

**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ**

**ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ**

**ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**ΕΠΙΧΕΙΡΗΜΑΤΙΚΕΣ ΚΑΙ ΟΙΚΟΝΟΜΙΚΕΣ  
ΕΦΑΡΜΟΓΕΣ  
ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ**

**ΚΑΝΕΛΛΗΣ ΝΙΚΟΛΑΟΣ**

**ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ: ΠΑΠΑΔΟΠΟΥΛΟΣ ΔΗΜΗΤΡΙΟΣ**

**ΠΑΤΡΑ, ΣΕΠΤΕΜΒΡΙΟΣ 2017**

## ΠΡΟΛΟΓΟΣ

Η συγγραφή αυτής της πτυχιακής εργασίας ξεκίνησε κατά τη διάρκεια της πρακτικής μου άσκησης ως αναλυτής δεδομένων σε μια επιχείρηση ηλεκτρονικού εμπορίου στο Sheffield της Αγγλίας, και ολοκληρώθηκε το Σεπτέμβριο του 2017, λίγο καιρό μετά την παρουσίαση μέρους της στο 5ο Διεθνές Συνέδριο Σύγχρονων Θεμάτων Μάρκετινγκ στη Θεσσαλονίκη. Κατά τη διάρκεια αυτής της περιόδου, είχα την ευκαιρία να δω πως εφαρμόζονται στην πράξη, σε ένα δραστήριο και απαιτητικό επιχειρηματικό περιβάλλον, οι διάφορες τεχνικές προβλέψεων και λήψης αποφάσεων που εξερεύνησα θεωρητικά στο τμήμα Διοίκησης Επιχειρήσεων του Τ.Ε.Ι. Δυτικής Ελλάδας. Μελετώντας παράλληλα την βιβλιογραφία πάνω στο θέμα της εργασίας, μου έγινε πλέον ξεκάθαρο ότι *οι επιχειρηματικές ανάγκες παράγουν έρευνα, και η έρευνα παράγει επιχειρηματικές ανάγκες*.

Κατά την ίδια περίοδο, οι εξελίξεις στον τομέα της τεχνητής νοημοσύνης και της μηχανικής μάθησης ήταν εντυπωσιακές. Πολύ χαρακτηριστικά παραδείγματα αποτελούν δύο προγράμματα τεχνητής νοημοσύνης (Libratus, AlphaGo), βασισμένα σε μεγάλο βαθμό σε τεχνολογία νευρωνικών δικτύων, τα οποία πέτυχαν συντριπτικές νίκες ενάντια σε ισχυρούς επαγγελματίες παίκτες στα παιχνίδια του Poker και του Go, τα οποία μέχρι πριν λίγο καιρό θεωρούνταν από τα “τελευταία προπύργια” της ανθρώπινης νοημοσύνης απέναντι στην τεχνητή.

Όλα τα παραπάνω μου έδωσαν σημαντικό κίνητρο για να ασχοληθώ με τους αλγορίθμους τεχνητών νευρωνικών δικτύων και τις εφαρμογές τους στις επιχειρήσεις, όχι μόνο στα πλαίσια της παρούσας εργασίας, αλλά και σαν γενικότερο ερευνητικό αντικείμενο. Κατά κάποιο τρόπο, η δομή αυτής της εργασίας περιγράφει την μαθησιακή πορεία που ακολούθησα πάνω σε αυτό το θέμα: από την εξοικείωση με τις βασικές αρχές της μηχανικής μάθησης, μέχρι την προσπάθεια να συνδέσω μια επιχειρηματική ανάγκη που εντόπισα στην πράξη με τις τεχνικές νευρωνικών δικτύων που μελέτησα θεωρητικά.

Ευχαριστώ θερμά τον εποπτεύοντα καθηγητή μου κ. Δημήτρη Παπαδόπουλο, που μου κίνησε την προσοχή για αυτό το σύγχρονο και ενδιαφέρον θέμα, και μου έδειξε απεριόριστη εμπιστοσύνη και υπομονή κατά τη διάρκεια της συγγραφής.

## ΠΕΡΙΛΗΨΗ

Τα Τεχνητά Νευρωνικά Δίκτυα, μια από τις πλέον διαδεδομένες μεθόδους μηχανικής μάθησης, βρίσκουν στις μέρες μας ιδανική εφαρμογή στους κλάδους της οικονομίας και των επιχειρήσεων. Η δυνατότητά τους να εντοπίζουν υποβόσκουσες σχέσεις αιτίου - αποτελέσματος σε μεγάλα σύνολα δεδομένων με πολλές παραμέτρους, αξιοποιώντας άριστα τις δυνατότητες παράλληλης επεξεργασίας των σύγχρονων υπολογιστών, προσφέρει στους αναλυτές ένα πολύ ισχυρό εργαλείο στην προσπάθειά τους για επίτευξη ανταγωνιστικού πλεονεκτήματος σε μια παγκόσμια διασυνδεδεμένη οικονομία, που καθημερινά παράγει έναν ωκεανό πληροφοριών προς εκμετάλλευση.

Στην παρούσα εργασία αρχικά παρουσιάζονται οι βασικές αρχές, η ιστορική εξέλιξη και το θεωρητικό υπόβαθρο της μηχανικής μάθησης και ειδικότερα των αλγορίθμων νευρωνικών δικτύων, περιγράφεται η διαδικασία ανάπτυξης ενός μοντέλου νευρωνικών δικτύων για την επίλυση ενός προβλήματος, και γίνεται σύγκριση των νευρωνικών δικτύων με άλλες μεθόδους μηχανικής μάθησης. Στην συνέχεια παρουσιάζονται κάποιες πραγματικές εφαρμογές τέτοιων μοντέλων σε επιχειρηματικά και οικονομικά προβλήματα. Τέλος, προτείνεται και περιγράφεται μια νέα επιχειρηματική εφαρμογή νευρωνικών δικτύων: ένας αλγόριθμος κατηγοριοποίησης παραγγελιών λιανικού ηλεκτρονικού εμπορίου, ο οποίος εντοπίζει αυτές που έχουν την μεγαλύτερη πιθανότητα να μην φτάσουν στον παραλήπτη τους, προσφέροντας στον έμπορο τη δυνατότητα να περιορίσει το μέσο κόστος αποστολής του.

**Λέξεις κλειδιά:** μηχανική μάθηση, τεχνητά νευρωνικά δίκτυα, ηλεκτρονικό εμπόριο, κόστος αποστολής.

## **ABSTRACT**

Artificial Neural Networks, one of the most popular Machine Learning methods, are finding ideal applications in the fields of economy and business these days. Their ability to pinpoint underlying cause-and-effect relationships in vast datasets with lots of parameters, utilising the parallel processing capabilities of modern computers very efficiently, provides analysts with a powerful tool in their effort to achieve a competitive advantage in a globally interconnected economy, which produces an ocean of information on a daily basis.

In this study, the basic principles, the history and the theoretical background of machine learning, and more specifically of neural network algorithms is initially demonstrated, then the process of developing a neural network model for the purposes of a specific problem is described, and neural networks are compared with other machine learning methods. Later on, some real applications of these models in economy and business related problems are showcased. Finally, a new business application of neural networks is proposed and described: a classification algorithm for retail e-commerce orders, that identifies the ones least likely to reach their intended recipient, giving the retailer an opportunity to reduce his average shipping cost.

**Keywords:** machine learning, artificial neural networks, e-commerce, shipping cost

# ΠΕΡΙΕΧΟΜΕΝΑ

<b>ΠΡΟΛΟΓΟΣ</b>	<b>1</b>
<b>ΠΕΡΙΛΗΨΗ</b>	<b>2</b>
<b>ABSTRACT</b>	<b>3</b>
<b>ΠΕΡΙΕΧΟΜΕΝΑ</b>	<b>4</b>
<b>ΕΙΣΑΓΩΓΗ</b>	<b>7</b>
<b>1.ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ</b>	<b>9</b>
Ορισμός και βασικά χαρακτηριστικά	10
Τύποι αλγορίθμων μάθησης	11
Μάθηση με επίβλεψη	12
Παλινδρόμηση	13
Ταξινόμηση	15
Μεροληψία - Διακύμανση (bias-variance tradeoff)	17
Πολυπλοκότητα του μοντέλου σε σχέση με την ποσότητα των εκπαιδευτικών παραδειγμάτων (model complexity)	18
Διάσταση του χώρου των δεδομένων εισόδου (input space dimensionality)	19
Ομοιομορφία των δεδομένων (homogeneity of the data)	19
Θόρυβος στις τιμές εξόδου (noise in the output values)	19
Μάθηση χωρίς επίβλεψη	19
Συσταδοποίηση	20
Κανόνες Συσχέτισης	22
Μάθηση με μερική επίβλεψη	26
Ενισχυτική μάθηση	28
Άλλοι τύποι προβλημάτων	29
Μείωση διαστάσεων	29
Ανίχνευση ανωμαλιών	30
<b>2.ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ</b>	<b>32</b>
Βασικά χαρακτηριστικά του τεχνητού νευρώνα	32

Συναρτήσεις ενεργοποίησης	34
Βηματική Συνάρτηση (binary step function)	35
Λογιστική Σιγμοειδής Συνάρτηση (logistic sigmoid function)	35
Συνάρτηση Υπερβολικής Εφαπτομένης (Hyperbolic tangent function)	36
Συνάρτηση Γραμμικού Ανορθωτή (Rectified linear unit - ReLU)	36
Επιλογή επιπέδων	37
Κανόνας Μάθησης (Learning Rule)	37
Οπισθοδιάδοση (Back-Propagation)	39
Οι βιολογικοί νευρώνες	39
Ιστορική αναδρομή	43
Ισχυρά σημεία και αδυναμίες των Νευρωνικών Δικτύων	44
<b>3.ΕΦΑΡΜΟΓΕΣ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ</b>	<b>45</b>
Εφαρμογές σε διάφορους τομείς	45
Γενικές εφαρμογές	45
Ιατρικές εφαρμογές	46
Βιολογικές - γεωλογικές - περιβαλλοντολογικές εφαρμογές	46
Αρχιτεκτονικές και βιομηχανικές εφαρμογές	47
Εφαρμογές στην φωτογραφική τέχνη	47
Εφαρμογές στις εμπορικές και βιομηχανικές επιχειρήσεις	47
Πωλήσεις και Μάρκετινγκ	48
Βελτιστοποίηση Επιχειρησιακών Λειτουργιών	49
Εφαρμογές στον τραπεζικό και χρηματιστηριακό τομέα.	49
Πρόβλεψη χρονοσειρών	50
Έλεγχος πιστοληπτικής ικανότητας	50
Εκτίμηση αξίας	51
Ανίχνευση απάτης	51
<b>4.ΠΡΟΤΕΙΝΟΜΕΝΗ ΕΦΑΡΜΟΓΗ</b>	<b>52</b>
Ταυτοποίηση πακέτων που ενδέχεται να μην φτάσουν στον παραλήπτη τους	52

Περιγραφή του προβλήματος	52
Παρακολούθηση πακέτων	53
Περιγραφή της εφαρμοζόμενης μεθόδου	54
Εκτίμηση των αποτελεσμάτων	57
Χρήση των αποτελεσμάτων από την επιχείρηση	57
Μελλοντικοί στόχοι της εφαρμογής	57
<b>5.ΣΥΜΠΕΡΑΣΜΑΤΑ</b>	<b>59</b>
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b>	<b>60</b>
Ελληνική Βιβλιογραφία	60
Ξένη Βιβλιογραφία	61
<b>ΠΑΡΑΡΤΗΜΑ</b>	<b>71</b>

## ΕΙΣΑΓΩΓΗ

*“Από αίτια που φαίνονται παρόμοια, περιμένουμε παρόμοια αποτελέσματα. Αυτό συνιστά το σύνολο των εμπειρικών μας συμπερασμάτων.”*

Η παραπάνω φράση ανήκει στο σπουδαίο φιλόσοφο του 18ου αιώνα David Hume (1748), ο οποίος σε πολλά δοκίμιά του είχε ασχοληθεί με τη σχέση αιτίου και αποτελέσματος. Από τους κύριους προβληματισμούς του ήταν η ορθότητα των συμπερασμάτων που κάθε άνθρωπος είναι σε θέση να βγάλει εμπειρικά, χωρίς να εφαρμόσει αυστηρά λογικούς κανόνες. Ένα παράδειγμα που χρησιμοποιούσε ήταν αυτό του ψωμιού: βλέποντας ένα αντικείμενο που *μοιάζει*, όπως το αντιλαμβανόμαστε μέσω των αισθήσεων μας, με ψωμί, η εκτίμηση μας είναι ότι θα μπορούμε να το φάμε, χωρίς όμως να μπορούμε να αποδείξουμε αυτήν την υπόθεση πριν το δοκιμάσουμε.

Καθημερινά συναντάμε περιπτώσεις όπου η ανθρώπινη διαίσθηση επιτυγχάνει να παράγει ορθά συμπεράσματα: κάθε φορά που ακούμε φράσεις όπως: *“απλά ακολούθησα το ένστικτό μου”*, *“κάτι μέσα μου έλεγε να μην πάρω αυτήν την απόφαση”*, *“από την πρώτη στιγμή είχα καταλάβει τον χαρακτήρα του”*, ευθέως αναγνωρίζουμε την αξία της εκτίμησης ή της απόφασης του συνομιλητή μας, χωρίς να απαιτούμε περαιτέρω ανάλυση του συλλογισμού ή των επιχειρημάτων του. Η προσλαμβανόμενη αξιοπιστία της διαίσθησης φαίνεται επίσης να ενισχύεται από την εμπειρία και την εξοικείωση με το εκάστοτε αντικείμενο: *“ασχολείται χρόνια με αυτό και ξέρει τι λέει”*, *“η εκτίμηση ενός ειδικού έχει μεγαλύτερη βαρύτητα”*, *“η γυναικεία διαίσθηση δεν κάνει λάθος σε τέτοιες περιπτώσεις”*, κλπ.

Σχεδόν τρεις αιώνες μετά τον Hume, ο άνθρωπος έχει μπει για τα καλά στην εποχή της πληροφορίας. Στα επιχειρηματικά περιβάλλοντα, οι συζητήσεις κυριαρχούνται από εκφράσεις όπως *big data analysis*, *data mining*, *knowledge management*. Η συντριπτική πλειοψηφία όλων των δεδομένων που υπάρχουν στον κόσμο, είτε σε παραδοσιακά είτε σε ψηφιακά μέσα αποθήκευσης, έχει δημιουργηθεί μόνο μέσα στα τελευταία λίγα χρόνια (SINTEF, 2013). Για αποφάσεις που έχουν να κάνουν με την ταξινόμηση, ανάλυση, και εξαγωγή συμπερασμάτων από τα διαρκώς διογκούμενα επιχειρηματικά δεδομένα, οι περισσότερες επιχειρήσεις στηρίζονται στις εκτιμήσεις ειδικών αναλυτών, οι οποίοι με τη σειρά τους συνεχώς αναζητούν νέα υπολογιστικά εργαλεία που θα διευκολύνουν την εργασία τους και θα ενισχύσουν την αξιοπιστία της.

Η ύπαρξη αυτού του εντυπωσιακού όγκου δεδομένων, η κατακόρυφη αύξηση της διαθέσιμης υπολογιστικής ισχύος, και η διαρκής ελάττωση του κόστους του αποθηκευτικού χώρου, σε συνδυασμό με την σχετικά πρόσφατη ανάπτυξη ολοκληρωμένων υπηρεσιών παράλληλης αποθήκευσης και επεξεργασίας δεδομένων όπως το *Apache Hadoop* ή το *Amazon Redshift*, αποτέλεσαν πρόσφορο έδαφος για την αναζήτηση σχέσεων αιτίου - αποτελέσματος, οι οποίες επί πολλά χρόνια ήταν καλά κρυμμένες σε μεγάλα σύνολα δεδομένων, αόρατες ακόμα για από τους πιο σχολαστικούς ερευνητές, ακόμη κι αν αυτοί υποψιάζονταν την ύπαρξή τους.



Τα Τεχνητά Νευρωνικά Δίκτυα, αλγόριθμοι που αναπτύχθηκαν αρχικά σαν μια απόπειρα προσομοίωσης της λειτουργίας των νευρώνων του ανθρώπινου εγκεφάλου, βρήκαν σε αυτό το τεχνολογικό περιβάλλον την ιδανική τους εφαρμογή. Η ικανότητά τους να προσαρμόζονται στα ίδια τα δεδομένα εισόδου τους, χωρίς να προαπαιτείται ένα πολύπλοκο σύνολο κανόνων από τον άνθρωπο - χρήστη, καθώς και η δυνατότητα παράλληλης επεξεργασίας τους, είναι μόνο δύο από τα πολλά χαρακτηριστικά που έχουν φέρει τα Νευρωνικά Δίκτυα στο προσκήνιο της σύγχρονης προσπάθειας των επιχειρήσεων για ανάλυση δεδομένων με στόχο την εξαγωγή πρωτότυπων συμπερασμάτων, μέσω των οποίων φιλοδοξούν να αποκτήσουν συγκριτικά πλεονεκτήματα απέναντι στους ανταγωνιστές τους.

Σχεδιάζοντας τη δομή αυτής της πτυχιακής εργασίας, κατέληξα στο συμπέρασμα ότι η παρουσίαση και η κατανόηση των εννοιών γύρω από τα νευρωνικά δίκτυα, καθώς και η περιγραφή των εφαρμογών τους, μπορούν να γίνουν με πολύ ομαλότερο τρόπο αν αναφερθώ πρώτα στην ευρύτερη κατηγορία αλγορίθμων στην οποία ανήκουν, την Μηχανική Μάθηση (Machine Learning). Έτσι, στο πρώτο κεφάλαιο παρουσιάζονται κάποιες στοιχειώδεις έννοιες σχετικές με την Μηχανική Μάθηση, αναφέρονται οι βασικές κατηγορίες προβλημάτων στις οποίες αυτοί οι αλγόριθμοι καλούνται να δώσουν απάντηση, και δίνεται μια μικρή περιγραφή για τους κυριότερους εξ αυτών.

Στην συνέχεια, στο δεύτερο κεφάλαιο, γίνεται μια αναλυτικότερη περιγραφή των βασικών χαρακτηριστικών, της ιστορίας, και του θεωρητικού υποβάθρου των Νευρωνικών Δικτύων, περιγράφεται η διαδικασία ανάπτυξης, λειτουργίας, και εκτίμησης των αποτελεσμάτων ενός τέτοιου αλγορίθμου, και αναφέρονται τα βασικά πλεονεκτήματα και μειονεκτήματά τους σε σύγκριση με παρόμοιες τεχνικές.

Στο τρίτο κεφάλαιο παρουσιάζονται κάποιες ήδη υπάρχουσες εφαρμογές των νευρωνικών δικτύων, με ιδιαίτερη έμφαση σε αυτές που αφορούν στον τομέα της οικονομικής επιστήμης και των επιχειρήσεων. Εφόσον έχουν ήδη γραφτεί πολλές εργασίες πάνω σε αυτό το θέμα, και πολλές δημοφιλείς εφαρμογές έχουν ήδη παρουσιαστεί επανειλημμένα σε διάφορα άρθρα και άλλες δημοσιεύσεις, αυτές που επέλεξα να παρουσιάσω είναι όσο το δυνατόν πιο πρόσφατες, με κάποιες εξαιρέσεις, όπου θεώρησα ότι η σπουδαιότητα και η χρησιμότητα κάποιων παλαιότερων εφαρμογών παραμένει πολύ μεγάλη.

Τέλος, στο τέταρτο κεφάλαιο παρουσιάζεται μια δική μου πρόταση εφαρμογής: ένας αλγόριθμος κατηγοριοποίησης παραγγελιών λιανικού ηλεκτρονικού εμπορίου, ο οποίος θα δίνει την δυνατότητα στον χρήστη του να εντοπίζει αυτές που έχουν την μεγαλύτερη πιθανότητα να μην φτάσουν στον παραλήπτη τους. Πιστεύω πως τα ιδιαίτερα χαρακτηριστικά αυτού του προβλήματος και ο γενικά ανεπαρκής τρόπος με τον οποίο αντιμετωπίζεται συνήθως, καθιστούν πολύ ελκυστική τη χρήση ενός νευρωνικού δικτύου. Είναι ένα επιχειρηματικό ζήτημα που δεν φαίνεται έχει απασχολήσει το ερευνητικό ενδιαφέρον προς το παρόν, αλλά θεωρώ ότι η ολοένα αυξανόμενη υιοθέτηση του ηλεκτρονικού εμπορίου από το ευρύ κοινό αναπόφευκτα θα στρέψει την προσοχή των αναλυτών σε αυτό και σε άλλα παρόμοια ζητήματα.

# 1.ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Η Μηχανική Μάθηση (Machine Learning), ένας κλάδος της Τεχνητής Νοημοσύνης, αποτελεί στις μέρες μας ένα ώριμο πεδίο έρευνας, με πολλά παρακλάδια και ένα τεράστιο εύρος πρακτικών εφαρμογών. Πολλά πανεπιστήμια σε όλο τον κόσμο περιλαμβάνουν πλέον τη Μηχανική Μάθηση στη λίστα των διαθέσιμων μαθημάτων τους, σε προπτυχιακό ή μεταπτυχιακό επίπεδο.

Αν κάναμε την παραπάνω δήλωση πριν μερικές δεκαετίες, θα απείχε αρκετά από την πραγματικότητα. Μέχρι και τις αρχές της δεκαετίας του '80, οι πιο γνωστές εφαρμογές τεχνητής νοημοσύνης δεν ήταν παρά αλγόριθμοι (expert systems) που συγκέντρωναν μεγάλα σύνολα λογικών κανόνων και συνθηκών if-then σε βάσεις δεδομένων (knowledge bases). Ένα από τα προβλήματα που έπρεπε να αντιμετωπίσουν αυτού του είδους οι αλγόριθμοι ήταν ότι στηρίζονταν στην ποιότητα της επικοινωνίας μεταξύ των ειδικών και των προγραμματιστών. Συχνά, οι ειδικοί δεν μπορούσαν να εκφράσουν όλο το εύρος των γνώσεών τους πάνω στο εκάστοτε αντικείμενο, και αντίστοιχα οι προγραμματιστές δεν ήταν σε θέση να κάνουν τις σωστές ερωτήσεις στους ειδικούς, ούτε μπορούσαν πάντα να μεταφράσουν τις απαντήσεις τους σε ένα αξιοποιήσιμο σύνολο λογικών κανόνων (Kubat, 2015). Επίσης, οι παραδοσιακοί αυτοί αλγόριθμοι δεν ήταν αρκετά ευέλικτοι, καθώς δεν προσαρμόζονταν εύκολα σε νέα δεδομένα, για τα οποία δεν υπήρχαν σαφώς καθορισμένοι κανόνες από τους ειδικούς (Διαμαντάρας, 2011).

Η απλή ιδέα που άλλαξε την πορεία της εξέλιξης της τεχνητής νοημοσύνης, και την οδήγησε σε αυτό που σήμερα αποκαλούμε Μηχανική Μάθηση, ήταν η εξής: Εφόσον για τα περισσότερα προβλήματα, το να περιγράψουμε ακριβώς τον τρόπο επίλυσης μέσα από έναν αλγόριθμο είναι αρκετά δύσκολο, άραγε θα μπορούσαμε να μεταδώσουμε έμμεσα αυτές τις πληροφορίες σε έναν υπολογιστή, δίνοντάς του παραδείγματα, μέσω των οποίων ο υπολογιστής θα μάθει από μόνος του πως να λύνει το εκάστοτε πρόβλημα;

Η προοπτική ενός συστήματος που μαθαίνει, βεβαίως, δεν ήταν νέα τη δεκαετία του '80. Ο Alan Turing, στο άρθρο του *Computing machinery and intelligence* (1950), προσπάθησε να υπερασπιστεί την ιδέα μιας μηχανής που θα μπορούσε να επιδείξει στοιχεία ανθρώπινης συμπεριφοράς στη συνομιλία σε τέτοιο βαθμό, ώστε να είναι πολύ δύσκολο να την ξεχωρίσουμε από έναν άνθρωπο στα πλαίσια του, διάσημου πλέον, *Imitation Game*, που οδήγησε στην καθιέρωση του *Turing Test* ως μέτρου αξιολόγησης συστημάτων τεχνητής νοημοσύνης. Ανάμεσα στα αντεπιχειρήματα που προσπάθησε να καταρρίψει, χαρακτηριστική ήταν η άποψη της Ada Lovelace (1842), η οποία περιγράφοντας την *Αναλυτική Μηχανή* του Charles Babbage σχολίαζε: “*Η Αναλυτική Μηχανή δεν μπορεί να δημιουργήσει κάτι πρωτότυπο. Έχει τη δυνατότητα να κάνει μόνο αυτά που εμείς γνωρίζουμε πώς να της ζητήσουμε να κάνει.*” Ο Turing, ως απάντηση, παραθέτει το παρακάτω σχόλιο του Douglas Hartree (1949): “*...αυτή η άποψη δεν αποκλείει το ενδεχόμενο της δημιουργίας μιας ηλεκτρονικής συσκευής που θα μπορούσε να σκεφτεί ή, χρησιμοποιώντας βιολογικούς όρους,*

να εφοδιαστεί με εξαρτημένα αντανεκλαστικά που θα αποτελέσουν τη βάση ενός τύπου μάθησης.”

Ο Turing, στο ίδιο άρθρο, έκανε και μερικές άλλες πολύ ενδιαφέρουσες παρατηρήσεις, όπως:

- Μια σκεπτόμενη μηχανή, αντί να προσπαθεί να προσομοιάζει το μυαλό ενός ενήλικου ατόμου, θα έπρεπε να σχεδιαστεί ώστε να προσομοιάζει το μυαλό ενός παιδιού, και κατόπιν να υποβληθεί σε μια διαδικασία *εκπαίδευσης*, ακολουθώντας έτσι την ίδια πορεία μάθησης με αυτήν του ανθρώπου.
- Ο “δάσκαλος” μιας τέτοιας μηχανής συχνά θα μπορεί να προβλέψει, σε ένα βαθμό, τα αποτελέσματα της λειτουργίας της, χωρίς απαραίτητα να μπορεί να προσδιορίσει τη λειτουργία της ανά πάσα στιγμή (σε αντίθεση με ένα παραδοσιακό υπολογιστικό αλγόριθμο).
- Ένα απαραίτητο χαρακτηριστικό μιας τέτοιας μηχανής θα ήταν η ύπαρξη κάποιου τυχαίου παράγοντα, ώστε να διαφοροποιούνται οι απαντήσεις που δίνει κάθε φορά.

Παρά την απλοϊκότητά τους, αυτές οι παρατηρήσεις μεταξύ άλλων αποτέλεσαν καθοριστικό παράγοντα στην διαμόρφωση των βάσεων της Μηχανικής Μάθησης, όπως τη γνωρίζουμε σήμερα.

## Ορισμός και βασικά χαρακτηριστικά

Η έννοια της μάθησης, όπως αυτή είναι αντιληπτή στην πραγματική ζωή, μπορεί να περιγραφεί από τις δύο ακόλουθες βασικές ικανότητες ενός οργανισμού που μαθαίνει (Βλαχάβας κ.α., 2011):

- η ικανότητα πρόσκτησης γνώσης κατά την αλληλεπίδραση με το περιβάλλον
- η ικανότητα βελτίωσης του τρόπου εκτέλεσης μιας ενέργειας, με την επανάληψή της

Ο Arthur Samuel, προγραμματιστής στην IBM τη δεκαετία του ‘50 και μετέπειτα καθηγητής στο πανεπιστήμιο του Stanford, όρισε τη Μηχανική Μάθηση ως:

*“Το πεδίο έρευνας που δίνει στους υπολογιστές τη δυνατότητα να μαθαίνουν, χωρίς να έχουν προγραμματιστεί ρητά για αυτό το σκοπό.”* (Samuel, 1959)

Ο Samuel, θέλοντας να αποδείξει ότι ένας υπολογιστής μπορεί να προγραμματιστεί ώστε να μάθει εκτελεί μια σύνθετη εργασία καλύτερα από τον ίδιο τον προγραμματιστή, σχεδίασε ένα πρόγραμμα το οποίο μπορούσε να παίζει ντάμα (checkers), και το έβαλε να παίζει 10000 φορές με τον εαυτό του. Το πρόγραμμα αποθήκευε κάθε θέση που εμφανιζόταν τουλάχιστον σε μια παρτίδα, και της ανέθετε μια βαθμολογία, ανάλογα με τις νίκες και τις ήττες που είχαν προκύψει από αυτήν τη θέση. Έτσι, κάθε φορά ήταν σε θέση να επιλέξει την καλύτερη κίνηση (βάσει βαθμολογίας) από το σύνολο των επιτρεπόμενων, σύμφωνα με την προηγούμενη *εμπειρία* του.

Η παραπάνω διαδικασία εκφράζεται με ακρίβεια και από τον μεταγενέστερο ορισμό ενός άλλου πρωτοπόρου της Μηχανική Μάθησης, του Nils Nilsson, επίσης καθηγητή του πανεπιστημίου Stanford:

*“Μηχανή μάθησης (learning machine), είναι οποιοσδήποτε μηχανισμός του οποίου οι ενέργειες επηρεάζονται από τις προηγούμενες εμπειρίες του.” (Nilsson, 1965)*

Παρόμοιο ορισμό έδωσε και ο Tom Mitchell, καθηγητής του Carnegie Mellon:

*“Το πεδίο της Μηχανικής Μάθησης ασχολείται με την κατασκευή προγραμμάτων  $H/Y$ , τα οποία βελτιώνονται αυτόματα με την εμπειρία τους” (Mitchell, 1997)*

Ο Mitchell, συγγραφέας του πρώτου εκπαιδευτικού βιβλίου για τη Μηχανική Μάθηση, προσπάθησε στην συνέχεια να ορίσει την μάθηση, ως εξής:

*“Ορίζουμε ότι ένα πρόγραμμα μαθαίνει από την εμπειρία  $E$  πάνω σε ένα τύπο προβλημάτων  $T$  δεδομένου ενός μέτρου απόδοσης  $P$ , όταν η απόδοσή του σε ένα πρόβλημα κατηγορίας  $T$ , μετρημένη κατά  $P$ , βελτιώνεται με την εμπειρία  $E$ .”*

Συνεπώς, τα βασικά χαρακτηριστικά ενός αλγορίθμου Μηχανικής Μάθησης είναι:

- ο τύπος του προβλήματος: (T) - task.
- η εμπειρία: (E) - experience.
- το μέτρο απόδοσης που θέλουμε να βελτιώσουμε: (P) - performance.

Για παράδειγμα, μπορούμε να περιγράψουμε τον πρώιμο αλγόριθμο του Samuel ως εξής:

- T: πως να παίζεις ντάμα
- E: 10000 παρτίδες του αλγόριθμου με τον εαυτό του
- P: το ποσοστό νικών απέναντι σε άλλους αντιπάλους.

Συνοπτικά λοιπόν, μπορούμε να πούμε ότι η Μηχανική Μάθηση είναι ένας κλάδος της Τεχνητής Νοημοσύνης με στόχο τη δημιουργία συστημάτων τα οποία μπορούν να *μαθαίνουν* ή *αλλιώς να εκπαιδεύονται* από την εμπειρία τους *αυτόματα*, επιτυγχάνοντας κατ' αυτό τον τρόπο στην βελτίωση της απόδοσής τους ως προς ένα προκαθορισμένο μέτρο απόδοσης.

## Τύποι αλγορίθμων μάθησης

Υπάρχουν διάφοροι τύποι αλγορίθμων μηχανικής μάθησης. Η επιλογή του τύπου που επιθυμούμε να χρησιμοποιήσουμε συνήθως εξαρτάται από τον τύπο του προβλήματος, καθώς και από τον τύπο των δεδομένων εξόδου που περιμένουμε από τον αλγόριθμο (Bell, 2015).

Οι περισσότεροι αλγόριθμοι μηχανικής μάθησης εμπίπτουν σε μια από τις παρακάτω κατηγορίες, ανάλογα με τη μέθοδο μάθησης που ακολουθείται:

- Μάθηση με επίβλεψη (supervised learning)

- Μάθηση χωρίς επίβλεψη (unsupervised learning)

Υπάρχουν βέβαια και κάποιες παραλλαγές σε αυτές τις μεθόδους μάθησης. Στην παρούσα εργασία θα αναφερθούν δύο από αυτές που είναι αρκετά δημοφιλείς τα τελευταία χρόνια:

- Μάθηση με μερική επίβλεψη (semi-supervised learning)
- Ενισχυτική Μάθηση (reinforcement learning)

## Μάθηση με επίβλεψη

Οι αλγόριθμοι που ανήκουν στην κατηγορία της μάθησης με επίβλεψη (supervised learning - αναφέρεται και ως επιτηρούμενη μάθηση), βασίζονται σε ένα σύνολο δεδομένων εισόδου τα οποία είναι ήδη *χαρακτηρισμένα (labeled)*, δηλαδή τους έχει ήδη αποδοθεί μια τιμή εξόδου ή έχουν ήδη ταξινομηθεί σε μια κατηγορία. Αυτό το σύνολο δεδομένων αποκαλείται *εκπαιδευτικό σύνολο (training set)*, και αποτελείται από αντικείμενα εισόδου, τα οποία συνήθως εκφράζονται ως διανύσματα που περιλαμβάνουν τις τιμές των διάφορων χαρακτηριστικών τους, καθώς και από τις αντίστοιχες τιμές εξόδου που συνήθως εκφράζονται ως απλοί αριθμοί. Κάθε αντικείμενο εισόδου μαζί με την τιμή εξόδου του αποτελούν ένα *εκπαιδευτικό παράδειγμα (training example)*. Βάσει αυτών των δεδομένων, ο αλγόριθμος προσπαθεί να κατασκευάσει ένα *μοντέλο (model)*: μια “συνάρτηση” η οποία μπορεί να αντιστοιχίζει τα αντικείμενα εισόδου του εκπαιδευτικού συνόλου με τις τιμές εξόδου τους, ώστε στη συνέχεια να είναι ικανό να αποδίδει από μόνο του τιμές εξόδου σε νέα αντικείμενα εισόδου, τα οποία δεν έχουν χαρακτηριστεί, σύμφωνα με αυτήν τη συνάρτηση.

Η παραπάνω διαδικασία, δηλαδή η προσπάθεια ανακάλυψης ενός γενικού κανόνα για ένα σύνολο βάσει ενός υποσυνόλου του, ή γενικότερα για ένα περιβάλλον βάσει μιας απλοποιημένης, αφαιρετικής εκδοχής του, ονομάζεται *επαγωγή (induction)*, και η δημιουργία ενός τέτοιου μοντέλου ονομάζεται αντίστοιχα *επαγωγική μάθηση (inductive learning)*.

Στην απλούστερη μορφή του, αυτό το μοντέλο μπορεί να εκφραστεί μαθηματικά ως εξής: Δεδομένου ενός συνόλου εισόδου  $X$ , και ενός συνόλου εξόδου  $Y$ , υποθέτουμε ότι υπάρχει ένα μοντέλο που μπορεί να τα αντιστοιχήσει, βάσει κάποιων παραμέτρων, ως εξής:

$$y = g(x|\theta)$$

όπου  $g(\cdot)$  είναι το μοντέλο και  $\theta$  είναι οι παράμετροί του (Alpaydin, 2010). Ο αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται προσπαθεί ουσιαστικά να βελτιστοποιήσει τις τιμές των παραμέτρων  $\theta$ , ώστε οι εκτιμήσεις του μοντέλου να πλησιάζουν όσο το δυνατόν περισσότερο τις πραγματικές τιμές εξόδου.

Η μάθηση με επίβλεψη είναι η πιο δημοφιλής μέθοδος μηχανικής μάθησης (Ng, 2011). Η αξιοπιστία των αλγορίθμων μηχανικής μάθησης με επίβλεψη, όσον αφορά στα παραδείγματα που δεν ανήκουν στο εκάστοτε εκπαιδευτικό σύνολο, βασίζεται στην *υπόθεση της επαγωγικής μάθησης (inductive learning hypothesis)*, σύμφωνα με την οποία “Κάθε υπόθεση  $h$  που προσεγγίζει καλά τη συνάρτηση στόχο για ένα αρκετά μεγάλο σύνολο παραδειγμάτων, θα

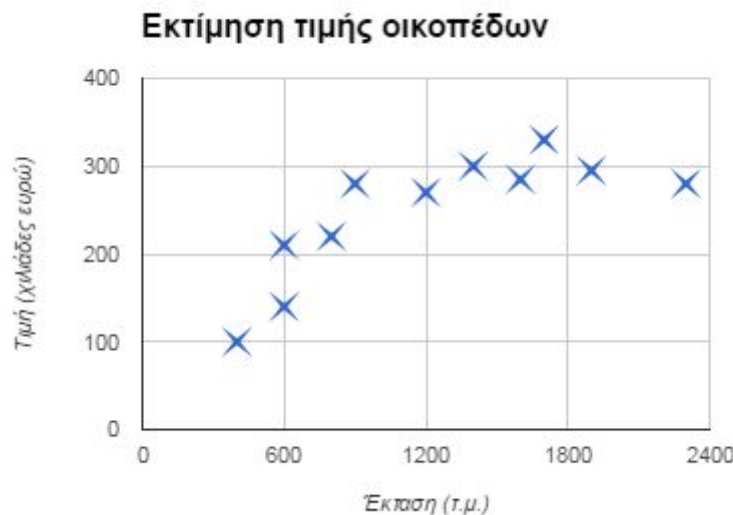
προσεγγίζει το ίδιο καλά τη συνάρτηση στόχο και για περιπτώσεις που δεν έχει εξετάσει.”  
(Βλαχάβας κ.α., 2011)

Αλγόριθμοι μάθησης με επίβλεψη συνήθως χρησιμοποιούνται για προβλήματα παλινδρόμησης ή ταξινόμησης. Στα προβλήματα παλινδρόμησης, η τιμή του  $y$  που αναφέρεται παραπάνω είναι ένας πραγματικός αριθμός, ενώ στα προβλήματα ταξινόμησης το  $y$  είναι ένας (συνήθως ακέραιος) αριθμός που αντιπροσωπεύει μια κατηγορία.

## Παλινδρόμηση

Τα προβλήματα παλινδρόμησης (regression - αναφέρεται και ως *παρεμβολή*), στα πλαίσια της μηχανικής μάθησης, εξετάζουν τη σχέση μεταξύ κάποιων ανεξάρτητων μεταβλητών (ή *εκτιμητών*), και μιας συνεχούς εξαρτημένης μεταβλητής. Το ζητούμενο σε αυτές τις περιπτώσεις είναι να προβλεφθεί η αλλαγή στην εκτιμώμενη τιμή της εξαρτημένης μεταβλητής, όταν μεταβάλλεται κάποια από τις ανεξάρτητες μεταβλητές.

Ένα πολύ απλό παράδειγμα είναι το ακόλουθο: έστω ότι προσπαθούμε να εκτιμήσουμε την τιμή αγοράς ενός οικοπέδου, γνωρίζοντας μόνο την έκτασή του. Αυτό είναι ένα πρόβλημα παλινδρόμησης, καθώς από ένα σύνολο ανεξάρτητων μεταβλητών (εν προκειμένω, έχουμε μόνο την μεταβλητή “έκταση”), προσπαθούμε να εκτιμήσουμε μια συνεχή μεταβλητή (την “τιμή αγοράς”). Γνωρίζοντας την τιμή και την έκταση για μερικά άλλα οικοπέδα, μπορούμε να κατασκευάσουμε ένα *εκπαιδευτικό σύνολο* όπως αυτό που απεικονίζεται παρακάτω:

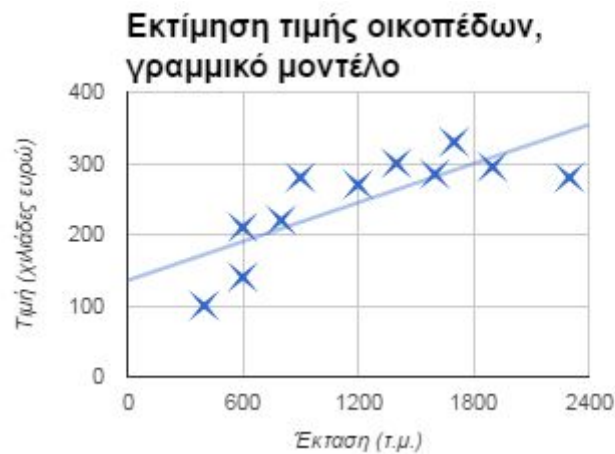


Κάθε σημείο σε αυτό το διάγραμμα αντιστοιχεί σε ένα *εκπαιδευτικό παράδειγμα*: το δεδομένο εισόδου είναι η έκταση του κάθε οικοπέδου, και η τιμή εξόδου (η οποία *χαρακτηρίζει* το εκπαιδευτικό παράδειγμα), είναι η ήδη γνωστή σε εμάς τιμή του κάθε οικοπέδου.

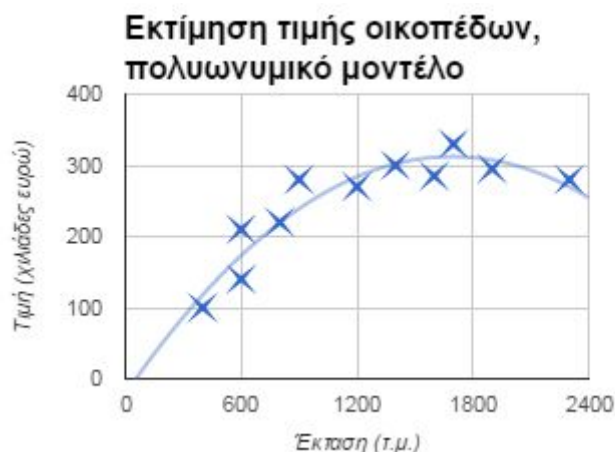
Ο ρόλος του αλγορίθμου είναι να διαμορφώσει μια συνάρτηση που θα εκτιμά την τιμή ενός οικοπέδου για οποιοδήποτε έκταση μπορεί να έχει, με όσο το δυνατόν λιγότερη απόκλιση

αυτής της τιμής, τόσο από τις τιμές των εκπαιδευτικών παραδειγμάτων, όσο και από τις πραγματικές τιμές των οικοπέδων για τα οποία αρχικά θα γνωρίζουμε μόνο την έκταση.

Αν, για παράδειγμα, χρησιμοποιήσουμε ένα αλγόριθμο απλής γραμμικής παλινδρόμησης, η συνάρτηση αυτή μπορεί να έχει την μορφή της ευθείας στο παρακάτω διάγραμμα:



Θέλοντας λοιπόν να εκτιμήσουμε την τιμή ενός οικοπέδου 1200 τ.μ., η συνάρτηση αυτή θα μας έδινε μια τιμή περίπου 250 χιλιάδες ευρώ. Εάν δε χρησιμοποιούσαμε έναν άλλο αλγόριθμο, *πολυωνυμικής παλινδρόμησης*, η μορφή της συνάρτησης μπορεί να ήταν κάπως έτσι:



Η εκτίμηση αυτής της συνάρτησης για ένα οικόπεδο έκτασης 1200 τ.μ. θα ήταν τότε περίπου 280 χιλιάδες ευρώ. Φυσικά ένα από τα δύο μοντέλα είναι μάλλον πιο ακριβές από το άλλο, ή με άλλα λόγια έχει μικρότερο *σφάλμα πρόβλεψης* από το άλλο, δηλαδή οι προβλεπόμενες τιμές του μοντέλου απέχουν λιγότερο από τις πραγματικές τιμές. Η επιλογή μεταξύ τους όμως δεν είναι όσο απλή όσο φαίνεται. Κάποιοι από τους παράγοντες που θα πρέπει να πάρουμε υπόψη στην εκτίμηση του κάθε μοντέλου, παρουσιάζονται στη συνέχεια.

Αλγόριθμοι παλινδρόμησης χρησιμοποιούνται συχνά για την πρόβλεψη τιμών μετοχών, θερμοκρασίας, βαθμολογίας, πωλήσεων και άλλων μεταβλητών τέτοιου τύπου.

## Ταξινόμηση

Τα προβλήματα ταξινόμησης (classification - αναφέρεται και ως *κατηγοριοποίηση*), εξετάζουν τη σχέση κάποιων ανεξάρτητων μεταβλητών με μια διακριτή (διατακτική, ονομαστική, ή λογική) μεταβλητή. Το ζητούμενο είναι, δεδομένων κάποιων χαρακτηριστικών μιας οντότητας, να προσδιορίσουμε την τάξη ή την κατηγορία της.

Μερικά παραδείγματα τέτοιων προβλημάτων είναι η αναγνώριση χειρόγραφων χαρακτήρων, η αξιολόγηση πελατών βάσει πιστωτικού προφίλ, και η διάγνωση ασθενειών από βιολογικά δεδομένα.

Μια παρατήρηση που κάνει σαφέστερη την διαφορά μεταξύ παλινδρόμησης και ταξινόμησης, είναι ότι στα προβλήματα ταξινόμησης όλες οι λάθος προβλέψεις είναι το ίδιο λανθασμένες: για παράδειγμα, αν ένας αλγόριθμος αναγνώρισης χειρόγραφων αριθμών εσφαλμένα ταξινομήσει ένα 4 ως 1, δεν είναι καλύτερο ή χειρότερο από το να ταξινομήσει ένα 4 ως 7. Στα προβλήματα παλινδρόμησης, όμως, τα σφάλματα δεν έχουν όλα την ίδια βαρύτητα. Αν η πραγματική τιμή ενός οικοπέδου είναι 150000€, μια πρόβλεψη για 170000€ είναι λογικά πολύ πιο ανεκτή από μια πρόβλεψη για 40000€.

Για να δώσουμε πάλι ένα παράδειγμα, έστω ότι θέλουμε να εκτιμήσουμε αν πρόκειται να βρέξει, βάσει της μέσης θερμοκρασίας και υγρασίας που μετρήθηκε την προηγούμενη μέρα. Εν προκειμένω, οι κατηγορίες στις οποίες μπορεί να εμπίπτει η μέρα είναι δύο (θα βρέξει - δεν θα βρέξει), ενώ παρατηρούμε ότι οι ανεξάρτητες μεταβλητές δεν είναι απαραίτητα διακριτές: η θερμοκρασία είναι συνεχής, και θα μπορούσαν να υπάρχουν και άλλα ποιοτικά χαρακτηριστικά όπως το αν είχε βρέξει την προηγούμενη μέρα (ναι - όχι), ή τι εποχή του χρόνου είναι κλπ.

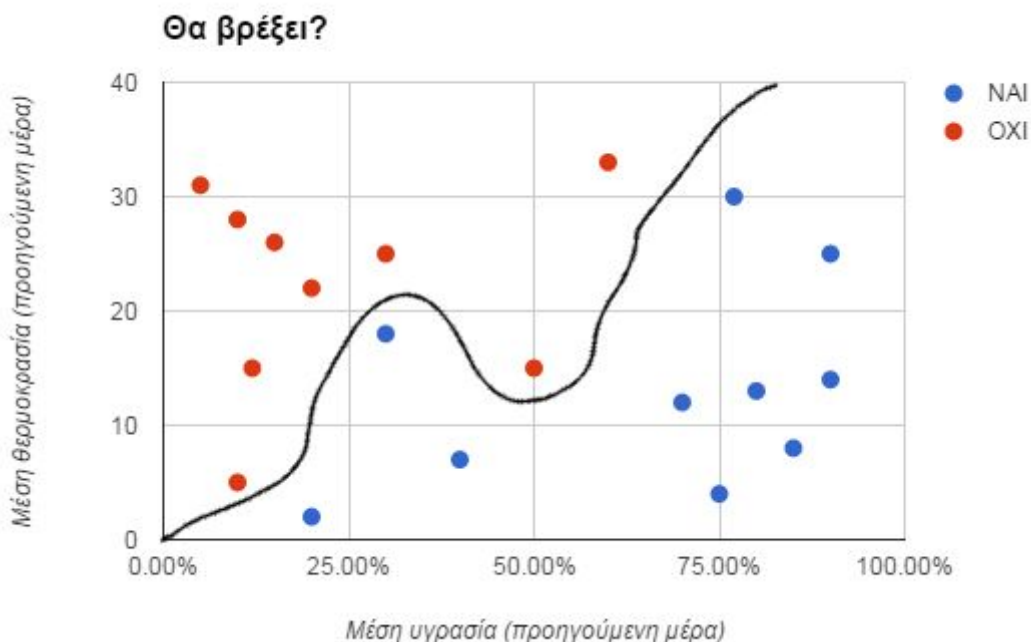
Στον παρακάτω πίνακα απεικονίζεται το εκπαιδευτικό σύνολο του προβλήματος:

Θερμοκρασία (προηγούμενη μέρα)	Σχετική Υγρασία (προηγούμενη μέρα)	Έβρεξε την επόμενη μέρα?
2	20.00%	ΝΑΙ
4	75.00%	ΝΑΙ
5	10.00%	ΟΧΙ
7	40.00%	ΝΑΙ
8	85.00%	ΝΑΙ
12	70.00%	ΝΑΙ
13	80.00%	ΝΑΙ
14	90.00%	ΝΑΙ
15	50.00%	ΟΧΙ



15	12.00%	OXI
18	30.00%	NAI
22	20.00%	OXI
25	30.00%	OXI
25	90.00%	NAI
26	15.00%	OXI
28	10.00%	OXI
30	77.00%	NAI
31	5.00%	OXI
33	60.00%	OXI

Αν βάλουμε αυτά τα δεδομένα σε ένα διάγραμμα διασποράς, θα διαπιστώσουμε ότι υπάρχουν πολλές συναρτήσεις που θα μπορούσαν να αποτελέσουν ένα καλό ταξινομητή για το αν θα βρέξει αύριο. Ο ταξινομητής (*classifier* ή *discriminant*) είναι μια συνάρτηση που ξεχωρίζει τα εκπαιδευτικά παραδείγματα βάσει κάποιου κανόνα. Η παρακάτω συνάρτηση, που απεικονίζεται ως η μαύρη καμπύλη του σχήματος, ταξινομεί τις ημέρες στις κλάσεις ΝΑΙ - θα βρέξει και ΟΧΙ - δεν θα βρέξει, βάσει της θέσης του γραμμικού συνδυασμού των χαρακτηριστικών τους:



Βάσει αυτής της συνάρτησης, μπορούμε να πούμε ότι αν την προηγούμενη μέρα είχε 30 βαθμούς και 25% υγρασία, τότε μάλλον δε θα βρέξει, ενώ αν είχε 10 βαθμούς και 50% υγρασία, μάλλον θα βρέξει. Πάλι όμως γεννιέται το ερώτημα: είναι αυτή η συνάρτηση ο πιο ακριβής ταξινομητής για την μεταβλητή που εξετάζουμε; Πριν δώσουμε απάντηση, θα

πρέπει να έχουμε εξετάσει τα παρακάτω σημαντικά ζητήματα που επηρεάζουν όλους τους αλγορίθμους μηχανικής μάθησης με επίβλεψη:

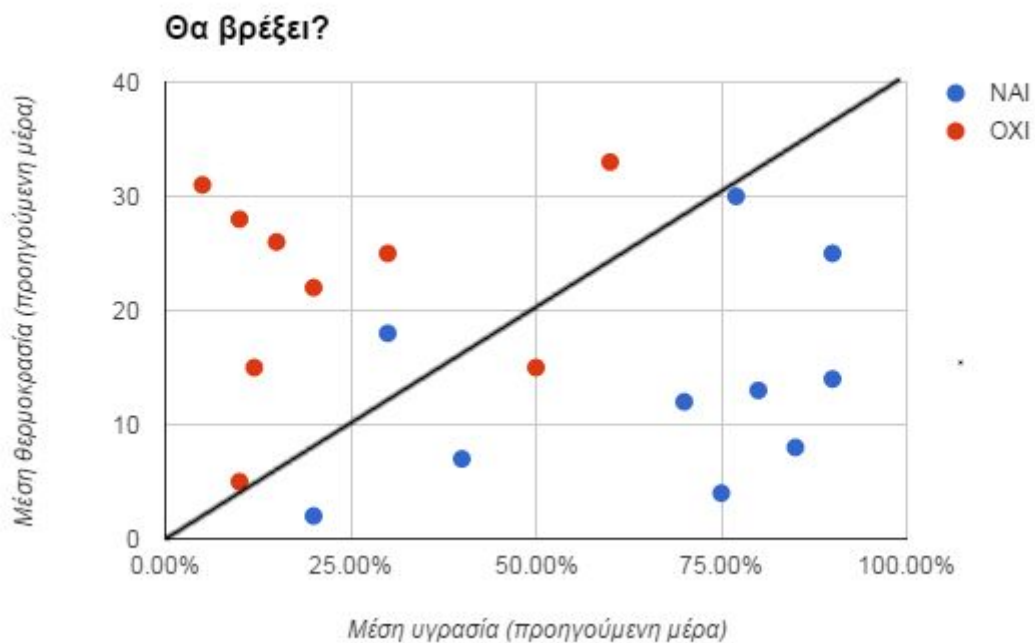
### Μεροληψία - Διακύμανση (bias-variance tradeoff)

Η μεροληψία και η διακύμανση είναι οι δύο πιο σημαντικές πηγές *σφαλμάτων* που ευθύνονται για την αδυναμία πολλών αλγορίθμων μάθησης με επίβλεψη να γενικεύσουν τα αποτελέσματά τους πέραν του εκπαιδευτικού συνόλου (Geman et al., 1992). Η μεροληψία αφορά στα σφάλματα που προκαλούνται από εσφαλμένες παραδοχές του αλγορίθμου μάθησης, ενώ η διακύμανση αφορά στα σφάλματα που προκαλούνται από ευαισθησία του αλγορίθμου μάθησης σε εκπαιδευτικά παραδείγματα που παρεκκλίνουν από τα υπόλοιπα, χωρίς αυτό να είναι αντιπροσωπευτικό για το σύνολο (τα οποία αποκαλούνται και *θόρυβος* του εκπαιδευτικού συνόλου).

Ένας αλγόριθμος με υψηλό βαθμό διακύμανσης μπορεί να οδηγήσει σε ένα μοντέλο πρόβλεψης που είναι σε θέση να απεικονίζει με ακρίβεια το εκπαιδευτικό σύνολο, αλλά παράλληλα να μην επιτυγχάνει να γενικεύσει με ακρίβεια σε παραδείγματα εκτός του εκπαιδευτικού συνόλου, επειδή προσαρμόζεται και σε εκπαιδευτικά δεδομένα που δεν είναι αντιπροσωπευτικά για το πρόβλημα, αποτελώντας τυχαίο θόρυβο. Αυτό το πρόβλημα ονομάζεται *υπερπροσαρμογή (overfitting)*.

Αντίθετα, ένας αλγόριθμος με υψηλό βαθμό μεροληψίας συνήθως οδηγεί σε απλούστερα μοντέλα που δεν υπερπροσαρμόζονται στο εκπαιδευτικό σύνολο, αλλά ενδεχομένως δεν καταφέρνουν να εκφράζουν κάποιες σημαντικές παρεκκλίσεις από το μοντέλο στο εκπαιδευτικό σύνολο, χάνοντας έτσι και ένα σημαντικό αριθμό σωστών προβλέψεων και εκτός αυτού. Αυτό το πρόβλημα ονομάζεται *υποπροσαρμογή (underfitting)*.

Επιστρέφοντας στο προηγούμενο παράδειγμα, ένας άλλος ταξινομητής θα μπορούσε να είναι η παρακάτω ευθεία γραμμή. Αυτός ο ταξινομητής έχει μεγαλύτερη μεροληψία (παρατηρούμε ότι ταξινομεί λάθος δύο εκπαιδευτικά παραδείγματα), αλλά είναι πιο απλός και ενδεχομένως να εκφράζει καλύτερα όλα τα παραδείγματα (εντός και εκτός εκπαιδευτικού συνόλου). Σε αυτή την περίπτωση, λέμε ότι ο πρώτος ταξινομητής είχε *υπερπροσαρμοστεί* στα δεδομένα. Φυσικά, αν τελικά ο πρώτος ταξινομητής αποδειχθεί πιο ακριβής, τότε λέμε ότι ο παρακάτω ταξινομητής *υποπροσαρμόζεται* στα δεδομένα, καθώς η αυξημένη πολυπλοκότητα του πρώτου ήταν χρήσιμη, εφόσον αποτύπωσε κάποιες σημαντικές παρεκκλίσεις από το μοντέλο της ευθείας γραμμής στο εκπαιδευτικό σύνολο.



Το άθροισμα του σφάλματος μεροληψίας και του σφάλματος διακύμανσης, μαζί με το αναπόφευκτο (*irreducible*) σφάλμα, αποτελούν αυτό που ονομάζουμε συνολικό σφάλμα πρόβλεψης ενός αλγορίθμου μηχανικής μάθησης (James, 2003).

Πολυπλοκότητα του μοντέλου σε σχέση με την ποσότητα των εκπαιδευτικών παραδειγμάτων (model complexity)

Στην πράξη, είναι σπάνιο να βρούμε ένα μοντέλο που να ελαχιστοποιεί και την μεροληψία, και την διακύμανση. Ο αναλυτής καλείται να επιλέξει ποιος τύπος σφάλματος είναι λιγότερο σημαντικός για το πρόβλημα που καλείται να αναλύσει, και προσαρμόζει το μοντέλο του ανάλογα. Γενικά, αν θέλουμε να μειώσουμε την διακύμανση, επειδή το μοντέλο μας περιέχει πολύ *θόρυβο*, για παράδειγμα, συνήθως προσπαθούμε να χρησιμοποιήσουμε όσο το δυνατόν περισσότερα εκπαιδευτικά δεδομένα, και όσο το δυνατόν λιγότερα χαρακτηριστικά. Αν όμως θέλουμε να μειώσουμε την μεροληψία, επειδή ίσως θεωρούμε ότι πολλές ιδιαιτερότητες του εκπαιδευτικού συνόλου είναι σημαντικές και θα πρέπει να ληφθούν υπόψη, ή γνωρίζουμε ότι το εκπαιδευτικό μας σύνολο είναι σχετικά μικρό για να αποτυπώσει με σαφήνεια όλες τις σημαντικές συσχετίσεις μεταξύ των μεταβλητών, τότε συνήθως θα πρέπει να επιλέξουμε ένα πιο πολύπλοκο μοντέλο.

Με λίγα λόγια, ένας καλός αλγόριθμος μηχανικής μάθησης θα πρέπει να προσαρμόζει το βαθμό μεροληψίας και διακύμανσης του μοντέλου ανάλογα με τον αριθμό των δεδομένων εισόδου σε σχέση με την εκτιμώμενη πολυπλοκότητα του μοντέλου.

### Διάσταση του χώρου των δεδομένων εισόδου (input space dimensionality)

Συνήθως, τα εκπαιδευτικά παραδείγματα εκφράζονται ως διανύσματα, με τις τιμές των  $n$  χαρακτηριστικών τους να αντιπροσωπεύουν τις συντεταγμένες του διανύσματος σε ένα  $n$ -διάστατο χώρο. Το πρόβλημα που μπορεί να δημιουργηθεί εδώ, είναι ότι αν τα χαρακτηριστικά που λαμβάνουμε υπόψη είναι πάρα πολλά, δηλαδή η διάσταση του χώρου των διανυσμάτων είναι πολύ μεγάλου βαθμού, τότε ο εκπαιδευτικός αλγόριθμος μπορεί να επηρεάζεται αρνητικά από κάποια χαρακτηριστικά που δεν είναι σημαντικά, αλλά είναι δυσανάλογα πολλά σε σχέση με τα σημαντικά χαρακτηριστικά. Σε αυτές τις περιπτώσεις ο αναλυτής καλείται να επιλέξει, από όλα τα διαθέσιμα χαρακτηριστικά, μόνο αυτά που θεωρεί σημαντικά, είτε από την εμπειρία του, είτε χρησιμοποιώντας έναν αλγόριθμο *επιλογής χαρακτηριστικών (feature selection)*.

### Ομοιομορφία των δεδομένων (homogeneity of the data)

Οι περισσότεροι αλγόριθμοι μηχανικής μάθησης δεν λειτουργούν σωστά όταν τα χαρακτηριστικά των παραδειγμάτων έχουν διαφορετικό εύρος τιμών μεταξύ τους, καθώς τα χαρακτηριστικά που έχουν μεγάλο αριθμητικό εύρος τιμών επηρεάζουν υπερβολικά το μοντέλο. Για αυτό το λόγο, πριν την εκπαίδευση ενός τέτοιου αλγορίθμου πρέπει να προηγείται η *κανονικοποίηση των χαρακτηριστικών*, δηλαδή η αναπαράστασή τους στο ίδιο εύρος τιμών (π. χ.  $[-1,1]$ ).

### Θόρυβος στις τιμές εξόδου (noise in the output values)

Συχνά, στο εκπαιδευτικό σύνολο υπάρχουν παραδείγματα με τιμές εξόδου σημαντικά διαφορετικές από την πλειοψηφία των παραδειγμάτων με παρόμοια χαρακτηριστικά, κάτι που μπορεί να συμβαίνει λόγω ανθρώπινου ή μηχανικού λάθους στο χαρακτηρισμό τους. Φυσικά, δεν θέλουμε το μοντέλο μας να προσαρμόζεται σε αυτά τα παραδείγματα, και για αυτό το λόγο είτε επιλέγουμε ένα μοντέλο με μεγάλο βαθμό μεροληψίας, αν πιστεύουμε ότι τα παραδείγματα αυτά είναι πολλά, ή δεν μπορούμε να τα εντοπίσουμε εύκολα, ή ακόμα καλύτερα, εντοπίζουμε τα εκπαιδευτικά παραδείγματα που οδηγούν σε αυτές τις τιμές και τα αφαιρούμε πριν την εκπαίδευση του μοντέλου. Αυτή η διαδικασία μπορεί να είναι χρονοβόρα και πολύπλοκη, αλλά αποδεδειγμένα μειώνει τα σφάλματα πρόβλεψης σε παραδείγματα εκτός του εκπαιδευτικού συνόλου (Brodley and Friedl, 1999).

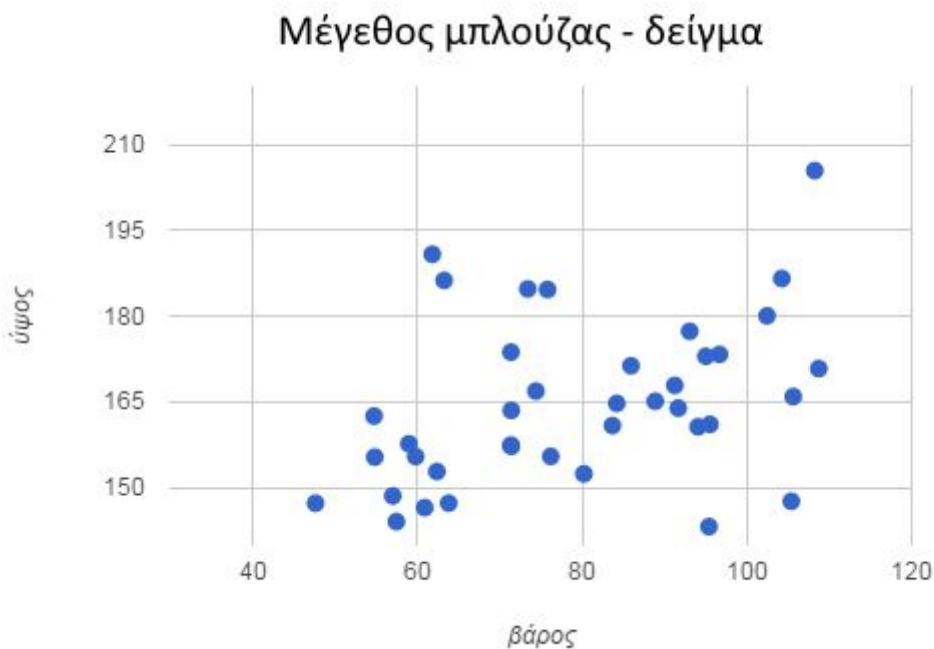
## Μάθηση χωρίς επίβλεψη

Οι αλγόριθμοι που ανήκουν στην κατηγορία της μάθησης χωρίς επίβλεψη (unsupervised learning - αναφέρεται και ως μη επιτηρούμενη μάθηση), σε αντίθεση με τους αλγορίθμους μάθησης με επίβλεψη, δεν απαιτούν *χαρακτηρισμένα* παραδείγματα εισόδου. Το μόνο που έχουμε στα χέρια μας είναι κάποια δεδομένα εισόδου, και αυτό που θέλουμε από τον αλγόριθμο είναι ο εντοπισμός κάποιων συγκεκριμένων μοτίβων σε αυτά. Συνήθως, χρησιμοποιούνται για προβλήματα *συσταδοποίησης, ή ανακάλυψης κανόνων συσχέτισης*.

## Συσταδοποίηση

Τα προβλήματα συσταδοποίησης (clustering - αναφέρεται συχνά και ως ομαδοποίηση), έχουν σαν στόχο τον εντοπισμό ομάδων με κοινά χαρακτηριστικά ανάμεσα στα δεδομένα εισόδου. Συνήθως θέλουμε να καταλήξουμε σε μια διαμέριση του συνόλου αντικειμένων εισόδου, έτσι ώστε αυτά τα αντικείμενα να έχουν μεγάλο βαθμό ομοιότητας με τα υπόλοιπα αντικείμενα της ομάδας τους, αλλά παράλληλα μικρό βαθμό ομοιότητας με τα αντικείμενα που βρίσκονται σε άλλες ομάδες. Μερικές πολύ δημοφιλείς εφαρμογές συσταδοποίησης είναι η οργάνωση των πελατών μιας επιχείρησης σε ομάδες για στοχευμένη διαφήμιση (τμηματοποίηση της αγοράς), η ανάλυση κοινωνικών δικτύων για τον εντοπισμό ατόμων που γνωρίζονται μεταξύ τους ή έχουν κοινά ενδιαφέροντα, και η ταξινόμηση ζωντανών οργανισμών βάσει μοτίβων στο γονιδίωμά τους.

Ένα παράδειγμα συσταδοποίησης είναι το εξής: έστω ότι μια εταιρία φτιάχνει μπλούζες, και διαθέτει τους πόρους και τα μέσα για να φτιάξει τρία μόνο διαφορετικά μεγέθη (small, medium, large). Ενδιαφέρεται λοιπόν να δει τι περίπου διαστάσεις θα πρέπει να έχουν αυτά τα μεγέθη ώστε να καλύπτουν όσο περισσότερους ανθρώπους γίνεται. Στη διάθεσή της έχει ένα σύνολο δεδομένων ύψους και βάρους για ένα τυπικό δείγμα ανθρώπων, όπως το ακόλουθο:

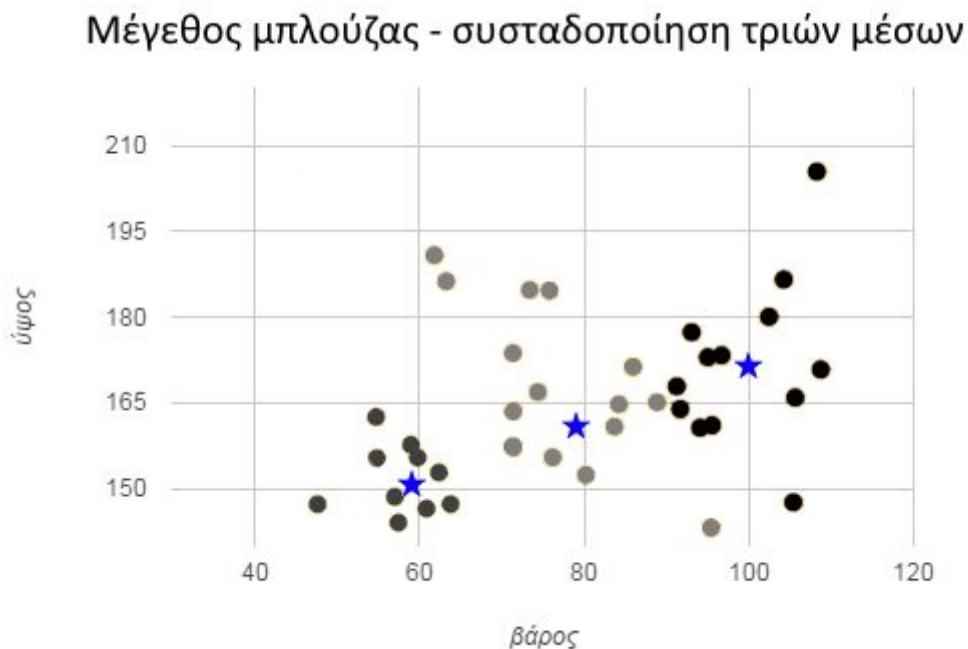


Μια απλή μέθοδος συσταδοποίησης που μπορούμε να χρησιμοποιήσουμε, είναι η μέθοδος συσταδοποίησης  $k$ -μέσων ( $k$ -means clustering), η οποία παρουσιάστηκε για πρώτη φορά από τον James MacQueen (1967). Στην πιο απλή του μορφή, ο αλγόριθμος δέχεται σαν είσοδο τα διανύσματα των δεδομένων εισόδου (εδώ, έχουν μόνο δύο διαστάσεις, ύψος και βάρος), την μέθοδο μέτρησης που θα χρησιμοποιήσει για να υπολογίσει την απόσταση μεταξύ τους (εδώ αρκεί η απλή ευκλείδεια απόσταση), τον αριθμό  $k$  που αντιστοιχεί στον αριθμό των ομάδων

που επιθυμούμε (τρεις), καθώς και μια τιμή *κατωφλιού* (*threshold*), της οποίας η χρησιμότητα θα γίνει κατανοητή σύντομα. Ο αλγόριθμος εκτελεί στη συνέχεια τα παρακάτω βήματα (Κιτμερίδης, 2013):

1. Επιλέγει  $k$  τυχαία διανύσματα από το σύνολο δεδομένων εισόδου, τα οποία τα ορίζει ως τα αρχικά *κέντρα* (*centroids*) των ομάδων.
2. Τα υπόλοιπα διανύσματα ταξινομούνται σε μια ομάδα, ανάλογα με το κέντρο από το οποίο απέχουν λιγότερο.
3. Όταν όλα τα διανύσματα έχουν ταξινομηθεί, υπολογίζεται εκ νέου το κέντρο κάθε ομάδας, σύμφωνα με το μέσο όρο όλων των διανυσμάτων που ανήκουν σ' αυτές.
4. Ο αλγόριθμος επιστρέφει στο δεύτερο βήμα, και επαναλαμβάνει αυτήν τη διαδικασία μέχρις ότου τα νέα κέντρα των ομάδων να απέχουν από τα προηγούμενα, ένα προς ένα, απόσταση μικρότερη από την τιμή *κατωφλιού* που όρισε ο χρήστης.
5. Όταν η παραπάνω συνθήκη εκπληρωθεί, θεωρούμε ότι ο αλγόριθμος συγκλίνει, οι ομάδες που δημιουργήθηκαν στην τελευταία επανάληψη είναι οι τελικές, και ο αλγόριθμος τερματίζει.

Το αποτέλεσμα ενός τέτοιου αλγορίθμου στο παράδειγμά μας, είναι να χωριστεί το δείγμα σε τρεις ομάδες, όπως φαίνονται στο παρακάτω σχήμα:



Τα αστέρια στο σχήμα συμβολίζουν τα *κέντρα* των τελικών ομάδων. Η εταιρία μπορεί να αξιοποιήσει αυτή τη συσταδοποίηση για να προσαρμόσει τις διαστάσεις των μεγεθών της ανάλογα. Φυσικά, αν ήθελε να έχει και άλλα μεγέθη (π.χ. extra small, extra large κλπ.) θα μπορούσε να εφαρμόσει έναν αλγόριθμο πέντε ή παραπάνω μέσων.

Οι αλγόριθμοι συσταδοποίησης χωρίζονται σε *επίπεδους* (ένα επίπεδο κατηγοριοποίησης, όπως στο παραπάνω παράδειγμα), και *ιεραρχικούς* (τα στοιχεία εισόδου χωρίζονται σε ομάδες, οι οποίες χωρίζονται σε υποομάδες κ.ο.κ.. Επίσης χωρίζονται σε *σαφείς* (κάθε στοιχείο ανήκει σε μια μόνο ομάδα) και *ασαφείς* (για κάθε στοιχείο βρίσκουμε την πιθανότητα να ανήκει σε μια από τις ομάδες). Τέλος, οι κανόνες συσταδοποίησης δεν έχουν πάντα σχέση μόνο με την απόσταση των παραδειγμάτων από κάποιο κέντρο, αλλά συχνά λαμβάνουν υπόψη και παράμετρους όπως η *πυκνότητα κατανομής* των παραδειγμάτων.

Εκτός από την συσταδοποίηση k-μέσων, μερικοί άλλοι πολύ χρήσιμοι αλγόριθμοι συσταδοποίησης είναι ο Mean Shift, η διαδικασία Dirichlet, η φασματική συσταδοποίηση και η μέθοδος μεγιστοποίησης προσδοκιών (E.M. - expectation-maximisation).

### Κανόνες Συσχέτισης

Η μάθηση κανόνων συσχέτισης (association rule learning), είναι μια μέθοδος ανακάλυψης ισχυρών σχέσεων μεταξύ μεταβλητών σε μεγάλες βάσεις δεδομένων, βάσει ενός μέτρου *ενδιαφέροντος*. Η ιδέα των κανόνων συσχέτισης αρχικά προτάθηκε (Agrawal et al., 1993) ως ένας τρόπος ανακάλυψης προϊόντων που αγοράζονται συχνά μαζί, από μεγάλες βάσεις δεδομένων που περιέχουν δεδομένα συναλλαγών σε μεγάλα εμπορικά καταστήματα (*ανάλυση καλαθιού αγοράς - basket analysis*).

Ας υποθέσουμε ότι έχουμε μια λίστα συναλλαγών που περιγράφει τα καλάθια των πελατών που ψώνισαν από ένα σουπερμάρκετ. Ένα απλοποιημένο παράδειγμα είναι το παρακάτω (Βλαχάβας κ.α., 2011). Τα 1 και 0 αντιστοιχούν στο αν αγόρασε αυτό το προϊόν ο πελάτης, ή όχι. Συνήθως δεν μας ενδιαφέρει η ποσότητα στα προβλήματα ανάλυσης καλαθιού αγοράς, αλλά αυτό που ψάχνουμε είναι σχέσεις του τύπου “αν ένας πελάτης αγοράσει καφέ και γάλα, τότε συνήθως αγοράζει και ζάχαρη”.

Καλάθι	Ψωμί	Καφές	Γάλα	Ζάχαρη
#001	1	0	1	0
#002	0	1	0	0
#003	1	0	1	1
#004	0	1	0	1
#005	1	0	1	1
#006	1	1	1	0
#007	1	0	0	1
#008	1	1	1	1

#009	0	0	1	1
#010	1	1	0	1

Έχουμε λοιπόν ένα σύνολο  $I = \{\psi\omega\mu\acute{\iota}, \kappa\alpha\phi\acute{\epsilon}\varsigma, \gamma\acute{\alpha}\lambda\alpha, \zeta\acute{\alpha}\chi\alpha\rho\eta\}$  από διακριτά προϊόντα, και ένα σύνολο  $D = \{\#001, \#002, \dots, \#010\}$  από συναλλαγές, όπου κάθε συναλλαγή  $T$  είναι ένα σύνολο από αντικείμενα, υποσύνολο του  $I$ .

Οι κανόνες συσχέτισης είναι προτάσεις της μορφής  $\{X_1, X_2, \dots, X_n\} \rightarrow Y$ , που σημαίνει ότι αν βρεθούν όλα τα  $X_1, X_2, \dots, X_n$  σε κάποιο καλάθι, τότε είναι πιθανό να βρεθεί και το  $Y$ . Με όρους θεωρίας συνόλων αυτές οι προτάσεις εκφράζονται ως εξής:

$$X \Rightarrow Y, \text{ where } X \subseteq I, Y \subseteq I \text{ and } X \cap Y = \emptyset$$

Το πρώτο μέρος του κανόνα ονομάζεται *υπόθεση (antecedent)*, ενώ το δεύτερο *συμπέρασμα (consequent)*. Βέβαια, για να θεωρήσουμε σημαντικό ένα κανόνα συσχέτισης, θα πρέπει να πληροί κάποια κριτήρια, τα οποία θεσπίζουμε βάσει των παρακάτω ποσοτικών μεγεθών (Bell, 2015):

- *Υποστήριξη (support)* ή *κάλυψη (coverage)*: Η υποστήριξη εκφράζει την πιθανότητα να βρεθεί το σύνολο  $\{X_1, X_2, \dots, X_n, Y\}$  στην βάση δεδομένων (λίστα συναλλαγών):

$$\text{supp}(X) = \frac{\text{basket containing } X}{\text{total number of baskets}}$$

- *Εμπιστοσύνη (confidence)* ή *ακρίβεια (accuracy)*: Η εμπιστοσύνη εκφράζει την πιθανότητα να βρεθεί το  $Y$  σε ένα καλάθι που περιέχει τα  $X_1, X_2, \dots, X_n$ , και ισούται με το λόγο των καλάθιων  $\{X_1, X_2, \dots, X_n, Y\}$  ως προς το σύνολο των καλάθιων που περιλαμβάνουν τα  $X_1, X_2, \dots, X_n$ :

$$\text{conf}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}$$

Το ζητούμενο της μάθησης κανόνων συσχέτισης είναι να ανακαλύψουμε όλους τους κανόνες που έχουν υποστήριξη μεγαλύτερη από μια ελάχιστη τιμή υποστήριξης που ονομάζεται *κατώφλι υποστήριξης (minsupp)* και εμπιστοσύνη μεγαλύτερη από μια ελάχιστη τιμή εμπιστοσύνης που ονομάζεται *κατώφλι εμπιστοσύνης (minconf)*, οι οποίες ορίζονται από τον χρήστη. Αυτή η αναζήτηση μπορεί να διαχωριστεί στα παρακάτω υποπροβλήματα (Κουρής, 2006):



1. Εύρεση όλων των αντικειμένων που έχουν υποστήριξη μεγαλύτερη από το κατώφλι υποστήριξης. Αυτά αποκαλούνται συχνά αντικείμενα, και όλα τα υπόλοιπα αποκαλούνται σπάνια.
2. Παραγωγή όλων των κανόνων συσχέτισης που ικανοποιούν το κατώφλι εμπιστοσύνης, από τα συχνά αντικείμενα.

Ο πιο κλασικός αλγόριθμος για την μάθηση κανόνων συσχέτισης είναι ο αλγόριθμος *Apriori* (Agrawal and Srikant, 1994), ο οποίος, όπως και οι περισσότεροι αλγόριθμοι αναζήτησης κανόνων συσχέτισης, στηρίζεται στην ιδιότητα της *μονοτονίας* (*monotonicity property*) ή αλλιώς ιδιότητα *a priori* (Βλαχάβας κ.α., 2011), σύμφωνα με την οποία “αν ένα σύνολο αντικειμένων είναι συχνό, τότε όλα τα υποσύνολα του είναι επίσης συχνά”. Ο αλγόριθμος αυτός ακολουθεί την διαδικασία αναζήτησης που αναφέρθηκε παραπάνω, ως εξής:

- Σε κάθε επίπεδο  $k$  του αλγορίθμου, εξετάζονται όλα τα υποψήφια σύνολα  $k$  συχνών αντικειμένων  $C_k$  και χαρακτηρίζονται ως συχνά όσα έχουν υποστήριξη μεγαλύτερη από το κατώφλι υποστήριξης. Μετά το πρώτο επίπεδο, εξετάζονται μόνο σύνολα που αποτελούνται από συχνά υποσύνολα (αξιοποιώντας την ιδιότητα *a priori*) και τα υπόλοιπα απορρίπτονται. Το σύνολο που προκύπτει κάθε φορά λέγεται σύνολο συχνών αντικειμένων και συμβολίζεται με  $L_k$ .
- Κάνοντας όλους τους δυνατούς συνδυασμούς μεταξύ των συχνών αντικειμένων, προκύπτει το σύνολο υποψηφίων συνόλων  $k$  αντικειμένων του επόμενου επιπέδου  $C_{k+1}$ .
- Η διαδικασία αυτή επαναλαμβάνεται μέχρι να βρεθούν τα συχνά σύνολα αντικειμένων ενός προκαθορισμένου επιπέδου (ή του μέγιστου δυνατού)
- Όταν φτάσουμε στο τελευταίο επίπεδο του αλγορίθμου, ελέγχεται η εμπιστοσύνη όλων των πιθανών κανόνων που μπορούν να προκύψουν από το μέγιστο συχνό σύνολο αντικειμένων, σύμφωνα με το κατώφλι εμπιστοσύνης.

Στην πράξη, αν εφαρμόσουμε τον αλγόριθμο *apriori* στο παράδειγμα του σουπερμάρκετ:

- Έστω ότι επιλέγουμε σαν κατώφλι υποστήριξης  $\text{minsupp} = 35\%$  και σαν κατώφλι εμπιστοσύνης  $\text{minconf} = 80\%$
- Αρχικά υπολογίζεται η υποστήριξη όλων των προϊόντων (που αποτελούν ουσιαστικά το  $C_1$ ), ένα προς ένα, και δημιουργείται το σύνολο  $L_1$ .
  - $\text{supp}(\{\psi\omega\mu\acute{\iota}\}) = 7/10 = 70\% > \text{minsupp}$
  - $\text{supp}(\{\kappa\alpha\phi\acute{\epsilon}\varsigma\}) = 5/10 = 50\% > \text{minsupp}$
  - $\text{supp}(\{\gamma\acute{\alpha}\lambda\alpha\}) = 6/10 = 60\% > \text{minsupp}$
  - $\text{supp}(\{\zeta\acute{\alpha}\chi\alpha\rho\eta\}) = 7/10 = 70\% > \text{minsupp}$
  - Άρα  $L_1 = \{\psi\omega\mu\acute{\iota}, \kappa\alpha\phi\acute{\epsilon}\varsigma, \gamma\acute{\alpha}\lambda\alpha, \zeta\acute{\alpha}\chi\alpha\rho\eta\}$
- Στη συνέχεια δημιουργείται το  $C_2$  από όλους τους δυνατούς συνδυασμούς του  $L_1$ 
  - $C_2 = \{ \{\psi\omega\mu\acute{\iota}, \kappa\alpha\phi\acute{\epsilon}\varsigma\}, \{\psi\omega\μ\acute{\iota}, \gamma\acute{\alpha}\lambda\alpha\}, \{\psi\omega\μ\acute{\iota}, \zeta\acute{\alpha}\chi\alpha\rho\eta\}, \{\kappa\alpha\phi\acute{\epsilon}\varsigma, \gamma\acute{\alpha}\lambda\alpha\}, \{\kappa\alpha\phi\acute{\epsilon}\varsigma, \zeta\acute{\alpha}\chi\alpha\rho\eta\}, \{\gamma\acute{\alpha}\lambda\alpha, \zeta\acute{\alpha}\chi\alpha\rho\eta\} \}$

- Τώρα υπολογίζεται η υποστήριξη των μελών του  $C_2$  και απορρίπτονται όσα υποσύνολα δεν ξεπερνούν το κατώφλι υποστήριξης, ώστε να δημιουργηθεί στη συνέχεια το  $L_2$ 
  - $\text{supp}(\{\psi\omega\acute{\mu}\acute{\iota}, \text{καφ}\acute{\epsilon}\varsigma\}) = 3/10 = 30\% < \text{minsupp}$  (απορρίπτεται)
  - $\text{supp}(\{\psi\omega\acute{\mu}\acute{\iota}, \text{γ}\acute{\alpha}\lambda\alpha\}) = 5/10 = 50\% > \text{minsupp}$
  - $\text{supp}(\{\psi\omega\acute{\mu}\acute{\iota}, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\}) = 4/10 = 40\% > \text{minsupp}$
  - $\text{supp}(\{\text{καφ}\acute{\epsilon}\varsigma, \text{γ}\acute{\alpha}\lambda\alpha\}) = 2/10 = 20\% < \text{minsupp}$  (απορρίπτεται)
  - $\text{supp}(\{\text{καφ}\acute{\epsilon}\varsigma, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\}) = 3/10 = 30\% < \text{minsupp}$  (απορρίπτεται)
  - $\text{supp}(\{\text{γ}\acute{\alpha}\lambda\alpha, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\}) = 4/10 = 40\% > \text{minsupp}$
  - Άρα  $L_2 = \{ \{\psi\omega\acute{\mu}\acute{\iota}, \text{γ}\acute{\alpha}\lambda\alpha\}, \{\psi\omega\acute{\mu}\acute{\iota}, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\}, \{\text{γ}\acute{\alpha}\lambda\alpha, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\} \}$
- Εδώ ενώνονται τα ζεύγη μεταξύ τους για να δημιουργήσουν τριάδες. Εν προκειμένω, υπάρχει μόνο μια:  $\{\psi\omega\acute{\mu}\acute{\iota}, \text{γ}\acute{\alpha}\lambda\alpha, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\}$ . Άρα  $C_3 = \{\psi\omega\acute{\mu}\acute{\iota}, \text{γ}\acute{\alpha}\lambda\alpha, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\}$ 
  - $\text{supp}(\{\psi\omega\acute{\mu}\acute{\iota}, \text{γ}\acute{\alpha}\lambda\alpha, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\}) = 3/10 = 30\% < \text{minsupp}$  (απορρίπτεται)
  - Άρα  $L_3 = \{ \}$
- Εφόσον το  $L_3$  είναι κενό, ο αλγόριθμος εύρεσης συχνών αντικειμένων σταματά και το μέγιστο συχνό σύνολο αντικειμένων είναι το  $L_2$ .
- Σε αυτό το σημείο ελέγχεται η εμπιστοσύνη όλων των πιθανών κανόνων που μπορούν να προκύψουν από τα υποσύνολα του  $L_2$ :
  - $\{\psi\omega\acute{\mu}\acute{\iota}, \text{γ}\acute{\alpha}\lambda\alpha\}$ 
    - $\text{conf}(\psi\omega\acute{\mu}\acute{\iota} \rightarrow \text{γ}\acute{\alpha}\lambda\alpha) = 5/7 = 71\% < \text{minconf}$  (απορρίπτεται)
    - $\text{conf}(\text{γ}\acute{\alpha}\lambda\alpha \rightarrow \psi\omega\acute{\mu}\acute{\iota}) = 5/6 = 83\% > \text{minconf}$  (εγκρίνεται)
  - $\{\psi\omega\acute{\mu}\acute{\iota}, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\}$ 
    - $\text{conf}(\psi\omega\acute{\mu}\acute{\iota} \rightarrow \text{ζ}\acute{\alpha}\chi\alpha\rho\eta) = 5/7 = 71\% < \text{minconf}$  (απορρίπτεται)
    - $\text{conf}(\text{ζ}\acute{\alpha}\chi\alpha\rho\eta \rightarrow \psi\omega\acute{\mu}\acute{\iota}) = 5/7 = 71\% < \text{minconf}$  (απορρίπτεται)
  - $\{\text{γ}\acute{\alpha}\lambda\alpha, \text{ζ}\acute{\alpha}\chi\alpha\rho\eta\}$ 
    - $\text{conf}(\text{γ}\acute{\alpha}\lambda\alpha \rightarrow \text{ζ}\acute{\alpha}\chi\alpha\rho\eta) = 4/6 = 66\% < \text{minconf}$  (απορρίπτεται)
    - $\text{conf}(\text{ζ}\acute{\alpha}\chi\alpha\rho\eta \rightarrow \text{γ}\acute{\alpha}\lambda\alpha) = 4/7 = 57\% < \text{minconf}$  (απορρίπτεται)
- Τελικά, παράγεται μόνο ένας κανόνας:  $\text{γ}\acute{\alpha}\lambda\alpha \rightarrow \psi\omega\acute{\mu}\acute{\iota}$ , δηλαδή *όποιος αγοράζει γάλα, συνήθως αγοράζει και ψωμί*.

Στην πράξη, ο αναλυτής που εφαρμόζει αυτήν τη μέθοδο ξεκινά από ένα υψηλά κατώφλια υποστήριξης και εμπιστοσύνης, τα οποία σταδιακά μειώνει μέχρι να εξάγει όσους κανόνες συσχέτισης επιθυμεί (Ρεφανίδης, 2011). Η μείωση του κατωφλιού υποστήριξης ανεβάζει σημαντικά το χρόνο λειτουργίας του αλγορίθμου, αλλά αυξάνει τα υποψήφια αντικείμενα, ενώ η μείωση του κατωφλιού εμπιστοσύνης αυξάνει τον αριθμό των κανόνων συσχέτισης που παράγονται. Για παράδειγμα, αν στην περίπτωση που μελετάμε εδώ μειώσουμε το κατώφλι εμπιστοσύνης στο 70%, θα παρατηρήσουμε ότι προκύπτουν άλλοι τρεις κανόνες συσχέτισης.

Φυσικά, υπάρχουν πιο εξελιγμένοι αλγόριθμοι από τον Apriori που εκτελούν την αναζήτηση κανόνων συσχέτισης πιο γρήγορα απαιτώντας λιγότερη μνήμη και υπολογιστική ισχύ, όπως ο

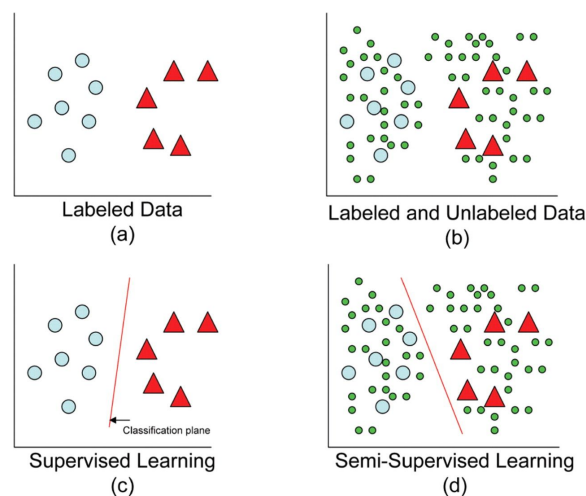
Equivalence Class Transformation (Zaki, 2000) και ο Frequent Pattern Growth (Han et al., 2004).

## Μάθηση με μερική επίβλεψη

Οι αλγόριθμοι που εντάσσονται στην κατηγορία της μάθησης με μερική επίβλεψη (semi-supervised learning, αναφέρεται και ως ημιεπιτηρούμενη μάθηση), συνήθως αφορούν προβλήματα ταξινόμησης, για τα οποία όμως θέλουμε να χρησιμοποιήσουμε σαν είσοδο και χαρακτηρισμένα, αλλά και μη χαρακτηρισμένα δεδομένα. Συνήθως, εφαρμόζουμε έναν αλγόριθμο ταξινόμησης σε ένα μικρό χαρακτηρισμένο σύνολο, και μετά προσαρμόζουμε αυτήν την ταξινόμηση σε ένα μη χαρακτηρισμένο σύνολο χρησιμοποιώντας έναν αλγόριθμο συσταδοποίησης, όπως τον E. M. Οι πιο σημαντικοί λόγοι για την χρήση μάθησης με μερική επίβλεψη είναι οι παρακάτω (Witten et al., 2011):

- Αν δεν έχουμε στην διάθεσή μας πολλά χαρακτηρισμένα δεδομένα, ένας αλγόριθμος κατηγοριοποίησης μπορεί να αντιμετωπίζει προβλήματα υψηλής διακύμανσης. Η διακύμανση μπορεί να μειωθεί προσθέτοντας και άλλα δεδομένα, και στην περίπτωση της μάθησης με μερική επίβλεψη, αυτά δεν χρειάζεται να είναι χαρακτηρισμένα.
- Η ακρίβεια πολλών αλγορίθμων συσταδοποίησης βελτιώνεται αν κάποια από τα δεδομένα τους είναι ήδη χαρακτηρισμένα (ανήκουν ήδη σε ομάδες).
- Όταν ο όγκος των δεδομένων είναι πολύ μεγάλος, συχνά είναι πολύ δύσκολο και χρονοβόρο να χαρακτηρίσουμε όλα τα εκπαιδευτικά παραδείγματα, για να εκτελέσουμε έναν αλγόριθμο ταξινόμησης. Επομένως προτιμούμε να χαρακτηρίσουμε μόνο μερικά αντιπροσωπευτικά από αυτά, ώστε τα υπόλοιπα να συσταδοποιηθούν αυτόματα με αυτά που τους αντιστοιχούν.

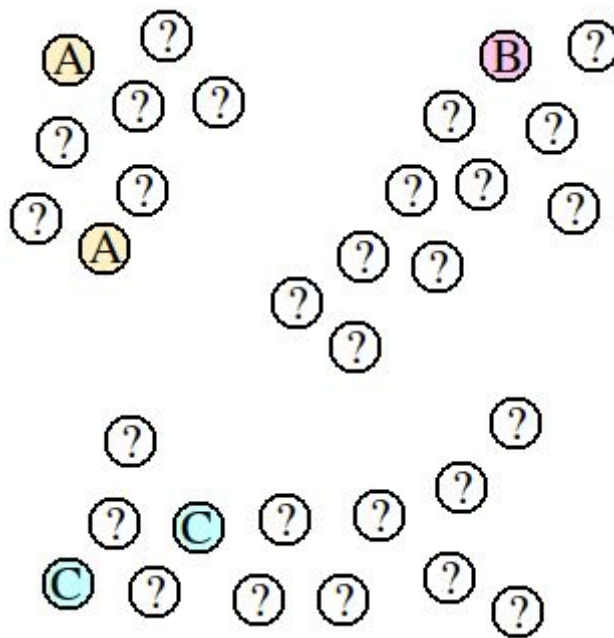
Στην παρακάτω εικόνα μπορούμε να δούμε ένα παράδειγμα βελτίωσης της ακρίβειας ενός ταξινομητή, με τη χρήση μη χαρακτηρισμένων δεδομένων:



(πηγή: <http://bioinformatics.oxfordjournals.org/content/24/6/783/F1.expansion.html>)

Υπάρχουν όμως και αλγόριθμοι μάθησης με μερική επίβλεψη που ακολουθούν μια διαφορετική μεθοδολογία: αντί να διαμορφώσουν ένα ταξινομητή βάσει των χαρακτηρισμένων δεδομένων, ο οποίος στη συνέχεια προσαρμόζεται στα μη χαρακτηρισμένα, προσπαθούν να χαρακτηρίσουν άμεσα κάθε μη χαρακτηρισμένο αντικείμενο βάσει των χαρακτηρισμένων αντικειμένων στη συστάδα του, χωρίς να δημιουργούν κάποιο γενικό κανόνα (ταξινομητή) για τον χαρακτηρισμό νέων αντικειμένων. Αυτή η διαδικασία μπορεί να βοηθήσει σε περιπτώσεις όπου η δημιουργία ενός γενικού κανόνα είτε είναι δύσκολη, είτε δεν είναι απαραίτητη. Το μειονέκτημα, φυσικά, είναι ότι για την εκτίμηση νέων δεδομένων η διαδικασία θα πρέπει να επαναληφθεί από την αρχή.

Στο παρακάτω παράδειγμα, υπάρχουν πολλά μη χαρακτηρισμένα εκπαιδευτικά παραδείγματα, και μόλις πέντε χαρακτηρισμένα. Ένας αλγόριθμος που επιχειρεί πρώτα ταξινόμηση βάσει των χαρακτηρισμένων παραδειγμάτων, δεν θα μπορούσε να δημιουργήσει ένα ιδιαίτερα αξιόπιστο μοντέλο από αυτά: για παράδειγμα, τα στοιχεία που βρίσκονται στο κέντρο του χώρου των παραδειγμάτων θα ταξινομούσαν λογικά σαν στοιχεία A, εφόσον υπάρχουν δυο στοιχεία A κοντά σε αυτήν την περιοχή. Ένας αλγόριθμος που λαμβάνει εξ αρχής υπόψη όλα τα εκπαιδευτικά δεδομένα και τα συσταδοποιεί, εύκολα θα διαπιστώσει ότι τα παραδείγματα που βρίσκονται στο κέντρο του χώρου ανήκουν σε μια προφανή κατηγορία, και στη συνέχεια θα τα χαρακτηρίσει ως B, σύμφωνα με τα χαρακτηρισμένα στοιχεία της κατηγορίας:



(πηγή: [https://en.wikipedia.org/wiki/Transduction\\_\(machine\\_learning\)](https://en.wikipedia.org/wiki/Transduction_(machine_learning)))

Εφόσον σε αυτούς τους αλγόριθμους δεν δημιουργείται κάποιο μοντέλο από τα δεδομένα, δεν μιλάμε πλέον για *επαγωγική μάθηση*: αυτό το είδος μάθησης ονομάζεται *μεταγωγή (transduction)*, και είναι ιδιαίτερα αποτελεσματικό όταν η γραφική αναπαράσταση των

δεδομένων υπονοεί μια συγκεκριμένη δομή, αλλά τα χαρακτηρισμένα δεδομένα είναι πολύ λίγα για να εξάγουν ένα χρήσιμο ταξινομητή (Chapelle et al., 2006).

Η μεταγωγή βασίζεται στην *αρχή του Varnik*: όταν προσπαθείς να λύσεις ένα πρόβλημα, δεν πρέπει να προσπαθείς να λύσεις ένα ακόμα δυσκολότερο πρόβλημα σαν ενδιάμεσο βήμα. Εν προκειμένω, ο Varnik θεωρεί ότι αν ο στόχος σου είναι να χαρακτηρίσεις μόνο ένα συγκεκριμένο σύνολο αντικειμένων, δεν υπάρχει λόγος να βρεις ένα μοντέλο που χαρακτηρίζει όλα τα πιθανά αντικείμενα στο χώρο τους. (Varnik, 1999).

## Ενισχυτική μάθηση

Στις τεχνικές της ενισχυτικής μάθησης (*reinforcement learning*), το σύστημα μαθαίνει μέσα από την άμεση αλληλεπίδραση με το περιβάλλον του, με ένα τρόπο που παραπέμπει στη μάθηση μέσω *επιβράβευσης και τιμωρίας*, η οποία παρατηρείται στους ζωικούς οργανισμούς.

Η βασική διαδικασία είναι η εξής: η οντότητα που μαθαίνει και παίρνει τις αποφάσεις ονομάζεται *πράκτορας (agent)*, ενώ οτιδήποτε άλλο εκτός του πράκτορα ονομάζεται *περιβάλλον (environment)*. Ο πράκτορας επιλέγει μια ενέργεια κάθε φορά, και το περιβάλλον ανταποκρίνεται σε αυτήν *αλλάζοντας κατάσταση (state)* ανάλογα. Επίσης, μετά από κάθε ενέργεια το περιβάλλον στέλνει στον πράκτορα ένα αριθμητικό σήμα που ονομάζεται *ανταμοιβή (reward)*, το οποίο περιγράφει κατά πόσο η απόφαση του πράκτορα επηρέασε το περιβάλλον θετικά ή αρνητικά, ανάλογα με το μέτρο που θέλουμε να μεγιστοποιήσουμε, ή να ελαχιστοποιήσουμε. Ο πράκτορας δοκιμάζει όλες τις ενέργειες που μπορεί να κάνει και καταγράφει την ανταμοιβή για κάθε μια, είτε βάσει της κατάστασης του περιβάλλοντος μετά από μια ενέργεια, είτε μετά από διαδοχικές ενέργειες, σε περίπτωση που μας ενδιαφέρει η ανταμοιβή σε βάθος χρόνου. Ανάλογα με την ανταμοιβή για κάθε ενέργεια, ο πράκτορας αναθέτει σε κάθε μια την πιθανότητα να την επιλέξει ώστε να προχωρήσει από την τρέχουσα κατάσταση του περιβάλλοντος (Βλαχάβας κ.α., 2011)

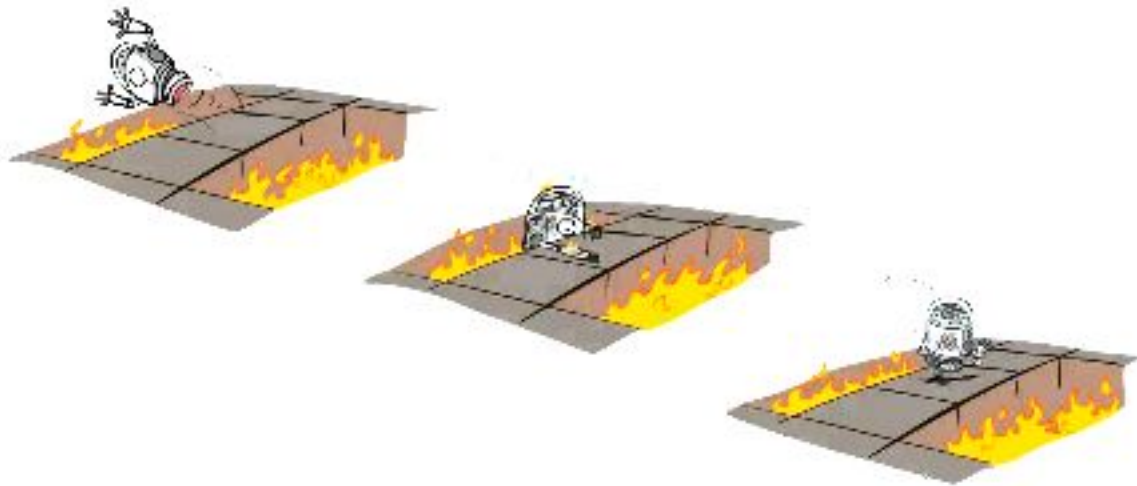
Σε αντίθεση με την μάθηση με επίβλεψη, το σύστημα δεν καθοδηγείται από κάποιον τρίτο μέσω θετικών και αρνητικών παραδειγμάτων, αλλά από μόνο του πρέπει να ανακαλύψει ποιες ενέργειες θα του επιφέρουν το μεγαλύτερο κέρδος. Λέμε λοιπόν ότι το σύστημα δεν μαθαίνει από ένα *δάσκαλο*, ο οποίος γνωρίζει τι πρέπει να γίνει και τι όχι, αλλά από ένα *κριτή (critic)*, ο οποίος γνωρίζει μόνο πόσο καλά τα πήγε ο αλγόριθμος στο παρελθόν, και συχνά αυτό μπορεί να μας το πει μόνο αφότου ο σκοπός έχει επιτευχθεί.

Για να δώσουμε ένα παράδειγμα, μπορούμε να υποθέσουμε ότι το πρόβλημά μας είναι η διάσχιση ενός λαβύρινθου από ένα ρομπότ: το ρομπότ (ο πράκτορας) μπορεί να κινηθεί κάθε φορά σε μια από τις τέσσερις κατευθύνσεις, η κατάσταση του περιβάλλοντος κάθε φορά είναι η θέση του ρομπότ στο λαβύρινθο, και μπορούμε να υποθέσουμε ότι θέλουμε να ελαχιστοποιήσουμε τον αριθμό των κινήσεων που θα κάνει το ρομπότ για να διασχίσει επιτυχώς το λαβύρινθο. Η ανταμοιβή γίνεται γνωστή στον πράκτορα κάθε φορά που φτάνει στο τέλος της διαδρομής: τότε μπορεί να αναθέσει μια αριθμητική τιμή ανταμοιβής σε κάθε κίνηση που έκανε, ανάλογα με το πόσες κινήσεις του πήρε για να φτάσει στο τέλος μετά από

κάθε κίνηση. Όσες περισσότερες φορές το πρόγραμμα διασχίσει επιτυχώς το λαβύρινθο αρκετές φορές, τόσο καλύτερα θα μπορεί να εκτιμήσει την αξία ανταμοιβής της κάθε κίνησης ξεχωριστά, και έτσι κάθε φορά θα μπορεί να διασχίσει το λαβύρινθο ακόμα πιο γρήγορα, μέχρι να καταλήξει στην γρηγορότερη διαδρομή (Alpaydin, 2010).

## Active Reinforcement Learning

---



*Ενισχυτική μάθηση στην πράξη!*

(πηγή: <http://rasas.me/learning/cs188-1x-artificial-intelligence>)

Συνήθως για αλγόριθμους τέτοιας μορφής, το περιβάλλον διαμορφώνεται ως μια διαδικασία απόφασης Markov. Επειδή οι αλγόριθμοι αυτού του είδους συνήθως λειτουργούν σε πραγματικό χρόνο, είναι σημαντικό να βρούμε μια ισορροπία ανάμεσα στο βαθμό εξερεύνησης του μοντέλου, και τον βαθμό αξιοποίησης της ήδη κεκτημένης γνώσης για το μοντέλο. Στο παράδειγμα του ρομπότ, αυτό σημαίνει ότι πρέπει να αποφασίσουμε σε τι βαθμό θα κάνει δοκιμές ο αλγόριθμος, εξετάζοντας όλες τις πιθανές ανεξερεύνητες διαδρομές, πριν αποφασίσει να κάνει μια κίνηση, ανάλογα με το αν μας ενδιαφέρει η ταχύτητα επίλυσης του προβλήματος, δηλαδή το να βρει γρήγορα μια αποδεκτά σύντομη διαδρομή, ή η ακρίβεια αυτής της λύσης, δηλαδή να βρει οπωσδήποτε την συντομότερη διαδρομή.

## Άλλοι τύποι προβλημάτων

### Μείωση διαστάσεων

Τα προβλήματα μείωσης διαστάσεων (dimensionality reduction) αφορούν στην προσπάθεια περιορισμού του αριθμού των μεταβλητών που θα λάβουμε υπόψη για να αντιμετωπίσουμε ένα πρόβλημα μηχανικής μάθησης. Υπάρχουν διάφοροι λόγοι για τους οποίους θέλουμε να

επιτύχουμε αυτήν τη “συμπύκνωση” ενός προβλήματος, με σημαντικότερους εκ των οποίων τους εξής (Witten, 2011):

- Η επιτάχυνση του αλγορίθμου μηχανικής μάθησης που θα χρησιμοποιηθεί.
- Η μείωση του απαιτούμενου αποθηκευτικού χώρου και της απαιτούμενης υπολογιστικής ισχύος.
- Η ευκολότερη παρουσίαση του προβλήματος και των αποτελεσμάτων.
- Η αύξηση της ακρίβειας των αποτελεσμάτων, λόγω της μείωσης της επιρροής των λιγότερο σημαντικών μεταβλητών.

Πέρα από τις στοιχειώδεις περιπτώσεις όπου η μείωση των διαστάσεων του προβλήματος μπορεί να γίνει πολύ απλά από τον ίδιο τον αναλυτή, για παράδειγμα όταν δύο ή περισσότερες μεταβλητές είναι ταυτόσημες ή είναι απόλυτα συσχετιζόμενες, η κύρια λειτουργία ενός αλγορίθμου μείωσης διαστάσεων είναι να εντοπίζει μεταβλητές που ενδεχομένως έχουν κάποιο βαθμό συσχέτισης, και να τις μετασχηματίζει σε νέες μεταβλητές με πολύ μικρότερο βαθμό συσχέτισης, μειώνοντας έτσι τον αριθμό των μεταβλητών (ή αν θέλουμε να το δούμε γεωμετρικά, μειώνοντας τις *διαστάσεις* του χώρου στον οποίο μπορούμε να αποτυπώσουμε τις διάφορες παρατηρήσεις του προβλήματος ως διανύσματα) χωρίς μεγάλη απώλεια ακρίβειας του μοντέλου που θα δημιουργηθεί από το νέο σύνολο μεταβλητών.

Η πιο δημοφιλής μέθοδος μείωσης διαστάσεων είναι η *Μέθοδος Ανάλυσης Κυρίων Συνιστωσών* (Principal Component Analysis - PCA). Οι βασικές αρχές της μεθόδου PCA είναι οι εξής (Ng, 2011):

- Ο στόχος είναι η μετατροπή όλων των μεταβλητών του μοντέλου σε μεταβλητές που δεν συσχετίζονται γραμμικά μεταξύ τους
- Η μετατροπή αυτή θα στηριχτεί πάνω στις λεγόμενες *κύριες συνιστώσες* του μοντέλου, δηλαδή σε αυτές που επιδεικνύουν την μεγαλύτερη απόκλιση τιμών στο αρχικό σύνολο δεδομένων (προϋποθέτοντας πάντα ότι οι τιμές των μεταβλητών έχουν πρώτα μετατραπεί σε παρόμοια κλίμακα).

Η επιλογή των κύριων συνιστωσών του προβλήματος, με απλά λόγια, μας επιτρέπει να απλοποιήσουμε το πρόβλημα ούτως ώστε η απώλεια πληροφορίας να είναι αντιστρόφως ανάλογη της επιρροής της κάθε μεταβλητής στο τελικό αποτέλεσμα.

Έρευνες όπως αυτή των Hinton και Salakhutdinov (2006), έχουν αποδείξει ότι για σύνολα δεδομένων με πάρα πολλές μεταβλητές (πολύ μεγάλης διάστασης), κάποιες μέθοδοι Νευρωνικών Δικτύων μπορούν να επιτύχουν πολύ καλύτερα αποτελέσματα από απλούστερες μεθόδους όπως η PCA σε προβλήματα μείωσης διαστάσεων.

## Ανίχνευση ανωμαλιών

Τα προβλήματα ανίχνευσης ανωμαλιών (anomaly detection), ή αλλιώς ανίχνευσης ακραίων τιμών (outlier detection), σχετίζονται με την εύρεση και ενδεχόμενη απομάκρυνση

παρατηρήσεων ενός συνόλου δεδομένων, οι οποίες δεν εναρμονίζονται με τις υπόλοιπες παρατηρήσεις, ή είναι εκτός του διαστήματος των αποδεκτών τιμών για αυτά τα δεδομένα, ή ακολουθούν κάποιο ασυνήθιστο μοτίβο για αυτόν τον τύπο δεδομένων. Αυτές οι ανωμαλίες, ή ακραίες τιμές, μπορούν να προκύψουν σε ένα σύνολο δεδομένων είτε λόγω κάποιας τεχνικής βλάβης στο σύστημα συλλογής των δεδομένων, είτε από ανθρώπινο λάθος, είτε λόγω κάποιας πραγματικής αλλά απροσδόκητης απόκλισης από την φυσιολογική συμπεριφορά του δείγματος, είτε ακόμα και από ενδεχόμενη κακόβουλη απόπειρα δολιοφθοράς του συνόλου δεδομένων (Hodge and Austin, 2004).

Οι ανωμαλίες δημιουργούν σημαντικά προβλήματα στην αξιοποίηση ενός συνόλου δεδομένων, καθώς μπορούν να παραπλανήσουν τον αναλυτή, οδηγώντας τον σε εντελώς διαφορετικά συμπεράσματα. Ακόμη χειρότερα, αν αυτές οι ανωμαλίες παρατηρούνται σε ένα ευαίσθητο σύνολο δεδομένων όπως η βάση των συναλλαγών μιας τράπεζας, ή η βάση του ιστορικού ασθενών ενός νοσοκομείου, μπορούν να έχουν πραγματικά καταστροφικές συνέπειες. Μια άλλη πολύ συχνή εφαρμογή αυτών των αλγορίθμων βρίσκεται στην ανάλυση δεδομένων χρήσης ενός συστήματος σε πραγματικό χρόνο: ενδεχόμενη απότομη αλλαγή στην συμπεριφορά του χρήστη μπορεί να είναι σημάδι εισβολής ενός ξένου στοιχείου στο σύστημα, όπως για παράδειγμα ενός hacker σε ένα δίκτυο υπολογιστών (Omar et al., 2013). Ο ρόλος ενός αλγορίθμου ανίχνευσης ανωμαλιών, συνεπώς, είναι να εντοπίζει έγκαιρα αυτές τις προβληματικές παρατηρήσεις στο σύνολο δεδομένων, ώστε αυτές να διορθώνονται ή να απομακρύνονται πριν το πρόβλημα αποκτήσει μεγαλύτερες διαστάσεις, καθώς και να διευκολύνει τον εντοπισμό των αιτίων τους.

Βεβαίως, οι ανωμαλίες σε ένα σύνολο δεδομένων δεν είναι πάντα κάτι κακό, αλλά αντίθετα μπορεί να είναι ακριβώς αυτό που επιθυμούμε να δούμε σε αυτό το σύνολο. Για παράδειγμα, ένας χρηματιστής μπορεί να αναζητεί απότομες αλλαγές σε ρυθμούς μεταβολής μετοχών, ή ένας ερευνητής μπορεί να αναζητεί ασυνήθιστες μοριακές δομές σε ένα σύνολο βιολογικών δεδομένων.

Στα πλαίσια της παρούσας εργασίας, εφόσον το κεντρικό θέμα είναι οι εφαρμογές αλγορίθμων νευρωνικών δικτύων, αξίζει να σημειωθεί ότι οι αλγόριθμοι ανίχνευσης ανωμαλιών χρησιμοποιούνται πολύ συχνά ως ένα πρώτο βήμα πριν την χρήση του ίδιου του νευρωνικού δικτύου, με στόχο τον “καθαρισμό” του δείγματος από ακραίες τιμές, κάτι που, ιδίως σε προβλήματα ταξινόμησης, μπορεί προσφέρει μεγαλύτερη ακρίβεια και ταύτιση με την πραγματικότητα στις τιμές εξόδου του νευρωνικού δικτύου (Smith and Martinez, 2011).



## 2. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Τα τεχνητά νευρωνικά δίκτυα είναι συστήματα επεξεργασίας πληροφοριών, των οποίων η γενική φιλοσοφία και δομή, καθώς και κάποια βασικά χαρακτηριστικά, είναι εμπνευσμένα από τα βιολογικά νευρωνικά δίκτυα των σύνθετων οργανισμών όπως οι άνθρωποι και τα ζώα, ή, ακριβέστερα, αποτελούν γενίκευση της εικόνας που έχουμε για τον τρόπο λειτουργίας των βιολογικών νευρωνικών δικτύων.

Αφήνοντας για λίγο στην άκρη την σύνδεση με τους βιολογικούς νευρώνες, και βλέποντας τα τεχνητά νευρωνικά δίκτυα από καθαρά προγραμματιστική σκοπιά, μπορούμε να βρούμε πολύ χρήσιμο και περιεκτικό τον λιτό ορισμό που έχει δώσει ο Robert Hecht-Nielsen, καθηγητής του πανεπιστημίου California San Diego και ένας από τους πρωτοπόρους ερευνητές στον κλάδο της επιστήμης του εγκεφάλου και των νευρωνικών δικτύων (Caudill, 1989):

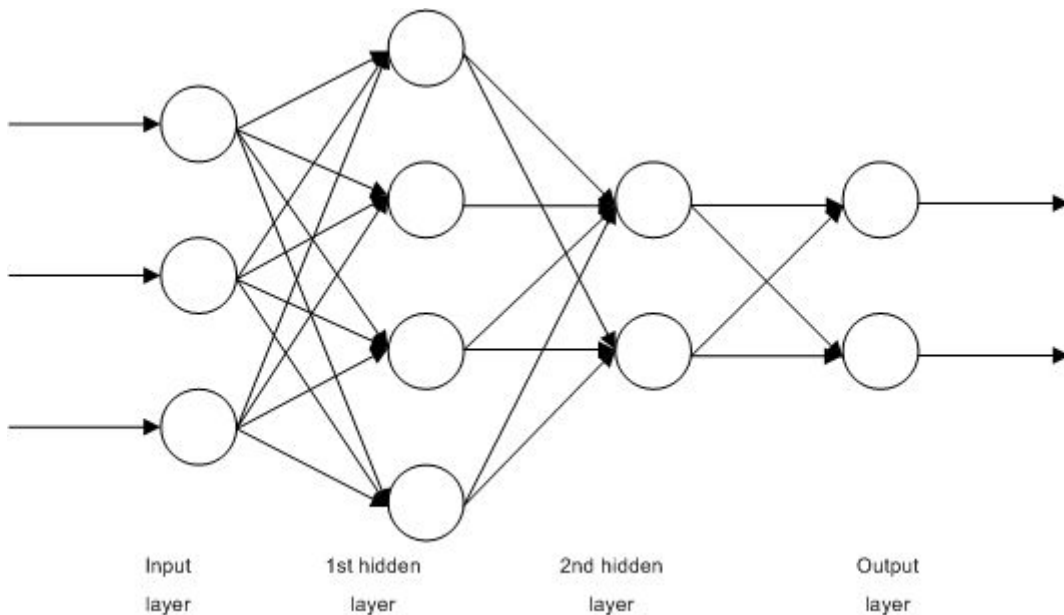
*“Νευρωνικό δίκτυο είναι ένα υπολογιστικό σύστημα που αποτελείται από έναν αριθμό απλών, πυκνά διασυνδεδεμένων μονάδων επεξεργασίας, οι οποίες επεξεργάζονται πληροφορίες μέσω της αντίδρασης της δυναμικής τους κατάστασης σε εξωτερικά ερεθίσματα.”*

### Βασικά χαρακτηριστικά του τεχνητού νευρώνα

Οι διασυνδεδεμένες μονάδες επεξεργασίας ενός τεχνητού νευρωνικού δικτύου ονομάζονται νευρώνες ή κόμβοι του Τ.Ν.Δ., και ο κάθε ένας εξ αυτών επικοινωνεί άμεσα με έναν αριθμό από άλλους νευρώνες του δικτύου μέσω συνδέσμων. Σε κάθε έναν από αυτούς τους συνδέσμους αντιστοιχεί και μια τιμή που αποκαλείται *βάρος* ή *βαρύτητα* (weight), ανάλογα με την οποία το σήμα που εκπέμπεται από τον έναν αντίστοιχο νευρώνα στον άλλο ενισχύεται ή αποδυναμώνεται. Τα βάρη αυτά καθορίζονται από τον *κανόνα μάθησης* (learning rule) του νευρωνικού δικτύου. Στις περισσότερες αρχιτεκτονικές Τ.Ν.Δ., οι κόμβοι οργανώνονται σε *επίπεδα* (layers), ως εξής:

- Οι κόμβοι που δεν δέχονται πληροφορία από άλλους κόμβους του Τ.Ν.Δ., παρά μόνο από εξωτερικές πηγές (όπως μια βάση δεδομένων), λέγονται *κόμβοι εισόδου*, και τους κατατάσσουμε στο πρώτο επίπεδο του Τ.Ν.Δ, το *επίπεδο εισόδου* (input layer). Οι πληροφορίες που εκπέμπουν αυτοί οι κόμβοι είναι συνήθως γνωστές στον χειριστή του Τ.Ν.Δ., και αποτελούν τα δεδομένα του εκάστοτε προβλήματος.
- Οι κόμβοι που δέχονται πληροφορία από άλλους κόμβους αλλά δεν στέλνουν πληροφορία σε επόμενους, λέγονται *κόμβοι εξόδου*, και τους κατατάσσουμε στο τελευταίο επίπεδο του Τ.Ν.Δ., το *επίπεδο εξόδου* (output layer).
- Όλοι οι υπόλοιποι κόμβοι δέχονται πληροφορίες από ένα σύνολο κόμβων και στέλνουν, κατόπιν επεξεργασίας, πληροφορίες σε ένα άλλο σύνολο κόμβων, λέγονται *κρυφοί κόμβοι*, και οργανώνονται σε *κρυφά επίπεδα* (hidden layers). Στη γενική περίπτωση, κάθε κόμβος που ανήκει σε ένα επίπεδο δέχεται πληροφορίες από όλους

τους κόμβους του προηγούμενου επιπέδου, και στέλνει πληροφορίες σε όλους τους κόμβους του επόμενου, όπως απεικονίζεται στο παρακάτω σχήμα:



Σχηματική αναπαράσταση ενός απλού μοντέλου νευρωνικού δικτύου. Οι κόμβοι είναι οργανωμένοι σε επίπεδα, και τα βέλη δείχνουν τη ροή της πληροφορίας κατά τη λειτουργία του.

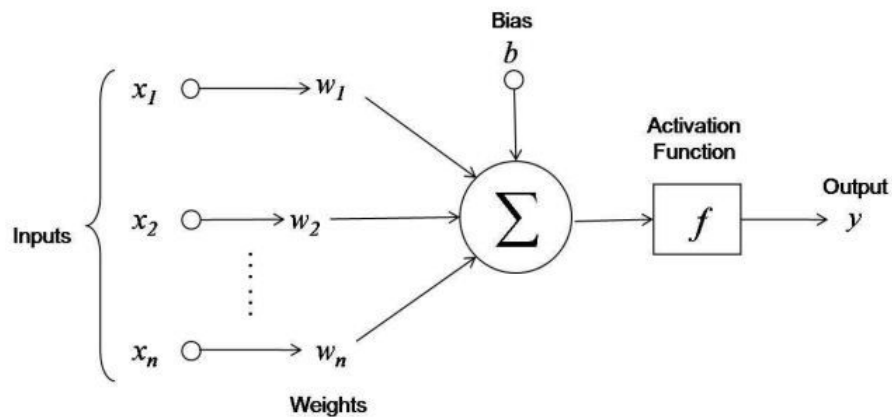
Κάθε νευρώνας έχει μια δυναμική εσωτερική κατάσταση που αποκαλείται *επίπεδο ενεργοποίησης* (activation ή *activity level*). Αυτή η κατάσταση προσδιορίζεται από μια *συνάρτηση ενεργοποίησης*, η οποία δέχεται σαν όρισμα το “άθροισμα” όλων των εξόδων του προηγούμενου επιπέδου του δικτύου, και η τιμή της αποτελεί το επίπεδο ενεργοποίησης, το οποίο τυπικά είναι και η έξοδος του νευρώνα. Το “άθροισμα” που αναφέρθηκε στην προηγούμενη συνάρτηση προκύπτει από μια άλλη συνάρτηση, την *συνάρτηση διάδοσης*, η οποία αθροίζει τα γινόμενα των εξόδων όλων των νευρώνων του προηγούμενου επιπέδου με το αντίστοιχο βάρος της κάθε σύνδεσης. Με μαθηματικούς όρους, το επίπεδο ενεργοποίησης  $a$  ενός νευρώνα  $j$  υπολογίζεται από την συνάρτηση ενεργοποίησης  $f$  ως εξής:

$$a = f(p_j)$$

όπου  $p_j$  είναι η τιμή της συνάρτησης διάδοσης για ένα νευρώνα  $j$ , η οποία λαμβάνοντας υπόψη τις εξόδους  $o_i$  των  $n$  νευρώνων του προηγούμενου επιπέδου που συνδέονται με τον νευρώνα  $j$ , καθώς και τα βάρη  $w_{ij}$  των συνάψεων μεταξύ του νευρώνα και του εκάστοτε προηγούμενού του, έχει τυπικά την παρακάτω μορφή:

$$p_j = \sum_i^n o_i w_{ij}$$

Οι υπολογισμοί αυτοί μπορούν να γίνουν παράλληλα για κάθε νευρώνα του ίδιου επιπέδου. Στο σχήμα που ακολουθεί δίνεται ένα παράδειγμα αυτής της διαδικασίας ενεργοποίησης ενός νευρώνα:



*Σχηματική αναπαράσταση ενός τεχνητού νευρώνα. Οι τιμές εισόδου  $x_1$ ,  $x_2$  και  $x_n$  πολλαπλασιάζονται με τα αντίστοιχα βάρη  $w_1$ ,  $w_2$  και  $w_n$ , και το άθροισμα αυτών των γινομένων μαζί με την πόλωση (bias) μπαίνουν ως όρισμα σε μια συνάρτηση ενεργοποίησης, από την οποία προκύπτει η έξοδος του συγκεκριμένου νευρώνα.*

Κάτι που δεν αναφέρθηκε πριν αλλά μπορεί να δει κάποιος στο παραπάνω σχήμα είναι η λεγόμενη μονάδα πόλωσης ή νευρώνας πόλωσης (bias unit - bias neuron). Η πόλωση, με πολύ απλά λόγια, είναι η τιμή που θα δεχθεί σαν όρισμα η συνάρτηση ενεργοποίησης αν όλες οι άλλες εισοδοί της είναι μηδέν. Στις περισσότερες δομές νευρωνικών δικτύων, η πόλωση αναπαρίσταται σαν ένας ξεχωριστός νευρώνας ο οποίος έχει μια σταθερή τιμή (συνήθως το 1), και συνδέεται με όλους τους νευρώνες του δικτύου, εκτός φυσικά από τους νευρώνες του επιπέδου εισόδου. Επομένως, δεν έχει ούτε εισόδους από άλλους νευρώνες, ούτε συνάρτηση ενεργοποίησης, αλλά η σύνδεσή του διέπεται κανονικά από ένα βάρους για τον εκάστοτε νευρώνα το οποίο επίσης μεταβάλλεται κανονικά από τον κανόνα μάθησης. Η πόλωση επιτρέπει σε ένα νευρώνα να είναι πιο ευέλικτος ως προς το εύρος των εξόδων που μπορεί να δώσει. Μπορούμε να φανταστούμε την χρησιμότητα της μονάδας πόλωσης σαν την χρησιμότητα της σταθεράς σε μια πρωτοβάθμια εξίσωση: ο συντελεστής του αγνώστου επηρεάζει την κλίση της ευθείας που αναπαριστά η εξίσωση, και η σταθερά καθορίζει την θέση της ευθείας στον χώρο.

## Συναρτήσεις ενεργοποίησης

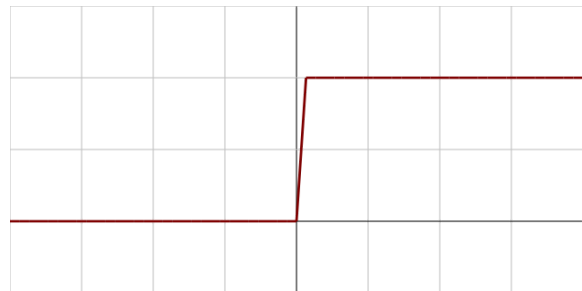
Συνήθως όλοι οι νευρώνες ενός νευρωνικού δικτύου, εκτός πάντα από τους νευρώνες εισόδου, έχουν την ίδια συνάρτηση ενεργοποίησης, αλλά και πάλι αυτό δεν είναι υποχρεωτικό. Μια γενική παρατήρηση είναι ότι μια συνάρτηση ενεργοποίησης πρέπει να είναι μη γραμμική, ώστε το δίκτυο να μπορεί να μοντελοποιεί μη γραμμικά φαινόμενα (Βλαχάβας κ.α., 2011). Επίσης μια γραμμική συνάρτηση καθιστά περιττή τη χρήση

παραπάνω από ενός κρυφού επιπέδου (Cybenko, 2006). Τέλος, είναι επιθυμητό μια συνάρτηση ενεργοποίησης να είναι παραγωγίσιμη σε όλο το πεδίο ορισμού της, καθώς αυτό επιτρέπει την χρήση κανόνων μάθησης που αξιοποιούν μεθόδους όπως την *μέθοδο απότομης καθόδου* (gradient descent) για να καταλήξουν στα βέλτιστα βάρη (Snyman, 2005).

Μερικές από τις πιο διαδεδομένες συναρτήσεις ενεργοποίησης, είναι οι παρακάτω:

### Βηματική Συνάρτηση (binary step function)

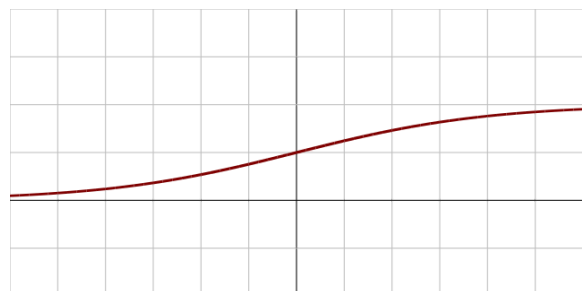
$$f(x) = \begin{cases} 1 & \text{for } x \geq \theta \\ 0 & \text{for } x < \theta \end{cases}$$



Η βηματική συνάρτηση είναι μια από τις απλούστερες μορφές συνάρτησης ενεργοποίησης, και χρησιμοποιείται κυρίως σε απλά νευρωνικά δίκτυα χωρίς κρυφό επίπεδο νευρώνων. Η έξοδος του νευρώνα είναι 1 αν το άθροισμα των εισόδων του είναι μεγαλύτερο ή ίσο από ένα κατώφλι  $\theta$ , ενώ σε αντίθετη περίπτωση είναι 0. Παρόμοια είναι και η *συνάρτηση προσήμου* (bipolar step function), η οποία παίρνει την τιμή -1 αν  $x < \theta$ .

### Λογιστική Σιγμοειδής Συνάρτηση (logistic sigmoid function)

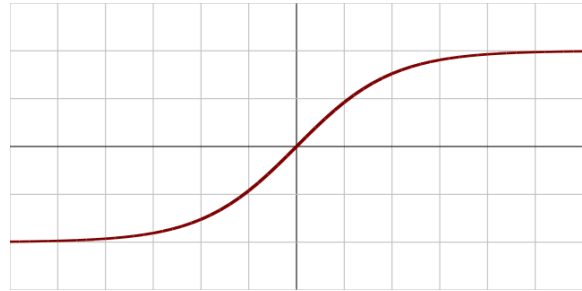
$$f(x) = \frac{1}{1 + e^{-\sigma x}}$$



Η λογιστική συνάρτηση χρησιμοποιείται σε Τ.Ν.Δ. πολλών επιπέδων όταν οι επιθυμητές τιμές εξόδου βρίσκονται στο διάστημα (0,1). Μεγάλο της θετικό προσόν είναι ότι παραγωγίζεται πολύ εύκολα:  $f'(x) = f(x)(1-f(x))$ , κάτι που μειώνει την υπολογιστική πολυπλοκότητα του αλγορίθμου, επιταχύνοντάς τον.

## Συνάρτηση Υπερβολικής Εφαπτομένης (Hyperbolic tangent function)

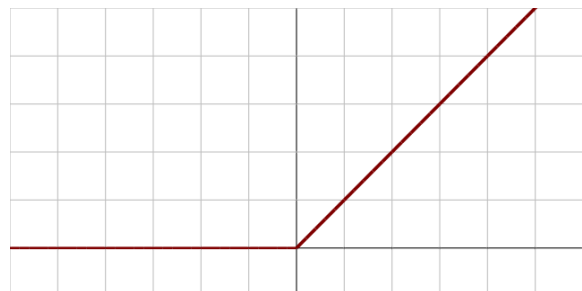
$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$$



Η συνάρτηση υπερβολικής εφαπτομένης επίσης χρησιμοποιείται σε Τ.Ν.Δ. πολλών επιπέδων, όταν οι επιθυμητές τιμές εξόδου βρίσκονται στο διάστημα  $(-1,1)$ . Έχει πολλά κοινά με την λογιστική συνάρτηση, όπως το θετικό προσόν της εύκολης παραγωγίσιμης:  $f'(x) = 1 - f(x)^2$

## Συνάρτηση Γραμμικού Ανορθωτή (Rectified linear unit - ReLU)

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



Η συνάρτηση γραμμικού ανορθωτή, γνωστή και ως συνάρτηση ράμπας (ramp function), από το χαρακτηριστικό της σχήμα, είναι από τις σχετικά νεότερες συναρτήσεις ενεργοποίησης ως προς την χρήση τους σε τεχνητά νευρωνικά δίκτυα, καθώς αυτή προτάθηκε για πρώτη φορά τον Hahnloser το 2000, ο οποίος δικαιολόγησε την χρήση της ως πιο συμβατή με τα περισσότερα βιολογικά φαινόμενα από τις σιγμοειδείς συναρτήσεις (Hahnloser et al., 2000). Παρ' ολ' αυτά, ήδη έχει εξελιχθεί στην πιο δημοφιλή συνάρτηση ενεργοποίησης, ιδιαίτερα για Τ.Ν.Δ. πολλών κρυφών επιπέδων που ανήκουν στον τομέα του *Deep Learning* (LeCun et al., 2015). Στα θετικά χαρακτηριστικά της συγκαταλέγεται επίσης η απλότητα των συνηθισμένων πράξεων κατά τη διάρκεια της λειτουργίας του δικτύου, ο μικρότερος αριθμός νευρώνων που ενεργοποιούνται κάθε φορά (καθώς πολλοί από αυτούς για αρκετή ώρα δίνουν μηδενική έξοδο, επιτρέποντας στον αλγόριθμο να εντοπίσει τους σημαντικότερους εξ

αυτών ταχύτερα), καθώς και το ότι δεν επηρεάζεται από την κλίμακα των δεδομένων όπως οι προηγούμενες συναρτήσεις ενεργοποίησης.

## Επιλογή επιπέδων

Η επιλογή του αριθμού των επιπέδων, καθώς και του αριθμού των κόμβων κάθε επιπέδου, είναι ένα πολύ περίπλοκο ζήτημα, καθώς η θεωρία δεν έχει μέχρι σήμερα δώσει ξεκάθαρη απάντηση για την καλύτερη μεθοδολογία επιλογής, και επίσης αυτές οι παράμετροι εξαρτώνται άμεσα από τη φύση του προβλήματος και τον αριθμό των σημαντικών μεταβλητών που πρέπει να ληφθούν υπόψη. Γενικά, ένα απλό T.N.Δ. τριών επιπέδων αποτελεί ήδη ένα πολύ ισχυρό εργαλείο για την λύση πολλών προβλημάτων που παρουσιάζονται στην πράξη, αλλά η αναζήτηση ολοένα ακριβέστερων και ισχυρότερων μεθόδων ποτέ δεν σταματά. Στην πράξη, για την επιλογή των επιπέδων και των νευρώνων ακολουθούνται συχνότερα οι παρακάτω κανόνες:

- **πειραματισμοί - δοκιμές (trial-and-error):** η απλούστερη μέθοδος, όπου δοκιμάζονται τυχαία διάφορες δομές και επιλέγεται αυτή που δίνει τα καλύτερα αποτελέσματα ανάμεσά τους
- **εμπειρικοί κανόνες:** σε διάφορες πηγές αναφέρονται εμπειρικοί κανόνες οι οποίοι συνήθως χρησιμοποιούνται ως βάση για μια μεθοδολογία trial-and-error, όπως για παράδειγμα ότι κάθε επόμενο επίπεδο πρέπει να έχει τους μισούς κόμβους από το προηγούμενο, ή περίπου όσο η ρίζα του αριθμού κόμβων του προηγούμενου, ή τη μέση τιμή ανάμεσα στον αριθμό των εισόδων και τον αριθμό των εξόδων, κ.ο.κ.
- **κλάδεμα (pruning):** Οι τεχνικές κλαδέματος ουσιαστικά είναι τεχνικές που ξεκινούν με έναν αυθαίρετα μεγάλο αριθμό επιπέδων και νευρώνων, και σταδιακά μειώνουν αυτόν τον αριθμό *κλαδεύοντας* (αφαιρώντας) τους νευρώνες και τα επίπεδα που φαίνεται να έχουν την μικρότερη επίδραση στην λειτουργία του T.N.Δ. Σημαντικοί αλγόριθμοι κλαδέματος με εφαρμογή στα T.N.Δ είναι ο Optimal Brain Damage (LeCun et al, 1990) και ο Optimal Brain Surgeon (Hassibi et al, 1993).

Μια αρκετά ενδιαφέρουσα εναλλακτική μέθοδος προτάθηκε πρόσφατα από τον Δ. Σταθάκη, ο οποίος προτείνει μια αναλυτική διαδικασία επιλογής της δομής ενός T.N.Δ με τη χρήση ενός γενετικού αλγόριθμου για την ανάδειξη της αποτελεσματικότερης (Stathakis, 2009).

## Κανόνας Μάθησης (Learning Rule)

Θα μπορούσε να πει κάποιος ότι τα βάρη ανάμεσα στους νευρώνες ενός νευρωνικού δικτύου αποτελούν την “ψυχή” του, και όχι μόνο επειδή οι τιμές τους καθορίζουν τις εξόδους των νευρώνων και όλου του δικτύου κατ’ επέκταση. Όταν θέλουμε να λύσουμε ένα πρόβλημα με ένα νευρωνικό δίκτυο, η απάντηση είναι το ίδιο το δίκτυο, ή με άλλα λόγια, η κατάλληλη δομή T.N.Δ. σε συνδυασμό με τα κατάλληλα βάρη ανάμεσα στους νευρώνες, έτσι ώστε κάθε φορά που το δίκτυο δέχεται μια είσοδο, να δίνει την επιθυμητή έξοδο.

Βασικό πλεονέκτημα των νευρωνικών δικτύων σε σχέση με άλλες μεθόδους όπως τα expert systems είναι ότι τα βάρη αυτά δεν χρειάζεται να τα προσδιορίσει ο ίδιος ο χρήστης, αλλά προσαρμόζονται αυτόματα στα δεδομένα βάσει ενός κανόνα μάθησης.

Ένας τρόπος να δούμε την διαδικασία της μάθησης στα πλαίσια των τεχνητών νευρωνικών δικτύων είναι ως μια διαδικασία βελτιστοποίησης. Πιο συγκεκριμένα, κατά τη διαδικασία της μάθησης αναζητούμε σε ένα πολυδιάστατο χώρο (αυτών των πιθανών τιμών των βαρών), για μια λύση (ή αλλιώς ένα διάλυμα του χώρου) η οποία σταδιακά βελτιστοποιεί μια συνάρτηση στόχο, την οποία αποκαλούμε συνήθως *συνάρτηση κριτηρίου*, και εκφράζει το επιθυμητό αποτέλεσμα που θέλουμε να έχει ο αλγόριθμος ώστε να προσφέρει λύση στον πρόβλημα. Σε αυτό το σημείο θα πρέπει να επανέρθουμε στους θεμελιώδεις τύπους μάθησης που αναλύθηκαν στο προηγούμενο κεφάλαιο.

Στην περίπτωση της μάθησης με επίβλεψη, ο στόχος του Τ.Ν.Δ. είναι να αντιστοιχίσει κάθε είσοδο του με μια επιθυμητή έξοδο. Για παράδειγμα, ένας αλγόριθμος αναγνώρισης χειρόγραφων αριθμητικών χαρακτήρων, θα πρέπει να αντιστοιχίζει κάθε εικόνα που δέχεται σαν είσοδο με τον αριθμό που αντιπροσωπεύει ο χειρόγραφος αριθμός της εικόνας. Σε αυτήν την περίπτωση, η συνάρτηση κριτηρίου είναι μια συνάρτηση που υπολογίζει το μέγεθος του σφάλματος, δηλαδή το πόσο απέχουν οι τιμές που δίνει το δίκτυο σαν έξοδο από τις πραγματικές λύσεις του προβλήματος, τις οποίες γνωρίζουμε για το εκπαιδευτικό σύνολο, καθώς μιλάμε για μάθηση με επίβλεψη. Η βελτιστοποίηση που επιθυμούμε είναι η ελαχιστοποίηση αυτής της συνάρτησης (η οποία ονομάζεται και *συνάρτηση κόστους - cost function*), έτσι ώστε οι τιμές των εξόδων του δικτύου να είναι όσο πιο κοντά γίνεται στις πραγματικές τιμές. Μια από τις πιο δημοφιλείς μεθόδους για την ελαχιστοποίηση της συνάρτησης κόστους είναι η μέθοδος *απότομης καθόδου* (gradient descent), η οποία μπορεί να αναλυθεί στα παρακάτω επιμέρους βήματα:

- 1) Δίνουμε τυχαίες τιμές στα βάρη και υπολογίζουμε τη συνάρτηση κόστους.
- 2) Υπολογίζουμε την μεταβολή της τιμής της συνάρτησης κόστους, όταν μεταβάλλουμε τα βάρη κατά μια πολύ μικρή προκαθορισμένη τιμή.
- 3) Αντικαθιστούμε τις τιμές των βαρών με τις νέες τιμές που μειώνουν την τιμή της συνάρτησης κόστους.
- 4) Επαναλαμβάνουμε τη διαδικασία μέχρι να φτάσουμε σε ένα σημείο όπου η τιμή της συνάρτησης κόστους δεν μπορεί να μειωθεί άλλο κατά αυτόν τον τρόπο.

Στην περίπτωση που έχουμε κάποιο πρόβλημα μάθησης χωρίς επίβλεψη, ο στόχος είναι συνήθως η οργάνωση των εκπαιδευτικών παραδειγμάτων του δείγματος σε ομάδες που έχουν κοινά χαρακτηριστικά. Εκεί, η συνάρτηση κριτηρίου έχει σαν στόχο την μικρότερη συνολική “απόσταση” όλων των μελών κάθε ομάδας από τα αντίστοιχα κέντρα της.

Τέλος, στην περίπτωση της ενισχυτικής μάθησης, η συνάρτηση που επηρεάζει την σύγκλιση των βαρών προς τις ιδανικές τιμές τους εξαρτάται από μια τρίτη συνάρτηση η οποία “εκτιμά”

πόσο καλά ανταποκρίνεται η λύση στους στόχους του προβλήματος, καθώς εδώ δεν έχουμε μια συγκεκριμένη “σωστή” λύση.

## Οπισθοδιάδοση (Back-Propagation)

Η οπισθοδιάδοση είναι μια επέκταση της μεθόδου *απότομης καθόδου* για νευρωνικά δίκτυα πολλών επιπέδων. Με κατεύθυνση από το επίπεδο εξόδου προς το επίπεδο εισόδου, για κάθε εσωτερικό νευρώνα υπολογίζεται η συμμετοχή του στα σφάλματα των νευρώνων εξόδου και γίνεται η αλλαγή των βαρών στην είσοδό του με την μέθοδο της *απότομης καθόδου*. Η συμμετοχή αυτή του κάθε νευρώνα στα σφάλματα των νευρώνων του επόμενου επιπέδου του είναι ανάλογη της τρέχουσας εισόδου του και των συντελεστών βαρύτητας που τον συνδέουν με τους νευρώνες του επόμενου επιπέδου. Κατ’ αυτό τον τρόπο, εντοπίζονται οι νευρώνες που “ευθύνονται” πρωτίστως για μια μη επιθυμητή έξοδο και μεταβάλλονται οι τιμές των βαρών τους ανάλογα, μειώνοντας έτσι την επίδραση της μάθησης στα βάρη των νευρώνων που συμπεριφέρονται “σωστά” και έτσι διατηρώντας την ικανότητα του T.N.Δ. να περιγράφει σχέσεις μεταξύ μεταβλητών σε δεύτερο ή μεγαλύτερο επίπεδο διασύνδεσης.

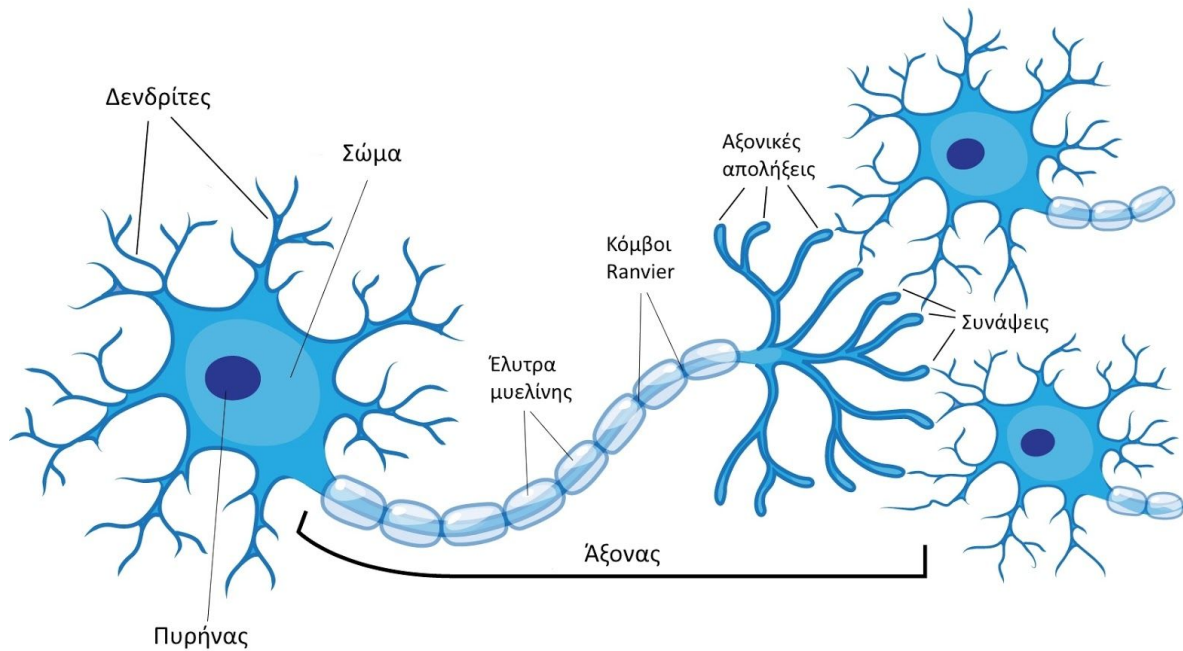
## Οι βιολογικοί νευρώνες

Εκτιμάται ότι ο μέσος εγκέφαλος ενός ενήλικου ανθρώπου περιέχει γύρω στα 86 δισεκατομμύρια νευρώνες (Azevedo et al., 2009). Οι νευρώνες αυτοί, μαζί με τους υπόλοιπους νευρώνες του ανθρώπινου κεντρικού νευρικού συστήματος, δημιουργούν ένα τεράστιο δίκτυο μεταξύ τους, μέσω του οποίου διαδίδονται ηλεκτροχημικά σήματα.

Τα δομικά στοιχεία που επιτρέπουν σε ένα βιολογικό νευρωνικό κύτταρο να επικοινωνεί με τους άλλους νευρώνες είναι οι το *σώμα* (που περικλείει τον *πυρήνα* του κυττάρου) ο *άξονας* (η *νευράξονας*) με τις *αξονικές απολήξεις* και οι *δενδρίτες*, όπως φαίνεται και στο παρακάτω σχήμα. Οι άκρες των δενδριτών αποκαλούνται *τερματικά κουμπιά*.

Οι δενδρίτες είναι λεπτές, διακλαδισμένες επεκτάσεις του κυτταρικού σώματος, των οποίων η κύρια λειτουργία είναι να συλλέγουν σήματα από το εξωτερικό του κυττάρου. Για να γίνει αυτό, οι επιφάνειες των δενδριτών περιέχουν υποδοχείς για χημικές ουσίες που αποκαλούνται *νευροδιαβιβαστές* και λειτουργούν ως “αγγελιοφόροι” ανάμεσα στους νευρώνες. Όταν ένας νευροδιαβιβαστής δεσμεύεται στον υποδοχέα ενός δενδρίτη, παράγεται μια μικρή ποσότητα ηλεκτρικής δραστηριότητας, η οποία στη συνέχεια μετακινείται μέσω του δενδρίτη στο σώμα του κυττάρου (Buhl et al., 2012).





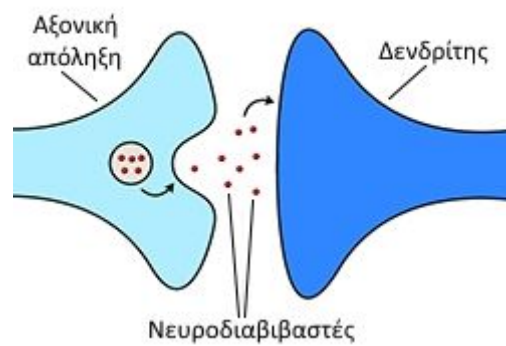
*Βασικά στοιχεία ενός βιολογικού νευρωνικού δικτύου.*

*πηγή εικόνας: <https://askabiologist.asu.edu/plosable/speed-human-brain>*

*(επεξεργασμένη και μεταφρασμένη)*

Κάθε νευρώνας έχει συνήθως πολλούς δενδρίτες. Είναι σημαντικό ότι ο αριθμός των δενδριτών που διεγείρονται μπορεί να διαφέρει από τη μία στιγμή στην άλλη. Έτσι, μπορεί να υπάρξει περισσότερη ή λιγότερη ηλεκτρική δραστηριότητα που ταξιδεύει προς το σώμα οποιαδήποτε στιγμή. Εάν η ηλεκτρική δραστηριότητα είναι πολύ μικρή, δεν συμβαίνει τίποτα. Εάν, ωστόσο, η ηλεκτρική δραστηριότητα ξεπεράσει κάποια κρίσιμη ποσότητα ή “κατώφλι”, τότε λέμε ότι “ενεργοποιείται” ο άξονας. Ο άξονας είναι η λεπτή προέκταση του νευρώνα που αποτελεί το μεγαλύτερο μέρος του μήκους του. Όταν ο άξονας ενεργοποιείται, ένα ηλεκτρικό σήμα ταξιδεύει μέσα του μέχρι να φτάσει στο άκρο του. Αυτή η ενεργοποίηση του άξονα αναφέρεται και ως *πυροδότηση* του νευρώνα. Στην άκρη του άξονα υπάρχει μια μικρή διόγκωση που ονομάζεται αξονική απόληξη. Από αυτό το σημείο, ο νευρώνας θα μεταδώσει το σήμα σε κάποιο άλλο κύτταρο, όπως έναν άλλο νευρώνα ή ένα μυϊκό κύτταρο.

Πολλοί επιστήμονες αρχικά πίστευαν ότι οι νευρώνες συνδέονται άμεσα μεταξύ τους, αλλά αργότερα αποδείχτηκε ότι ένας νευρώνας δεν αγγίζει ποτέ πραγματικά έναν άλλο. Αντίθετα, οι αξονικές απολήξεις του ενός διαχωρίζονται από τα τερματικά κουμπιά του άλλου από ένα πολύ μικρό κενό μεταξύ τους που ονομάζεται *συναπτική σχισμή* ή *συναπτικό χάσμα*. Οι νευροδιαβιβαστές που εκκρίνονται από την αξονική απόληξη διασχίζουν τη συναπτική σχισμή για να φτάσουν στους υποδοχείς που βρίσκονται στις άκρες των δενδριτών, στα τερματικά κουμπιά. (Feinberg, 2012).



Σχηματική απεικόνιση μιας σύναψης μεταξύ δύο νευρώνων.

Οι νευρώνες δεν είναι συνδεδεμένοι φυσικά, αλλά η πληροφορία μεταδίδεται με την έκκριση χημικών ουσιών που αποκαλούνται νευροδιαβιβαστές.

πηγή εικόνας:: <https://askabiologist.asu.edu/plosable/speed-human-brain>

(επεξεργασμένη και μεταφρασμένη)

Κάτι τελευταίο που πρέπει να αναφερθεί για τους βιολογικούς νευρώνες, καθώς σχετίζεται άμεσα με το θέμα αυτής της εργασίας, είναι ο τρόπος με τον οποίο αυτοί προσαρμόζονται σταδιακά στα ερεθίσματα που δέχονται ώστε να οδηγήσουν σε αυτό που αποκαλούμε μάθηση. Ο Donald Hebb, στο πρωτοπόρο βιβλίο του *Organization of Behavior* (1949), διατύπωσε την άποψη ότι “όταν ένα νευρωνικό κύτταρο ενεργοποιεί συχνά και επίμονα ένα άλλο, τότε λαμβάνει χώρα μια διαδικασία ανάπτυξης ή μια μεταβολική αλλαγή, με στόχο την αποτελεσματικότερη λειτουργία του πρώτου ως ενεργοποιητής του δεύτερου”. Η άποψη αυτή, που ονομάζεται κανόνας του Hebb, και που ουσιαστικά προσπαθεί να ερμηνεύσει γιατί, όπως αναφέρθηκε νωρίτερα, “μάθηση είναι η ικανότητα βελτίωσης του τρόπου εκτέλεσης μιας ενέργειας, με την επανάληψή της”, επαληθεύτηκε με την ανακάλυψη μιας βιολογικής διαδικασίας που ονομάζεται *spike-timing-dependent plasticity*, η οποία προτάθηκε από τον M. M. Taylor (1973) και αποδείχτηκε πειραματικά, μεταξύ άλλων, από τον Henry Markram (Markram et al, 1997). Με πολύ απλά λόγια, αν η ενεργοποίηση ενός νευρώνα τείνει να συμβαίνει αμέσως πριν την ενεργοποίηση ενός άλλου, τότε η ισχύς αυτού του σήματος ενισχύεται, ενώ αλλά σήματα που λαμβάνονται μετά την ενεργοποίηση του νευρώνα αποδυναμώνονται, έτσι ώστε τα σήματα που αποτελούν την αιτία της ενεργοποίησης του νευρώνα να είναι πιο πιθανό να επαναληφθούν στο μέλλον, ενώ όλα τα υπόλοιπα να έχουν μικρότερη πιθανότητα να επαναληφθούν, καταλήγοντας έτσι αναπόφευκτα σε μια δομή ανάμεσα στους νευρώνες που προάγει και επιταχύνει σχέσεις αιτίου - αποτελέσματος.

Σε αυτό το σημείο, μπορούμε σταματήσουμε και να συνοψίσουμε τις βασικές ιδιότητες των βιολογικών νευρωνικών δικτύων που προσομοιάζουν τα τεχνητά νευρωνικά δίκτυα:

- Η επεξεργασία των πληροφοριών γίνεται σε στοιχειώδεις μονάδες επεξεργασίας. Οι νευρώνες - κόμβοι του Τ.Ν.Δ. αντιστοιχούν στα νευρωνικά κύτταρα .

- Κάθε νευρώνας συνδέεται παράλληλα με πολλούς άλλους νευρώνες. Οι κόμβοι Τ.Ν.Δ. δέχονται τιμές από όλους τους νευρώνες στο προηγούμενο επίπεδο και στέλνουν πληροφορία σε όλους τους νευρώνες του επόμενου, όπως ένα νευρωνικό κύτταρο έχει πολλές αξονικές απολήξεις που στέλνουν νευροδιαβιβαστές και πολλούς δενδρίτες που δέχονται νευροδιαβιβαστές από άλλα κύτταρα.
- Κάθε νευρώνας εκπέμπει πληροφορίες στους επόμενούς του όταν ξεπεραστεί ένα συγκεκριμένο *κατώφλι ενεργοποίησης*: είτε η ποσότητα ηλεκτροχημικής ενέργειας στους βιολογικούς νευρώνες, είτε η τιμή της συνάρτησης ενεργοποίησης στα Τ.Ν.Δ.
- Το δίκτυο *εκπαιδύεται* με την επανάληψη κάποιας διαδικασίας που επιδρά πάνω στην ισχύ της σύνδεσης μεταξύ των κόμβων του: η βιολογική διαδικασία *spike-timing-dependent plasticity* αντιστοιχεί στον κανόνα μάθησης ενός νευρωνικού δικτύου που επιδρά πάνω στις βαρύτητες των συνάψεων μεταξύ των κόμβων του δικτύου.

Στα πλαίσια της *γνωσιακής επιστήμης* (cognitive science), και πιο συγκεκριμένα της θεωρίας του *συνδεδετισμού* (connectionism), αυτές οι ιδιότητες είναι αρκετές για να χαρακτηρίσουν αμφότερα τα βιολογικά και τα τεχνητά νευρωνικά δίκτυα ως *μοντέλα παράλληλης και κατακευματισμένης επεξεργασίας - Parallel Distributed Processing Models* (Medler, 1998). Υπάρχουν βέβαια και άλλες ομοιότητες των νευρωνικών δικτύων με τα βιολογικά, όπως, για παράδειγμα, είναι η *ανοχή στα σφάλματα* (fault tolerance): σύνθετοι οργανισμοί όπως οι άνθρωποι και τα ζώα γεννιούνται με δεκάδες δισεκατομμύρια νευρώνες, και οι περισσότεροι από αυτούς δεν αναπληρώνονται από το σώμα κατά τη διάρκεια ζωής του οργανισμού (Johnson and Brown, 1988). Αυτό σημαίνει ότι οι οργανισμοί *μαθαίνουν* σε όλη τη διάρκεια της ζωής τους, παρά το γεγονός τα νευρωνικά τους κύτταρα όλο και λιγοστεύουν. Ακόμα και σε περιπτώσεις όπου ένας τραυματισμός οδηγεί σε απώλεια μεγάλου αριθμού νευρώνων, έχει παρατηρηθεί ότι άλλοι νευρώνες μπορεί να αναλάβουν την λειτουργία που εκτελούσαν οι χαμένοι νευρώνες. Ενδεικτικό είναι ένα πρωτοποριακό πείραμα ερευνητών του MIT, κατά το οποίο συνδέσαν τα μάτια μικρών ζώων με το ακουστικό κέντρο του εγκεφάλου τους και διαπίστωσαν ότι οι νευρώνες του ακουστικού κέντρου αντιδρούσαν σε οπτικά ερεθίσματα (Sharma et al., 2000). Παρομοίως, ένα τεχνητό νευρωνικό μπορεί να σχεδιαστεί έτσι ώστε να είναι ευέλικτο σε περίπτωση σφαλμάτων στις συνδέσεις ή στα δεδομένα του, και επίσης μπορεί να επανεκπαιδευτεί για διαφορετικά δεδομένα (Fausett, 1994).

Γενικά, η προσπάθεια για απόλυτη προσομοίωση των βιολογικών νευρωνικών δικτύων δεν αποτελεί αυτοσκοπό ή υψηλή προτεραιότητα για τους δημιουργούς των τεχνητών νευρωνικών δικτύων. Στόχος είναι, βέβαια, η επιλογή του κατάλληλου τύπου δικτύου για το εκάστοτε πρόβλημα. Παρ' ολ' αυτά, υπάρχουν περιπτώσεις όπου έχει αποδειχτεί εκ των υστέρων ότι μια ιδανική δομή Τ.Ν.Δ. για κάποιο πρόβλημα προσομοιάζει παρόμοιες φυσικές δομές ανθρώπινων νευρώνων (Rogers and Kabrisky, 1991).

## Ιστορική αναδρομή

Το 1943, ο νευρολόγος Warren McCulloch και ο μαθηματικός Walter Pitts σχεδίασαν το πρώτο νευρωνικό δίκτυο, εμπνευσμένοι από το ανθρώπινο νευρικό σύστημα και βασισμένοι στην ιδέα ότι ένα δίκτυο υπολογιστικών μονάδων που στέλνουν σήμα στην επόμενη κατά σειρά μονάδα αν η τιμή τους είναι μεγαλύτερη από ένα συγκεκριμένο κατώφλι έχει τη δυνατότητα να υπολογίσει με μεγάλη ταχύτητα πολλές τυπικές συναρτήσεις (McCulloch and Pitts, 1943). Σημαντική ώθηση σε αυτές τις πρώιμες προσπάθειες έδωσε το βιβλίο *The Organization of Behaviour* του Donald Hebb, στο οποίο προτάθηκε ο “κανόνας του Hebb” που αναφέρθηκε νωρίτερα. Το δίκτυο των McCulloch και Pitts είχε στατικά, προκαθορισμένα βάρη και δεν εμπειριείχε την έννοια της *μάθησης*, αλλά ο Rosenblatt (1958), εμπνευσμένος από τη δουλειά του Hebb δημιούργησε τα *perceptrons*, τα οποία έμοιαζαν με τα δίκτυα νευρώνων των McCulloch και Pitts αλλά ακολουθούσαν ένα κανόνα μάθησης που βελτιώνει την ιδέα του Hebb. Παράλληλα, με την ανάπτυξη των ηλεκτρονικών υπολογιστών κατά τη δεκαετία του 1950, έγινε δυνατή η εφαρμογή αυτών των ιδεών και η δημιουργία των πρώτων τέτοιων μοντέλων που έλυσαν πρακτικά προβλήματα. Συγκεκριμένα, ο Bernard Widrow και ο Marcian Hoff δημιούργησαν το μοντέλο ADALINE (ADaptive LINear Element) το οποίο μετεξελιχθηκε στο MADALINE (Multiple ADaptive LINear Elements). Τα μοντέλα αυτά διέφεραν από τα perceptrons του Rosenblatt, καθώς δεν είχαν συνάρτηση ενεργοποίησης αλλά απλώς προσέθεταν τα βάρη μεταξύ τους (Widrow and Hoff, 1960), και χρησιμοποιήθηκαν σαν βάση μιας εφαρμογής που ξεχώριζε την ομιλία από την ηχώ σε μια τηλεφωνική γραμμή.

Παρά αυτές τις πρώιμες επιτυχίες, η ανάπτυξη των νευρωνικών δικτύων ως τεχνικών μηχανικής μάθησης επισκιάστηκε από την ακόμα μεγαλύτερη ανάπτυξη των παραδοσιακών μοντέλων ηλεκτρονικών υπολογιστών. Επίσης, αυτές οι πρώιμες επιτυχίες ώθησαν τους επιστήμονες σε μεγαλόστομες υποσχέσεις για το μέλλον της τεχνητής νοημοσύνης και των Τ.Ν.Δ. οι οποίες ήταν απλά ανέφικτες με την τεχνολογία της εποχής. Κρατώντας επιφυλακτική στάση απέναντι σε όλα αυτά, οι Marvin Minsky και Seymour Papert το 1969 δημοσίευσαν ένα βιβλίο το οποίο καταδείκνυε τις αδυναμίες των perceptrons, με κυριότερη την αδυναμία τους να προσαρμοστούν σε συναρτήσεις που δεν είναι γραμμικά διαχωρίσιμες. Όλοι οι παραπάνω παράγοντες οδήγησαν σε μια μεγάλη παύση ενδιαφέροντος και χρηματοδότησης για μελέτες σχετιζόμενες με την τεχνητή νοημοσύνη την δεκαετία του 1970.

Την δεκαετία του 1980, επιστήμονες όπως ο Werbos, ο Parker και ο LeCun παρουσίασαν την ιδέα της οπισθοδιάδοσης (back-propagation), η οποία επέτρεψε την δημιουργία των πρώτων πολυεπίπεδων νευρωνικών δικτύων, λύνοντας έτσι το πρόβλημα της γραμμικής διαχωρισιμότητας. Η εφαρμογή ενός δικτύου οπισθοδιάδοσης για την μηχανική αναγνώριση χειρόγραφων ταχυδρομικών κωδικών (LeCun et al, 1989) έδωσε νέα ώθηση στην έρευνα πάνω στα νευρωνικά δίκτυα. Παρ’ όλ αυτά, τα πολυεπίπεδα νευρωνικά δίκτυα οπισθοδιάδοσης είχαν και αυτά ένα πρόβλημα που τα εμπόδιζε από το να επιλύσουν τα

ολοένα και πιο πολύπλοκα προβλήματα που αντιμετώπιζαν, το λεγόμενο *vanishing gradient*: όσο περισσότερα κρυφά επίπεδα είχε ένα δίκτυο, τόσο μικρότερη ήταν η “συμμετοχή στο σφάλμα” που αντιστοιχούσε στους νευρώνες των πρώτων επιπέδων, στερώντας το νόημα από τη χρήση πολλών επιπέδων.

Η διαπίστωση αυτή οδήγησε σε μια νέα περίοδο ισχρών αγελάδων για τους ερευνητές των νευρωνικών δικτύων, η οποία κράτησε μέχρι το 2006, όταν μια ομάδα ερευνητών με επικεφαλής τον Geoffrey Hinton απέδειξε ότι αν τα αρχικά βάρη ενός πολυεπίπεδου T.N.Δ. δεν είναι τυχαία, αλλά έχουν προκαθοριστεί “εκπαιδεύοντας” κάθε κρυφό επίπεδο ξεχωριστά με μεθόδους μάθησης χωρίς επίβλεψη, τότε αποφεύγεται το πρόβλημα του *vanishing gradient* και το T.N.Δ. μπορεί να επιλύσει πολύπλοκα προβλήματα με πολύ μεγαλύτερη ταχύτητα και ακρίβεια από ένα T.N.Δ. μικρού αριθμού επιπέδων (Hinton et al, 2006). Η δημοσίευση αυτή, σε συνδυασμό με την κατακόρυφη αύξηση της υπολογιστικής ισχύος που προσέφεραν οι υπερυπολογιστές και το *cloud computing* οδήγησαν στην τωρινή “έκρηξη” την δημοφιλίας και των εφαρμογών πολυεπίπεδων τεχνητών νευρωνικών δικτύων. Στην εποχή μας, έχει επικρατήσει πλέον το (σαφώς πιο εμπορικό) όνομα *Deep Learning* για να περιγράψει τους αλγορίθμους αυτού του τύπου, που συνδυάζουν πολλά επίπεδα με εναλλαγές στους κανόνες μάθησης για την βελτιστοποίηση της δομής τους.

## Ισχυρά σημεία και αδυναμίες των Νευρωνικών Δικτύων

Στα ισχυρά σημεία των νευρωνικών δικτύων συγκαταλέγονται:

- η δυνατότητά τους να εντοπίζουν συσχετίσεις που δεν είναι ευνόητες και μοτίβα κανονικότητας που δεν είναι προφανή
- η δυνατότητά τους να διαχειρίζονται τεράστιους όγκους δεδομένων με μεγάλο αριθμό χαρακτηριστικών
- η προσαρμοστικότητά τους στα ίδια τα δεδομένα, που απαλλάσσει τον ερευνητή από την υποχρέωση να θεσπίσει ο ίδιος ένα σύνολο κανόνων

Ως αδυναμίες τους, μπορούμε να εντοπίσουμε τις εξής:

- τα νευρωνικά δίκτυα προσπαθούν να *προσεγγίσουν* ένα μεγάλο βαθμό ακρίβειας για προβλήματα με πολλά δεδομένα και πολλές μεταβλητές, κάτι που τα καθιστά ιδανικά για προβλήματα στα οποία ένας μικρός βαθμός λάθους είναι ανεκτός, αλλά ίσως όχι ιδανικά για προβλήματα στα οποία απαιτείται απόλυτη ακρίβεια
- τα νευρωνικά δίκτυα αποτελούν μια μεθοδολογία “μαύρου κουτιού” (black box). Αυτό σημαίνει ότι ο χρήστης τους μπορεί να δώσει κάποια δεδομένα εισόδου και να λάβει τα επιθυμητά αποτελέσματα, αλλά δύσκολα θα είναι σε θέση να εξηγήσει τον λόγο για τον οποίο το σύστημα επέλεξε αυτά τα αποτελέσματα. Υπό άλλη οπτική γωνία, μπορούμε να πούμε ότι ένα νευρωνικό δίκτυο δεν μας παράγει μια ξεκάθαρη συνάρτηση που συσχετίζει τις εισόδους με τις εξόδους του, αλλά το ίδιο το δίκτυο είναι η λύση του προβλήματος της συσχέτισής τους.

### 3.ΕΦΑΡΜΟΓΕΣ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Οι μεγάλες δυνατότητες των αλγορίθμων νευρωνικών δικτύων, ιδιαίτερα σε προβλήματα ταξινόμησης/κατηγοριοποίησης, δίνουν λύσεις σε ένα τεράστιο εύρος προβλημάτων και έχουν πρακτικές εφαρμογές σχεδόν σε όλους τους τομείς της σύγχρονης έρευνας και ανάπτυξης τεχνολογίας. Σε αυτό το κεφάλαιο παρουσιάζονται συνοπτικά μερικές ενδεικτικές εφαρμογές, με έμφαση σε αυτές που άπτονται του οικονομικού και επιχειρηματικού τομέα.

#### Εφαρμογές σε διάφορους τομείς

Οι εφαρμογές που θα παρουσιαστούν σε αυτήν την ενότητα δεν σχετίζονται άμεσα με το θέμα της παρούσας εργασίας, αλλά είναι σαφές ότι το διεπιστημονικό εύρος και η αποτελεσματικότητα κάποιων από αυτές προσέφεραν μεγάλη αναγνωρισιμότητα και κύρος στις μεθοδολογίες Τ.Ν.Δ., επιτρέποντας έτσι την σταδιακή αποδοχή τους από τις επιχειρήσεις και την εξάπλωσή τους στον οικονομικό τομέα.

#### Γενικές εφαρμογές

Μια από τις πρώτες εφαρμογές των Τ.Ν.Δ. που ακόμα παραμένει σημαντική και χρήσιμη σε διάφορους κλάδους είναι η επεξεργασία σήματος. Αναφερθήκαμε ήδη στην ιστορική εφαρμογή των Widrow και Hoff (1960), η οποία αφαιρούσε την ηχώ από μια τηλεφωνική συνομιλία. Η αφαίρεση ανεπιθύμητου θορύβου από ηχητικά σήματα παραμένει μέχρι και σήμερα σημαντική, με διάφορες προσεγγίσεις των αλγορίθμων Τ.Ν.Δ. να δοκιμάζονται όπως δείχνουν οι εργασίες των Badri (2009) και Al-Allaf (2015). Στα πλαίσια της επεξεργασίας σήματος εντάσσεται και η πολύ διαδεδομένη εφαρμογή της αναγνώρισης φωνής, με στόχο είτε την υπαγόρευση κειμένου αντί δακτυλογράφησης, είτε την χρήση συσκευών με προφορικές εντολές. Μια προσέγγιση Τ.Ν.Δ. στο συγκεκριμένο πρόβλημα κάνουν οι Dede και Sazli (2010), με πολύ ικανοποιητικά αποτελέσματα.

Η αμέσως επόμενη κατά σειρά ιστορικότητας εφαρμογή είναι η αναγνώριση χειρόγραφων χαρακτήρων. Αναφέραμε φυσικά την πολύ σημαντική πρώτη εφαρμογή του LeCun (1989) με τους χειρόγραφους ταχυδρομικούς κώδικες. Εκτός από μια εφαρμογή που ακόμα έχει μεγάλη χρησιμότητα (όπως στα συστήματα αναγνώρισης χαρακτήρων OCR), έχει κατά κάποιο τρόπο εξελιχθεί σε μέτρο σύγκρισης μεταξύ διαφορετικών μεθοδολογιών Τ.Ν.Δ. Ο Yann LeCun διατηρεί μια κλασσική πλέον βάση δεδομένων με 70000 χειρόγραφους χαρακτήρες η οποία λέγεται MNIST (<http://yann.lecun.com/exdb/mnist/>) και αποτελεί σημείο αναφοράς για κάθε τύπο νευρωνικών δικτύων. Μια σύγχρονη εφαρμογή deep learning, η οποία εκμεταλλεύεται την ταχύτητα και τις δυνατότητες παράλληλης επεξεργασίας των σύγχρονων καρτών γραφικών (Ciresan et al, 2012) μπορεί πλέον να έχει ακρίβεια 99,77% στην αναγνώριση των χαρακτήρων της βάσης MNIST! Βέβαια, εκτός από την αναγνώριση χειρόγραφων χαρακτήρων, υπάρχουν και άλλες εφαρμογές αναγνώρισης εικόνας, όπως η αναγνώριση προσώπων, η οποία μεταξύ άλλων χρησιμοποιείται σε εφαρμογές βιομετρικής ασφάλειας, η σε εγκληματολογικές αναλύσεις βίντεο (Kasar et al, 2016). Ένα στάδιο μετά

την αναγνώριση εικόνας είναι η αναγνώριση δραστηριοτήτων σε βίντεο. Αυτές οι εφαρμογές αποτελούν την αιχμή της τεχνολογίας Deep Learning και δείχνουν το δρόμο για το τι θα ακολουθήσει. Εντυπωσιακό παράδειγμα είναι η πολύ πρόσφατη εφαρμογή διαβάσματος χειλιών από βίντεο *LipNet*, που χρηματοδοτήθηκε από το παράρτημα τεχνητής νοημοσύνης της Google, το *Google Deepmind*. Το LipNet είναι η πρώτη εφαρμογή αυτού του είδους που επιτυγχάνει μεγαλύτερη απόδοση και ακρίβεια από έναν άνθρωπο εκπαιδευμένο στο διάβασμα των χειλιών, καθώς μπορεί να διαβάσει τα λόγια ομιλητών ακόμα κι αν μιλούν ο ένας πάνω στον άλλο, σε επίπεδο πρότασης και όχι λέξη προς λέξη, και με ακρίβεια 95,2%! Η εφαρμογή αυτή εκτός των άλλων αποτελεί και έναν οιώνό για αυτό που εκτιμάται ότι θα ακολουθήσει σαν τάση στις εφαρμογές νευρωνικών δικτύων: από την αναγνώριση και εξαγωγή χαρακτηριστικών από στατικές φωτογραφίες, περνάμε στην αναγνώριση και εξαγωγή *δραστηριοτήτων* από βίντεο.

Υπάρχουν πάρα πολλές άλλες εφαρμογές T.N.Δ. γενικής χρήσης. Για να μην επεκταθούμε, θα αναφέρουμε δύο ακόμα παραδείγματα που αποτελούν την τελευταία λέξη της τεχνολογίας νευρωνικών δικτύων. Η Google Deepmind προετοιμάζει ήδη τον διάδοχο της γνωστής υπηρεσίας μετάφρασής της, Google Translate, με τη χρήση ενός αλγορίθμου νευρωνικών δικτύων που επιτυγχάνει έως και 60% καλύτερες μεταφράσεις από την προηγούμενη εφαρμογή (Wu et al., 2016). Επίσης, τα εργαστήρια της NVIDIA βελτιώνουν συνεχώς τους ήδη εφαρμοζόμενους αλγορίθμους αυτόματης οδήγησης, που βρίσκονται σε οχήματα όπως τα ηλεκτροκίνητα Tesla, με τεχνικές που επιτρέπουν την καλύτερη αναγνώριση μικρών δρόμων χωρίς σήμανση (Bojarski et al., 2016).

### Ιατρικές εφαρμογές

Πάρα πολλές είναι οι εφαρμογές και στον ιατρικό τομέα. Κυριότερες εξ αυτών, και χαρακτηριστικά παραδείγματα μηχανικής μάθησης γενικότερα, είναι οι εφαρμογές που αναλύουν εικόνες ή ενδείξεις ιατρικών οργάνων με στόχο τον εντοπισμό ανωμαλιών όπως κακοηθείς όγκους. Πρόσφατα παραδείγματα είναι μια νέα μέθοδος ανάλυσης μαστογραφιών για εντοπισμό όγκων που υποδεικνύουν καρκίνο του μαστού, (Addiou et al., 2016), και μια μέθοδος εντοπισμού πιθανών μελανωμάτων στο δέρμα με ανάλυση απλών φωτογραφιών που μπορεί να βγάλει οποιοσδήποτε με το κινητό του (αντί για εικόνες δερμοσκοπίου), επιτρέποντας έτσι την εύκολη και έγκαιρη πρόληψη (Jafari et al., 2017). Ακόμα και στον οργανωσιακό τομέα ενός ιατρικού κέντρου, τα νευρωνικά δίκτυα μπορούν να προσφέρουν πολύτιμα εργαλεία, όπως μια μέθοδος που προτάθηκε από τους Gul και Guneri (2015) η οποία εκτιμά τη διάρκεια παραμονής ενός ασθενή στο τμήμα επειγόντων περιστατικών ενός νοσοκομείου, βάσει παραγόντων όπως η ηλικία, το φύλο, ο τρόπος προσέλευσης, τα αποτελέσματα των πρώτων ιατρικών δοκιμών κ.α.

### Βιολογικές - γεωλογικές - περιβαλλοντολογικές εφαρμογές

Οι εφαρμογές στον φυσικό κόσμο μπορούν να ξεκινούν από το κυτταρικό επίπεδο, όπως η ταξινόμηση αλυσίδων DNA με σκοπό τον εντοπισμό μοτίβων και ανωμαλιών (Garro et al.,

2016) και να φτάνουν μέχρι τον εντοπισμό οργανισμών που ζουν στα βάθη των ωκεανών από σκοτεινές φωτογραφίες του βυθού (Hollis et al., 2016), ή την ακριβή πρόβλεψη της στάθμης μιας λίμνης σε καθημερινή βάση (Shiri et al., 2016).

### Αρχιτεκτονικές και βιομηχανικές εφαρμογές

Τα νευρωνικά δίκτυα εξυπηρετούν αυτούς τους κλάδους με δυο συνήθως τρόπους: εφαρμογές στατιστικού ποιοτικού ελέγχου, και προσομοιώσεις. Ενδεικτικά αναφέρουμε δύο πολύ πρόσφατες εφαρμογές: εκτίμηση της ταλάντωσης πολυκατοικιών σε περίπτωση σεισμού (Suryanita et al., 2016), και πρόβλεψη της εξάπλωσης της σκόνης σε έργα γεωτρήσεων - εκσκαφών (Nagesha et al., 2015).

### Εφαρμογές στην φωτογραφική τέχνη

Πέρα από την αναγνώριση εικόνας που αναφέραμε σε προηγούμενη υποενότητα, οι αλγόριθμοι τεχνητών νευρωνικών δικτύων αποτελούν ένα πολύ χρήσιμο εργαλείο για την επεξεργασία εικόνων, χάρις στην ικανότητά τους να προσαρμόζονται στις ιδιότητες των δεδομένων εισόδου τους. Μια εξαιρετικά ενδιαφέρουσα πρόσφατη εφαρμογή αντιμετωπίζει το κλασικό πρόβλημα του επαναχρωματισμού ασπρόμαυρων εικόνων. Αυτή η εφαρμογή εκπαιδεύεται πάνω σε ένα μεγάλο σύνολο έγχρωμων φωτογραφιών, και αντιστοιχίζει χρώματα σε διάφορα αντικείμενα στις εικόνες, αλλά και σε ευρύτερα χαρακτηριστικά των εικόνων όπως η ηλιοφάνεια, η ύπαρξη τεχνητού φωτός κ.α. Χωρίς ανθρώπινη παρέμβαση, η εφαρμογή επιτυγχάνει αρκετά αληθοφανή χρωματικά αποτελέσματα σε νέες εισόδους ασπρόμαυρων εικόνων (Iizuka et al., 2016). Πηγαίνοντας ένα βήμα παραπέρα, μια άλλη πρόσφατη εφαρμογή μεταφέρει το στυλ ζωγραφικής ενός πίνακα σε μια φωτογραφία δημιουργώντας εντυπωσιακά καλλιτεχνικά αποτελέσματα (Gatys et al., 2016).

### Εφαρμογές στις εμπορικές και βιομηχανικές επιχειρήσεις

Ο τομέας των επιχειρήσεων είναι ιδανικό περιβάλλον για την άνθιση εφαρμογών νευρωνικών δικτύων, καθώς τα νευρωνικά δίκτυα εκπαιδεύονται πολύ καλύτερα όταν το σύνολο των εκπαιδευτικών τους παραδειγμάτων είναι μεγάλο και ανομοιογενές, και οι επιχειρήσεις διατηρούν έτσι κι αλλιώς τεράστιες βάσεις δεδομένων με στοιχεία πελατών, συναλλαγών, χρηματοοικονομικών δεικτών κ.λ.π. Φυσικά αυτοί οι αλγόριθμοι δεν είναι πανάκεια: για απλά θέματα όπως ο υπολογισμός της μισθοδοσίας, των χρηματοροών και του λογιστικού ισοζυγίου μιας επιχείρησης, οι παραδοσιακοί υπολογιστές είναι πολύ πιο δόκιμοι, ενώ όταν υπάρχουν συγκεκριμένοι κανόνες που πρέπει να ακολουθεί μια διεργασία, σίγουρα είναι προτιμώμε ένα *expert system*. Από την άλλη μεριά, οι αλγόριθμοι Γ.Ν.Δ. αντιμετωπίζουν εξαιρετικά θέματα όπως η κατηγοριοποίηση δεδομένων σε κλάσεις, η πρόβλεψη χρονοσειρών και η εύρεση μοτίβων ή ανωμαλιών σε μεγάλα σύνολα δεδομένων. Λόγω της πολυπλοκότητάς τους, οι επιχειρήσεις που καταφέρνουν να βρουν νέες, ακριβέστερες και αποτελεσματικότερες λύσεις σε τέτοια προβλήματα μπορούν να αποκτήσουν σημαντικά ανταγωνιστικά πλεονεκτήματα.



Σε αυτήν την ενότητα θα παρουσιάσουμε διάφορα παραδείγματα εφαρμογών, που εμπίπτουν στις παρακάτω βασικές κατηγορίες επιχειρηματικών δραστηριοτήτων που επωφελούνται περισσότερο από την χρήση αλγορίθμων Τ.Ν.Δ.

## Πωλήσεις και Μάρκετινγκ

Χάρης στην δυνατότητά τους να συνεκτιμούν πολλούς διαφορετικούς παράγοντες όπως η ζήτηση ενός προϊόντος, το μέγεθος και οι οικονομικές δυνατότητες της αγοράς-στόχου, ή η τιμή των συμπληρωματικών προϊόντων, τα νευρωνικά δίκτυα έχουν πολύ καλά αποτελέσματα σε εφαρμογές πρόβλεψης πωλήσεων. Φυσικά, εδώ και δεκαετίες οι επιχειρήσεις χρησιμοποιούν άλλες, παραδοσιακές στατιστικές μεθόδους για την πρόβλεψη των πωλήσεών τους όπως η γραμμική παλινδρόμηση και η εποχική ανάλυση, αλλά μελέτες όπως αυτή των Zhang και Qi (2002) απέδειξαν ότι ένας συνδυασμός αυτών των παραδοσιακών τεχνικών με τεχνητά νευρωνικά δίκτυα αποδίδει πολύ καλύτερα.

Το δε μάρκετινγκ είναι ίσως η επιχειρησιακή λειτουργία με τις περισσότερες ευκαιρίες για χρήση αλγορίθμων Τ.Ν.Δ. Βασικός στόχος του μάρκετινγκ είναι ο εντοπισμός των πελατών που ενδέχεται να αντιδράσουν θετικά σε ένα προϊόν ή μια υπηρεσία, ώστε η διαφήμιση αυτού του προϊόντος ή της υπηρεσίας να είναι πιο εστιασμένη και να έχει μεγαλύτερο ποσοστό επιτυχίας. Μια εφαρμογή Τ.Ν.Δ. κατηγοριοποίησης μπορεί τμηματοποιήσει την αγορά βάσει δημογραφικών, κοινωνικών και οικονομικών παραγόντων, γεωγραφικής θέσης, αγοραστικής συμπεριφοράς κ.α (Badea, 2014). Παράδειγμα τέτοιας εφαρμογής είναι η εργασία των Πετρουλάκη και Μιαουδάκη, πάνω στην τμηματοποίηση της αγοράς κινητών τηλεφώνων (Petroulakis and Miaoudakis, 2007). Ακόμα μεγαλύτερο ενδιαφέρον έχουν όμως εφαρμογές Τ.Ν.Δ. μάθησης χωρίς επίβλεψη, οι οποίες μπορούν να τμηματοποιήσουν την αγορά βάσει μοτίβων που βρίσκουν αυτόματα και που δεν ανήκουν στα τυπικά χαρακτηριστικά που αναφέραμε πριν, υποδεικνύοντας έτσι μερίδια αγοράς που ανταποκρίνονται στην πραγματικότητα χωρίς να είναι άμεσα προφανή. Νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν και στο αμέσως επόμενο βήμα, την επίδειξη της σωστής διαφήμισης στον σωστό πελάτη, όπως για παράδειγμα με ένα σύστημα προτάσεων αγοράς (recommender system) σαν αυτό που περιγράφουν οι Liu και Singh (2015).

Μια άλλη σημαντική λειτουργία του μάρκετινγκ είναι η διερεύνηση της αγοραστικής συμπεριφοράς των καταναλωτών, με στόχο την καλύτερη στόχευση στρατηγικών όπως τα bundle deals, το upselling και το crossselling. Κλασική τέτοια εφαρμογή στον τομέα της μηχανικής μάθησης χωρίς επίβλεψη είναι η *ανάλυση καλάθιού (basket analysis)*, στην οποία αναφερθήκαμε δίνοντας και σχετικό παράδειγμα στην ενότητα που μιλούσε για τους κανόνες συσχέτισης. Αναλύοντας τις συναλλαγές με τους πελάτες τους και εντοπίζοντας προϊόντα που αγοράζονται συχνά μαζί, μεγάλες επιχειρήσεις λιανικού εμπορίου όπως η Tesco και η Target μπορούν να κάνουν προτάσεις συμπληρωματικών προϊόντων σε πραγματικό χρόνο στους πελάτες τους, η και πολύ απλούστερα πράγματα όπως το να βάζουν προϊόντα που αγοράζονται συχνά μαζί σε κοντινά ράφια, δίνοντας έτσι κίνητρα στους πελάτες για να αγοράσουν περισσότερα αντικείμενα κατά την επίσκεψή τους στο (ηλεκτρονικό ή μη)

κατάστημα (Kar and De, 2009). Μια άλλου τύπου εφαρμογή που εστιάζεται στην καταναλωτική συμπεριφορά μπορεί να διαχωρίζει τους πιο σημαντικούς πελάτες από τους λιγότερο σημαντικούς, με κριτήρια όπως την μέση αξία αγορών ή την συχνότητά τους, ώστε να προσφέρονται περισσότερο στοχευμένες διαφημίσεις και προσφορές στους καλούς πελάτες (Ip et al, 2002).

## Βελτιστοποίηση Επιχειρησιακών Λειτουργιών

Νευρωνικά δίκτυα έχουν χρησιμοποιηθεί με μεγάλη επιτυχία και στον τομέα της διαχείρισης λειτουργιών, και ιδιαίτερα στους τομείς του σχεδιασμού και χρονοπρογραμματισμού δραστηριοτήτων. Ο υπολογισμός της ποσότητας παραγωγής, η οργάνωση της αλυσίδας παραγωγής και η διαχείριση ομάδων παραγωγής με διακριτούς ρόλους είναι θέματα με συχνά πολύ μεγάλο αριθμό παραμέτρων, που καλούν για λύσεις με Τ.Ν.Δ. (Zhang et al., 2016). Μια εφαρμογή Τ.Ν.Δ. μπορεί, βασιζόμενη σε ιστορικά στοιχεία αποθηκευμένα στο σύστημα ERP της επιχείρησης να εντοπίσει που δαπανάται ο περισσότερος χρόνος η ο λιγότερος χρόνος κατά την διεκπεραίωση ενός έργου, βοηθώντας έτσι στο χρονοπρογραμματισμό του (Relich and Muszyński, 2014).

Η κοστολόγηση ενός προϊόντος είναι επίσης μια διαδικασία που εξαρτάται από όλες τις φάσεις παραγωγής του και περνάει μέσα από διάφορα τμήματα της επιχείρησης. Μια πολύ ενδιαφέρουσα μελέτη των Kim και Han δείχνει ότι ένας συνδιασμός γενετικών αλγορίθμων και Τ.Ν.Δ. είναι ξεκάθαρα αποτελεσματικότερος στον εντοπισμό των σημαντικότερων παραγόντων κοστολόγησης ενός προϊόντος από απλούστερες στατιστικές μεθόδους (Kim and Han, 2003).

Τέλος, τα νευρωνικά δίκτυα μπορούν να προσφέρουν εκπληκτικές λύσεις σε θέματα ελέγχου ποιότητας όπου οι τυπικές στατιστικές μέθοδοι τύπου six sigma δεν είναι αρκετές. Ως παράδειγμα μπορούμε να αναφέρουμε μια τεχνική ελέγχου ποιότητας ραφών υφασμάτων η οποία δεν χρειάζεται καμία μέτρηση ως είσοδο, παρά μόνο φωτογραφίες των εν λόγω ραφών, για να προσδιορίσει αν αυτές ακολουθούν τα προκαθορισμένα πρότυπα ποιότητας (Bahlmann et al., 1999).

## Εφαρμογές στον τραπεζικό και χρηματιστηριακό τομέα.

Τα τραπεζικά και χρηματιστηριακά θέματα είναι κατεξοχήν ποσοτικά. Οι ερευνητές και οι επιχειρηματίες σε αυτούς τους τομείς ενδιαφέρονται πρωτίστως για το μέγεθος και την πορεία μεταβλητών όπως οι τιμές πρώτων υλών και μετοχών, τα επιτόκια και οι ρυθμοί μεταβολής τους, και οι πιθανότητες χρεοκοπίας ή αδυναμίας κάλυψης υποχρεώσεων φυσικών προσώπων και επιχειρήσεων. Ο λόγος που τα νευρωνικά δίκτυα διαπρέπουν και σε αυτούς τους τομείς είναι ότι όλες οι παραπάνω μεταβλητές αλλάζουν διαρκώς σε ένα περιβάλλον τόσο μεγάλης αβεβαιότητας και εξαρτώνται από τόσους πολλούς παράγοντες, που οι κλασσικές στατιστικές μέθοδοι δεν είναι πλέον επαρκείς για να δώσουν αρκετά ακριβείς προβλέψεις και εκτιμήσεις.

Θα εξετάσουμε τέσσερις κατηγορίες τραπεζικών και χρηματιστηριακών θεμάτων, τα οποία έχουν τα εξής κοινά χαρακτηριστικά που υποδεικνύουν ανάγκη εφαρμογής αλγορίθμων τεχνητών νευρωνικών δικτύων:

- τεράστιος όγκο δεδομένων που διογκώνεται καθημερινά (*big data*)
- μεγάλος αριθμός παραγόντων που τα επηρεάζουν
- αβεβαιότητα σε βαθμό που εντελώς ακριβείς λύσεις είναι ανέφικτες, αλλά καλές προσεγγίσεις της πραγματικότητας είναι εξαιρετικά αξιοποιήσιμες.

### Πρόβλεψη χρονοσειρών

Όσον αφορά στον χρηματιστηριακό τομέα, η πιο δημοφιλής εφαρμογές είναι φυσικά αυτές που επιχειρούν προβλέψεις τιμών μετοχών. Εφαρμογές όπως το KRANG (Roghani, 2015) καταφέρνουν ένα αξιοπρεπή βαθμό ακρίβειας, με κύριο προτέρημά τους το ότι ο χρήστης δεν χρειάζεται να εισάγει ένα πολύπλοκο σύστημα λογικών κανόνων, παρά μόνο να επιλέξει τους παράγοντες που θεωρεί ότι μπορούν να επηρεάσουν την χρονοσειρά και να αφήσει το T.N.Δ. να εκπαιδευτεί πάνω στα ιστορικά δεδομένα για να καταλήξει στο συμπέρασμα για τους παράγοντες που είναι λιγότερο ή περισσότερο σημαντικοί. Επίσης, τεχνικές όπως η *Μέθοδος Ανάλυσης Κυρίων Συνιστωσών* (PCA) μπορούν να συνδιαστούν με αλγορίθμους T.N.Δ., επιτυγχάνοντας μικρότερο σφάλμα πρόβλεψης από παραδοσιακές στατιστικές μεθόδους (Grigoryan, 2015) Για λιγότερο συγκεκριμένους δείκτες όπως ο γενικός δείκτης ενός χρηματιστηρίου, στην βιβλιογραφία βρίσκουμε εφαρμογές που ισχυρίζονται τρομερά ποσοστά ακρίβειας (Kar, 1990).

Ένας άλλος πολύ σημαντικός δείκτης, με μεγάλες ευκαιρίες κερδοφορίας από την ακριβή πρόβλεψή του είναι η συναλλαγματική ισοτιμία μεταξύ νομισμάτων. Και εδώ υπάρχουν έρευνες που δείχνουν ότι δεδομένης της συνεχούς τροφοδοσίας ενός νευρωνικού δικτύου με πρόσφατα ιστορικά στοιχεία, αυτό μπορεί να έχει ακριβέστερες προβλέψεις από μια κλασική μέθοδο προβλέψεων όπως το μοντέλο ARIMA, ακόμη και χωρίς εξωτερικούς παράγοντες ως δεδομένα εισόδου (Yao and Tan, 2002).

### Έλεγχος πιστοληπτικής ικανότητας

Ο έλεγχος πιστοληπτικής ικανότητας είναι βασικό μέλημα των τραπεζών πριν δανείσουν ένα φυσικό πρόσωπο ή μια επιχείρηση. Καθώς ουσιαστικά αποτελεί ένα πρόβλημα κατηγοριοποίησης (αιτήσεων) βάσει ανομοιογενών χαρακτηριστικών (προφίλ πελάτη), χωρίς ιστορικά στοιχεία για τον ίδιο τον πελάτη κάθε φορά, αλλά με ιστορικά στοιχεία για το σύνολο των δανείων που έχει δώσει ο οργανισμός η ή τράπεζα, αποτελεί ένα κλασσικό πεδίο εφαρμογής αλγορίθμων T.N.Δ.

Οι περισσότερες τέτοιες εφαρμογές κατατάσσουν τις αιτήσεις, ανάλογα με τα διαθέσιμα χαρακτηριστικά και το ιστορικό της εταιρίας, σε ασφαλείς, αβέβαιες, ή επισφαλείς, επιταχύνοντας έτσι την διαδικασία λήψης αυτής της απόφασης, και επιτρέποντας ενδεχομένως και σε κατώτερα στελέχη τη δυνατότητα να αναλάβουν αυτήν την απόφαση με

ένα καλό επίπεδο σιγουριάς (De Nittis et al., 1998). Σε ένα παρόμοιο άρθρο παρατηρείται ότι μια τέτοια διαδικασία θα μπορούσε να γίνει σε δύο στάδια: σε πρώτο στάδιο να γίνονται αποδεκτές ή να απορρίπτονται αυτόματα οι ασφαλείς ή οι επισφαλείς αιτήσεις αντίστοιχα, και στη συνέχεια να γίνεται αξιολόγηση εκ νέου για τις περιπτώσεις δεν μπορούν να ταξινομηθούν με σιγουριά, ώστε ο λήπτης της απόφασης να έχει μια ιδέα για την πλευρά στην οποία κλίνουν (West and Muchineuta, 2002).

Σχετικό είναι και το θέμα της πρόβλεψης χρεωκοπίας, κυρίως στο επίπεδο των επιχειρήσεων. Με παρόμοιο τρόπο ένα T.N.Δ. μπορεί να εκπαιδευτεί πάνω σε ιστορικά δεδομένα για να εκτιμά την πιθανότητα χρεωκοπίας μιας επιχείρησης βάσει χρηματοοικονομικών δεικτών της (Jones et al., 2017).

### Εκτίμηση αξίας

Η εκτίμηση της αξίας ενός ακινήτου ή άλλου περιουσιακού στοιχείου είναι σημαντική για τις τράπεζες, ιδιαίτερα σε περιπτώσεις υποθηκών ή πλειστηριασμών. Ιδιαίτερα στην περίπτωση των ακινήτων, μέθοδοι T.N.Δ. που λαμβάνουν υπόψη παράγοντες όπως οι καιρικές και περιβαλλοντικές συνθήκες, ή η πρόσβαση σε μέσα μαζικής μεταφοράς κάνουν πολύ καλύτερες προβλέψεις από κλασικά μοντέλα πρόβλεψης χρονοσειρών (Chiarazzo et al., 2014).

### Ανίχνευση απάτης

Ο εντοπισμός ύποπτων τραπεζικών συναλλαγών, κινήσεων λογαριασμών ή αγορών που καταδεικνύουν ενδεχόμενη εξαπάτηση, κλοπή, παραβίαση ενός ηλεκτρονικού συστήματος ή ακόμα και επικίνδυνου ανθρώπινου λάθους, είναι μια γενικά πολύ δύσκολη διαδικασία. Όσο ικανός και να είναι ένας ελεγκτής, δεν μπορεί να εκτιμήσει κάθε περίπτωση ξεχωριστά δεδομένου του τεράστιου όγκου δεδομένων που μπορεί να έχει μια βάση συναλλαγών η οποία να ανανεώνεται μάλιστα με ταχύτατο ρυθμό σε πραγματικό χρόνο. Η δυνατότητα των T.N.Δ. να βρίσκουν στατιστικά σημαντικά μοτίβα σε μεγάλα σύνολα δεδομένων, λύνει τα χέρια των ελεγκτών, καθώς το T.N.Δ. μπορεί να εκπαιδευτεί μόνο του πάνω στα ιστορικά δεδομένα και να παρουσιάζει αυτόματα στον ελεγκτή περιπτώσεις που θεωρεί στατιστικά ασυνήθιστες, ή που παρουσιάζουν κοινά χαρακτηριστικά με προηγούμενες κακόβουλες συναλλαγές (Fawcett et al., 1998). Εφαρμογές ανίχνευσης απάτης αυτού του τύπου μπορούν να χρησιμοποιηθούν όχι μόνο για να εντοπίζουν απάτες που έχουν ήδη διαπραχθεί, αλλά και ως προληπτικές μέθοδοι που επισημαίνουν περιπτώσεις που παρουσιάζουν ύποπτα χαρακτηριστικά (Estévez et al., 2006).

## 4.ΠΡΟΤΕΙΝΟΜΕΝΗ ΕΦΑΡΜΟΓΗ

Κλείνοντας την παρουσίαση των οικονομικών και επιχειρηματικών εφαρμογών των νευρωνικών δικτύων, προτείνουμε μια νέα εφαρμογή σε ένα πεδίο επιχειρηματικού ενδιαφέροντος στο οποίο δεν έχει δοθεί μεγάλη προσοχή από τους ερευνητές, τουλάχιστον σύμφωνα με την τρέχουσα διαθέσιμη βιβλιογραφία. Φυσικά, είναι πολύ πιθανό μια μέθοδος δομημένη με παρεμφερή τρόπο με αυτόν που προτείνεται εδώ να εφαρμόζεται ήδη σε κάποιες επιχειρήσεις, αλλά να μην έχει λάβει δημοσιότητα, λόγω της κοινής πρακτικής των περισσότερων επιχειρήσεων να διατηρούν στο εσωτερικό τους ως επιχειρηματικά μυστικά μεθοδολογίες που εν δυνάμει τους προσφέρουν ανταγωνιστικό πλεονέκτημα. Η πρόταση εφαρμογής που θα περιγράψουμε παρουσιάστηκε για πρώτη φορά από το συγγραφέα αυτής της πτυχιακής εργασίας στο διεθνές συνέδριο *ICCM I 2017 - 5<sup>th</sup> International Conference on Contemporary Marketing Issues* στην Θεσσαλονίκη, υπό τον τίτλο *Identification of potentially undelivered packages with an artificial neural network method* (Kanellis and Papadopoulos, 2017).

### Ταυτοποίηση πακέτων που ενδέχεται να μην φτάσουν στον παραλήπτη τους

Η προτεινόμενη εφαρμογή έχει να κάνει με την χρήση αλγορίθμου νευρωνικών δικτύων ο οποίος εκτιμά αν ένα πακέτο λιανικού εμπορίου ενδέχεται να χαθεί κατά την μεταφορά πριν φτάσει στον παραλήπτη του. Το νευρωνικό δίκτυο δέχεται διάφορα χαρακτηριστικά του εμπορεύματος, του πελάτη και της παραγγελίας, και εκπαιδεύεται πάνω στο ιστορικό παραγγελιών της επιχείρησης. Η έξοδος του δικτύου αποτελεί μια εκτίμηση για το αν το πακέτο θα φτάσει στον παραλήπτη του ή όχι. Στην συνέχεια η επιχείρηση μπορεί να αξιοποιήσει τις εξόδους του δικτύου ώστε να διαμορφώσει την πολιτική αποστολής της για κάθε πακέτο ξεχωριστά, μειώνοντας έτσι τα κόστη αποστολής της, καθώς και ενδεχόμενες απώλειες από επιστροφές εμπορευμάτων ή χρημάτων.

### Περιγραφή του προβλήματος

Στην προσπάθειά τους να επιβιώσουν στην σύγχρονη, ευρέως παγκοσμιοποιημένη οικονομία, οι περισσότερες επιχειρήσεις λιανικού εμπορίου απομακρύνονται σταδιακά από το παραδοσιακό πρότυπο της τοπικής αγοράς και επιδιώκουν εξάπλωση στην διεθνή αγορά. Σε μεγάλο βαθμό, αυτό οφείλεται στις μεγάλες επιχειρήσεις λιανικού ηλεκτρονικού εμπορίου (Amazon, Alibaba, eBay κ.α.) οι οποίες τα τελευταία χρόνια με την εξάπλωση του διαδικτύου κατάφεραν να έρθουν κοντά σε εκατομμύρια πελάτες και να κερδίσουν την εμπιστοσύνη τους. Αυτές οι επιτυχίες ώθησαν πολλές μικρότερες επιχειρήσεις στην υιοθέτηση του ηλεκτρονικού εμπορίου, με την ελπίδα να εξελιχθούν από τοπικούς σε διεθνείς προμηθευτές αγαθών. Ανεξαρτήτως του μεγέθους της επιχείρησης, το ηλεκτρονικό εμπόριο φαίνεται να έχει κάτι παραπάνω να προσφέρει: μια μεγάλη εμπορική επιχείρηση μπορεί να πολλαπλασιάσει τις πωλήσεις της, εκμεταλλευόμενη μεταξύ άλλων και διαφορές

στις νομισματικές ισοτιμίες, ενώ μια μικρότερη μπορεί να καταφέρει να αποκτήσει μέσω του διαδικτύου ένα κοινό αρκετό για να την κρατήσει ζωντανή, κάτι που ενδεχομένως να μην ήταν εφικτό αν στηριζόταν μόνο σε τοπικούς πελάτες.

Η υιοθέτηση μιας παγκόσμιας αγοράς - στόχου, μέσω του ηλεκτρονικού εμπορίου, βέβαια, ενέχει και κάποια νέα κόστη για την επιχείρηση, με ίσως σημαντικότερο εξ αυτών το κόστος αποστολής σε παραλήπτες εκτός συνόρων. Η διαχείριση αυτού του κόστους είναι ιδιαίτερα σημαντική για μικρομεσαίες επιχειρήσεις, οι οποίες δεν έχουν δικά τους κανάλια διανομής και εξαρτώνται από τρίτους διανομείς. Για τους δε καταναλωτές, το υψηλό κόστος αποστολής ενός πακέτου αποτελεί μια από τις συχνότερες αιτίες προτίμησης ενός τοπικού καταστήματος από ένα κατάστημα ηλεκτρονικού εμπορίου σε άλλη χώρα. Το συμπέρασμα είναι ότι μια επιχείρηση που μπορεί να μειώσει αυτά τα επιμέρους κόστη αποστολής, ή ακόμα και να προσφέρει δωρεάν αποστολή στους πελάτες της, μπορεί να επιτύχει ξεκάθαρο ανταγωνιστικό πλεονέκτημα σε σχέση με άλλες παρεμφερείς επιχειρήσεις.

Το κόστος αποστολής ενός πακέτου αναλύεται σε πολλά επιμέρους κόστη. Τα περισσότερα από αυτά έχουν βέβαια να κάνουν με παράγοντες που δεν ελέγχει η ίδια η επιχείρηση, όπως η τιμολογιακή πολιτική της μεταφορικής με την οποία συνεργάζεται ο έμπορος, ενδεχόμενοι τελωνειακοί δασμοί κ.α. Η συγκεκριμένη πρόταση εφαρμογής επικεντρώνεται στην προσπάθεια περιορισμού ενός από αυτά τα επιμέρους κόστη, που καθορίζεται βάσει απόφασης της ίδιας της εμπορικής επιχείρησης, το οποίο είναι η αξιοποίηση πρόσθετων υπηρεσιών που παρέχονται από την μεταφορική, όπως η *παρακολούθηση πακέτου* ή η *επιβεβαίωση παραλαβής*. Ανάλογα με την εκάστοτε μεταφορική, το όνομα και η ακριβής περιγραφή αυτών των υπηρεσιών μπορεί να διαφέρει, αλλά στην παρούσα εργασία, χάριν απλότητας και συντομίας, θα χρησιμοποιήσουμε τον γενικό όρο *παρακολούθηση πακέτου*, εννοώντας οποιαδήποτε πρόσθετη αμειβόμενη υπηρεσία της μεταφορικής που επιτρέπει στον έμπορο να εξακριβώσει αν ένα πακέτο που έστειλε έφτασε στον παραλήπτη του, ή χάθηκε κατά την μεταφορά.

## Παρακολούθηση πακέτων

Για να αποδείξει ότι ένα πακέτο έφτασε στον παραλήπτη του, η μεταφορική συνήθως θα πρέπει να το παραδώσει στον παραλήπτη αυτοπροσώπως και να του ζητήσει να υπογράψει για την επιτυχή παραλαβή του. Επειδή αυτό κοστίζει σε πόρους και εργατοώρες, οι περισσότερες μεταφορικές επιχειρήσεις δεν προσφέρουν πάγια αυτήν την υπηρεσία, παρά μόνο με πρόσθετη χρέωση. Φυσικά, κάθε πρόσθετη χρέωση, όσο μικρή και αν είναι ανά αντικείμενο, περιορίζει αναπόφευκτα το περιθώριο κέρδους της επιχείρησης. Έτσι, με την εξαίρεση κάποιων επιχειρήσεων που θεωρούν την παροχή αυτής της υπηρεσίας στους πελάτες τους ως κάτι το αδιαπραγμάτευτο, οι περισσότερες επιχειρήσεις λιανικού εμπορίου θα προτιμούσαν, αν είναι δυνατό, να αποφύγουν εντελώς αυτό το επιπλέον έξοδο.

Φυσικά, η απόφαση για τη χρήση ή όχι αυτής της υπηρεσίας δεν είναι τόσο απλή. Κάθε πακέτο που δεν παρακολουθείται προϋποθέτει την ανάληψη συγκεκριμένου ρίσκου από την

επιχείρηση: σε περίπτωση που ο πελάτης της δηλώσει ότι το πακέτο δεν έφτασε ποτέ στα χέρια του, η επιχείρηση δεν έχει τρόπο να εξακριβώσει αν η δήλωση αυτή είναι αληθής ή ψευδής, κάτι που την καθιστά ευάλωτη τόσο σε σφάλματα κατά τη μεταφορά του πακέτου, όσο και σε απόπειρες εξαπάτησής της από τον πελάτη! Η πολιτική των περισσότερων εταιριών σε περιπτώσεις απώλειας πακέτου κατά την μεταφορά είναι να επιστρέφουν τα χρήματα στον πελάτη, ή (σπανιότερα) να του στέλνουν εκ νέου ένα ίδιο πακέτο χωρίς επιπρόσθετη χρέωση. Σε κάθε περίπτωση, η επιχείρηση χάνει όλη την αξία του εμπορεύματός της, κάτι που θα μπορούσε να είχε αποφευχθεί αν είχε στα χέρια της απόδειξη παραλαβής του πακέτου, ή στοιχεία που θα της επέτρεπαν να ζητήσει αποζημίωση από την μεταφορική λόγω σφάλματος κατά την μεταφορά.

Στην πράξη, οι επιχειρήσεις προσπαθούν να ελαχιστοποιήσουν τις απώλειές τους και από τα πρόσθετα κόστη αποστολής αλλά και από ενδεχόμενες αποζημιώσεις σε πελάτες ακολουθώντας κάποιους αυθαίρετους, πρακτικούς κανόνες κατά την αποστολή των πακέτων όπως:

- παρακολούθηση πακέτων μόνο για αντικείμενα με τιμή μεγαλύτερη από κάποιο καθορισμένο ποσό
- παρακολούθηση πακέτων μόνο για αντικείμενα βαρύτερα ή μεγαλύτερα σε διαστάσεις από ένα συγκεκριμένο όριο, το οποίο μπορεί να καθορίζεται και βάσει της τιμολογιακής πολιτικής της μεταφορικής
- παρακολούθηση πακέτων μόνο για πελάτες σε χώρες που θεωρούνται “υψηλής επικινδυνότητας” λόγω υψηλού ποσοστού αποτυχημένων αποστολών που παρατηρήθηκε σε προηγούμενες συναλλαγές με πελάτες αυτών των χωρών, κ.ο.κ.

Τέτοιοι κανόνες είναι συνήθως απλοί στην εφαρμογή τους και πρακτικοί, αλλά θεωρούμε ότι δεν θα μπορούσαν να είναι ακριβέστεροι, και συνεπώς αποτελεσματικότεροι, από μια πιο εκλεπτυσμένη μέθοδο σαν έναν *αλγόριθμο ταξινόμησης*. Ιδανικά, η επιχείρηση θα ήθελε να έχει πληρώσει για την παρακολούθηση μόνο των πακέτων που “χάθηκαν”, και να έχει αποφύγει αυτό το επιπλέον έξοδο για όλα τα υπόλοιπα. Ο στόχος, λοιπόν, του αλγόριθμου ταξινόμησης, είναι να εκτιμά, βάσει παραμέτρων σχετικών με το εμπόρευμα, τον πελάτη και την παραγγελία, εάν αυτή αναμένεται να φτάσει ή όχι στον πελάτη.

## Περιγραφή της εφαρμοζόμενης μεθόδου

Ο αλγόριθμος ταξινόμησης που προτείνουμε και περιγράφουμε σε αυτήν την είναι ένα *πολυεπίπεδο νευρωνικό δίκτυο οπισθοδιάδοσης (multilayer backpropagation neural network)*, χωρίς να αποκλείουμε, σε αυτή τη φάση, και τη χρήση ενός πιο εξειδικευμένου τύπου νευρωνικών δικτύων. Θεωρούμε ότι οι ιδιαίτερες δυνατότητες των αλγορίθμων νευρωνικών δικτύων μπορούν να βρουν εξαιρετική εφαρμογή στα ιδιαίτερα χαρακτηριστικά του προβλήματος. Πιο συγκεκριμένα:

- η δυνατότητα προσαρμογής πολλών μεταβλητών στο μοντέλο, ακόμα και αν αυτές είναι εκατοντάδες, μας επιτρέπει να πειραματιστούμε με ένα τεράστιο μείγμα

χαρακτηριστικών πελάτη, εμπορεύματος και παραγγελίας πριν ο αλγόριθμος καταλήξει σε αυτά που μας οδηγούν στις ακριβέστερες προβλέψεις

- η πιθανότητα ανακάλυψης εξαρτήσεων μεταξύ φαινομενικά ασύνδετων μεταξύ τους μεταβλητών, μπορεί να οδηγήσει και σε πολύ ενδιαφέροντα για την επιχείρηση παράπλευρα συμπεράσματα
- η δυνατότητα εκπαίδευσης του δικτύου πάνω σε νέα δεδομένα βοηθά στον έγκαιρο εντοπισμό νέων τάσεων και μοτίβων στην συμπεριφορά των πελατών και στην αξιοπιστία των μεταφορικών εταιριών, κ.α.

Το νευρωνικό δίκτυο αποτελείται από ένα αρκετά μεγάλο επίπεδο εισόδου, ένα (ή περισσότερα) κρυφά επίπεδα, και ένα επίπεδο εξόδου με μόνο ένα κόμβο ο οποίος θα έχει δυαδική έξοδο: είτε το πακέτο εκτιμάται να φτάσει στον παραλήπτη, είτε όχι.

Το επίπεδο εισόδου δέχεται ως είσοδο μεταβλητές που είναι διαθέσιμες στην εταιρία κατά την στιγμή της συναλλαγής, όπως:

- **μεταβλητές σχετιζόμενες με το εμπόρευμα**
  - τιμή
  - περιθώριο κέρδους
  - όγκος
  - βάρος
  - κατηγορία εμπορεύματος
- **μεταβλητές σχετιζόμενες με την παραγγελία**
  - χρονική περίοδος της παραγγελίας (μήνας ή εβδομάδα του έτους)
  - μέθοδος παραγγελίας (ηλεκτρονικό κατάστημα, ηλεκτρονικό *marketplace*, τηλέφωνο κ.α.)
- **μεταβλητές σχετιζόμενες με τον πελάτη**
  - χώρα διαμονής του παραλήπτη
  - ασυμφωνία μεταξύ διεύθυνσης παραλήπτη (shipping address) και διεύθυνσης χρέωσης (billing address)
  - αριθμός προηγούμενων παραγγελιών από τον συγκεκριμένο πελάτη βάσει ιστορικού
  - ποσοστό αποτυχημένων παραγγελιών του συγκεκριμένου πελάτη

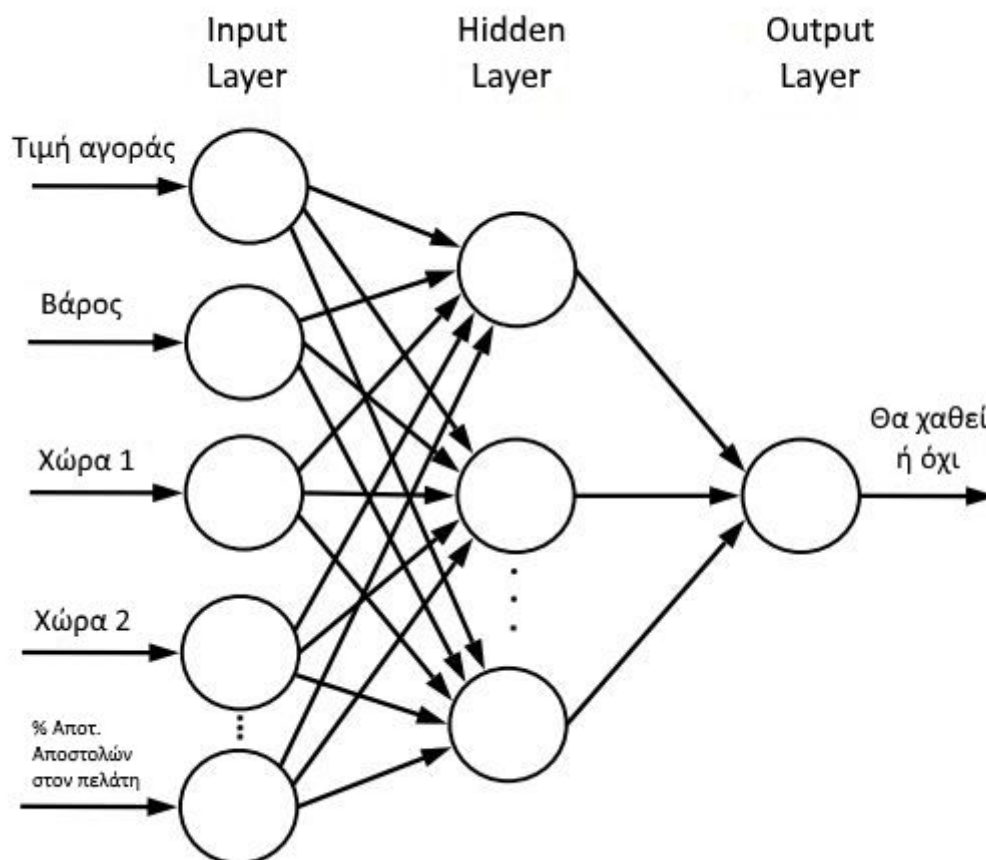
Οι παραπάνω μεταβλητές θα πρέπει να μεταγραφούν σε μια μορφή που θα μπορεί να αξιοποιηθεί από ένα Τ.Ν.Δ. Για τις κατηγορικές μεταβλητές, μπορούμε να δημιουργήσουμε μια δυαδική μεταβλητή για κάθε πιθανή τιμή της κατηγορικής μεταβλητής (μια πρακτική που οι προγραμματιστές αποκαλούν *dummy coding*). Έτσι, αν εξυπηρετούμε 100 χώρες, και μια παραγγελία αποστέλλεται στην Ελλάδα, τότε όσον αφορά στη συγκεκριμένη παραγγελία, η μεταβλητή που αναφέρεται στην Ελλάδα ως χώρα παραλήπτη θα πάρει την τιμή 1, και όλες



οι άλλες 99 μεταβλητές που αντιπροσωπεύουν χώρες παραλήπτη θα πάρουν την τιμή 0. Μεταβλητές όπως το βάρος ή ο όγκος μπορούν να εκφραστούν ως ιεραρχικές μεταβλητές για τους σκοπούς του αλγορίθμου (π.χ. 1 → πολύ μικρό, 2 → μικρό, 3 → μεσαίο, 4 → μεγάλο, 5 → πολύ μεγάλο κ.ο.κ), μια πρακτική που όχι μόνο απλοποιεί τους υπολογισμούς, αλλά μπορεί να εκφράζει παράλληλα και τις βαθμίδες τιμολόγησης της μεταφορικής επιχείρησης. Στο παράρτημα της παρούσας εργασίας δίνεται ένα παράδειγμα της μορφής που μπορεί να έχουν τα δεδομένα πριν και μετά την διαδικασία αυτής της μεταγραφής.

Ο κόμβος εξόδου του δικτύου θα παίρνει μια δυαδική τιμή: 0, αν εκτιμά ότι η παραγγελία δεν θα φτάσει στον παραλήπτη της, ή 1 αν εκτιμά ότι θα φτάσει.

Στο παρακάτω διάγραμμα μπορούμε να δούμε σχηματικά πως διαγράφεται η δομή του προτεινόμενου αλγορίθμου T.N.Δ.:



*Μια (συμπυκνωμένη) επισκόπηση της δομής του T.N.Δ.*

Το δίκτυο εκπαιδεύεται πάνω σε ιστορικά δεδομένα παραγγελιών της επιχείρησης (*μάθηση με επίβλεψη*). Ένα κομμάτι αυτής της βάσης δεδομένων χρησιμοποιείται ως εκπαιδευτικό σύνολο, και το υπόλοιπο ως *δοκιμαστικό σύνολο*, πάνω στο οποίο εξετάζεται η ακρίβεια ταξινόμησης του δικτύου σε νέα, μη χαρακτηρισμένα εκπαιδευτικά παραδείγματα. Χάρης στην μέθοδο της οπισθοδιάδοσης, μετά από αρκετές επαναλήψεις, οι τιμές των βαρών του δικτύου θα έχουν προσαρμοστεί έτσι ώστε να ελαχιστοποιείται η απόκλιση των τιμών εξόδου

- προβλέψεων από τις πραγματικές τιμές. Σε δεύτερο στάδιο, αυτές οι τιμές μπορούν να μας υποδείξουν και συσχετίσεις ανάμεσα σε μεταβλητές που δεν γνωρίζαμε από πριν, ενδεχομένως κατόπιν εφαρμογής ενός αλγορίθμου *εξαγωγής κανόνων (rule extraction)* πάνω στο ήδη εκπαιδευμένο νευρωνικό δίκτυο.

## Εκτίμηση των αποτελεσμάτων

Η μέγιστη ακρίβεια που θα επιτύχει ο αλγόριθμος επί του *δοκιμαστικού συνόλου*, θα πρέπει να συγκριθεί με την ακρίβεια που έχουν οι αυθαίρετοι, πρακτικοί κανόνες που αναφέρθηκαν νωρίτερα και χρησιμοποιούνται κατά κόρον στις επιχειρήσεις, με την τεχνική παραδοχή ότι τα πακέτα για τα οποία η επιχείρηση πληρώνει παρακολούθηση είναι τα πακέτα που εκτιμά ότι θα χαθούν. Θεωρούμε πως η ακρίβεια του αλγορίθμου θα είναι μεγαλύτερη, γνωρίζοντας την αποτελεσματικότητα των T.N.Δ. σε προβλήματα ταξινόμησης, αλλά ενδεχομένως η διαφορά ακρίβειας να μην είναι αρκετά σημαντική ώστε να δικαιολογεί ένα πρόσθετο, πολύπλοκο βήμα κατά την λήψη απόφασης τύπου αποστολής για το εκάστοτε εμπόρευμα.

## Χρήση των αποτελεσμάτων από την επιχείρηση

Θα πρέπει να επισημανθεί ότι η πρόβλεψη του αλγορίθμου δεν ισοδυναμεί με την απόφαση της επιχείρησης για χρήση ή μη της υπηρεσίας παρακολούθησης πακέτων. Ο τελικός αλγόριθμος σίγουρα θα αποτελέσει ένα χρήσιμο εργαλείο για την λήψη αυτής της απόφασης, αλλά η πρόβλεψή του έχει να κάνει με το αν θα φτάσει το πακέτο στον παραλήπτη, και όχι με το αν αξίζει να παρακολουθηθεί. Η απόφαση της επιχείρησης επηρεάζεται και από άλλες παραμέτρους εκτός του αλγορίθμου. Οι δύο πιθανές περιπτώσεις που αναφέρονται παρακάτω θα κάνουν αυτήν την παρατήρηση πιο ξεκάθαρη:

- ένα εμπόρευμα που επισημαίνεται από τον αλγόριθμο ως πιθανό να χαθεί κατά τη μεταφορά, μπορεί να έχει τόσο χαμηλό περιθώριο κέρδους, που μια πρόσθετη χρέωση θα στερούσε το νόημα από την πώλησή του. Σε τέτοιες περιπτώσεις, ο έμπορος θα προτιμήσει να στείλει το πακέτο χωρίς παρακολούθηση, προτιμώντας να αναλάβει το ρίσκο παρά να καταστήσει την πώληση ζημιογόνα εξ ορισμού.
- ένα εμπόρευμα μπορεί να είναι τόσο ακριβό, ή δύσκολο στο να αντικατασταθεί, ώστε ακόμη και αν ο αλγόριθμος θεωρεί την αποστολή του ασφαλής, ο έμπορος προτιμά να πληρώσει κάτι παραπάνω για την παρακολούθηση του πακέτου, παρά να αναλάβει το ρίσκο μιας λανθασμένης πρόβλεψης του αλγορίθμου, που θα του επιφέρει ζημιά πολύ μεγαλύτερη από το κόστος που θα είχε η υπηρεσία παρακολούθησης πακέτου.

## Μελλοντικοί στόχοι της εφαρμογής

Η εφαρμογή που περιγράφεται σε αυτό το κεφάλαιο δεν έχει δοκιμαστεί πρακτικά από τον συγγραφέα. Ο κυριότερος λόγος είναι η έλλειψη ενός κατάλληλου συνόλου δεδομένων: οι επιχειρήσεις είναι κατά κανόνα πολύ φειδωλές στην παροχή δεδομένων συναλλαγών, πόσο μάλλον στην περίπτωση που μιλάμε για αγαθά που δεν έφτασαν στους παραλήπτες τους!

Επίσης η αυθαίρετη δημιουργία ενός “τυχαίου” συνόλου δεδομένων για τους στόχους της έρευνας καταστρατηγεί έναν από τους βασικούς λόγους για τους οποίους θεωρούμε ότι αυτή η μέθοδος έχει νόημα για το συγκεκριμένο πρόβλημα: τον εντοπισμό μη προφανών συσχετίσεων μεταξύ χαρακτηριστικών που εντοπίζονται σε πραγματικά δεδομένα. Ελπίζω στο μέλλον να καταφέρω να εξασφαλίσω ένα κατάλληλο σύνολο δεδομένων ώστε να μπορέσω να συγκρίνω την αποτελεσματικότητα της μεθόδου με αυτήν που έχουν οι πρακτικοί τρόποι αντιμετώπισής του.

Σε δεύτερο στάδιο, θα είχε μεγάλο ενδιαφέρον η ενδεχόμενη αυτοματοποίηση του επόμενου βήματος της λήψης της απόφασης σχετικά με την χρήση ή όχι της υπηρεσίας παρακολούθησης πακέτου από τον έμπορο, όπως αυτό περιγράφεται στην προηγούμενη υποενότητα. Αυτό προϋποθέτει πιθανότατα την χρήση ενός δεύτερου αλγόριθμου ο οποίος θα δέχεται σαν όρισμα την τιμή εξόδου του νευρωνικού δικτύου, αλλά θα συνεκτιμά και άλλες παραμέτρους όπως το κόστος της υπηρεσίας, το περιθώριο κέρδους, και το μέγεθος της ζημιάς σε περίπτωση λανθασμένης πρόβλεψης.

Τέλος, θεωρούμε ότι μια ολοκληρωμένη εφαρμογή αυτού του τύπου θα μπορούσε να βρει εύκολα θέση σε ένα επιχειρησιακό πληροφοριακό σύστημα ERP, επιτρέποντας την απρόσκοπτη και αποτελεσματική αξιοποίηση των προβλέψεών της.

## 5.ΣΥΜΠΕΡΑΣΜΑΤΑ

Η σχεδόν καθολική χρησιμότητα των Τ.Ν.Δ., ιδιαίτερα σε κεντρικούς τομείς μιας επιχείρησης όπως ο σχεδιασμός στρατηγικής μάρκετινγκ και οι εκτιμήσεις πωλήσεων, υποχρεώνουν ουσιαστικά τις επιχειρήσεις που φιλοδοξούν να αποκτήσουν ανταγωνιστικό πλεονέκτημα, να εξερευνήσουν τις δυνατότητες που μπορούν να τους παρέχουν οι αλγόριθμοι Τ.Ν.Δ.

Ιδιαίτερα σε επιχειρήσεις που δραστηριοποιούνται σε κλάδους με έντονο το στοιχείο της συγκομιδής “μεγάλων δεδομένων”, όπως το ηλεκτρονικό εμπόριο ή ο τραπεζικός τομέας, θα πρέπει να αναγνωρισθεί η μετατόπιση του ερευνητικού ενδιαφέροντος των οικονομικών αναλυτών από τις παραδοσιακές στατιστικές μεθόδους προς μεθόδους τεχνητής νοημοσύνης όπως τα Τ.Ν.Δ., έτσι ώστε να γίνει μια αντίστοιχη αλλαγή στον τρόπο στελέχωσης και λειτουργίας των τμημάτων ανάλυσης και υποστήριξης αποφάσεών τους.

Όσον αφορά στους οικονομικούς αναλυτές, είναι βέβαιο πως η αποδοχή και η καθιέρωση της χρήσης των μεθόδων Τ.Ν.Δ. δημιουργεί πολύ καλές ευκαιρίες για όσους αποφασίσουν να επενδύσουν σε μια εξειδίκευση σε μεθόδους Τ.Ν.Δ., πιθανότατα με την χρήση μιας γλώσσας προγραμματισμού όπως η Python ή η MATLAB.

Η εφαρμογή που προτείνεται στην παρούσα εργασία δείχνει πως υπάρχουν πάντα περιθώρια βελτιστοποίησης, ακόμα και στις πιο απλές επιχειρηματικές διεργασίες, αρκεί να υπάρχει θέληση για αξιοποίηση σύγχρονων μεθόδων όπως τα Τ.Ν.Δ. σε κάθε βήμα της εφοδιαστικής αλυσίδας.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

### Ελληνική Βιβλιογραφία

- Ατζαλάκη, Ζ. και Κουτζαγιώτη, Α. (2005) *Τεχνητά νευρωνικά δίκτυα και πρόβλεψη χρονοσειρών*. Πτυχιακή Εργασία, Τ.Ε.Ι. Πατρών, τμ. Επιχειρηματικού Σχεδιασμού και Πληροφοριακών Συστημάτων
- Αφούρας, Τ. (2015) *Ταξινόμηση με Νευρωνικά Δίκτυα Ανάδρασης και Ενισχυτική Μάθηση*. Πτυχιακή Εργασία, Αριστοτέλειο Πανεπιστήμιο Μακεδονίας, τμ. Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
- Βλαχάβας, Ι., Κεφάλας, Π., Βασιλειάδης, Ν., Κόκκορας, Φ. και Σακελλαρίου, Η. (2011) *Τεχνητή Νοημοσύνη*. Γ' έκδοση. Εκδόσεις Πανεπιστημίου Μακεδονίας
- Βλαχοπούλου, Ε. (2014) *Η χρήση υπολογιστικών μεθόδων για την πρόβλεψη των τιμών μιας μετοχής*. Πτυχιακή Εργασία, Τ.Ε.Ι. Δυτικής Ελλάδας, τμ. Διοίκησης Επιχειρήσεων
- Διαμαντάρας, Κ. (2011) *Μηχανική Μάθηση*. Σημειώσεις μαθήματος, Αλεξάνδρειο Τ.Ε.Ι. Θεσσαλονίκης, τμ. Πληροφορικής
- Διαμαντάρας, Κ. και Γουλιάνας, Κ. (2011) *Εργαστηριακές Ασκήσεις Μηχανικής Μάθησης*. Σημειώσεις μαθήματος, Αλεξάνδρειο Τ.Ε.Ι. Θεσσαλονίκης, τμ. Πληροφορικής
- Διαμαντάρας, Κ. και Γουλιάνας, Κ. (2013) *Σύντομο Εγχειρίδιο MATLAB*. Σημειώσεις μαθήματος, Αλεξάνδρειο Τ.Ε.Ι. Θεσσαλονίκης, τμ. Πληροφορικής
- Κιτμερίδης, Ν. (2013) *Μελέτη προσεγγίσεων ομαδοποίησης μεγάλων συνόλων αρχείων κειμένου που προέρχονται από πηγές του Παγκόσμιου Ιστού σε κατακευματωμένα υπολογιστικά περιβάλλοντα*. Πτυχιακή Εργασία, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, τμ. Πληροφορικής
- Κόλλιας, Δ. (2015) *Τεχνικές Βαθιάς Μηχανικής Μάθησης και Γνώσης για Ανάλυση Συναισθήματος στην Αλληλεπίδραση Ανθρώπου Μηχανής*. Διπλωματική Εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, τμ. Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
- Κοντοπούλου, Ε. και Γαλλόπουλος, Ε. (2015) *Επιστημονικός Υπολογισμός*. Σημειώσεις Μαθήματος, Πανεπιστήμιο Πατρών, τμ. Μηχανικών Η/Υ και Πληροφορικής

- Κουρής, Γ. Ν. (2006) *Εφαρμογή Τεχνικών Data Mining σε Συστήματα Ηλεκτρονικού Εμπορίου*. Διδακτορική Διατριβή, Πανεπιστήμιο Πατρών, τμ. Μηχανικών Η/Υ και Πληροφορικής
- Κυρίτσης, Κ. (2014) *Νευρωνικά Δίκτυα και Μηχανές Διανυσματικής Υποστήριξης*. Διπλωματική Εργασία, Πανεπιστήμιο Πατρών, τμ. Μαθηματικών
- Ρεφανίδης, Ι. (2011) *Νευρωνικά Δίκτυα*. Σημειώσεις μαθήματος, Πανεπιστήμιο Μακεδονίας, τμ. Εφαρμοσμένης Πληροφορικής
- Σεϊντής, Ι και Πούλιος, Ν. (2014) *Εφαρμογή τεχνητών νευρωνικών δικτύων στην διαχείριση χαρτοφυλακίου*. Πτυχιακή εργασία, Τ.Ε.Ι. Δυτικής Ελλάδας, τμ. Διοίκησης Επιχειρήσεων
- Ταμπάκης, Π. (2004) *Τεχνητά νευρωνικά δίκτυα και πρόβλεψη οικονομικών χρονοσειρών*. Πτυχιακή εργασία, Τ.Ε.Ι. Πατρών, τμ. Επιχειρηματικού Σχεδιασμού και Πληροφοριακών Συστημάτων

## Ξένη Βιβλιογραφία

- Addioui, A., Benabbou, F., El Filali, M. and El Aroussi, M. (2016) Breast cancer mammography diagnosis approach using dual tree complex wavelet transform and artificial neural networks. *International Journal of Imaging and Robotics*, Vol.14, No.4, pp. 62-68
- Agrawal, R., Imieliński, T. and Swami, A. (1993) Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*, p. 207
- Agrawal, R. and Srikant, R. (1994) Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, pp. 487-499
- Al-Allaf, O.N.A. (2015) Removing Noise from Speech Signals Using Different Approaches of Artificial Neural Networks. *I.J. Information Technology and Computer Science*, Vol.07, pp. 8-18
- Alpaydin, E. (2010) *Introduction to Machine Learning*. United Kingdom: MIT Press
- Angelov, P. and Sperduti, A., (2016) Challenges in deep learning. *Proceedings of the 24th European symposium on artificial neural networks (ESANN)*, pp. 489-495
- Anh, D. T. (2011) *Neural Networks in Business Applications*. Course notes, Ho Chi Minh City of Technology, Faculty of Computer Science and Engineering

- Asael, Y. M., Shillingford, B., Whiteson, S. and de Freitas, N. (2016) *LipNet: End-to-end Sentence-level Lipreading*. available online at: <https://arxiv.org/abs/1611.01599>
- Azevedo, F., Carvalho, L., Grinberg, L., Farfel, J., Ferretti, R., Leite, R., Filho, W., Lent, R. and Herculano-Houzel, S. (2009) Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology*, Vol.513, No.5, pp. 532-541
- Badea, L. M. (2014) Predicting Consumer Behavior with Artificial Neural Networks. *Procedia Economics and Finance*, Vol.15, pp. 238-246
- Badri, L. (2009) Development of Neural Networks for Noise Reduction. *The International Arab Journal of Information Technology*, Vol.7, No.3, pp. 289-294
- Bahlmann, C., Heidemann, G. and Ritter, H. (1999) Artificial Neural Networks for Automated Quality Control of Textile Seams. *Pattern Recognition*, Vol.32, No.6, pp. 1049–1060
- Bassis, S., Esposito, A. and Morabito, F. C. (2015) *Advances in Neural Networks: Computational and Theoretical Issues*. Switzerland: Springer International Publishing
- Bell, J. (2015) *Machine Learning: Hands-On for Developers and Technical Professionals*. United States: John Wiley & Sons, Inc.
- Brodley, C. E. and Friedl, M. A. (1999) Identifying and Eliminating Mislabeled Training Instances, *Journal of Artificial Intelligence Research*, Vol.11, pp. 131-167
- Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L.D., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J. and Zieba, K. (2016) End-to-end Learning for Self Driving Cars. available online at: <https://arxiv.org/abs/1604.07316>
- Bose, I. and Agarwal, A. (2002) Prediction of Survival and Attrition of Click-and-Mortar Corporations. in Smith, K. A. and Gupta, J. N. D. (2002) *Neural Networks in Business: Techniques and Applications*. United Kingdom: Idea Group Publishing
- Buhl, L. K., Dekker, J. and Strichartz, G. R. (2012) Principles of Cellular Excitability and Electrochemical Transmission, *Principles of Pharmacology*, Vol.3, pp. 82-92, United States: Lippincott Williams & Wilkins
- Caudill, M. (1987) Neural networks primer, part I. *AI Expert*, Vol.2, No.12, pp. 46-52
- Chapelle, O., Schölkopf, B. and Zien, A. (2006) *Semi-supervised learning*. United States: MIT Press

- Chiarazzo, V., Caggiani, L., Marinelli, M. and Ottomanelli, M. (2014) A Neural Network based model for real estate price estimation considering environmental quality of property location. *Transportation Research Procedia*, Vol.3, pp. 810-817
- Chow, T.W.S. and Cho, S.Y. (2007) *Neural Networks and Computing: Learning Algorithms and Applications*. United Kingdom: Imperial College Press
- Ciresan, D., Meier, U. and Schmidhuber, J. (2012) Multi-column Deep Neural Networks for Image Classification. *Descente du gradient*, Vol.12, pp. 13-18
- Corazza, M., Fasano, G. and Mason, F. (2014) An artificial neural network-based technique for on-line hotel booking. *Procedia Economics and Finance*, Vol.15, pp. 45-55
- Costea, A. (2014) Applying fuzzy logic and machine learning techniques in financial performance predictions, *Procedia Economics and Finance*, Vol.10, pp. 4-9
- Cybenko, G.V. (2006). Approximation by Superpositions of a Sigmoidal function. In van Schuppen, Jan H. *Mathematics of Control, Signals, and Systems*. pp. 303-314, Springer International
- De Nittis, F., Tecchiolli, G. and Zorat, A. (1998) Consumer Loan Classification Using Artificial Neural Networks. *ICSC EIS*, Vol.98
- Dede, G. and Sazli, M.H. (2010) Speech Recognition with Artificial Neural Networks. *Digital Signal Processing*, Vol.20, pp. 3642-3649
- Downey, A.B. (2011) *Physical Modeling in MATLAB*. United States: Green Tea Press
- Duda, R.O. and Hart, P.E. (1973) *Pattern classification and scene analysis*. United States: John Wiley & Sons
- Estévez, P. A., Held, C. M. and Perez, C. A. (2006) Subscription fraud prevention in telecommunications using fuzzy rules and neural networks. *Expert Systems with Applications*, Vol.31, No.2, pp. 337-344
- Fausett, L. (1994) *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. United States: Prentice-Hall International Inc.
- Fawcett, T., Haimowitz, I., Provost, F. and Stolfo, S. (1998) AI approaches to fraud detection and risk management. *AI Magazine*, Vol.19, No.2, p.107
- Feinberg, A (2012) *Neurons and electrochemical transmission*. Available online at: <https://sites.google.com/site/abrahamfeinberg/psych-101-example>



- Fitzpatrick, J. M. and Crocetti, J. D. (2011) *Introduction to Programming with MATLAB*. Course notes, Vanderbilt University, Electrical Engineering and Computer Science dept.
- Gales, M. (2015) *Module 4F10: Statistical Pattern Processing*, Course notes, University of Cambridge, Engineering dept.
- Garro, B.A., Rodriguez, K. and Vasquez, R.A. (2016) Classification of DNA microarrays using artificial neural networks and ABC algorithms. *Applied Soft Computing*, Vol.38, pp. 548-560
- Gatys, L. A. Ecker, A. S. and Bethge, M. (2016) Image Style Transfer Using Convolutional Neural Networks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition - CVPR 2016*, pp. 2414-2123
- Geman, S., Bienenstock, E. and Doursat, R. (1992) Neural networks and the bias/variance dilemma. *Neural Computation*, Vol.4, pp.1-58
- Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. United States: MIT Press. Available online at: <http://www.deeplearningbook.org>
- Grigoryan, H. (2015) Stock Market Prediction using Artificial Neural Networks. Case Study of TALIT, Nasdaq OMX Baltic Stock. *Database Systems Journal BOARD*, Vol.14
- Gul, M. and Guneri, A.F. (2015) Forecasting Patient Length of Stay in an Emergency Department by Artificial Neural Networks. *Journal of Aeronautics and Space Technologies*, Vol.8, No.2, pp. 43-48
- Hahnloser, R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., Seung, H.S. (2000) Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, Vol.405, pp. 947–951
- Han, J., Pei, J., Yin, Y. and Mao, R. (2004) Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. *Data Mining and Knowledge Discovery*, Vol.8, pp. 53-87
- Hartree, D. (1949) *Calculating Instruments and Machines*. United States: University of Illinois Press
- Hassibi, B. Stork, D. G., Wolff, G. and Watanabe, T. (1993) Optimal brain surgeon: extensions and performance comparisons. *Proceedings of the 6th International Conference on Neural Information Processing Systems (NIPS'93)*, pp. 263-270.
- Hebb, D. O. (1949). *The Organization of Behavior*. United States: Wiley & Sons.

- Herbrich, R., Keilbach, M., Graepel, T., Bollmann-Sdorra, P. and Obermayer, K. (1999) Neural networks in economics. *Computational Techniques for Modelling Learning in Economics*. United States: Springer Science + Business Media, pp. 169–196
- Hinton, G. E. and Salakhutdinov, R. R. (2006) Reducing the Dimensionality of Data with Neural Networks, *Science*, Vol.313, pp. 504-507
- Hinton, G. E., Osindero, S. and Teh, Y.W. (2006) A fast learning algorithm for deep belief nets. *Neural Computation*, Vol.18, No.7, pp. 1527-1554
- Hodge, V. J. and Austin, J. (2004) A survey of outlier detection methodologies. *Artificial Intelligence Review*, Vol.22, No.2, pp. 85-126
- Hollis, D.J., Edgington, D. and Cline, D. (2006) Automated Detection of Deep-Sea Animals. *STEM Teacher and Research Presentations*, Vol.370
- Hřebíček, J., Soukopová, J. and Trenz, O. (2014) Current trends of economic Modelling of sustainable corporate performance and reporting – review and research agenda. *Procedia Economics and Finance*, Vol.12, pp. 234–242
- Huang, G., Liu, Z., van der Maaten, L. and Weinberger, K. (2017) Densely Connected Convolutional Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR Oral) 2017*
- Hume, D. (1748) *Philosophical essays concerning human understanding*. United Kingdom: A. Millar
- Iizuka, S., Simo-Serra, E. and Ishikawa, H (2016) Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorisation with simultaneous classification. *ACM Transactions on Graphics (TOG)*, Vol.35, No.4
- Ip, E., Johnson, J., Yada, K., Hamuro, Y., Katoh, N. and Cheung, S. (2002) A Neural Network Application to Identify High-Value Customers for a Large Retail Store in Japan. in Smith, K. A. and Gupta, J. N. D. (2002) *Neural Networks in Business: Techniques and Applications*. United Kingdom: Idea Group Publishing
- Jafari, M.H., Nasr-Esfahari, E., Karim, N., Reza Sordushmehr, S.M., Samavi, S. and Najarian, K. (2017) Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma. *International Journal of Computer Assisted Radiology and Surgery*. Vol.12, No.6, pp. 1021-1030
- James, G. (2003) Variance and Bias for General Loss Functions. *Machine Learning*, Vol.51, pp. 115-135

- Jones, S., Johnstone, D. and Wilson, R. (2017) Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks. *Journal of Business Finance & Accounting*, Vol.44, pp. 3–34
- Kanellis, N. and Papadopoulos, D. (2017) Identification of potentially undelivered packages with an artificial neural network method. *Proceedings of the 5th International Conference on Contemporary Marketing Issues (ICCM) 2017*, pp. 314-319
- Kar, A. (1990) Stock prediction using artificial neural networks. available online at: [https://people.eecs.berkeley.edu/~akar/IITK\\_website/EE671/report\\_stock.pdf](https://people.eecs.berkeley.edu/~akar/IITK_website/EE671/report_stock.pdf)
- Kar, A.K., De, S. K. (2009) Using neural networks for pattern association for the online purchase of products. *Sprouts: Working Papers on Information Systems*, Vol.9, No.27
- Kasar, M. M., Bhattacharyya, D. and Kim, T. (2016) Face Recognition Using Neural Network: A Review. *International Journal of Security and its applications*, Vol.10, No.3, pp.81-100
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) ImageNet classification with deep convolutional neural networks. *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 1097 - 1105
- Kubat, M. (2015) *An Introduction to Machine Learning*. Switzerland: Springer International Publishing
- Kim, K. and Han, I. (2003) Application of a hybrid genetic algorithm and neural network approach in activity-based costing, *Expert Systems with Applications*, Vol.24, pp. 73-77
- Langtangen, H.P. (2014) *A Primer on Scientific Programming with Python*. Germany: Springer-Verlag Berlin and Heidelberg GmbH & Co. K.
- LeCun, Y., Bengio, Y. and Hinton, G. E. (2015) Deep Learning, *Nature*, Vol.521, pp 436-444
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. and Jackel, L. D. (1990) Handwritten digit recognition with a back-propagation network, *Advances in Neural Information Processing Systems (NIPS 1989)*, Vol.2
- LeCun, Y., Bottu, L., Bengio, Y. and Haffner, P. (1998) Gradient-Based Learning Applied to Document Recognition, *Proceedings of the IEEE*, Vol.86, No.11, pp. 2278-2324

- LeCun, Y., Denker, J. S. and Solla, S. A. (1990) Optimal brain damage. *Advances in neural information processing systems*, Vol.2
- Li, E. (1994) Artificial neural networks and their business applications. *Information & Management*, Vol.27, No.5, pp.303-313
- Liu, D. Z. and Singh, G. (2015) A Recurrent Neural Network Based Recommendation System. available online at: <https://cs224d.stanford.edu/reports/LiuSingh.pdf>
- Lovelace, A. (1842) Translator's notes to an article on Babbage's Analytical Engine. *Scientific Memoir* (ed. by R. Taylor), Vol.3, pp. 691-731
- MacQueen, J.B. (1967) Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281-297
- Markram, H., Lubke, J., Frotscher, M. and Sakmann, B. (1997) Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science*, Vol.275, pp. 213-215
- McCulloch, W.S. and Pitts, W.H. (1943) A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Vol.5, No.4, pp. 115-133
- Medler, D.A. (1998) A Brief History of Connectionism. *Neural Computing Surveys*, Vol.1, pp. 61-101
- Minsky, M. and Papert, S. (1969) *Perceptrons: An Introduction to Computational Geometry*. United States: MIT Press
- Mitchell, T. M. (1997) *Machine Learning*. 7th edn. United States: McGraw Hill Higher Education
- Nagesha, K.V., Ram Chandler, K. and Sastry, V. R. (2015) Prediction of Dust Dispersion by Drilling Operation Using Artificial Neural Networks. *International Journal of Prevention and Control of Industrial Pollution*, Vol.1, No.2
- Ng, A. (2011) *Machine Learning*. Course notes, Stanford University, Computer Science dept.
- Nielsen, M.A. (2015) *Neural Networks and Deep Learning*. United States: Determination Press
- Nilsson, N.J. (1965) *Learning machines*. United States: McGraw-Hill Inc.
- Nilsson, N.J. (1990) *The mathematical foundations of learning machines*. United States: Morgan Kaufmann Publishers Inc.

- Nilsson, N.J. (1998) *Introduction to Machine Learning*. Course notes, Stanford University, Computer Science dept. Available online at: <http://ai.stanford.edu/~nilsson/mlbook.html>
- Omar, S., Ngadi, A. and Jebur, H. H. (2013) Machine Learning Techniques for Anomaly Detection: An Overview, *International Journal of Computer Applications*, Vol.79, No.2, pp. 33 - 41
- Papadopoulou, P., Kanellis, P. and Martakos, D. (2001) Investigating trust in e-commerce: A literature review and a model for its formation in customer relationships. *Proceedings of the Seventh Americas Conference on Information Systems*, pp. 791-798
- Petroulakis, N. and Miaoudakis, A. (2007) An Application of Neural Networks in Market Segmentation, *Proceedings of the PanHellenic Conference in New Technologies and Marketing, Ierapetra, 24-25 May 2007*, pp. 185 - 190
- Relich, M. and Muszyński, W. (2014) The Use of Intelligent Systems for Planning and Scheduling of Product Development Projects. *Procedia Computer Science*, Vol.35, pp.1586-1595
- Rogers, S. K. and Kabrisky, M. (1991) *An introduction to biological and artificial neural networks for pattern recognition*. United States: SPIE Optical Engineering Press
- Roghani, A. (2015) *Artificial Neural Networks: Applications in Financial Forecasting*. United Kingdom: CreateSpace Independent Publishing Platform
- Rosenblatt, F. (1958) The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, Vol.65, No.6, pp. 386-408
- Safa, N.S., Ghani, N.A. and Ismail, M.A. (2014) An artificial neural network classification approach for improving accuracy of customer identification in e-commerce. *Malaysian Journal of Computer Science*, Vol.27, No.3, pp. 171-185
- Samuel, A. (1959) Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, Vol.3, No.3, pp. 210-229.
- Serrano-Cinca, C. (1996) Self organizing neural networks for financial diagnosis. *Decision Support Systems*, Vol.17, No.3, pp. 227-238
- SINTEF (2013) *Big data, for better or worse: 90% of world's data generated over last two years*. available online at: <https://www.sciencedaily.com/releases/2013/05/130522085217.htm>

- Sharma, J., Angelucci, A. and Sur, M. (2000) Induction of visual orientation modules in auditory cortex. *Nature*, Vol.404, pp. 841-847
- Shiri, J., Shanshirband, S., Kisi, O., Karimi, S. and Bateni, S.M. (2016) Prediction of Water Level in the Urmia Lake using the Extreme Learning Machine Approach. *Water Resources Management*, Vol.30, No.14, pp. 5217-5229
- Smith, K. A. and Gupta, J. N. D. (2002) *Neural Networks in Business: Techniques and Applications*. United Kingdom: Idea Group Publishing
- Smith, M. R. and Martinez, T. (2011) Improving classification accuracy by identifying and removing instances that should be misclassified. *The 2011 International Joint Conference on Neural Networks*, pp. 2690 - 2697
- Smola, A. and Vishwanathan, S.V.N. (2008) *Introduction to Machine Learning*. United Kingdom: Cambridge University Press
- Snyman, Jan (2005) *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. United States: Springer Science & Business Media
- Stathakis, D. (2009) How many hidden layers and nodes? *International Journal of Remote Sensing*, Vol.30, No.8, pp. 2133-2147
- Suryanita, R., Jingga, H., and Yuniarto, F. (2016) The application of Artificial Neural Networks in Predicting Structural Response of Multistory Building in the Region of Sumatra Island. *ICoSE Conference on Instrumentation, Environment and Renewable Energy*
- Taylor, M. M. (1973) The problem of stimulus structure in the behavioural theory of perception. *South African Journal of Psychology*, Vol.3, pp. 23-45
- Turing, A.M. (1950) Computing machinery and intelligence. *MIND (the Journal of the Mind Association)*, Vol.LIX, No.236, pp. 433-460
- Vapnik, V. (1999) An Overview of Statistical Learning Theory. *IEEE Transactions on Neural Networks*, Vol.10, No.5, pp. 988-999
- Vinyals, O., Ewals, T., Bartunov, S., Georgiev, P., Vezhnevets, A. S., Yeo, M., Makhzani, A., Kuttler, H., Agapiou, J., Schrittwieser, J., Gaffney, S., Petersen, S., Simonyan, K., Schaul, T., van Hasselt, H., Silver, D., Lillicrap, T., Calderone, K., Keet, P., Brunasso, A., Lawrence, D., Ekermo, A., Repp, J. and Tsing, R. (2017) *StarCraft II: A New Challenge for Reinforcement Learning*. Available online at: <https://deepmind.com/blog/deepmind-and-blizzard-open-starcraft-ii-ai-research-environment/>

- West, D. and Muchineuta, C. (2002) Credit Scoring Using Supervised and Unsupervised Neural Networks. in Smith, K. A. and Gupta, J. N. D. (2002) *Neural Networks in Business: Techniques and Applications*. United Kingdom: Idea Group Publishing
- Widrow, B and Hoff, M. E. Jr (1960) Adaptive Switching Circuits. *IRE WESCON Convention Record, part 4, pp. 96-104*
- Witten, I. H., Frank, E. and Hall, M.A. (2011) *Data mining: Practical machine learning tools and techniques*. 3rd edn. United States: Morgan Kaufmann Publishers
- Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnic, A., Vinyals, O., Corrado, G., Hughes, M. and Dean, J. (2016) *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. Available online at: <https://arxiv.org/abs/1609.08144>
- Yao, J. and Tan, C. L. (2002) Neural Networks for Technical Forecasting of Foreign Exchange Rates. in Smith, K. A. and Gupta, J. N. D. (2002) *Neural Networks in Business: Techniques and Applications*. United Kingdom: Idea Group Publishing
- Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, Vol.12, No.3, pp. 372-390
- Zhang, G.P. and Qi, M. (2002) Predicting Consumer Retail Sales Using Neural Networks. in Smith, K. A. and Gupta, J. N. D. (2002) *Neural Networks in Business: Techniques and Applications*. United Kingdom: Idea Group Publishing
- Zhang, J., Wang J. and Qin, W. (2016). Artificial Neural Networks in Production Scheduling and Yield Prediction of Semiconductor Wafer Fabrication System. available online at:  
<https://www.intechopen.com/books/artificial-neural-networks-models-and-applications/artificial-neural-networks-in-production-scheduling-and-yield-prediction-of-semiconductor-wafer-fabr>

## ΠΑΡΑΡΤΗΜΑ

Στις παρακάτω εικόνες δίνεται ένα παράδειγμα της μορφής που έχουν τα δεδομένα εισόδου της προτεινόμενης εφαρμογής, πριν και μετά την μεταγραφή τους σε μορφή αξιοποιήσιμη από το νευρωνικό δίκτυο. Και στις δύο εικόνες, αυτό που βλέπουμε είναι μόνο ένα κομμάτι της βάσης δεδομένων: οι μεταβλητές στην πρώτη περίπτωση είναι δεκάδες, στην δεύτερη εκατοντάδες (καθώς κάποιες κατηγορικές μεταβλητές αναλύθηκαν σε επιμέρους *dummy variables*), ενώ οι εγγραφές μπορεί να είναι χιλιάδες, ιδιαίτερα αν μιλάμε για μια επιχείρηση με μεγάλο ετήσιο τζίρο.

	A	B	C	D	E	F	G	H
1	insertion_date	payment_status_name	order-id	SKU	price	shipping-fee	total-price	ship-country
2	2015-03-19 15:06:35	Replacement Sent	9596586411011	PS3G3157	14.99	0.49	15.48	Bolivia
3	2015-03-19 15:06:35	Replacement Sent	9596586411011	PS3G3157	14.99	0.49	15.48	Bolivia
4	2015-04-26 14:21:36	Despatched	6296361921011	TY3109	29.99	1.99	31.98	Estonia
5	2015-04-26 14:21:36	Despatched	6296361921011	BOG1887	37.99	1.99	39.98	Estonia
6	2015-04-26 15:05:26	Replacement Sent	9596361921011	PS3G1298	16.99	0.99	17.98	Germany
7	2015-04-26 16:55:33	Despatched	1717371921011	PS4A0005	8.99	0	8.99	United Kingdom
8	2015-04-27 22:48:32	Despatched	9574581921011	CL6222	9.99	0.99	10.98	United Kingdom
9	2015-04-27 22:48:32	Replacement Sent	9574581921011	CL9165	9.99	0.99	10.98	United Kingdom
10	2015-04-28 11:16:36	Despatched	6218591921011	CAG0954	15.99	0	15.99	United Kingdom
11	2015-04-28 14:12:29	Refunded	8441691921011	PS3C0041	128.69	0	128.69	United Kingdom
12	2015-04-28 14:12:29	Refunded	8441691921011	PS3A0444	7.49	0	7.49	United Kingdom
13	2015-04-28 14:56:33	Despatched	7352691921011	360G1305	6.99	1.99	8.98	Italy
14	2015-04-28 14:56:33	Despatched	7352691921011	360G1723	6.99	1.99	8.98	Italy
15	2015-04-28 19:09:28	Despatched	8485691921011	COL3163	10.99	0	10.99	United Kingdom
16	2015-04-28 19:09:28	Despatched	8485691921011	COL3164	10.99	0	10.99	United Kingdom
17	2015-04-28 19:09:28	Despatched	8485691921011	COL3165	10.99	0	10.99	United Kingdom
18	2015-04-28 19:42:28	Refunded	2895691921011	PS4G0225	44.99	0	44.99	United Kingdom
19	2015-04-28 19:42:28	Refunded	2895691921011	PS4G0225	44.99	0	44.99	United Kingdom
20	2015-04-28 19:42:28	Refunded	2895691921011	PS4A0086	60.99	0	60.99	United Kingdom
21	2015-04-28 20:37:27	Despatched	8496691921011	PCG0608	6.99	0	6.99	United Kingdom
22	2015-04-28 20:37:27	Despatched	8496691921011	PCG0766	9.99	0	9.99	United Kingdom
23	2015-04-28 20:37:27	Despatched	8496691921011	PCG1175	9.49	0	9.49	United Kingdom

Στην πρώτη εικόνα, βλέπουμε τη μορφή που έχουν τα ιστορικά δεδομένα παραγγελιών μιας τυπικής επιχείρησης ηλεκτρονικού εμπορίου. Παρατηρούμε την ημερομηνία, την κατάσταση της παραγγελίας απ' όπου μπορούμε να εξάγουμε το αν έφτασε ή δεν έφτασε στον παραλήπτη, το Stock Keeping Unit που είναι ο αριθμός ταυτοποίησης του προϊόντος και απο εκεί μπορούμε να εξάγουμε την κατηγορία στην οποία ανήκει, την τιμή, τα έξοδα αποστολής, την χώρα του παραλήπτη κ.α.



	A	B	C	D	E	F	G	H	I	J	K	L
1	id	price	month	Argentina	Australia	Austria	Belarus	Belgium	Board Games	Books	Card Games	Collectibles
2	1	1	7	0	0	0	0	0	0	0	0	0
3	2	2	11	0	1	0	0	0	0	0	0	0
4	3	2	11	0	0	0	0	0	0	0	0	0
5	4	2	11	0	0	1	0	0	0	0	0	0
6	5	2	11	0	0	0	0	0	0	0	0	0
7	6	2	11	1	0	0	0	0	0	0	0	0
8	7	2	11	0	0	0	0	0	0	1	0	0
9	8	2	11	0	0	0	0	0	0	0	1	0
10	9	2	11	0	0	0	0	0	0	0	0	1
11	10	2	11	0	0	0	0	0	0	0	0	0
12	11	2	11	0	0	0	0	0	0	0	0	0
13	12	2	11	0	0	0	0	0	0	0	0	0
14	13	2	11	0	0	0	0	0	0	0	0	0
15	14	2	11	0	0	0	0	0	0	0	0	0
16	15	2	11	0	0	0	0	0	0	0	0	0
17	16	2	11	0	0	0	0	0	0	0	0	0
18	17	2	11	0	0	0	0	0	0	0	0	0
19	18	2	11	0	0	0	0	0	0	0	0	0
20	19	2	11	0	0	0	0	0	0	0	0	0
21	20	2	11	0	0	0	0	0	0	0	0	0
22	21	2	11	0	0	0	0	0	0	0	0	0
23	22	2	11	0	0	0	0	0	0	0	0	0
24	23	2	11	0	0	0	0	0	0	0	0	0

Στην δεύτερη εικόνα βλέπουμε ένα παράδειγμα των επεξεργασμένων δεδομένων, όπως αυτά θα εισαχθούν στο νευρωνικό δίκτυο. Παρατηρούμε ότι έχει εξαχθεί ο μήνας που έγινε η παραγγελία, καθώς και τα dummy variables που προέκυψαν από την ανάλυση των κατηγορικών μεταβλητών “χώρα παραλήπτη” και “κατηγορία εμπορεύματος”.