



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Σχολή Ηλεκτρολόγων Μηχανικών  
και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Χρήση συνελικτικών νευρωνικών δικτύων για τον εντοπισμό του Αλ-  
τςχαϊμερ σε ασθενείς μέσα από μαγνητικές εγκεφάλου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΙΩΡΓΟΣ Ε. ΚΑΡΟΖΗΣ

Επιβλέπων : Ανδρέας-Γεώργιος Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π

Συνεπιβλέπων : Γεώργιος Σιόλας

Αθήνα, Σεπτέμβριος 2021





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Σχολή Ηλεκτρολόγων Μηχανικών

και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Χρήση συνελικτικών νευρωνικών δικτύων για τον εντοπισμό του Αλ-  
τςχαϊμερ σε ασθενείς μέσα από μαγνητικές εγκεφάλου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΙΩΡΓΟΣ Ε. ΚΑΡΟΖΗΣ

Επιβλέπων : Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π

Συνεπιβλέπων : Γεώργιος Σιόλας

Αθήνα, Σεπτέμβριος 2021



(Υπογραφή)

.....  
**Γιώργος Καρόζης**

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

Copyright © Γεώργιος Καρόζης, 2021.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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



## Περίληψη

Εδώ και αρκετά χρόνια η μηχανική μάθηση αποτελεί έναν τομέα απαραίτητο για την επιστήμη της ιατρικής τόσο για την περίθαλψη όσο και για την πρόληψη αρκετών ασθενειών. Μέσα σε αυτές τις ασθένειες περιλαμβάνεται και η νόσος του Alzheimer.

Το 2019 εκτιμήθηκε πως ο αριθμός των ανθρώπων που έχουν άνοια είναι περίπου 48 εκατομμύρια, καθώς επίσης προβλέπεται ότι ο αριθμός αυτός θα αυξηθεί αρκετά στο άμεσο μέλλον. Το Alzheimer αποτελεί την πιο κοινή μορφή άνοιας και επιπλέον μια μάστιγα των Η.Π.Α, εφόσον 6,2 εκατομμύρια άνθρωποι πάσχουν από την συγκεκριμένη νόσο. Για αυτούς τους λόγους οι προσπάθειες καταπολέμησης της γίνονται πιο εντατικές. Βέβαια αξίζει να αναφερθεί πως δεν έχει βρεθεί κάποια θεραπεία, αλλά η πρόωρη διάγνωση της μπορεί να καταστείλει την ασθένεια σε μεγάλο βαθμό και για μεγαλύτερο χρονικό διάστημα, καθώς επίσης να βοηθήσει τους ειδικούς να αντιληφθούν με ποιον τρόπο λειτουργεί, ώστε να βρεθεί η θεραπεία στο μέλλον. Για το κομμάτι της πρόληψης του Alzheimer, λοιπόν, χρησιμοποιήσαμε μοντέλο βαθιάς μάθησης.

Πιο συγκεκριμένα, τα πειράματά μας εκτελούνται πάνω στο αποθετήριο δεδομένων Alzheimer's Disease Neuroimaging Initiative (ADNI). Το ADNI αποτελεί μια προσεγμένη βάση δεδομένων με ιατρικά στοιχεία ασθενών, η οποία δημιουργήθηκε από τον Michael W. Weiner το 2003. Σκοπός του ADNI είναι η μελέτη του κατά πόσον οι μαγνητικές τομογραφίες, οι τομογραφίες εκπομπής ποζιτρονίων, άλλοι βιολογικοί δείκτες και νευροψυχολογικές αξιολογήσεις μπορούν να συνδυαστούν για να μετρηθεί η πρόοδος των MCI και AD.

Στον σκοπό της παραπάνω προσπάθειας δημιουργήθηκε ένα συνελικτικό νευρωνικό δίκτυο(CNN), το οποίο στοχεύει στον διαχωρισμό των υγιών ατόμων(CN), των ατόμων που έχουν κάποια έλλειψη μνήμης(MCI), καθώς και εκείνων που έχουν διαγνωστεί με Alzheimer (AD). Το συνελικτικό νευρωνικό δίκτυο λαμβάνει ως δεδομένα μαγνητικές εγκεφάλου (MRI), προκειμένου να γίνει ο διαχωρισμός. Οι μαγνητικές αυτές λήφθηκαν από το προαναφερθέν αποθετήριο δεδομένων ADNI.

## Λέξεις Κλειδιά

Μηχανική μάθηση, βαθιά μάθηση, νευρωνικά δίκτυα, συνελικτικά νευρωνικά δίκτυα, Alzheimer, ταξινόμηση εικόνων, μαγνητικές εγκεφάλου, κατάτμηση εγκεφαλικών εικόνων





## Abstract

For many years machine learning forms an essential field of medical science and is useful for healthcare as well as the prevention of many diseases. One of those diseases is Alzheimer's disease.

In 2019 it was estimated that the number of people that have dementia is about 48 million, as well it is predicted that this number will increase dramatically around the corner. Alzheimer's disease comprises the most common form of dementia and blight of the U.S.A. since 6.2 million people suffer because of it. Many efforts take place to cure this disease for all these reasons. It is worth mentioning that there is no specific cure for this disease. However early prevention can confine it for a long time and additionally can help specialists realize the way it is functioning, in order to find the cure in the future. For the sake of prevention, we used a deep learning model.

More specifically the experiments were made through the library of Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI is a carefully created database that consists of medical documentation of patients and was created by Michael W. Weiner in 2003. The overall goal of ADNI is to test whether serial magnetic resonance imaging (MRI), position emission tomography (PET), other biological markers, and clinical and neuropsychological assessment could be combined to measure the progression of MCI and early AD.

For this purpose, a convolutional neural network (CNN) was designed, in order to distinguish control normals (CN), people with mild cognitive impairment (MCI), and those who have been diagnosed with Alzheimer's disease(AD). The convolutional neural network takes magnetic resonance images as an input for the above purpose. All these MRIs were taken from the ADNI database.

## Key Words

Machine Learning, Deep Learning, Neural Networks, Convolutional Neural Networks, Alzheimer, Image Classification, Magnetic resonance images, Brain Mri Segmentation



## Ευχαριστίες

Θα ήθελα αρχικά να ευχαριστήσω τον επιβλέποντα καθηγητή αυτής της εργασίας, κύριο Ανδρέα Σταφυλοπάτη που μου έδωσε την δυνατότητα να εκπονήσω την διπλωματική μου εργασία στο Εργαστήριο Ευφυών Συστημάτων του τομέα Τεχνολογίας Πληροφορικής και Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Ιδιαίτερες ευχαριστίες θα ήθελα να δώσω στον κύριο Γεώργιο Σιόλα για την πολύτιμη στήριξη και καθοδήγηση του καθ' όλη την διάρκεια της εκτέλεσης της διπλωματικής εργασίας, καθώς και την άμεση ανταπόκριση του σε όποια δυσκολία αντιμετώπιζα.

Στην συνέχεια θα ήθελα να ευχαριστήσω τους καθηγητές του Εθνικού Μετσόβιου Πολυτεχνείου, οι οποίοι, στο μεγαλύτερο ποσοστό, είχαν μεράκι για την δουλειά τους και την μετάδοση της γνώσης στους φοιτητές.

Η συνεισφορά ειδικών και επιστημόνων έπαιξαν καθοριστικό παράγοντα. Ανφέρομαι στην βοήθεια που έδωσαν εν αγνοία τους μέσα από τον κόσμο του διαδικτύου, η οποία ήταν πολύτιμη.

Κλείνοντας θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου που είναι δίπλα μου και με στηρίζουν όλα αυτά τα χρόνια.



# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>19</b>
1.1	Κίνητρο . . . . .	20
1.2	Νόσος Αλτςχάϊμερ . . . . .	21
1.2.1	Διάγνωση . . . . .	21
1.2.2	Ήπια νοητική βλάβη . . . . .	22
1.3	Μαγνητική τομογραφία . . . . .	22
1.4	Στόχοι της διπλωματικής εργασίας . . . . .	23
1.5	Σχετική έρευνα . . . . .	23
<b>2</b>	<b>Μηχανική Μάθηση</b>	<b>25</b>
2.1	Ο ρόλος της μηχανικής μάθησης στην ιατρική . . . . .	25
2.2	Τι είναι μηχανική μάθηση . . . . .	26
2.2.1	Προβλήματα που αντιμετωπίζει η μηχανική μάθηση . . . . .	27
2.2.2	Τύποι μάθησης . . . . .	27
2.2.3	Πώς λειτουργεί η επιβλεπόμενη μάθηση . . . . .	28
2.2.4	Διαφορά προβλημάτων ταξινόμησης και παλινδρόμησης . . . . .	29
2.3	Στάδια Μάθησης ( Learning Stages) . . . . .	32
<b>3</b>	<b>Αλγόριθμοι Μηχανικής Μάθησης</b>	<b>35</b>
3.1	Ο αλγόριθμος Naive-Bayes . . . . .	35
3.2	Μηχανές διανυσματικής υποστήριξης SVM . . . . .	38
3.3	kNN . . . . .	41
3.4	Δέντρα αποφάσεων . . . . .	42
3.5	Τυχαία Δάση - Random Forests . . . . .	44
<b>4</b>	<b>Νευρωνικά δίκτυα</b>	<b>47</b>
4.1	Τεχνητά νευρωνικά δίκτυα . . . . .	47
4.2	Συναρτήσεις ενεργοποίησης . . . . .	48

4.3	Δίκτυα Perceptron . . . . .	49
4.4	Δίκτυα Πρόσθιας Τροφοδότησης . . . . .	50
4.4.1	Αλγόριθμος οπισθοδιάδοσης σφάλματος . . . . .	51
<b>5</b>	<b>Συνελικτικά νευρωνικά δίκτυα (CNN)</b>	<b>55</b>
5.1	Αρχές Λειτουργίας . . . . .	55
5.1.1	Συνελιξεις . . . . .	56
5.1.2	Συγκέντρωση(Pooling) . . . . .	57
5.1.3	Μη γραμμικότητα . . . . .	57
5.2	Dropout . . . . .	58
5.2.1	Normalization . . . . .	59
5.3	Προ-εκπαίδευση . . . . .	61
<b>6</b>	<b>Πειραματική διαδικασία</b>	<b>63</b>
6.1	Συλλογή και Προεπεξεργασία δεδομένων . . . . .	63
6.1.1	Συλλογή . . . . .	63
6.1.2	Προεπεξεργασία . . . . .	64
6.2	Μεθοδολογία . . . . .	68
6.2.1	Επιλογή Μοντέλου . . . . .	68
6.2.2	Πρόωρο Σταμάτημα (Early Stopping) . . . . .	70
6.2.3	Αποτελέσματα-Εκτίμηση του μοντέλου . . . . .	75
<b>7</b>	<b>Συμπεράσματα και Μελλοντική Έρευνα</b>	<b>81</b>

# Κατάλογος Σχημάτων

1.1	Πρόβλεψη αύξησης πληθυσμού . . . . .	20
1.2	Saggital, axial and coronal planes . . . . .	22
2.1	Πρόβλεψη της βαθμολογίας με ταξινόμηση και παλινδρόμηση . . . . .	30
3.1	Η συνάρτηση $\text{sign}(x)$ . . . . .	39
3.2	Μοντέλο SVM για διανύσματα χαρακτηριστικών δύο διαστάσεων . . . .	40
3.3	Σχήμα knn για $k=1$ και $k=3$ . . . . .	41
3.4	Δέντρο απόφασης . . . . .	43
3.5	random forest . . . . .	46
4.1	Ένα απλό Τεχνητό Νευρωνικό Δίκτυο . . . . .	48
4.2	Γραφική παράσταση σιγμοειδούς συνάρτησης . . . . .	49
4.3	Κλασσικό δίκτυο πρόσθιας τροφοδότησης . . . . .	51
4.4	Εκπαίδευση δικτύου με οπισθοδιάδοση . . . . .	52
5.1	Εφαρμογή φίλτρου σε ένα patch . . . . .	56
5.2	Εφαρμογή MaxPooling . . . . .	57
5.3	Εφαρμογή Dropout . . . . .	59
6.1	Αρχική μαγνητική . . . . .	65
6.2	Κατόπιν αφαίρεσης των εικονοστοιχείων . . . . .	65
6.3	Μαγνητική του σετ δεδομένων . . . . .	66
6.4	Στραμμένη κατά 10 μοίρες . . . . .	66
6.5	Στραμμένη κατά -10 μοίρες . . . . .	66
6.6	Μαγνητική του σετ δεδομένων . . . . .	67
6.7	Μετατοπισμένη εικόνα . . . . .	67
6.8	Εικόνα με θόρυβο . . . . .	67
6.9	Πίνακας σύγκρισης πολλών συνελικτικών επιπέδων . . . . .	69

6.10	Train and validation loss before dropout . . . . .	70
6.11	Train and validation accuracy after dropout in dense layer . . . . .	71
6.12	Train and validation loss after dropout in dense layer . . . . .	71
6.13	Train and validation accuracy after dropout in conv layers . . . . .	72
6.14	Train and validation loss after dropout in conv layers . . . . .	72
6.15	Train and validation accuracy after dropout + regularization . . . . .	73
6.16	Train and validation loss after dropout + regularization . . . . .	73
6.17	Train and validation accuracy for new dataset . . . . .	74
6.18	Train and validation loss for new dataset . . . . .	74
6.19	Πίνακας σύγκρισης του μοντέλου μας . . . . .	76
6.20	Μετρικές της πρώτης περίπτωσης . . . . .	77
6.21	35 epochs training (accuracy) . . . . .	77
6.22	35 epochs training (loss) . . . . .	78
6.23	Πίνακας σύγκρισης δεύτερης περίπτωσης . . . . .	78
6.24	Μετρικές δεύτερης περίπτωσης . . . . .	79
6.25	Πίνακας σύγκρισης τρίτης περίπτωσης . . . . .	79
6.26	Μετρικές τρίτης περίπτωσης . . . . .	80



# Κατάλογος Πινάκων

6.1	Dataset review . . . . .	66
6.2	Dataset review2 . . . . .	67
6.3	Train, Test and Validation Sets . . . . .	68
6.4	Χαρακτηριστικά συνελκτικού δικτύου . . . . .	69
6.5	New Train, Test and Validation Sets . . . . .	74



# Κεφάλαιο 1

## Εισαγωγή

Ζούμε σε μια εποχή στην οποία μπορούμε να αντιληφθούμε την έντονη ανάγκη της συνεργασίας μεταξύ ιατρικής και τεχνολογίας. Η έξαρση της πανδημίας του κορονοϊού μας δείχνει πόσο σημαντική είναι η συμβολή της τεχνολογίας στον τομέα της ιατρικής. Έτσι και η μηχανική μάθηση με την σειρά της παίζει πολύ σημαντικό ρόλο στην πρόληψη και περίθαλψη πολλών ασθενειών. Στην συγκεκριμένη διπλωματική θα ασχοληθούμε με την ασθένεια του Alzheimer.

Πρόκειται για τον κυριότερο εκπρόσωπο μια ομάδας νευροεκφυλιστικών νοσημάτων που προκαλεί άνοια. Άνοια είναι η κατάσταση που χαρακτηρίζεται από εξασθένηση των νοητικών λειτουργιών, όπως διαταραχή της μνήμης, διαταραχή του λόγου και διαταραχή της αντίληψης του χώρου. Όντας και εγώ λοιπόν ένας άνθρωπος που πάσχει από κάποιο αυτοάνοσο νόσημα ήταν μονόδρομος για εμένα να χρησιμοποιήσω την επιστήμη της τεχνολογίας προς όφελος της ιατρικής στην διπλωματική μου.

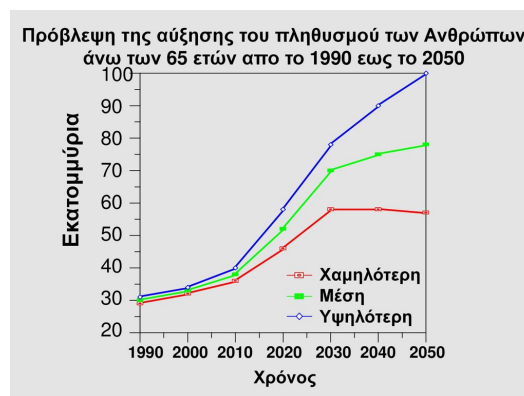
Γίνεται, λοιπόν, μια προσπάθεια διάγνωσης της νόσου του Alzheimer μέσω μηχανικής μάθησης και νευρωνικών δικτύων. Πιο συγκεκριμένα ερευνούμε τρόπους και μεθόδους για την βελτίωση της απόδοσης ενός συνελικτικού νευρωνικού δικτύου. Αρχικά, πραγματοποιείται ανάλυση των δεδομένων, τα οποία πρόκειται για μαγνητικές εγκεφάλου smri , τα οποία υφίστανται κάποια προεπεξεργασία. Οι smri εικόνες μας δίνουν πληροφορίες σχετικά με την δομή του εγκεφάλου, καθώς όπως φαίνεται οι ασθενείς που πάσχουν από την νόσο τείνουν να έχουν κάποιες διαφοροποιήσεις σε σχέση με τους υπόλοιπους ανθρώπους. Έτσι με βάση τις διαστάσεις των εικόνων δημιουργήθηκε ένα συνελικτικό νευρωνικό δίκτυο, το οποίο αποσκοπεί στην ταξινόμηση (classification) των μαγνητικών εγκεφάλου στις τρεις κατηγορίες ανθρώπων, δηλαδή τους ανθρώπους που πάσχουν από Αλτςχάϊμερ (AD), αυτούς με ήπια νοητική βλάβη (MCI) και τους υγιείς(CN).(AD vs MCI vs CN)

## 1.1 Κίνητρο

Η ασθένεια του Αλτςχάϊμερ είναι μία από τις μεγαλύτερες προκλήσεις του 21ου αιώνα, καθώς είναι η κύρια αιτία πρόκλησης της άνοιας. Στο σύνολο 40 εκατομμύρια άνθρωποι ανά τον κόσμο εκτιμάται πως υποφέρουν από άνοια και αυτός ο αριθμός πρόκειται να αυξηθεί στο διπλάσιο για κάθε 20 χρόνια, μέχρις ότου το 2050, λόγω και της αύξησης του μέσου όρου θνησιμότητας (Βλ. Σχήμα 1.1). [1] Παρόλο που δεν υπάρχει κάποια θεραπεία ή κάποια αγωγή που να καθυστερεί σημαντικά την εξέλιξη της νόσου για την συγκεκριμένη ασθένεια, υπάρχουν θεραπείες που βοηθούν στην αντιμετώπιση ορισμένων παρενεργειών της. [2]

Η πρόωρη διάγνωση του Αλτςχάϊμερ είναι σημαντική, από την στιγμή που επιτρέπει στους ανθρώπους με άνοια και στις οικογένειές τους να είναι καλύτερα προετοιμασμένοι για την εξέλιξη της ασθένειας. Επιπλέον όμως δίνει την ευκαιρία να χρησιμοποιηθούν οι διαθέσιμες θεραπείες οι οποίες θα υπονομεύσουν κάποιες από τις παρενέργειες της άνοιας και θα αυξήσουν την ποιότητα ζωής μέσω της καλύτερης διαχείρισης της ασθένειας. Σε μακροχρόνια βάση η πρόωρη διάγνωση της ασθένειας είναι κρίσιμη, αφού μπορεί να γίνει πιο εύκολα κατανοητό στου γιατρούς πώς λειτουργεί, προκειμένου να οδηγηθούμε σε κάποια μελλοντική θεραπεία στα πρώιμα στάδια της.

Παραδοσιακές διαγνώσεις της άνοιας γίνονται προς τα τελευταία στάδια, και προκειμένου να δοθεί σχετικά λογική διάγνωση θα πρέπει να γίνει λεπτομερής εξέταση του ιατρικού ιστορικού, καθώς και της σωματικής και πνευματικής κατάστασης του ασθενούς. Διαφορετικοί βιοδείκτες εξετάζονται, ώστε να δοθεί κάποια πιο αξιόπιστη πρόωρη διάγνωση. Η χρήση μαγνητικών τομογραφιών (MRI) του εγκεφάλου σε συνδυασμό με την μηχανική μάθηση δείχνουν πολλά υποσχόμενα αποτελέσματα ώστε να διακριθούν διάφορα είδη άνοιας συγκριτικά με ανθρώπους που δεν πάσχουν από Αλτςχάϊμερ. [3] [4] [5]



Σχήμα 1.1: Πρόβλεψη αύξησης πληθυσμού

## 1.2 Νόσος Αλτςχάϊμερ

Πρόκειται για μια νευροεκφυλιστική νόσο, η οποία χειροτερεύει στην περίοδο του χρόνου. Αυτή η επιδείνωση μπορεί να είναι είτε αργή είτε γρήγορη. [6] Η νόσος αυτή είναι υπεύθυνη για διαταραχές της μνήμης, του λόγου και της αντίληψης του χώρου, καθώς και δυσκολία προγραμματισμού πιο σύνθετων λειτουργιών, απραξία και ψυχολογικά συμπτώματα.

Αν και οι ακριβείς μοριακές διαδικασίες δεν είναι ακόμα απόλυτα γνωστές, αυτό που χαρακτηρίζει ιστολογικά τη νόσο Alzheimer είναι η σταδιακή εναπόθεση συσσωματωμάτων πρωτεϊνών, οι οποίες σχηματίζουν δύο ιδιαίτερες δομές υπό το μικροσκόπιο [7]:

- Πλάκες αμυλοειδούς(amyloid plaques): Αποτελούνται από εναποθέσεις ενός πεπτιδίου που ονομάζεται β-αμυλοειδές. Οι πλάκες αυτές βρίσκονται ανάμεσα στα νευρικά κύτταρα και σχετίζονται με τις βιοχημικές διαδικασίες επεξεργασίας της πρόδρομης πρωτεΐνης του αμυλοειδούς(amyloid precursor protein).
- Νευροϊνιδιακοί σωροί(neurofibrillary tangles): Αποτελούνται από την πρωτεΐνη tau (ή αλλιώς πρωτεΐνη "τ") και βρίσκονται εντός των νευρικών κυττάρων. Η πρωτεΐνη αυτή είναι πολύ σημαντική για την λειτουργία των νευρώνων και εμπλέκεται σε πολλές νευροεκφυλιστικές νόσους

Αυτές οι αλλοιώσεις ξεκινούν από συγκεκριμένα σημεία του εγκεφάλου και επεκτείνονται σε παρακείμενες περιοχές, καθώς η νόσος εξελίσσεται. Παράλληλα, παρατηρούνται μεταβολές σε αρκετά άλλα κυτταρικά συστήματα, μείωση του αριθμού των νευρώνων και ατροφία του εγκεφάλου.

Δεν υπάρχει συγκεκριμένη εξέταση μέσω της οποίας γίνεται η διάγνωση του Αλτςχάϊμερ αλλά μόνο προσεγγίσεις.

### 1.2.1 Διάγνωση

Παραδείγματα προσεγγίσεων είναι [8]:

1. Επίβλεψη του ατομικού ιατρικού ιστορικού, του ιστορικού της οικογένειας, καθώς και προτροπές των υπόλοιπων μελών της οικογένειας σχετικών με αλλαγές στην συμπεριφορά και της ορθής σκέψης του ασθενούς.
2. Ολοκλήρωση γνωσιακών και νευρολογικών εξετάσεων τύπου Mini Mental State Examination (MMSE)
3. Υποβολή του ασθενή σε εξετάσεις προκειμένου να αποκλειστούν άλλοι λόγοι πρόκλησης των συμπτωμάτων (όπως κάποιος όγκος)

4. Απεικονιστικές εξετάσεις όπως μαγνητική τομογραφία (MRI), αξονική τομογραφία (CT) και ποζιτρονική τομογραφία (Pet) βοηθούν στην διάγνωση όλων των μη φυσιολογικών εικόνων του εγκεφάλου.

### 1.2.2 Ήπια νοητική βλάβη

Ως ήπια νοητική βλάβη (Mild Cognitive Impairment) ορίζονται οι νοητικές αλλαγές ενός ατόμου, οι οποίες είναι αρκετά εμφανείς, ώστε να γίνουν αντιληπτές από το ίδιο το άτομο ή από κοντινούς συγγενείς και φίλους. Παρόλα αυτά δεν είναι αρκετά σημαντικές, ώστε να ταξινομηθούν ως άνοια και όχι αρκετά ικανές να επηρεάσουν την καθημερινή ζωή ενός ανθρώπου. [9]

Η ήπια νοητική βλάβη (MCI) ίσως είναι πρόδρομος του Αλτςχάιμερ. Μια στατιστική ανάλυση του Mitchell και του Shirifeshki έδειξαν ότι το ποσοστό μετατροπής των ασθενών MCI σε ασθενείς με άνοια που παρακολουθούνταν παραπάνω από 5 χρόνια ήταν στο 38%. [10]

## 1.3 Μαγνητική τομογραφία

Αξίζει να δούμε λίγα πράγματα για τις μαγνητικές τομογραφίες, αφού αποτελούν το σετ δεδομένων μας. Η μαγνητική τομογραφία αποτελεί μία μη επεμβατική τεχνική ιατρικής απεικόνισης που χρησιμοποιείται για την παραγωγή λεπτομερών εικόνων τόσο της ανατομίας όσο και της φυσιολογίας του ανθρώπινου σώματος. [11]

Μια πολύ ακριβής εικόνα μπορεί να δημιουργηθεί από μια προηγμένη συσκευή ε-

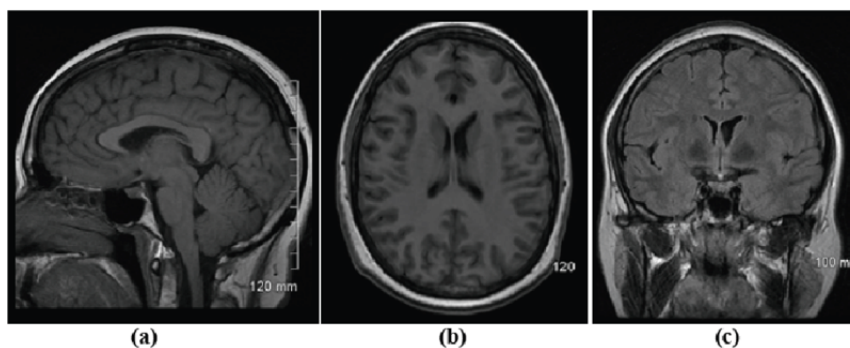


Figure 1.2: Saggital, axial and coronal planes

ξοπλισμένη με έναν πολύ ισχυρό μαγνήτη και μια ποικιλία εξειδικευμένων πηνίων που εκπέμπουν και δέχονται κύματα ραδιοσυχνότητας που αλληλεπιδρούν με άτομα υδρογόνου μέσα στο σώμα μας. Μπορεί να χρησιμοποιηθεί για την διερεύνηση σχεδόν σε

κάθε μέρος του σώματος μας και ειδικά τον εγκέφαλο. [12]

Οι μαγνητικοί τομογράφοι με ισχυρό μαγνητικό πεδίο 1.5 Τέσλα απεικονίζουν με πολύ υψηλή ακρίβεια, περιορίζουν αρκετά τον χρόνο εξέτασης στο  $\frac{1}{8}$  των παλαιότερων μηχανημάτων, μειώνουν σημαντικά τον θόρυβο και προσφέρουν μεγαλύτερη άνεση στον εξεταζόμενο. Με την συνεχή αναβάθμιση των δυνατοτήτων τους μπορούν να προσφέρουν σύγχρονες εξειδικευμένες εξετάσεις, όπως μαστογραφία, προστάτη, τρακτογραφία και έλεγχο αιμάτωσης. Αξίζει να αναφερθεί ότι ο εξεταζόμενος δεν εκτίθεται σε ιοντίζουσα ακτινοβολία.

## 1.4 Στόχοι της διπλωματικής εργασίας

Η διπλωματική εργασία έχει τους εξής στόχους:

1. Να εξεταστεί σε ποιον βαθμό μπορεί να επιτευχθεί η ορθή πρόωρη διάγνωση της ασθένειας του Αλτςχάιμερ μέσω χρήσης συνελικτικών νευρωνικών δικτύων τριών διαστάσεων (3D convolutional neural networks) με είσοδο δεδομένων μαγνητικές τομογραφίες (MRI)
2. Η απλή εξήγηση στον μέσο προπτυχιακό φοιτητή του προβλήματος προς επίλυση. Σκοπός είναι η βήμα βήμα επεξήγηση του προβλήματος και της θεωρίας αυτού, ώστε να μπορεί να τα αντιληφθεί χωρίς προϋπάρχουσα γνώση τι πραγματεύεται η διπλωματική μου.

## 1.5 Σχετική έρευνα

Το 2018 στον διεθνή διαγωνισμό για αυτόματη πρόβλεψη των MCI η ελληνική ομάδα του Δ.Λιπαρά και λοιποί [13] κατέκτησαν την πρώτη θέση στο πρόβλημα ταξινόμησης τεσσάρων κλάσεων του ADNI . Οι 4 διακριτές κλάσεις ήταν οι υγιείς εξεταζόμενοι, οι εξεταζόμενοι που είχαν ήπιες νοητικές βλάβες (MCI) , εκείνοι που είχαν όμοιες βλάβες και έγιναν ασθενείς Αλτςχάιμερ(cMCI) και τέλος εξεταζόμενοι με Αλτςχάιμερ. Κατάφεραν μέσω ενός αλγορίθμου Random Forest να επιτύχουν ακρίβεια 61.9% σε αυτό το πολύ δύσκολο πρόβλημα.

Το 2020 σε αντίστοιχο πρόβλημα τριών κλάσεων (δεν υπήρχαν οι cMCI) η Pereira et al [14] έχτισαν ένα συνελικτικό νευρωνικό δίκτυο δύο διαστάσεων αποσπώντας 64·64 τμήματα των δισδιάστατων μαγνητικών εικόνων. Κατάφεραν να επιτύχουν ακρίβεια της τάξης του 68.6%.

Τέλος το 2015 οι Payan και Montana [15] πρότειναν μια μέθοδο για το πρόβλημα

ταξινόμησης των τριών κλάσεων χρησιμοποιώντας τόσο δισδιάστατα όσο τρισδιάστατα συνελικτικά νευρωνικά δίκτυα μαζί με αραιούς αυτο-κωδικοποιητές (auto-encoders) . Πέτυχαν τα καλύτερα αποτελέσματα χρησιμοποιώντας τρισδιάστατα συνελικτικά νευρωνικά δίκτυα για το πρόβλημα των τριών κλάσεων της τάξης του 89.47%.



## Κεφάλαιο 2

# Μηχανική Μάθηση

### 2.1 Ο ρόλος της μηχανικής μάθησης στην ιατρική

Είναι πλέον γνωστό στο μεγαλύτερο ποσοστό των ανθρώπων η τεράστια δύναμη που έχουν οι υπολογιστές. Αυτή η δύναμη αναφέρεται στην δύναμη επεξεργασίας, στον τεράστιο χώρο της μνήμης καθώς και τον πρωτοφανή πλούτο σε δεδομένα που μπορεί να βρει κανείς στο internet. Όλα αυτά καθιστούν τον υπολογιστή ένα όπλο απέναντι σε πολλούς τομείς της καθημερινότητας μας, όπου ο ανθρώπινος παράγοντας αδυνατεί να βρει λύσεις.

Ένας τέτοιος τομέας είναι και ο τομέας της ιατρικής. Οι αλγόριθμοι μηχανικής μάθησης έχουν την δυνατότητα να ανιχνεύουν μοτίβα σχετικά με διάφορες ασθένειες και καταστάσεις υγείας μέσω της επεξεργασίας χιλιάδων καταγραφών και διαφορετικού είδους δεδομένων (εικόνες, βίντεο, ηχογραφήσεις, κείμενα). Πρόσφατες εξελίξεις πάνω στην βαθειά μάθηση μπορούν να αυξήσουν την ιατροφαρμακευτική περίθαλψη στις αναπτυσσόμενες χώρες για διάφορες ασθένειες ακόμα και του καρκίνου. Το site [entrepreneur](#) ισχυρίζεται πως ένα μοντέλο βαθειάς μάθησης που αναπτύχθηκε στο Ινστιτούτο Τεχνολογίας της Μασαχουσέτης μπορεί να προβλέψει τον καρκίνο στο στήθος των γυναικών χρόνια πριν την εμφάνιση του. Ομοίως ένα [άρθρο](#) της AMA Journal of ethics τρέφει την άποψη πως εφαρμογές της τεχνητής νοημοσύνης μπορούν να διαγνώσουν τον καρκίνο του δέρματος με μεγαλύτερη ακρίβεια από ότι ένας ιατρός.

Όπως βλέπουμε από τα παραπάνω υπάρχει διττός ρόλος στην προσπάθεια συνεργασίας μεταξύ μηχανικής μάθησης και ιατρικής. Από την μία πλευρά υπάρχει η ανάγκη αποθήκευσης της όλης χρησιμοποιούμενης πληροφορίας, η οποία βέβαια είναι κατά κύριο λόγο δουλειά των ιατρών. Από την άλλη πλευρά είναι ευθύνη των data scientists να εκμεταλλευτούν στο έπακρο αυτήν την πληροφορία προκειμένου να χτίσουν μοντέλα που θα αποδώσουν στο βέλτιστο. Η συνεργασία μεταξύ των δύο ειδικοτήτων είναι απαρα-

ίτητη για την ομαλή έκβαση αυτής της εργασίας, σκεπτόμενοι ότι η πληροφορία πρέπει να είναι σε κάποια συγκεκριμένη επεξεργάσιμη μορφή.

## 2.2 Τι είναι μηχανική μάθηση

Η μηχανική μάθηση είναι ένα υποπεδίο της επιστήμης των υπολογιστών, που αναπτύχθηκε από την μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. Η μηχανική μάθηση μπορεί να οριστεί με βάση τον Mehryar Mohri [16] ως οι υπολογιστικές μέθοδοι, οι οποίες χρησιμοποιούν την εμπειρία, προκειμένου να βελτιώσουν την απόδοση ή να πραγματοποιήσουν ακριβείς προβλέψεις. Η εμπειρία αναφέρεται στην προηγούμενη πληροφορία που γίνεται διαθέσιμη στον μαθητευόμενο (learner) , η οποία κατά κανόνα έχει την μορφή ηλεκτρονικών δεδομένων που συλλέχθηκαν και είναι διαθέσιμα για ανάλυση. Σε πιο απλή εξήγηση η μηχανική μάθηση επεξεργάζεται (συνήθως μεγάλο) όγκο δεδομένων για να ανακαλύπτει μοτίβα, τα οποία βοηθούν στην λήψη συγκεκριμένων αποφάσεων. Στο δικό μας παράδειγμα τα δεδομένα είναι οι μαγνητικές τομογραφίες εγκεφάλου και η απόφαση που πρέπει να παρθεί είναι εάν ο ασθενής πάσχει ή όχι από Alzheimer .

Οι αλγόριθμοι μηχανικής μάθησης είναι σχεδιασμένοι έτσι ώστε να μαθαίνουν από τα δεδομένα ανεξάρτητα, χωρίς την παρέμβαση του ανθρώπινου παράγοντα. Αυτή η εκμάθηση γίνεται σταδιακά με την πάροδο του χρόνου και γίνεται σαφής η βελτίωση της ακρίβειας των αποτελεσμάτων χωρίς περαιτέρω προγραμματισμό. Η μηχανική μάθηση αποτελείται από 3 στάδια, την αναπαράσταση, την εκτίμηση και την βελτιστοποίηση του μοντέλου.

Η μηχανική μάθηση εκτός των άλλων περιλαμβάνει αποδοτικούς και ακριβείς αλγόριθμους πρόβλεψης. Εντούτοις εκτός από την χρονική και χωρική πολυπλοκότητα που πρέπει να λάβει υπόψιν ο προγραμματιστής, όπως σε κάθε τομέα της επιστήμης των υπολογιστών, χρειάζεται επίσης αντίληψη για την πολυπλοκότητα δείγματος (sample complexity), ώστε να εκτιμήσει το απαιτούμενο δείγμα δεδομένων που χρειάζεται ο αλγόριθμος για να μάθει σωστά από αυτό.

Όπως καταλαβαίνουμε η επιτυχία ενός αλγορίθμου μηχανικής μάθησης εξαρτάται και από τα δεδομένα. Ως εκ τούτου το συγκεκριμένο πεδίο της επιστήμης των υπολογιστών σχετίζεται εγγενώς με την ανάλυση δεδομένων και την στατιστική. Πολλές τεχνικές μάθησης στηρίζονται σε θεμελιώδεις έννοιες σχετικές με την στατιστική, τις πιθανότητες και της βελτιστοποίησης.

### 2.2.1 Προβλήματα που αντιμετωπίζει η μηχανική μάθηση

Η μηχανική μάθηση διαχειρίζεται ένα ευρύ φάσμα πρακτικών εφαρμογών, που περιλαμβάνει τα ακόλουθα:

- Text ή document classification. Πρόκειται για την επεξεργασία κειμένου, όπως για παράδειγμα εάν το περιεχόμενο μιας ιστοσελίδας είναι ακατάλληλο.
- Natural language processing (NLP). Η επεξεργασία φυσικής γλώσσας περιλαμβάνει την κατανόηση φυσικής γλώσσας, δηλαδή την προσπάθεια να καταστούν ικανοί οι υπολογιστές να εξάγουν νοήματα από ανθρώπινα ή γλωσσικά δεδομένα, αλλά και την παραγωγή φυσικής γλώσσας.
- Speech processing applications, όπως η αναγνώριση λόγου και η επαλήθευση ομιλητή.
- Computer vision applications. Πρόκειται για την αναγνώριση αντικειμένων και προσώπων.
- Πολλά ακόμα προβλήματα, όπως είναι η αναγνώριση της απάτης (fraud detection), εισβολή του διαδικτύου (network intrusion), εκμάθηση παιχνιδιών και η ιατρική διάγνωση (medical diagnosis).

Πρέπει να ειπωθεί πως η παραπάνω λίστα απέχει πολύ από το να είναι περιεκτική. Καθημερινά υπάρχουν ολοένα και περισσότερα προβλήματα στα οποία συνεισφέρει η μηχανική μάθηση και τα παραπάνω δίνουν απλώς μια εικόνα στον αναγνώστη.

### 2.2.2 Τύποι μάθησης

Οι εργασίες μηχανικής μάθησης συνήθως ταξινομούνται σε τρεις μεγάλες κατηγορίες, ανάλογα με την φύση του εκπαιδευτικού "σήματος" ή την "ανατροφοδότηση" που είναι διαθέσιμα σε ένα σύστημα εκμάθησης. Αυτές είναι:

- **Επιβλεπόμενη Μάθηση** (αλλιώς επιτηρούμενη μάθηση ή μάθηση με επίβλεψη) (Supervised learning): Είναι η διαδικασία όπου ο αλγόριθμος κατασκευάζει μια συνάρτηση που απεικονίζει δεδομένες εισόδους (σύνολο εκπαίδευσης) σε γνωστές επιθυμητές εξόδους, με απώτερο στόχο τη γενίκευση της συνάρτησης αυτής και για εισόδους με άγνωστη έξοδο. Χρησιμοποιείται σε προβλήματα ταξινόμησης (Classification), πρόγνωσης (Prediction), διερμηνείας (Interpretation).
- **Μη Επιβλεπόμενη Μάθηση** (αλλιώς μη επιτηρούμενη μάθηση ή μάθηση χωρίς επίβλεψη) (Unsupervised Learning): Εδώ ο αλγόριθμος κατασκευάζει ένα μοντέλο για κάποιο σύνολο εισόδων υπό μορφή παρατηρήσεων χωρίς να γνωρίζει

τις επιθυμητές εξόδους. Χρησιμοποιείται για προβλήματα Ανάλυσης συσχετισμών (*Association Analysis*) και Ομαδοποίησης (*Clustering*).

- **Ενισχυτική Μάθηση** (Reinforcement Learning), όπου ο αλγόριθμος μαθαίνει μια στρατηγική ενεργειών μέσα από άμεση αλληλεπίδραση με το περιβάλλον. Χρησιμοποιείται κυρίως σε προβλήματα Σχεδιασμού (Planning). Σχετικά παραδείγματα είναι ο έλεγχος κίνησης ρομπότ, η βελτιστοποίηση εργασιών σε εργοστασιακούς χώρους, η οδήγηση ενός οχήματος και το να μάθει να παίζει ένα παιχνίδι εναντίον κάποιου αντιπάλου.

Αξίζει να γίνει επίσης αναφορά στην ημι-επιτηρούμενη μάθηση όπου ο δάσκαλος δίνει ένα ελλιπές εκπαιδευτικό σήμα (ένα σύνολο εκπαίδευσης με κάποια από τα αποτελέσματα στόχους να λείπουν). Επίσης η αναπτυξιακή μάθηση είναι ένας σημαντικός τύπος μάθησης, η οποία έχει αναπτυχθεί για την εκμάθηση ρομπότ. Δημιουργεί την δική της ακολουθία μαθησιακών καταστάσεων, ώστε το ρομπότ συσσωρευτικά αποκτά ποικιλία δεξιοτήτων μέσω της αυτόνομης αυτοεξερεύνησης και της κοινωνικής αλληλεπίδρασης με ανθρώπους εκπαιδευτές και χρησιμοποιώντας μηχανισμούς καθοδήγησης, όπως η ενεργητική μάθηση, η ωρίμανση και η μίμηση.

### 2.2.3 Πώς λειτουργεί η επιβλεπόμενη μάθηση

Στην επιβλεπόμενη μάθηση όλα ξεκινούν με την εκπαίδευση ενός μοντέλου, το οποίο πρόκειται για μια μαθηματική συνάρτηση ικανή να τροποποιεί συνεχώς τον τρόπο με τον οποίο λειτουργεί, προκειμένου να κάνει ακριβείς προβλέψεις μέσα από την χρήση δεδομένων.

Πριν βέβαια ξεκινήσει η εκπαίδευση, πρώτα από όλα πρέπει να γίνει η επιλογή για το ποια δεδομένα πρέπει να συλλέξουμε, καθώς και να αποφασίσουμε ποια χαρακτηριστικά των δεδομένων κυρίως μας ενδιαφέρουν. Ειδικά στα προβλήματα ταξινόμησης (classification problems) όπως και το δικό μας κάθε ομάδα χαρακτηριστικών έχει και μια ετικέτα (label) που τα συνοδεύει και οι οποίες πρέπει βεβαίως να αποθηκευτούν.

Μετά την συλλογή των απαραίτητων δεδομένων γίνεται ένας διαχωρισμός των δεδομένων σε σύνολο εκπαίδευσης (train set) και σε σύνολο δοκιμής (test set) . Στην φάση της εκπαίδευσης του μοντέλου, το σύστημα τροφοδοτείται με τις ετικέτες που προαναφέραμε, οι οποίες συμβουλεύουν το σύστημα ποια έξοδος ταιριάζει σε ποιά είσοδο. Κατόπιν στο εκπαιδευμένο πλέον μοντέλο τροφοδοτείται το σύνολο δοκιμής, του οποίου οι ετικέτες είναι άγνωστες για τον αλγόριθμο. Έτσι το σύστημα κάνει προβλέψεις για

την έξοδο με βάση τα χαρακτηριστικά εισόδου του συνόλου δοκιμής και εξετάζει κατά πόσο το μοντέλο προέβλεψε σωστά ή όχι την έξοδο μέσω των ετικετών των δεδομένων. Το σύνολο δοκιμής πρέπει να εξαιρείται από όλη την διαδικασία εκπαίδευσης του μοντέλου ώστε το μοντέλο μας να είναι αμερόληπτο όταν πρόκειται να αξιολογήσει την έξοδο.

Τι ακριβώς όμως συμβαίνει στην φάση της εκπαίδευσης; Κάθε νευρωνικό δίκτυο λαμβάνει ως είσοδο τα χαρακτηριστικά των δεδομένων του συνόλου εκπαίδευσης αποκλειστικά. Στην βαθιή μάθηση τα χαρακτηριστικά αυτά είναι αριθμοί (ούτως ή άλλως οι υπολογιστές μονάχα αριθμούς μπορούν να καταλάβουν). Σε κάθε έναν αριθμό δίνεται ξεχωριστή βαρύτητα ανάλογα με την συνάρτηση ενεργοποίησης και μάλιστα σε κάθε ένα στρώμα του δικτύου. Με βάση την συνάρτηση απωλειών (\*loss function) το μοντέλο τροποποιεί τα βάρη που δίνεται για κάθε ένα χαρακτηριστικό ώστε να ταιριάζει με την αναμενόμενη έξοδο. Αυτό συμβαίνει για κάθε ένα παράδειγμα του συνόλου εκπαίδευσης.

Τα παραπάνω αποτελούν προϋποθέσεις απαραίτητες για να λειτουργήσει ένα δίκτυο μηχανικής μάθησης. Παρόλα αυτά υπάρχουν κάποια πράγματα που αξίζει να αναφερθούν. Εκτός από τον διαχωρισμό των δεδομένων σε σύνολο εκπαίδευσης και σύνολο αξιολόγησης, σε μεγάλα σετ δεδομένων, αναγκαίος είναι και ο διαχωρισμός των δεδομένων σε ένα επιπλέον σύνολο. Το σύνολο επικύρωσης (validation set) είναι υπεύθυνο να υποδεικνύει τις κατάλληλες υπερπαραμέτρους, προκειμένου να αποδώσει κάποιο νευρωνικό δίκτυο στο μέγιστο (για τις συγκεκριμένες τιμές εισόδου).

Επιπρόσθετα μεταξύ του διαχωρισμού των συνόλων σε σύνολα εκπαίδευσης, αξιολόγησης και επικύρωσης στις περισσότερες περιπτώσεις για την καλύτερη απόδοση του νευρωνικού δικτύου οφείλει ο προγραμματιστής να κάνει προεπεξεργασία δεδομένων (preprocessing). Τέτοιου είδους προεπεξεργασία είναι για παράδειγμα η αφαίρεση δύο ή παραπάνω ταυτόσημων στοιχείων, καθώς κάτι τέτοιο μπορεί να οδηγούσε στην υπερπροσαρμογή (overfitting) ή την αφαίρεση κάποιων εικονοστοιχείων (pixels) τα οποία προκαλούν θόρυβο και δυσκολεύουν το δίκτυο μας.

#### 2.2.4 Διαφορά προβλημάτων ταξινόμησης και παλινδρόμησης

Και οι δύο αυτές κατηγορίες ανήκουν στα προβλήματα επιβλεπόμενης μάθησης. Όπως είδαμε και στο κεφάλαιο 1.2.2 ο στόχος ενός προβλήματος μάθησης είναι η εκμάθηση μιας απεικόνισης  $f$  από τα στοιχεία εισόδου του  $x$  χώρου δειγμάτων, στα στοιχεία εξόδου  $y$  του χώρου ετικετών. Το κριτήριο με το οποίο κατηγοριοποιούνται τα προβλήματα επιβλεπόμενης μάθησης στις υποκατηγορίες τους είναι το είδος των τιμών που μπορεί να

λάβει το σύνολο του χώρου ετικετών  $Y$ . Έτσι, αν τα στοιχεία του συνόλου ετικετών παίρνουν διακριτές τιμές, δηλαδή  $y \in \{1, \dots, C\}$  τότε μιλάμε για ένα πρόβλημα ταξινόμησης των δειγμάτων  $x$  σε  $C$  κλάσεις. Ωστόσο, αν τα στοιχεία του συνόλου εξόδου δύνανται να λάβουν τιμές από ένα συνεχές σύνολο για παράδειγμα  $y = \mathbb{R}$  τότε επιλύουμε ένα πρόβλημα παλινδρόμησης.

Ο διαχωρισμός στις δύο αυτές κατηγορίες μπορεί εύκολα να εξηγηθεί μέσω ενός παραδείγματος βαθμολόγησης ενός φοιτητή, όπως αυτό παρουσιάζεται στο Σχήμα 2.1. Στην περίπτωση του προβλήματος παλινδρόμησης καλούμαστε να προβλέψουμε την ακριβή βαθμολογία που μπορεί να λάβει ο φοιτητής στο μάθημα σε ένα συνεχές διάστημα τιμών από 0-10. Από την άλλη, στην περίπτωση του προβλήματος ταξινόμησης, έχουμε χωρίσει το συνεχές διάστημα πιθανών τιμών της βαθμολογίας σε δύο κλάσεις, δύο σύνολα τιμών δηλαδή με ετικέτες FAIL, PASS και καλούμαστε να προβλέψουμε αν ο φοιτητής με βάση κάποια χαρακτηριστικά του θα περάσει ή όχι το μάθημα. Άλλο ένα παράδειγμα είναι η πρόβλεψη της θερμοκρασίας. Στην περίπτωση της ταξινόμησης θα προβλέπαμε εάν πρόκειται να έχει κρύο ή ζέστη, ενώ στην περίπτωση της παλινδρόμησης θα προσεγγίζαμε μια πιο ακριβή τιμή της θερμοκρασίας σε ένα συνεχές διάστημα.

### Το πρόβλημα της ταξινόμησης

Έστω ότι έχουμε ένα πρόβλημα ταξινόμησης, όπως αυτό που περιγράφηκε παραπάνω, δηλαδή που έχει στόχο την εκμάθηση μιας απεικόνισης  $f$  από τον χώρο δειγμάτων εισόδου  $x$  στον χώρο ετικετών  $y$  όπου  $y \in \{1, \dots, C\}$  και σαφώς  $C$  είναι το πλήθος

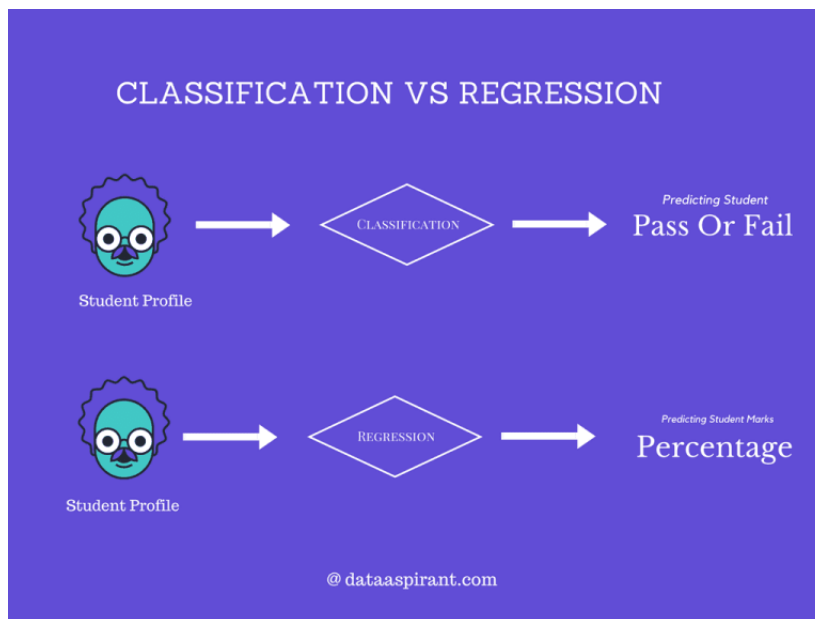


Figure 2.1: Πρόβλεψη της βαθμολογίας με ταξινόμηση και παλινδρόμηση

των κλάσεων του προβλήματος. Τότε ανάλογα με αυτόν τον αριθμό των κλάσεων ενός προβλήματος ταξινομήσης έχουμε τις εξής κατηγορίες:

1. Αν  $C = 2$  τότε έχουμε πρόβλημα δυαδικής ταξινόμησης (Binary Classification). Σε αυτήν την περίπτωση έχουμε συνήθως μια κατάσταση η οποία θεωρείται φυσιολογική και η δεύτερη που θεωρείται αφύσικη. Τα στοιχεία εισόδου πρέπει να αντιστοιχούν σε μια μοναδική έξοδο. Το πρόβλημα μοντελοποιείται με ένα πολύ απλό πείραμα τύχης, την δοκιμή Bernoulli. Πρόκειται για ένα πείραμα τύχης με μόνο δύο, αμοιβαίως αποκλειόμενα, δυνατά αποτελέσματα. Έτσι το μοντέλο προβλέπει μια πιθανότητα ενός παραδείγματος να ανήκει στην κλάση ένα ή μηδέν. Δημοφιλείς αλγόριθμοι που μπορούν να χρησιμοποιηθούν στην δυαδική ταξινόμηση είναι οι Logistic Regression, k-Nearest Neighbors, Decision Trees, SVM (Support Vector Machine), και Naive Bayes.
2. Αν  $C > 2$  και οι κλάσεις είναι αμοιβαία αποκλειόμενες, τότε μιλάμε για πολλαπλών κλάσεων ταξινόμηση (Multi-Class Classification). Εδώ τα παραδείγματα μας ταξινομούνται με βάση εάν ανήκουν σε μία και μόνο μία κλάση από ένα εύρος γνωστών κλάσεων. Ο αριθμός των κλάσεων μπορεί να είναι πολύ μεγάλος. Για παράδειγμα ένα μοντέλο μπορεί να θέλει να προβλέψει ένα πρόσωπο ανάμεσα σε χιλιάδες άλλα πρόσωπα. Το πρόβλημα μοντελοποιείται με την διακριτή πολυωνυμική κατανομή (Multinoulli Distribution), η οποία αποτελεί την γενίκευση της δοκιμής Bernoulli για παραπάνω από 2 κλάσεις. Δημοφιλείς αλγόριθμοι που μπορούν να χρησιμοποιηθούν στην πολλαπλών κλάσεων ταξινόμηση είναι οι k-Nearest Neighbors, Decision Trees, Naive-Bayes, Random Forest και Gradient Boosting.
3. Αν  $C > 2$  και οι κλάσεις δεν είναι αμοιβαία αποκλειόμενες, τότε μιλάμε για πολλαπλών ετικετών ταξινόμηση (Multi-Class Classification). Εδώ τα παραδείγματα μας ταξινομούνται με βάση εάν ανήκουν ενδεχομένως σε παραπάνω από μία κλάση από ένα εύρος γνωστών κλάσεων. Για παράδειγμα ένα μοντέλο μπορεί να θέλει να προβλέψει εάν σε μια φωτογραφία υπάρχει ένα ή παραπάνω από ένα διακριτά αντικείμενα. Το πρόβλημα μοντελοποιείται με ένα μοντέλο που προβλέπει πολλαπλές εξόδους. Ουσιαστικά το μοντέλο αυτό κάνει πολλαπλές δυαδικές προβλέψεις ταξινόμησης για κάθε παράδειγμα. Δημοφιλείς αλγόριθμοι που μπορούν να χρησιμοποιηθούν στην πολλαπλών κλάσεων ταξινόμηση είναι οι Multi-label Decision Trees, Multi-label Random Forests και Multi-label Gradient Boosting.

## 2.3 Στάδια Μάθησης ( Learning Stages)

Σε αυτήν την ενότητα θα κάνουμε χρήση του δυαδικού προβλήματος ανίχνευσης ενοχλητικής αλληλογραφίας (spam detection) , προκειμένου να διευκρινίσουμε κάποιους βασικούς ορισμούς και να περιγράψουμε την χρησιμότητα και την εκτίμηση των αλγορίθμων μηχανικής μάθησης στην πράξη, συμπεριλαμβάνοντας τα διαφορετικά του στάδια.

Η ανίχνευση της ενοχλητικής αλληλογραφίας είναι το πρόβλημα κατά το οποίο θέλουμε να ταξινομήσουμε ποια μηνύματα ηλεκτρονικής αλληλογραφίας (e-mail) είναι ενοχλητικά ή όχι (SPAM or non-SPAM) . Οι παρακάτω ορισμοί σηνύτως χρησιμοποιούνται στην μηχανική μάθηση:

- *Παραδείγματα (Examples)*: Αντικείμενα ή παραδείγματα των δεδομένων για εκμάθηση (learning) ή εκτίμηση. Στο πρόβλημα της ενοχλητικής αλληλογραφίας, τα παραδείγματα αυτά αντιστοιχούν στην συλλογή των e-mail που θα χρησιμοποιήσουμε για εκμάθηση και έλεγχο.
- *Χαρακτηριστικά (Features)*: Το σύνολο των γνωρισμάτων, που συχνά εκπροσωπείται ως διάνυσμα και είναι συνδεδεμένο με ένα παράδειγμα. Στο παράδειγμα μας τέτοια χαρακτηριστικά είναι το μήκος μηνύματος, το όνομα του αποστολέα, τα χαρακτηριστικά της επικεφαλίδας, ορισμένες λέξεις κλειδιά στο κείμενο κ.ο.κ.
- *Ετικέτες (Labels)*: Τιμές ή κατηγορίες αναθετημένες στα παραδείγματα. Στην συγκεκριμένη περίπτωση οι ετικέτες-κατηγορίες είναι SPAM, no-SPAM . Στα προβλήματα παλινδρόμησης όπως είδαμε στο κεφάλαιο 2.2.4 οι ετικέτες αποτελούν πραγματικές τιμές.
- *Υπερπράμετροι (Hyperparameters)*: Ελεύθερες παράμετροι που δεν καθορίζονται κατά την διάρκεια της εκπαίδευσης, αλλά δίνονται ως είσοδοι στον αλγόριθμο εκμάθησης
- *Δείγμα Εκπαίδευσης (Training sample)*: Δεδομένα που χρησιμοποιούνται για την εκπαίδευση του αλγορίθμου μάθησης. Στο πρόβλημα ενοχλητικής αλληλογραφίας, το δείγμα εκπαίδευσης περιλαμβάνει τα δεδομένα με τα χαρακτηριστικά τους και τις αντίστοιχες ετικέτες.
- *Δείγμα Επικύρωσης (Validation sample)*: Παραδείγματα δεδομένων που χρησιμοποιούνται για τον συντονισμό των παραμέτρων του αλγορίθμου εκμάθησης
- *Δείγμα Ελέγχου (Test sample)*: Δεδομένα που χρησιμοποιούνται για την εκτίμηση του μοντέλου εκμάθησης. Το συγκεκριμένο σετ δεδομένων πρέπει να είναι διαφορετικό από εκείνο της επικύρωσης και της εκπαίδευσης και να μην είναι διαθέσιμο



κατά την φάση της εκπαίδευσης. Στο πρόβλημα μας το Δείγμα Ελέγχου αποτελείται από μια συλλογή από e-mail για τα οποία ο αλγόριθμος εκμάθησης πρέπει να προβλέψει σωστά τις ετικέτες του μέσα από τα χαρακτηριστικά του. Οι προβλέψεις που γίνονται για τις ετικέτες αυτού του δείγματος ύστερα συγκρίνονται με τις πραγματικές ετικέτες του για να μετρηθεί η απόδοση του μοντέλου.

- *Συνάρτηση απώλειας (Loss Function)*: Η συνάρτηση μέτρησης της διαφοράς μεταξύ των προβλεπόμενων ετικετών και των πραγματικών ετικετών. Έστω ότι έχουμε το σετ δεδομένων για όλες τις ετικέτες ως  $Y$  και των προβλεπόμενων ετικετών ως  $Y'$ . Τότε ορίζουμε ως συνάρτηση απώλειας  $L$  την απεικόνιση  $L : Y \times Y' \rightarrow \mathbb{R}$ . Στις περισσότερες περιπτώσεις ισχύει ότι  $Y = Y'$  και η συνάρτηση απώλειας  $L$  είναι οριοθετημένη, αλλά αυτό δεν ισχύει πάντοτε. Συνηθισμένα παραδείγματα συναρτήσεων απώλειας είναι η zero-one που ορίζεται ως  $\{-1, +1\} \times \{-1, +1\}$  από την  $L(y, y') = 1_{y \neq y'}$ .



## Κεφάλαιο 3

# Αλγόριθμοι Μηχανικής Μάθησης

### 3.1 Ο αλγόριθμος Naive-Bayes

Ο ταξινομητής Naive-Bayes απλουστεύει την εκμάθηση μέσω της υπόθεσης ότι τα χαρακτηριστικά εισόδου είναι ανεξάρτητα της δοσμένης κλάσης. Αν και η συγκεκριμένη υπόθεση φαίνεται παράδοξη πολλές φορές ο συγκεκριμένος αλγόριθμος δίνει πολύ καλά αποτελέσματα και συγκρίνεται με πιο βαθυστόχαστους αλγορίθμους. [17]

Ο ταξινομητής Bayes διαλέγει την πιο πιθανή κλάση με βάση ένα συγκεκριμένο παράδειγμα, το οποίο περιγράφεται από το διάνυσμα των χαρακτηριστικών του (έννοιες που περιγράφηκαν στο κεφάλαιο 2.3). Με βάση όσα αναφέρθηκαν ο Naive-Bayes υποθέτει ότι τα χαρακτηριστικά είναι ανεξάρτητα της δοσμένης κλάσης, δηλαδή ότι  $P(X|C) = \prod_{i=1}^n P(X_i|C)$ , όπου  $X = (X_1, \dots, X_n)$  είναι το διάνυσμα χαρακτηριστικών και  $C$  είναι μια κλάση. Ο αλγόριθμος *Naive – Bayes* έχει αποδειχθεί ιδιαίτερα αποτελεσματικός σε πολλές πρακτικές εφαρμογές, όπως η ταξινόμηση κειμένου (text classification), η ιατρική διάγνωση (medical diagnosis) και η διαχείριση της απόδοσης συστημάτων (systems performance management).

Στο γενικότερο πρόβλημα της αναγνώρισης προτύπων, όπως αναφέραμε και προηγουμένως, καλούμαστε να επιλέξουμε μία κλάση  $c_j$  στην οποία θεωρούμε ότι ανήκει ένα πρότυπο με βάση το διάνυσμα χαρακτηριστικών του, έστω  $x$ . Η επιλογή μας γίνεται μέσα από  $N$  πιθανές κλάσεις  $c_1, c_2, \dots, c_N$ . Αν ορίσουμε τη δεσμευμένη πιθανότητα:  $P(c_j|x), j = 1, \dots, N$ , ως την πιθανότητα το πρότυπο  $x$  να ανήκει στην κλάση  $c_j$ , γνωστή και ως εκ των υστέρων πιθανότητα (a posteriori probability), τότε η διαίσθησή μας λέει να επιλέξουμε για το  $x$ , την κλάση που μεγιστοποιεί την παραπάνω a posteriori πιθανότητα, έστω την κλάση  $k$ . Δηλαδή θεωρούμε τον ακόλουθο κανόνα απόφασης: [18]

Το πρότυπο  $x$  αντιστοιχίζεται στην κλάση  $c_k$ , όπου:

$$k = \arg \max_j P(c_j|x), j = 1, \dots, N$$

Αυτός ακριβώς είναι ο κανόνας απόφασης στον ταξινομητή *Naive – Bayes*, και για αυτό ονομάζεται και Maximum A Posteriori (MAP) ταξινομητής.

Η πιθανότητα  $P(c_j|x)$  εφαρμόζοντας το θεώρημα του *Bayes* υπολογίζεται ως εξής:

$$P(c_j|x) = \frac{P(c_j, x)}{P(x)} = \frac{P(x|c_j)P(c_j)}{P(x)}$$

Όπου:

- $P(c_j)$  είναι η πρότερη πιθανότητα (prior probability) της κλάσης  $j$
- $P(x|c_j)$  είναι η πιθανότητα του χαρακτηριστικού  $x$  δεδομένης της κλάσης  $c_j$  (class conditional probability density function)

Παρατηρούμε πώς η πιθανότητα  $P(x)$  δεν χρειάζεται να υπολογιστεί διότι στην εφαρμογή του κανόνα *Naive – Bayes* εμφανίζεται ως σταθερή ποσότητα, ανεξάρτητη του  $j$  και δεν επηρεάζει την μεγιστοποίηση.

Στο σημείο αυτό έρχεται να εφαρμοστεί και η υπόθεση της ανεξαρτησίας μεταξύ των χαρακτηριστικών δεδομένης της κλάσης  $c_j$ , οπότε η πιθανότητα  $P(x|c_j)$  υπολογίζεται ως το γινόμενο των επιμέρους  $P(x_i|c_j)$ . Δηλαδή:

$$P(x|c_j) = \prod_{i=1}^n P(x_i|c_j)$$

Για την απόφαση του ταξινομητή λοιπόν, αρκεί ο υπολογισμός των πιθανοτήτων  $P(c_j)$  και  $P(x_i|c_j)$ . Οι πιθανότητες αυτές εκτιμώνται κάνοντας χρήση της εκτίμησης μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation – MLE) πάνω στο training set. Σύμφωνα με τη MLE, οι παράμετροι ενός στατιστικού μοντέλου επιλέγονται έτσι ώστε να συμφωνούν με τα δεδομένα που έχουμε στη διάθεσή μας. Έτσι, η πιθανότητα  $P(c_j)$  υπολογίζεται ως το ποσοστό των προτύπων στο training set που ανήκουν στη κλάση  $c_j$  και η πιθανότητα  $P(x|c_j)$  υπολογίζεται από τις επιμέρους πιθανότητες  $P(x_i|c_j)$  οι οποίες εκτιμώνται ίσες με τις αντίστοιχες συχνότητες των χαρακτηριστικών στο ίδιο training set.

Υπάρχουν διάφορες παραλλαγές του αλγορίθμου Naive Bayes. Η διαφορά τους έγκειται μόνο στην υπόθεση που κάνουν σχετικά με την κατανομή  $P(x_i|c_j)$ . Κάποιες εκδόσεις Naive Bayes είναι:

- Gaussian Naive Bayes. Εδώ γίνεται η υπόθεση ότι η κατανομή  $P(x_i|c_j)$  είναι συνεχής και μάλιστα Gaussian. Δηλαδή:

$$P(x_i|c_j) = \frac{e^{-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}}}{\sqrt{2\pi\sigma_j^2}}$$

όπου η μέση τιμή  $\mu_j$  και η τυπική απόκλιση  $\sigma_j$  υπολογίζονται μέσω εκτίμησης μέγιστης πιθανοφάνειας. Ο αλγόριθμος αυτός συνήθως δεν βρίσκει εφαρμογή σε tasks επεξεργασίας φυσικής γλώσσας, όπως αυτό του sentiment analysis και δεν θα μας απασχολήσει στη συνέχεια.

- Multinomial Naive Bayes. Η έκδοση αυτή υλοποιεί τον αλγόριθμο Naive Bayes για πολυωνυμικά κατανεμημένα δεδομένα και είναι μία από τις δύο κλασικές παραλλαγές του αλγορίθμου Naive Bayes που χρησιμοποιούνται στην ταξινόμηση κειμένου. Εδώ τα δεδομένα αναπαριστώνται συνήθως ως μετρήσεις απλών λέξεων ή n-grams ανάλογα με τη θεωρία. Η κατανομή παραμετροποιείται από τα διανύσματα  $\theta_{c_j} = (\theta_{c_j1}, \dots, \theta_{c_jn})^T$ , όπου ο αριθμός  $n$  των χαρακτηριστικών για την ταξινόμηση κειμένου ισούται με το μέγεθος του λεξιλογίου και  $\theta_{c_j1}$  είναι η πιθανότητα  $P(x_i|c_j)$  του χαρακτηριστικού-token  $i$  να εμφανιστεί σε ένα δείγμα της κλάσης  $c_j$ . Τα στοιχεία του διανύσματος  $\theta_{c_j}$  υπολογίζονται μέσω μίας εξομαλυμένης εκδοχής *MLE* ως εξής:

$$\theta_{c_j i} = \frac{N_{c_j i} + \alpha}{N_{c_j} + \alpha n}$$

όπου  $N_{c_j i}$  είναι ο αριθμός των φορές που το χαρακτηριστικό  $i$  εμφανίζεται στα δείγματα της κλάσης  $c_j$  στο training set  $D$  και  $N_{c_j}$  είναι το συνολικό πλήθος των χαρακτηριστικών για τη κλάση  $c_j$ . Η παράμετρος ομαλοποίησης  $\alpha$  εισάγεται για την αντιμετώπιση χαρακτηριστικών που δεν εμφανίζεται καθόλου στο σύνολο εκπαίδευσης και εμποδίζει τη διάδοση μηδενικών πιθανοτήτων στους υπολογισμούς. Η παραπάνω τεχνική αν  $\alpha = 1$  ονομάζεται εξομάλυνση Laplace (Laplace smoothing ή add-one smoothing), αλλιώς αν  $\alpha < 1$  ονομάζεται εξομάλυνση Lidstone (Lidstone smoothing). Αν αντί να μετράμε όλες τις εμφανίσεις μιας λέξης ή ενός n-gram στο κείμενο, τις μετράμε μόνο μία φορά, τότε προκύπτει η δυαδικοποιημένη (binarized) εκδοχή του Multinomial Naive Bayes που ονομάζεται και Boolean Multinomial Naive Bayes.

- Bernoulli Naive Bayes. Πρόκειται για τη δεύτερη κλασική παραλλαγή του αλγορίθμου Naive Bayes που χρησιμοποιείται στην ταξινόμηση κειμένου. Εδώ κάθε όρος του λεξιλογίου ισούται με 1 εάν εμφανίζεται στο κείμενο αλλιώς με 0. Η διαφορά

του από τον Boolean Naive Bayes είναι ότι λαμβάνει υπόψη τους όρους που δεν εμφανίζονται στο κείμενο. Ενώ στο μοντέλο Boolean Multinomial οι όροι που δεν εμφανίζονται αγνοούνται τελείως, στο μοντέλο Bernoulli οι όροι αυτοί παραγοποιούνται όταν υπολογίζονται οι δεσμευμένες πιθανότητες και άρα η απουσία των όρων συνυπολογίζεται. Η πιθανότητα  $P(x_i|c_j)$  υπολογίζεται ως:

$$P(x_i|c_j) = P(i|c_j)x_i + (1 - P(i|c_j))(1 - x_i)$$

Συνήθως ο Multinomial Naive Bayes χρησιμοποιείται όταν οι πολλαπλές εμφανίσεις των λέξεων είναι σημαντικές στο πρόβλημα ταξινόμησης. Ένα τέτοιο παράδειγμα είναι όταν προσπαθούμε να κάνουμε ταξινόμηση με βάση το θέμα (topic classification). Ο δυαδικοποιημένος Multinomial Naive Bayes χρησιμοποιείται όταν οι συχνότητες των λέξεων δεν παίζουν σημαντικό ρόλο στην ταξινόμησή μας. Ένα τέτοιο παράδειγμα είναι η ανάλυση συναισθήματος, όπου δεν ενδιαφέρει τόσο το πόσες φορές αναφέρει κάποιος τη λέξη “bad” αλλά περισσότερο το γεγονός ότι απλά την αναφέρει. Τέλος, ο Bernoulli Naive Bayes μπορεί να χρησιμοποιηθεί όταν στο πρόβλημά μας η απουσία κάποιας συγκεκριμένης λέξης παίζει ρόλο. Για παράδειγμα, ο Bernoulli Naive Bayes χρησιμοποιείται συνήθως στην ανίχνευση spam ή στην ανίχνευση περιεχομένου για ανηλίκους με πολύ καλά αποτελέσματα.

### 3.2 Μηχανές διανυσματικής υποστήριξης SVM

Οι μηχανές διανυσματικής υποστήριξης (support vector machines) αποτελούν εργαλεία μάθησης, τα οποία εισήγαγε στην επιστημονική κοινότητα ο V.Vapnik το 1992 [19]. Η ευρεία χρήση τους στο χώρο της μηχανικής μάθησης οφείλεται σε τρεις βασικούς παράγοντες [20]:

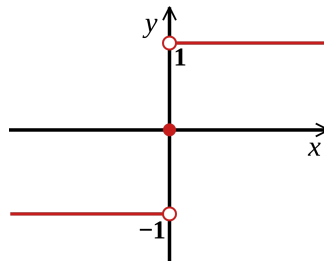
1. Η δυνατότητά τους να επιδεικνύουν ικανοποιητικά αποτελέσματα μαθαίνοντας από διανύσματα εισόδου  $\mathbf{x}$  με μικρό πλήθος χαρακτηριστικών.
2. Στην ανθεκτικότητα που επιδεικνύουν στη πραγματοποίηση λαθών.
3. η μικρή υπολογιστική επιβάρυνση που επιφέρουν συγκριτικά με άλλα εργαλεία όπως τα τεχνητά νευρωνικά δίκτυα.

Ο αλγόριθμος των μηχανών διανυσματικής υποστήριξης στηρίζει την λειτουργία του στην αποτύπωση των χαρακτηριστικών του διανύσματος εισόδου  $\mathbf{x}$  σε ένα χώρο διάστασης ίσης με το πλήθος των χαρακτηριστικών. Λαμβάνοντας υπόψιν την αναπαράσταση αυτή, ο αλγόριθμος προσπαθεί να διαχωρίσει τα διανύσματα χαρακτηριστικών, μέσω ενός υπερεπιπέδου, σε δύο χωρία που αντιστοιχίζονται στις κλάσεις ενδιαφέροντος. Αυτό

το διαχωριστικό υπερεπίπεδο μεταξύ των εμπλεκόμενων κλάσεων ονομάζεται σύνορο απόφασης (decision boundary). [21]

Η εξίσωση του υπερεπιπέδου εκφράζεται μέσω δύο παραμέτρων, ενός διανύσματος πραγματικών τιμών  $w$ , ίσης διάστασης με το διάνυσμα χαρακτηριστικών εισόδου  $x$  και ενός πραγματικού αριθμού  $b$  από τη σχέση  $w x - b = 0$ .

Επιπλέον η εκτιμώμενη ετικέτα  $y$  για ένα διάνυσμα χαρακτηριστικών  $w$  δίνεται από τον τύπο  $y = \text{sign}(wx - b)$  και λαμβάνει τις τιμές  $-1$  και  $+1$  όπως φαίνεται και στο γράφημα της συνάρτησης (Σχήμα 3.1) Σκοπός κάθε αλγορίθμου μάθησης και των μηχανών



Σχήμα 3.1: Η συνάρτηση  $\text{sign}(x)$

διανυσματικής υποστήριξης εν προκειμένω είναι να εκμεταλλευτούν το σετ δεδομένων ώστε να προσδιορίσουν τις βέλτιστες τιμές  $(w', b')$  για τις παραμέτρους  $(w, b)$ . Μόλις ο αλγόριθμος μάθησης εντοπίσει αυτές τις τιμές για τις παραμέτρους, τότε το μοντέλο  $h(x)$  ορίζεται ως :

$$h(x) = \text{sign}(w'x - b')$$

Πώς βρίσκει ο αλγόριθμος τις ζητούμενες τιμές  $(w', b')$ . Ουσιαστικά προσπαθεί να λύσει ένα πρόβλημα βελτιστοποίησης, το οποίο δομείται υπό περιορισμούς. Υπενθυμίζω πως κάθε παράδειγμα εκπαίδευσης δίνεται από το ζεύγος  $(x_i, y_i)$  με το  $x_i$  να είναι το διάνυσμα χαρακτηριστικών για το  $i$  και το  $y_i$  να είναι η ετικέτα που στο εν λόγω παράδειγμα λαμβάνει τιμές  $-1$  και  $+1$ .

- $w x_i - b \geq 1$  εάν  $y_i = +1$
- $w x_i - b \leq -1$  εάν  $y_i = -1$

Επιπλέον επιθυμούμε το υπερεπίπεδο που διαχωρίζει τα θετικά από τα αρνητικά παραδείγματα να είναι αυτό με το μεγαλύτερο περιθώριο (margin) μεταξύ των κλάσεων. Με τον όρο περιθώριο αναφερόμαστε στην απόσταση μεταξύ των δύο κοντινότερων παραδειγμάτων των δύο κλάσεων όπως αυτές διαχωρίζονται από το σύνορο απόστασης. Ένα μεγάλο περιθώριο συνεισφέρει στην καλύτερη δυνατότητα του μοντέλου για γενίκευση της λειτουργίας του, κατηγοριοποιώντας αποτελεσματικά μελλοντικά παραδείγματα εισόδου. Για να επιτευχθεί κάτι τέτοιο χρειάζεται να ελαχιστοποιήσουμε τη λεγόμενη

Ευκλείδεια νόρμα του  $\mathbf{w}$ .

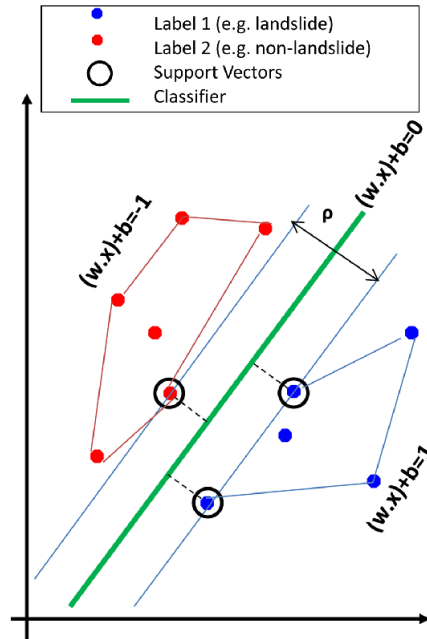
$$\|\mathbf{w}\| = \sqrt{\sum_{j=1}^d (\mathbf{w}^{(j)})^2} \quad (3.1)$$

Τελικά το πρόβλημα βελτιστοποίησης, το οποίο καλείται να επιλύσει το υπολογιστικό σύστημα συνοψίζεται στην εξής πρόταση:

Ελαχιστοποίησε  $\|\mathbf{w}\|$  με την εξάρτηση  $y_i(\mathbf{w}\mathbf{x}_i - b) \geq +1$  για  $i = 1, \dots, m$ .

Η λύση του παραπάνω προβλήματος που όπως είπαμε δίνεται από το ζεύγος  $(\mathbf{w}', \mathbf{b}')$  ονομάζεται στατιστικό μοντέλο (statistical model) ή απλά μοντέλο (model). Η διαδικασία εύρεσης της λύσης και δημιουργίας του μοντέλου ονομάζεται εκπαίδευσης (training).

Για διανύσματα χαρακτηριστικών δύο διαστάσεων, το πρόβλημα και η λύση οπτικοποιούνται στο Σχήμα 3.2. Οι μπλε και πορτοκαλί κύκλοι αναπαριστούν αντίστοιχα τα θετικά και αρνητικά παραδείγματα και η γραμμή που δίνεται από την εξίσωση  $\mathbf{w}\mathbf{x} - b = 0$  αποτελεί το σύνορο απόφασης. Ειδικότερα, μπορούμε και γεωμετρικά να δούμε γιατί ελαχιστοποιώντας την νόρμα του  $\mathbf{w}$  βρίσκουμε το υπερεπίπεδο που μεγιστοποιεί το περιθώριο μεταξύ των δύο κλάσεων. Οι εξισώσεις  $\mathbf{w}\mathbf{x} - b = +1$  και  $\mathbf{w}\mathbf{x} - b = -1$  ορίζουν δύο παράλληλα υπερεπίπεδα με τη μεταξύ τους απόσταση να είναι  $\frac{2}{\|\mathbf{w}\|}$



Σχήμα 3.2: Μοντέλο SVM για διανύσματα χαρακτηριστικών δύο διαστάσεων



### 3.3 kNN

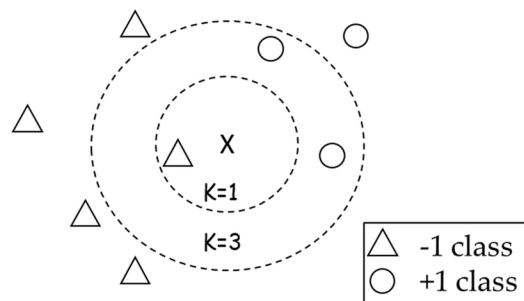
Ο αλγόριθμος του knn είναι μία πολύ γνωστή και ευρεία χρησιμοποιούμενη τεχνική κατηγοριοποίησης που στηρίζεται στη χρήση μέτρων βασισμένων στην απόσταση. Η κεντρική ιδέα είναι πως η τιμή της συνάρτησης-στόχου για ένα νέο στιγμιότυπο βασίζεται αποκλειστικά και μόνο στις αντίστοιχες τιμές των  $k$  πιο “κοντινών” στιγμιότυπων εκπαίδευσης, τα οποία αποτελούν τους “γείτονες” του. Δύο ζητήματα πρέπει να αποφασιστούν για να καθοριστεί πλήρως ο αλγόριθμος:

1. Ο ορισμός της απόστασης μεταξύ δύο στιγμιότυπων, δηλαδή μιας τιμής πάνω στο χώρο των στιγμιότυπων, που θα εκφράζει την εγγύτητα, ή αλλιώς την «ομοιότητα» μεταξύ των στιγμιότυπων.
2. Η τιμή του  $k$

Για το πρώτο ζήτημα, υπάρχουν πολλές εναλλακτικές επιλογές. Η απόφαση εξαρτάται από τα ειδικά χαρακτηριστικά του χώρου στιγμιότυπων του προβλήματος. Ιδιαίτερη σημασία έχει αν στην αναπαράσταση των στιγμιότυπων περιλαμβάνονται αριθμητικά ή συμβολικά χαρακτηριστικά. Στον «παραδοσιακό»  $k$ -NN αλγόριθμο, στον οποίο τα στιγμιότυπα θεωρούνται πως ανήκουν στον  $n$ -διάστατο χώρο  $\mathbb{R}^n$ , μια μέτρηση που υιοθετείται συχνά είναι η γνωστή Ευκλείδεια απόσταση. Πιο συγκεκριμένα, αν τα στιγμιότυπα αναπαρίστανται ως διανύσματα από χαρακτηριστικά που παίρνουν τιμές πραγματικών αριθμούς, δηλαδή το στιγμιότυπο  $x$  αναπαρίσταται από το διάνυσμα:

$$\langle a_1(x), a_2(x), \dots, a_n(x) \rangle$$

όπου  $a_r(x)$  δηλώνει την τιμή του  $r$ -οστού χαρακτηριστικού του  $x$ , τότε η απόσταση  $d(x_i, x_j)$  μεταξύ δύο σημείων  $x_i$  και  $x_j$  ορίζεται ως  $d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$ . Βέβαια μπορούν να επιλεγθούν και άλλα μέτρα ομοιότητας-αντιομοιότητας αντί της Ευκλείδειας. Στην εικόνα φαίνεται η λειτουργία του  $k$ -NN. Τα “+” και τα “-” δείχνουν τα



Σχήμα 3.3: Σχήμα knn για  $k=1$  και  $k=3$

στιγμιότυπα εκπαίδευσης της κάθε κλάσης και το  $x$  ένα στιγμιότυπο προς κατάταξη. Φαίνεται πως ο 1-NN κατατάσσει το  $x$  ως “-” ενώ ο 3-NN το κατατάσσει ως “+”.

Ένα μειονέκτημα που παρουσιάζεται στο προηγούμενο παράδειγμα είναι πως όλα τα χαρακτηριστικά θεωρούνται ισοδύναμα κατά τον υπολογισμό της απόστασης. Αυτό είναι ιδιαίτερα προβληματικό αν δεν είναι όλα τα χαρακτηριστικά σχετικά με τη συγκεκριμένη συνάρτηση-στόχο που επιδιώκεται να προσεγγιστεί, αλλά και γενικότερα, οποτεδήποτε υπάρχουν σημαντικές διαφορές μεταξύ των χαρακτηριστικών ως προς την αξία τους στον προσδιορισμό της συνάρτησης. Σε μία τέτοια περίπτωση, οι παραπάνω μετρήσεις είναι παραπλανητικές, από την άποψη πως μερικά στιγμιότυπα που σχετίζονται πραγματικά μεταξύ τους, είναι δυνατόν να θεωρούνται απομακρυσμένα λόγω των διαφορών τους σε άσχετα ή ασήμαντα χαρακτηριστικά.

Μια λύση σε αυτό το πρόβλημα είναι κάθε χαρακτηριστικό να αποτιμάται διαφορετικά στον υπολογισμό της απόστασης, ανάλογα με την αξία του. Η μέθοδος αυτή λέγεται αποτίμηση των χαρακτηριστικών (feature weighting). Με βάση αυτήν, ο τύπος της Ευκλείδειας θα μπορούσε να γίνει:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n w_r (a_r(x_i) - a_r(x_j))^2}$$

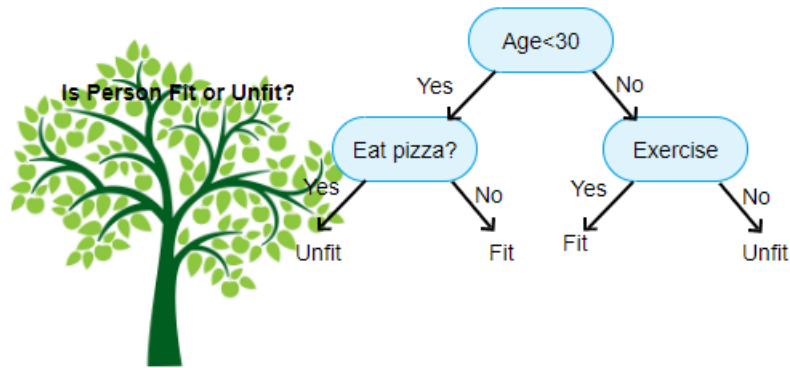
όπου  $w_r$  είναι το βάρος του χαρακτηριστικού  $a_r$ .

### 3.4 Δέντρα αποφάσεων

Τα δέντρα αποφάσεων (desicion trees) αποτελεί έναν αλγόριθμο μηχανικής μάθησης για προβλήματα σχετικά με την επιβλεπόμενη μάθηση (supervised learning) [22]. Ένα δέντρο αποφάσεων θυμίζει πολύ στην δομή, από την επιστήμη των υπολογιστών, ενός κλασσικού διαγράμματος ροής [23] (βλ. Σχήμα 3.4), στο οποίο κάθε κόμβος αποτελεί κάποιου είδους έλεγχο. Στην συνέχεια κάθε κλαδί αποτελεί το αποτέλεσμα ενός ελέγχου και τέλος κάθε φύλλο (ή τερματικός κόμβος) αποτελεί την ετικέτα κλάσης. Η κορυφή του δέντρου αποτελεί την ρίζα του δέντρου (root node). Το δέντρο απόφασης μπορεί να χρησιμοποιηθεί για την οπτική και ανλυτική παρουσίαση αποφάσεων.

Είναι εύκολο να τα κατανοήσει κανείς και μπορεί να χειριστεί τόσο αριθμητικά όσο και κατηγορικά δεδομένα. Αυτός είναι και ο λόγος που καθίστανται κατάλληλα τόσο για προβλήματα ταξινόμησης όσο και προβλήματα παλινδρόμησης. Επίσης μπορεί να χρησιμοποιηθεί για προβλήματα Multi-class Classification αφού μπορεί να υποστηρίξει προβλήματα πολλαπλών εξόδων. Δεν χρειάζεται ιδιαίτερη προσπάθεια για να γίνει η προετοιμασία των δεδομένων από τον προγραμματιστή και μη γραμμικές σχέσεις μεταξύ των παραμέτρων δεν επηρεάζουν την απόδοσή τους.

Πως λειτουργεί όμως ένα δέντρο απόφασης. Ξεκινώντας από τον αρχικό κόμβο το



Σχήμα 3.4: Δέντρο απόφασης

σύστημα υποβάλλει ερωτήσεις σχετικά με τα δεδομένα εισόδου. Με βάση την απάντηση που έχει λάβει συνεχίζει σε ένα από τα δύο πιθανά κλαδιά και υποβάλλει εκ νέου μια ερώτηση, ώστε να πάει στο επόμενο επίπεδο ερωτήσεων. Ο αλγόριθμος τελειώνει όταν μετά από κάποια απάντηση βρεθεί σε τερματικό κόμβο, ο οποίος του υποδηλώνει με βάση τις απαντήσεις που του δόθηκαν νωρίτερα σε ποια ετικέτα (ή εύρος τιμών εάν μιλάμε για regression problem) ανήκουν τα δεδομένα εισόδου. Κάθε διαχωρισμός- απόφαση που παίρνουμε έχει κάποιο κόστος ανάλογα με μια επιλεγμένη συνάρτηση. Σε κάθε διαχωρισμό επιλέγουμε με βάση ποια απόφαση θα έχει το ελάχιστο κόστος στην ακρίβεια.

Και στην περίπτωση των προβλημάτων ταξινόμησης αλλά και των προβλημάτων παλινδρόμησης οι συναρτήσεις κόστους προσπαθούν να βρουν ομοιογενή κλαδιά, ή κλαδιά με σύνολα που έδωσαν παρόμοιες απαντήσεις. Έτσι θα είμαστε σίγουροι πως ένα σύνολο ελέγχου (test set) θα ακολουθήσει συγκεκριμένες διακλαδώσεις. (Κάτι οδηγεί πολλές φορές στην υπερπροσαρμογή overfitting) Ενδεικτικές συναρτήσεις κόστους είναι

$$\text{Regression} : \sum_{i=1}^n (x_i - \text{prediction})^2$$

Έστω ότι θέλουμε να κάνουμε προβλέψεις για την τιμή των ακινήτων σπιτιών. Το δέντρο θα χωριστεί λαμβάνοντας υπόψιν όλα τα χαρακτηριστικά του συνόλου εισόδου. Ο μέσος όρος των απαντήσεων του συνόλου εισόδου ενός συγκεκριμένου γκρουπ θεωρείται η πρόβλεψη του γκρουπ. Η παραπάνω συνάρτηση εφαρμόζεται σε όλα τα δεδομένα και το κόστος υπολογίζεται για κάθε υποψήφιο διαχωρισμό. Και πάλι επιλέγεται το ελάχιστο κόστος.

$$\text{Classification} : \sum_{i=1}^n p_k \cdot (1 - p_k)$$

Η συνάρτηση Gini [24] δίνει μια καλή ιδέα του πόσο καλός είναι ένας διαχωρισμός από τον υπολογισμό της πιθανότητας ενός τυχαίου χαρακτηριστικού να κατηγοριοποιηθεί λανθασμένα όταν αυτό επιλέγεται τυχαία. Το  $pk$  είναι η αναλογία των εισόδων της ίδιας κλάσης να ανήκουν σε ένα συγκεκριμένο γκρουπ.

### 3.5 Τυχαία Δάση - Random Forests

Μία νέα και εξαιρετική μέθοδος ταξινόμησης δεδομένων είναι τα «Τυχαία Δάση» (Random Forests), η οποία είναι αλληλένδετη με την μέθοδο των Δέντρων Απόφασης/ταξινόμησης, που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Τα Random Forests είναι ουσιαστικά μια επιλογή από decision trees. Εμπνευστής της μεθόδου των Random Forests είναι ο Leo Breiman. [25] Τα παρακάτω βήματα, δίνουν μία γενική ιδέα για το πώς λειτουργούν τα random forests :

- Αρχικά, αναπτύσσονται πολλά classification decision trees – Το πώς ακριβώς αναπτύσσονται θα αναφερθεί αναλυτικά στη συνέχεια του κεφαλαίου.
- Κάθε tree δίνει μία ταξινόμηση – «Το δέντρο ψηφίζει αυτήν την κλάση».
- Έτσι, κάθε κλάση έχει έναν αριθμό «ψηφών» (votes).
- Η τελική και οριστική ταξινόμηση γίνεται με το «δάσος» να διαλέγει την κλάση με τις περισσότερες votes. Εύκολα προκύπτει το συμπέρασμα, ότι στα παραπάνω βήματα κύριο ρόλο παίζει το πώς ακριβώς αναπτύσσονται τα decision trees. Κάθε δέντρο, λοιπόν, αναπτύσσεται σύμφωνα με τον παρακάτω αλγόριθμο [26]:

1. Θέσε  $N$  = αριθμός των objects ή cases του training set και  $M$  = αριθμός των μεταβλητών-χαρακτηριστικών (input variables)
2. Για κάθε δέντρο επέλεξε από το data set  $N$  cases στην τύχη με εναπόθεση (with replacement). – bootstrap sampling. Αυτές οι  $N$  cases αποτελούν το training set για κάθε δένδρο

Η μέθοδος του bootstrap sampling αποτελεί μία τεχνική συλλογής δειγμάτων από ένα data set και λειτουργεί ως εξής:

Έστω ότι υπάρχει ένα καλάθι με 5 μπάλες (samples) με ονόματα (labels) :

$$\{A, B, C, D, E\}$$

Από το καλάθι επιλέγεται στην τύχη μία μπάλα και καταγράφεται το όνομα της, έστω η  $B$ . Στη συνέχεια, επανατοποθετείται η μπάλα  $B$  στο καλάθι και ξανά επιλέγεται μία στην τύχη. Η διαδικασία αυτή επαναλαμβάνεται όσες φορές είναι επιθυμητό.

Στο τέλος η καταγραφή των δειγμάτων που επιλέχθηκε μπορεί να μοιάζει σαν και αυτή:

$$[B, E, D, B, C, C, A, D, E, B, A, A, E, C, E, D]$$

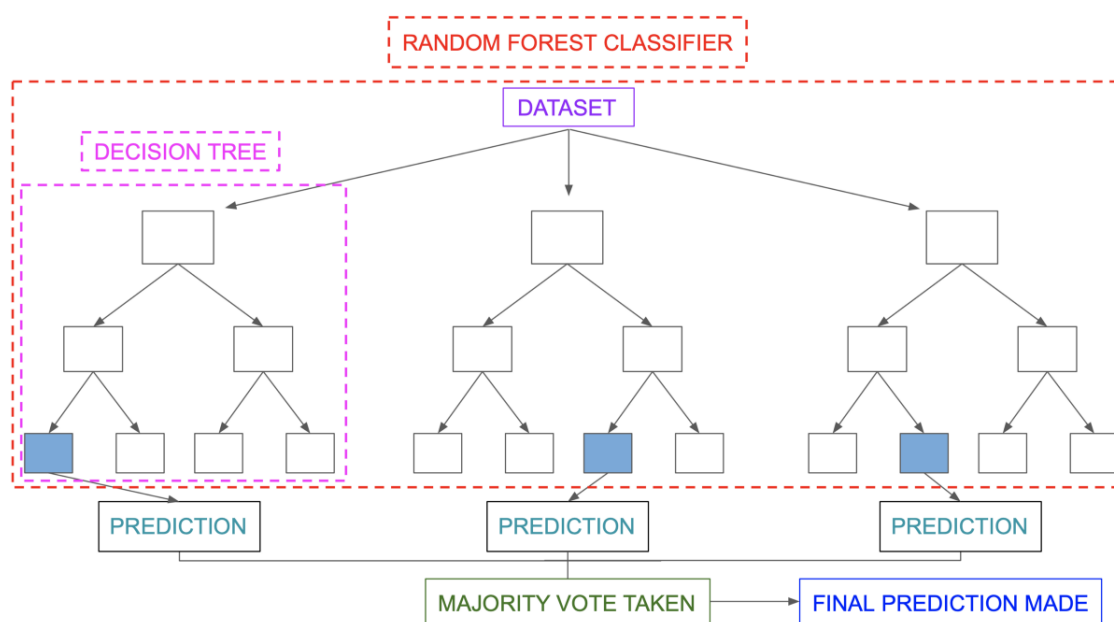
Η τεχνική αυτή ονομάζεται bootstrap sampling ή sampling with replacement και το τελικό δείγμα bootstrap sample (το οποίο είναι το input vector που δέχεται και ξεκινάει την ανάπτυξη του κάθε ένα δένδρο).

Εφαρμόζοντας την τεχνική αυτή, σε μεθόδους ταξινόμησης, δίνεται η δυνατότητα να δημιουργηθούν από ένα data set περισσότερα από ένα training sets. Αυτός είναι και ο λόγος που χρησιμοποιείται στα random forests, καθώς θέλουμε να δημιουργήσουμε πολλά διαφορετικά δένδρα και επομένως πολλά training sets (εφόσον κάθε δένδρο πρέπει να έχει το δικό του training set).

3. Επέλεξε τυχαία  $m \ll M$  από τις  $M$  μεταβλητές και το καλύτερο split που μπορεί να γίνει σε αυτές τις  $m$  μεταβλητές επιλέγεται για να χρησιμοποιηθεί στο node. Η τιμή του  $m$  παραμένει σταθερή κατά τη διάρκεια κατασκευής όλου του δάσους και παίζει πολύ σημαντικό ρόλο για το classification error του δάσους.
4. Κάθε δένδρο αναπτύσσεται στο μεγαλύτερο δυνατό βαθμό χωρίς να πραγματοποιείται «κλάδεμα» (no pruning).

Για το λόγο του ότι αναπτύσσονται πάρα πολλά decision trees το σφάλμα γενίκευσης (generalization error) περιορίζεται σημαντικά. Κάτι τέτοιο σημαίνει ότι είναι αδύνατο να εμφανιστεί το φαινόμενο της Υπέρ-ειδίκευσης (over-fitting), γεγονός που αποτελεί έναν πολύ ενισχυτικό παράγοντα για την πρόβλεψη των ταξινομητών (prediction)

Στην Εικόνα 3.5 παρακάτω, βλέπουμε το διάγραμμα ενός εκπαιδευμένου δάσους. Κάθε decision tree δέχεται ξεχωριστό data point (input vector of attributes) και δίνει διαφορετική ταξινόμηση «ψηφίζοντας» μία κλάση.



Σχήμα 3.5: random forest

## Κεφάλαιο 4

# Νευρωνικά δίκτυα

### 4.1 Τεχνητά νευρωνικά δίκτυα

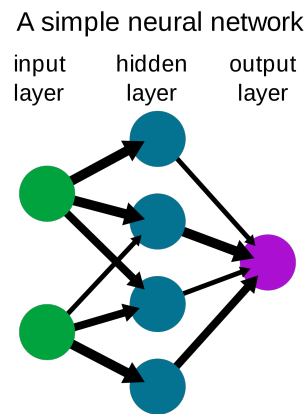
Τεχνητό νευρωνικό δίκτυο ονομάζεται ένα κύκλωμα διασυνδεδεμένων μονάδων επεξεργασίας που ονομάζουμε **Νευρώνες**. Στους υπολογιστές είναι ένα υπολογιστικό μοντέλο που χρησιμοποιείται για την επίλυση κάποιων υπολογιστικών προβλημάτων. [27]

Το νευρωνικό δίκτυο ονομάζεται δίκτυο καθώς αποτελείται από υπολογιστικούς κόμβους που συνδέονται μεταξύ τους. Κάθε υπολογιστικός κόμβος δέχεται ένα σύνολο αριθμητικών εισόδων (από άλλους νευρώνες είτε από κάποια άλλη είσοδο), εκτελεί έναν υπολογισμό με βάση αυτές τις εισόδους και παράγει μία έξοδο. Η έξοδος από αυτόν τον κόμβο μπορεί είτε να αποτελέσει μέρος της συνολικής εξόδου του ΤΝΔ είτε να διοχετευτεί σε άλλους κόμβους. Γενικά (αν και δεν είναι υποχρεωτική η διάκριση αυτή) θεωρούμε πως υπάρχουν τριών ειδών νευρώνες:

1. Οι νευρώνες εισόδου, των οποίων η εργασία είναι να διοχετεύσουν στους υπολογιστικούς νευρώνες την είσοδο του προβλήματος (πχ πρότυπα).
2. Οι νευρώνες εξόδου, χρησιμοποιούνται για να παρουσιάσουν στο περιβάλλον την απάντηση του ΤΝΔ σε κάποιο πρόβλημα, όπως για παράδειγμα την εκτίμηση της κατηγορίας ενός προβλήματος κατηγοριοποίησης.
3. Οι υπολογιστικοί νευρώνες ή κρυμμένοι νευρώνες, οι οποίοι πολλαπλασιάζουν κάθε είσοδο που δέχονται από νευρώνες εισόδου ή από άλλους νευρώνες επεξεργασίας με μια τιμή συσχετισμένη με αυτούς που ονομάζεται βάρος. Το συνολικό αποτέλεσμα εισάγεται σε μια συνάρτηση που θα ονομάζουμε συνάρτηση ενεργοποίησης και είτε παρουσιάζεται στην έξοδο είτε δίδεται σε κάποιον άλλο νευρώνα επεξεργασίας.

Ένα σχηματικό παράδειγμα ενός απλού ΤΝΔ παρουσιάζεται στο Σχήμα 4.1. Στο σχήμα αυτό υπάρχουν δύο νευρώνες εισόδου, τέσσερις νευρώνες επεξεργασίας και ένας νευ-

ρώνας εξόδου. Αν θέλουμε να εκφράσουμε την έξοδο ενός νευρώνα επεξεργασίας  $k$  θα μπορούσαμε να χρησιμοποιήσουμε την παρακάτω εξίσωση:



Σχήμα 4.1: Ένα απλό Τεχνητό Νευρωνικό Δίκτυο

$$y_k = f\left(\sum_{i=1}^d w_i x_i + b_k\right)$$

1.  $x_i$  είναι η  $i$  είσοδος από τις  $d$  που έχει το πρόβλημα
2.  $w_i$  είναι το βάρος της διασύνδεσης με την  $i$  είσοδο
3.  $b_k$  είναι η πόλωση για τον νευρώνα  $k$ . Η τιμή της πόλωσης συνήθως είναι ανεξάρτητη από το πρόβλημα
4. Η συνάρτηση  $f(x)$  είναι η συνάρτηση ενεργοποίησης και στην βιβλιογραφία χρησιμοποιείται μια πληθώρα συναρτήσεων ενεργοποίησης που παρουσιάζονται στην επόμενη ενότητα. [28] [29] [30]

## 4.2 Συναρτήσεις ενεργοποίησης

### Βηματική συνάρτηση

Είναι η πιο απλή συνάρτηση ενεργοποίησης και δίνεται από τον τύπο

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

Επειδή δεν είναι παραγωγίσιμη συνάρτηση δεν χρησιμοποιείται αρκετά συχνά.



### Γραμμική συνάρτηση

$$f(x) = x$$

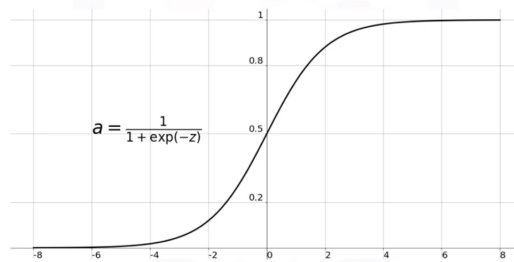
### Σιγμοειδής συνάρτηση

Είναι η συνάρτηση που χρησιμοποιείται πιο συχνά και ορίζεται από τον τύπο

$$\text{sig}(x) = \frac{1.0}{1 + e^{-x}}$$

. Έχει την γραφική παράσταση που φαίνεται στο Σχήμα 4.2.

#### Sigmoid Function



Σχήμα 4.2: Γραφική παράσταση σιγμοειδούς συνάρτησης

## 4.3 Δίκτυα Perceptron

Είναι η απλούστερη μορφή ΤΝΔ στα οποία υπάρχει μόνο ένας νευρώνας επεξεργασίας και έχει εφαρμογή κυρίως σε προβλήματα δύο κατηγοριών. Σε αυτήν την περίπτωση γίνεται ο υπολογισμός

$$u(x) = w^T x + b$$

όπου  $x$  είναι το διάνυσμα εισόδου,  $w$  είναι τα βάρη του νευρώνα και  $b$  είναι η πόλωση. Αν  $u(x) > 0$  τότε το πρότυπο πάει στην πρώτη κατηγορία και αν  $u(x) < 0$  στην δεύτερη. Τα βάρη  $w$  αναπροσαρμόζονται σύμφωνα με την εξίσωση:

$$w^{(k+1)} = w^{(k)} + n(y(x) - o(x))$$

όπου

1.  $y(x)$  είναι η πραγματική έξοδος για το πρότυπο  $x$  (+1 και -1)
2.  $o(x)$  είναι +1 αν  $u(x) > 0$  και -1 αν  $u(x) < 0$

3.  $n$  είναι ο ρυθμός μάθησης με  $n < 1$

Συνοπτικά ο αλγόριθμος μάθησης Perceptron [31] έχει ως ακολούθως:

1. Αρχικοποίηση των βαρών  $w_i^{(0)}, i = 1..d+1$  (+1 για την πόλωση). Η αρχικοποίηση μπορεί να σημαίνει είτε μηδενισμός των βαρών είτε αρχικοποίησή τους σε πολύ μικρές τυχαίες τιμές.
2. Καθορισμός  $n$
3. Καθορισμός  $EMAX$ , μεγίστου αριθμού επαναλήψεων
4.  $E = 0$  (αριθμός επαναλήψεων)
5. Για κάθε πρότυπο  $x = [1, x_1, x_2, \dots, x_d]$  στο σύνολο εκπαίδευσης (το 1 είναι για την πόλωση)
  - (α') Υπολογισμός της ποσότητας  $u(x) = w^T x$
  - (β') Αν  $u(x) < 0, o(x) = 1$  αλλιώς  $o(x) = -1$
  - (γ') Ενημέρωση βαρών:
 
$$w^{k+1} = w^k + n(y(x) - o(x))$$
  - (δ')  $k = k + 1$
6.  $E = E+1$
7. Αν δεν υπάρχει μεγάλη μεταβολή βαρών από γεννιά σε γεννιά ή  $E > EMAX$  τότε ο αλγόριθμος τερματίζει.

#### 4.4 Δίκτυα Πρόσθιας Τροφοδότησης

Τα δίκτυα πρόσθιας τροφοδότησης (Feed-Forward Neural networks) ή αλλιώς πολυ-επίπεδα Perceptron (Multi-layer perceptron) αποτελεί μια επέκταση του κλασσικού Perceptron. [32] Αποτελείται από τρεις τύπους επιπέδων, τα οποία είναι το επίπεδο εισόδου, το επίπεδο εξόδου και το κρυφό επίπεδο, όπως φαίνεται και στο σχήμα 4.3. Το επίπεδο εισόδου λαμβάνει ως είσοδο το σήμα εισόδου που πρόκειται να επεξεργαστεί. Η ζητούμενη εργασία της πρόβλεψης και της ταξινόμησης πραγματοποιείται στο επίπεδο εξόδου. Ένας αυθαίρετος αριθμός από κρυφά επίπεδα υπάρχουν μεταξύ του επιπέδου εισόδου και του επιπέδου εξόδου και τα οποία αποτελούν την πραγματική υπολογιστική μηχανή του πολυ-επίπεδου Perceptron. Τα δεδομένα μεταφέρονται από το επίπεδο εισόδου μόνο προς τα μπροστά προς τα κρυφά επίπεδα με την σειρά που έχουν τοποθετηθεί και

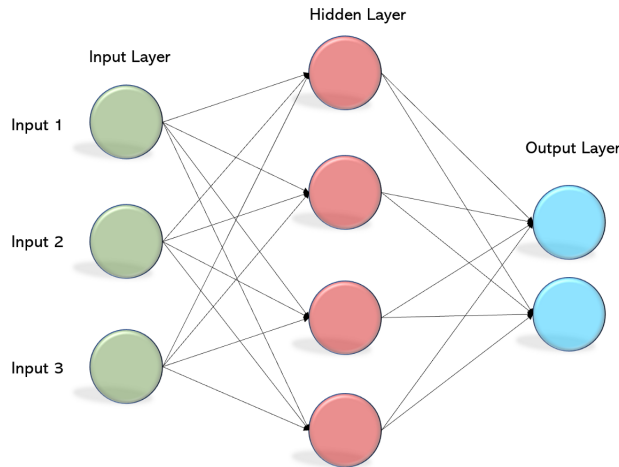
τέλος στο επίπεδο εξόδου. Οι νευρώνες ενός MLP εκπαιδεύονται μέσω του αλγορίθμου back-propagation. Αυτά τα δίκτυα σχεδιάζονται προσεγγιστικά με οποιαδήποτε συνεχή συνάρτηση και μπορούν να λύσουν προβλήματα τα οποία είναι μη γραμμικά διαχωρίσιμα. [33] Η κύρια χρήση των MLP είναι για μοτίβα ταξινόμησης, αναγνώρισης, πρόβλεψης και προσέγγισης.

Οι υπολογισμοί που γίνονται στους νευρώνες των κρυφών επιπέδων και του επιπέδου εξόδου έχουν ως εξής:

$$(1) o(x) = G(b(2) + W(2)h(x))$$

$$(2) h(x) = F(X) = s(b(1) + W(1)x)$$

Διανύσματα πόλωσης  $b(1), b(2)$ : πίνακες με βάρη  $W(1), W(2)$  και συναρτήσεις ενεργοποίησης  $G$  και  $s$ . Το σύνολο παραμέτρων προς εκμάθηση είναι το σύνολο  $j = W(1), b(1), W(2), b(2)$ . Τυπικές επιλογές για το  $s$  συμπεριλαμβάνουν την υπερβολική εφαιπτομένη  $\tanh$  με  $\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$  ή την σιγμοειδή συνάρτηση, δηλαδή  $\text{sigmoid}(a) = \frac{1}{1+e^{-a}}$



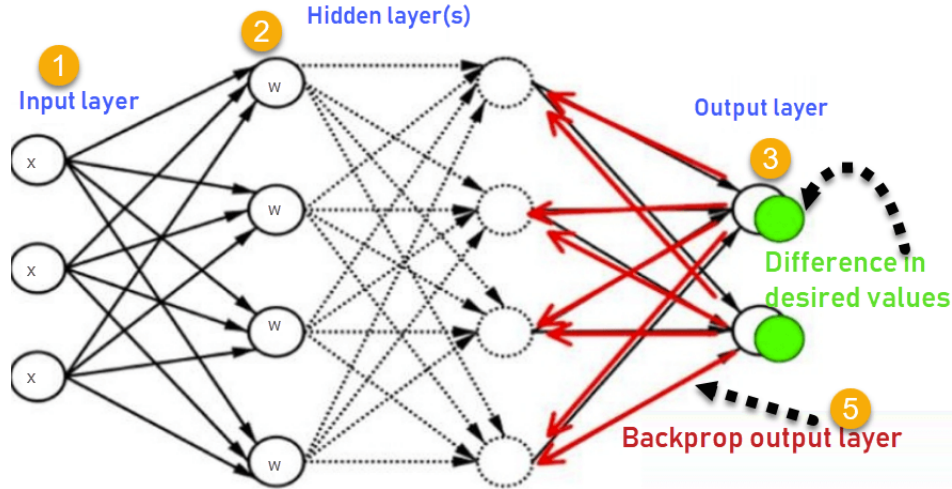
Σχήμα 4.3: Κλασσικό δίκτυο πρόσθιας τροφοδότησης

#### 4.4.1 Αλγόριθμος οπισθοδιάδοσης σφάλματος

Στα πολυεπίπεδα ΤΝΔ, ο συνηθέστερος τρόπος επιβλεπόμενης μάθησης είναι η μάθηση με οπισθοδιάδοση σφάλματος (backpropagation algorithm). Τα πολυεπίπεδα ΤΝΔ στα οποία ακολουθείται αυτή η μέθοδος εκπαίδευσης, καλούνται Back-Propagation ΤΝΔ. [34] [35]

Κατά την εκπαίδευση των συγκεκριμένων δικτύων, για κάθε είσοδο που δίνεται στο δίκτυο υπολογίζονται οι έξοδοι με εφαρμογή των συναρτήσεων μετάβασης σε κάθε μονάδα κρυφού ή εξωτερικού επιπέδου (βλέπε Σχήμα 4.4). Για κάθε μονάδα εξωτερικού

επιπέδου λαμβάνονται υπόψη οι διαφορές μεταξύ του υπολογιζόμενου και του επιθυμητού αποτελέσματος και διαδίδονται προς τα πίσω στις μονάδες του κρυφών επιπέδων, έτσι ώστε να καθορίσουν τις απαραίτητες αλλαγές στα βάρη σύνδεσης μεταξύ των μονάδων. Οι αλλαγές αυτές γίνονται βάσει του κανόνα εκμάθησης οπισθοδιάδοσης και ως σκοπό έχουν τη μείωση του εμφανιζόμενου στην έξοδο σφάλματος. Η συμμετοχή μιας μονάδας



Σχήμα 4.4: Εκπαίδευση δικτύου με οπισθοδιάδοση

στα σφάλματα των μονάδων του επόμενου επιπέδου της είναι ανάλογη της τρέχουσας εισόδου της και των συντελεστών βαρύτητας που τη συνδέουν με τις μονάδες του επόμενου επιπέδου. Το δίκτυο εν συνεχεία εφαρμόζει εκ νέου τις συναρτήσεις μετάβασης, για να υπολογίσει το νέο σφάλμα.

Το πραγματικό σφάλμα  $E_k$  μιας μονάδας εξόδου  $k$  ενός παραδείγματος  $p$  υπολογίζεται καταρχάς όπως και στον Perceptron, δηλαδή:

$$E_k = (a_{kp} - o_{kp})$$

Στην συνέχεια πολλαπλασιάζεται επί την παράγωγο της συνάρτησης ενεργοποίησης στη μονάδα  $k$  ( $u_k$ ), σύμφωνα με τον γενικευμένο κανόνα δέλτα, για να υπολογιστεί το λεγόμενο προσαρμοσμένο σφάλμα νευρώνα:

$$\delta_k = (a_{kp} - o_{kp}) \cdot g'(u_k)$$

Το αντίστοιχο σφάλμα σε μια μονάδα κρυφού επιπέδου  $i$  υπολογίζεται από τα προσαρμοσμένα σφάλματα στις  $k$  μονάδες του επόμενου επιπέδου με τις οποίες η μονάδα συνδέεται με βάρη  $w_{ik}$ , ως εξής:

$$\delta(i) = g'(u_i) \sum_{k=1}^k w_{ik} \cdot \delta_k$$

Η διαδικασία της εκπαίδευσης περιλαμβάνει πλήθος τέτοιων κύκλων διόρθωσης σφάλματος και λήγει με την μείωση του σφάλματος κάτω από ένα επιθυμητό όριο. Εναλλακτικά, ως συνθήκη τερματισμού μπορεί να θεωρηθεί η πραγματοποίηση ενός συγκεκριμένου αριθμού κύκλων εκπαίδευσης ή η πάροδος ενός συγκεκριμένου χρονικού διαστήματος.

### Αλλαγές στα βάρη

Αφού έχει υπολογιστεί για κάθε μονάδα  $i$  το σφάλμα  $\delta_i$ , η αλλαγή στα βάρη εισόδου σε όλους τους νευρώνες γίνεται ως εξής: [36]

$$\Delta w_{ji} = -d \cdot \delta_i \cdot a_j$$

Δηλαδή, η αλλαγή στο βάρος από τον νευρώνα  $i$  στον νευρώνα  $j$  εξαρτάται από το σφάλμα του νευρώνα  $i$ , την έξοδο του νευρώνα  $j$  και τον ρυθμό μάθησης (learning rate)  $d$ .

Υπάρχουν 2 τρόποι αλλαγής των βαρών κατά την παρουσίαση των παραδειγμάτων εκπαίδευσης:

- **Αυξητική εκπαίδευση** (incremental training), όπου για κάθε δείγμα που παρουσιάζεται γίνονται άμεσα οι αλλαγές των βαρών.
- **Μαζική εκπαίδευση** (batch training), όπου πρώτα παρουσιάζονται όλα τα παραδείγματα μια φορά, υπολογίζονται μόνο οι αλλαγές των βαρών για κάθε δείγμα και εφαρμόζονται ταυτόχρονα αφού παρουσιαστούν όλα τα παραδείγματα.

Η παρουσίαση όλων των παραδειγμάτων μια φορά (ανεξαρτήτως του τρόπου αλλαγής των βαρών) αντιστοιχεί σε έναν κύκλο εκπαίδευσης και ονομάζεται **εποχή εκπαίδευσης** (training-epoch) [37]



## Κεφάλαιο 5

# Συνελικτικά νευρωνικά δίκτυα (CNN)

Τα συνελικτικά νευρωνικά δίκτυα (CNN ή ConvNet) είναι μια ειδική κατηγορία των βαθιών δικτύων εμπρόσθιας τροφοδότησης και δημιουργήθηκαν με σκοπό να επεξεργάζονται δεδομένα με τοπολογία πλέγματος (grid) , όπως είναι οι ψηφιακές εικόνες και τα βίντεο. Σχεδιάστηκαν για πρώτη φορά το 1988 από τον Fukushima [38], αλλά ουσιαστική χρήση τους έγινε για πρώτη φορά το 1998 από την επιστημονική ομάδα του Lecun [39], με σκοπό την αυτοματοποιημένη αναγνώριση χειρόγραφων αριθμών.

Παρά την μεγάλη επιτυχία των απλών νευρωνικών δικτύων σε απλά προβλήματα μάθησης, οι πρώτες δυσκολίες έκαναν την εμφάνιση τους μαζί με την είσοδο πιο πολύπλοκων δεδομένων εισόδου, όπως είναι οι εικόνες, τα βίντεο και οι ηχογραφήσεις.

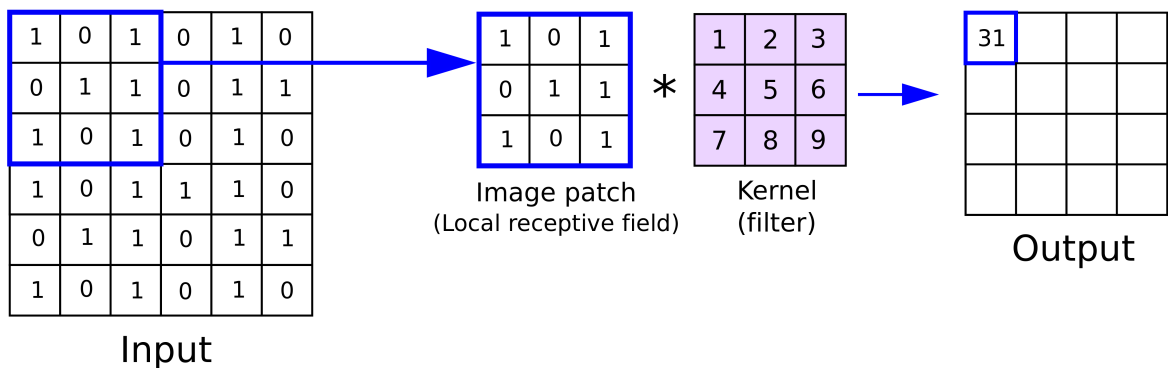
### 5.1 Αρχές Λειτουργίας

Η ψηφιακή εικόνα ουσιαστικά αποτελείται από πολλά ψηφία, τα οποία όταν συνδυαστούν μαζί μας δίνουν αυτό το οπτικό αποτέλεσμα. Κάθε ψηφίο περιλαμβάνει έναν αριθμό, ο οποίος υποδηλώνει την φωτεινότητα του συγκεκριμένου ψηφίου και το ακριβές χρώμα το οποίο περιλαμβάνει. Οι εικόνες μπορούν να είναι είτε grayscale είτε RGB (Red-Green-Blue). Αυτό σημαίνει πως εάν για παράδειγμα έχουμε μία  $100 \cdot 100$  grayscale εικόνα τότε η εικόνα είναι ασπρόμαυρη και τα 10000 ψηφία με τον αριθμό τους είναι αρκετά για προσδιοριστεί τελείως η εικόνα. Ωστόσο εάν είναι έγχρωμη τότε εκτός των  $100 \cdot 100$  ψηφίων θα μας χρειαστεί και μια τρίτη διάσταση των τριών χρωμάτων (κόκκινου, πράσινου και μπλε) τα οποία με την σειρά τους καθορίζουν πλήρως το χρώμα ενός pixel και συνεπώς καθορίζουν πλήρως την εικόνα. Σε αυτήν την περίπτωση μια εικόνα θα αποτελείται από πληροφορία  $100 \cdot 100 \cdot 3$  όπου  $100 \cdot 100$  τα ψηφία και 3 τα τρία χρώματα των ψηφίων που

προαναφέραμε. Κάθε ψηφίο έχει μια τιμή από 0-1 η οποία μας δίνει την φωτεινότητα του συγκεκριμένου ψηφίου, όπου 0 είναι το λιγότερο φωτεινό και 1 το περισσότερο. Με την ίδια λογική κάθε χρώμα του ψηφίου καθορίζεται από το 0-1, όπου για παράδειγμα μας δίνει με πόσο κόκκινο πρέπει να αναμείξουμε μια παλέτα χρωμάτων, 0 για το καθόλου και 1 για το πολύ.

### 5.1.1 Συνελίξεις

Έχουμε ήδη μιλήσει για τα κρυφά επίπεδα ενός (MLP) . Αυτά τα κρυφά επίπεδα των συνελικτικών νευρωνικών δικτύων που διαφοροποιούνται σε σχέση με τα υπόλοιπα νευρωνικά δίκτυα είναι τα επίπεδα της συνέλιξης (convolutional layers) . Τα επίπεδα αυτά είναι υπεύθυνα ώστε να γίνονται αντιληπτά κάποια μοτίβα σε μια εικόνα. Για παράδειγμα θέλουμε σε μια εικόνα να βρούμε που υπάρχουν γωνίες, κύκλοι, τρίγωνα και άλλα γεωμετρικά σχήματα. Αυτά τα 'μοτίβα' τα βρίσκουμε μέσα από τα συνελικτικά επίπεδα. Για να καταλάβουμε τι συμβαίνει από μαθηματικής απόψεως θα χρησιμοποιήσουμε το σχήμα 5.1. Όπως βλέπουμε από το σχήμα η εικόνα με τις διάφορες τιμές της (input)



Σχήμα 5.1: Εφαρμογή φίλτρου σε ένα patch

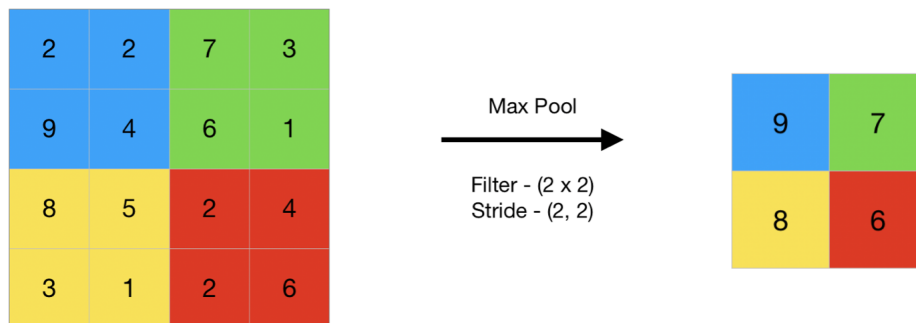
χωρίζεται σε τμήματα (patches). Παρατηρούμε επίσης το φίλτρο (kernel-filter) το οποίο είναι 3x3 και εφαρμόζεται σε κάθε 3x3 υπο-κομμάτι της εικόνας, ώστε να μας δώσει μια συγκεκριμένη έξοδο. Για παράδειγμα το συγκεκριμένο φίλτρο μπορεί να εστιάζει στην εύρεση μιας κάθετης γραμμής. Η πράξη που γίνεται μεταξύ του image patch και του φίλτρου είναι η πράξη της συνέλιξης. Στα επόμενα συνελικτικά επίπεδα τα φίλτρα που εφαρμόζονται μπορούν να γενικεύσουν πιο εύκολα για την εύρεση για παράδειγμα των



ματιών της μύτης ή των δοντιών ενός ζώου και στα ακόμη βαθύτερα να καταλήξει εάν στην εικόνα απεικονίζεται λιοντάρι, μαϊμού ή κάποιο άλλο ζώο.

### 5.1.2 Συγκέντρωση(Pooling)

Πολλές φορές τα επίπεδα συνέλιξης ακολουθούνται από ένα επίπεδο συγκέντρωσης. [40] Σκοπός αυτών των επιπέδων είναι να μειώσουν την διαστατικότητα (dimensionality). Εδώ το μέγεθος του φίλτρου εφαρμόζεται με τον ίδιο τρόπο όπως και στα επίπεδα της συνέλιξης. Βέβαια μπορεί να καθορίσει κανείς κάποιον βηματισμό με τον οποίο θα εφαρμόζεται το φίλτρο στα δεδομένα εισόδου. Ο βηματισμός(stride) δείχνει πόσα βήματα πρέπει να μεταβούμε από το συγκεκριμένο ψηφίο για να εφαρμόσουμε το φίλτρο της συγκέντρωσης (το φίλτρο μπορεί να είναι κάποια γνωστή μαθηματική συνάρτηση). Όλα θα γίνουν πιο κατανοητά με ένα παράδειγμα. Έστω το σχήμα 5.2, στο οποίο έχουμε δεδομένα εισόδου (πιθανότατα αποτέλεσμα από ένα συνελικτικό επίπεδο) μεγέθους  $4 \times 4$ . Στο συγκεκριμένο παράδειγμα εφαρμόζουμε συγκέντρωση με φίλτρο max (MaxPooling2D)  $2 \times 2$  και stride (2,2). Το φίλτρο μας δίνει στα πόσα ψηφία θα εφαρμοστεί η συγκέντρωση και ο βηματισμός πόσο θα μετακινηθούμε για να εφαρμόσουμε το επόμενο φίλτρο από την περιοχή που βρισκόμαστε. Ουσιαστικά εδώ μεταφερόμαστε από μπλοκ  $2 \times 2$  σε μπλοκ  $2 \times 2$ . Τέλος το φίλτρο μας λέει ποια συνάρτηση θα εφαρμόσουμε. Στο συγκεκριμένο παράδειγμα το *MaxPooling* μας λέει να πάρουμε το μέγιστο κάθε μπλοκ. Μπορεί σε άλλο παράδειγμα να λαμβάναμε τον μέσο όρο του μπλοκ (Average-Pooling2D)



Σχήμα 5.2: Εφαρμογή MaxPooling

### 5.1.3 Μη γραμμικότητα

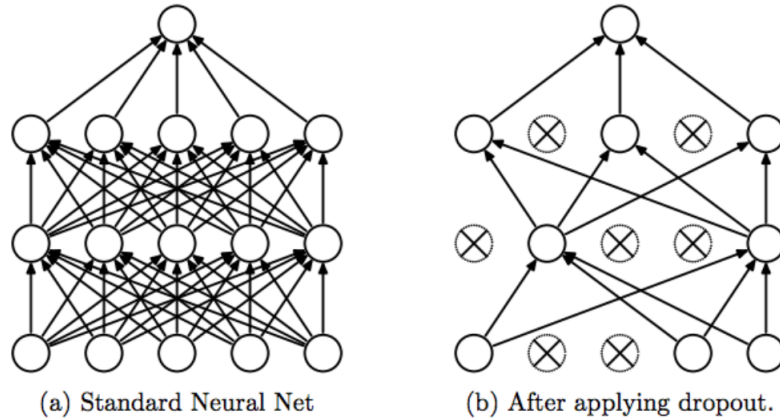
Η εισαγωγή της μη γραμμικότητας (non-linearity) δίνει σοβαρό πλεονέκτημα στα CNN έναντι άλλων γνωστών μεθόδων αντιμετώπισης πολλών προβλημάτων. Σαν παράδειγμα αναφέρουμε την περίπτωση ενός μη γραμμικού συστήματος και την προσπάθεια

πρόβλεψης αυτού, όπου οι γραμμικές μέθοδοι αδυνατούν να δώσουν καλά αποτελέσματα, ιδιαίτερα όταν το σύστημα εμφανίζει χαοτική συμπεριφορά. Έτσι, γίνεται αμέσως αντιληπτό το πλεονέκτημα των μη γραμμικών νευρωνικών δικτύων έναντι άλλων γνωστών γραμμικών μεθόδων. Μερικά παραδείγματα μη γραμμικότητας αποτελούν η , η sigmoid και η ευρέως χρησιμοποιούμενη Rectified Linear Unit (ReLU)

Πιο κατάλληλη θεωρείται η ReLU, για το λόγο ότι απλοποιεί τον αλγόριθμο back-propagation και τον επιταχύνει, εφόσον οι πράξεις που χρειάζεται να εκτελεστούν είναι πολύ λιγότερες. [41]

## 5.2 Dropout

Υπάρχουν διάφορες μέθοδοι για να μειώσουμε τα σφάλματα εκπαίδευσης ενός Νευρωνικού Δικτύου, όπως για παράδειγμα να συγκρίνουμε τις προβλέψεις πολλών και διαφορετικών μοντέλων. Για μεγάλα Νευρωνικά Δίκτυα των οποίων η εκπαίδευση μπορεί να κρατήσει αρκετές ημέρες αυτή η μέθοδος είναι χρονοβόρα και σε ορισμένες περιπτώσεις ίσως και αδύνατη. Προς αντιμετώπιση αυτού του προβλήματος έχει αναπτυχθεί κάποια μέθοδος σύγκρισης μοντέλων αρκετά αποτελεσματική, το υπολογιστικό κόστος της οποίας είναι πολύ χαμηλό. Η μέθοδος αυτή, η οποία ονομάζεται “Dropout” συνίσταται στην ανάθεση ως “0” της εξόδου κάθε κρυφού νευρώνα με πιθανότητα 0.5. Οι νευρώνες που συμμετέχουν με αυτό τον τρόπο στο “Dropout” δεν συνεισφέρουν στη διάδοση προς τα εμπρός των σημάτων εκπαίδευσης και δεν συμμετέχουν στην διαδικασία του back-propagation. Έτσι, κάθε φορά που μια είσοδος παρουσιάζεται στο Δίκτυο, εκείνο χρησιμοποιεί διαφορετική αρχιτεκτονική αλλά όλες αυτές οι αρχιτεκτονικές μοιράζονται τα ίδια βάρη. Αυτή η τεχνική μειώνει τις περίπλοκες συν-προσαρμογές των νευρώνων κι έτσι το Δίκτυο αποκτά τη δυνατότητα εκμάθησης ισχυρότερων χαρακτηριστικών. Κατά τη διαδικασία της επαλήθευσης (test) χρησιμοποιούμε όλους τους νευρώνες πολλαπλασιάζοντας όμως τις εξόδους τους με 0.5 προκειμένου να πάρουμε τον γεωμετρικό μέσο των κατανομών πρόβλεψης. Τέλος, αναφέρουμε ότι η χρήση του Dropout σχεδόν διπλασιάζει τον αριθμό των επαναλήψεων που απαιτούνται για σύγκλιση. Ένα παράδειγμα στο οποίο φαίνονται οι “παγωμένοι” νευρώνες λόγω εφαρμογής του Dropout φαίνεται στο σχήμα 5.3.



Σχήμα 5.3: Εφαρμογή Dropout

### 5.2.1 Normalization

Η κανονικοποίηση (Normalization), μπορεί να εφαρμοσθεί εντός ενός χαρακτηριστικού ή και μεταξύ ενός συνόλου χαρακτηριστικών και προσφέρει μικρότερη διακύμανση στα δεδομένα. Επίσης μπορεί να εφαρμοσθεί πριν ή μετά το pooling χωρίς ιδιαίτερη διαφορά μεταξύ των περιπτώσεων, ενώ δεν είναι απαραίτητη πάντοτε. Οι πιο συχές μορφές κανονικοποίησης είναι οι L1 και L2.

#### Κανονικοποίηση L1

Η ιδέα της κανονικοποίησης L1 είναι η αλλαγή της συνάρτησης μη κανονικοποιημένου κόστους προσθέτοντας το άθροισμα των απόλυτων τιμών των βαρών όπως φαίνεται στη σχέση παρακάτω. Αυτό που θέλουμε να πετύχουμε είναι να 'τιμωρήσουμε' τα μεγάλα βάρη και να κάνουμε το τεχνητό νευρωνικό δίκτυο να προτιμά βάρη με μικρότερη τιμή.

$$C' = C + \frac{\lambda}{n} \sum_w |w|$$

Διαφορίζοντας την παραπάνω σχέση παίρνουμε

$$\frac{\partial C}{\partial w_k} = \frac{\partial C}{\partial w} + \frac{\lambda}{n} \cdot \text{sgn}(w)$$

, όπου  $\text{sgn}(w)$  είναι το πρόσημο του  $w$  δηλαδή  $+1$  αν το  $w$  είναι θετικό και  $-1$  εάν το  $w$  είναι αρνητικό.

Χρησιμοποιώντας την παραπάνω σχέση μπορούμε εύκολα να τροποποιήσουμε τον κανόνα μάθησης, ώστε να χρησιμοποιεί την κανονικοποίηση L1. Ο κανόνας ενημέρωσης των βαρών είναι ο παρακάτω:

$$w_k \leftarrow w_k - \frac{\eta \lambda}{n} \text{sgn}(w) - \eta \frac{\partial C}{\partial w_k}$$

### Κανονικοποίηση L2

Η ιδέα της κανονικοποίησης L2 (φθορά βάρους) είναι να προστεθεί ένας επιπλέον όρος στην προϋπάρχουσα συνάρτηση κόστους, έναν όρο που ονομάζεται όρος ρύθμισης. Η δράση της κανονικοποίησης L2 φαίνεται στη σχέση παρακάτω, όπου ο πρώτος όρος είναι η προϋπάρχουσα συνάρτηση κόστους και ο δεύτερος όρος είναι το άθροισμα των τετραγώνων όλων των βαρών του τεχνητού νευρωνικού δικτύου. Αυτός όρος κλιμακώνεται από το παράγοντα  $\frac{\lambda}{2n}$  όπου το  $\lambda$  είναι πάντα θετικό ( $\lambda > 0$ ) και ονομάζεται παράγοντας κανονικοποίησης και το  $n$  είναι το μέγεθος του συνόλου εκπαίδευσης. Εδώ αξίζει να σημειωθεί ότι ο παράγοντας κανονικοποίησης δεν περιλαμβάνει τις παραμέτρους μεροληψίας.

$$C' = C + \frac{\lambda}{2n} \sum_w w^2$$

Διαισθητικά, το αποτέλεσμα της κανονικοποίησης L2 είναι να επηρεάσει το δίκτυο ώστε να προτιμά να μαθαίνει μικρά βάρη. Τα μεγάλα βάρη θα αλλάζουν μόνο εάν βελτιωθεί σημαντικά το πρώτο μέρος της συνάρτησης κόστους.

Με άλλο τρόπο, η κανονικοποίηση L2 μπορεί να θεωρηθεί ως ένας τρόπος συμβιβασμού μεταξύ της εξεύρεσης μικρών βαρών και της ελαχιστοποίησης της αρχικής συνάρτησης κόστους. Η σχετική σημασία των δύο στοιχείων του συμβιβασμού εξαρτάται από την τιμή του  $\lambda$ :

όταν  $\lambda$  είναι μικρό τότε αρχική συνάρτηση κόστους ελαχιστοποιείται, όταν όμως το  $\lambda$  είναι μεγάλο προτιμούμε τα μικρά βάρη. Έτσι οι κανόνες μάθησης γίνονται αυτοί των παρακάτω σχέσεων.

$$w_k \rightarrow w_{k'} = w_k - \eta \frac{\partial C}{\partial w_k} - \frac{\eta \lambda}{n} w_k = (1 - \frac{\eta \lambda}{n}) w_k - \eta \frac{\partial C}{\partial w_k}$$

$$b_l \rightarrow b_{l'} = b_l - \eta \frac{\partial C}{\partial b_l}$$

Εάν συγκρίνουμε αυτούς τους κανόνες με τους κανόνες των σχέσεων της L1 παρατηρούμε ότι ο κανόνας μάθησης για τις παραμέτρους μεροληψίας παραμένει ο ίδιος ενώ ο κανόνας μάθησης για τα βάρη πολλαπλασιάζει το προηγούμενο βάρος με το παράγοντα  $(1 - \frac{\eta \lambda}{n})$ . Με την πρώτη ματιά φαίνεται ότι αυτό σημαίνει ότι τα βάρη οδηγούνται προς το μηδέν. Αλλά αυτό δεν ισχύει πάντα, καθώς ο άλλος όρος μπορεί να οδηγήσει στην αύξηση των βαρών, αν κάτι τέτοιο προκαλέσει μείωση της μη κανονικοποιημένης συνάρτησης κόστους.

Συγκρίνοντας τις σχέσεις των κανονικοποιήσεων L1 και L2 αντίστοιχα, παρατηρούμε ότι έχουν κάποια κοινά χαρακτηριστικά καθώς και οι δύο τεχνικές προσπαθούν να συρρικνώσουν τα βάρη. Αυτό συμφωνεί με τον σκοπό χρήσης τους καθώς και στις δύο

θέλουμε να ‘τιμωρήσουμε’ τα μεγάλα βάρη. Αλλά ο τρόπος με τον οποίο τα βάρη συρρικνώνεται είναι διαφορετικός. Στην κανονικοποίηση L1, τα βάρη συρρικνώνονται κατά ένα σταθερό ποσό προς το 0. Στην κανονικοποίηση L2, τα βάρη συρρικνώνονται κατά ένα ποσό το οποίο είναι ανάλογο του βάρους  $w$ . Έτσι, όταν ένα συγκεκριμένο βάρος έχει ένα μεγάλο μέγεθος,  $|w|$ , η κανονικοποίηση L1 συρρικνώνει το βάρος πολύ λιγότερο από ότι κάνει η κανονικοποίηση L2. Αντίθετα, όταν η τιμή  $|w|$  είναι μικρή, τότε η κανονικοποίηση L1 συρρικνώνει το βάρος πολύ περισσότερο από την κανονικοποίηση L2. Το καθαρό αποτέλεσμα είναι ότι η κανονικοποίηση L1 τείνει να συγκεντρώνει το βάρος του δικτύου σε σχετικά μικρό αριθμό συνδέσεων μεγάλης σημασίας, ενώ τα άλλα βάρη τείνουν προς το μηδέν (Nielsen, 2015).

### 5.3 Προ-εκπαίδευση

Κατά την εκπαίδευση ενός Συνελικτικού Νευρωνικού Δικτύου, η διαδικασία της οπισθοδιάδοσης σφάλματος συνίσταται στη διάδοση κλίσεων στα προηγούμενα επίπεδα του Δικτύου για τη διόρθωση των βαρών των φίλτρων που έχουν υπολογιστεί. [42] Μερικά από τα πρώτα επίπεδα ενός πολύ-επίπεδου Νευρωνικού Δικτύου αλλάζουν ελάχιστα την τιμή τους λόγω της μειωμένης διάδοσης της κλίσης (εκπαιδεύονται ανεπαρκώς) και αυτό μπορεί να αποτελέσει σημαντικό πρόβλημα. Η λύση έρχεται από την διάθεση ενός μεγάλου αριθμού δειγμάτων, τουλάχιστον 5000 δείγματα ανά κλάση και μετά από αρκετό χρόνο εκπαίδευσης το δίκτυο καταφέρνει να αποκτήσει την τελική και αποδοτική του μορφή.

Όταν τα δεδομένα εκπαίδευσης είναι ανεπαρκή, απαιτείται προ-εκπαίδευση και η τεχνική που ακολουθείται είναι η παροχή στο Δίκτυο διανυσμάτων-εικόνων μεγάλης βάσης δεδομένων από τομέα σχετικό με την εφαρμογή που θέλουμε να δημιουργήσουμε, ώστε να επέλθει σύγκλιση των παραμέτρων του Δικτύου. Στη συνέχεια, κατά την εκπαίδευση, δίνονται ως είσοδοι εικόνες-διανύσματα από τα δεδομένα που έχουμε στη διάθεσή μας. Στην περίπτωση των εικόνων, ένα επιπλέον βήμα που μπορεί να ακολουθηθεί, είναι η επέκταση της βάσης δεδομένων μέσα από επεξεργασία των εικόνων εκμεταλλευόμενοι του στοιχείου της “τοπικότητας” των Συνελικτικών Νευρωνικών Δικτύων. Έτσι, για κάθε εικόνα μπορούμε να δημιουργήσουμε μερικά ελαφρώς παραλλαγμένα αντίγραφα, όπου έχουμε αλλάξει τις τιμές αριθμού τυχαίων pixel πάνω στην εικόνα, πχ ανάθεση σε ορισμένα pixel της τιμής “0” ή “1”, περιστροφές των εικόνων για ελάχιστες μοίρες (-10 έως +10 μοίρες) και μετατόπιση των εικόνων σε δυο διευθύνσεις με βήμα ένα μικρό αριθμό pixels (translation). Με αυτό τον τρόπο, δίνουμε στο Δίκτυο μια άλλη “οπτική” της ίδιας πληροφορίας, όπως συμβαίνει και στις εικόνες RGB όπου το Δίκτυο έχει στη

διάθεσή του τρείς διαφορετικές εκδοχές της ίδιας εικόνας, με αποτέλεσμα η εκπαίδευσή του να γίνεται πιο αποτελεσματική. [43]

Γενικά η προ-εκπαίδευση:

- Κάνει την βελτιστοποίηση του δικτύου πιο εύκολη
- Μειώνει την υπερπροσαρμογή (overfitting)

Σε άλλες μορφές νευρωνικών δικτύων, για παράδειγμα στους Autoencoders και τις RBMs, η προ-εκπαίδευση είναι αρκετά σημαντική αφού ένα από τα θεμελιώδη ζητήματα για την ορθή εκπαίδευσή τους αποτελεί η ανάθεση κατάλληλων αρχικών τιμών των βαρών τους.

## Κεφάλαιο 6

# Πειραματική διαδικασία

### 6.1 Συλλογή και Προεπεξεργασία δεδομένων

#### 6.1.1 Συλλογή

Σκοπός της παρούσας εργασίας είναι η εφαρμογή ενός μοντέλου όπως αυτό του κεφαλαίου 5. Πιο συγκεκριμένα να αναπτυχθεί ένα συνελικτικό νευρωνικό δίκτυο, με σκοπό να καταφέρουμε την επίτευξη του διαχωρισμού πασχόντων, από κάποια νευροεκφυλιστική ασθένεια, και μη πασχόντων ατόμων εξετάζοντας τις μαγνητικές τους τομογραφίες. Στο συγκεκριμένο πρόβλημα μελετήθηκαν τομογραφίες ανθρώπων με την ασθένεια του Alzheimer . Η ασθένεια του Alzheimer πρόκειται για μία μη θεραπεύσιμη και θανατηφόρα νευροεκφυλιστική νόσο, στην οποία οφείλεται η πλειοψηφία των περιπτώσεων άνοιας.

Τα δεδομένα των μαγνητικών τομογραφιών εγκεφάλου για τα πειράματά μας λήφθηκαν από την βάση δεδομένων ADNI (Alzheimer’s Disease Neuroimaging Initiative). Η βάση αυτή έχει δημιουργηθεί από τον Michael W. Weiner το 2003 και αποτελείται από ιατρικά στοιχεία ασθενών. Σκοπός της δημιουργίας της είναι η μελέτη του κατά πόσον οι τομογραφίες εκπομπής ποζιτρονίων, άλλοι βιολογικοί δείκτες και νευροψυχολογικές αξιολογήσεις μπορούν να συνδυαστούν για να μετρηθεί η πρόοδος των MCI και AD(Alzheimer Disease). Επομένως γίνεται αντιληπτό πως τα δεδομένα αυτά έχουν να κάνουν αποκλειστικά με την καταπολέμηση του Alzheimer και όχι με άλλη ασθένεια.

Για την λήψη των μαγνητικών ήταν απαραίτητη η εγγραφή στην ιστοσελίδα <https://ida.loni.usc.edu/>, η οποία είναι μια σελίδα φτιαγμένη για αποθήκευση εικόνων σε συνεργασία με το πανεπιστήμιο της Νότιας Καλιφόρνιας. Η πρόσβαση στα δεδομένα μου δόθηκε μέσω της ακαδημαϊκής μου ταυτότητας κατόπιν αίτησης. Για το ‘κατέβασμα’ download των δεδομένων στον σέρβερ της σχολής έπρεπε να συνδεθώ μέσω VPN στον

server και μέσω X-11 forwarding στον chrome για να συνδεθώ με τα διαπιστευτήρια (credentials) στο ADNI.

Για την συγκεκριμένη μελέτη αξιοποιήθηκαν οι T1 τομές από ασθενείς που είχαν χαρακτηριστεί CN (Normal), MCI, ή AD. Οι μαγνητικές αυτές είχαν Acquisition Type 3D και Acquisition Plane Sagittal. Έγινε αρχικά μια προσπάθεια να συγχωνευθούν και άλλα Acquisition Planes, δηλαδή Coronal-Axial αλλά η προσπάθεια αυτή εγκαταλείφθηκε γρήγορα, καθώς υπήρχαν αρκετές διαφορές στις μαγνητικές αυτές, οι οποίες θα μπερδευαν στο έπακρο το συνελικτικό μας δίκτυο. Έχουμε λοιπόν το πρόβλημα τριών κλάσεων ADvsMCIvsCN. Οι μαγνητικές που λάβαμε σε αριθμούς αφορούσαν 94, 207 και 149 μαγνητικές ανθρώπων που περιγράφηκαν ως AD, MCI και CN αντίστοιχα. Οι μαγνητικές αυτές είχαν διαστάσεις (256,256,170) όλες για λόγους συνέχειας των δεδομένων και πιο εύκολης προεπεξεργασίας, όπου 170 είναι το βάθος της μαγνητικής και (256,256) το ύψος και το πλάτος αυτής. Δεν λήφθηκαν υπόψιν οι επιδόσεις των ασθενών στα σχετικά τεστ (FAQ,GDSCALE,Global CDR(Clinical Dementia Rating),MMSE, NPI-Q)) τα οποία ίσως υποδηλώνουν την ενδεχόμενη αλλαγή διάγνωσης ενός ασθενούς από μια κλάση σε κάποια άλλη.

### 6.1.2 Προεπεξεργασία

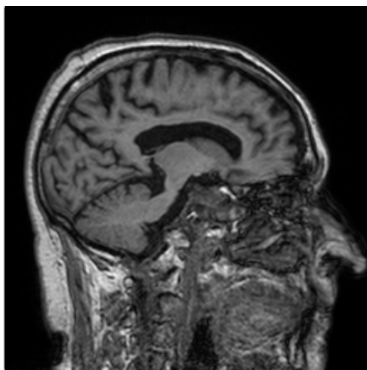
Οι παραπάνω μαγνητικές ήταν αρχεία της μορφής dual Nifty format(.nii) και για την επεξεργασία τους χρησιμοποιήσαμε την βιβλιοθήκη nibabel (Η γλώσσα προγραμματισμού που χρησιμοποιήθηκε για τα πειράματα ήταν η Python 2.7.18), η οποία μέσω της συνάρτησης getfdata() μας έδωσε ένα numpy array διαστάσεων (256,256,170) για να έχουμε την πληροφορία την τρισδιάστατης εικόνα. Η συγκεκριμένη βιβλιοθήκη εκτός της μαγνητικής μας δίνει πληροφορίες σχετικά με την επικεφαλίδα (header) και τους αφινικούς μετασχηματισμούς (affine transformation) σε περίπτωση που θέλουμε να τις κάνουμε πιο εύκολα resize.

Στο σχήμα 6.1 έχουμε μια μαγνητική σε μέσο βάθος και εύκολα γίνεται αντιληπτό ότι μας δίνεται επιπλέον πληροφορία που δεν μας ενδιαφέρει (παρατηρούμε την μύτη, το στόμα ή το κρανίο του ασθενούς). Επίσης στα πρώτα και στα τελευταία επίπεδα της μαγνητικής μας δίνονται μόνο μαύρες εικόνες, οι οποίες δεν δίνουν κάποια χρήσιμη πληροφορία στο δίκτυο μας. Όπως παρατηρήθηκε αυτές οι εικόνες αποτελούν τα πρώτα 45 και τα τελευταία 55 επίπεδα του άξονα z. Επιπλέον στα επίπεδα βάθους (depth) που τελικά αποσπάσαμε μας ενδιαφέρει μονάχα το κομμάτι του εγκεφάλου. Οπότε περιορίσαμε την εικόνα σε αυτήν του σχήματος 6.1. Στον άξονα των x λάβαμε τα εικονοστοιχεία



(pixels) από 50-210 (από το σύνολο 256) και στον άξονα των  $y$  από 40-150. Έτσι έχουμε διαστάσεις των επεξεργασμένων εικόνων (110,160,88). Η παραπάνω προεπεξεργασία ήταν απαραίτητη, καθώς τα στοιχεία που αφαιρέσαμε δεν έδιναν κάποια πληροφορία για τον εγκέφαλο και ανέβαζε την πολυπλοκότητα και τις παραμέτρους του νευρωνικού δικτύου μας. Στα σχήματα 6.1 και 6.2 φαίνεται το είδος της προεπεξεργασίας. Η ιδέα για τα παραπάνω προέκυψε από την μέθοδο της αφαίρεσης εγκεφάλου skull-stripping κατά την οποία αφαιρούνται τα κομμάτια του κρανίου προκειμένου να γίνει πιο εύκολη η επεξεργασία του εγκεφάλου. [45]

Επίσης τα δεδομένα μας δεν είναι πολλά ώστε να θεωρηθούν αρκετά για να δικαιολο-



Σχήμα 6.1: Αρχική μαγνητική

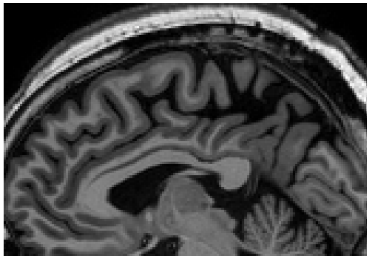


Σχήμα 6.2: Κατόπιν αφαίρεσης των εικονοστοιχείων

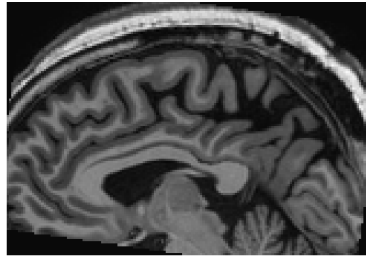
γήσουν την εκπαίδευση ενός συνελικτικού νευρωνικού δικτύου. Επιπλέον οι μαγνητικές που προαναφέραμε σε κάποιες περιπτώσεις ανήκουν στο ίδιο άτομο για τον λόγο ότι εκείνος έκανε παραπάνω από μία επισκέψεις. Επομένως από τις 94 μαγνητικές με την ετικέτα Αλτςχάιμερ έχουμε συμμετοχή 49 ατόμων. Ομοίως οι 207 MCI μαγνητικές ανήκαν σε 115 ανθρώπους και οι 149 CN ανήκουν σε 80 ανθρώπους. Επειδή λοιπόν, παρόλο που υπάρχουν αλλαγές στις μαγνητικές με την πάροδο του χρόνου, ελοχεύει ο κίνδυνος της υπερεκπαίδευσης (overfitting) λόγω του ότι τα δεδομένα παρουσιάζουν ομοιότητες, εφαρμόστηκε επαύξηση δεδομένων (data augmentation). Η επαύξηση δεδομένων είναι κατάλληλη για προβλήματα στα οποία δεν έχουμε αρκετά δεδομένα ή δεν μπορούμε να βρούμε αρκετά δεδομένα ή στις περιπτώσεις που τα δεδομένα μοιάζουν μεταξύ τους και αναζητούμε έναν μεγαλύτερο πλουραρισμό στα δεδομένα μας για να αποφύγουμε την υπερεκπαίδευση [44]. Με άλλα λόγια δημιουργήσαμε νέα δεδομένα από τα ήδη υπάρχοντα για να βοηθήσουμε το δίκτυο μας να εκπαιδευτεί πιο εύκολα στην φάση του training και για να αποφύγει την υπερεκπαίδευση στην φάση της αξιολόγησης. Αξίζει να αναφερθεί πως συνήθως η επαύξηση δεδομένων γίνεται στο σύνολο της εκπαίδευσης, προκειμένου να είμαστε σίγουροι ότι το νευρωνικό μας δίκτυο γενικεύει και σε άλλες περιπτώσεις

εικόνων.

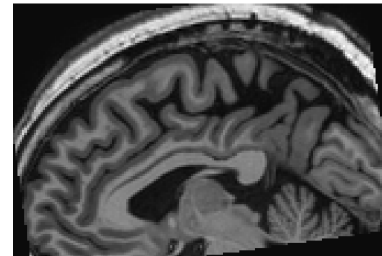
Για κάθε μοναδικό άνθρωπο στην κλάση του AD εφαρμόσαμε περιστροφή της μίας μαγνητικής του κατά +10 και -10 μοίρες, ώστε να αλλάξουμε λιγάκι τα δεδομένα αλλά όχι αρκετά για να αλλοιώσουμε την πληροφορία τους. Επομένως προσθέσαμε 98 ακόμα εικόνες στις ήδη υπάρχουσες φτάνοντας τις 192. Ίδια τακτική εφαρμόσαμε και στους ανθρώπους που χαρακτηρίστηκαν MCI-CN αλλά πήραμε λιγότερους και όχι όλους, καθώς δεν θέλαμε να δημιουργήσουμε unbalanced dataset. Παράδειγμα μιας τέτοιας εφαρμογής στην πράξη φαίνεται στις παρακάτω 3 εικόνες. Φαίνεται η διαφοροποίηση που



Σχήμα 6.3: Μαγνητική του σετ δεδομένων



Σχήμα 6.4: Στραμμένη κατά 10 μοίρες



Σχήμα 6.5: Στραμμένη κατά -10 μοίρες

υπάρχει στις εικόνες δίχως όμως να αλλοιώνεται η απαιτούμενη πληροφορία. Με αυτόν τον τρόπο καταφέραμε να αυξήσουμε τα λίγα δεδομένα μας και μάλιστα να βελτιώσουμε το μοντέλο μας.

Έτσι καταλήξαμε να έχουμε τον αριθμό δεδομένων που φαίνεται στον πίνακα 6.1 όπου η τελευταία γραμμή δίνει τον τελικό αριθμό δεδομένων του dataset.

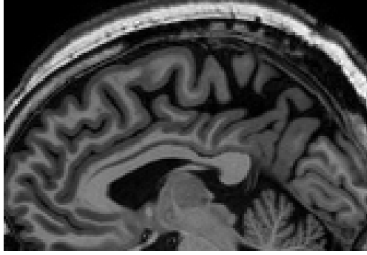
	AD	MCI	CN
Unique Subject	49	115	80
Number of Images	94	207	149
Final Number of Images	192	307	259

Table 6.1: Dataset review

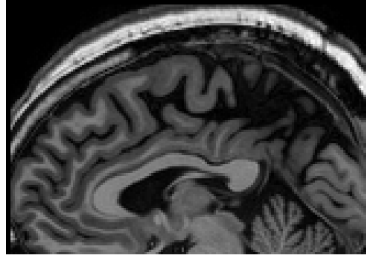
Στην συνέχεια μέσω της βιβλιοθήκης `scipy` επιτύχαμε την περιστροφή (rotation) των εικόνων κατά -10 και +10 μοιρών αντίστοιχα προσθέτοντας τις μαγνητικές των εικόνων 6.2 και 6.3 στο σετ δεδομένων μας. Πιο συγκεκριμένα κάναμε `import` από την `scipy` την `ndimage.rotation` η οποία κάνει την περιστροφή της τρισδιάστατης εικόνας δίχως να χανθεί πολύτιμη πληροφορία.

Βέβαια η τεχνική της περιστροφής των εικόνων δεν είναι η μοναδική όταν μιλάμε για `data augmentation`. Αφού λάβαμε τα αποτελέσματα του μοντέλου από την παραπάνω

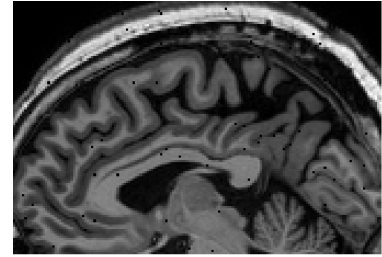
διαδικασία, προσθέσαμε κάποιες επιπλέον εικόνες χρησιμοποιώντας και άλλες τεχνικές. Στην συγκεκριμένη εργασία προσθέσαμε σε κάποιες εικόνες μια αντίστοιχη μετατοπισμένη κατά 2 pixels (shifted) σε κάθε έναν άξονα. Επίσης προσθέσαμε θόρυβο salt & pepper, μέσω του οποίου αντικαθιστούμε τυχαία εικονοστοιχεία με τιμή 0. Στις εικόνες 6.6, 6.7, 6.8 φαίνεται η εφαρμογή αυτών των τεχνικών.



Σχήμα 6.6: Μαγνητική του σετ δεδομένων



Σχήμα 6.7: Μετατοπισμένη εικόνα



Σχήμα 6.8: Εικόνα με θόρυβο

Τα δεδομένα που λάβαμε μετά την εφαρμογή αυτών των τεχνικών φαίνονται παρακάτω:

	AD	MCI	CN
Unique Subject	49	115	80
Number of Images	94	207	149
Final Number of Images	290	327	269

Table 6.2: Dataset review2

Για την κανονικοποίηση των εικόνων-numpy arrays στην κλίμακα 0-1 χρησιμοποιήθηκε ο MinMaxScaler κατά τον οποίο λαμβάνεται η μικρότερη τιμή του dataset, που στην περίπτωση μας είναι το 0 αφού μιλάμε για εικόνες και η μέγιστη τιμή τους. Στην συνέχεια σε κάθε ένα στοιχείο του πίνακα που αποτελείται η εικόνα εφαρμόζεται η πράξη

$$\frac{data - min}{max - min}$$

, όπου data κάποιο εικονοστοιχείο της μαγνητικής.

Τέλος, αξίζει να αναφερθεί πως ο ανθρώπινος εγκέφαλος στους πολύ μεγάλους ανθρώπους (άνω των 87) τείνει να μοιάζει με εκείνο των ασθενών με Αλτσχάιμερ παρά το γεγονός ότι μπορεί να είναι υγιείς [46]. Για τον λόγω αυτό στην προεπισκόπηση του dataset αφαιρέθηκαν από το σετ δεδομένων μας αυτοί οι άνθρωποι για την διευκόλυνση του νευρωνικού μας δικτύου.

## 6.2 Μεθοδολογία

Στο πρόβλημα των τριών κλάσεων του ADNI γίνεται μια προσπάθεια δημιουργίας ενός συνελκτικού νευρωνικού δικτύου για των διαχωρισμό τους στην συγκεκριμένη εργασία.

### 6.2.1 Επιλογή Μοντέλου

Στην αρχή η πρώτη μας σκέψη ήταν να αναπτύξουμε ένα συνελκτικό νευρωνικό δίκτυο 2 διαστάσεων (Conv2D). Η ιδέα στηρίχτηκε στο γεγονός του χωρισμού κάθε τρισδιάστατης μαγνητικής σε δισδιάστατα κομμάτια, τα οποία θα επαυξήσουν επαρκώς τα δεδομένα μας και θα έδιναν την ευκαιρία εκμάθησης του νευρωνικού δικτύου πιο εύκολα. Ωστόσο η προσπάθεια αυτής της προσέγγισης εγκαταλείφθηκε γρήγορα για τον λόγο ότι υπήρχε ένας μεγαλύτερος κίνδυνος υπερκαίδευσης του μοντέλου αφού τα δισδιάστατα κομμάτια μιας και μόνο μαγνητικής θα έμοιαζαν αρκετά μεταξύ τους. Άλλος ένας λόγος ήταν ότι οι πληροφορίες που πρέπει να πάρει τον νευρωνικό μας δίκτυο πρέπει να είναι μιας ολόκληρης μαγνητικής για να βγάλει πόρισμα, εφόσον το Αλτζχάϊμερ προσβάλλει στις περισσότερες περιπτώσεις συγκεκριμένες περιοχές του εγκεφάλου (περιοχές της αμυγδαλής, του ενδορινικού φλοιού, του ιππόκαμπου και της παραϊπποκάμπειας έλικας). [47]

Ως εκ τούτου καταλήξαμε στην δημιουργία ενός συνελκτικού νευρωνικού δικτύου τριών διαστάσεων (Conv3D) για το πρόβλημα ταξινόμησης. Για αυτόν τον σκοπό χωρίσαμε το σύνολο των δεδομένων μας σε 60% train set 20% test set και μόλις 20% validation set.

	AD	MCI	CN
train set	117	179	164
test set	34	65	53
validation	41	63	46

Table 6.3: Train, Test and Validation Sets

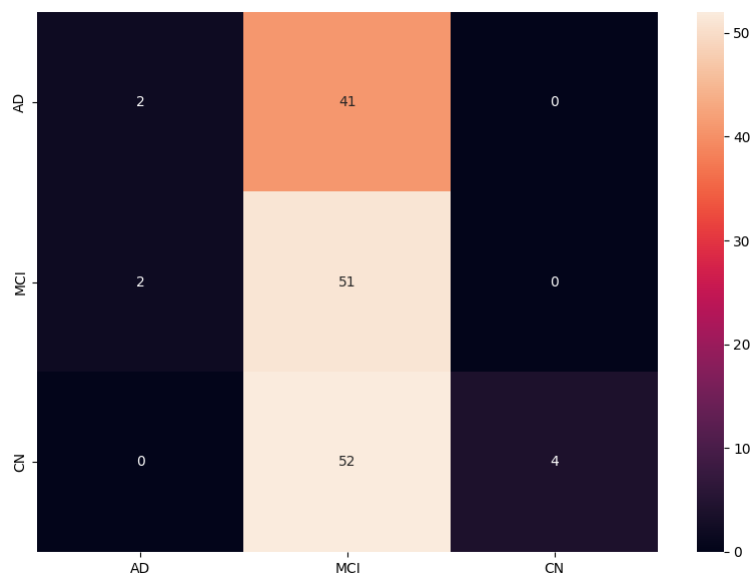
Το δίκτυο μας δεν αποτελείται από ιδιαιτέρως πολλά επίπεδα νευρώνων, καθώς σε άλλη περίπτωση τείνει να προβλέψει μια συγκεκριμένη κλάση. Τα χαρακτηριστικά του δικτύου φαίνονται στο παρακάτω πίνακάκι:

Αρχικά προσπαθήσαμε να δημιουργήσουμε αρκετά επίπεδα νευρώνων συνέλιξης μαζί με πολλές παραμέτρους. Κάτι τέτοιο βέβαια οδηγούσε το μοντέλο να προβλέπει συγκεκριμένα μία εκ των τριών κλάσεων (σχήμα 6.5) με αποτέλεσμα να χρησιμοποιήσουμε

Επίπεδα	Σχήμα Εξόδου	Αριθμός παραμέτρων
Conv3D	(None, 108, 158, 86, 8)	224
MaxPooling3D	(None,54,79,43,8)	0
Conv3D	(None, 52, 77, 41, 16)	3472
MaxPooling3D	(None, 26, 38, 20, 32)	0
Dropout	(None, 26, 38, 20, 16)	0
Conv3D	(None, 24, 36, 18, 64)	27712
MaxPooling3D	(None, 12, 18, 9, 64)	0
Dropout	(None, 12, 18, 9, 64)	0
Flatten	(None, 124416)	0
Dense	(None, 64)	7962688
Dropout	(None, 64)	0
Dense	(None,3)	195

Table 6.4: Χαρακτηριστικά συνελικτικού δικτύου

μικρότερο αριθμό επιπέδων συνέλιξης.



Σχήμα 6.9: Πίνακας σύγκρισης πολλών συνελικτικών επιπέδων

Στο παραπάνω σχήμα βλέπουμε πώς όταν έχουμε λίγο πιο περίπλοκο νευρωνικό δίκτυο εκείνο τείνει να προβλέψει την κλάση MCI πιθανώς επειδή είναι η πιο συχνή κλάση στο σετ δεδομένων μας. Έτσι καταλήξαμε στα συνελικτικά επίπεδα του πίνακα 6.3.

### 6.2.2 Πρόωρο Σταμάτημα (Early Stopping)

Για την σωστή αξιολόγηση του μοντέλου πρέπει να γίνει και σωστή επιλογή των epochs κατά τα οποία θα γίνεται η εκπαίδευση του μοντέλου. Αυτή η επιλογή επίσης στηρίζεται στα γραφήματα απεικόνισης των train και validation accuracies και losses. Τα παρακάτω πειράματα έγιναν για τα δεδομένα με επαύξηση δεδομένων μόνο με περιστροφή (πίνακας 6.1 και πίνακας 6.3).

Χωρίς την χρήση του Dropout το πρόωρο σταμάτημα γινόταν σε προηγούμενη εποχή, γεγονός που δεν άφηνε το δίκτυο να μάθει στο μέγιστο των δυνατοτήτων του. Έτσι εκ των υστέρων προστέθηκαν τα Dropout layers. Αυτό καταδεικνύει και το σχήμα 6.10, στο οποίο το πρόωρο σταμάτημα ξεκινά από το 11ο epoch:

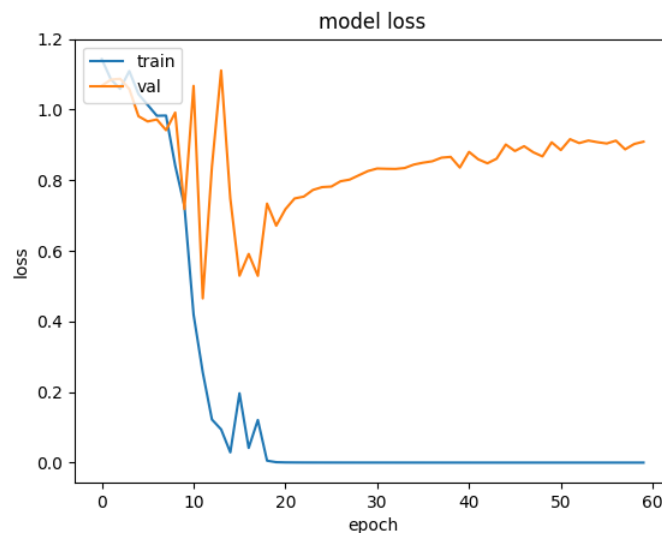


Figure 6.10: Train and validation loss before dropout

Ας ρίξουμε μια ματιά στα παρακάτω 2 γραφήματα, τα οποία είναι τα αποτελέσματα του μοντέλου μας (Πίνακας 6.2) για 60 εποχές και εφαρμογή του Dropout μόνο στο Dense Layer:

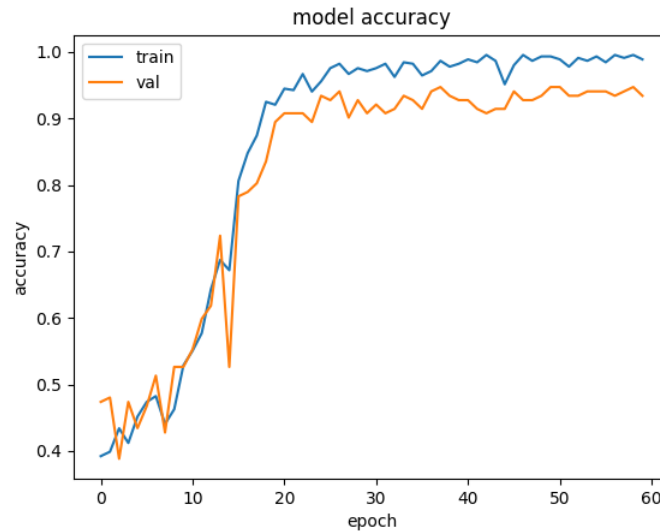


Figure 6.11: Train and validation accuracy after dropout in dense layer

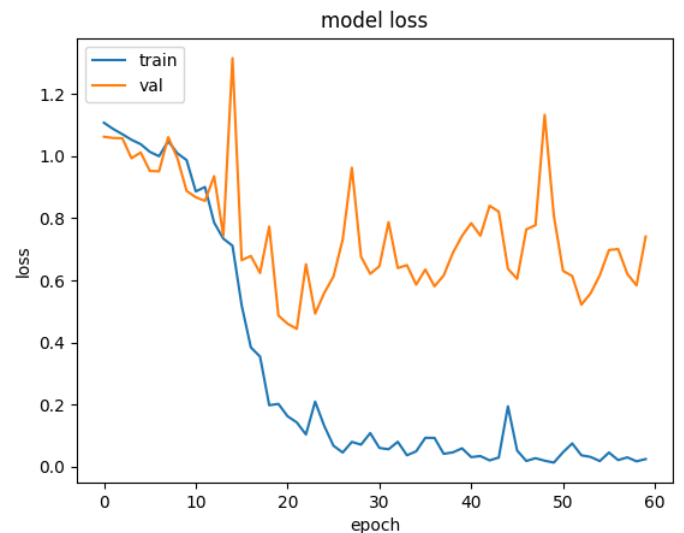
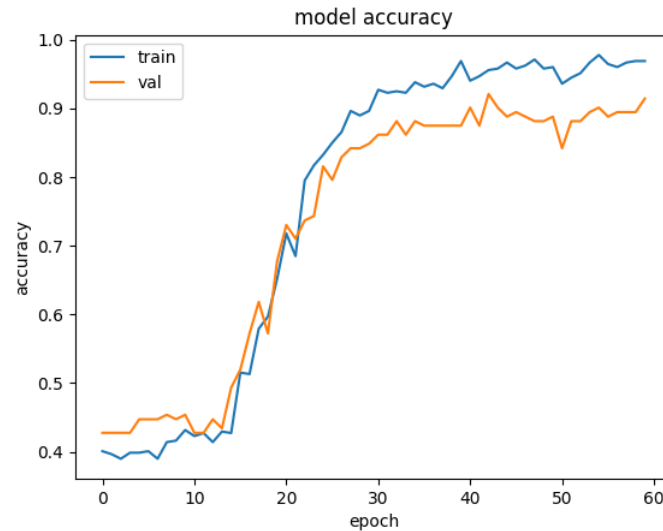


Figure 6.12: Train and validation loss after dropout in dense layer

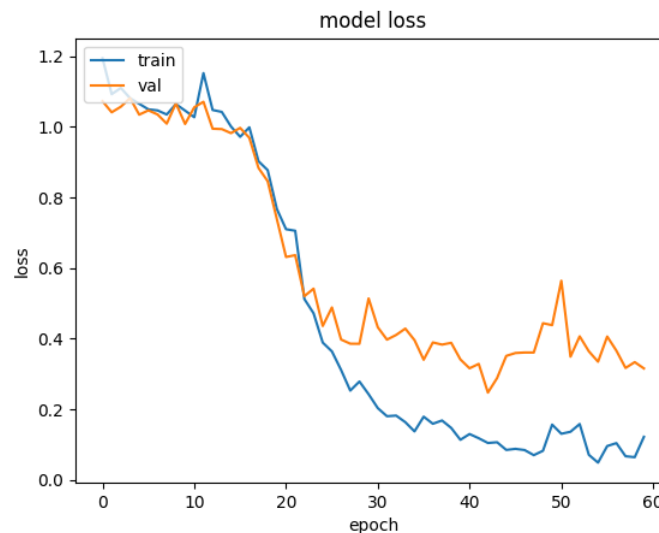
Παρατηρούμε κάτι κοινό και στα 2 γραφήματα. Μετέπειτα της 14ης εποχής στο σχήμα 6.11 η ακρίβεια του validation set αρχίζει να σταθεροποιείται και πιθανόν να πέφτει μετά τις 60. Στο σχήμα 6.12 μετά την 14η εποχή οι 2 γραφικές παραστάσεις για τα 2 σετ δεδομένων ξεκινούν να αποκλίνουν η μία από την άλλη, ενώ μέχρι εκεί είχαν παρόμοια πορεία. Αυτή η συμπεριφορά ονομάζεται πρόωρο σταμάτημα (early stopping) και υποδεικνύει την εποχή στην οποία πρέπει να σταματήσουμε την εκπαίδευση του μοντέλου. Εάν η εκπαίδευση συνεχιζόταν τότε θα είχαμε υπερεκπαίδευση και τα αποτελέσματα ε-

κτίμησης του τεστ σετ δεν θα αντικατόπτριζαν την πραγματικότητα.

Ωστόσο για το ίδιο σετ δεδομένων εφαρμόσαμε Dropout και μετά τα συνελικτικά δίκτυα. Οι γραφικές παραστάσεις απώλειας και ακρίβειας φαίνονται καλύτερες κατόπιν αυτής της εφαρμογής, αφού το πρόωρο σταμάτημα ξεκινάει στην 22η εποχή:



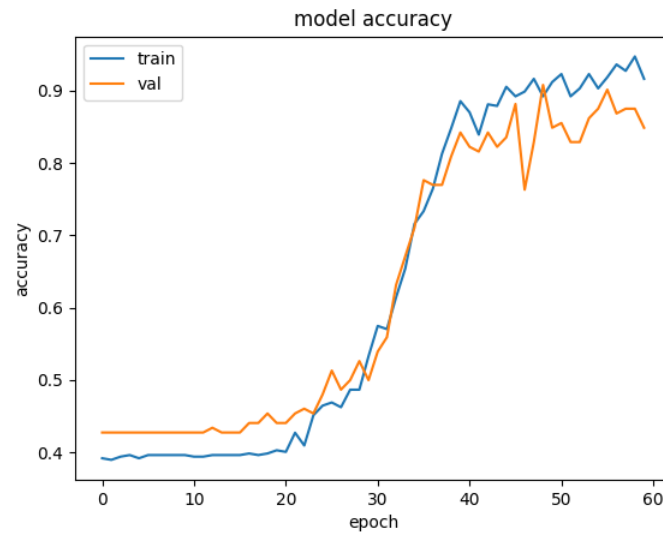
Σχήμα 6.13: Train and validation accuracy after dropout in conv layers



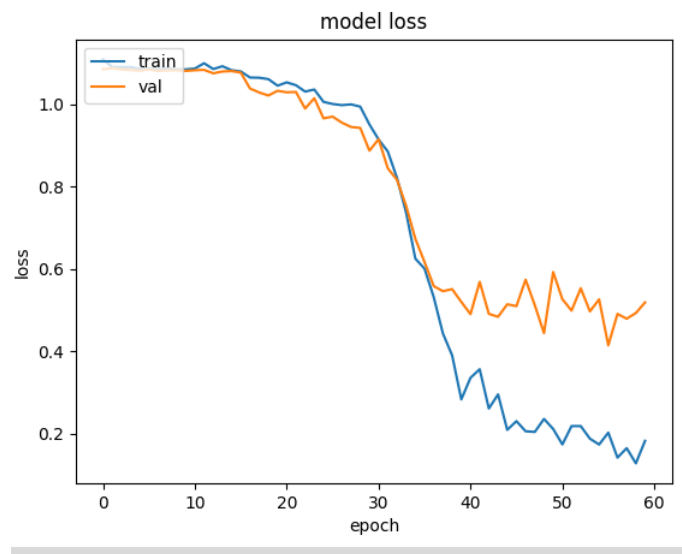
Σχήμα 6.14: Train and validation loss after dropout in conv layers

Τις καλύτερες γραφικές παραστάσεις τις λάβαμε μέσω εφαρμογής της κανονικοποίησης L2 στο προτελευταίο dense layer . Μας έδωσε την δυνατότητα εκπαίδευσης του μοντέλου μέχρι τις 35 εποχές. Παρακάτω φαίνονται οι παραστάσεις:





Σχήμα 6.15: Train and validation accuracy after dropout + regularization



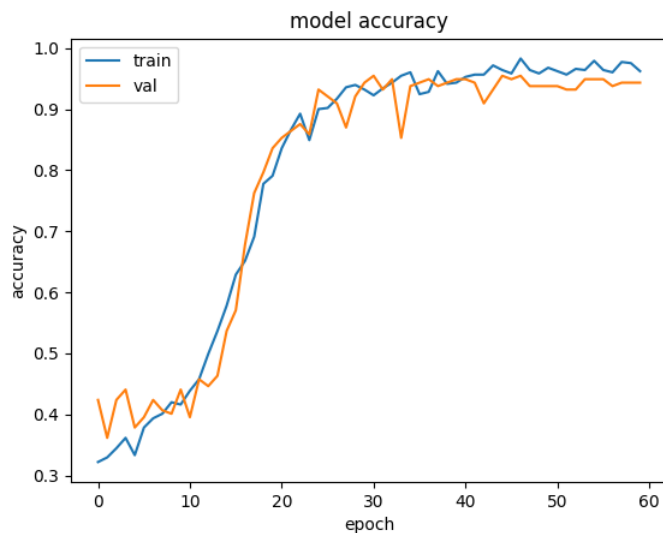
Σχήμα 6.16: Train and validation loss after dropout + regularization

Τέλος αυξήσαμε και άλλο τα ήδη υπάρχοντα δεδομένα στο σύνολο εκπαίδευσης, προσθέτοντας τις μετατοπισμένες εικόνες και τις εικόνες με τον θόρυβο salt and pepper. Το σύνολο δεδομένων περιείχε τα εξής:

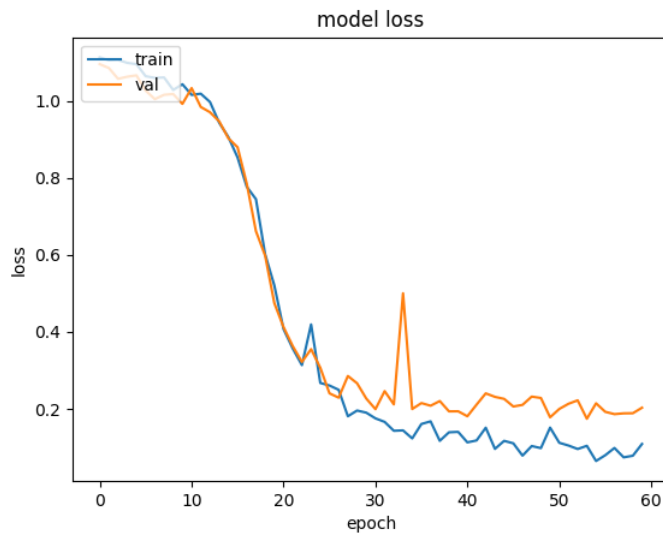
Οι γραφικές παραστάσεις απωλειών και ακρίβειας για το ίδιο μοντέλο με παραπάνω αλλά διαφοροποιημένο dataset φαίνονται παρακάτω:

	AD	MCI	CN
train set	117	179	164
test set	34	65	53
validation	41	63	46

Table 6.5: New Train, Test and Validation Sets



Σχήμα 6.17: Train and validation accuracy for new dataset



Σχήμα 6.18: Train and validation loss for new dataset

Βλέπουμε ότι η εκπαίδευση πρέπει να σταματήσει περίπου στην 22η-23η εποχή σε αυτήν την περίπτωση.

### 6.2.3 Αποτελέσματα-Εκτίμηση του μοντέλου

Έτσι σύμφωνα με τα προηγούμενα υποκεφάλαια αυτού του κεφαλαίου διαλέξαμε το νευρωνικό δίκτυο του πίνακα 6.3. Επιπλέον πήραμε τα αποτελέσματα από την εφαρμογή του μοντέλου αυτού για τρεις περιπτώσεις:

1. Το μοντέλο χωρίς την εφαρμογή κανονικοποίησης στο dense layer μόνο για τα γνήσια δεδομένα και εκείνα με εφαρμογή της περιστροφής στο σύνολο εκπαίδευσης (train-set).(πίνακας 6.3)
2. Το μοντέλο με εφαρμογή κανονικοποίησης στο dense layer μόνο για τα γνήσια δεδομένα και εκείνα με εφαρμογή της περιστροφής στο σύνολο εκπαίδευσης(train-set).(πίνακας 6.3)
3. Το μοντέλο με εφαρμογή κανονικοποίησης στο dense layer μόνο για τα γνήσια δεδομένα και εκείνα με εφαρμογή τεχνικών περιστροφής, μετατόπισης και θορύβου salt pepper στο σύνολο εκπαίδευσης(train-set).(πίνακας 6.5)

Για συνάρτηση ενεργοποίησης των επιπέδων συνέλιξης διαλέξαμε την συνάρτηση relu και για το επίπεδο εξόδου (Dense) την συνάρτηση softmax (Βλέπε κεφάλαιο 4.2).

#### Πρώτη περίπτωση

Για τον λόγο του πρόωρου σταματήματος και όπως άλλωστε μας υποδεικνύουν τα σχήματα 6.13 και 6.14 η καλύτερη επιλογή για τα epochs είναι η τιμή 26.

Η εκτίμηση του μοντέλου που έγινε για το πρόβλημα τριών κλάσεων έγινε σύμφωνα με την "αντιπαλότητα" ADvsMCIvsCN . Οι μετρικές που χρησιμοποιήθηκαν για την αξιολόγηση του μοντέλου ήταν οι accuracy, η οποία παρουσιάζεται και με πιο ωραίο τρόπο εμμέσως μέσω του πίνακα σύγχυσης (confusion matrix), η ακρίβεια (precision), η ανάκληση (recall) και η f1 score.

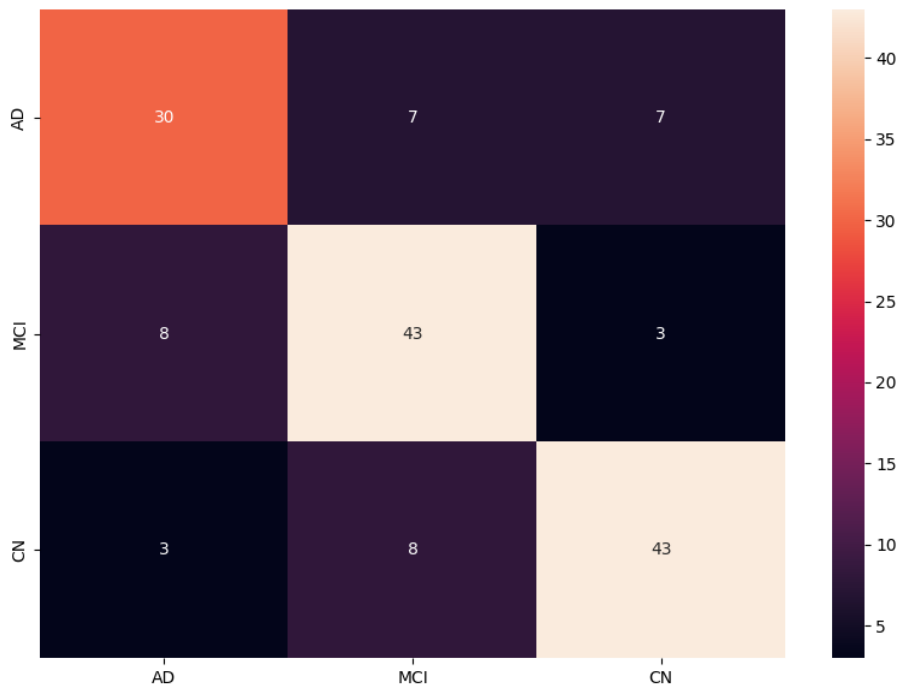
Αρχικά πρέπει να καθοριστούν ορισμένοι όροι:

- TP (True positives): Είναι τα παραδείγματα του test set που προβλέθηκαν με την ετικέτα της θετικής κλάσης σωστά. Στο πρόβλημα τριών κλάσεων ως θετική κλάση είναι η κλάση που αξιολογούμε και αρνητικές οι υπόλοιπες
- FP (False Positives): Είναι τα παραδείγματα του test set που προβλέθηκαν με την ετικέτα της θετικής κλάσης λανθασμένα.
- TN (True Negative): Είναι τα παραδείγματα του test set που προβλέθηκαν με την ετικέτα της αρνητικής κλάσης σωστά. Δηλαδή με κάποια άλλη κλάση εκτός της

επιλεγμένης.

- FN (False Negative): Είναι τα παραδείγματα του test set που προβλέθηκαν με την ετικέτα της αρνητικής κλάσης λανθασμένα.

Η μετρική accuracy είναι  $acc = \frac{TP+TN}{TP+TN+FN+FP}$ , δηλαδή είναι το κλάσμα των σωστών προβλέψεων δια του αριθμού των προβλέψεων και είναι πολύ σημαντική μετρική για το πείραμα μας. Επιτύχαμε accuracy με τιμή **76,3%**. Ο πίνακας σύγχυσης δείνει τις σωστές προβλέψεις κάθε κλάσης. Η κύρια διαγώνιος του πίνακα αυτού πρέπει να λαμβάνει μεγάλες τιμές. Ο δικός μας πίνακας σύγχυσης είναι ο παρακάτω (Σχήμα 6.9):

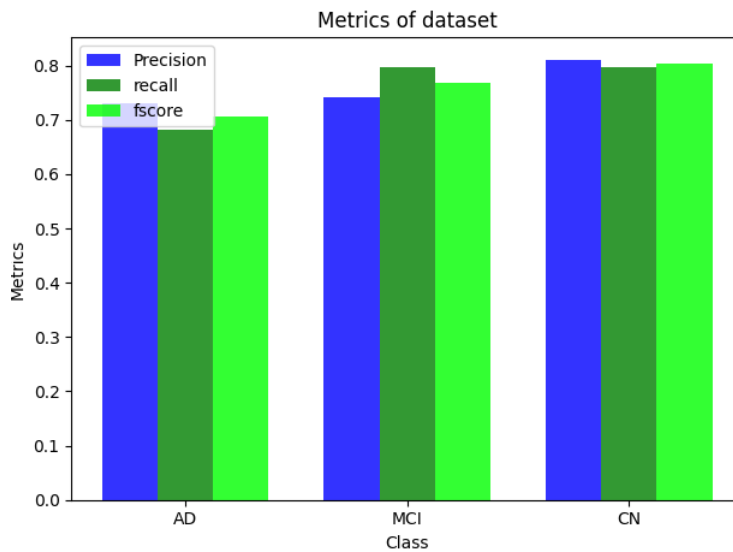


Σχήμα 6.19: Πίνακας σύγχυσης του μοντέλου μας

Στην συνέχεια παρουσιάζουμε τις μετρικές precision, recall και f1score για κάθε μία κλάση. Ο ορισμός των μετρικών αυτών είναι ο εξής:

- Precision: Είναι το κλάσμα  $\frac{TP}{TP+NP}$  και υπολογίζει το ποσοστό των 'θετικών' ενδείξεων της επιλεγμένης κλάσης που προβλέφθηκαν σωστά.
- Recall: Είναι το κλάσμα  $\frac{TP}{TP+NF}$  και υπολογίζει το ποσοστό των 'θετικών' προβλέψεων της επιλεγμένης κλάσης που κατάφεραν να προβλεφθούν σωστά.
- FScore: Είναι το κλάσμα  $\frac{2 \cdot recall \cdot precision}{precision + recall}$  και συνδυάζει την ακρίβεια και την ανάκληση. Συνήθως πιο χρήσιμη μετρική στα μη ισορροπημένα datasets.

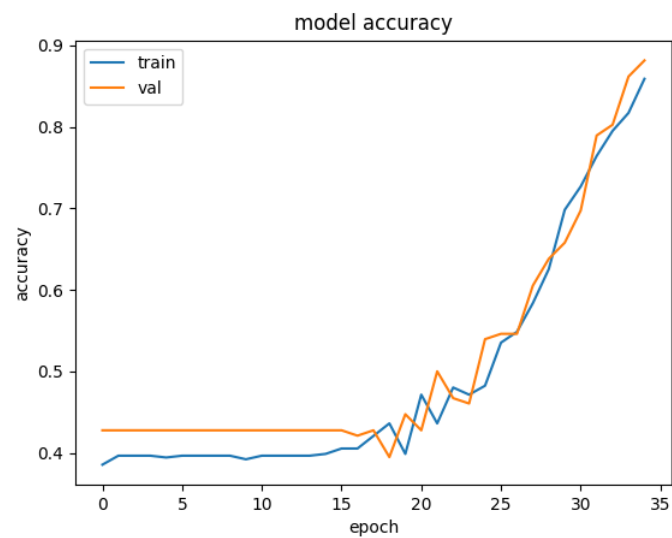
Έχουμε το εξής ραβδόγραμμα που μας δίνει τις μετρικές για κάθε κλάση (Σχήμα 6.10):



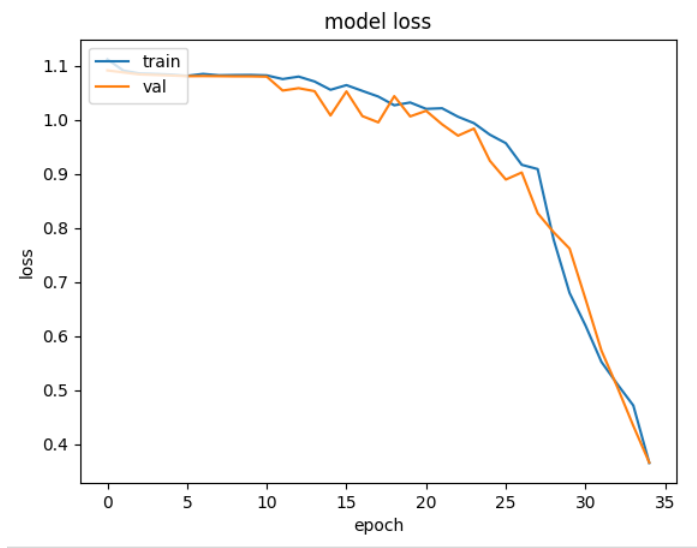
Σχήμα 6.20: Μετρικές της πρώτης περίπτωσης

### Δεύτερη περίπτωση

Τα σχήματα 6.15 και 6.16 μας υποδεικνύουν ότι ο αριθμός των εποχών εκπαίδευσης είναι οι 35 εποχές.

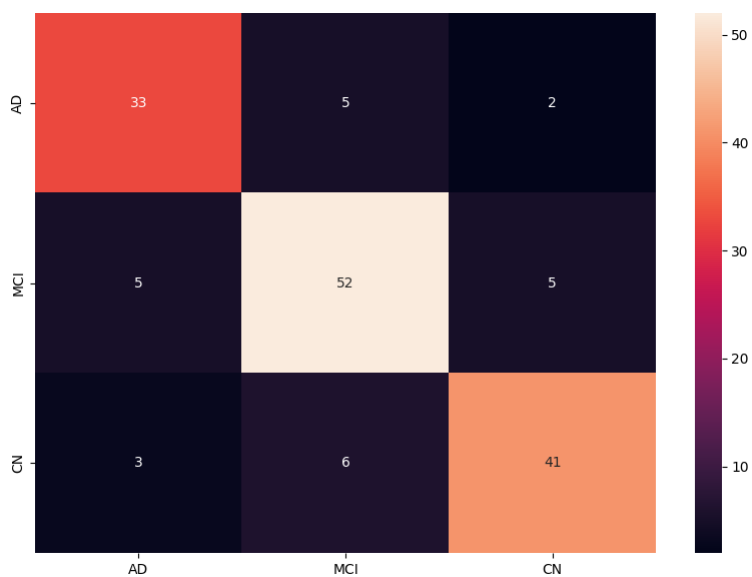


Σχήμα 6.21: 35 epochs training (accuracy)



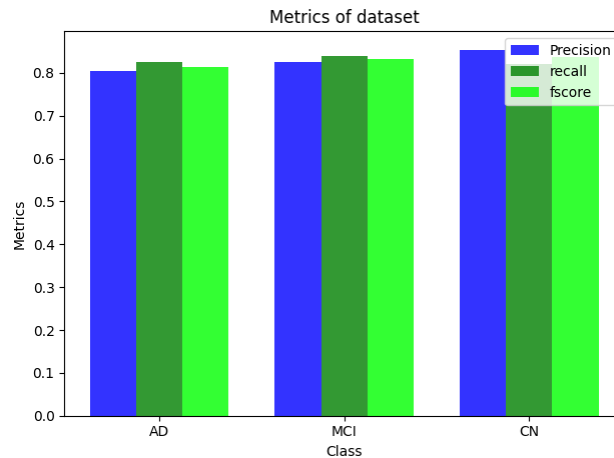
Σχήμα 6.22: 35 epochs training (loss)

Επιτύχαμε accuracy με τιμή **82,8%**. Ο πίνακας σύγχυσης μετά την κανονικοποίηση είναι ο παρακάτω:



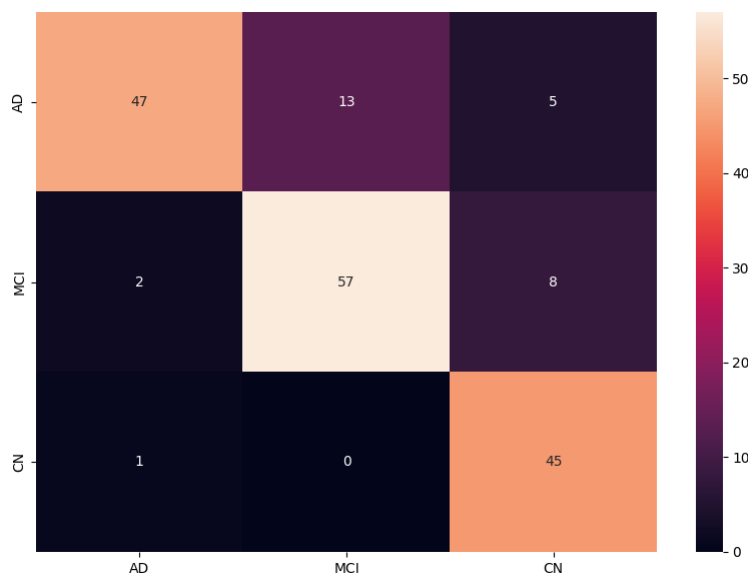
Σχήμα 6.23: Πίνακας σύγχυσης δεύτερης περίπτωσης

Το ραβδόγραμμα των μετρικών για κάθε κλάση είναι το παρακάτω:



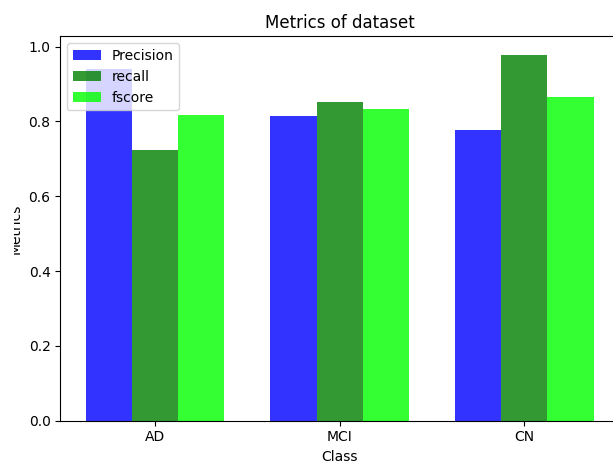
Σχήμα 6.24: Μετρικές δεύτερης περίπτωσης

**Τρίτη Περίπτωση** Τα σχήματα 6.17, 6.18 μας υποδεικνύουν ότι η εκπαίδευση πρέπει να σταματήσει στις 23 εποχές. Μετά την προσθήκη αυτών των δεδομένων στο σύνολο εκπαίδευσης-train set (δηλαδή μετά την μετατόπιση κάποιων εικόνων και προσθήκης θορύβου), το accuracy ανέβηκε στο **83,5%**. Ο πίνακας σύγχυσης είναι ο παρακάτω:



Σχήμα 6.25: Πίνακας σύγχυσης τρίτης περίπτωσης

Το ραβδόγραμμα των μετρικών για κάθε κλάση είναι το παρακάτω:



Σχήμα 6.26: Μετρικές τρίτης περίπτωσης



## Κεφάλαιο 7

# Συμπεράσματα και Μελλοντική Έρευνα

### Συμπεράσματα

- Παρατηρούμε πως το μοντέλο μας έχει πολύ καλή απόδοση και ξεχωρίζει με πολύ καλή ακρίβεια τους CN ανθρώπους τόσο από το ραβδόγραμμα όσο και από τον πίνακα σύγχησης
- Η επαύξηση δεδομένων στα προβλήματα ταξινόμησης εικόνων είναι μια πάρα πολύ χρήσιμη μέθοδος και η βιβλιοθήκη `scipy` την καθιστά εύκολα εφαρμόσιμη. Ακόμα και σε ένα σύνολο δεδομένων, το οποίο είναι σχετικά μικρό μπορεί να δώσει χρήσιμες λύσεις και να παρέχει πολυφωνία. Με αυτόν τον τρόπο αποφεύγεται και η υπερπροσαρμογή του μοντέλου μας, αφού εκπαιδεύεται για περισσότερα παραδείγματα, γεγονός που ευνοεί την γενίκευση του.
- Στα προβλήματα βαθειάς μάθησης η εφαρμογή του Dropout συνήθως σε ένα 50% μπορεί να αυξήσει τον αριθμό των εποχών προς εκπαίδευση του μοντέλου. Δηλαδή το πρόωρο σταμάτημα (early stopping) να ξεκινήσει σε μια πιο μετέπειτα εποχή (epoch).
- Δεν είναι απαραίτητο να επιλεγθούν πολλά συνελικτικά επίπεδα για την εκπαίδευση του μοντέλου. Όπως είδαμε και από τον πίνακα σύγχησης (confusion matrix) του σχήματος 6.5 κάτι τέτοιο μπορεί να οδηγήσει το δίκτυο μας στην πρόβλεψη της πιο συχνής κλάσης.

## Μελλοντική Έρευνα

Παρακάτω παραθέτω κάποιες προσωπικές προτάσεις για μελλοντική έρευνα για την ταξινόμηση του Alzheimer με συνελικτικά νευρωνικά δίκτυα:

- Σημαντικό ρόλο στην αποφυγή της υπερκπαίδευσης μπορεί να παίζει η επιλογή μαγνητικών τομογραφιών από ασθενείς σε μία μόνο επίσκεψη τους και όχι σε παρπάνω. Αυτό θα ήταν καλό για 2 λόγους:
  1. Αρχικά είναι πολύ πιθανό οι μαγνητικές τομογραφίες του ίδιου ασθενούς να μοιάζουν πολύ μεταξύ τους με αποτέλεσμα να έχουμε εικόνες σχεδόν ίδιες στο ίδιο σετ δεδομένων και να είναι πιο εύκολο να υπάρξει υπερεκπαίδευση.
  2. Επιπρόσθετα οι τρεις κλάσεις του προβλήματος μας δεν είναι και οι μοναδικές. Για παράδειγμα υπάρχουν κατηγορίες ασθενών όπως είναι οι cMCI και οι sMCI (stable MCI) οι οποίες είναι οι άνθρωποι που από MCI έγιναν AD σε κάποια επόμενη επίσκεψη τους και αυτοί που παρέμειναν MCI αντίστοιχα. Εάν όμως κάποιος ασθενής κατηγορίας MCI γίνει σε κάποια μετέπειτα επίσκεψη του AD λόγω αύξησης του global CDR (clinical dementia rating) εκείνος θα ανήκει ακόμα στην κατηγορία AD.
- Το συνελικτικό νευρωνικό δίκτυο μπορεί να λάβει ως δεδομένα και μαγνητικές που είναι Axial και coronal, αφού μέσω της εντολής της Python swapaxes μπορούν να 'μετατραπούν' σε Sagittal.
- Μεγαλύτερος πλουραλισμός στα δεδομένα θα επιφέρει καλύτερα και πιο εύκολα αποτελέσματα, καθώς είναι και κάτι που μας παρέχει το ADNI.
- Η εφαρμογή του Age correction. Εμείς στην παρούσα διπλωματική απλώς αφαιρέσαμε τις μαγνητικές των ανθρώπων που ήταν άνω των 87 για να μην μπερδέψει το δίκτυο μας. Ωστόσο υπάρχουν διάφοροι αλγόριθμοι που μπορεί να λάβει κάποιος υπόψιν που θεωρούν την ηλικία ως μεταβλητή που επηρεάζει το αποτέλεσμα (co-variate), αφού ο εγκέφαλος των μεγάλων σε ηλικία ανθρώπων τείνουν να μοιάσουν με εκείνο των ασθενών με Αλτςχάϊμερ.
- Εφαρμογή τρισδιάστατων συνελικτικών νευρωνικών δικτύων στο πρόβλημα τεσσάρων κλάσεων.
- Εφαρμογή τρισδιάστατων συνελικτικών νευρωνικών δικτύων στο πρόβλημα τριών κλάσεων μέσα από λειτουργικές μαγνητικές τομογραφίες (functional magnetic

resonance imaging-fmri) ή τομογραφίες εκπομπής ποζιτρονίων (positron emission tomography).

- Πειραματισμός με χρήση μεταφερόμενης μάθησης πάνω σε εικόνες με μορφή παρόμοια με αυτήν των δομικών μαγνητικών τομογραφιών (smri).



# Bibliography

- [1] Yiannopoulou Konstantina G and Papageorgiou Sokratis G. Current and future treatments in alzheimer disease: an update. *Journal of central nervous system disease*, 12:1179573520907397, 2020.
- [2] Andrew Kingston, Adelina Comas-Herrera, Carol Jagger, et al. Forecasting the care needs of the older population in england over the next 20 years: estimates from the population ageing and care simulation (pacsim) modelling study. *The Lancet Public Health*, 3(9):e447–e455, 2018.
- [3] Shahamat Hossein and Abadeh Mohammad Saniee. Brain mri analysis using a deep learning based evolutionary approach. *Neural Networks*, 126:218–234, 2020.
- [4] Cong Shan, Yao Xiaohui, Huang Zhi, Risacher Shannon L, Nho Kwangsik, Saykin Andrew J, Shen Li, UK Brain Expression Consortium, Alzheimer’s Disease Neuroimaging Initiative, et al. Volumetric gwas of medial temporal lobe structures identifies an erc1 locus using adni high-resolution t2-weighted mri data. *Neurobiology of Aging*, 95:81–93, 2020.
- [5] Li Yingjie, Zhang Liangliang, Bozoki Andrea, Zhu David C, Jongeun Cho, and Maiti Taps. Early prediction of alzheimer’s disease using longitudinal volumetric mri data from adni. *Health Services and Outcomes Research Methodology*, 20(1):13–39, 2020.
- [6] Gotaro Kojima, Ann Liljas, Steve Iliffe, and Kate Walters. Prevalence of frailty in mild to moderate alzheimer’s disease: a systematic review and meta-analysis. *Current Alzheimer Research*, 14(12):1256–1263, 2017.
- [7] Alberto Serrano-Pozo, Matthew P Frosch, Eliezer Masliah, and Bradley T Hyman. Neuropathological alterations in alzheimer disease. *Cold Spring Harbor perspectives in medicine*, 1(1):a006189, 2011.

- [8] Eric B Larson, Marie-Florence Shadlen, Li Wang, Wayne C McCormick, James D Bowen, Linda Teri, and Walter A Kukull. Survival after initial diagnosis of alzheimer disease. *Annals of internal medicine*, 140(7):501–509, 2004.
- [9] Ping Liu, Li Wu, Guoping Peng, Yuqiu Han, Ruiqi Tang, Jianping Ge, Lijiang Zhang, Longfei Jia, Siqing Yue, Kai Zhou, et al. Altered microbiomes distinguish alzheimer’s disease from amnestic mild cognitive impairment and health in a chinese cohort. *Brain, behavior, and immunity*, 80:633–643, 2019.
- [10] Alex J Mitchell and Mojtaba Shiri-Feshki. Rate of progression of mild cognitive impairment to dementia—meta-analysis of 41 robust inception cohort studies. *Acta psychiatrica scandinavica*, 119(4):252–265, 2009.
- [11] Nehmat Houssami, Robin M Turner, and Monica Morrow. Meta-analysis of pre-operative magnetic resonance imaging (mri) and surgical treatment for breast cancer. *Breast cancer research and treatment*, 165(2):273–283, 2017.
- [12] Andreas Benjamin Schmidt, S Berner, W Schimpf, C Müller, T Lickert, N Schwaderlapp, S Knecht, JG Skinner, A Dost, P Rovedo, et al. Liquid-state carbon-13 hyperpolarization generated in an mri system for fast imaging. *Nature communications*, 8(1):1–8, 2017.
- [13] Stavros I Dimitriadis, Dimitris Liparas, Magda N Tsolaki, Alzheimer’s Disease Neuroimaging Initiative, et al. Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological mri measures to discriminate among healthy elderly, mci, cmci and alzheimer’s disease patients: From the alzheimer’s disease neuroimaging initiative (adni) database. *Journal of neuroscience methods*, 302:14–23, 2018.
- [14] Mariana Pereira, Irene Fantini, Roberto Lotufo, and Leticia Rittner. An extended-2d cnn for multiclass alzheimer’s disease diagnosis through structural mri. In *Medical Imaging 2020: Computer-Aided Diagnosis*, volume 11314, page 113141V. International Society for Optics and Photonics, 2020.
- [15] Adrien Payan and Giovanni Montana. Predicting alzheimer’s disease: a neuroimaging study with 3d convolutional neural networks. *arXiv preprint arXiv:1502.02506*, 2015.
- [16] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

- [17] Irina Rish et al. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46, 2001.
- [18] Εμμανουήλ Παπαδάκης. Ανάλυση συναισθήματος από κείμενο με τεχνικές μηχανικής μάθησης και χρήση λεξικού. 2016.
- [19] Vladimir Vapnik. The support vector method of function estimation. In *Nonlinear modeling*, pages 55–85. Springer, 1998.
- [20] Raoof Gholami and Nikoo Fakhari. Support vector machine: principles, parameters, and applications. In *Handbook of Neural Computation*, pages 515–535. Elsevier, 2017.
- [21] Tasadduq Imam, Kai Ming Ting, and Joarder Kamruzzaman. z-svm: An svm for improved classification of imbalanced data. In *Australasian Joint Conference on Artificial Intelligence*, pages 264–273. Springer, 2006.
- [22] Bahzad Charbuty and Adnan Abdulazeez. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01):20–28, 2021.
- [23] Nathan Ensmenger. The multiple meanings of a flowchart. *Information & Culture*, 51(3):321–351, 2016.
- [24] B Chandra and P Paul Varghese. Fuzzifying gini index based decision trees. *Expert Systems with Applications*, 36(4):8549–8559, 2009.
- [25] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [26] Steven L Salzberg. C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993, 1994.
- [27] Bernhard Mehlig. Artificial neural networks. *arXiv e-prints*, pages arXiv–1901, 2019.
- [28] Qingshan Liu and Jun Wang. A one-layer recurrent neural network with a discontinuous hard-limiting activation function for quadratic programming. *IEEE transactions on neural networks*, 19(4):558–570, 2008.
- [29] P Sibi, S Allwyn Jones, and P Siddarth. Analysis of different activation functions using back propagation neural networks. *Journal of theoretical and applied information technology*, 47(3):1264–1268, 2013.

- [30] Yingying Wang, Yibin Li, Yong Song, and Xuewen Rong. The influence of the activation function in a convolution neural network model of facial expression recognition. *Applied Sciences*, 10(5):1897, 2020.
- [31] Binh Thai Pham, Manh Duc Nguyen, Kien-Trinh Thi Bui, Indra Prakash, Kamran Chapi, and Dieu Tien Bui. A novel artificial intelligence approach based on multi-layer perceptron neural network and biogeography-based optimization for predicting coefficient of consolidation of soil. *Catena*, 173:302–311, 2019.
- [32] S Abirami and P Chitra. Energy-efficient edge based real-time healthcare support system. In *Advances in Computers*, volume 117, pages 339–368. Elsevier, 2020.
- [33] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.
- [34] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier, 1992.
- [35] Yves Chauvin and David E Rumelhart. *Backpropagation: theory, architectures, and applications*. Psychology press, 2013.
- [36] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in neural information processing systems*, pages 3981–3989, 2016.
- [37] Martin Zinkevich, Markus Weimer, Alexander J Smola, and Lihong Li. Parallelized stochastic gradient descent. In *NIPS*, volume 4, page 4. Citeseer, 2010.
- [38] Kunihiro Fukushima. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, 1(2):119–130, 1988.
- [39] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [40] Jawad Nagi, Frederick Ducatelle, Gianni A Di Caro, Dan Cireşan, Ueli Meier, Alessandro Giusti, Farrukh Nagi, Jürgen Schmidhuber, and Luca Maria Gambardella. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 342–347. IEEE, 2011.



- [41] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.
- [42] Sing-Tze Bow. *Pattern recognition and image preprocessing*. CRC press, 2002.
- [43] Hyeok-June Jeong, Kyeong-Sik Park, and Young-Guk Ha. Image preprocessing for efficient training of yolo deep learning networks. In *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 635–637. IEEE, 2018.
- [44] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13001–13008, 2020.
- [45] Florent Ségonne, Anders M Dale, Evelina Busa, Maureen Glessner, David Salat, Horst K Hahn, and Bruce Fischl. A hybrid approach to the skull stripping problem in mri. *Neuroimage*, 22(3):1060–1075, 2004.
- [46] Juergen Dukart, Matthias L Schroeter, Karsten Mueller, and Alzheimer’s Disease Neuroimaging Initiative. Age correction in dementia–matching to a healthy brain. *PloS one*, 6(7):e22193, 2011.
- [47] Kirsi Juottonen, Mikko P Laakso, Kaarina Partanen, and Hilikka Soininen. Comparative mr analysis of the entorhinal cortex and hippocampus in diagnosing alzheimer disease. *American Journal of Neuroradiology*, 20(1):139–144, 1999.