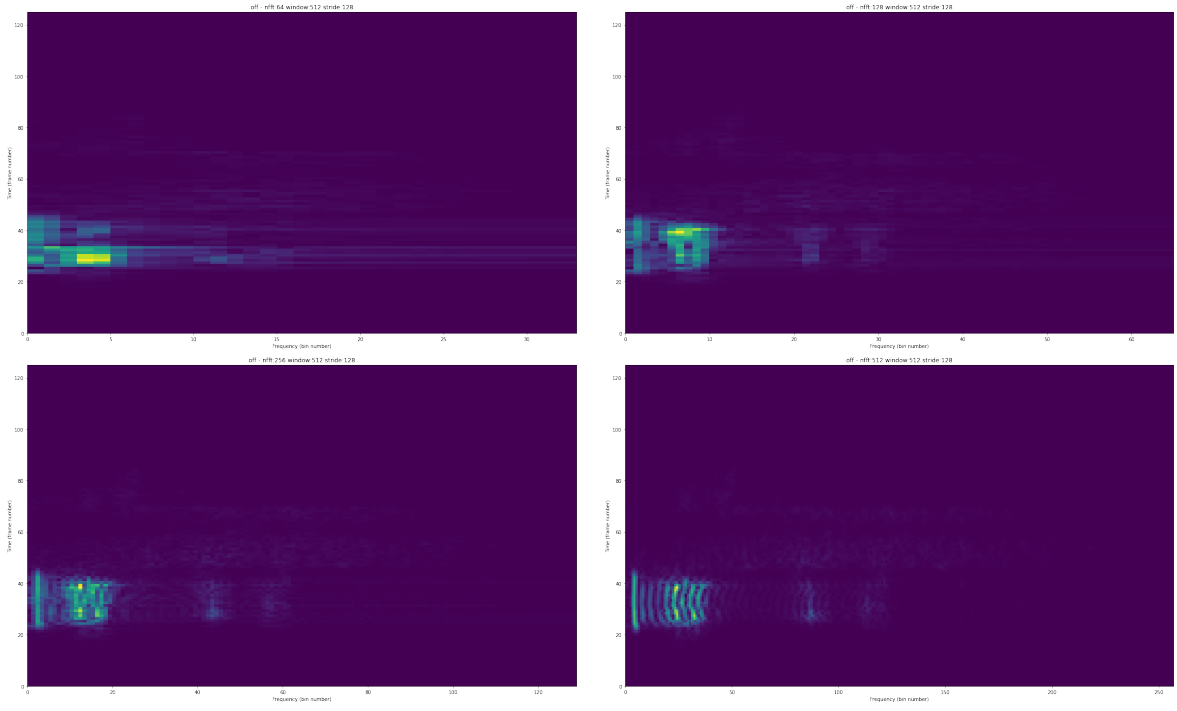
Στις παρακάτω εικόνες φαίνεται μια σύγκριση των παραγόμενων φασματογραφήμάτων με διαφορετικές παραμέτρους σε κυματομορφή που αποτελείται από 16000 δείγματά..

Στην παρακάτω εικόνα (εικ. 3.4) κρατήσαμε το window και το nfft σταθερό ενώ αλλάξαμε την τιμή του stride. Όπως βλέπουμε η αλλαγή του stride επηρεάζει αρκετά την εμφάνιση του φασματογραφήματος που παράγεται. Παρατηρούμε ότι με μεγάλες τιμές stride έχουμε λιγότερη πληροφορία για το ποιες συχνότητες είναι σε κάποια χρονική στιγμή αλλά γίνεται καθαρότερος ο διαχωρισμός των συχνοτήτων.



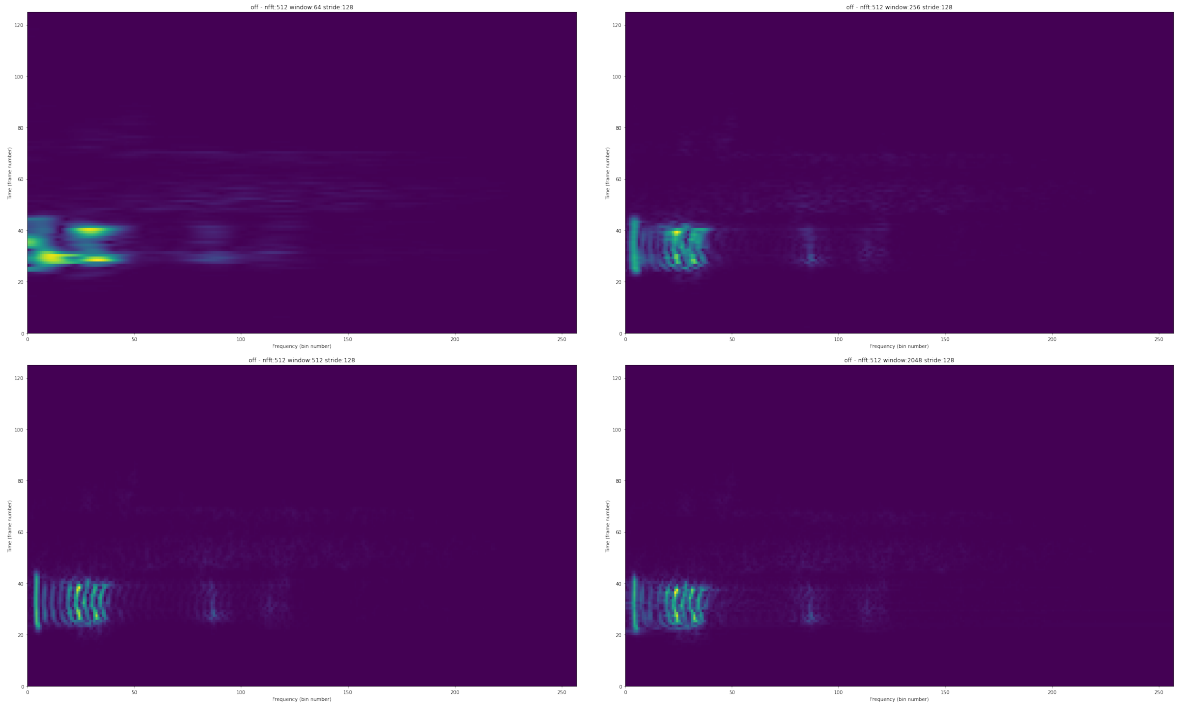
Σχήμα 3.4: Φασματογράφημα της εικ. 3.3 με διαφορετικές τιμές stride (64, 128, 1024, 2048)

Στην συνέχεια κρατώντας σταθερό το stride και το window μεταβάλουμε την τιμή του nfft (εικ. 3.5). Παρατηρούμε ότι έχουμε αλλαγή στον άξονά της συχνότητας και συγκεκριμένα αυξάνοντάς το nfft αυξάνουμε τον αριθμό των τμημάτων που διαχωρίζουμε την συχνότητα.



Σχήμα 3.5: Φασματογράφημα της εικ. 3.3 με διαφορετικές τιμές nfft (64, 128, 256, 512)

Στο τέλος δοκιμάσαμε να αλλάξουμε την παράμετρο window κρατώντας τις άλλες δυο σταθερές (εικ. 3.6).



Σχήμα 3.6: Φασματογράφημα της εικ. 3.3 με διαφορετικές τιμές window (64, 256, 512, 2048)

Οι τελικές παράμετροι που επιλέξαμε για την δημιουργία των δικών μας φασματογραφημάτων είναι nfft=256, window=512, stride=128 διότι παρατηρήσαμε ότι υπάρχει ένας καλός διαχωρισμός των συχνοτήτων χωρίς να έχουμε μεγάλη αβεβαιότητά στον χρόνο. Επιπλέον με αυτές τις παραμέτρους καταλαμβάνουν μικρή ποσότητά μνήμης το οποίο μπορεί να επιταχύνει την εκπαίδευση του μοντέλου μας.

Κεφάλαιο 4

Το μοντέλο για κατηγοριοποίηση ήχου

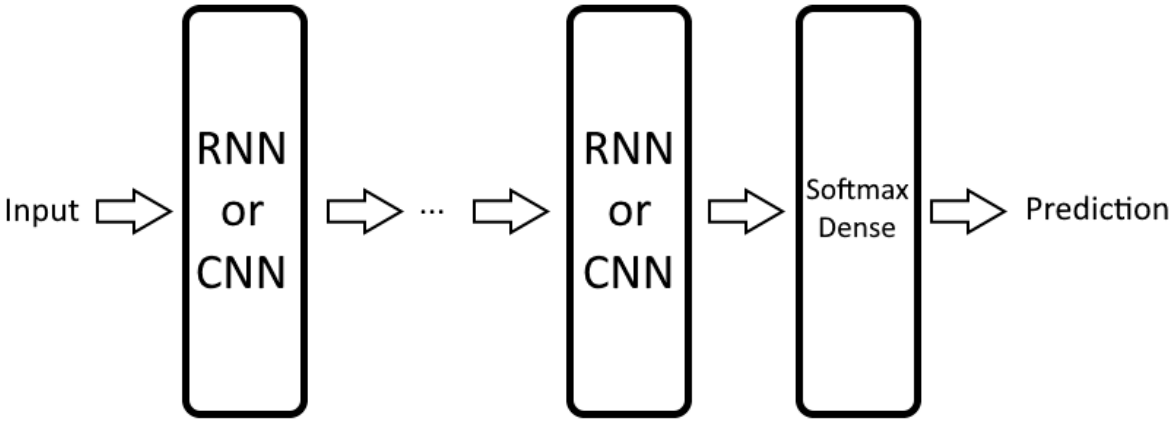
Στο πλαίσιο αυτής της διπλωματικής αναπτύχθηκαν μερικά μοντέλα βασισμένα σε νευρωνικά δίκτυα και για την ανάπτυξή τους χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python και συγκεκριμένα η βιβλιοθήκη TensorFlow. Τα μοντέλα και ο υπόλοιπος κώδικας μπορεί να βρεθεί στο παρακάτω repository `https://github.com/ikerastas/speech_commands_recognition`.

Το TensorFlow είναι μια μαθηματική βιβλιοθήκη, ανοιχτού κωδικά, που δημιουργήθηκε από την Google με σκοπό την διευκόλυνση ανάπτυξης κωδικά μαθηματικών πράξεων. Βρήκε πολύ μεγάλη χρήση στην δημιουργία μοντέλων μηχανικής μάθησης και συγκεκριμένα νευρωνικών δικτύων.

4.1 Αρχιτεκτονική του δικτύου

Η βασική ιδέα είναι ότι θα δημιουργήσουμε ένα δίκτυο οπού θα παίρνει σαν είσοδο το spectogram οπού περιγράψαμε στο προηγούμενο κεφάλαιο και θα αποτελείται από μερικά επίπεδά, είτε αναδρομικά είτε συνελικτικά, τα οποία θα μάθουν να κάνουν την κατηγοριοποίηση και στο τέλος θα υπάρχει ένα επίπεδο, dense, με συνάρτησή ενεργοποίησής την softmax οπού θα μας δίνει τις πιθανότητες να ανήκει η είσοδος σε κάποια από τις 12 κατηγορίες που έχουμε.

Έγιναν αρκετές δοκιμές με διαφορετικές αρχιτεκτονικές ή διαφορετικό αριθμό επίπεδων για την δημιουργία του δικτύου και μερικές παρουσιάζονται παρακάτω. Οι παρακάτω δοκιμές γίνανε με μέγεθος παρτίδας 256 και ρυθμό εκμάθησης με αρχική τιμή $0,01$. Επιπλέον ο ρυθμός εκμάθησης μειώνεται αυτόματα σε περίπτωση οπού δεν υπήρχε βελτίωσή στο δίκτυο



Σχήμα 4.1: Γενική δομή του δικτύου

για 4 συνεχόμενες εποχές μέχρι να φτάσει την τιμή $0,0001$ και η εκπαίδευσή σταματάει όταν για 12 συνεχόμενες εποχές το δίκτυο δεν παρουσιάσει βελτίωσή.

4.1.1 Αναδρομικό Νευρωνικό Δίκτυο

Αρχικά έγιναν δοκιμές με αναδρομικά νευρωνικά δίκτυα δίνοντας σαν είσοδο τα φασματογραφήματα που περιγράψαμε στην προηγουμένη ενότητα. Ένα φασματογράφημα μπορεί να θεωρηθεί μια χρονοσειρά και συνεπώς μπορεί να δοθεί σας είσοδο σε ένα τέτοιο δίκτυο.

Ενός επίπεδου

Όπως έχουμε αναφέρει προηγούμενος, για την δημιουργία αναδρομικών νευρωνικών δικτύων χρησιμοποιούμε είτε LSTM είτε GRU. Σαν πρώτο πείραμά κάναμε μια δοκιμή για να συμπεράνουμε ποιο από αυτά τα δυο έχει καλύτερή απόδοσή στα δικά μας δεδομένα. Για αυτόν τον σκοπό δημιουργήσαμε δυο ίδια νευρωνικά δίκτυα οπού αποτελούνται από ένα αναδρομικό επίπεδο με 256 νευρώνες, είτε GRU είτε LSTM, και ένα επίπεδο dense.

Για το δίκτυο GRU (εικ. 4.2) μπορούμε να παρατηρήσουμε ότι έχει 300300 παραμέτρους και η εκπαίδευσή σταματάει μετά από 29 εποχές με τα παρακάτω αποτελέσματά (πιν. 4.1).

Το ίδιο δίκτυο (εικ. 4.3) με νευρώνες LSTM έχει 398348 παραμέτρους και η εκπαίδευσή σταμάτησε μετά απο 32 εποχές. Επίσης λόγο των παραπάνω παραμέτρων κάθε εποχή σε αυτό το δίκτυο ήταν περίπου 20% με 25% πιο αργή. Τα αποτελέσματα παρουσιάζονται παρακάτω (πιν. 4.2). Από τα αποτελέσματα συμπεραίνουμε ότι είχαμε μια μικρή αύξησή στην ακρίβειά των προβλέψεων, η αύξηση αυτή όμως συνεπάγεται με μεγαλύτερο χρόνο εκπαίδευσης.

Παρατηρούμε ότι δεν υπάρχουν μεγάλες διαφορές μεταξύ των αποτελεσμάτων του δι-

1-Layer GRU with 256 neurons		
Split	Loss	Accuracy
Training	0.0288	0.9934
Validation	0.2898	0.9334
Testing	0.2933	0.9369

Πίνακας 4.1: Αποτελέσματα δοκιμής GRU με ένα επίπεδο και 256 νευρώνες.

```
Model: "gru_model_256"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
gru (GRU)                    (None, 256)               297216
_________________________________________________________________
output_dense (Dense)         (None, 12)                3084
=================================================================
Total params: 300,300
Trainable params: 300,300
Non-trainable params: 0
```

Σχήμα 4.2: Δομή του δικτύου GRU

κτύου GRU και του δικτύου LSTM. Όμως καλύτερο από τα δύο θεωρούμε το δίκτυο LSTM διότι στα δεδομένα που κρατήσαμε για να κάνουμε την τελική μας δοκιμή είχε ακρίβεια $95,29\%$ ενώ το GRU $93,69\%$.

Πολλών επιπέδων

Στην συνέχεια δοκιμάσαμε διάφορές αρχιτεκτονικές LSTM και GRU πολλών επίπεδων για να ελέγχουμε πως επηρεάζεται το αποτέλεσμά αν προσθέσουμε παραπάνω αναδρομικά επίπεδα. Δοκιμάσαμε πολυεπίπεδα LSTM και GRU μέχρι τέσσερα αναδρομικά επίπεδά και παρακάτω παρουσιάζονται τα αποτελέσματά στα δεδομένα που κρατήσαμε για την τελική δοκιμή μας.

Στα νευρωνικά δίκτυα που αποτελούνται απο GRU παρατηρούμε (πιν. 4.3) ότι η αύξησή των αναδρομικών επίπεδων σε δυο βελτιώνει το ποσοστό σωστών προβλέψεων κατά 2% περίπου. Όμως αυτή η βελτίωση δεν είναι αρκετή για να ξεπεράσουν το δίκτυο LSTM ενός επίπεδου και επιπλέον η αύξησή των επίπεδων σε παραπάνω από δυο δεν παρουσιάζει καμία βελτίωση και ο χρόνος εκπαίδευσής αυξάνεται αρκετά χωρίς κανένα πλεονέκτημά.

Στα LSTM νευρωνικά δίκτυα παρατηρούμε (πιν. 4.4) ότι η αύξησή των αναδρομικών επίπεδων σε δυο δεν βελτιώνει καθόλου το ποσοστό σωστών προβλέψεων, ενώ η αύξηση σε 3 και 4 παρουσιάσει βελτίωση κατά 1% περίπου. Όμως αυτή η βελτίωση έχει το μειονέκτημα

1-Layer LSTM with 256 neurons		
Split	Loss	Accuracy
Training	0.0239	0.9941
Validation	0.3472	0.9330
Testing	0.2423	0.9529

Πίνακας 4.2: Αποτελέσματα δοκιμής LSTM με ένα επίπεδο και 256 νευρώνες.

```
Model: "lstm_model_256"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 256)               395264
_________________________________________________________________
output_dense (Dense)         (None, 12)                3084
=================================================================
Total params: 398,348
Trainable params: 398,348
Non-trainable params: 0
```

Σχήμα 4.3: Δομή του δικτύου LSTM

του μεγαλυτέρου χρόνου εκπαίδευσης και για τόσο μικρή βελτίωση ο παραπάνω χρόνος εκπαίδευσής πιθανόν να μην αξίζει.

Αμφίδρομα επίπεδα

Μια σημαντική βελτίωσή μπορεί να παρουσιαστεί αν αλλάξουμε τον τρόπο που παρουσιάζουμε τα δεδομένα στο κάθε επίπεδο. Τα αναδρομικά νευρωνικά δίκτυα μπορούν να εξάγουν παραπάνω πληροφορίες από τα ήδη υπάρχοντα δεδομένα που έχουμε, αν τους τα παρουσιάζουμε και με αντίθετή χρονολογική σειρά.

Για αυτόν τον σκοπό υπάρχει το αμφίδρομο επίπεδο (bidirectional layer) το οποίο αποτελείται από δύο αναδρομικά επίπεδά, ένα ”βλέπει” τα δεδομένα μας με την κανονική σειρά και ένα οπού να βλέπει με την αντίθετή. Στο τέλος το αποτέλεσμα από αυτά τα δύο επίπεδά συνενώνεται και παρουσιάζεται σαν ένα.

Στις δοκιμές που έγιναν δεν παρατηρήθηκε βελτίωσή στο ποσοστό ακρίβειας χρησιμοποιώντας αμφίδρομα επίπεδα. Ειδικότερα χρησιμοποιώντας ένα αμφίδρομο επίπεδο LSTM (εικ. 4.5) είδαμε (πιν. 4.5) αντίστοιχα ή και χειρότερα αποτελέσματά σε σχέση με την χρήση ενός επίπεδου LSTM (πιν. 4.2) ενώ ταυτόχρονα ο χρόνος εκπαίδευσής αυξήθηκε αρκετά λόγο της αυξημένης πολυπλοκότητας του δικτύου (200% παραπάνω παράμετροι).

Επιπλέον αυξάνοντάς τον αριθμό των επίπεδων σε δυο και σε τρία (πιν. 4.6), η ακρίβεια

Multi-Layer GRU with 256 neurons		
Number of layers	Testing loss	Testing accuracy
2	0.2153	0.9520
3	0.1908	0.9584
4	0.1954	0.9573

Πίνακας 4.3: Αποτελέσματα δοκιμής GRU με πολλά επίπεδα και 256 νευρώνες.

Multi-Layer LSTM with 256 neurons		
Number of layers	Testing loss	Testing accuracy
2	0.2084	0.9566
3	0.1920	0.9644
4	0.2121	0.9652

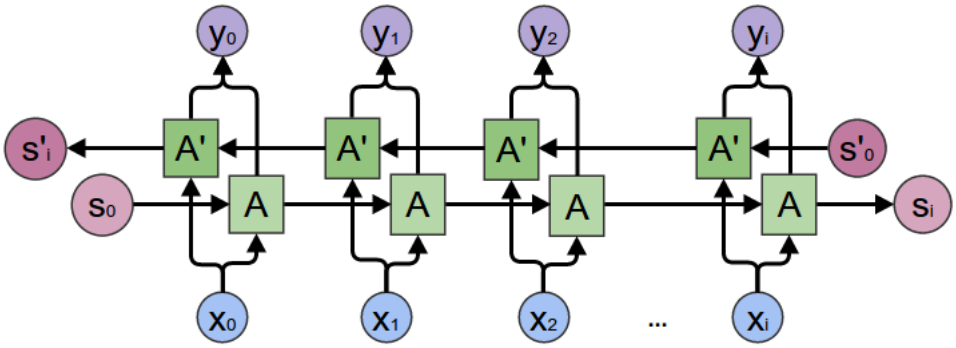
Πίνακας 4.4: Αποτελέσματα δοκιμής LSTM με πολλά επίπεδα και 256 νευρώνες.

σε σχέση με πριν είχε μια μικρή βελτίωσή αλλά αυτή η βελτίωση μπορεί να οφείλεται σε πολλούς διάφορους παράγοντες (πχ. αρχικοποίηση βαρών) και όχι στην αύξηση των επίπεδων. Σε σύγκριση με τα απλά LSTM δεν παρουσιάστηκε βελτίωση και επιπλέον τα πολυεπίπεδα LSTM παρουσίασαν μεγαλύτερη ακρίβειά και γρηγορότερους χρόνους εκπαίδευσης ανά εποχή.

Ακριβώς ίδια αποτελέσματά είχαμε και με αμφίδρομα GRU (πιν. 4.7), αν και είχανε, οριακά, καλύτερή απόδοση απο τα αντίστοιχα αμφίδρομα LSTM.

Αριθμός νευρώνων και ακρίβεια

Στην συνέχεια θέλουμε να ελέγχουμε πως επηρεάζεται η ακρίβεια του δικτύου αλλάζοντάς τον αριθμό των νευρώνων και πραγματοποιήσαμε δοκιμές μόνο με ένα δίκτυο LSTM ενός επίπεδου και με ένα δίκτυο αμφίδρομου LSTM δύο επίπεδων. Διαλέξαμε μονό αυτά διότι λόγω της αύξησης των νευρώνων στα πολυεπίπεδα δίκτυα η πολυπλοκότητα του δικτύου και ο αριθμός των παραμέτρων αυξάνονται ραγδαία (πιν. 4.10) και με τους διαθέσιμου πόρους δεν ήταν εφικτή η εκπαίδευση μεγαλυτέρων δικτύων.



Σχήμα 4.4: Δομή αμφίδρομου επίπεδου

```
Model: "lstm_bi_model_256"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
bidirectional_lstm (Bidirectional)  (None, 512)           790528
_________________________________________________________________
output_dense (Dense)         (None, 12)                6156
=================================================================
Total params: 796,684
Trainable params: 796,684
Non-trainable params: 0
```

Σχήμα 4.5: Δομή του αμφίδρομου δικτύου LSTM

4.1.2 Συνελικτικό νευρωνικό δίκτυο

Στην συνέχεια έγινε δοκιμή ενός δικτύου με συνελικτικά επίπεδα. Όπως έχει αναφερθεί, η είσοδος στο δίκτυο μας είναι φασματογραφήματα απο τα αρχεία ήχου που έχουμε. Αυτά τα φασματογραφήματα είναι πίνακες 2 διαστάσεων και μπορούν να θεωρηθούν εικόνες και συνεπώς μπορούμε να τα παρουσιάσουμε σαν είσοδο σε ένα συνελικτικό νευρωνικό δίκτυο.

Έγιναν δοκιμές με δυο συνελικτικά νευρωνικά δίκτυα με την μονή διαφορά να είναι στον αριθμό των νευρώνων σε κάθε επίπεδο. Η γενικότερη δομή των συνελικτικών δικτύων, που δοκιμάσαμε, παρουσιάζετε στην παρακάτω εικόνα (εικ. 4.6).

Στην πρώτη δοκιμή δημιουργήσαμε ένα δίκτυο οπού αποτελείται από τέσσερα συνελικτικά επίπεδα μεγέθους 32, 64, 128, 256 (πιν. 4.11) ενώ στην δεύτερη με πέντε επίπεδα μεγέθους 32, 64, 128, 256, 512 (πιν. 4.12).

Παρατηρήσαμε ότι και αυτά τα δίκτυα είχανε αποτελέσματα εφάμιλλα ή και καλύτερα από τα αναδρομικά δίκτυα.

1-Layer BI-LSTM with 256 neurons		
Split	Loss	Accuracy
Training	0.0183	0.9957
Validation	0.3075	0.9374
Testing	0.2578	0.9438

Πίνακας 4.5: Αποτελέσματα δοκιμής Bi-LSTM με ένα επίπεδο και 256 νευρώνες.

2-Layer Bi-LSTM with 256 neurons		
Split	Loss	Accuracy
Training	0.0056	0.9986
Validation	0.3103	0.9454
Testing	0.2709	0.9526

3-Layer BI-LSTM with 256 neurons		
Split	Loss	Accuracy
Training	0.0028	0.9994
Validation	0.2912	0.9526
Testing	0.2518	0.9564

Πίνακας 4.6: Αποτελέσματα δοκιμής Bi-LSTM με δύο και τρία επίπεδα και 256 νευρώνες.

Multi-Layer BI-GRU with 256 neurons		
Number of layers	Testing loss	Testing accuracy
1	0.2592	0.9459
2	0.2560	0.9533
3	0.2230	0.9604

Πίνακας 4.7: Αποτελέσματα δοκιμής BI-GRU με πολλά επίπεδα και 256 νευρώνες.

Single-Layer LSTM		
Number of neurons	Testing loss	Testing accuracy
512	0.3208	0.9439
1024	0.3106	0.9520

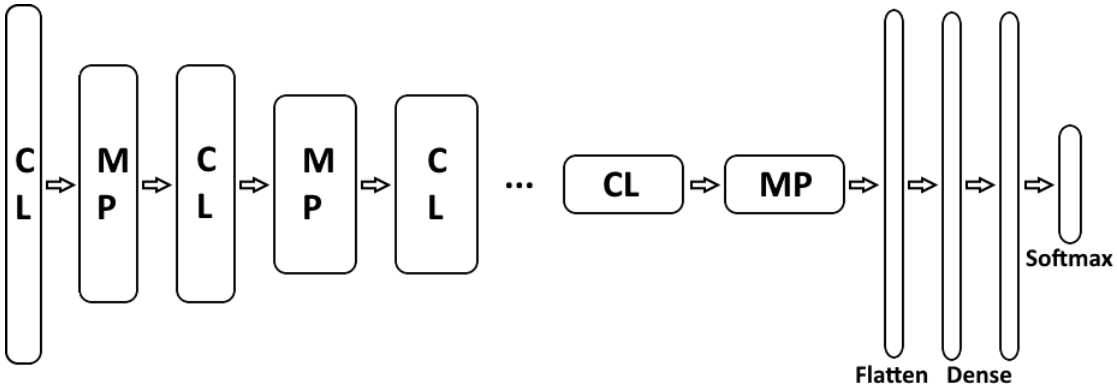
Πίνακας 4.8: Αποτελέσματα δοκιμής LSTM τεσσάρων επίπεδων

2-Layer BI-LSTM		
Number of neurons	Testing loss	Testing accuracy
512	0.2609	0.9548
1024	0.2000	0.9693

Πίνακας 4.9: Αποτελέσματα δοκιμής BI-LSTM δύο επίπεδων

Αριθμός παραμέτρων δικτύου LSTM ενός επίπεδου	
Number of neurons	Παράμετροι
128	267,276
256	796,684
512	2,641,932
1024	9,478,156

Πίνακας 4.10: Εκθετική αύξηση παραμέτρων LSTM



Σχήμα 4.6: Δομή συνελικτικού νευρωνικού δικτύου. CL: Convolutional Layer, MP: Max Pooling

Συνελικτικό Νευρωνικό Δίκτυο 32-64-128-256		
Split	Loss	Accuracy
Training	0.0527	0.9844
Validation	0.2469	0.9498
Testing	0.2057	0.9542

Πίνακας 4.11: Αποτελέσματα δοκιμής συνελικτικού νευρωνικού 4 επίπεδων

Συνελικτικό Νευρωνικό Δίκτυο 32-64-128-256-512		
Split	Loss	Accuracy
Training	0.0095	0.9998
Validation	0.2922	0.9596
Testing	0.2620	0.9614

Πίνακας 4.12: Αποτελέσματα δοκιμής συνελικτικού νευρωνικού 5 επίπεδων

Κεφάλαιο 5

Συμπεράσματα

Σε αυτό το κεφάλαιο θα γίνει μια σύνοψη των αποτελεσμάτων και θα συγκριθούν τα διάφορα μοντέλα που δοκιμάστηκαν με διάφορα κριτήρια όπως η ακρίβεια και ο χρόνος εκπαίδευσης.

5.1 Σύνοψη και συμπεράσματα

Σε αυτήν την διπλωματική δοκιμάστηκαν δυο είδη νευρωνικών δικτύων για να αντιμετωπίσουμε το πρόβλημα της αναγνώρισης φωνητικών εντολών. Το πρόβλημα αυτό είναι πολυδιάστατο και απαιτούνται αρκετά βήματα ώστε να λυθεί. Το πρώτο βήμα είναι να γίνει προ-επεξεργασία των δεδομένων ήχου που έχουμε ώστε να μπορέσουμε να εξάγουμε χαρακτηριστικά που θα επιτρέψουν σε κάποιον αλγόριθμο μηχανικής μάθησης να κατηγοριοποιήσει τα δεδομένα μας.

Η τρόπος που επιλέξαμε να εξάγουμε χαρακτηριστικά είναι μέσω της δημιουργίας ενός φασματογραφήματος για κάθε αρχείο ήχου. Η διαδικασία αυτή βασίζεται πάνω στον μετασχηματισμό Fourier συντόμου χρόνου και απαιτεί τον καθορισμό αρκετών παραμέτρων που μπορούν να επηρεάσουν το τελικό φασματογράφημα.

Στην συνέχεια σχεδιάσαμε και δοκιμάσαμε διάφορα νευρωνικά δίκτυα που βασίζονται είτε σε αναδρομικά επίπεδα, είτε σε συνελικτικά επίπεδα. Στον πίνακα (πιν. 5.1) αναφέρονται περιληπτικά όλα τα αποτελέσματα. Μπορούμε να παρατηρήσουμε ότι σε γενικότερες γραμμές τα δίκτυα που βασίζονται σε LSTM είχανε καλύτερα αποτελέσματα. Η χρήση αμφίδρομων επίπεδων βελτίωσε ακόμα παραπάνω την ακρίβεια αλλά προκάλεσε μεγάλη αύξηση στον αριθμό των παραμέτρων και συνεπώς και στον χρόνο εκπαίδευσης. Για αυτόν

τον λόγο αμφίδρομα δίκτυα με περισσότερα επίπεδά ή μεγαλύτερο αριθμό νευρώνων δεν ήταν δυνατό να εκπαιδεύσουμε λόγω περιορισμένων πόρων που είχαμε στην διάθεση μας. Επιπλέον πολύ καλά αποτελέσματα είδαμε και απο τα συνελικτικά δίκτυα τα οποία μπόρεσαν και πλησίασαν αρκετά τα καλύτερα αναδρομικά δίκτυα. Όμως και αυτά είχανε αρκετά μεγάλο αριθμό παραμέτρων και ο χρόνος εκπαίδευσης ήταν αυξημένος άλλα όχι στο ίδιο επίπεδο με τα πολυεπίπεδα αμφίδρομα αναδρομικά.

Σύνοψη των αποτελεσμάτων	
Είδος δικτύου	Ακρίβεια
1-Layer-GRU	0.9369
2-Layer-GRU	0.9520
3-Layer-GRU	0.9584
4-Layer-GRU	0.9573
1-Layer-LSTM	0.9529
2-Layer-LSTM	0.9566
3-Layer-LSTM	0.9644 ***
4-Layer-LSTM	0.9652 **
1-Layer-BI-GRU	0.9459
2-Layer-BI-GRU	0.9533
3-Layer-BI-GRU	0.9604
1-Layer-BI-LSTM	0.9438
2-Layer-BI-LSTM	0.9526
3-Layer-BI-LSTM	0.9564
1-Layer-LSTM-512	0.9439
1-Layer-LSTM-1024	0.9520
2-Layer-BI-LSTM-512	0.9548
2-Layer-BI-LSTM-1024	0.9693 *
4-Layer-CNN	0.9542
5-Layer-CNN	0.9614 ****

Πίνακας 5.1: Το είδος του δικτύου και η ακρίβεια στο κομμάτι των δεδομένων δοκιμής. Με * σημειώνονται τα δίκτυα με την μεγαλύτερή ακρίβειά.

5.2 Μελλοντικές επεκτάσεις

Μια επέκταση αυτής της διπλωματικής θα μπορούσε να είναι στον τομέα του Ίντερνετ των πραγμάτων (IoT). Η αναγνώριση φωνητικών εντολών σε τέτοιες συσκευές θα ενισχύσει αρκετά την εμπειρία που έχει ο χρήστης και τον τρόπο οπού θα αλληλεπιδρά με αυτές. Βέβαια σε συσκευές IoT, οπού είναι αρκετά περιορισμένες σε πόρους, θα πρέπει να ακολουθηθεί μια διαφορετική λογική ώστε να βρεθεί το δίκτυο οπού απαιτεί τους λιγότερους πόρους άλλα και ταυτόχρονα έχει την καλύτερή επίδοση. Επιπλέον διαφορετικά βήματα προ-επεξεργασίας θα μπορούσαν να βοηθήσουν την απόδοσή του δικτύου, όπως για παράδειγμά αντί για δημιουργία ενός φασματογραφήματος θα μπορούσε να δημιουργηθεί ένα φασματογράφημα κλίμακάς mel, ή να γίνει εξαγωγή συντελεστών mel-frequency cepstrum (MFCCs).

Βιβλιογραφία

[1] Displayr. https://www.displayr.com/what-is-a-decision-tree/. Ημερομηνία πρόσβασης: 24-2-2021.

[2] Wikipedia - cluster analysis. https://en.wikipedia.org/wiki/Cluster_analysis.

[3] Scikit-learn.org. https://scikit-learn.org/. Ημερομηνία πρόσβασης: 24-2-2021.

[4] Wikipedia - neuron. https://en.wikipedia.org/wiki/Neuron. Ημερομηνία πρόσβασης: 24-2-2021.

[5] allaboutcircuits - how to train a multilayer perceptron neural network. https://www.allaboutcircuits.com/technical-articles/how-to-train-a-multilayer-perceptron-neural-network/. Ημερομηνία πρόσβασης: 24-2-2021.

[6] A comprehensive guide to convolutional neural networks. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[7] Basics of the classic cnn. https://towardsdatascience.com/basics-of-the-classic-cnn-a3dce1225add. Ημερομηνία πρόσβασης: 24-2-2021.

[8] Wikipedia - convolutional neural network. https://en.wikipedia.org/wiki/Convolutional_neural_network. Ημερομηνία πρόσβασης: 24-2-2021.

[9] Understanding lstm networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/. Ημερομηνία πρόσβασης: 24-2-2021.

[10] Wikipedia - long short-term memory. `https://en.wikipedia.org/wiki/Long_short-term_memory`. Ημερομηνία πρόσβασης: 24-2-2021.

[11] Wikipedia - gated recurrent unit. `https://en.wikipedia.org/wiki/Gated_recurrent_unit`. Ημερομηνία πρόσβασης: 24-2-2021.

[12] Yifan Zhang, Peter Fitch, Maria Vilas, and Peter Thorburn. Applying multi-layer artificial neural network and mutual information to the prediction of trends in dissolved oxygen. *Frontiers in Environmental Science*, 7, 04 2019.

[13] Gradient descent algorithm and its variants. `https://towardsdatascience.com/gradient-descent-algorithm-and-its-variants-10f652806a3`. Ημερομηνία πρόσβασης: 24-2-2021.

[14] M.J.F. Gales and Steve Young. The application of hidden markov models in speech recognition. *Foundations and Trends in Signal Processing*, 1:195–304, 01 2007.

[15] H. Lee, S. Chang, D. Yook, and Y. Kim. A voice trigger system using keyword and speaker recognition for mobile devices. *IEEE Transactions on Consumer Electronics*, 55(4):2377–2384, 2009.

[16] C. Hale and C. Nguyen. Voice command recognition using fuzzy logic. *Proceedings of WESCON'95*, pages 608–, 1995.

[17] Patrick Jansson. Single-word speech recognition with convolutional neural networks on raw waveforms. In *theseus.fi*, 2018.

[18] Sushan Poudel and Anu Radha. Speech command recognition using artificial neural networks. *JOIV : International Journal on Informatics Visualization*, 4, 04 2020.

[19] Mariusz Kubanek, Janusz Bobulski, and Joanna Kulawik. A method of speech coding for speech recognition using a convolutional neural network. *Symmetry*, 11:1185, 09 2019.

[20] Π. Καραντζιάς. Εφαρμογή αλγορίθμων μηχανικής μάθησης σε σύνολα δεδομένων και αποτίμηση των αποτελεσμάτων. Master's thesis, Πανεπιστήμιο Πειραιά, Φεβρουάριος 2019.

[21] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, September 1995.

[22] Σ. Παπαποστόλου. Κατηγοριοποίηση με μηχανές διανυσμάτων υποστήριξης. Master's thesis, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 2017.

[23] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, pages 65–386, 1958.

[24] Λύκας - Τεχνητά Νευρωνικά Δίκτυα. `http://www.cs.uoi.gr/~arly/courses/nn/slides/K4.pdf`. Ημερομηνία πρόσβασης: 24-2-2021.

[25] Ι. Δημαρίδης. Εξαγωγή Χαρακτηριστικών με Βαθιά Συνελικτικά Δίκτυα για την Αισθητική Αξιολόγηση Εικόνων. Διπλωματική εργασία, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Μάρτιος 2020.

[26] Razvan Pascanu, Tomás Mikolov, and Yoshua Bengio. Understanding the exploding gradient problem. *CoRR*, abs/1211.5063, 2012.

[27] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[28] Γ. Μήτσιος. Αναδρομικά Νευρωνικά Δίκτυα και αυτόματη παραγωγή hashtag από tweet του twitter. Διπλωματική εργασία, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.

[29] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014.

[30] Γ. Καραντώνης. Ανάπτυξή συστήματος απάντησης ερωτημάτων. Διπλωματική εργασία, Πανεπιστήμιο Πατρών.

[31] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.

[32] Α. Παπαδόπουλος. Πρόβλεψη Τροχιών σε Δεδομένα Κίνησης με Βαθιά Νευρωνικά Δίκτυα. Πανεπιστήμιο Πειραιά, Δεκέμβριος 2018.

[33] Θ. Φλωράκης. Ανάπτυξη Αρχιτεκτονικών Βαθιάς Μάθησης με Χρήση Δικτύων Μακράς Βραχυπρόθεσμης Μνήμης Για Ανάλυση Συναισθήματος. Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, Νοέμβριος 2018.

[34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning Internal Representations by Error Propagation*, page 318–362. MIT Press, Cambridge, MA, USA, 1986.

[35] Ταξιάρχης Διαμαντόπουλος Θεόδωρος Λώτης. *Μουσική Πληροφορική & Μουσική με υπολογιστές*. Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, 2015.

[36] Wikipedia - Ανάλυση Φουριέ. `https://el.wikipedia.org/wiki/%CE%91%CE%BD%CE%AC%CE%BB%CF%85%CF%83%CE%B7_%CE%A6%CE%BF%CF%85%CF%81%CE%B9%CE%AD`. Ημερομηνία πρόσβασης: 24-2-2021.

[37] Ψηφιακή Επεξεργασία Σήματος 26. `https://www.csd.uoc.gr/~hy370/w19/Slides/HY370-Lec26.pdf`. Ημερομηνία πρόσβασης: 15-1-2021.

[38] P. Warden. Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition. *ArXiv e-prints*, April 2018.