Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Ηλεκτρολογων Μηχανικων και
Μηχανικων Υπολογιστων

Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Μοντελοποίηση Αβεβαιότητας σε Βαθιά Νευρωνικά Δίκτυα

Μια Σύγχρονη Προσέγγιση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΛΕΞΑΝΔΡΟΣ Κ. ΜΠΑΡΜΠΕΡΗΣ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2022

Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Ηλεκτρολογων Μηχανικων και
Μηχανικων Υπολογιστων

Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Μοντελοποίηση Αβεβαιότητας σε Βαθιά Νευρωνικά Δίκτυα

Μια Σύγχρονη Προσέγγιση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΛΕΞΑΝΔΡΟΣ Κ. ΜΠΑΡΜΠΕΡΗΣ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 26η Ιουνίου 2022.

.....	.....	.....
Στέφανος Κόλλιας Καθηγητής Ε.Μ.Π.	Ανδρέας-Γεώργιος Σταφυλοπάτης Καθηγητής Ε.Μ.Π.	Γιώργος Στάμου Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2022

Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων

Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δε μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(υπογραφή)

.....

Αλέξανδρος Μπαρμπέρης

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

Αθήνα, Ιούνιος 2022

Περίληψη

Το

Λέξεις κλειδιά

πρώτη λέξη

Abstruct

Cloud computing is the dominant approach to compute infrastructure,

Keywords

enwords

Ευχαριστίες

Για την παρούσα εργασία, που σηματοδοτεί την ολοκλήρωση μίας πορείας ετών, θα ήθελα να ευχαριστήσω

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

«Η επιτυχημένη δημιουργία [γενικευμένης] Τεχνητής Νοημοσύνης θα είναι το μεγαλύτερο γεγονός στην ανθρώπινη ιστορία. Δυστυχώς, ίσως είναι και το τελευταίο εάν δε μάθουμε πώς να αποφεύγουμε τα ρίσκα.» — Stephen Hawking

1.1 Επισκόπηση του Κόσμου της Τεχνητής Νοημοσύνης

Είναι πλέον γεγονός, η Τεχνητή Νοημοσύνη (Artificial Intelligence - AI) εντοπίζεται σε πολλές εκφάνσεις της καθημερινότητας των περισσότερων ανθρώπων [;]. Δεν αποτελεί απλά έναν ακόμα μοδάτο όρο που καταχράται στον χώρο της αγοραστικής (marketing). Εν αντιθέσει, διαδραματίζει καθοριστικό ρόλο στην εργασία μας, στη μετακίνησή μας και στην ψυχαγωγία μας. Για παράδειγμα, εντοπίζεται στις μηχανές αναζήτησης όπως η Google, στους ηλεκτρονικούς χάρτες πλοήγησης αλλά και στα συστήματα συστάσεων (Recomender systems) όπως αυτό του YouTube, του Twitter και του Netflix [;] που εξατομικεύουν το προβαλλόμενο περιεχόμενο στα ενδιαφέροντα του χρήστη.

Η επιρροή που έχει η Τεχνητή Νοημοσύνη είναι ακόμα πιο ευδιάκριτη αν τη μελετήσει κανείς υπό μια συλλογική σκοπιά. Για παράδειγμα, στον χώρο του επιχειρείν, η αξιοποίηση τεχνολογιών Τεχνητής Νοημοσύνης έχει αποδειχθεί ότι αυξάνει την επιχειρηματική αξία (business value) μέσα από τη βελτίωση της επίδοσης τόσο στο οικονομικό (financial), αγοραστικό (marketing) και διοικητικό (administrative) επίπεδο όσο και στο επίπεδο επιχειρηματικών διαδικασιών (business process) [;]. Χαρακτηριστικό παράδειγμα αποτελεί η χρήση των συστημάτων συστάσεων αφού με το να φιλτράρουν το περιεχόμενο και να παρουσιάζουν στον χρήστη μόνο αυτό που του είναι οικείο και θεμιτό, αυξάνουν τον βαθμό ενασχόλησής του με κίνδυνο την παγίδευσή του σε μια «φούσκα προκατειλημμένου φιλτραρίσματος» («biased filter bubble») [;]. Άλλωστε, όπως δηλώνει η ομάδα ανάπτυξης του εν λόγω συστήματος για λογαριασμό της συνδρομητικής υπηρεσίας streaming, Netflix, «Πιστεύουμε ότι αθροιστικά, η επίδραση της εξατομίκευσης και των συστάσεων μας εξοικονομεί ένα δισεκατομμύριο δολάρια τον χρόνο» [;].

Στον εργασιακό χώρο, πολλές δουλειές που περιλαμβάνουν επαναλαμβανόμενες, προβλέψιμες εργασίες κυρίως στον τομέα της βιομηχανίας και της γεωργίας αντικαθίστανται από αυτοματισμούς

Τεχνητής Νοημοσύνης (AI automations) εκτοπίζοντας έτσι τον άνθρωπο. Η μείωση των διαθέσιμων θέσεων εργασίας στους τομείς αυτούς δοκιμάζει τα όρια του κοινωνικού οικοδομήματος: τα «εκτοπισμένα» άτομα καλούνται να αποκτήσουν νέες δεξιότητες προκειμένου να βρουν απασχόληση στις πιο δημιουργικές (και συνάμα λιγότερο τυποποιήσιμες) θέσεις του εξελισσόμενου τομέα των υπηρεσιών [;]. Βέβαια, η μείωση των θέσεων εργασίας επαναλαμβανόμενης φύσης είναι μόνο η μια πλευρά του νομίσματος. Σύμφωνα με αναλύσεις, μέχρι την επόμενη δεκαετία οι εφαρμογές της Τεχνητής Νοημοσύνης εν δυνάμει θα αυξήσουν το παγκόσμιο Ακαθάριστο Εθνικό Προϊόν (Gross Domestic Product - GDP) κατά 26% (δεκαπέντε τρισεκατομμύρια δολάρια) [;]. Αυτό, με τη σειρά του, θα οδηγήσει στη δημιουργία πολλών νέων θέσεων εργασίας έτσι ώστε να μην παρατηρηθεί αύξηση στους δείκτες ανεργίας [;, ;].

Τέλος, δε θα μπορούσαμε να παραλείψουμε την επιρροή που έχει η Τεχνητή Νοημοσύνη στον χώρο της υγείας. Οι εφαρμογές είναι ατελείωτες: από συστήματα για πρόωρη διάγνωση ασθενειών μέχρι ρομποτικά συστήματα υποβοήθησης χειρουργείου [;]. Αξιοσημείωτη είναι επίσης η επιτυχημένη εφαρμογή της για την πρόβλεψη της τρισδιάστατης δομής των πρωτεϊνών [;] - ένα θέμα με σημαντικές προεκτάσεις που απασχολούσε την επιστημονική κοινότητα για 50 χρόνια. Παρόλα αυτά, μαζί με την προσπάθεια για αξιοποίηση των νέων τεχνολογιών στην κλινική πράξη προκύπτουν νέες προκλήσεις. Μια πρώτη δυσκολία είναι η ανάπτυξη συστημάτων Τεχνητής Νοημοσύνης που απαιτούν μεγάλο όγκο δεδομένων σε χώρους προβλημάτων όπου αυτά σπανίζουν (όπως για παράδειγμα, στην περίπτωση μιας ασυνήθιστης ασθένειας όπου ο αριθμός των ιατρικών υποθέσεων είναι ελάχιστος). Αυτή η δυσκολία εντείνεται αφενός λόγω της έλλειψης ασφαλών υποδομών για τη συλλογή ιατρικών δεδομένων [;] και αφετέρου λόγω της απόρρητης φύσης αυτών, κάτι που δυσχεραίνει τον ελεύθερο διαμοιρασμό τους. Μια τελευταία δυσκολία αποτελεί το γεγονός ότι πολλά συστήματα Τεχνητής Νοημοσύνης που έχουν αναπτυχθεί σε περιβάλλον εργαστηρίου (Lab setting) δεν παρέχουν αρκετά κίνητρα για μεταστροφή της καθιερωμένης κλινικής πράξης [;]. Για να επιτευχθεί κάτι τέτοιο, μεταξύ άλλων θα πρέπει τα συστήματα να αποδίδουν αποδεδειγμένα τόσο καλά όσο και το καταρτισμένο προσωπικό στο συγκεκριμένο πεδίο εφαρμογής τους και να παρέχουν πληροφορίες που θα τα καθιστούν περισσότερο έμπιστα π.χ. αιτιολογώντας την απόφασή τους (explainability), παρέχοντας μια μετρική αβεβαιότητας (uncertainty) ή δίνοντας τη δυνατότητα αλληλεπίδρασης [;].

Αντιλαμβανόμενοι το εύρος των εφαρμογών της Τεχνητής Νοημοσύνης, η εκτίμηση από την International Data Corporation - IDC πως οι Ευρωπαϊκές δαπάνες σε τέτοιες εφαρμογές θα έχουν σχεδόν τριπλασιαστεί μέσα στα επόμενα τρία χρόνια δε θα πρέπει να μας εκπλήσσει. Ωστόσο, με τη μεγάλη ισχύ έρχεται και μεγάλη ευθύνη. Είναι αλήθεια, η Τεχνητή Νοημοσύνη είναι ήδη εδώ και θα συνεχίζει να επιδρά όλο και εντονότερα στην καθημερινότητα μας και στην κοινωνία. Βέβαια, η σημερινή Τεχνητή Νοημοσύνη είναι εντελώς διαφορετική από αυτό που φαντάζονταν η κοινή γνώμη τις προηγούμενες δεκαετίες (σαφώς επηρεασμένη από ταινίες επιστημονικής φαντασίας όπως το Terminator). Θα λέγαμε, αντίθετα, πως εμφανίζεται περισσότερο με μια περιορισμένη μορφή στην εκάστοτε συγκεκριμένη εφαρμογή (narrow AI). Έτσι, ένα «εφυές» σύστημα για μια εργασία δεν μπορεί να «γενικεύσει» και να εφαρμοστεί σε άλλο χώρο προβλημάτων. Ούτε λόγος δε για αισθήματα και υπαρξιακή συνείδηση· αυτά (ακόμα) ανήκουν στην επιστημονική φαντασία. Αυτό όμως δε σημαίνει ότι η επιπόλαιη χρήση της Τεχνητής Νοημοσύνης δεν ελοχεύει κινδύνους. Σύμφωνα με το περιοδικό Spectrum της IEEE [;] προτού

επιτευχθεί Τεχνητή Νοημοσύνη επιπέδου ανθρώπου (human-like Artificial Inteligence) - αυτή στην οποία αναφέρεται ο Stephen Hawking - υπάρχουν ήδη πολλά σενάρια όπου εφαρμογές της μπορούν να αποβούν μοιραίες. Ενδεικτικά, ένα από αυτά είναι τα deepfakes - ψεύτικα πολυμέσα βίντεο και εικόνας κατασκευασμένα από εφαρμογές Τεχνητής Νοημοσύνης - έχουν υπονομεύσει την εμπιστοσύνη στα συστήματα πληροφόρησης. Επιπρόσθετα, ένα ακόμα καταστροφικό σενάριο σχετίζεται με την ιδιωτικότητα (privacy) και την ελεύθερη βούληση (free will). Με την παραχώρηση ευαίσθητων δεδομένων σε επιχειρήσεις και κυβερνήσεις τους παρέχουμε τη δυνατότητα να μας εποπτεύουν ακόμα και να μας χειραγωγούν. Ένα τελευταίο σενάριο για το οποίο διαδραματίζουν άμεσο ρόλο τα κοινωνικά δίκτυα είναι αυτό του μειωμένου διαστήματος προσοχής (short attention span) ως απόρροια της εκμετάλλευσης του μηχανισμού επιβράβευσης του εγκεφάλου ώστε οι χρήστες να εθίζονται σε αυτά. Το περιοδικό καλεί τον αναγνώστη να αναλογιστεί τις συνέπειες της συνεχόμενης βελτίωσης των μηχανισμών που μας καθηλώνουν από τη νέα τεχνολογία. Συμπερασματικά, η Τεχνητή Νοημοσύνη αν και δεν προσομοιάζει την ανθρώπινη νοημοσύνη δεν παύει να αποτελεί μια πολύ ισχυρή τεχνολογία που μπορεί να αποβεί είτε σωτήρια είτε μοιραία ανάλογα με τον τρόπο αξιοποίησής της.

Είναι λοιπόν απαραίτητη η εξασφάλιση της συνετής χρήσης αυτών των τεχνολογιών μέσω μιας σειράς κανονισμών. Στην Ευρωπαϊκή Ένωση, μια σειρά από διατάξεις επιχειρούν να θέσουν ένα νομοθετικό πλαίσιο ώστε να ωθήσουν στην αξιοποίηση της Τεχνητής Νοημοσύνης διασφαλίζοντας παράλληλα την ασφάλεια των θεμελιωδών δικαιωμάτων [;]. Άλλωστε, σύμφωνα με την von der Lein [;], «Η Τεχνητή Νοημοσύνη πρέπει να εξυπηρετεί τους ανθρώπους και συνεπώς, πρέπει πάντα να συμμορφώνεται με τα δικαιώματά τους. Αυτός είναι ο λόγος που ένα άτομο πρέπει πάντα να έχει τον έλεγχο στην περίπτωση κρίσιμων αποφάσεων[...] Εφαρμογές της Τεχνητής Νοημοσύνης που μπορεί να παρέμβουν στα ανθρώπινα δικαιώματα θα πρέπει να ελέγχονται και να πιστοποιούνται πριν φτάσουν στην Ευρωπαϊκή αγορά. » Αν και οι αυστηροί διακανονισμοί καθυστερούν τη μετάβαση εφαρμογών Τεχνητής Νοημοσύνης από το εργαστήριο στην αγορά, εξασφαλίζουν την ασφάλειά τους συμβάλλοντας στην αξιοπιστία τους.

Με την παραπάνω σύντομη εισαγωγή καλύψαμε εμπεριστατωμένα πολλά από τα μη τεχνικά θέματα που σχετίζονται με την Τεχνητή Νοημοσύνη (Artificial Inteligence). Αναλυτικότερα, αρχίσαμε από παραδείγματα εντοπισμού της στην καθημερινή ζωή σε ατομικό και σε συλλογικό επίπεδο με τα οποία αντιληφθήκαμε τη σημασία της. Συνεχίσαμε με τους κινδύνους που ελοχεύει η απερίσκεπτη εφαρμογή της σε συγκεκριμένες εργασίες επισημαίνοντας ταυτόχρονα ότι η τεχνητή νοημοσύνη απέχει από την ανθρώπινη. Κλείσαμε, με μερικές από τις προσπάθειες που γίνονται σε Ευρωπαϊκό επίπεδο για την αποφυγή αυτών των κινδύνων. Πολλά από τα προαναφερθέντα στοιχεία πιθανότατα να είναι ήδη γνωστά σε έναν έμπειρο αναγνώστη. Εντούτοις, εξυπηρετούν σε μια ομαλή εισαγωγή για τον αρχάριο και σε μια υπενθύμιση για τον έμπειρο αναγνώστη του κόσμου της Τεχνητής Νοημοσύνης. Στην επόμενη υπο-ενότητα αυτού του κεφαλαίου θα παρουσιάσουμε μια ιστορική αναδρομή της τεχνητής νοημοσύνης. Έτσι, ο αναγνώστης θα κατανοήσει σε βάθος την έννοια γύρω από την οποία εκτυλίσσεται η παρούσα διπλωματική προτού εισαχθεί στο συγκεκριμένο τεχνικό της θέμα. Έπειτα, περιγράφεται το κίνητρο που με ωθούσε καθ'όλη τη διάρκεια συγγραφής του έργου. Τέλος, αναφερόμαστε στην τεχνική συνεισφορά της παρούσας εργασίας και στην οργάνωση του τόμου.

1.2 Ιστορική Αναδρομή Τεχνητής Νοημοσύνης

Οι εφευρέτες οραματίζονται εδώ και χιλιετίες τη δημιουργία μηχανών που σκέφτονται. Ήδη, γύρω στο 700 π.Χ. αναφέρεται από τον Ησίοδο ο Τάλος: ο μυθικός χάλκινος γίγαντας φτιαγμένος από τον Ήφαιστο με αποστολή να προστατεύσει το νησί της Κρήτης από τους επιδρομείς [;]. Παρόμοια παραδείγματα αποτελούν αυτά της Πανδώρας και της Γαλάτειας. Η μακρόβια επιθυμία για απομίμηση της νοημοσύνης μαρτυρά την αξία που της δίνει ο άνθρωπος. Γεγονός απόλυτα δικαιολογημένο αφού η νοημοσύνη - η νοητική ικανότητα που μας επιτρέπει να σκεπτόμαστε λογικά, να επιλύουμε προβλήματα και να μαθαίνουμε - έχει συμβάλει σημαντικά στην επιβίωση του είδους από τον διαειδικό ανταγωνισμό (interspecific competition)[1]. Σε τελική ανάλυση, ο όρος homo sapiens - άνθρωπος ο σοφός - οφείλεται στη σημασία που έχει η νοημοσύνη στη ζωή μας.

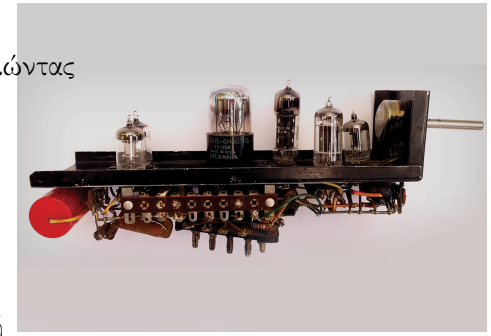
Το πρώτο ευρέως αναγνωρισμένο έργο προς την επίτευξη Τεχνητής Νοημοσύνης είναι αυτό της μαθηματικής μοντελοποίησης της λειτουργίας ενός νευρώνα από τους Warren McCulloch και Walter Pitts (1943) [;]. Αναλυτικότερα, βασιζόμενοι στην υπόθεση ότι η κατάσταση λειτουργίας ενός νευρώνα είναι δυαδική («all-or-none») αναπαρέστησαν κάθε νευρώνα ενός δικτύου ως μια πρόταση (proposition) της προτασιακής λογικής (propositional logic). Όπως περιγράφουν, η διέγερση (excitation) του μοντέλου ενός νευρώνα είναι ταυτόσημη με το να είναι η πρόταση του νευρώνα θετική, κάτι που εξαρτάται από την κατάσταση των γειτονικών νευρώνων. Όσο περισσότεροι, διεγερτικά διασυνδεδεμένοι (excitatory connected) προσυναπτικοί νευρώνες (presynaptic neurons) είναι ενεργοποιημένοι τόσο πιο πιθανή είναι η ενεργοποίηση του μετασυναπτικού νευρώνα (postsynaptic neuron). Μεταξύ άλλων, απέδειξαν ότι κάθε συνάρτηση που μπορεί να υπολογιστεί από μια μηχανή Turing μπορεί να υπολογιστεί και από ένα δίκτυο από διασυνδεδεμένους τεχνητούς νευρώνες ύστερα από την κατάλληλη παραμετροποίησή του. Το έργο των Warren McCulloch και Walter Pitts ήταν πρωτοπόρο για την εποχή του αφού έθεσε τις βάσεις όχι μόνο για την σημερινή Τεχνητή Νοημοσύνη αλλά και για την Υπολογιστική Νευρωεπιστήμη (Computational Neuroscience). Ωστόσο, οι συγγραφείς δεν παρουσίασαν κανέναν αλγόριθμο για την αλλαγή της τοπολογίας και των παραμέτρων του τεχνητού νευρωνικού δικτύου αν και φαίνεται πως αναγνώριζαν τη σημασία του για τη μάθηση.

Στο βιβλίο του [;] ο D. Hebb το 1949 επιδίωξε να ενώσει τις αποκλίνουσες θεωρίες της ψυχολογίας και της νευροεπιστήμης θέτοντας κοινές βάσεις για την ερμηνεία της ανθρώπινης συμπεριφοράς. Προηγούμενες θεωρίες απέφευγαν να δώσουν εξήγηση στις διεργασίες του εγκεφάλου ως μεσάζοντα μεταξύ του αισθητηριακού ερεθίσματος (sensory stimuli) και της πιθανής, καθυστερημένης απόκρισης. Αντίθετα, κατέφευγαν στη φιλοσοφία για την ανάλυση των χαρακτηριστικών της ανθρώπινης συμπεριφοράς όπως η προσοχή (attention), το ενδιαφέρον (interest) και η «προσδοκία» (expectancy). Ο Hebb όμως, εργάστηκε στο να αποδείξει ότι η ανθρώπινη συμπεριφορά μπορεί

[1]Για την ακρίβεια, ενώ σύμφωνα με τη Δαρβινική θεώρηση η αφηρημένη νοημοσύνη μπορεί να προκύψει άμεσα από τη θεωρία της εξέλιξης των ειδών, νεότερες έρευνες το διαψεύδουν αφού το χαρακτηριστικό της αφηρημένης σκέψης ήταν αχρείαστο στην πραγματιστική παλαιολιθική εποχή. Στην προσπάθειά τους να ερμηνεύσουν την εμφάνιση νοημοσύνης στους ανθρώπους ορίζουν τον όρο «διανοητική βιοθέση» (cognitive niche) για να περιγράψουν όλα τα ζωολογικά ασυνήθιστα χαρακτηριστικά (zoologically unusual traits) που εμφανίζει ο άνθρωπος με τα κυριότερα να είναι η κοινωνικότητα και η λογική αιτίου - αποτελέσματος (cause-and-effect reasoning) [;,;]. Υποστηρίζουν λοιπόν, ότι η εμφάνιση της διανοητικής βιοθέσης αποτέλεσε τον καταλύτη για την εξέλιξη της ανθρώπινης νοημοσύνης [;]

να γίνει κατανοητή υπό το πρίσμα της φυσιολογίας ("[expectancy] can be a physiologically intelligible process"). Φαινομενικά, ενώ το έργο του απασχολεί μόνο την επιστήμη της Νευροφυσιολογίας (Neurophysiology) συνεισέφερε σημαντικά στο κίνημα του διασυνδετισμού (Connectionism) - το κίνημα μελέτης των διανοητικών διεργασιών με τη χρήση τεχνητών νευρωνικών δικτύων. Για παράδειγμα, η θεώρηση της διαδικασίας μάθησης ως της επαναλαμβανόμενης, ταυτόχρονης πυροδότησης γειτονικών νευρώνων με αποτέλεσμα [την ενδυνάμωση των δεσμών και] τη διαμόρφωση νευρικών συστάδων (cell assemblies) είναι η λογική πίσω από πολλούς αλγορίθμους μάθησης τεχνητών νευρωνικών δικτύων. Επίσης, ένα στοιχείο που θα μας φανεί χρήσιμο στη συνέχεια είναι η παρατήρησή του ότι η κίνηση των ματιών δεν είναι τυχαία αλλά σχετίζεται με τη διαδικασία αντίληψής των θωρούμενων αντικειμένων. Τα παραπάνω δύο έργα έθεσαν το απαραίτητο θεωρητικό υπόβαθρο εμπνέοντας τους ερευνητές στην πειραματική υλοποίηση της Τεχνητής Νοημοσύνης.

Αρκετές ήταν οι απόπειρες δημιουργίας Τεχνητής Νοημοσύνης. Ο πρώτος υπολογιστής τεχνητών νευρωνικών δικτύων ονομάζονταν SNARC (Stochastic Neural Analog Reinforcement Calculator) και κατασκευάστηκε από τους Marvin Minsky και Dean Edmonds το 1950 [;]. Χρησιμοποιώντας 3000 λυχνίες κενού και έναν μηχανισμό αυτόματου πιλότου εξομοίωσαν τη λειτουργία 40 διασυνδεδεμένων νευρώνων. Χρησιμοποιώντας έναν απλό μηχανισμό επιβράβευσης τα «βάρη» του δικτύου - υπο τη μορφή ποτενσιόμετρων - προσαρμόζονταν στο πρόβλημα του λαβυρίνθου στο οποίο δοκιμάστηκε. Ακόμα ένα παράδειγμα τεχνητής νοημοσύνης μπορεί να θεωρηθεί το πρόγραμμα του Christopher Strachey στον υπολογιστή Manchester Mark 1 [;] που αργότερα θεωρήθηκε το πρώτο βιντεοπαιχνίδι. Ήταν ένα παιχνίδι ντάμας που πιθανώς χρησιμοποιούσε κάποιον μη πλήρη (incomplete) αλγόριθμο αναζήτησης στον χώρο των επιτρεπτών ακολουθιών κινήσεων (action sequences). Παρόλα αυτά, η δυνατότητά του να ανταγωνίζεται αποτελεσματικά τον άνθρωπο οδήγησε στη θεώρησή του ως Τεχνητή Νοημοσύνη. Άλλωστε, η σύγχυση για το θέμα ήταν ακόμα μεγαλύτερη την εποχή εκείνη.



Σχήμα 1.1: Ένας από τους 40 νευρώνες του SNARC. Χορηγία της κα. Margaret Minsky [;]

Ο αρχικός ενθουσιασμός για το επιστημονικό πεδίο τράβηξε τα βλέμματα πολλών επιφανών ερευνητών της εποχής. Ο πατέρας της Επιστήμης των Υπολογιστών (Computer Science) και της Τεχνητής Νοημοσύνης, Alan Turing, στην προσπάθειά του να διασαφηνίσει το ερώτημα του «αν οι μηχανές σκέφτονται» επινόησε το επονομαζόμενο Turing Test. Σύμφωνα με τη δημοσίευσή του [;] το 1950, πρόκειται για μια δοκιμασία που εμπλέκει έναν «ανακριτή» ο οποίος διατυπώνει γραπτές ερωτήσεις σε δύο «μάρτυρες»: έναν άνθρωπο και μια μηχανή. Η δοκιμασία θεωρείται επιτυχής όταν ο ανακριτής - χωρίς να έχει οπτική επαφή με τους «μάρτυρες» - δεν μπορεί να ξεχωρίσει τον άνθρωπο από τη μηχανή. Στην ίδια δημοσίευση τόνισε τη σημασία της μάθησης για την ανάπτυξη της Τεχνητής Νοημοσύνης. Υποστήριζε ότι αντί να επιχειρείται η εξονυχιστική συγγραφή ενός προγράμματος που θα μοιάζει με τη σκέψη ενός ώριμου ενήλικα (με αμέτρητες προγραμματισμένες εντολές) είναι προτιμότερη και ταχύτερη η προσομοίωση της νοημοσύνης

ενός παιδιού που μέσα από μηχανισμούς εκπαίδευσης, έμμεσα, αποκτά ώριμη σκέψη. Επίσης, εναπόθεσε τους σπόρους για τους γενετικούς αλγορίθμους, ενώ σε επόμενη δημοσίευσή του [;] μελέτησε τους τρόπους με τους οποίους μια μηχανή με νοημοσύνη θα μπορούσε να λειτουργεί. Ένα ακόμα δημοφιλές όνομα, ο John von Neumann συνεισέφερε στον χώρο αναπτύσσοντας τα «τεχνητά αυτόματα» (artificial automata) [;] ενώ η συμβολή του πιθανώς θα ήταν ακόμα μεγαλύτερη αν προλάβαινε να ολοκληρώσει το βιβλίο του «Ο Υπολογιστής και το Μυαλό» (The Computer and the Brain). Μια τελευταία απόδειξη της προσοχής που έλαβε η Τεχνητή Νοημοσύνη διαφαίνεται στα δέκα μέλη σχετικού σεμιναρίου (workshop) που έλαβε χώρα το καλοκαίρι του 1955 στο Dartmouth College [;]. Ίσως το ποιο σημαντικό πόρισμα αυτής της συνάθροισης ήταν η ανάπτυξη του Logic Theorist από τους Allen Newell και Herbert Simon, ενός συστήματος για την απόδειξη θεωρημάτων στα μαθηματικά.

Η δεκαετία που ακολούθησε χαρακτηρίζεται από έντονη αισιοδοξία για τις δυνατότητες της Τεχνητής Νοημοσύνης: γενναιόδωρες επενδύσεις σε ερευνητικά προγράμματα ενθάρρυναν τη δημιουργία ποικίλων προγραμμάτων που υποδείκνυαν κάποια μορφή νοημοσύνης. Πιο συγκεκριμένα, αν εξαιρέσουμε την εξέχουσα δουλειά του Arthur Samuel όπου ανέπτυξε ένα παιχνίδι ντάμας χρησιμοποιώντας ενισχυτική μάθηση (Reinforcement Learning), οι περισσότερες προσπάθειες εστίασαν στον χώρο της μίμησης της ανθρώπινης συλλογιστικής (reasoning). Η ιδέα είναι ότι με μια τυπική γλώσσα (formal language) για την αναπαράσταση της γνώσης (knowledge representation) στον υπολογιστή μαζί με την εφαρμογή απλών κανόνων λογικής συμπερασματολογίας (logical inference) σε αυτή καθιστούν δυνατή την εξαγωγή πορισμάτων. Καθοριστικό ρόλο σε αυτή τη «σχολή» είχε η - αναχρονιστική - υπόθεση ότι η νοημοσύνη είναι άρρηκτα συνδεδεμένη με τη δυνατότητα χειρισμού συμβόλων οργανωμένων σε δομές δεδομένων (physical symbol system hypothesis).

Σε αυτήν την κατεύθυνση, δηλαδή της συμβολικής Τεχνητής Νοημοσύνης (Symbolic Artificial Inteligence), εργάστηκαν αρκετοί επιστήμονες της εποχής. Για παράδειγμα, οι δύο ερευνητές πίσω από το Logic Theorist επινόησαν το General Problem Solver. Πρόκειται ουσιαστικά για έναν αλγόριθμο ο οποίος δέχεται σαν είσοδο μια τυποποιημένη περιγραφή του προβλήματος και το επιλύει ακολουθώντας μια στρατηγική ευρετικής αναζήτησης (heuristic search) της λύσης [;]. Στο ίδιο μήκος κύματος εργάστηκε και ο John McCarthy. Εκτός από το ότι ανέπτυξε τη γλώσσα προγραμματισμού Lisp, ειδικά φτιαγμένη για εφαρμογές Τεχνητής Νοημοσύνης, εξέλιξε το πεδίο με το δημοσίευμά του «Programs with Common Sense» (1958) στο οποίο περιέγραφε το Advice Taker. Αυτό ήταν ένα πρόγραμμα για την επίλυση προβλημάτων μέσω της εφαρμογής «κοινής λογικής» σε προτάσεις διατυπωμένες σε τυπική γλώσσα. Για παράδειγμα, δοθέντος μιας σειράς υποθέσεων σχετικά με το περιβάλλον του προβλήματος διατυπωμένων σε τυπική γλώσσα (π.χ. «Εγώ είμαι στο γραφείο.», «Θέλω να πάω αεροδρόμιο.» κτλ.) ο αλγόριθμος εξήγαγε ένα πλάνο με τα βήματα που πρέπει να ακολουθηθούν για τη μετάβαση στο αεροδρόμιο [;]. Το έργο του συνέχισε ο J. A. Robinson όπου και επινόησε μια πλήρη μέθοδο επίλυσης (complete resolution method) για προβλήματα εκφρασμένα σε λογική πρώτης τάξης. Οι εφαρμογές του ήταν πολλές: από συστήματα μαθηματικού λογισμού (James Slagle's SAINT program [;] και Daniel Bobrow's STUDENT program) μέχρι εφαρμογές ερωταπαντήσεων (Cordell Green's question-answering and planning systems) και ρομποτικής (Shakey Robotics Project). Τέλος, πολλές εφαρμογές της Συμβολικής Τεχνητής Νοημοσύνης αναπτύχθηκαν για το «παιχνίδι» blocks world:

ένα περιβάλλον αποτελούμενο από τουβλάκια που αποσκοπούσε στον πειραματισμό συστημάτων αναπαράστασης γνώσης και συλλογιστικής [;].

Την ίδια εποχή, ειδικά για τον χώρο των νευρωνικών δικτύων υπήρξε σημαντική πρόοδος με τα έργα Perceptron και ADALINE. Το πρώτο συγγράφηκε από τον F. Rosenblatt το 1958 και αποτέλεσε το πρώτο μοντέλο νευρωνικού δικτύου με δυνατότητα επιβλεπόμενης μάθησης supervised learning. Πιο συγκεκριμένα, το Perceptron είναι ένας ταξινομητής γραμμικά διαχωρίσιμων προτύπων με ένα μεμονωμένο τεχνητό νευρώνα του οποίου οι ελεύθεροι παράμετροι - τα προσυναπτικά βάρη (presynaptic weights) και η πόλωση (bias) - προσαρμόζονται στα δεδομένα εισόδου σύμφωνα με έναν αλγόριθμο μάθησης (perceptron rule) [;]. Σε μια εκτενή παρουσίαση του έργου [;], ο Rosenblatt βασίστηκε στη θεωρία του D. Hebb και την επέκτεινε προτείνοντας ένα μοντέλο (το Perceptron) με το οποίο η συμπεριφορά (καμπύλη εκμάθησης) μπορεί να προβλεφθεί από τη νευροφυσιολογία του συστήματος (τα συναπτικά βάρη). Παρόμοιο ήταν και το έργο ADALINE (Adaptive Linear Neuron) του B. Widrow στο οποίο περιγράφεται και πάλι ένας αλγόριθμος μάθησης για την προσαρμογή των βαρών. Αυτή τη φορά όμως, είναι ο (γνωστός) αλγόριθμος στοχαστικής καθόδου κλίσης stochastic gradient descent που χρησιμοποιείται ακόμα και σήμερα στον αλγόριθμο γραμμικής παλινδρόμησης (linear regression). Μια ακόμα αξιοσημείωτη διαφορά έγκειται στη συνάρτηση ενεργοποίησης όπου ενώ στο πρώτο έργο είναι η βηματική συνάρτηση (step function), στο δεύτερο έργο είναι η γραμμική συνάρτηση (linear activation function - identity function) που καθιστά τον αλγόριθμο κατάλληλο για την πρόβλεψη πραγματικών τιμών [;, ;]. Συνεπώς, ενώ το πρώτο έργο αποτελεί όπως προαναφέρθηκε έναν αλγόριθμο ταξινόμησης, το δεύτερο έργο ανήκει στην κατηγορία αλγορίθμων γραμμικής παλινδρόμησης. Βέβαια, μολονότι και τα δύο έργα είχαν καθοριστική σημασία στην εξέλιξη της Τεχνητής Νοημοσύνης με τη μορφή που τη συναντάμε σήμερα, όπως θα δούμε στη συνέχεια, η έντονη κριτική που ακολούθησε τα επισκίασε για μια ολόκληρη δεκαετία.

Γύρω στο 1970, το επιστημονικό πεδίο της Τεχνητής Νοημοσύνης διήλθε μια εποχή «χειμώνα» (AI winter). Η χρηματοδότηση ερευνητικών προγραμμάτων πάγωσε και έτσι το ενδιαφέρον στράφηκε αλλού. Κοιτώντας πίσω, είναι εύκολο να εντοπίσει κανείς τα αίτια αυτού του «χειμώνα». Καταρχάς, ένας λόγος ήταν (και είναι) το ελλιπές επιστημονικό υπόβαθρο σε ό,τι αφορά την ανθρώπινη νοημοσύνη [;]. Αναμενόμενο, αφού για μια επιτυχημένη μίμηση της ανθρώπινης νοημοσύνης απαιτείται πρώτα η κατανόησή της. Βέβαια, ίσως ο σημαντικότερος λόγος ήταν η απογοήτευση που προκλήθηκε όταν φιλόδοξες υποσχέσεις για τις δυνατότητες της Τεχνητής Νοημοσύνης στο εγγύς μέλλον δεν μπόρεσαν να ικανοποιηθούν. Όπως αποδείχθηκε, η μετάβαση της Τεχνητής Νοημοσύνης από εφαρμογές παιδικών κόσμων όπως το blocks world σε πραγματικά προβλήματα δεν ήταν απλώς ζήτημα γραμμικής αύξησης της υπολογιστικής δύναμης. Για παράδειγμα, στην περίπτωση της συμβολικής Τεχνητής Νοημοσύνης, με την ωρίμανσή της θεωρίας πολυπλοκότητας (computational complexity αναδείχθηκε το θέμα της συνδυαστικής έκρηξης (combinatorial explosion) αποκαλύπτοντας έτσι τη δυσεπίλυτη (intractable) φύση πολλών προβλημάτων του αληθινού κόσμου. Αντίστοιχα εμπόδια έκαναν την εμφάνισή τους στον χώρο των Νευρωνικών Δικτύων. Το σημαντικότερο ήταν αυτό της αδυναμίας ενός μεμονωμένου Perceptron με δύο εισόδους να αναπαραστήσει πολλές συναρτήσεις όπως τη συνάρτηση XOR [;] - περιγράφηκε από το βιβλίο Perceptrons των M. Minsky και S. Papert το 1969. Συνολικά, αν και ορισμένα από τα ανωτέρω αίτια είναι ακόμα σε ισχύ, το ενδιαφέρον σύντομα αναζωπυρώθηκε.

Παρά τις ανωτέρω αδυναμίες της Τεχνητής Νοημοσύνης την εποχή εκείνη, αυτό δεν εμπόδισε την αξιοποίησή της σε εξειδικευμένες εφαρμογές. Πιο συγκεκριμένα, τις δεκαετίες του 1970 και 1980 αναπτύχθηκαν τα «έμπειρα συστήματα» (expert systems). Πρόκειται για προγράμματα που εφαρμόζουν κανόνες συλλογιστικής σε εξειδικευμένη βάση γνώσης (domain-specific knowledge base) μιμούμενοι τη διαδικασία λήψης αποφάσεων ενός εμπειρογνώμονα. Η εξειδικευμένη βάση γνώσης περιόριζε σημαντικά τον χώρο αναζήτησης λύσεων (search space) έτσι ώστε η συνδυαστική έκρηξη να μην αποτελεί πρόβλημα. Αυτό, επέτρεψε να αναπτυχθούν πολλά έμπειρα συστήματα για εμπορική χρήση - κυρίως στον χώρο της υγείας - αποδεικνύοντας για πρώτη φορά έμπρακτα τα οφέλη της Τεχνητής Νοημοσύνης. Ενδεικτικά, δύο δημοφιλή παραδείγματα είναι το MYCIN και το INTERNIST. Το MYCIN αποσκοπούσε στη διάγνωση βακτηριακών μολύνσεων μέσω ενός αλγορίθμου συλλογιστικής που μοντελοποιούσε την αβεβαιότητα των λογικών υποθέσεων και συμπερασμάτων. Το INTERNIST από την άλλη βοηθούσε στη διάγνωση ασθενειών μετά από την περιγραφή των εκδηλούμενων συμπτωμάτων [;]. Αν και τα έμπειρα συστήματα ανανέωσαν το ενδιαφέρον για την Τεχνητή Νοημοσύνη, αυτό δε διήρκησε πολύ λόγω προβλημάτων που εμφάνιζαν με το κυριότερο να είναι η έλλειψη «κοινής λογικής» [;].

Η έρευνα στον χώρο της Τεχνητής Νοημοσύνης αποκαταστάθηκε στα συνήθη υψηλά επίπεδα σε σύντομο χρονικό διάστημα. Αυτό μπορεί σε μεγάλο βαθμό να αποδοθεί σε ένα μεμονομένο έργο· το Parallel Distributed Processing που συγγράφηκε από τους David E. Rumelhart et al. και δημοσιεύτηκε το 1986 [;]. Οι συγγραφείς, μεταξύ των οποίων και ο Geoff Hinton εμπνεόμενοι από τα παλαιότερα έργα πάνω στη γνωστική νευροεπιστήμη (cognitive neuroscience) έστρεψαν την έρευνα του χώρου από πειραματικές «αλχημείες» (π.χ. αυτή της συμβολικής λογικής) σε μια πιο επίσημη, φορμαλιστική διαδικασία βασιζόμενη λιγότερο στη φιλοσοφία και περισσότερο στις θετικές επιστήμες[2]. Με αυτόν τον τρόπο, η σχολαστική συγγραφή αναρίθμητων κανόνων προτασιακής λογικής για τη δημιουργία βάσεων γνώσης εγκαταλείφθηκε, μαζί της και η θεωρία της συμβολικής Τεχνητής Νοημοσύνης. Άλλωστε, η τεχνολογία των έμπειρων συστημάτων είχε παρακμάσει αφού όπως φάνηκε από την απουσία «κοινής λογικής» σε αυτά, ήταν εξαιρετικά περιοριστική η χρήση προτασιακής λογικής για την περιγραφή του πραγματικού, αβέβαιου κόσμου [;, ;].

Τη δεκαετία του 1980 τη θέση της συμβολικής Τεχνητής Νοημοσύνης πήρε το κίνημα του κονεκτιβισμού (connectionist movement). Όπως θα δούμε, αυτό συνέβαλλε καθοριστικά στη διαμόρφωση του σημερινού κλάδου των νευρωνικών δικτύων. Τυπικά, ο κονεκτιβισμός είναι το κίνημα της γνωστικής επιστήμης (cognitive science) που επιχειρεί να εξηγήσει τις διανοητικές διεργασίες με τη χρήση ενός δικτύου με βάρη (weighted network) που διασυνδέει απλές μονάδες επεξεργασίες [;]. Σε αυτήν τη θεωρία καταπιάστηκαν και οι συγγραφείς του έργου Parallel Distributed Processing [;] οι οποίοι εδραίωσαν τις ιδέες που σήμερα θεμελιώνουν τη θεωρία των νευρωνικών δικτύων. Η πρώτη σημαντική ιδέα που περιγράφεται λεπτομερώς στο έργο είναι η κατανεμημένη αναπαράσταση (distributed representation) σύμφωνα με την οποία κάθε είσοδος στο σύστημα αναπαρίσταται από πολλά χαρακτηριστικά κατανεμημένα στο δίκτυο και ανάποδα, δηλαδή κάθε μεμονωμένο χαρακτηριστικό μπορεί να αποτελεί μέρος της περιγραφής πολλών, ετερογενών εισόδων. Μια ακόμα σημαντική ιδέα είναι αυτή της μηχανικής μάθησης (machine learning) με την οποία η επίδοση ενός συστήματος βελτιώνεται (μαθαίνει) από την εμπειρία.

[2]Στη βιβλιογραφία αυτό το γεγονός αναφέρεται σαν «η νίκη των καθαρών» (victory of the neats) [;].

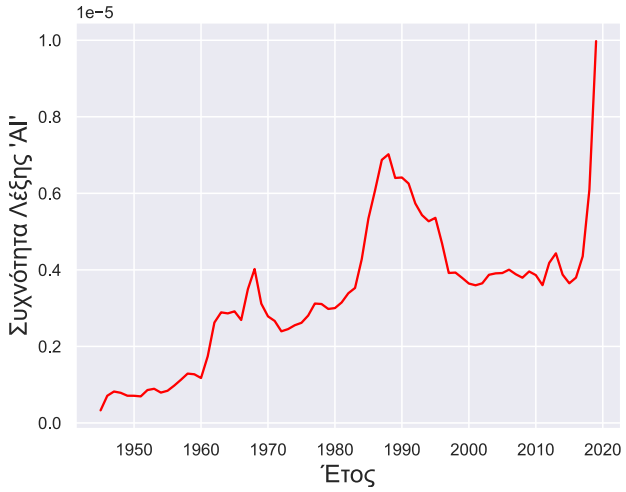
Για τον σκοπό αυτό, παρουσιάζουν έναν επαναστατικό αλγόριθμο ο οποίος αυτοματοποιεί τη διαδικασία μηχανικής μάθησης στα νευρωνικά δίκτυα. Πρόκειται για τον αλγόριθμο ανάστροφης διάδοσης σφάλματος (back propagation) ο οποίος, παραδόξως, ενώ είχε αναπτυχθεί περίπου το 1960 γνώρισε ευρεία χρήση από το 1980 και μετά. Συνεπώς, η σημασία του κονεκτιβισμού είναι καθοριστική αφού αναβίωσε τις ιδέες της γνωστικής επιστήμης επανατοποθετώντας κατά αυτόν τον τρόπο τον κλάδο της Τεχνητής Νοημοσύνης σε μια πιο επιστημονική τροχιά.

Η στροφή σε μια πιο επιστημονική προσέγγιση του κλάδου της Τεχνητής Νοημοσύνης το 1980 δε συνέβαλλε μόνο στην ανάπτυξη του κλάδου των νευρωνικών δικτύων. Για παράδειγμα, ο κλάδος της επεξεργασίας φυσικής γλώσσας επωφελήθηκε σημαντικά από την επιτυχημένη μοντελοποίηση ακολουθιών με τη χρήση κρυφών Μαρκοβιανών μοντέλων (hidden Markov models) και αργότερα, το 1997 με μονάδες μακράς-βραχέας μνήμης (Long-Short Term Memory block - LSTM). Επίσης, ο χώρος της όρασης υπολογιστών επωφελήθηκε από τη σύγκλιση της Τεχνητής Νοημοσύνης με τις θετικές επιστήμες. Την πρόοδο μαρτυρά η εμφάνιση των πρώτων εφαρμογών οπτικής αναγνώρισης χαρακτήρων (optical character recognition) τη δεκαετία του 1980 και ύστερα, των τυποποιημένων βάσεων δεδομένων για ανάπτυξη και μέτρηση απόδοσης οπτικών συστημάτων αναγνώρισης μοτίβων (π.χ. MNIST). Επίσης, σημαντικά αναπτύχθηκε ο χώρος της ταξινόμησης προτύπων με τη δημιουργία ή βελτίωση αρκετών μοντέλων όπως οι μηχανές διανυσματικής υποστήριξης με μέθοδο πυρήνα (Support Vector Machines with kernel trick) και τα δίκτυα ακτινικής βάσης (Radial Basis Networks). Ακόμα και ο χώρος της συλλογιστικής για τον οποίο κάναμε λόγο σε προηγούμενες παραγράφους εμπλουτίστηκε με μια σχολαστική και αποδοτική - αυτή τη φορά - μοντελοποίηση αβεβαιότητας της γνώσης μέσω της ανάπτυξης Μπεϋζιανών δικτύων (Bayesian networks). Τέλος, ο χώρος της στατιστικής γνώρισε πρόοδο αφού στις παραδοσιακές τεχνικές συμπερασματολογίας προστέθηκαν αυτές βασιζόμενες σε μηχανική μάθηση.

Προς τα τέλη της δεκαετίας του 1990 και τις αρχές του 2000 η έρευνα είχε εστιάσει σε κλάδους της Τεχνητής Νοημοσύνης που δε σχετίζονταν με τα νευρωνικά δίκτυα. Η κατάσταση αυτή όμως σύντομα αναστράφηκε. Αρχικά, η μεγάλη υπολογιστική ισχύ που απαιτούσαν αλγόριθμοι εκπαίδευσης βαθιών νευρωνικών δικτύων δε διευκόλυναν τον πειραματισμό [;]. Παρόλα αυτά, χάρη σε τρεις ερευνητές (Geoffrey Hinton, Yann LeCun και Yoshua Bengio) που χρηματοδοτούνταν από το Καναδικό Ινστιτούτο για προηγμένη έρευνα (CIFAR) η ενασχόληση με τα νευρωνικά δίκτυα κρατήθηκε ζωντανή οδηγώντας τελικά σε αξιοσημείωτη πρόοδο. Το πρώτο ορόσημο των βαθιών νευρωνικών δικτύων ήταν το 2006 όπου οι Hinton et al. [;] απέδειξαν ότι ένα είδος βαθέος νευρωνικού δικτύου - τα βαθιά δίκτυα πίστης (deep belief networks) - μπορούν να εκπαιδευτούν αποδοτικά και γρήγορα μέσω ενός άπληστου (greedy) αλγορίθμου. Το δεύτερο ορόσημο που απέδειξε τις προοπτικές των βαθιών νευρωνικών δικτύων ήταν η επιτυχία αυτής της τεχνολογίας σε διαγωνισμούς ταξινόμησης εικόνων της βάσης ImageNet το 2010 και το 2012. Στη δημοσίευσή τους ImageNet Classification with Deep Convolutional Neural Networks, οι A. Krizhevsky et al. [;] περιγράφουν μια νέα αρχιτεκτονική νευρωνικών δικτύων (βαθιά συνελικτικά δίκτυα) αλλά και πρωτοπόρες μεθόδους για αποδοτική εκπαίδευση (dropout, ReLU activation function κλπ.). Έκτοτε, το ενδιαφέρον απογειώθηκε και σε συνδυασμό με την αυξημένη διαθεσιμότητα δεδομένων, εξειδικευμένων συσκευών για παράλληλη επεξεργασία και αποδοτικών αλγορίθμων εκπαίδευσης δημιούργησαν το πιο πρόσφορο έδαφος για την άνθιση της Τεχνητής Νοημοσύνης.

Σήμερα, η εξέλιξη της Τεχνητής Νοημοσύνης και δη των νευρωνικών δικτύων είναι ραγδαία. Για παράδειγμα, με τη χρήση μοντέλων νευρωνικών δικτύων βασισμένων στον μηχανισμό προσοχής (attention-based neural networks) όπως τα λεγόμενα transformers, σημειώθηκε αξιοσημείωτη πρόοδος τόσο στον κλάδο της επεξεργασίας φυσικής γλώσσας (natural language processing) (βλέπε GPT -3) όσο και στον χώρο της όρασης υπολογιστών (βλέπε Vision Transformer και CoAtNet). Αναλυτικότερα για το Generative Pre-trained Transformer 3 - GPT 3, πρόκειται για το γλωσσικό μοντέλο που μπορεί να δημιουργήσει κείμενο σε φυσική γλώσσα ακόμα και να διατηρήσει για εύλογο χρόνο συζήτηση με έναν άνθρωπο. Ακόμα, σημαντική πρόοδος παρατηρείται στη μέθοδο μάθησης με αυτο-επίβλεψη (self-supervision) επιτρέποντας έτσι την εκπαίδευση δικτύων χωρίς να απαιτείται η σχολαστική και χρονοβόρα ανάθεση ετικετών στα δεδομένα εκπαίδευσης. Τέλος, μεταξύ άλλων, σπουδαία εξέλιξη υπήρξε πρόσφατα στις εφαρμογές όπου δοθέντων μερικών εικόνων από ένα αντικείμενο συνθέτουν εικόνες αυτού από νέες γωνίες θέασης (Novel View Synthesis). Ενδεικτικά, πρωτοπόρα έργα στον χώρο αυτό είναι το NeRF και το GIRAFFE [;].

Κλείνοντας το ιστορικό αυτό σημείωμα, δεν μπορώ παρά να αντικρίσω με δέος το μέλλον που επιφέρει η τεχνολογία της Τεχνητής Νοημοσύνης. Ανύπαρκτη πριν έναν αιώνα, σήμερα είναι μέρος της καθημερινότητά μας με τεράστια ορμή που δε φαίνεται να κατευνάζει. Αν μέσα σε μερικές δεκαετίες έχει τόσες δυνατότητες, στο εγγύς μέλλον η δύναμή της θα είναι τεράστια. Δύναμη που θα δημιουργήσει μια επίγεια ουτοπία ή θα αποτελέσει ένα ακόμα τουβλάκι στην οικοδόμηση της κοινωνίας του ρίσκου [3](;)... ο χρόνος θα δείξει.



Σχήμα 1.2: Γραφική παράσταση της συχνότητας του όρου AI σε βιβλία γραμμένα στην αγγλική γλώσσα ανά έτος (από 1945 μέχρι και 2019). Είναι εμφανείς οι τρεις περίοδοι ακμής του κλάδου. *Παράχθηκε από το Google Ngram Viewer.*

1.3 Κίνητρο

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam

[3]Ο όρος «κοινωνία του ρίσκου» (risk society) είναι δανεισμένος από το ομώνυμο βιβλίο του Ulrich Beck [;] όπου περιγράφεται το χαρακτηριστικό των μοντέρνων κοινωνιών να οργανώνονται γύρω από νέες μορφές, απαρατήρητου ρίσκου (όπως αυτό της κλιματικής αλλαγής).

nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.4 Συνεισφορά Εργασίας

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent

blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.5 Οργάνωση του Τόμου

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Στο παρόν κεφάλαιο θα οικοδομήσουμε την απαραίτητη γνώση στην οποία βασίζεται η έρευνα των επόμενων ενοτήτων. Αρχικά, θα παρουσιαστούν συνοπτικά τα τεχνητά νευρωνικά δίκτυα [1] υπό μια μαθηματική σκοπιά. Έπειτα, θα αναλυθούν τα νευρωνικά δίκτυα με κάψουλες (capsule networks) τα οποία και αποτελούν το κύριο θέμα της εργασίας. Τέλος, θα γίνει αναφορά σε νέες τεχνικές και αλγορίθμους που χρησιμοποιήθηκαν στο παρόν έργο ώστε η μετέπειτα εισαγωγή των μεθόδων μας για την εξέλιξη των νευρωνικών δικτύων με κάψουλες να είναι περισσότερο ομαλή και κατανοητή.

2.1 Τεχνητά Νευρωνικά Δίκτυα

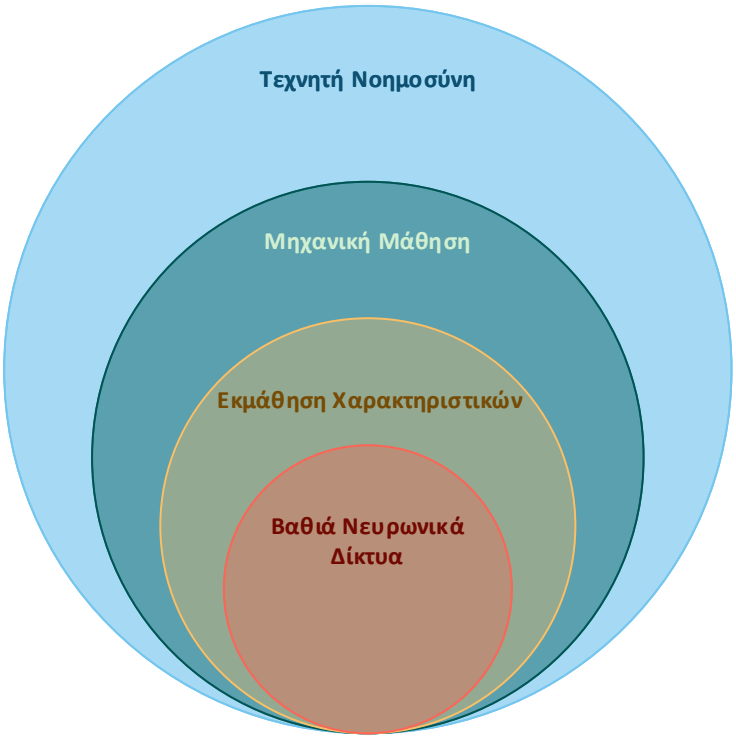
Τα σημερινά τεχνητά νευρωνικά δίκτυα, όπως είναι αναμενόμενο, απέχουν σημαντικά από το πρώτο μοντέλο των Warren McCulloch και Walter Pitts [;] που συζητήσαμε στην ενότητα 1.2. Με την ωρίμανση της τεχνολογίας, αυτή ανεξαρτητοποιήθηκε από την (υπολογιστική) νευροεπιστήμη και εντάχθηκε στην Τεχνητή Νοημοσύνη υπό μια ιεραρχική δομή. Κρίνεται λοιπόν σκόπιμο να παρουσιάσουμε αυτήν την ιεραρχική δομή οργάνωσης της Τεχνητής Νοημοσύνης και μετέπειτα να αναφερθούμε στα επιμέρους στοιχεία της.

Όπως βλέπουμε στο σχήμα 2.1 τα νευρωνικά δίκτυα πολλών επιπέδων (βαθιά νευρωνικά δίκτυα) είναι ένα μέρος του κλάδου της εκμάθησης χαρακτηριστικών (feature learning ή representation learning) που είναι ένα μέρος της μηχανικής μάθησης η οποία με τη σειρά της ανήκει στο ευρύτερο επιστημονικό πεδίο της τεχνητής νοημοσύνης. Φυσικά, η τεχνητή νοημοσύνη περιλαμβάνει αρκετούς άλλους κλάδους εκτός από αυτόν της μηχανικής μάθησης[2]. Μια χρήσιμη παρατήρηση είναι ότι οι σχέσεις υποσυνόλου συμπίπτουν με τη χρονική αλληλουχία ανάπτυξης του κάθε κλάδου. Δηλαδή, κάθε υποσύνολο αναπτύχθηκε ταυτόχρονα ή αργότερα από το οποιοδήποτε υπερσύνολό του.

Στη συνέχεια, θα γίνει λόγος για τα στοιχεία εκείνα που περιλαμβάνουν την τεχνολογία των βαθιών νευρωνικών δικτύων προκειμένου ο αναγνώστης να αποκτήσει μια εποπτικότερη εικόνα.

[1] Από εδώ και στο εξής, με τον όρο «νευρωνικά δίκτυα» θα αναφερόμαστε στα «τεχνητά νευρωνικά δίκτυα».

[2] Βέβαια, ο κλάδος της μηχανικής μάθησης είναι σήμερα ο γρηγορότερα αναπτυσσόμενος.



Σχήμα 2.1: Διάγραμμα Venn όπου απεικονίζει τη θέση των νευρωνικών δικτύων στην οργάνωση της τεχνητής νοημοσύνης. *Παράχθηκε από το Microsoft Visio™.*

2.1.1 Μηχανική Μάθηση

Όπως προδίδει ο όρος, σε αδρές γραμμές τα συστήματα μηχανικής μάθησης έχουν τη δυνατότητα να μαθαίνουν μια εργασία χωρίς να έχουν προγραμματιστεί με ρητές εντολές για τη συγκεκριμένη εργασία αυτή[3]. Ίσως, ο πιο πλήρης ορισμός δίνεται από τον Tom M. Mitchell [;] σύμφωνα με τον οποίο, ένα υπολογιστικό πρόγραμμα λέγεται ότι μαθαίνει από μια εμπειρία E, ως προς ένα σύνολο εργασιών T και ένα μέτρο απόδοσης P, εάν η απόδοσή του σε εργασίες του T, όπως αυτή μετριέται από το P, βελτιώνεται με την E. [4]

Σύμφωνα με τον ανωτέρω ορισμό διακρίνουμε τρία βασικά συστατικά ενός συστήματος μηχανικής μάθησης. Αυτά είναι τα παρακάτω:

Εργασία - T Είναι το πρόβλημα το οποίο επιθυμούμε να λύσουμε.

Μέτρο Απόδοσης - P Αποτελεί μια μετρική του στόχου ως ένδειξη ποιότητας της λύσης μας. Από μαθηματική σκοπιά, είναι αυτό που ο αλγόριθμος μάθησης βελτιστοποιεί.

Εμπειρία - E Πρόκειται για τα δεδομένα εισόδου που λαμβάνει το σύστημα υπό τη μορφή

[3] Η δυνατότητα αυτή είναι πολύ σημαντική αφού, όπως διαπιστώσαμε στην ενότητα 1.2 όταν έγινε λόγος για τα έμπειρα συστήματα, για πολλές εργασίες είναι πρακτικός αδύνατο να περιγραφούν ρητά και ντετερμινιστικά οι λύσεις τους.

[4] Ο ορισμός αυτός εξηγεί γιατί για παράδειγμα η λήψη μιας ιστοσελίδας της βικιπέδιας και η αποθήκευσή της τοπικά στον υπολογιστή δεν αποτελεί μηχανική μάθηση. Όπως προκύπτει, η «γνώση» αυτή δεν καθιστά καλύτερο τον υπολογιστή σε κάποια εργασία [;].

παραδειγμάτων ή ως ερεθίσματα ανάδρασης από το περιβάλλον. Όπως θα δούμε στη συνέχεια, ο τρόπος απόκτησης αυτών των δεδομένων αλλά και η φύση τους καθορίζει το είδος της μάθησης.

Βασικά Είδη Συστημάτων Μηχανικής Μάθησης

Τα είδη των συστημάτων μηχανικής μάθησης μπορούν να ταξινομηθούν ανάλογα με το:

- *Αν εκπαιδεύονται με ανθρώπινη επίβλεψη.*
  Ανάλογα με αυτό το κριτήριο έχουμε τις εξής βασικές κατηγορίες: επιβλεπόμενη (supervised), μη-επιβλεπόμενη (un-supervised) και ενισχυτική μάθηση (reinforcement learning).

- *Αν μαθαίνουν σταδιακά (incrementally) και «στον αέρα» (on the fly).*
  Σε αυτήν την περίπτωση χωρίζουμε τα συστήματα μηχανικής μάθησης σε αυτά που πραγματοποιούν μάθηση σε ζωντανό χρόνο (online learning) και σε αυτά που μαθαίνουν κατά δέσμες (batch learning).

- *Αν κατασκευάζουν μοντέλα προσαρμοσμένα στα δεδομένα.*
  Με αυτό το κριτήριο χωρίζονται σε συστήματα βασισμένα σε μοντέλο (model-based) ή σε συστήματα βασισμένα σε παραδείγματα (instance-based). [;]

Προφανώς, κάθε δυνατός συνδυασμός των παραπάνω κριτηρίων είναι αποδεκτός, οδηγώντας έτσι στην ταξινόμηση των συστημάτων μηχανικής μάθησης σε μια πληθώρα από διαφορετικές κατηγορίες. Κρίνεται χρήσιμο, να αναφέρουμε σε όλη την έκταση του έργου τις κατηγορίες στις οποίες ανήκει το κάθε σύστημα που παρουσιάζουμε. Για αυτόν τον λόγο, παροτρύνουμε τον αναγνώστη που δεν είναι εξοικειωμένος με τους ανωτέρω όρους να διαβάσει τους αντίστοιχους ορισμούς στο παράρτημα Α΄.

2.1.2 Εκμάθηση Χαρακτηριστικών

Η ανάπτυξη των πρώτων συστημάτων μηχανικής μάθησης απεμπόλησε την ανάγκη των ευφυών εφαρμογών για σχολαστική και ρητή (hard-coded) αναπαράσταση του χώρου του προβλήματος (π.χ. με την χρήση προτασιακής λογικής). Με τα νέα συστήματα, η γνώση για το πρόβλημα μαθαίνονταν αυτοματοποιημένα μέσω αλγορίθμων μάθησης από το σύνολο δεδομένων εκπαίδευσης. Με άλλα λόγια, τα αλγοριθμικά κατασκευάσματα μάθαιναν να αντιστοιχίζουν με αυτοματοποιημένο τρόπο τα δεδομένα εισόδου (κωδικοποιημένα σε μια μορφή αναπαράστασης) σε τιμές εξόδου.

Παρόλα αυτά, τα πρώτα, απλά συστήματα μηχανικής μάθηση δεν έλυσαν όλα τα προβλήματα. Όπως είναι εμφανές από την ανωτέρω περιγραφή, αν και δεν απαιτούνταν η λεπτομερή συγγραφή βάσεων γνώσης, παρέμενε η ανάγκη για αναπαράσταση των δεδομένων εισόδου με μια αποδοτική μορφή. Είναι γεγονός, άλλωστε, ότι η αναπαράσταση σε πολλά συστήματα επηρεάζει καθοριστικά την απόδοση του συστήματος[5]. Για αυτόν τον λόγο, εξελίχθηκαν διαδικασίες «μηχανικής χαρακτηριστικών» (feature engineering) όπου αξιοποιώντας την τεχνική γνώση του χώρου του προβλήματος (domain knowledge) στόχος είναι η αναπαράσταση των ακατέργαστων δεδομένων εκπαίδευσης ως σύνολο

[5]Η σημασία της αναπαράστασης δεδομένων στην απόδοση των αλγοριθμικών κατασκευασμάτων δε θα πρέπει να μας εκπλήσσει αφού κάτι αντίστοιχο ισχύει και στους ανθρώπους. Για παράδειγμα, οι περισσότεροι είναι πολύ πιο αποδοτικοί στην αριθμητική χρησιμοποιώντας την αραβική αναπαράσταση αριθμών απ ότι τη λατινική [;].

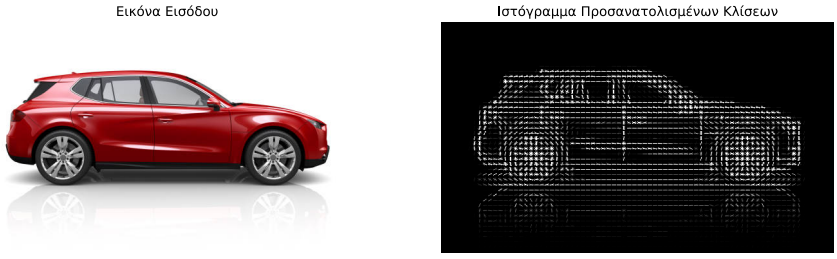
(συνήθως διάνυσμα) από κατάλληλα χαρακτηριστικά. Η καταλληλότητα έγκειται στο πόσο χρήσιμη πληροφορία παρέχουν τα χαρακτηριστικά υπό τον περιορισμό να είναι όσο το δυνατόν περισσότερο ανεξάρτητα μεταξύ τους ώστε να αποπλέκουν (disentagle) τους παράγοντες διακύμανσης (factors of variation) των δεδομένων που επηρεάζουν την τιμή εξόδου [;].

Οι ανωτέρω έννοιες μπορούν να καταστούν περισσότερο κατανοητές με ένα παράδειγμα συστήματος εκτίμησης τιμών κατοικιών [;] (πρόβλημα παλινδρόμησης, επίλυση με επιβλεπόμενη μάθηση κατά δέσμες). Πιο συγκεκριμένα, δοθέντος ενός συνόλου ακατέργαστων δεδομένων που αφορούν την αγορά σπιτιών σε μια περιοχή, το σύστημα, μέσω μηχανικής μάθησης, θα είναι ικανό να εκτιμήσει την τιμή με την οποία μια κατοικία θα πρέπει να κοστολογηθεί για να βγει στην αγορά. ··Όπως εξηγήσαμε, προτού τροφοδοτήσουμε το σύστημα με το σύνολο δεδομένων, είναι σκόπιμο να εφαρμόσουμε διαδικασίες μηχανικής χαρακτηριστικών σε αυτά και να δημιουργήσουμε μια νέα αναπαράσταση. Τα ακατέργαστα δεδομένα εκπαίδευσης αποτελούνται από μια λίστα όπου κάθε γραμμή αντιστοιχεί σε μια οικία με όλες τις προδιαγραφές της και την τιμή πώλησής της. Στο πρόβλημα του παραδείγματος:

- Ένας παράγοντας διακύμανσης θα μπορούσε να είναι η ακρίβεια της συγκεκριμένης περιοχής. Εντούτοις, σαν προδιαγραφές ας υποθέσουμε ότι αναφέρονται μόνο το γεωγραφικό πλάτος και γεωγραφικό μήκος με αποτέλεσμα η ακρίβεια της περιοχής να μην είναι άμεσα παρατηρήσιμη (συνηθισμένο φαινόμενο στους παράγοντες διακύμανσης). Θα μπορούσαμε λοιπόν να μετατρέψουμε τις συντεταγμένες σε ένα νέο χαρακτηριστικό: την «κλάση» της περιοχής. Ένας ακόμα παράγοντας διακύμανσης που είναι όμως άμεσα παρατηρήσιμος είναι το εμβαδόν επιφάνειας της κατοικίας.

- Μη χρήσιμη πληροφορία θα μπορούσε να είναι ο προσανατολισμός της οικίας. Σε αυτή την περίπτωσή, η δημιουργία μιας νέας αναπαράσταση δεδομένων χωρίς το παρόν προσδιορισμό θα βοηθούσε την επίδοση του συστήματος.

- Δύο αλληλοεξαρτώμενα χαρακτηριστικά (με υψηλή συν—διακύμανση) θα μπορούσαν να είναι ο αριθμός των υπνοδωματίων και ο αριθμός των μπάνιων όπου τότε η επιλογή της συγχώνευσής τους πιθανότατα βελτίωνε την απόδοση.



Σχήμα 2.2: Παράδειγμα εξαγωγής χαρακτηριστικών σε εικόνα ενός αυτοκινήτου με τη μέθοδο του Ιστογράμματος Προσανατολισμένων Κλίσεων (Histogram of Oriented Gradients). *Παράχθηκε τοπικά. ΤΟΔΟ*

Αν και στο παραπάνω πρόβλημα ήταν σχετικά εύκολη η «χειροναχτική» εξαγωγή χαρακτηριστικών, υπάρχουν πολλοί χώροι προβλημάτων όπου κάτι τέτοιο είναι από πολύ απαιτητικό έως απίθανο. Ενδεικτικά, σε ένα πρόβλημα οπτικής αναγνώρισης ζώων και αντικειμένων (όπως αυτό του CIFAR-10 [;]) είναι εξαιρετικά δύσκολη η περιγραφή χαρακτηριστικών που θα λαμβάνουν μια αναπαράσταση σε εικονοστοιχεία (pixel) και θα παράγουν μια χρήσιμη αναπαράσταση. Μολονότι υπάρχουν γενικευμένες μέθοδοι για την εξαγωγή χαρακτηριστικών σε εικόνες όπως αυτή του σχήματος 2.2, αυτές δεν είναι βέλτιστα προσαρμοσμένες στον χώρο του εκάστοτε προβλήματος. Απόρροια αυτού μεταξύ άλλων είναι η απόρριψη στοιχείων των ακατέργαστων δεδομένων που μπορεί να είναι χρήσιμα (π.χ. σε ένα πρόβλημα αναγνώρισης μάρκας και χρώματος αυτοκινήτου, η μέθοδος Ιστογράμματος Προσανατολισμένων Κλίσεων θα απέρριπτε το χρώμα, στοιχείο χρήσιμο για τον χώρο του προβλήματος).

Η λύση για την αντιμετώπιση των προβλημάτων της χειροναχτικής εξαγωγής χαρακτηριστικών είναι η χρήση των αλγορίθμων μηχανικής μάθησης για την εκμάθηση όχι μόνο για της αντιστοίχησης των δεδομένων εκπαίδευσης στην επιθυμητή έξοδο αλλά και των ίδιων των αναπαραστάσεων των δεδομένων. Αν και συνήθως, οι προκύπτουσες αναπαραστάσεις μετά τον μετασχηματισμό των ακατέργαστων δεδομένων δεν είναι κατανοητές από τον άνθρωπο, εφόσον η εκπαίδευση γίνει επιτυχημένα, αποτελούνται από χρήσιμα χαρακτηριστικά (που κωδικοποιούν τους παράγοντες διακύμανσης). Χαρακτηριστικό παράδειγμα συστήματος για την εκμάθηση χαρακτηριστικών είναι ο Αυτοκωδικοποιητής (Autoencoder).

2.1.3 Πολυεπίπεδα Νευρωνικά Δίκτυα

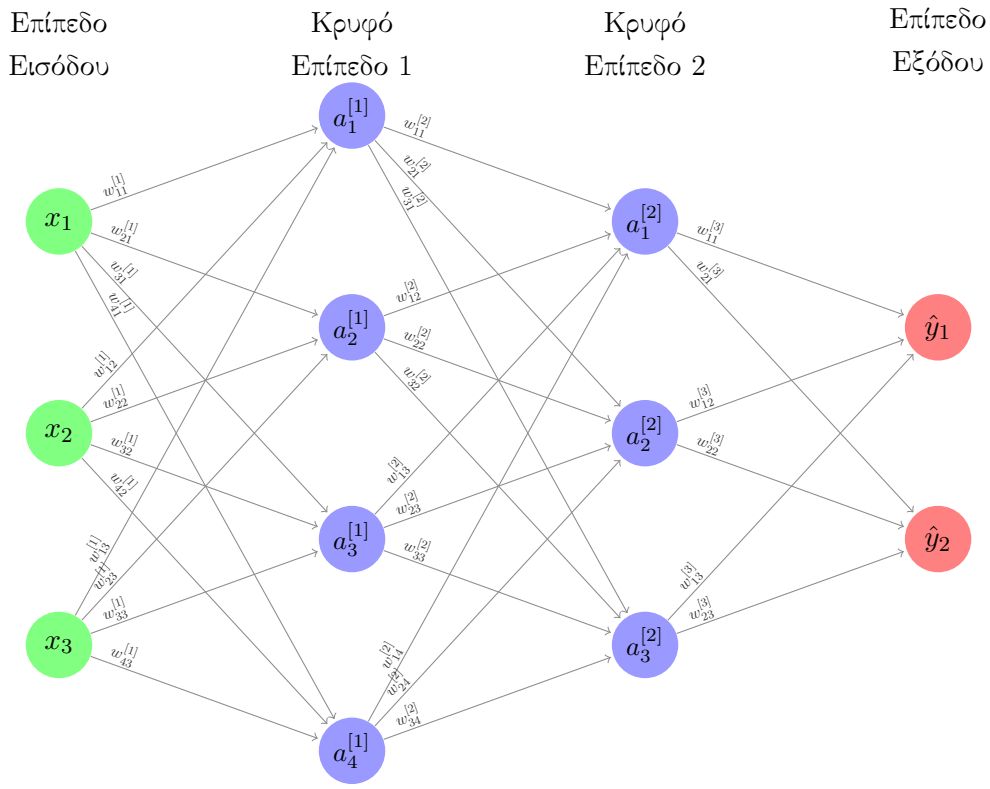
Η εκμάθηση χαρακτηριστικών σε συνδυασμό με τις ιδέες του κονεκτιβισμού περί κατανεμημένης αναπαράστασης (βλ. ενότητα 1.2) μας οδηγεί αναπόφευκτα στη βαθιά μάθηση (deep learning). Υπό μια αφαιρετική σκοπιά, πρόκειται για τα λεγόμενα «πολυεπίπεδα νευρωνικά δίκτυα» τα οποία συνδυάζουν τόσο τον μετασχηματισμό της αναπαράστασης των δεδομένων εισόδου όσο και την αντιστοίχηση αυτών των νέων αναπαραστάσεων στην τιμή εξόδου. Τα συστήματα αυτά, όπως θα δούμε στη συνέχεια, είναι δομημένα από απλές υπολογιστικές μονάδες που τους επιτρέπουν να δημιουργούν σύνθετες αναπαραστάσεις μέσω μιας σειράς από εμφωλευμένες, απλούστερες αναπαραστάσεις. Σημειώνουμε ότι πραγματοποιούν μάθηση με κατασκευή μοντέλου (model–based learning systems) και χρησιμοποιούνται τόσο σε προβλήματα ταξινόμησης όσο και παλινδρόμησης. Στις επόμενες παραγράφους, θα περιγράψουμε με μεγαλύτερη λεπτομέρεια τον χώρο των νευρωνικών δικτύων[6].

Δομή Απλών Νευρωνικών Δικτύων

Τα Νευρωνικά Δίκτυα στην πιο βασική τους μορφή (Feedforward Neural Networks ή Multilayer Perceptron) αποτελούνται από απλούς τεχνητούς νευρώνες διασυνδεδεμένους μεταξύ τους με συνάψεις σχηματίζοντας μια πολυεπίπεδη διάταξη. Το πρώτο επίπεδο ονομάζεται επίπεδο εισόδου (input layer) ενώ το τελευταίο ονομάζεται επίπεδο εξόδου (output layer). Όλα τα ενδιάμεσα επίπεδα λέγονται κρυφά επίπεδα (hidden layers) διότι οι τιμές τους δε δίνονται από τα δεδομένα [;]. Στην απλή περίπτωση που εξετάζουμε, κάθε νευρώνας δέχεται ως είσοδο τιμές από όλους τους νευρώνες του αμέσως προηγούμενου επιπέδου (fully connected layer) και αφού κάνει

[6]Για έναν τυπικό ορισμό, παραπέμπουμε τον αναγνώστη στο παράρτημα Α΄

υπολογισμούς με αυτές στέλνει το αποτέλεσμα σε όλους τους νευρώνες του αμέσως επόμενου επιπέδου.



Σχήμα 2.3: Διάγραμμα Τεχνητού Νευρωνικού Δικτύου με δύο κρυφά επίπεδα. Οι αγκύλες στους εκθέτες προσδιορίζουν τον αριθμό του επιπέδου. *Παράχθηκε από το LaTeX πακέτο neuralnetwork. Το πακέτο τροποποιήθηκε και επεκτάθηκε τοπικά.*

Η αναπαράσταση των νευρωνικών δικτύων γίνεται με έναν γράφο από ακμές (συνάψεις) και κόμβους (νευρώνες ή τιμές εισόδου). Κοιτώντας κανείς το σχήμα 2.3 μπορεί να παρατηρήσει πως οι κόμβοι των επιπέδων εισόδου και εξόδου ξεχωρίζουν από τους κόμβους των κρυφών επιπέδων. Αυτό έχει γίνει για να τονιστεί η ξεχωριστή λειτουργία τους. Πιο συγκεκριμένα, στην περίπτωση του επιπέδου εισόδου, αυτό περιέχει τόσους κόμβους όσος είναι και ο αριθμός των χαρακτηριστικών που περιγράφουν το κάθε παράδειγμα (δηλαδή όσο και το μήκος του διανύσματος εισόδου). Ουσιαστικά, οι κόμβοι εισόδου απλά λαμβάνουν τις τιμές των χαρακτηριστικών και, χωρίς να τις μεταβάλλουν, τις δρομολογούν στους κόμβους του επόμενου επιπέδου (για αυτό και αποφεύγεται η επωνομασία αυτών των κόμβων ως νευρώνες). Στην περίπτωση του επιπέδου εξόδου, ο αριθμός των κόμβων είναι τόσος όσος και ο αριθμός των χαρακτηριστικών για την περιγραφή της πρόβλεψης (τόσο όσο το μήκος του διανύσματος εξόδου). Οι κόμβοι εξόδου συνήθως επιβάλουν περιορισμούς στις τιμές των χαρακτηριστικών εξόδου ώστε αυτές να ανήκουν σε ένα φραγμένο σύνολο αριθμών (π.χ. το [0,1]).

Ένα νευρωνικό δίκτυο χωρίς κρυφά επίπεδα δε διαφέρει από έναν γραμμικό ταξινομητή. Είναι γεγονός ότι οι εκπληκτικές δυνατότητες των νευρωνικών δικτύων αποδίδονται στα κρυφά επίπεδα. Χάρη σε αυτά είναι δυνατή η σταδιακή σύνθεση αφηρημένων αναπαραστάσεων από επίπεδο σε επίπεδο που κωδικοποιούν τους παράγοντες διακύμανσης. Τα κρυφά επίπεδα τα απαρτίζουν οι

κόμβοι κρυφού επιπέδου[7]. Ο κάθε ένας από αυτούς υπολογίζει την έξοδο μιας μη γραμμικής συνάρτησης με είσοδο ένα γραμμικό συνδυασμό των τιμών των κόμβων του προηγούμενου επιπέδου. Αξίζει να αναφερθεί στο σημείο αυτό πως δεν υπάρχει κάποιος συγκεκριμένος περιορισμός για τον αριθμό των κόμβων των κρυφών επιπέδων.

Η φορμαλιστική περιγραφή των παραμέτρων[8] και των υπολογισμών που λαμβάνουν χώρα κατά τη διαδικασία πρόβλεψης ενός νευρωνικού δικτύου περιγράφονται παρακάτω.

Έστω ένα παράδειγμα εισόδου το οποίο περιγράφεται από n_x χαρακτηριστικά με το διάνυσμα $X = [x_1, x_2, x_3, \ldots, x_{n_x}]^T$. Όλα τα δεδομένα εκπαίδευσης, έστω M, μπορούν να ομαδοποιηθούν σε έναν πίνακα $\boldsymbol{X}$ ως εξής:

$$\boldsymbol{X} = \underbrace{\begin{bmatrix} | & | & & | \\ X^{(1)} & X^{(2)} & \ldots & X^{(M)} \\ | & | & & | \end{bmatrix}}_{(n_x \times M)}. \tag{2.1}$$

Όπου οι παρενθέσεις στους εκθέτες δηλώνουν τον αριθμό του παραδείγματος.

Αφού προσδιορίσαμε μια μαθηματική αναπαράσταση για τα δεδομένα εισόδου, πάμε να προσδιορίσουμε με φορμαλιστικό τρόπο τις παραμέτρους του νευρωνικού δικτύου. Με L θα συμβολίζουμε τον αριθμό των επιπέδων του νευρωνικού δικτύου (χωρίς να μετράμε το επίπεδο εισόδου). Για παράδειγμα, στο δίκτυο του σχήματος 2.3 ισχύει $L = 3$. Επίσης, ο αριθμός των κόμβων ενός επιπέδου, έστω l, θα συμβολίζεται με $n^{[l]}$. Προφανώς, θα ισχύει ότι $l \in [0, L]$. Στο παράδειγμα του σχήματος 2.3 ισχύει $n^{[0]} = n_x = 3, n^{[1]} = 4, n^{[2]} = 3, n^{[3]} = n_y = 2$. Έχοντας δώσει έναν συμβολισμό ορισμένων βασικών υπερπαραμέτρων[9], μπορούμε αποδώσουμε φορμαλιστικά τις παραμέτρους του δικτύου οι οποίες είναι:

- Τα βάρη των ακμών (weights) που συνδέουν δύο διαδοχικά επίπεδα.
  Τα βάρη μεταξύ διαδοχικών επιπέδων $l-1$ και l μπορούμε να τα οργανώσουμε σε μια μορφή πίνακα ως εξής:

$$\boldsymbol{W^{[l]}} = \underbrace{\begin{bmatrix} w^{[l]}_{11} & w^{[l]}_{12} & \ldots & w^{[l]}_{1n^{[l-1]}} \\ w^{[l]}_{21} & w^{[l]}_{22} & \ldots & w^{[l]}_{2n^{[l-1]}} \\ \vdots & \vdots & \ddots & \vdots \\ w^{[l]}_{n^{[l]}1} & w^{[l]}_{n^{[l]}2} & \ldots & w^{[l]}_{n^{[l]}n^{[l-1]}} \end{bmatrix}}_{(n^{[l]} \times n^{[l-1]})}. \tag{2.2}$$

- Τα δυναμικά πόλωσης (biases) του κάθε νευρώνα.
  Όπως είναι λογικό, οι κόμβοι του επιπέδου 0 (επίπεδο εισόδου) δε διαθέτουν δυναμικά πόλωσης. Για όλους τους άλλους κόμβους σε κάθε επίπεδο (έστω l) έχουμε το εξής

[7] Εφεξής θα αποκαλούνται ως κόμβοι.

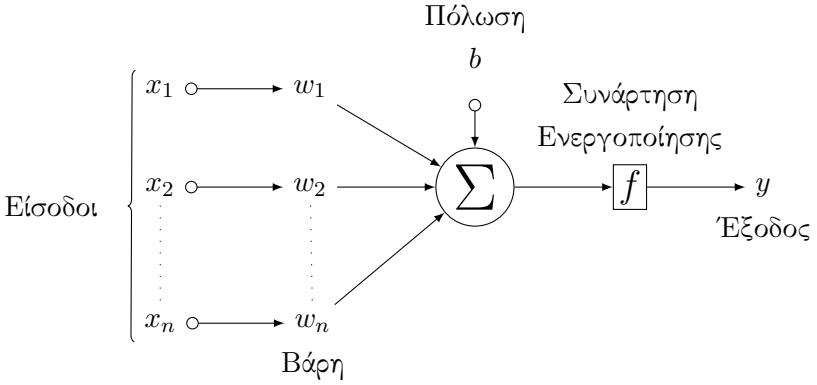
[8] Πρόκειται για μεταβλητές των οποίων οι τιμές μαθαίνονται κατά τη διάρκεια της εκπαίδευσης. Έτσι το νευρωνικό δίκτυο λέμε ότι προσαρμόζεται στα δεδομένα.

[9] Σε αντίθεση με τις παραμέτρους, οι υπερπαράμετροι είναι μεταβλητές που ορίζει ο χρήστης και δε μεταβάλλονται κατά την εκπαίδευση. Ονομάζονται έτσι διότι ελέγχουν έμμεσα τις τιμές των παραμέτρων.

διάνυσμα στήλη:

$$\boldsymbol{b}^{[l]} = \underset{(n^{[l]}\times 1)}{\begin{bmatrix} b_1^{[l]} \\ b_2^{[l]} \\ \vdots \\ b_{n^{[l]}}^{[l]} \end{bmatrix}}. \tag{2.3}$$

Τώρα, είμαστε σε θέση να περιγράψουμε τους υπολογισμούς που πραγματοποιεί κάθε νευρώνας. Για τον σκοπό αυτό, παρουσιάζουμε μια αναπαράστασή του στο σχήμα 2.4.



Σχήμα 2.4: Διάγραμμα ενός τεχνητού νευρώνα. *Παράχθηκε από το πακέτο tikz.*

Εσωτερικά ο κάθε νευρώνας δέχεται ως είσοδο τις τιμές των νευρώνων του προηγούμενου επιπέδου ή (στην περίπτωση του δεύτερου επιπέδου) τις τιμές του παραδείγματος εκπαίδευσης και πραγματοποιεί υπολογισμούς με αυτές για να παράξει μια τιμή εξόδου. Την κάθε τιμή εισόδου τη συμβολίζουμε με x ενώ την τιμή εξόδου του μεμονωμένου νευρώνα τη συμβολίζουμε με y. Η πράξη που επιτελείται εσωτερικά είναι:

$$y = f\Big(\sum_{i=1}^{n} w_i \times x_i + b\Big) \tag{2.4}$$

Όπου $f(x)$ είναι η συνάρτηση ενεργοποίησης (activation function). Αυτές οι συναρτήσεις, μεταξύ άλλων, επιτρέπουν στο σύστημα να μοντελοποιήσει μη γραμμικές σχέσεις εισόδου–εξόδου. Παραδείγματα αποτελούν η βηματική συνάρτηση (όπως στο Perceptron του H. Rosenblat [**;**]), η σιγμοειδής, η υπερβολική συνάρτηση εφαπτομένης (tanh) κ.τ.λ.

Στο σημείο αυτό να αναφέρουμε πως το όρισμα της συνάρτησης ενερογοποίησης, δηλαδή την ποσότητα $\sum_{i=1}^{n} w_i \times x_i + b$ τη συμβολίζουμε με z ενώ την τιμή του y την ονομάζουμε και τιμή ενεργοποίησης (συμβολίζοντάς τη με a). Στην περίπτωση δε που ο νευρώνας ανήκει στο επίπεδο εξόδου, την τιμή y την ονομάζουμε τιμή πρόβλεψης και την αναπαριστούμε με το γράμμα $\hat{y}$.

Κοιτώντας τώρα εποπτικά τη διαδικασία παραγωγής νέων προβλέψεων ενός τυπικού νευρωνικού δικτύου, μπορούμε να την περιγράψουμε χρησιμοποιώντας αναπαράσταση με διανύσματα (vector notation). Αναλυτικότερα, αν συγκεντρώσουμε όλες τις τιμές ενεργοποίησης a ενός επιπέδου l σε ένα διάνυσμα $A^{[l]} = [a_1^{[l]}, a_2^{[l]}, a_3^{[l]}, \ldots, a_{n^{[l]}}^{[l]}]^T$ και κάνουμε το ίδιο για τα ορίσματα z των συναρτήσεων ενεργοποίησης δηλαδή $Z^{[l]} = [z_1^{[l]}, z_2^{[l]}, z_3^{[l]}, \ldots, z_{n^{[l]}}^{[l]}]^T$ τότε συνολικά, γράφουμε:

- Τα στοιχεία a προκύπτουν από τα στοιχεία z του ίδιου επιπέδου ως εξής:

$$A^{[l]} = F^{[l]}\big(Z^{[l]}\big) \tag{2.5}$$

Όπου η συνάρτηση F εφαρμόζει την f σε κάθε στοιχείο του ορίσματός της ξεχωριστά (elementwise).

- Τα στοιχεία z υπολογίζονται από τα στοιχεία a του προηγούμενου επιπέδου μέσω της σχέσης:

$$Z^{[l]} = W^{[l]} \times A^{[l-1]} + \boldsymbol{b}^{[l]} \tag{2.6}$$

Όπου $A^0 = X$, το διάνυσμα χαρακτηριστικών ενός παραδείγματος ενώ $A^{[L]} = \hat{Y}$ το διάνυσμα εξόδου.

Η ανωτέρω ανάλυση πραγματοποιήθηκε για ένα μεμονωμένο παράδειγμα. Θα βοηθήσει για την επεξήγηση του τρόπου εκπαίδευσης των νευρωνικών δικτύων αν παρουσιάσουμε τώρα μια μορφή μαθηματικών τύπων που συγκεντρώνουν τις τιμές όλων των παραδειγμάτων του συνόλου εκπαίδευσης. Θυμηθήτε, ότι κάθε στήλη του πίνακα δεδομένων εισόδου $\boldsymbol{X}$ περιέχει ένα διάνυσμα παραδείγματος με αποτέλεσμα ο αριθμός των στηλών να ισούται με τον αριθμό των παραδειγμάτων, M. Κατά αυτόν τον τρόπο έχουμε:

- Για τα z ενός επιπέδου l για το κάθε παράδειγμα εισόδου:

$$\boldsymbol{Z^{[l]}} = \underset{(n^{[l]} \times M)}{\begin{bmatrix} | & | & & | \\ Z^{[l](1)} & Z^{[l](2)} & \dots & Z^{[l](M)} \\ | & | & & | \end{bmatrix}}. \tag{2.7}$$

- Αντίστοιχα, για τα a ενός επιπέδου l για το κάθε παράδειγμα εισόδου:

$$\boldsymbol{A^{[l]}} = \underset{(n^{[l]} \times M)}{\begin{bmatrix} | & | & & | \\ A^{[l](1)} & A^{[l](2)} & \dots & A^{[l](M)} \\ | & | & & | \end{bmatrix}}. \tag{2.8}$$

Όπως προκύπτουν από τις σχέσεις

$$\underset{(n^{[l]} \times M)}{\boldsymbol{Z}^{[l]}} = \underset{(n^{[l]} \times n^{[l-1]})}{W^{[l]}} \times \underset{(n^{[l-1]} \times M)}{\boldsymbol{A}^{[l-1]}} + \underset{(n^{[l]} \times 1)}{\boldsymbol{b}^{[l]}} \tag{2.9}$$

και

$$\underset{(n^{[l]} \times M)}{\boldsymbol{A}^{[l]}} = F^{[l]}\big(\underset{(n^{[l]} \times M)}{\boldsymbol{Z}^{[l]}}\big) \tag{2.10}$$

αντίστοιχα, όπου η μόνη διαφορά με τις 2.6, 2.5 είναι ότι αντικαταστήσαμε τα διανύσματα στήλες Z και A με τους πίνακες $\boldsymbol{Z}$ και $\boldsymbol{A}$. Και πάλι, $\boldsymbol{A}^0 = \boldsymbol{X}$, ο πίνακας όλων των δεδομένων εισόδου ενώ $\boldsymbol{A}^{[L]} = \hat{\boldsymbol{Y}}$ το σύνολο διανυσμάτων εξόδου.

Εκπαίδευση Νευρωνικών Δικτύων

Στην προηγούμενη παράγραφο διατυπώσαμε τους μαθηματικούς τύπους σύμφωνα με τους οποίους ένα νευρωνικό δίκτυο, δοθέντος ενός συνόλου δεδομένων $\boldsymbol{X}$ παράγει ένα σύνολο από προβλέψεις

$\hat{\boldsymbol{Y}}$. Η διαδικασία αυτή ονομάζεται και πρόσθια διάδοση (forward propagation). Παρόλα αυτά, δεν αναφερθήκαμε καθόλου στη διαδικασία μάθησης του δικτύου. Υποθέσαμε σιωπηρά ότι αυτό ήταν ήδη εκπαιδευμένο και οι παράμετροί του (τα βάρη και τα δυναμικά πόλωσης) ήταν σταθερά. Με τη μέχρι τώρα παρουσίαση, το νευρωνικό δίκτυο δεν είναι τίποτα άλλο παρά μια μη γραμμική συνάρτηση. Σε αυτήν την παράγραφο όμως, θα κάνουμε τη σύνδεση των νευρωνικών δικτύων με τη μηχανική μάθηση αναλύοντας τον μηχανισμό εκπαίδευσής τους.

Όπως έχουμε αναφέρει, τα νευρωνικά δίκτυα είναι πολυδύναμα συστήματα μηχανικής μάθησης βασισμένα σε μοντέλο ενώ καμία επιπλέον υπόθεση δεν μπορεί να γίνει εκ των προτέρων. Με άλλα λόγια, υπάρχουν νευρωνικά δίκτυα που ανήκουν σε όλες τις υπόλοιπες κατηγορίες που παρατέθηκαν στην ενότητα 2.1.1. Παρόλα αυτά, για τον σκοπό της παρούσας παραγράφου, θα περιορίσουμε αυτά τα πολυδύναμα συστήματα σε αυτά που πραγματοποιούν επιβλεπόμενη μάθηση. Ευτυχώς, αυτή είναι η πιο βασική κατηγορία και οι ιδέες που θα παρουσιαστούν υπό αυτή εύκολα μεταφέρονται και στις υπόλοιπες.

Στο πλαίσιο της επιβλεπόμενης μάθησης, εκτός από το σύνολο δεδομένων εισόδου $\boldsymbol{X}$ παρέχεται, και ένα σύνολο δεδομένων εξόδου $\boldsymbol{Y}$. Το τελευταίο αποτελείται από ένα επιθυμητό διάνυσμα (ή τιμή) στόχο για κάθε παράδειγμα του συνόλου δεδομένων εισόδου. Έτσι, (όπως αναφέρουμε και στο παράρτημα Α΄) σχηματίζονται ζεύγη διανυσμάτων εισόδου–επιθυμητής εξόδου. Στόχος του νευρωνικού δικτύου σε αυτήν την περίπτωση είναι να δημιουργήσει μια συνάρτηση που θα κάνει την αντιστοίχηση από τα παραδείγματα X στις προβλέψεις του στόχου, $\hat{Y}$ να είναι όσο πιο πιστή γίνεται στην αντιστοίχηση X σε Y. Με μαθηματικούς όρους, έστω η συνάρτηση του νευρωνικού δικτύου: $\mathcal{F}(X; \overline{W}, \overline{b})$, όπου με $\overline{W}$ και $\overline{b}$ συμβολίζεται αντίστοιχα το σύνολο των βαρών και δυναμικών πόλωσης σε όλα τα επίπεδα[10]. Θέλουμε για τη συνάρτηση $\mathcal{F}(X; \overline{W}, \overline{b}) : X \to \hat{Y}$ να ισχύει $\mathcal{F}(X; \overline{W^*}, \overline{b^*}) \approx \mathcal{G}^*(X)$ όπου $\mathcal{G}$ η (άγνωστη) συνάρτηση από την οποία (θεωρητικά) παράχθηκαν τα ζεύγη εισόδου–εξόδου και οι δείκτες αστερίσκοι συμβολίζουν τις ιδανικές τιμές των παραμέτρων[11].

Για να προσαρμόσουμε με βέλτιστο τρόπο τις παραμέτρους του μοντέλου στην εμπειρία απαιτείται (σύμφωνα με τον ορισμό της μηχανικής μάθησης) μια μετρική η οποία θα μας δείχνει πόσο κατάλληλη είναι η προσέγγισή μας (fittness). Αυτή η μετρική ονομάζεται συνάρτηση απώλειας (loss function) και στη γενική της μορφή είναι $\mathsf{L}(\hat{Y}, Y) = \mathsf{L}(\mathcal{F}(\boldsymbol{X}; \overline{W}, \overline{b}), Y)$. Ανάλογα με το είδος του προβλήματος, τυπικές συναρτήσεις απώλειας είναι:

- Η (binary cross entropy loss): $\mathsf{L}(\hat{Y}, Y) = -\frac{1}{M} \sum_{i=1}^{M} \left(y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \right)$ στην περίπτωση προβλήματος δυαδικής ταξινόμησης (όπου η έξοδος $\hat{y}$ είναι ίση με την πιθανότητα το παράδειγμα x να ανήκει στην κλάση 1)

[10] Το σύμβολο «;» διαβάζεται ως «παραμετροποιημένο από».

[11] Ουσιαστικά, το κύριο πρόβλημα που καλούνται να λύσουν τα νευρωνικά δίκτυα (και τα συστήματα μηχανικής μάθησης γενικότερα) πηγάζει από το γεγονός ότι η συνάρτηση $\mathcal{G}$ είναι άγνωστη. Αντ΄ αυτής, διατίθεται ένα σύνολο ζευγών $X - Y$ που αποτελεί υποσύνολο του πληθυσμού όλων των δυνατών εισόδων και το σύστημα επιδιώκει από αυτό το υποσύνολο να μάθει να γενικεύει σε παραδείγματα που δεν έχει δει. Αυτός είναι και ο λόγος που, υπό την αυστηρά μαθηματική έννοια, η διαδικασία της εκπαίδευσης αποκαλείται και συμπερασματολογία (inference). Σε τελική ανάλυση, τα περισσότερα νευρωνικά δίκτυα εκπαιδεύονται χρησιμοποιώντας συμπερασματολογία βασισμένη στη μέγιστη πιθανοφάνεια (maximum likelihood inference) [;]. Δηλαδή, προσπαθούν να μεγιστοποιήσουν την ποσότητα $p(y|x; \overline{W}, \overline{b})$ και γιαυτό ονομάζονται μοντέλα διάκρισης (descriminative models).

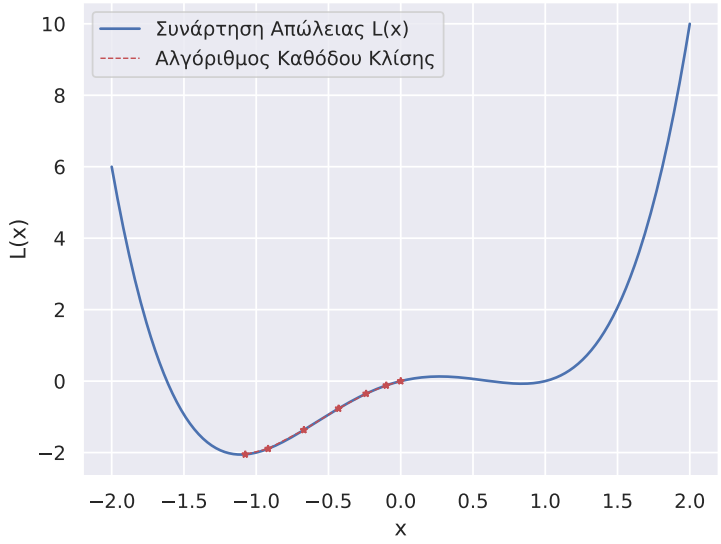
- Η (mean square error loss): $\mathsf{L}(\hat{Y}, Y) = \frac{1}{M}\sum_{i=1}^{M} \left\|y^{(i)} - \hat{y}^{(i)}\right\|^2$
  στην περίπτωση προβλήματος παλινδρόμησης (όπου $\|x\|$ συνήθως είναι η $L2$ νόρμα).

Να σημειώσουμε ότι ορίσαμε τη συνάρτηση απώλειας ώστε να δέχεται ένα σύνολο από προβλέψεις και επιθυμητές τιμές στόχους. Αντίστοιχα, θα μπορούσαμε να ορίσουμε συνάρτηση απώλειας που συγκρίνει την πρόβλεψη με την έξοδο στόχο ενός συγκεκριμένου παραδείγματος (και να μην αθροίζει όλα τα παραδείγματα). Τότε, θα είχαμε $L(\hat{y}, y) = L(\mathcal{F}(X; \overline{W}, \overline{b}), y)$.

Έχοντας στη διάθεσή μας μια μετρική απόδοσης, το πρόβλημα της εκπαίδευσης του νευρωνικού δικτύου μπορεί να διατυπωθεί ως πρόβλημα βελτιστοποίησης. Με μαθηματικούς όρους έχουμε:

$$\overline{W^*}, \overline{b^*} = \underset{\overline{W}, \overline{b}}{\operatorname{argmax}}(\mathsf{L}(\mathcal{F}(\boldsymbol{X}; \overline{W}, \overline{b}), Y)) \tag{2.11}$$

Εξετάζοντας τη μετρική L σαν συνάρτηση των $\overline{W}$ και $\overline{b}$, για την επίλυση του ανωτέρω προβλήματος αρκεί να βρούμε το σημείο στον χώρο των παραμέτρων που την ελαχιστοποιεί. Λόγω των μη γραμμικών στοιχείων όμως, δεν υπάρχει κλειστός τύπος (closed form) για την εύρεση του σημείου αυτού. Όπως φαίνεται και από το σχήμα 2.5, η συνάρτηση απώλειας είναι μη κυρτή. Αυτό μας οδηγεί στη χρήση επαναληπτικών μεθόδων για την εύρεση κάποιου (τοπικού) ελάχιστου.



Σχήμα 2.5: Γραφική παράσταση στην οποία εφαρμόζεται ο αλγόριθμος καθόδου κλίσης σε μια μη κυρτή συνάρτηση με μια παράμετρο (την x).

Να προσθέσω λινκ σε κώδικα.

Παράχθηκε τοπικά.

Ο πιο δημοφιλής αλγόριθμος για αυτόν τον σκοπό είναι ο αλγόριθμος καθόδου κλίσης (gradient descent). Σύμφωνα με τον αλγόριθμο αυτό, πραγματοποιούνται επαναλαμβανόμενα βήματα «καθόδου» προς την κατεύθυνση με τη μεγαλύτερη κλίση. Διαισθητικά, φαίνεται λογικό σε κάθε βήμα να υπολογίζουμε σημειακά την κλίση της συνάρτησης που θέλουμε να ελαχιστοποιήσουμε και να κινούμαστε προς την κατεύθυνση με τη μικρότερη κλίση. Πιο συγκεκριμένα, πρώτα αρχικοποιούνται

ανεξάρτητα όλες οι παράμετροι σε τυχαίες τιμές $\overline{W}_0$, $\overline{b}_0$ και έπειτα ξεκινά μια επαναληπτική διαδικασία όπου σε κάθε βήμα αυτής (έστω i):

1. Υπολογίζονται οι μερικοί παράγωγοι (η κλήση) της συνάρτησης απώλειας ως προς όλες τις παραμέτρους σημειακά:

$$dw_i = \left.\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial w}\right|_{(\overline{W},\overline{b})=(\overline{W}_{i-1},\overline{b}_i-1)}, \forall w \tag{2.12}$$

και

$$db_i = \left.\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial b}\right|_{(\overline{W},\overline{b})=(\overline{W}_{i-1},\overline{b}_i-1)}, \forall b \tag{2.13}$$

Όπου $\mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})$ η συνάρτηση απώλειας υπολογισμένη για ένα σύνολο δεδομένων $\boldsymbol{X}$ υπό τις παραμέτρους $\overline{W}, \overline{b}$. Συνοπτικά, αν συγκεντρώσουμε όλες τις παραμέτρους $\overline{W}$ και $\overline{b}$ στο διάνυσμα στήλη $\overline{W_{all}}$ τότε γράφουμε:

$$d\overline{W_{all}}_i = \nabla \mathsf{L}(\overline{\mathsf{W}_{\mathsf{all}}})\big|_{\overline{W_{all}}=\overline{W_{all}}_{i-1}} \tag{2.14}$$

2. Μετακινείται το σημείο στον χώρο παραμέτρων προς την κατεύθυνση της μεγαλύτερης κλίσης σύμφωνα με τον κανόνα ενημέρωσης (update rule) των παραμέτρων[12]. Ο κανόνας είναι ο εξής:

$$w_i = w_{i-1} - \alpha \times dw_i, \forall w \tag{2.15}$$

και

$$b_i = b_{i-1} - \alpha \times db_i, \forall b \tag{2.16}$$

Όπου α ο ρυθμός μάθησης (learning rate): μια υπερπαράμετρος που καθορίζει το μέγεθος του βήματος κατά την ενημέρωση των παραμέτρων. Αντίστοιχα με πριν, συνοπτικά, έχουμε:

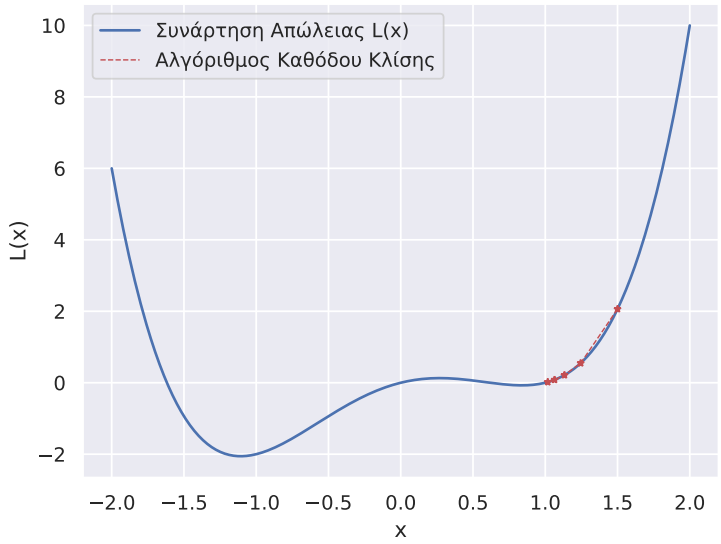
$$\overline{W_{all}}_i = \overline{W_{all}}_{i-1} - \alpha \times d\overline{W_{all}}_i \tag{2.17}$$

Ο αλγόριθμος τελειώνει είτε όταν οι ενημερώσεις είναι πλέον αμελητέες και η τιμή της συνάρτησης απώλειας δε μειώνεται άλλο από επανάληψη σε επανάληψη (ο αλγόριθμος έχει βρει ένα τοπικό ελάχιστο) είτε όταν ξεπεραστεί ο μέγιστος αριθμός επαναλήψεων.

Να σημειώσουμε ότι ο αλγόριθμος καθόδου κλίσης δε βρίσκει πάντα το ολικό ελάχιστο της συνάρτησης. Ανάλογα με την αρχικοποίηση των παραμέτρων και τις τιμές των υπερπαραμέτρων (ρυθμός μάθησης, αριθμός επαναλήψεων) οδηγούμαστε κάθε φορά σε διαφορετικά αποτελέσματα. Για παράδειγμα, στο σχήμα 2.6 αρχικοποιήσαμε την τιμή της παραμέτρου x με την τιμή $+1.5$ με αποτέλεσμα ο αλγόριθμος να τερματίσει στο τοπικό ελάχιστο της συνάρτησης. Ανεξάρτητα από τον αριθμό επαναλήψεων, δε θα εντόπιζε ποτέ το ολικό ελάχιστο υπό αυτήν την αρχικοποίηση. Η αδυναμία εγγύησης για την εύρεση του ολικού ελαχίστου αποτελεί τον λόγο για τον οποίο σε αρκετά προβλήματα μηχανικής μάθησης οι αλγόριθμοι εκπαιδεύονται πολλές φορές με διαφορετική όμως αρχικοποίηση των παραμέτρων τους. Ευτυχώς, στα πολυεπίπεδα νευρωνικά δίκτυα δεν ενδιαφέρει η εύρεση του ολικού ελαχίστου[13] αλλά ενός τοπικού ελαχίστου [;].

[12] Φανταστείτε τις παραγώγους ως «ευθύνες» της κάθε παραμέτρου για τις σωστές ή λάθος προβλέψεις. Όσο πιο μεγάλη η παράγωγος, τόσο πιο καθοριστικό ρόλο παίζει η μεταβλητή στη διαμόρφωση της τιμής της συνάρτησης απώλειας.

[13] Καθώς οδηγεί σε overfitting.



Σχήμα 2.6: Γραφική παράσταση στην οποία εφαρμόζεται ο αλγόριθμος καθόδου κλίσης σε μια μη κυρτή συνάρτηση με μια παράμετρο (την x) αρχικοποιημένη όμως στην τιμή 1.5 με αποτέλεσμα να βρίσκεται ένα τοπικό ελάχιστο.

Ο αλγόριθμος της καθόδου κλίσης δε θα χρησίμευε στην εκπαίδευση των νευρωνικών δικτύων αν δεν υπήρχε η δυνατότητα αποδοτικού υπολογισμού των μερικών παραγώγων. Ευτυχώς, τη λειτουργία αυτή την επιτελεί η μέθοδος οπισθοδιάδοσης σφάλματος (back propagation). Με λίγα λόγια, πρόκειται για μια μέθοδο η οποία χρησιμοποιώντας τον κανόνα της αλυσίδας υπολογίζει την παράγωγο της συνάρτησης απώλειας ως προς όλες τις παραμέτρους του δικτύου (σημειακά), ξεκινώντας από αυτές του τελευταίου επιπέδου και τερματίζοντας σε αυτές του πρώτου.

Παρακάτω παρατίθενται οι υπολογισμοί που λαμβάνουν χώρα κατά τη διάρκεια εύρεσης των μερικών παραγώγων ως προς τις παραμέτρους ενός επιπέδου $L-1$ μέσω της οπισθοδιάδοσης σφάλματος για την i–οστή επανάληψη του αλγορίθμου καθόδου κλίσης. Αν και οι παράγωγοι υπολογίζονται σημειακά για τις τιμές των παραμέτρων $\overline{W}_i$ και $\bar{b}_i$, για λόγους ευκολότερης ανάγνωσης αυτό δε θα απεικονίζεται κατά τη διατύπωση των παρακάτω μερικών παραγώγων. Ξεκινώντας από το επίπεδο L έχουμε:

- Η παράγωγος της συνάρτησης απώλειας ως προς τα βάρη από τον κόμβο k του επιπέδου $L-1$ στον κόμβο j του επιπέδου L είναι:

$$\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \bar{\mathsf{b}})}{\partial w_{jk}^{[L]}} = \frac{\partial \mathsf{z}_{\mathsf{j}}^{[\mathsf{L}]}}{\partial w_{jk}^{[L]}} \times \frac{\partial \mathsf{a}_{\mathsf{j}}^{[\mathsf{L}]}}{\partial z_j^{[L]}} \times \frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \bar{\mathsf{b}})}{\partial a_j^{[L]}} \tag{2.18}$$

  Όπου ο όρος $\frac{\partial \mathsf{L}(\overline{\mathsf{W}},\bar{\mathsf{b}})}{\partial a_j^{[L]}}$ υπολογίζεται άμεσα από την επιλεγμένη συνάρτηση απώλειας.

  Ο όρος $\frac{\partial \mathsf{a}_{\mathsf{j}}^{[\mathsf{L}]}}{\partial z_j^{[L]}}$ υπολογίζεται άμεσα από την επιλεγμένη συνάρτηση ενεργοποίησης.

  Τέλος, η μερική παράγωγος $\frac{\partial \mathsf{z}_{\mathsf{j}}^{[\mathsf{L}]}}{\partial w_{jk}^{[L]}}$ υπολογίζεται λαμβάνοντας την παράγωγο του γραμμικού συνδυασμού

- Η παράγωγος της συνάρτησης απώλειας ως προς τα δυναμικά πόλωσης είναι:

$$\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial b_j^{[L]}} = \frac{\partial \mathsf{z}_\mathsf{j}^{[\mathsf{L}]}}{\partial b_j^{[L]}} \times \frac{\partial \mathsf{a}_\mathsf{j}^{[\mathsf{L}]}}{\partial z_j^{[L]}} \times \frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial a_j^{[L]}} \tag{2.19}$$

  Στην περίπτωση αυτή, η μερική παράγωγος $\frac{\partial \mathsf{z}_\mathsf{j}^{[\mathsf{L}]}}{\partial b_j^{[L]}} = 1$.

- Τέλος, η παράγωγος της συνάρτησης απώλειας ως προς τις τιμές ενεργοποίησης του προηγούμενου επιπέδου $a_k^{[L-1]}$ είναι:

$$\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial a_k^{[L-1]}} = \sum_{j=1}^{n^{[L]}} \frac{\partial \mathsf{z}_\mathsf{j}^{[\mathsf{L}]}}{\partial a_k^{[L-1]}} \times \frac{\partial \mathsf{a}_\mathsf{j}^{[\mathsf{L}]}}{\partial z_j^{[L]}} \times \frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial a_j^{[L]}} \tag{2.20}$$

Παρατηρούμε ότι οι μερικοί παράγωγοι των μεταβλητών ενός επιπέδου $l-1$ εξαρτώνται από το επίπεδο l. Για αυτό και όπως προαναφέραμε, οι υπολογισμοί ξεκινούν από το τελευταίο επίπεδο. Επαγωγικά, με τη χρήση της 2.20 στις 2.18 και 2.19 μπορούμε να βρούμε τις μερικές παραγώγους ως προς τις παραμέτρους όλων των επιπέδων.

Συγκεντρωτικά, χρησιμοποιώντας την αναπαράσταση με χρήση πίνακα που παρουσιάσαμε στην προηγούμενη παράγραφο, οι σχέσεις 2.18, 2.19 και 2.20 γράφονται αντίστοιχα:

$$\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial \boldsymbol{W}^{[l]}} = \frac{1}{M} \times \Big(\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial \boldsymbol{A}^{[l]}} \odot \frac{\partial \boldsymbol{A}^{[\mathsf{l}]}}{\partial \boldsymbol{Z}^{[l]}}\Big) \times \boldsymbol{A}^{[l-1]^T} \tag{2.21}$$

$$\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial \boldsymbol{b}^{[l]}} = \frac{1}{M} \times \Big(\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial \boldsymbol{A}^{[l]}} \odot \frac{\partial \boldsymbol{A}^{[\mathsf{l}]}}{\partial \boldsymbol{Z}^{[l]}}\Big) \times \mathbf{1}_M^T \tag{2.22}$$

$$\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial \boldsymbol{A}^{[l-1]}} = \boldsymbol{W}^{[l]^T} \times \Big(\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial \boldsymbol{A}^{[l]}} \odot \frac{\partial \boldsymbol{A}^{[\mathsf{l}]}}{\partial \boldsymbol{Z}^{[l]}}\Big) \tag{2.23}$$

Όπου ο τελεστής $\odot$ συμβολίζει το γινόμενο στοιχείο προς στοιχείο (elementwise product) ενώ το $\mathbf{1}_n = \big[1, 1, 1, \ldots, 1\big] \in \Re^n$. Για τον όρο στην παρένθεση που συναντάται συχνά στους ανωτέρω τύπους ισχύει $\big(\frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial \boldsymbol{A}^{[l]}} \odot \frac{\partial \boldsymbol{A}^{[\mathsf{l}]}}{\partial \boldsymbol{Z}^{[l]}}\big) = \frac{\partial \mathsf{L}(\overline{\mathsf{W}}, \overline{\mathsf{b}})}{\partial \boldsymbol{Z}^{[l]}}$.

Σαν τελικά σχόλια σχετικά με την εκπαίδευση των νευρωνικών δικτύων είναι ωφέλιμο να κάνουμε δύο παρατηρήσεις:

- Κατά την εφαρμογή του αλγορίθμου καθόδου κλίσης σε ένα νευρωνικό δίκτυο, σε κάθε βήμα αυτού γίνονται δύο περάσματα: μια πρόσθια διάδοση που περιγράφεται από τις εξισώσεις 2.9 και 2.10 για τον υπολογισμό της συνάρτησης απώλειας και μια οπισθοδιάδοση που περιγράφεται από τις εξισώσεις 2.21, 2.22 και 2.23 για τον υπολογισμό των παραγώγων που χρησιμοποιούνται στον κανόνα ενημέρωσης.

- Επειδή το σύνολο δεδομένων εισόδου μπορεί να είναι πολύ μεγάλο, αντί να λαμβάνονται όλα τα παραδείγματα M για τον υπολογισμό του $d\overline{W_{all}}$ με βάση τη συνάρτησης απώλειας, συνηθίζεται να χωρίζεται σε μικρά πακέτα (mini batches) από m παραδείγματα το καθένα. Έτσι, πραγματοποιείται ένα βήμα ενημέρωσης για κάθε μικρό πακέτο δεδομένων. Όταν εφαρμόζεται αυτή η τακτική, σιωπηρά γίνεται η υπόθεση ότι το κάθε δείγμα των m παραδειγμάτων είναι επαρκώς αντιπροσωπευτικό ώστε η συνάρτηση απώλειας υπολογισμένη στα m παραδείγματα

να είναι καλή προσέγγιση της συνάρτησης υπολογισμένης στα M παραδείγματα. Ακραία μορφή αυτού είναι ο στοχαστικός αλγόριθμος καθόδου κλίσης (stochastic gradient descent) στον οποίο $m = 1$. Οι μαθηματικοί τύποι που παραθέσαμε σε αυτό το κεφάλαιο ισχύουν σε κάθε περίπτωση μετά την κατάλληλη ανάθεση της υπερπαραμέτρου M.

Συνελικτικά Νευρωνικά Δίκτυα

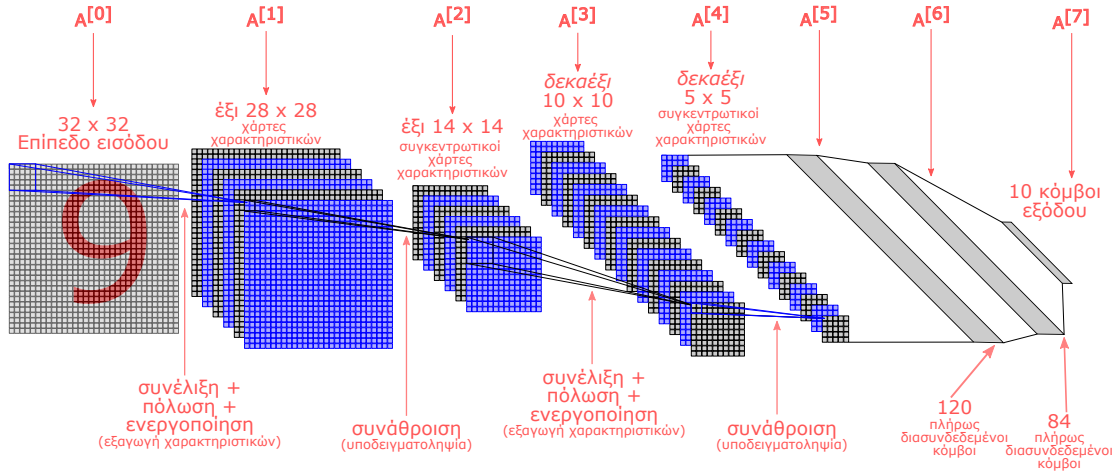
Έχοντας περιγράψει τη δομή και την εκπαίδευση των απλών νευρωνικών δικτύων πρόσθιας διάδοσης, εύκολα μπορούμε να κατανοήσουμε μερικές από τις παραλλαγές του. Μια από τις σημαντικότερες, είναι αυτή των Συνελικτικών Νευρωνικών Δικτύων (Convolutional Neural Networks) που χρησιμοποιείται συστηματικά στον χώρο της όρασης υπολογιστών. Πρόκειται για την υποκατηγορία των νευρωνικών δικτύων πρόσθιας διάδοσης στην οποία οδηγήθηκε η επιστημονική κοινότητα αφενός επιδιώκοντας να λύσει ορισμένα από τα πρακτικά προβλήματα της εφαρμογής νευρωνικών δικτύων στον χώρο της όρασης υπολογιστών και αφετέρου μελετώντας τη νεύρο–φυσιολογία του οπτικού φλοιού (visual cortex).

Από τη σκοπιά της νευροεπιστήμης, οι David H. Hubel και Torsten Wiesel μετά από μια σειρά πειραμάτων σε γάτες [;,;] γύρω στο 1960 και αργότερα, σε πιθήκους [;] έριξαν φως στη δομή του οπτικού φλοιού, εμπνέοντας έτσι το κίνημα του διασυνδετισμού (connectionism). Σύμφωνα με το έργο τους, (για το οποίο τιμήθηκαν με το βραβείο nobel το 1981) πολλοί νευρώνες του οπτικού φλοιού έχουν μικρά, τοπικά πεδία υποδοχής (receptive fields) που μπορεί να επικαλύπτονται μεταξύ τους. Πιο συγκεκριμένα, ο κάθε νευρώνας αφορά ένα περιορισμένο τμήμα του οπτικού πεδίου αλλά όλοι μαζί, καλύπτουν το σύνολό του. Επιπλέον, μετά από πειράματα οπτικής αναγνώρισης σχημάτων (ορθογώνιο παραλληλόγραμμο σε μορφή μπάρας) σε διάφορες γεωμετρίες παρατηρήθηκε ότι διαφορετικοί νευρώνες με το ίδιο πεδίο υποδοχής ενεργοποιούνται ανάλογα με τη γεωμετρία του σχήματος (κάποιοι νευρώνες ενεργοποιούνται κατά τον κάθετο προσανατολισμό της μπάρας ενώ άλλοι με τον οριζόντιο προσανατολισμό της). Τέλος, επισήμαναν ότι ορισμένοι νευρώνες ενεργοποιούνται με την αναγνώριση πιο περίπλοκων μοτίβων όπως προκύπτουν από τη σύνθεση απλών γεωμετριών χαμηλότερου επιπέδου [;].

Από πρακτικής σκοπιάς, η υπολογιστική πολυπλοκότητα που προκύπτει από την τροφοδότηση ενός πλήρως διασυνδεδεμένου νευρωνικού δικτύου με εικόνες είναι απαγορευτική. Για παράδειγμα, έστω ότι διατίθεται ένα σύνολο από ασπρόμαυρες εικόνες μεγέθους 100×100 εικονοστοιχεία. Αυτό συνεπάγεται ότι το επίπεδο εισόδου θα διαθέτει τόσους κόμβους όσα είναι και τα χαρακτηριστικά (τα εικονοστοιχεία), δηλαδή 10.000[14]. Στην κλασσική περίπτωση, ο αριθμός των κόμβων του πρώτου κρυφού επιπέδου θα είναι περίπου ίσος με τον αριθμό των χαρακτηριστικών εισόδου, δηλαδή πάλι 10.000. Αυτό σημαίνει ότι μόνο στο πρώτο επίπεδο του νευρωνικού δικτύου θα υπήρχαν 10.000×10.000 βάρη και 10.000 δυναμικά πόλωσης, ένας αριθμός, πολύ μεγάλος. Θα μπορούσαμε, φυσικά, αντί να τροφοδοτήσουμε το νευρωνικό δίκτυο με ολόκληρη την εικόνα σε ακατέργαστη μορφή, να εξάγαμε με ντετερμινιστικό τρόπο ορισμένα χαρακτηριστικά ώστε να καταλήξουμε με ένα μικρότερο διάνυσμα χαρακτηριστικών που θα εσωκλείει όλη τη χρήσιμη πληροφορία. Μια τέτοια διαδικασία χειρονακτικής εξαγωγής χαρακτηριστικών απεικονίζεται στο σχήμα 2.2. Παρόλα αυτά, για τους λόγους που αναφέραμε στην ενότητα 2.1.2 θα επιθυμούσαμε

[14]Χρησιμοποιούμε κόμμα για τη διάκριση των δεκαδικών (decimal comma separator) και τελεία για τη διάκριση των χιλιάδων.

την αυτοματοποιημένη εκμάθησή χαρακτηριστικών.



Σχήμα 2.7: Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου (LeNet-5) [;]. Η αναπαράστασή τους διαφέρει από αυτήν των απλών νευρωνικών δικτύων αφού εδώ δίνεται έμφαση στους πίνακες τιμών ενεργοποίησης $A^{[l]}$. Οι χάρτες χαρακτηριστικών αναπαριστώνται με τετράγωνα ενώ τα βάρη με ακμές. Τα δύο τελευταία επίπεδα είναι πλήρως διασυνδεδεμένα. *Παράχθηκε από το Inkscape τροποποιώντας αυτήν την εικόνα.*

Η λύση στο πρόβλημα δόθηκε μέσω της αξιοποίησης της τοπικής χωρικής συνεκτικότητας (local spacial coherence) και της ιεραρχικής δομής (hierarchical structure) των δεδομένων εικόνων. Εμπνευσμένη από τις ανωτέρω επιστημονικές παρατηρήσεις, ενσωματώθηκε η γνώση του χώρου προβλημάτων με εικόνες στη δομή των νευρωνικών δικτύων οδηγώντας έτσι στη δημιουργία των συνελικτικών νευρωνικών δικτύων (βλ. σχήμα 2.7). Οι δομικές διαφορές των συνελικτικών νευρωνικών δικτύων που τους διακρίνουν από τα νευρωνικά δίκτυα που παρουσιάστηκαν στην προηγούμενη ενότητα μπορούν να συνοψιστούν ως εξής:

- Μια πρώτη διαφορά που επιλύει το πρόβλημα της απαγορευτικής υπολογιστικής πολυπλοκότητας έγκειται στο τρόπο διασύνδεσης των κόμβων ενός επιπέδου με τους κόμβους του αμέσως προηγούμενου. Αντί να είναι πλήρως διασυνδεδεμένοι με αυτούς του προηγούμενου επιπέδου όπως στην περίπτωση των απλών νευρωνικών δικτύων, ενώνονται με βάρη μόνο με αυτούς που ανήκουν στο λεγόμενο πεδίο υποδοχής. Με άλλα λόγια, κάθε νευρώνας επιπέδου l δέχεται σαν είσοδο ενα διαφορετικό και περιορισμένο τμήμα του πίνακα $A^{[l-1]}$.

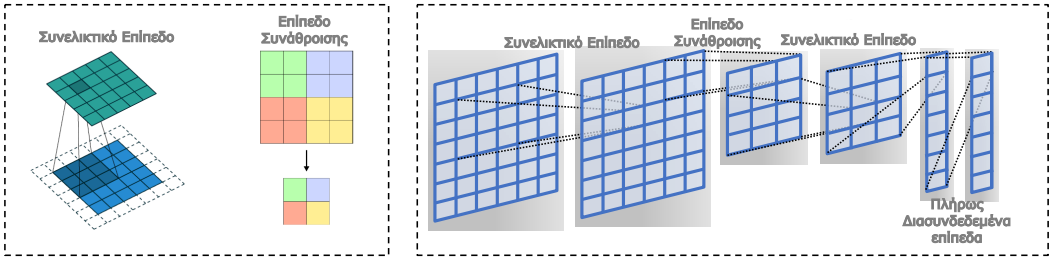
- Στα συνελικτικά νευρωνικά δίκτυα, οι κόμβοι του κάθε επιπέδου είναι οργανωμένοι σε όγκους τριών διαστάσεων με πλάτος, ύψος και βάθος. Με αυτόν τον τρόπο, διατηρείται η τοπική χωρική συνεκτικότητα. Αναλυτικότερα, οι κόμβοι εισόδου, για παράδειγμα, οργανώνονται όπως τα εικονοστοιχεία σε μια εικόνα: το βάθος του επιπέδου αντιστοιχεί στον αριθμό των καναλιών της εικόνας (π.χ. RGB) ενώ το ύψος και το πλάτος του επιπέδου στο ύψος και πλάτος της εικόνας. Έτσι, το νευρωνικό δίκτυο έχει τη δυνατότητα να αντλήσει εύκολα πληροφορία από μια χωρική γειτονιά της εικόνας (το πεδίο υποδοχής κάποιου νευρώνα) αφού οι αποστάσεις μεταξύ των εικονοστοιχείων διατηρούνται αναλλοίωτες. Αν όμως λαμβάναμε την εικόνα και την αναπτύσσαμε σε μια διάσταση (flatten) σημιουργόντας

ένα μεγάλο διάνυσμα, τότε οι σχετικές αποστάσεις των στοιχείων εισόδου δε θα διατηρούνταν. Ανάλογες παρατηρήσεις ισχύουν και για τα κρυφά επίπεδα. Δηλαδή, και στα επόμενα επίπεδα οι κόμβοι οργανώνονται σε τρισδιάστατες δομές οι οποίες διατηρούν τοπικό χαρακτήρα. Η διαφορά έγκειται στο ότι η τιμή των κόμβων των κρυφών επιπέδων δεν είναι η τιμή του εκάστοτε εικονοστοιχείου στη θέση αυτή. Αντίθετα, είναι η τιμή ενός (σύνθετου) χαρακτηριστικού της περιοχής που έχει υπολογιστεί από την επεξεργασία απλούστερων χαρακτηριστικών προηγούμενων επιπέδων. Σχετικά με την ιδιότητα της ιεραρχικής δομής των εικόνων του πραγματικού κόσμου, αυτή αξιοποιείται μέσω διαδοχικών κρυφών επιπέδων που σταδιακά διευρύνουν το οπτικό πεδίο και συνθέτουν ολοένα και πιο σύνθετα χαρακτηριστικά. Έχει δειχθεί, ότι το σύστημα μαθαίνει να εξάγει μέσω των πρώτων επιπέδων απλά χαρακτηριστικά (π.χ. οριζόντιες και κάθετες ακμές) τα οποία σε επόμενα επίπεδα συνδυάζει για να εξάγει πιο περίπλοκα χαρακτηριστικά **;;**. Αφού οι κόμβοι είναι οργανωμένοι σε όγκους, προκύπτει φυσικά ότι το διάνυσμα ενεργοποίησης $A^{[l]}$ του κάθε επιπέδου l που κατασκευάζεται από την έξοδο κάθε κόμβου έχει τη μορφή πίνακα τριών διαστάσεων. Διαισθητικά για τα κρυφά επίπεδα, το ύψος και το πλάτος του διανύσματος ενεργοποίησης κωδικοποιούν αμυδρά τη θέση του χαρακτηριστικού στην εικόνα ενώ το βάθος κωδικοποιεί τα διάφορα χαρακτηριστικά (π.χ. βάθος 1: οριζόντιες ακμές, βάθος δύο: κατακόρυφες ακμές[15]). Στην περίπτωση των δικτύων που εξετάζουμε, το $A^{[l]}$ λέμε ότι αποτελείται από επίπεδα φύλλα τα οποία στοιβάζονται στη διάσταση z και ονομάζονται χάρτες χαρακτηριστικών (feature maps).

- Μια ακόμα δομική διαφορά είναι ότι στα συνελικτικά νευρωνικά δίκτυα η εκμάθηση και εξαγωγή των χαρακτηριστικών δε γίνεται ανεξάρτητα σε κάθε περιοχή της εικόνας. Νευρώνες των οποίων τα βάρη προσαρμόζονται ώστε να αναγνωρίζουν και να εξάγουν γενικά χαρακτηριστικά της εικόνας όπως τα χαρακτηριστικά ακμών θα ήταν ασύμφορο να είχαν εφαρμογή μόνο στο πεδίο υποδοχής τους και όχι σε όλη την εικόνα. Έτσι, οδηγούμαστε στην έννοια του διαμοιρασμού παραμέτρων (weight sharing). Σύμφωνα με αυτήν την έννοια, αντί οι κόμβοι ενός επιπέδου να είναι διασυνδεδεμένοι με τους κόμβους του προηγούμενου επιπέδου με ξεχωριστά βάρη και δυναμικά πόλωσης, οι παράμετροι αυτές μοιράζονται μεταξύ των κόμβων. Έτσι, οι κόμβοι ενός επιπέδου επιτελούν τον ίδιο γραμμικό συνδυασμό $y = f\Big(\sum_{i=1}^{n} w_i \times x_i + b\Big)$ αλλά με διαφορετικό διάνυσμα εισόδου X που εξαρτάται από το οπτικό πεδίο. Σημειώνουμε δε ότι ο διαμοιρασμός βαρών θα ήταν δύσκολο να εφαρμοστεί στην περίπτωση που η είσοδος αποτελούνταν από δομημένα δεδομένα καθώς αυτά μπορεί να είχαν πλήρως ετερογενή χαρακτηριστικά.

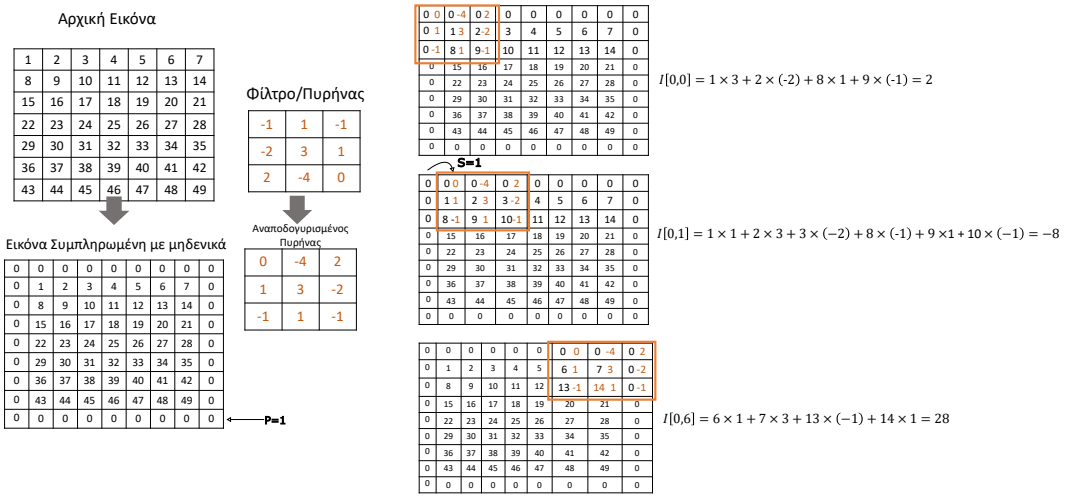
Πρακτικά, αν εξαιρέσουμε τα τελευταία, πλήρως διασυνδεδεμένα επίπεδα ενός συνελικτικού νευρωνικού δικτύου, οι ανωτέρω δομικές διαφορές υλοποιούνται με τη χρήση αφενός των συνελικτικών επιπέδων και αφετέρου των επιπέδων υποδειγματοληψίας. Αναφορικά με τα πρώτα, η εσωτερική τους λειτουργία απεικονίζεται στο αριστερό τμήμα του σχήματος 2.8. Το οπτικό πεδίο ανπαρίσταται με ένα σκούρο παραλληλόγραμμο επάνω στον χάρτη χαρακτηριστικών του προηγούμενου επιπέδου. Τα βάρη, είναι ευκολότερο να τα φανταστεί κανείς σαν ένα παραλληλόγραμμο (ή ένα ορθογωνικό κυβοειδές σε τρεις διαστάσεις) το οποίο έχει τις ίδιες διαστάσεις με το οπτικό πεδίο πάνω στο

[15]Πρακτικά στη διαδικασία εκμάθησης χαρακτηριστικών είναι δύσκολο να εκφράσουμε με σαφήνεια τι αναπαριστά το καθένα.



Σχήμα 2.8: Μεμονωμένο συνελικτικό επίπεδο και επίπεδο υποδειγματοληψίας (αριστερά). Συνδυασμός των επιπέδων για την κατασκευή ενός συνελικτικού νευρωνικού δικτύου (δεξιά). *Παράχθηκε από το Inkscape τροποποιώντας αυτήν την εικόνα.*

προηγούμενο επίπεδο. Η τιμή ενεργοποίησης κάθε στοιχείου του τρισδιάστατου πίνακα $A^{[l]}$ υπολογίζεται ως το αποτέλεσμα της εφαρμογής της συνάρτησης ενεργοποίησης στον γραμμικό συνδυασμό των στοιχείων του πίνακα $A^{[l-1]}$ που βρίσκονται εντός του οπτικού πεδίου με βάρη τα στοιχεία του $W^{[l]}$ και το δυναμικό πόλωσης $b^{[l]}$. Στην ουσία, τα βάρη υπερτίθενται στο οπτικό πεδίο και επιτελείται γινόμενο μεταξύ των πινάκων στοιχείο προς στοιχείο (elementwise product). Αν υπήρχε ένας πίνακας από βάρη για κάθε στοιχείο του $A^{[l]}$, τότε δε θα είχαμε διαμοιρασμό βαρών. Αντιθέτως, ο διαμοιρασμός βαρών έγκειται στην ολίσθηση αυτού του παραλληλόγραμμου (ή ορθογωνικού κυβοειδούς στις τρεις διαστάσεις) στο ύψος και πλάτος της εικόνας, όπως φαίνεται και στο σχήμα 2.9 για την περίπτωση των δύο διαστάσεων. Για την περίπτωση που έχουμε πολλούς χάρτες χαρακτηριστικών, παραπέμπουμε τον αναγνώστη στο σχήμα 2.10. Αυτή η διαδικασία της ολίσθησης του πίνακα βαρών ονομάζεται δισδιάστατη συνέλιξη[16], ενώ το κυλιόμενο παράθυρο ονομάζεται και φίλτρο (filter) ή πυρήνας (kernel).



Σχήμα 2.9: Παράδειγμα συνέλιξης σε δύο διαστάσεις *Παράχθηκε από το Inkscape τροποποιώντας και διορθώνοντας αυτήν την εικόνα.*

Με κάθε συνελικτικό επίπεδο εισάγεται μια σειρά από παραμέτρους πέρα από αυτές που υπήρχαν σε κάθε απλό νευρωνικό δίκτυο. Για κάθε επίπεδο l, μεταξύ δύο στοιβαγμένων συνόλων

[16]Τυπικά, η πράξη ονομάζεται διασταυρούμενη συσχέτιση (cross corelation) και είναι ίδια με τη συνέλιξη αν στην πρώτη περίπτωση αναποδογυρίσουμε τον πυρήνα (γυρνώντας τον ως προς την κύρια και δευτερεύουσα διαγώνιο).

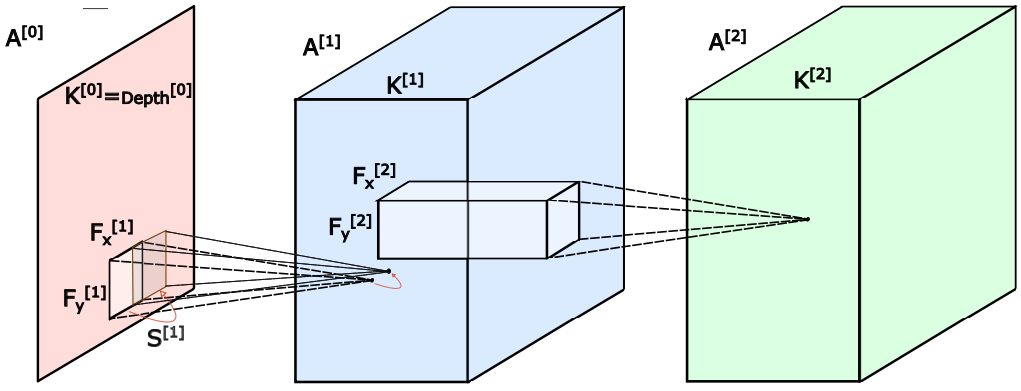
από χάρτες χαρακτηριστικών $A^{[l-1]}$ και $A^{[l]}$ με διαστάσεις $Width^{[l-1]} \times Height^{[l-1]} \times Depth^{[l-1]}$ και $Width^{[l]} \times Height^{[l]} \times Depth^{[l]}$ αντίστοιχα πρέπει να ορίσουμε:

- Το μέγεθος του κυλιόμενου πυρήνα (ή φίλτρου). Αυτό εισαγάγει τις παραμέτρους ${F_x}^{[l]}$ και ${F_y}^{[l]}$. Το βάθος του φίλτρου δεν αποτελεί υπερπαράμετρο και είναι ίσο με $Depth^{[l-1]}$.

- Ο αριθμός των φίλτρων, $K^{[l]}$. Όπως έχει γίνει σαφές από τα ανωτέρω σχήματα, το κάθε φίλτρο $k \in [1, K^{[l]}]$ λαμβάνει σαν είσοδο έναν όγκο από χαρακτηριστικά και έχει ως έξοδο έναν αριθμό $a \in \Re$. Για να έχει ο παραγόμενος χάρτης χαρακτηριστικών βάθος, θα πρέπει να γίνει χρήση πολλαπλών φίλτρων που θα εξάγουν διαφορετικά χαρακτηριστικά. Έτσι, μπορούμε να καθορίσουμε το $Depth^{[l]}$ θέτοντας το $K^{[l]}$ αφού ισχύει ότι $K^{[l]} = Depth^{[l]}$. Για κάθε ένα φίλτρο παράγεται ένας χάρτης χαρακτηριστικών.

- Το βήμα της συνέλιξης, $s_x^{[l]}$ κατά τον x άξονα και $s_y^{[l]}$ κατά τον y άξονα (συνήθως, οι δύο ποσότητες είναι ίσες και συμβολίζονται ως $s^{[l]}$). Αυτή η ποσότητα καθορίζει το βήμα της ολίσθησης του φίλτρου πάνω στο χάρτη χαρακτηριστικών του επιπέδου $l-1$. Για παράδειγμα, μετά την εφαρμογή του στο άνω αριστερά άκρο της εισόδου στο σχήμα 2.9 θα μετακινηθεί με βήμα 1 μια θέση δεξιά (ή κάτω, αργότερα) για να υπολογίσει την επόμενη έξοδο. Στην περίπτωση που η τιμή του βήματος είναι ίση με το μέγεθος του φίλτρου (στον x ή y άξονα), τότε δεν υπάρχει επικάλυψη μεταξύ των οπτικών πεδίων. Μεγαλύτερη τιμή από αυτή δεν είναι επιθυμητή καθώς οδηγεί σε συστηματική αγνόηση στοιχείων του $A^{[l-1]}$.

- Το ποσό της περιμετρικής συμπλήρωσης της εισόδου $A^{[l-1]}$ με μηδενικά[17], $P_x^{[l]}$ και $P_y^{[l]}$. Αν και κάτι τέτοιο δεν είναι απαραίτητο, συνήθως γίνεται για τον έλεγχο των διαστάσεων ύψους και πλάτους του χάρτη χαρακτηριστικών της εξόδου (ή των χαρτών χαρακτηριστικών της εξόδου αν $K^{[l]} > 1$). Μετά από αυτήν τη διαδικασία, καταλήγουμε με ένα σύνολο χαρτών χαρακτηριστικών $\grave{A}^{[l-1]}$ του οποίου οι νέες διαστάσεις είναι $\grave{Width}^{[l-1]} = Width^{[l-1]} + 2 \times P^{[l]}$ και $\grave{Height}^{[l-1]} = Height^{[l-1]} + 2 \times P^{[l]}$



Σχήμα 2.10: Συνελικτικά επίπεδα στη σειρά με έμφαση στην περίπτωση όπου οι πίνακες A είναι τριών διαστάσεων. Στο σχήμα φαίνονται οι υπερπαράμετροι των συνελικτικών επιπέδων. *Παράχθηκε από το Inkscape τροποποιώντας αυτήν την εικόνα.*

Συγκεντρωτικά, το συνελικτικό επίπεδο, για την παραγωγή ενός συνόλου χαρτών χαρακτηριστικών

[17] Υπάρχουν και άλλες επιλογές στον τρόπο περιμετρικής συμπλήρωσης όπως αυτός που περιλαμβάνει ανάκλαση αλλά η χρήση μηδενικών είναι η πιο συνηθισμένη μέθοδος και για αυτό εστιάζουμε σε αυτή.

εφαρμόζει τις εξής διαδικασίες:

1. Συνέλιξη πάνω στους χάρτες χαρακτηριστικών $\grave{A}^{[l-1]}$ (αφού έχουν ίσως συμπληρωθεί με μηδενικά) με τον (τρισδιάστατο) πίνακα βαρών $W_k^{[l]}$.

2. Σημειακή πρόσθεση της τιμής πόλωσης ($b_k^{[l]}$) σε κάθε στοιχείο του προηγούμενου πίνακα με αποτέλεσμα την παραγωγή ενός πίνακα $Z_k^{[l]}$ με τις ίδιες διαστάσεις.

3. Εφαρμογή συνάρτησης ενεργοποίησης $F^{[l]}$ σημειακά ώστε τελικά να παραχθεί ο χάρτης χαρακτηριστικών $A_k^{[l]}$ με διαστάσεις ίδιες με τον $Z_k^{[l]}$, δηλαδή:

$$Width^{[l]} = \frac{Width^{[l-1]} - {F_x}^{[l]} + 2 \times P_x^{[l]}}{S^{[l]}} + 1, \tag{2.24}$$

και

$$Height^{[l]} = \frac{Height^{[l-1]} - {F_y}^{[l]} + 2 \times P_y^{[l]}}{S^{[l]}} + 1 \tag{2.25}$$

4. Επανάληψη από το βήμα 1 $K^{[l]}$ φορές, όσο και το βάθος του $A^{[l]}$.

5. Στοίβαξη των παραχθέντων χαρτών χαρακτηριστικών ως προς τον άξονα z ώστε να κατασκευαστεί ο τρισδιάστατος πίνακας $A^{[l]}$. Τελικά, το σύνολο των χαρτών χαρακτηριστικών $A^{[l]}$ έχει διαστάσεις μήκους και πλάτους ίδιες με αυτές του $A_k^{[l]}$ αλλά το βάθος τώρα, αντί για μονάδα είναι:

$$Depth^{[l]} = K^{[l]} \tag{2.26}$$

Με μαθηματικούς όρους, οι υπολογισμοί που εκτελούνται σε ένα συνελικτικό επίπεδο είναι οι εξής:

$$Z_k^{[l]} = {W_k^{[l]}}^{T^\tau} \underset{step=S^{[l]}}{*} A^{[l-1]} + b_k^{[l]} \tag{2.27}$$

και

$$A_k^{[l]} = F^{[l]}(Z_k^{[l]}). \tag{2.28}$$

Όπου το σύμβολο τ στον εκθέτη ενός πίνακα δηλώνει την αναστροφή του πίνακα υπό τη δευτερεύουσα διαγώνιο ενώ το σύμβολο $\underset{step=S^{[l]}}{*}$ δηλώνει τη συνέλιξη με βήμα ίσο με $S^{[l]}$.

Ένα δεύτερο είδος επιπέδου αποτελεί αυτό της υποδειγματοληψίας (ή συνάθροισης) όπως φαίνεται στο σχήμα 2.8. Το συγκεκριμένο είδος δε διαθέτει καμία παράμετρο αφού η μόνη λειτουργία του είναι να πραγματοποιεί υποδειγματοληψία στο χάρτη χαρακτηριστικών. Ο τρόπος εφαρμογής του είναι παρόμοιος με αυτόν του συνελικτικού επιπέδου. Δηλαδή, και πάλι υπάρχει ένα κυλιόμενο παράθυρο πάνω στον χάρτη χαρακτηριστικών το οποίο συναθροίζει τα στοιχεία στα οποία υπερτίθεται σε ένα στοιχείο υπό μια προκαθορισμένη στρατηγική. Πιο συγκεκριμένα, ανάλογα με το αν επιλέγεται σαν έξοδος το μεγαλύτερο στοιχείο στη γειτονιά συνάθροισης (το οπτικό πεδίο) ή μια μέση τιμή αυτών, έχουμε τη μέγιστη συνάθροιση (max pooling) ή τη μέση συνάθροιση (average pooling) αντίστοιχα. Σε κάθε περίπτωση, μια διαφορά με τα συνελικτικά επίπεδα είναι ότι το κυλιόμενο παράθυρο υπερτίθεται σε κάθε «φύλλο» της εισόδου $A^{[l-1]}$ ξεχωριστά (δηλαδή, σε κάθε $A_k^{[l-1]}$). Με άλλα λόγια, παρόλο που ο πίνακας $A^{[l-1]}$ μπορεί να είναι τρισδιάστατος και να αποτελείται από πολλούς χάρτες χαρακτηριστικών στοιβαγμένους στον z άξονα, η γειτονιά συνάθροισης θα είναι πάντα ένα δισδιάστατο παράθυρο. Τέλος, να

σημειώσουμε ότι αυτό το επίπεδο συμβάλλει στην ευρωστία του συστήματος σε μικρές διακυμάνσεις της θέσης των αντικειμένων, μια ιδιότητα που θα αναλύσουμε περαιτέρω στην επόμενη ενότητα.

Σε κάθε επίπεδο συνάθροισης l έχουμε τις εξής υπερπαραμέτρους:

- Το μέγεθος του πυρήνα υποδειγματοληψίας ${Pk_x}^{[l]}$ και ${Pk_y}^{[l]}$. Καθορίζει τη γειτονιά συνάθροισης αλλά και το μέγεθος του συναθροισμένου χάρτη χαρακτηριστικών.
- Τη στρατιγική του επιπέδου συνάθροισης. Όπως αναφέραμε, εδώ οι στρατηγικές είναι δύο: μέγιστη συνάθροιση και μέση συνάθροιση.
- Το βηματισμό του κυλιόμενου παραθύρου $s_x^{[l]}$ κατά τον x άξονα και $s_y^{[l]}$ κατά τον y άξονα (όπως και στα συνελικτικά επίπεδα, συνήθως, οι δύο ποσότητες είναι ίσες και συμβολίζονται ως $s^{[l]}$). Ήθιστε, το βήμα να ισούται με το μέγεθος του πυρήνα.

Αναφορικά με το μέγεθος της εξόδου ενός επιπέδου συνάθροισης l, με είσοδο έναν χάρτη χαρακτηριστικών $A^{[l-1]}$ με διαστάσεις $Width^{[l-1]} \times Height^{[l-1]} \times Depth^{[l-1]}$ ισχύει:

$$Width^{[l]} = \frac{Width^{[l-1]} - {Pk_x}^{[l]}}{S^{[l]}} + 1, \tag{2.29}$$

$$Height^{[l]} = \frac{Height^{[l-1]} - {Pk_y}^{[l]}}{S^{[l]}} + 1 \tag{2.30}$$

και

$$Depth^{[l]} = Depth^{[l-1]} \tag{2.31}$$

Έχοντας καλύψει πλήρως τα νευρωνικά δίκτυα και την υποκατηγορία τους η οποία χρησιμοποιείται στην όραση υπολογιστών, είμαστε σε θέση να περιγράψουμε ένα νεότερο είδος νευρωνικών δικτύων για τον ίδιο σκοπό, τα λεγόμενα νευρωνικά δίκτυα με κάψουλες.

2.2 Νευρωνικά Δίκτυα με Κάψουλες

Τα τελευταία χρόνια, διερευνάται μια ακόμα παραλλαγή των νευρωνικών δικτύων για εφαρμογές όρασης υπολογιστών: αυτή των νευρωνικών δικτύων με κάψουλες (capsule networks). Η ιδέα πίσω από τη νέα αρχιτεκτονική παρουσιάστηκε από τον Geoffrey Hinton, το ίδιο άτομο που είχε συμβάλει καθοριστικά στην ανάπτυξη και εδραίωση του πρώτου συνελικτικού δικτύου ;;. Αυτή τη φορά όμως, στα σχετικά έργα του (ΤΟΔΟ) τονίζει ορισμένες αδυναμίες της εδραιωμένης, πλέον, τεχνολογίας ενώ προτείνει μια νέα αρχιτεκτονική που θα τις αντιμετωπίζει. Ένας έμπειρος αναγνώστης μπορεί να επισημάνει ότι η σύλληψη της ιδέας των νευρωνικών δικτύων με κάψουλες δεν είναι νέα (2011). Παρόλα αυτά, όπως θα διαπιστώσουμε στο κεφάλαιο 3 μόλις πρόσφατα άρχισε να λαμβάνει πρακτική υπόσταση με την ανάπτυξή σύνθετων αρχιτεκτονικών που πραγματώνουν τις βασικές ιδέες.

Στην ενότητα αυτή θα ξεκινήσουμε κάνοντας αναφορά σε ορισμένα στοιχεία του ανθρώπινου μηχανισμού αναγνώρισης προτύπων εικόνων που αποτέλεσαν πηγή έμπνευσης για τα νευρωνικά δίκτυα με κάψουλες. Στην συνέχεια θα αναφερθούμε στις αδυναμίες που παρουσιάζουν τα συνελικτικά νευρωνικά δίκτυα της προηγούμενης ενότητας και τις οποίες η νέα αρχιτεκτονική καλείται να ξεπεράσει. Τέλος, βασιζόμενοι στα κύρια έργα του Geoffrey Hinton σχετικά με τα

νευρωνικά δίκτυα με κάψουλες στο πλαίσιο επιβλεπόμενης μάθησης, θα εμβαθύνουμε στις αρχές λειτουργίας τους.

2.2.1 Στοιχεία Έμπνευσης των Νευρωνικών Δικτύων με Κάψουλες

Εδώ θα πεις για το biological motivation. Ο ανθρωπο πάντα κάνει ιμποσε ανα ιμπλιξιτ φραμε οφ ρεφερενςε. Ο άνθρωπος έχει μια ιεραρχική δομή για τηην αναπαράσταση αντικειμένων (οι νευρώνες δημιουργούν ιεραρχίες). Έτσι έχουμε τα νευρωνικά δίκτυα με κάψουλες. Αυτές οι ιεραρχίες δεν υπάρχουν στα συνελικτικά νευρωνικά δίκτυα. Επίσης, οι νευρώνες στον ανθρώπινο εγκέφαλο οργανώνονται σε συστάδες (έτσι και τα ςαπσυλες).

2.2.2 Βασικές Ανεπάρκειες των Συνελικτικών Νευρωνικών Δικτύων

Θα πεις και κάποια θετικά αναέδώ των συνελικτικών νευρωνικών δικτύων (κυρίως για τα συνελικτικά επίπδα και το ωειγητ σηαρινγ). Δισενταγλινγ τηε φαςτορς οφ αριατιον βεττερ. τηεψ αρε νοτ δεαλινγ ιν α πρινςιπαλεδ ωαψ ωιτη τηε ατιατιονς οφ ιεωποιντ τηατ ςηανγε τηε ιμαγε...

2.2.3 Αρχές Λειτουργείας Νευρωνικών Δικτύων με Κάψουλες

ςολαβορατιε φιλτερινγ, ρουτινγ βψ αγρεεμεντ, ωηατ ις α ςαπσυλε ...

2.3 Μηχανισμός Προσοχής

2.4 Μετασχηματιστές

2.5 Χάρτες Αυτο-οργάνωσης

Κεφάλαιο 3

Σχετικές Ερασίες

Κεφάλαιο 4

Μέθοδος

Κεφάλαιο 5

Πειράματα

Κεφάλαιο 6

Επίλογος

Παράρτημα Α΄

Ορισμοί Εννοιών

Το παρόν παράρτημα περιέχει ορισμούς εννοιών που εισάγονται κατά τη διάρκεια της παρούσας εργασίας. Κατά αυτόν τον τρόπο, δε διακόπτεται η ροή του κυρίως κειμένου. [;, ;, ;, ;]

Τεχνητή Νοημοσύνη

Έχουν υπάρξει πολλοί διαφορετικοί ορισμοί της Τεχνητής Νοημοσύνης: Μερικοί την περιγράφουν σαν εσωτερική διαδικασία της σκέψης που προσομοιάζει αυτή του ανθρώπου ενώ άλλοι ως εξωτερική διαδικασία μαθηματικά βέλτιστης συμπεριφοράς. Σύμφωνα με το κυρίαρχο μοντέλο, η Τεχνητή Νοημοσύνη ασχολείται κυρίως με τη λογική δράση. Ένας ιδανικός ευφυής πράκτορας δρα βέλτιστα σε κάθε περίσταση. Έτσι λοιπόν, η μελέτη της δημιουργίας ευφυών πρακτόρων μπορεί να τεθεί ως ορισμός της Τεχνητής Νοημοσύνης.

Μηχανική Μάθηση

Με λίγα λόγια, πρόκειται για τον κλάδο της τεχνητής νοημοσύνης ο οποίος ασχολείται με την ανάπτυξη υπολογιστικών συστημάτων ικανών να μαθαίνουν από παραδείγματα. Αναλυτικότερα, μπορούν και προσαρμόζονται χωρίς να ακολουθούν ρητές εντολές αλλά μέσω αλγορίθμων και στατιστικών μοντέλων που τους επιτρέπουν να αναλύουν και να εξάγουν συμπεράσματα από μοτίβα σε δεδομένα. Χαρακτηριστικό γνώρισμα των συστημάτων μηχανικής μάθησης είναι η ικανότητά τους να βελτιώνουν την απόδοσή τους σε μια εργασία (όπως αυτή μετράται με κάποια κατάλληλη μετρική) όσο η «εμπειρία» τους σε αυτήν αυξάνεται [;].

Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα αποτελούν ένα αλγοριθμικό κατασκεύασμα από απλούς υπολογιστικούς κόμβους διασυνδεδεμένους μεταξύ τους μέσω ακμών κάτω από μια συγκεκριμένη τοπολογία (συνήθως οργανώνονται σε επίπεδα, βλ. 2.1.3). Εμπνευσμένα από τα βιολογικά νευρωνικά δίκτυα, οι κόμβοι μπορούν να παρομοιαστούν με κύτταρά νευρώνων ενώ οι ακμές με νευρικές συνάψεις.

Τα τεχνητά νευρωνικά δίκτυα είναι παράδειγμα συστήματος μηχανικής μάθησης αφού μετά την κατάλληλη εκπαίδευσή τους, γενικεύουν από τα δεδομένα (inference). Τελικά, μετά την ανάπτυξή τους, υπό μια αφαιρετική σκοπιά αποτελεί το καθένα μια συνάρτηση που αντιστοιχίζει δεδομένα από τον χώρο εισόδου σε «προβλέψεις» του χώρου εξόδου.

Βαθιά Μάθηση

Αποτελεί μια υποκατηγορία μηχανικής μάθησης όπου χρησιμοποιούνται πολυεπίπεδα νευρωνικά δίκτυα. Τα πολλαπλά επίπεδα που διαθέτουν τους επιτρέπουν να μαθαίνουν και να αναγνωρίζουν εσωτερικά, γενικευμένα χαρακτηριστικά των δεδομένων εισόδου.

Υπολογιστική Νευροεπιστήμη (Computational Neuroscience)

Πρόκειται για τον κλάδο της Νευρωεπιστήμης που χρησιμοποιεί μαθηματικά μοντέλα, μαθηματική ανάλυση και προσεγγιστικά προς τον εγκέφαλο συστήματα για να κατανοήσει τις αρχές ανάπτυξης, δομής, φυσιολογίας καθώς και των γνωστικών (cognitive) ικανοτήτων του νευρικού συστήματος.

Επιβλεπόμενη Μάθηση

Στους αλγορίθμους επιβλεπόμενης μάθησης, ως είσοδος παρέχεται ένα σύνολο δεδομένων μαζί με τους επιθυμητούς στόχους. Δηλαδή, τα δεδομένα δίνονται σε ζεύγη (παράδειγμα εισόδου—επιθυμητή τιμή εξόδου). Με βάση αυτά, το σύστημα καλείται να εξάγει μια συνάρτηση η οποία θα έχει μάθει να μοντελοποιεί τη σχέση εισόδου–εξόδου μέσα από τα παραδείγματα και τελικά θα είναι ικανή να προβλέψει την τιμή εξόδου σε νέα παραδείγματα για τα οποία η τιμή στόχος είναι άγνωστη. Συνήθως, εκτός από τα δεδομένα για την εκπαίδευση υπάρχουν και άλλα σύνολα δεδομένων για τον έλεγχο της απόδοσης του συστήματος πρόβλεψης.

Ανάλογα με το αν η τιμή στόχος είναι διακριτή η συνεχής, έχουμε αντίστοιχα το πρόβλημα ταξινόμησης (classification) ή της παλινδρόμησης (regression). Παράδειγμα συστήματος ταξινόμησης επιβλεπόμενης μάθησης είναι το φίλτρο ανεπιθύμητης αλληλογραφίας το οποίο αφού εκπαιδεύτηκε με ένα σύνολο επισημασμένων αλληλογραφιών ως ανεπιθύμητων ή επιθυμητών έμαθε να εντοπίζει νέα εισερχόμενη ανεπιθύμητη αλληλογραφία. Ένα παράδειγμα συστήματος παλινδρόμησης επιβλεπόμενης μάθησης είναι αυτό της πρόβλεψης τιμών μετοχών καθώς ο στόχος (κόστος μετοχής) είναι συνεχής αριθμός.

Μη-επιβλεπόμενη Μάθηση

Οι αλγόριθμοι μη–επιβλεπόμενης μάθησης, σε αντιδιαστολή με τους αλγορίθμους επιβλεπόμενης μάθησης, δέχονται ως είσοδο ένα σύνολο δεδομένων που περιλαμβάνει παραδείγματα, χωρίς όμως να συνοδεύονται από αντίστοιχες τιμές–στόχους. Στην περίπτωση αυτή, το υπό εκπαίδευση σύστημα επιχειρεί να μάθει πρότυπα στα δεδομένα εισόδου χωρίς κάποιο μηχανισμό ανατροφοδότησης. Συνήθεις εφαρμογές μη–επιβλεπόμενης μάθησης είναι αυτές της ομαδοποίησης των δεδομένων σε συστάδες ή της αναπαράστασής τους με ένα γράφημα.

Ενισχυτική Μάθηση

Στην ενισχυτική μάθηση, στο σύστημα (το οποίο καλείται «ευφυής πράκτορας» στο πλαίσιο αυτό) δεν παρέχεται κάποιο σύνολο δεδομένων αλλά η όποια εμπειρία αποκτάται μέσω της αλληλεπίδρασής του με το περιβάλλον. Ο πράκτορας έχει τη δυνατότητα να παρατηρήσει το περιβάλλον του και τη (πιθανή) κατάστασή του και ανάλογα με μια στρατηγική (policy) να δράσει σε αυτό. Το περιβάλλον του, με κάθε δράση (και ανάλογα την κατάσταση) παρέχει την απαραίτητη εμπειρία υπό τη μορφή επιβράβευσης (reward) ή ποινής (punishment). Έτσι, ο πράκτορας μαθαίνει από την εμπειρία προσαρμόζοντας τη στρατηγική του ώστε να μεγιστοποιεί την επιβράβευση την οποία λαμβάνει και τελικά να πετυχαίνει τον στόχο του.

Παράδειγμα ενός τέτοιου πράκτορα είναι ένα σύστημα το οποίο παίζει σκάκι.

Μάθηση Κατά Δέσμες

Αφορά το είδος συστημάτων μηχανικής μάθησης που δεν έχουν τη δυνατότητα να μαθαίνουν σταδιακά αλλά εκπαιδεύονται μονομιάς χρησιμοποιώντας όλο το σύνολο δεδομένων στην είσοδό τους. Σε περίπτωση που προστεθούν νέα δεδομένα στα οποία θα επιθυμούσαμε το σύστημα να προσαρμοστεί, απαιτείται εκ νέου εκπαίδευση στο καινούριο σύνολο δεδομένων το οποίο θα περιέχει τόσο τα παλαιά όσο και τα επιπρόσθετα δεδομένα (διαδικασία χρονοβόρα και υπολογιστικά κοστοβόρα). Συνήθως, σε τέτοιες περιπτώσεις το σύστημα πρέπει να σταματήσει να λειτουργεί και να μεταβεί στη φάση σχεδιασμού. Παραδείγματα αυτών των μεθόδων αποτελούν ο αλγόριθμος Expectation Maximization και ο Self-organizing map όπως περιγράφονται στην ενότητα ;; και 2.5.

Μάθηση σε Ζωντανό Χρόνο

Πρόκειται για τα συστήματα μηχανικής μάθησης που, σε αντίθεση με αυτά που μαθαίνουν κατά δέσμες, είναι ικανά να εκπαιδεύονται σταδιακά, είτε με ένα παράδειγμα τη φορά είτε με μικρές δέσμες παραδειγμάτων στην είσοδό τους. Το θετικό σε αυτά τα συστήματα είναι η δυνατότητα προσαρμογής τους σε νέα δεδομένα με πολύ μικρό χρονικό και υπολογιστικό κόστος. Αποτέλεσμα αυτού είναι ότι υπάρχει (συνήθως) η δυνατότητα η εκπαίδευσή τους να γίνει ζωντανά (online) χωρίς να σταματήσει η λειτουργία του συστήματος. Παράδειγμα αποτελούν οι εφαρμογές πρόβλεψης τιμών μετοχών όπου απαιτείται συνεχής προσαρμογή του συστήματος στα νέα δεδομένα της αγοράς.

Μάθηση Βασισμένη σε Παραδείγματα

Είναι μια οικογένεια απλών συστημάτων μηχανικής μάθησης που αφορά τον τρόπο με τον οποίο ένα σύστημα γενικεύει από τα παραδείγματα του συνόλου εισόδου. Στα συγκεκριμένα, όταν τα τροφοδοτούμε με κάποιο νέο παράδειγμα, το συγκρίνουν με τα δεδομένα εισόδου (ή ένα υποσύνολο αυτών) τα οποία έχουν αποθηκευθεί στη μνήμη τους κατά την εκπαίδευση. Ένα χαρακτηριστικό μειονέκτημα αυτών των συστημάτων είναι ότι ο χώρος που απαιτείται για την αποθήκευση του μοντέλου (του συστήματος μάθησης μετά την εκπαίδευσή του) αυξάνεται με το μέγεθος του συνόλου εισόδου (συνήθως με γραμμικό τρόπο). Ενδεικτικά, ένα σύστημα που γενικεύει κατά αυτόν τον τρόπο είναι το K-nearest neighbors.

Μάθηση Βασισμένη σε μοντέλο

Είναι μια άλλη οικογένεια συστημάτων όπου η μηχανική μάθηση γίνεται μέσω της προσαρμογής (fitting) ενός μοντέλου στα δεδομένα εισόδου. Έχοντας εκφράσει το σύνολο των δεδομένων εκπαίδευσης (ή τη σχέση αυτών με την επιθυμητή έξοδο) χρησιμοποιώντας ένα κατάλληλα εκφραστικό (expressive) μοντέλο, λέμε ότι το σύστημα μαθαίνει να «γενικεύει» από τα παραδείγματα. Έτσι, για να παράξει προβλέψεις σε νέα δεδομένα, δεν απαιτείται η αποθήκευση όλων των δεδομένων εκπαίδευσης αλλά μόνο των παράμετρων του μοντέλου που εκφράζει.

Γνωστική Νευροεπιστήμη

A σηορτ ονε-λινε δεσςριπτιον.

Ταξινόμηση Προτύπων

A σηορτ ονε-λινε δεσςριπτιον.

Γραμμικά Διαχωρίσιμες Κλάσσεις
: Α σηορτ ονε-λινε δεσςριπτιον.

Γραμμικά Μοντέλα
: Α σηορτ ονε-λινε δεσςριπτιον.

Γενετικοί Αλγόριθμοι
: Α σηορτ ονε-λινε δεσςριπτιον.

Νευρωνικά Δίκτυα με Κάψουλες (Capsule Networks)
: Είναι αυτά...

Παράρτημα Β΄

Απόδοση Ξενόγλωσσων Όρων

Ξενόγλωσσος όρος	**Ελληνική απόδοση**
batch learning	μάθηση κατά δέσμες
online learning	μάθηση σε ζωντανό χρόνο
supervised learning	επιβλεπόμενη μάθηση
unpervised learning	μη-επιβλεπόμενη μάθηση
reinforcement learning	ενισχυτική μάθηση
capsule networks	νευρωνικά δίκτυα με κάψουλες
instance based	βασισμένο σε παραδείγματα
model based	βασισμένο σε μοντέλο

Παράρτημα Γ΄

Συντομογραφίες - Ακρωνύμια

Γ΄.1 Ελληνικά

Συντομογραφία ή Ακρωνύμιο	Πλήρης όρος
δφθφεωργερρεωγτε	γσδφ εργωεγφεωρφτωεγ
δφθφεωργεργεργρεωγτε	γσγφεωρφτωεγ
δφθφεωργεγτε	γσδφ εωρφτωεγ
δφθφεωργερρεγεργεργτεργγεργρεωγτε	γσδφ εργωεγφεωρφτωεγεγεργεργερωγερφ

Γ΄.2 Αγγλικά

Συντομογραφία ή Ακρωνύμιο	Πλήρης όρος
δφθφεωργερρεωγτε	γσδφ εργωεγφεωρφτωεγ
δφθφεωργεργεργρεωγτε	γσγφεωρφτωεγ
δφθφεωργεγτε	γσδφ εωρφτωεγ
δφθφεωργερρεγεργεργτεργγεργρεωγτε	γσδφ εργωεγφεωρφτωεγεγεργεργερωγερφ