



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΙΩΑΝΝΙΝΩΝ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΕΚΠΑΙΔΕΥΣΗ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΜΕ ΣΤΟΧΑΣΤΙΚΕΣ ΤΕΧΝΙΚΕΣ

Παρασκευή Τοκμακίδου

Επιβλέπων: Ιωάννης Γ. Τσούλος
Καθηγητής

Άρτα, Οκτώβριος, 2024

TRAINING OF ARTIFICIAL NEURAL NETWORKS WITH STOCHASTIC TECHNIQUES

Εγκρίθηκε από τριμελή εξεταστική επιτροπή

Άρτα, Οκτώβριος 2024

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Επιβλέπων καθηγητής

Ιωάννης Γ. Τσούλος

2. Μέλος επιτροπής

Βασίλειος Χαριλόγης

3. Μέλος επιτροπής

Νικόλαος Γιαννακέας

© Τοκμακίδου, Παρασκευή, 2024.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Δήλωση μη λογοκλοπής

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί Πνευματικής Ιδιοκτησίας, ότι η παρούσα πτυχιακή εργασία είναι εξ ολοκλήρου αποτέλεσμα δικής μου ερευνητικής εργασίας, δεν αποτελεί προϊόν αντιγραφής ούτε προέρχεται από ανάθεση σε τρίτους. Όλες οι πηγές που χρησιμοποιήθηκαν (κάθε είδους, μορφής και προέλευσης) για τη συγγραφή της περιλαμβάνονται στη βιβλιογραφία.

Τοκμακίδου, Παρασκευή

Υπογραφή

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Ιωάννη Τσούλο, για την πολύτιμη καθοδήγηση, τις συμβουλές και την υπομονή καθ' όλη τη διάρκεια όχι μόνο της εκπόνησης της πτυχιακής εργασίας αλλά και των σπουδών μου. Αποτέλεσε σημαντικό παράγοντα για την εξέλιξη μου στον τομέα. Επιπλέον, θα ήθελα να ευχαριστήσω όλους τους καθηγητές της σχολής που ο καθένας με τον δικό του τρόπο, συνέβαλε όχι μόνο στην απόκτηση πολύτιμων γνώσεων αλλά και στη συλλογή εμπειριών, καθοριστικές για την εξέλιξη μου. Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την εμπιστοσύνη που μου έδειξαν σε κάθε επιλογή μου, τον Β. Α. που ήταν δίπλα μου σε κάθε εμπόδιο, καθώς και τους φίλους μου για την υποστήριξη τους.

ΠΕΡΙΛΗΨΗ

Ο νευρώνας είναι το πιο σημαντικό στοιχείο ενός εγκεφάλου και είναι υπεύθυνο για ευφυείς διαδικασίες. Γεγονός που αποτέλεσε έμπνευση για τη ανάπτυξη δικτύων με σκοπό την προσομοίωση τέτοιων διαδικασιών, με κυριότερο εκπρόσωπο, την διαδικασία της μάθησης. Σκοπός αυτής της πτυχιακής εργασίας είναι η κατανόηση των βασικών εννοιών του κλάδου των ΤΝΔ καθώς και η μελέτη κάποιων βασικών στοχαστικών αλγορίθμων με έμφαση στο μοντέλο και όχι στη μαθηματική ανάλυση.

Στο πρώτο κεφάλαιο περιγράφονται κάποιες θεμελιώδεις έννοιες, απαραίτητες για την κατανόηση των ΤΝΔ. Στο δεύτερο κεφάλαιο, αναλύεται ο τρόπος λειτουργίας κάποιων ΤΝΔ με επίβλεψη, με αφετηρία το πιο ιστορικό δίκτυο, το Perceptron του Rosenblatt. Στη συνέχεια, αναλύεται το Adaline δίκτυο και έπειτα το δίκτυο Multi-Layer Perceptron. Συνδυαστικά με το τελευταίο δίκτυο, περιγράφονται δύο βασικοί μέθοδοι για την εκπαίδευση του: Back Propagation και Gradient Descent. Στο τρίτο κεφάλαιο, αναλύονται οι γενετικοί αλγόριθμοι. Ένα άλλο είδος αλγορίθμων, οι οποίοι έχουν την δυνατότητα βελτιστοποίησης των ΤΝΔ.

Παράλληλα με το κείμενο της πτυχιακής εργασίας, έγινε υλοποίηση μίας εφαρμογής σε γλώσσα JAVA στην οποία υπάρχει η δυνατότητα εκπαίδευσης ενός δικτύου MLP με τις μεθόδους Back Propagation και Gradient Descent αλλά και με τη αξιοποίηση ενός γενετικού αλγορίθμου με σκοπό την βέλτιστη αρχικοποίηση των βαρών του MLP δικτύου. Σε αυτή την εφαρμογή υπάρχουν ήδη έτοιμα κάποια σύνολα δεδομένων για δοκιμή, αλλά υπάρχει και η δυνατότητα να φορτώσει ο χρήστης κάποιο που επιθυμεί, με την προϋπόθεση ότι πληρεί κάποιες προϋποθέσεις. Η υλοποίηση είναι διαθέσιμη στον ιστότοπο <https://github.com/Paraskevi-Tokmakidou/artificial-neural-networks>.

Λέξεις-κλειδιά: Τεχνητά Νευρωνικά Δίκτυα, Γενετικοί Αλγόριθμοι, Στοχαστικές μέθοδοι

ABSTRACT

The neuron is the most important component of the brain and is responsible for intelligent processes. This fact was an inspiration for the development of networks to simulate such processes, with the main representative being the learning process. The purpose of this thesis is to understand the basic concepts of the ANN branch and to study some basic stochastic algorithms with an emphasis on the model and not on the mathematical analysis.

In the first chapter, a description of fundamental concepts necessary for the basic understanding of ANN is given. In the second chapter, the operational mode of some supervised ANN is analyzed, starting with the most historical network, Rosenblatt's Perceptron. Subsequently, the Adaline network and the Multi-Layer Perceptron are examined. Along with the latter network, two basic methods for its training are described, Back Propagation and Gradient Descent. In the third chapter, genetic algorithms are described - another type of algorithm that has the potential of optimising ANN.

In conjunction with the thesis text, an application implemented with Java language, which includes the ability to train an MLP network using the Back Propagation and Gradient Descent methods, as well as with the utilization of a genetic algorithm with the purpose of optimizing the initial weights of the MLP network. In this application, some datasets are already prepared for testing, but the user can also load any file, provided it meets specific conditions. The implementation is available on the website <https://github.com/Paraskevi-Tokmakidou/artificial-neural-networks>.

Keywords: Artificial Neural Networks, Genetic Algorithms, Stochastic techniques

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΥΧΑΡΙΣΤΙΕΣ	vi
ΠΕΡΙΛΗΨΗ	vii
ABSTRACT	viii
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	ix
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	xi
ΚΑΤΑΛΟΓΟΣ ΜΑΘΗΜΑΤΙΚΩΝ ΤΥΠΩΝ	xii
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	xiii
ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ	xiv
ΓΛΩΣΣΑΡΙΟ	xv
1. Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα	1
1.1 Ιστορική Αναδρομή	1
1.2 Μετάβαση από τους βιολογικούς στους τεχνητούς νευρώνες	2
1.3 Συναρτήσεις ενεργοποίησης	5
1.4 Τεχνητά Νευρωνικά Δίκτυα: Μία σφαιρική εικόνα	8
1.5 Μάθηση με επίβλεψη	10
2. Δομή και Λειτουργία των Τεχνητών Νευρωνικών Δικτύων	12
2.1 Δίκτυα Perceptron	12
2.2 Δίκτυα Adaline	15
2.3 Δίκτυα MLP	17
2.4 Μέθοδος Back Propagation	21
2.5 Μέθοδος Gradient Descent	22
2.6 Εφαρμογές νευρωνικών δικτύων	23
3. Γενετικοί αλγόριθμοι	25

3.1 Ιστορική Αναδρομή	25
3.2 Δομή και Λειτουργία των Γενετικών Αλγορίθμων	26
3.3 Μέθοδοι κωδικοποίησης	27
3.4 Συνάρτηση καταλληλότητας	28
3.5 Γενετικοί τελεστές	29
3.6 Εφαρμογές γενετικών αλγορίθμων	32
4. Μέθοδος - αποτελέσματα	34
4.1 Τα dataset που χρησιμοποιήθηκαν	34
4.2 Η προτεινόμενη μέθοδος	35
4.3 Πειραματικά αποτελέσματα	36
5. Συμπεράσματα	39
BIBΛΙΟΓΡΑΦΙΑ	40
ΠΑΡΑΡΤΗΜΑ	44
ΠΑΡΑΡΤΗΜΑ Α: Κώδικας σε Octave για δημιουργία γραφικών παραστάσεων	44
ΠΑΡΑΡΤΗΜΑ Β: Αλγόριθμοι με την μορφή ψευδοκώδικα	46

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1 Αναπαράσταση βιολογικού νευρώνα (Βλαχάβας, 2013)	3
Εικόνα 2 Αναπαράσταση τεχνητού νευρώνα (Βλαχάβας, 2013)	5
Εικόνα 3 Γραφική παράσταση της γραμμικής συνάρτησης	6
Εικόνα 4 Γραφική παράσταση της βηματικής συνάρτησης 0/1	6
Εικόνα 5 Γραφική παράσταση της βηματικής συνάρτησης -1/1	7
Εικόνα 6 Γραφική παράσταση της σιγμοειδής συνάρτησης	7
Εικόνα 7 Σχηματική απεικόνιση ενός απλού ΤΝΔ	9
Εικόνα 8 Perceptron	13
Εικόνα 9 MLP	19
Εικόνα 10 Single point crossover	30
Εικόνα 11 Double point crossover	31

ΚΑΤΑΛΟΓΟΣ ΜΑΘΗΜΑΤΙΚΩΝ ΤΥΠΩΝ

Τύπος 1 Υπολογισμός πλήθους βαρών για ένα Perceptron δίκτυο	14
Τύπος 2 Ενημέρωση βαρών του δικτύου Perceptron	15
Τύπος 3 Τετραγωνικό σφάλμα	17
Τύπος 4 Υπολογισμός πλήθους βαρών για ένα MLP δίκτυο	19
Τύπος 5 Υπολογισμός εξόδου ενός MLP δικτύου (Τσούλος, Τζάλλας & Τσαλικάκης, 2018)	20
Τύπος 6 Ενημέρωση των βαρών με χρήση Gradient Descent	23
Τύπος 7 Μέσο Τετραγωνικό Σφάλμα	29

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1	Πειραματικά αποτελέσματα για το σύνολο δεδομένων της ιονόσφαιρας	37
Πίνακας 2	Πειραματικά αποτελέσματα για το σύνολο δεδομένων των κρασιών	38
Πίνακας 3	Πειραματικά αποτελέσματα για το σύνολο δεδομένων των λουλουδιών	39

ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ

TNΔ	Τεχνητά Νευρωνικά Δίκτυα
ANN	Artificial Neural Networks
MLP	Μοντέλο Perceptron πολλών στρωμάτων
LMS	Least Mean Squares

ΓΛΩΣΣΑΡΙΟ

On-Line Αλγόριθμος Μάθησης

Στοχαστική μέθοδος κατά την οποία η εκπαίδευση των βαρών γίνεται με κάθε πέρασμα ενός προτύπου και όχι μετά την ολοκλήρωση μίας ολόκληρης εποχής.

ΠΡΟΤΥΠΑ

Το σύνολο των δεδομένων εισόδου.

ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΠΡΟΤΥΠΟΥ

Κάθε τιμή που αναπαριστά μία ιδιότητα.

ΜΗ ΓΡΑΜΜΙΚΑ ΔΙΑΧΩΡΙΣΙΜΑ ΠΡΟΤΥΠΑ

Πρότυπα που δεν μπορούν να διαχωριστούν με μία ευθεία γραμμή κατά την αναπαράστασή τους.

ΕΠΟΧΗ

Ένας πλήρης κύκλος χρήσης όλων των προτύπων.

ΚΛΕΙΣΤΟ ΔΙΑΣΤΗΜΑ

Το κλειστό διάστημα $[α,β]$ μίας μεταβλητής x σημαίνει πως $α \leq x \leq β$.

ΝΤΕΤΕΡΜΙΝΙΣΤΙΚΟ - ΑΙΤΙΟΚΡΑΤΙΚΟ ΜΟΝΤΕΛΟ

Μοντέλο κατά το οποίο η έξοδος είναι γνωστή για συγκεκριμένη είσοδο καθώς η διαδικασία είναι απόλυτα προβλέψιμη και καθορισμένη, χωρίς καμία τυχαιότητα.

ΣΤΟΧΑΣΤΙΚΟ ΜΟΝΤΕΛΟ

Μοντέλο κατά το οποίο δεν είναι προβλέψιμη η έξοδος για συγκεκριμένη είσοδο, λόγω τυχαιότητας.

1. Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα

1.1 Ιστορική Αναδρομή

Τη δεκαετία του 1940 σημειώθηκε σημαντική πρόοδος στη μελέτη των βιολογικών νευρωνικών δικτύων και στη μαθηματική μοντελοποίησή τους. Η περίοδος των ΤΝΔ ξεκίνησε το 1943 με την ανάπτυξη ενός βασικού μοντέλου τεχνητού νευρώνα, από τους Αμερικανούς επιστήμονες McCulloch και Pitts. Το μοντέλο αυτό συνέδεσε τις επιστήμες της Νευροφυσιολογίας και της Μαθηματικής Ανάλυσης, μιμούμενο τη λειτουργία ενός βιολογικού νευρώνα. Αξιοσημείωτο είναι πως το 1949, ο von Neumann χρησιμοποίησε τα στοιχεία αυτού του τεχνητού νευρώνα για την κατασκευή του EDVAC, του πρώτου υπολογιστή γενικού σκοπού.

Οκτώ χρόνια αργότερα, το 1957, ο ψυχολόγος Frank Rosenblatt ανέπτυξε το δίκτυο Perceptron στο Αεροναυτικό Εργαστήριο του Κορνέλλ (Cornell Aeronautical Laboratory). Αποτελεί μία καινοτόμα μέθοδο μάθησης με επίβλεψη και είναι μία ελαφρώς τροποποιημένη έκδοση του νευρώνα των McCulloch και Pitts. Το δίκτυο αυτό, δημοσιεύτηκε το 1958 και προτάθηκε ως ένας μηχανισμός που μπορεί να εκπαιδευτεί στην κατηγοριοποίηση προτύπων και εξακολουθεί να υφίσταται μέχρι και σήμερα. Η πρώτη εμπεριστατωμένη κριτική του Perceptron παρουσιάστηκε από τους Minsky και Selfridge κατά την οποία έγινε επισήμανση πως αυτό το δίκτυο δεν θα μπορούσε να υποστηρίξει γενικεύσεις. Έτσι, το 1969 στο περίφημο βιβλίο “Perceptrons” των Minsky και Papert, αποδείχθηκε με μαθηματικό τρόπο, πως ΤΝΔ ενός επιπέδου όπως είναι ο Perceptron, δεν μπορούν να λύσουν προβλήματα που αφορούν μη γραμμικά διαχωρίσιμα πρότυπα. Αυτόν τον περιορισμό ήρθε να καταργήσει το δίκτυο Adaline. Αναπτύχθηκε το 1960, από τους Widrow και Hoff και αποτελεί μία τροποποίηση του δικτύου Perceptron. Δίκτυο, που για πρώτη φορά μπορεί να εφαρμοστεί σε μη γραμμικά διαχωρίσιμα πρότυπα με επιτυχία. Έτσι, αποτέλεσε ένα από τα πρώτα δίκτυα που εφαρμόστηκαν σε πρακτικά προβλήματα. Ειδικότερα, χρησιμοποιήθηκε ως φίλτρο για την εξάλειψη της ηχούς σε τηλεφωνικές γραμμές καθώς και για την πρόβλεψη και την ισοστάθμιση καναλιών επικοινωνίας.

Τη δεκαετία του 1970, η έρευνα των ΤΝΔ παρουσίασε ύφεση. Όμως, κατά τη δεκαετία του 1980, σημειώθηκε έκρηξη δραστηριότητας με την εισαγωγή περισσότερων νευρώνων στο δίκτυο Perceptron. Το νέο δίκτυο που προέκυψε ονομάστηκε MLP και προσφέρει ακόμη και σήμερα απεριόριστες δυνατότητες αναπαράστασης συναρτήσεων και διαχωρισμού κλάσεων ξεπερνώντας τους υπολογιστικούς περιορισμούς που είχαν παρουσιαστεί για το Perceptron.

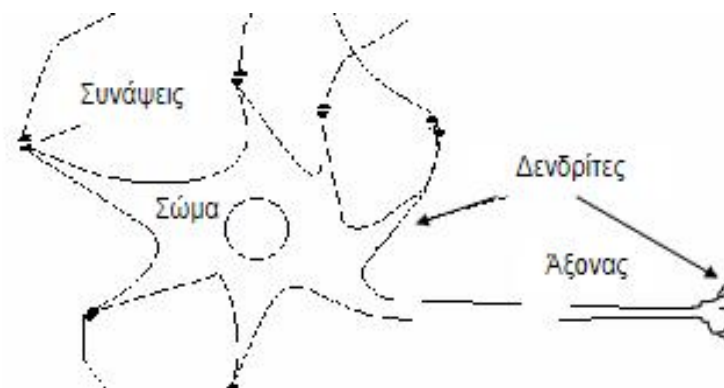
Λίγο αργότερα, οι μέθοδοι Back Propagation και Gradient Descent προτάθηκαν συνδυαστικά με σκοπό την εκπαίδευση των MLP δικτύων. Βιβλιογραφικά, γνωρίζουμε πως η πρόταση της μεθόδου Back Propagation έγινε από τον Paul Werbos. Και οι δύο (2) μέθοδοι αυτοί υπήρχαν ήδη, αλλά δεν είχαν σχεδιαστεί για να θυμίζουν ή να έχουν σχέση με τα ΤΝΔ. Η ενσωμάτωση τους ωστόσο σε αυτά, αποτέλεσε ορόσημο για την εξέλιξη του κλάδου των ΤΝΔ.

Τα ΤΝΔ σίγουρα έχουν διανύσει πολύ δρόμο από την εποχή των McCulloch και Pitts, και έχουν εγκαθιδρυθεί ως διεπιστημονικό αντικείμενο με βαθιές ρίζες στην ψυχολογία, στα μαθηματικά, στις φυσικές επιστήμες και στη μηχανική. (Haykin, 2010)· (Γεωργούλη, 2015)· (Διαμαντάρας, 2007)· (Λυκοθανάσης και Κουτσημητρόπουλος, 2021)· (Ντούνης, 2024)· (Τσίπουρας, 2015-2016)· (Τσούλος, Τεχνητά Νευρωνικά Δίκτυα, χ.χ.)

1.2 Μετάβαση από τους βιολογικούς στους τεχνητούς νευρώνες

Ο ανθρώπινος εγκέφαλος αποτελεί το πιο πολύπλοκο όργανο του ανθρώπινου σώματος και είναι υπεύθυνος για τις πιο σύνθετες και εξελιγμένες λειτουργίες που αυτό εκτελεί όπως η μνήμη, η σκέψη, η ανάλυση και η μάθηση. Λαμβάνει πληροφορίες μέσω των αισθήσεων όπως είναι η όραση και η ακοή και εκπέμπει κινητικές διεγέρσεις για τον συντονισμό των κινήσεων καθώς και τη ρύθμιση των σωματικών λειτουργιών. (Δούνιας και Καραμπότης, 2023) Ο Ramón y Cajal ήταν ο πρώτος που βοήθησε προκειμένου να γίνει πιο εύκολη η κατανόηση του εγκεφάλου, εισάγοντας την ιδέα των νευρώνων ως δομικά συστατικά του. Ο νευρώνας είναι ένα μεγάλο σε μέγεθος κύτταρο, και παρά την ύπαρξη αρκετών παραλλαγών του, ανατομικά αποτελείται από τέσσερα βασικά συστατικά. Τους δενδρίτες, το σώμα - που αποτελεί και τον πυρήνα του νευρώνα, τον άξονα και τις συνάψεις. Οι δενδρίτες λειτουργούν ως πύλες εισόδου του νευρώνα, λαμβάνοντας σήματα από γειτονικούς νευρώνες, με τη μορφή ηλεκτρικών παλμών. Το σώμα του νευρώνα, επεξεργάζεται αυτά τα εισερχόμενα σήματα με την πάροδο του χρόνου και τα αποστέλλει,

σε άλλους νευρώνες μέσω του άξονα - ο οποίος έχει το σχήμα μακρόστενης κλωστής. Ανάμεσα στον άξονα του νευρώνα και στους δενδρίτες άλλων νευρώνων, υπάρχει ένα σημείο ένωσης, που ονομάζεται σύναψη. Οι συνάψεις αυτές - συχνά αναφέρονται και ως νευρικές απολήξεις, είναι στοιχειώδεις δομικές και λειτουργικές μονάδες που παίζουν διαμεσολαβητικό ρόλο κατά τις αλληλεπιδράσεις μεταξύ των νευρώνων. Στόχος τους είναι η μετάδοση της ηλεκτρικής δραστηριότητας από τον άξονα - αποστολέα στους δενδρίτες - παραλήπτες, δημιουργώντας έτσι ένα νευρωνικό δίκτυο. Χαρακτηριστικά όπως το πλάτος της σύναψης και η απόστασή της από τον δενδρίτη, επηρεάζουν την ευκολία με την οποία μεταδίδεται εν τέλει το φορτίο. Η ευκολία αυτή εκφράζεται με μία τιμή, την οποία αποκαλούμε συναπτικό βάρος. Κάθε νευρώνας αξιολογεί τα εισερχόμενα σήματα, χρησιμοποιώντας τα αντίστοιχα συναπτικά βάρη προκειμένου να υπολογίσει το τελικό φορτίο. Όσο πιο ισχυρή είναι η σύνδεση μεταξύ τους, τόσο πιο έντονα συμμετέχει το συγκεκριμένο φορτίο εισόδου στο συνολικό άθροισμα. Αν το άθροισμα του φορτίου ξεπερνάει κάποιο κατώφλι (threshold) τότε ο άξονας αρχίζει να παράγει ηλεκτρικούς παλμούς με μεγάλη συχνότητα, οπότε λέμε ότι ο νευρώνας πυροβολεί και θεωρούμε τη σύναψη ενισχυτική. Αν όμως το άθροισμα του φορτίου δεν ξεπερνάει το κατώφλι αυτό, τότε ο νευρώνας εκπέμπει σε τυχαία χρονικά διαστήματα πολύ αραιά παλμούς και λέμε ότι ο νευρώνας είναι αδρανής και έτσι θεωρούμε τη σύναψη ως ανασταλτική. (Βλ. Εικόνα 1.1 για τη δομή του βιολογικού νευρώνα) (Haykin, 2010)· (Βλαχάβας, Κεφάλας, Βασιλειάδης, Κόκκορας και Σακελλαρίου, 2006)· (Διαμαντάρας, 2007)



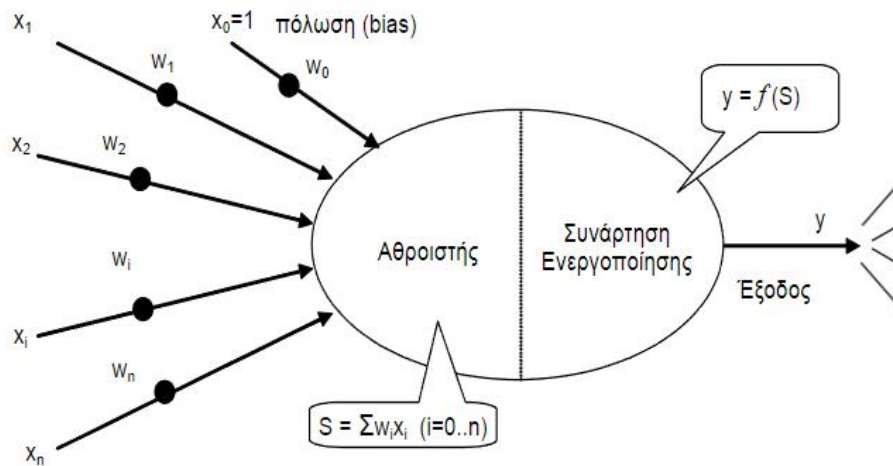
Εικόνα 1 Αναπαράσταση βιολογικού νευρώνα (Βλαχάβας, 2013)

Ο εγκέφαλος έχει την ικανότητα να λαμβάνει πολύπλοκες αποφάσεις με εκπληκτική ταχύτητα. Παρά το γεγονός ότι οι νευρώνες είναι πιο αργοί από τις λογικές πύλες που υλοποιούνται με τεχνολογία πυριτίου, αυτό επιτυγχάνεται καθώς η υπολογιστική ικανότητά του αλλά και η πληροφορία που αυτός περιέχει, είναι

διαμοιρασμένα σε όλο του τον όγκο. Αυτά τα χαρακτηριστικά αποτελούν και το μεγαλύτερο κίνητρο έτσι ώστε τα ΤΝΔ να μοντελοποιήσουν τον εγκέφαλο. (Haykin, 2010)· (Βλαχάβας et al., 2006)· (Πούλος, 2015) Αξίζει να σημειωθεί πως τα συνήθη ΤΝΔ χρησιμοποιούν απλοποιημένα μοντέλα νευρώνων, τα οποία διατηρούν μόνο τα πολύ βασικά χαρακτηριστικά των λεπτομερών μοντέλων που χρησιμοποιούνται στη νευρολογία. Θα έλεγε κανείς πως οι τεχνητοί νευρώνες που χρησιμοποιούμε για την κατασκευή των ΤΝΔ είναι αρκετά πρωτόγονοι σε σύγκριση με τους νευρώνες του ανθρώπινου εγκεφάλου. Ωστόσο, υπάρχει η πεποίθηση πως παρά την απλότητα τους, οι λεπτομέρειες δεν έχουν ιδιαίτερη σημασία στην κατανόηση της ευφυούς συμπεριφοράς των βιολογικών νευρωνικών συστημάτων. Αυτά τα απλά μοντέλα νευρώνων αξιοποιώντας δύο (2) βασικά χαρακτηριστικά, είναι ικανά να δημιουργήσουν ιδιαίτερος ενδιαφέροντα δίκτυα. Το πρώτο χαρακτηριστικό είναι η διευκόλυνση της μάθησης με τη χρήση ρυθμιζόμενων παραμέτρων στους νευρώνες, το οποίο είναι γνωστό και ως πλαστικότητα των νευρώνων. Το δεύτερο χαρακτηριστικό, είναι πως το δίκτυο αποτελείται από μεγάλο πλήθος νευρώνων ώστε να επιτυγχάνεται παραλληλισμός της επεξεργασίας και κατανομή της πληροφορίας. (Διαμαντάρας, 2007)· (Ηλιάδης, χ.χ.)

Επομένως, ο τεχνητός νευρώνας είναι ένα υπολογιστικό μοντέλο, του οποίου τα μέρη μπορούν να αντιστοιχιστούν άμεσα με αυτά του βιολογικού νευρώνα. Δέχεται αντί για ηλεκτρικούς παλμούς, συνεχείς μεταβλητές ως σήματα εισόδου. Κάθε σήμα εισόδου, μεταβάλλεται από μία τιμή βάρους η οποία είναι πραγματικός αριθμός, και αντιστοιχεί στον ρόλο της σύναψης σε έναν βιολογικό νευρώνα. Το σώμα του τεχνητού νευρώνα αποτελείται από δύο τμήματα. Το πρώτο περιλαμβάνει τον αθροιστή (adder), ο οποίος προσθέτει τα επηρεασμένα από τα βάρη σήματα εισόδου. Το δεύτερο τμήμα εμπεριέχει τη συνάρτηση ενεργοποίησης (activation function), η οποία λειτουργεί ως φίλτρο και στόχος της είναι η διαμόρφωση της τελικής τιμής του σήματος εξόδου. Δεδομένου ότι περιορίζει το επιτρεπτό εύρος πλάτους του σήματος εξόδου σε κάποια πεπερασμένη τιμή, συχνά αναφέρεται και ως συνάρτηση περιορισμού/συμπίεσης (squashing function). Συνήθως θεωρούμε πως υπάρχει και ένα επιπλέον βάρος, το οποίο χαρακτηρίζεται ως πόλωση (bias) ή αλλιώς παράγοντας προδιάθεσης του νευρώνα. Η μόνη διαφορά του από τα υπόλοιπα βάρη, είναι ότι επιδρά συνεχώς σε μία τιμή εισόδου με τιμή ένα (1). Ο όρος αυτός πρόκειται για εξωτερικό ερέθισμα το οποίο προστίθεται με τα εισερχόμενα σήματα, επομένως δεν πρέπει να αποδίδεται στο εσωτερικό του νευρώνα. Μπορούμε να πούμε πως ο νευρώνας είναι ντετερμινιστικό/αιτιοκρατικό μοντέλο, διότι είναι επακριβώς καθορισμένη η συμπεριφορά του για όλες τις εισόδους. Δηλαδή για κάθε είσοδο, υπάρχει

μία προκαθορισμένη έξοδος. (Βλ. Εικόνα 1.2 για τη δομή του τεχνητού νευρώνα) (Haykin, 2010)· (Βλαχάβας et al., 2006)· (Θεοδωρίδης και Κουτρούμπας, 2012)· (Πούλος, 2015)



Εικόνα 2 Αναπαράσταση τεχνητού νευρώνα (Βλαχάβας, 2013)

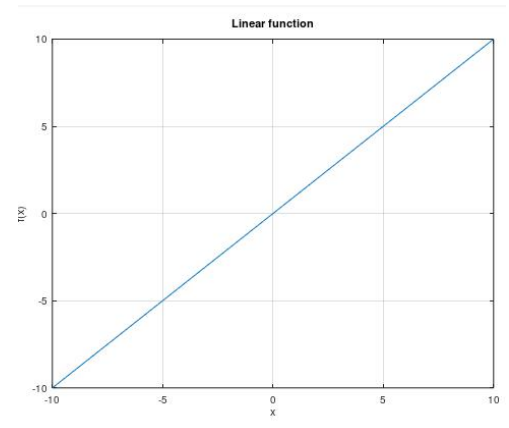
1.3 Συναρτήσεις ενεργοποίησης

Υπάρχουν αρκετές προσεγγίσεις για τη μοντελοποίηση του νευρώνα, οι οποίες διαφέρουν από το μοντέλο McCulloch και Pitts. Η πιο σημαντική διαφορά είναι στη μορφή της συνάρτησης ενεργοποίησης που χρησιμοποιείται. Όπως αναφέρθηκε στο υποκεφάλαιο 1.2, η συνάρτηση αυτή - η οποία συμβολίζεται ως $f(x)$, αποτελεί ένα από τα δύο βασικά τμήματα του σώματος του τεχνητού νευρώνα, καθώς ορίζει την έξοδο του. Υπάρχουν αρκετά είδη συναρτήσεων που μπορούν να χρησιμοποιηθούν, με πιο δημοφιλή τα παρακάτω:

Γραμμική συνάρτηση:¹

(Βλ. Παράρτημα Α, Κώδικας 1)

$$f(x) = x$$

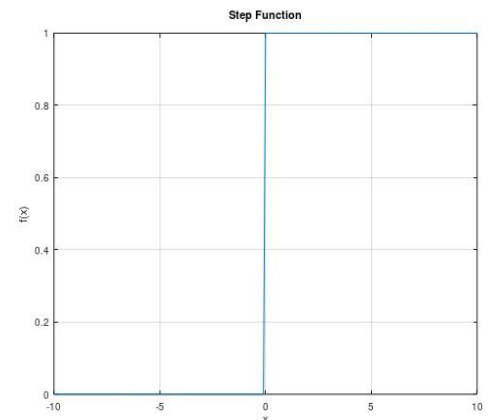


Εικόνα 3 Γραφική παράσταση της γραμμικής συνάρτησης

Βηματική συνάρτηση:²

(Βλ. Παράρτημα Α, Κώδικας 2)

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$$



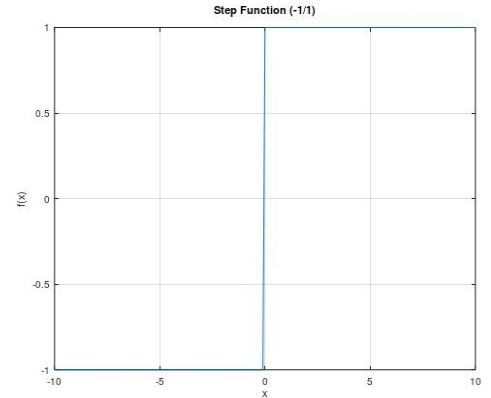
Εικόνα 4 Γραφική παράσταση της βηματικής συνάρτησης 0/1

¹ Χρησιμοποιείται στο δίκτυο Adaline (Διαμαντάρας, 2007)· (Δούνιας και Καραμπότσης, 2023)· (Τσούλος, Τεχνητά Νευρωνικά Δίκτυα, χ.χ.)

² Αλλιώς αναφέρεται και ως συνάρτηση κατωφλίου (Λύκας, χ.χ.) ή συνάρτηση Heaviside. (Haykin, 2010) Χρησιμοποιείται στον νευρώνα McCulloch & Pitts, αλλά και στο δίκτυο Perceptron (Δούνιας και Καραμπότσης, 2023) Δεν προτιμάται στο δίκτυο MLP. Είναι η πιο απλή συνάρτηση ενεργοποίησης και για αυτόν τον λόγο, χρησιμοποιείται περισσότερο σε θεωρητικές μελέτες. (Chiarandini, χ.χ.)· (Τσούλος, Τεχνητά Νευρωνικά Δίκτυα, χ.χ.) Η έξοδος είναι ένας δυαδικός αριθμός είτε με την κλασική μορφή (0/1) είτε με τη λεγόμενη διπολική μορφή (-1/1). Η επιλογή ωστόσο μεταξύ αυτών των μορφών είναι ήσσονος σημασίας (Διαμαντάρας, 2007).

Αλλά και με την μορφή ³
(Βλ. Παράρτημα Α, Κώδικας 3)

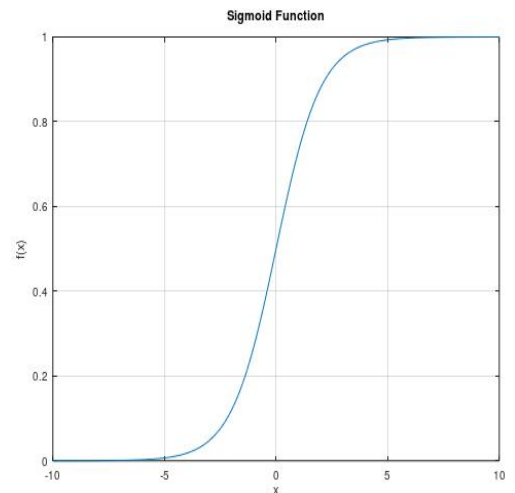
$$f(x) = \begin{cases} -1 & x \leq 0 \\ +1 & x > 0 \end{cases}$$



Εικόνα 5 Γραφική παράσταση της βηματικής συνάρτησης -1/1

Σιγμοειδής συνάρτηση:⁴
(Βλ. Παράρτημα Α, Κώδικας 4)

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



Εικόνα 6 Γραφική παράσταση της σιγμοειδής συνάρτησης

³ Η συνάρτηση αυτή συμβολίζεται και ως sgn από το λατινικό ‘signum’ που θα πει πρόσημο (Διαμαντάρας, 2007).

⁴ Χρησιμοποιείται περισσότερο σε πρακτικές εφαρμογές. (Chiarandini, χ.χ.) Η γραφική της παράσταση έχει το σχήμα “S” και είναι η πιο κοινή μορφή συνάρτησης ενεργοποίησης που χρησιμοποιείται για την κατασκευή ΤΝΔ. (Haykin, 2010)· (Διαμαντάρας, 2007)· (Τσούλος, Τεχνητά Νευρωνικά Δίκτυα, χ.χ.)

1.4 Τεχνητά Νευρωνικά Δίκτυα: Μία σφαιρική εικόνα

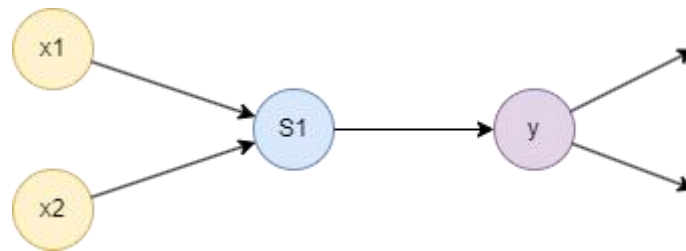
Τα ΤΝΔ μοιράζονται δύο (2) χαρακτηριστικά με τον ανθρώπινο εγκέφαλο. Το πρώτο από αυτά είναι πως το δίκτυο προσλαμβάνει τη γνώση από το περιβάλλον του, μέσω μίας διαδικασίας μάθησης. Το δεύτερο χαρακτηριστικό είναι πως η αποκτηθείσα γνώση αποθηκεύεται μέσω της προσαρμογής της ισχύος των συνδέσεων μεταξύ των νευρώνων. Χάριν συντομίας, τα δίκτυα αυτά σύμφωνα με τον Haykin (2010), αποκαλούνται και νευρωνικά δίκτυα. Σύμφωνα με τους τους Βλαχάβα et al. (2006), έχει διατυπωθεί ο παρακάτω ορισμός από τους Aleksander και Morton:

Ένα νευρωνικό δίκτυο είναι ένας τεράστιος παράλληλος επεξεργαστής με κατανεμημένη αρχιτεκτονική, ο οποίος αποτελείται από απλές μονάδες επεξεργασίας και έχει από τη φύση του τη δυνατότητα να αποθηκεύει εμπειρική γνώση και να την καθιστά διαθέσιμη για χρήση.

Αυτές οι μονάδες επεξεργασίας, δηλαδή οι νευρώνες, συνήθως είναι οργανωμένοι σε μία σειρά από επίπεδα (layers) ή αλλιώς στρώματα. Το πρώτο από αυτά χρησιμοποιείται για την εισαγωγή των δεδομένων και ονομάζεται στρώμα εισόδου (input layer). Η χρήση του όρου νευρώνες σε αυτό το στρώμα είναι καταχρηστική καθώς δεν γίνεται κάποια επεξεργασία σε αυτούς. Επιπλέον, μπορεί να υπάρχουν προαιρετικά ένα ή περισσότερα κρυφά στρώματα (hidden layer) ενώ στο τέλος υπάρχει το στρώμα εξόδου (output layer).

Υπάρχουν δύο κατηγορίες νευρώνων ανάλογα με τον τρόπο διασύνδεσης τους στο δίκτυο. Η πιο συνηθισμένη περίπτωση είναι αυτή των πλήρως συνδεδεμένων νευρώνων κατά την οποία οι νευρώνες ενός στρώματος είναι πλήρως συνδεδεμένοι με τους νευρώνες του επόμενου στρώματος χωρίς όμως να είναι συνδεδεμένοι με νευρώνες είτε του προηγούμενου είτε του ίδιου στρώματος. Χαρακτηριστικό παράδειγμα αυτής της κατηγορίας, είναι τα δίκτυα πρόσθιας τροφοδότησης (feed-forward) τα οποία χρησιμοποιούνται στην πλειοψηφία των εφαρμογών των ΤΝΔ. Το όνομα τους οφείλεται στο ότι η ροή της πληροφορίας μέσα στο δίκτυο είναι μονής κατεύθυνσης. Σε αυτά τα δίκτυα υπάρχουν τα στρώματα εισόδου και εξόδου και προαιρετικά ένα ή περισσότερα κρυφά στρώματα. (Βλαχάβας et al., 2006) Για τις περιπτώσεις που δεν υπάρχουν κρυφά στρώματα, πολλές φορές χρησιμοποιείται για αυτά, ο όρος Perceptron. (Γεωργούλη, 2015)

Σε οποιαδήποτε άλλη περίπτωση θεωρούμε ότι οι νευρώνες είναι μερικώς συνδεδεμένοι. (Λυκοθανάσης, χ.χ.) Στο παρακάτω σχήμα, παρουσιάζεται μία απλή μορφή ΤΝΔ πρόσθιας τροφοδότησης. Στο στρώμα εισόδου υπάρχουν δύο (2) νευρώνες, στο κρυφό στρώμα ένας (1) νευρώνας επεξεργασίας και στο στρώμα εξόδου ένας (1) νευρώνας. Σε αυτό το σημείο καλό θα ήταν να διευκρινίσουμε πως παρόλο που ο νευρώνας εξόδου έχει δύο (2) εξόδους, αυτές θα πρέπει να έχουν τις ίδιες ακριβώς τιμές.



Εικόνα 7 Σχηματική απεικόνιση ενός απλού ΤΝΔ

Υπάρχουν τέσσερις (4) ιδιότητες που συνδέονται άρρηκτα με τα ΤΝΔ. Η πρώτη από αυτές είναι η ικανότητά τους να μαθαίνουν μέσω παραδειγμάτων (learn by example). Η δεύτερη ιδιότητα αφορά τη δυνατότητα τους να θεωρηθούν ως κατανεμημένη μνήμη (distributed memory) και ως μνήμη συσχέτισης (associative memory). Αυτό οφείλεται στο γεγονός ότι η κωδικοποίηση που δημιουργείται είναι κατανεμημένη σε όλα τα βάρη της συνδεσμολογίας τους. Επίσης, αντί για τον παραδοσιακό τρόπο αποθήκευσης της πληροφορίας, χρησιμοποιούν κατάλληλες συσχετίσεις που προκύπτουν από τα δεδομένα εκπαίδευσης. Έτσι, η ανάκληση της πληροφορίας γίνεται βάσει του περιεχομένου και όχι της διεύθυνσης, όπως ακριβώς συμβαίνει και στον ανθρώπινο εγκέφαλο. Η τρίτη ιδιότητα είναι η μεγάλη τους ανοχή σε σφάλματα (fault-tolerant) καθώς ορισμένα είδη ΤΝΔ είναι σε θέση να παράγουν τη σωστή έξοδο ακόμη και αν τα δεδομένα εισόδου είναι λίγο διαφορετικά. Μία τέτοια κατάσταση μπορεί να προκληθεί είτε λόγω θορύβου είτε επειδή τα δεδομένα είναι ελλιπή. Επιπλέον, ακόμη και σε περίπτωση δομικών σφαλμάτων, η δυσλειτουργία ή ακόμη και η καταστροφή ενός νευρώνα ή ορισμένων συνδέσεων δε μπορεί να διαταράξει σημαντικά τη λειτουργία του δικτύου. Αυτό οφείλεται στο γεγονός ότι η πληροφορία που εσωκλείουν δεν είναι συγκεντρωμένη σε συγκεκριμένο σημείο αλλά είναι διάχυτη σε ολόκληρο το δίκτυο. Η τελευταία ιδιότητα είναι η ικανότητά τους για αναγνώριση προτύπων (pattern recognition). Από τη στιγμή που ένα ΤΝΔ εκπαιδεύεται στο να αναγνωρίζει συνθήκες και καταστάσεις, τότε προκειμένου να προσδιοριστεί μία

συγκεκριμένη κατάσταση, απαιτείται ένας μόνο κύκλος λειτουργίας του. Οι δύο τελευταίες ιδιότητες καθιστούν τα ΤΝΔ ιδανικά για εφαρμογές αυτοματισμού, οι οποίες πρέπει να είναι σε θέση να λειτουργήσουν σε αντίξοες συνθήκες όπως για παράδειγμα σε διαστημικές αποστολές, σε χώρους με ραδιενέργεια, ακόμη και σε πεδία μάχης. (Βλαχάβας et al., 2006)· (Λύκας, χ.χ)

1.5 Μάθηση με επίβλεψη

Τα ΤΝΔ αντιμετωπίζουν την πρόκληση της εύρεσης κατάλληλων αλγορίθμων με σκοπό την εκπαίδευση τους καθώς και την ανάκληση της πληροφορίας που περιέχεται. Αυτό έχει ως στόχο την προσομοίωση ευφυνών διαδικασιών όπως είναι η αναγνώριση προτύπων, φυσικής γλώσσας αλλά και προσώπων, η πλοήγηση ρομπότ σε φυσικά εμπόδια, η ανάπτυξη βέλτιστων στρατηγικών για την επίλυση προβλημάτων και η εκτέλεση συλλογισμών για την εξαγωγή λογικών συμπερασμάτων. Αξίζει να σημειωθεί ότι τα ΤΝΔ στοχεύουν επιπλέον, στην αυτοπροσαρμογή σε νέες καταστάσεις και γνωστά περιβάλλοντα, καθώς και στη μάθηση από την εμπειρία τους. Προκειμένου να επιτευχθούν τα παραπάνω, απαιτείται ο ορισμός του κατάλληλου περιβάλλοντος εκπαίδευσης, μία διαδικασία που μπορεί να κατηγοριοποιηθεί ανάλογα με τον τρόπο λειτουργίας της. Οι πιο βασικές κατηγορίες είναι αυτή της μάθησης με εκπαιδευτή και αυτή χωρίς εκπαιδευτή.

Σε αυτή την πτυχιακή εργασία, θα εστιάσουμε στη μάθηση με εκπαιδευτή, την κατηγορία που χρησιμοποιείται στα περισσότερα δίκτυα. Η μάθηση με εκπαιδευτή αναφέρεται επίσης ως μάθηση με επίβλεψη ή αλλιώς επιβλεπόμενη μάθηση. Σε αυτή την κατηγορία μάθησης, το περιβάλλον είναι άγνωστο στο νευρωνικό δίκτυο και υποθέτουμε πως δεν υπάρχει κάποια εξωτερική "αυθεντία" η οποία γνωρίζει τις ορθές παραμέτρους για τη λύση του προβλήματος. Ωστόσο, μπορούμε να θεωρήσουμε ότι ο εκπαιδευτής έχει γνώση του περιβάλλοντος, η οποία αντιπροσωπεύεται από ένα δείγμα προτύπων, έχοντας αντιστοίχιση μεταξύ των εισόδων (ή αλλιώς ερεθισμάτων) - γνωστές και ως ανεξάρτητες μεταβλητές και της αντίστοιχης επιθυμητής απόκρισης (στόχος) - γνωστή και ως εξαρτημένη μεταβλητή. (Βλαχάβας et al., 2006) Αυτή η πληροφορία είναι γνωστή και ως a-priori (εκ των προτέρων) και τα πρότυπα αναφέρονται ως χαρακτηρισμένα (labeled). Αξίζει να σημειωθεί πως η εργασία της συλλογής αυτών των προτύπων είναι χρονοβόρα και ακριβή, ειδικά όταν αντιμετωπίζουμε μεγάλης κλίμακας προβλήματα μάθησης.

Όπως αναφέρθηκε παραπάνω, σε αυτή την κατηγορία μάθησης υπάρχει αντιστοίχιση ανάμεσα στην είσοδο και στην έξοδο. Έτσι, το δίκτυο συνέχεια επεξεργάζεται τις εισόδους με τα αντίστοιχα συναπτικά βάρη και συγκρίνει το αποτέλεσμα αυτό, με το επιθυμητό αποτέλεσμα της εξόδου. Στις περιπτώσεις κατά τις οποίες δεν υπάρχει αντιστοίχιση ανάμεσα στο πραγματικό και το επιθυμητό αποτέλεσμα της εξόδου, το σύστημα αναπροσαρμόζει τα βάρη του. Υπάρχουν όμως και περιπτώσεις, κατά τις οποίες η αναπροσαρμογή των βαρών γίνεται μετά από κάθε πέρασμα κάποιου προτύπου ανεξάρτητα από την έξοδο του στο δίκτυο. Αυτή είναι μία επαναλαμβανόμενη διαδικασία, καθώς τα βάρη βελτιώνονται συνεχώς. Το σύστημα θεωρούμε πως έχει εκπαιδευτεί με επιτυχία όταν θα έχουμε μία ολόκληρη εποχή χωρίς να χρειάζεται αναπροσαρμογή των βαρών. (Haykin, 2010). (Διαμαντάρας, 2007). (Θεοδωρίδης και Κουτρούμπας, 2012). (Κίτρου, χ.χ.). (Λυκοθανάσης και Κουτσομητρόπουλος, 2021) Πρέπει να σημειωθεί βέβαια ότι κάποια δίκτυα δεν μαθαίνουν ποτέ. Ένας βασικός λόγος που μπορεί να προκύψει ένα τέτοιο αποτέλεσμα, είναι από λανθασμένα σύνολα δεδομένων ή μη επαρκείς πληροφορίες. Αξίζει να σημειωθεί πως όσο περισσότερα δεδομένα διαθέτουμε, τόσο το καλύτερο, διότι αρκετά δεδομένα μπορούν να κρατηθούν για δοκιμή. (Anderson & McNeill, 1992)

Σύμφωνα με τον Διαμαντάρα (2007), οι κύριοι αλγόριθμοι εκπαίδευσης με επίβλεψη περιλαμβάνουν τους Perceptron και Adaline. Παράλληλα οι Βλαχάβας et al. (2006) αναφέρουν ότι στη μάθηση με επίβλεψη διακρίνονται δύο είδη προβλημάτων. Το πρώτο είδος αναφέρεται σε προβλήματα ταξινόμησης (classification), των οποίων στόχος είναι η δημιουργία μοντέλων πρόβλεψης διακριτών κλάσεων όπως για παράδειγμα η ομάδα αίματος. Το δεύτερο είδος αφορά τα προβλήματα παρεμβολής (regression), τα οποία αφορούν τη δημιουργία μοντέλων πρόβλεψης αριθμητικών τιμών. Τέλος, σύμφωνα με τον Κίτρου (χ.χ.), η μάθηση με επίβλεψη χρησιμοποιείται και σε προβλήματα πρόγνωσης αλλά και διερμηνείας.

2. Δομή και Λειτουργία των Τεχνητών Νευρωνικών Δικτύων

2.1 Δίκτυα Perceptron

Τα δίκτυα Perceptron καταλαμβάνουν μία ειδική θέση στην ιστορική εξέλιξη των ΤΝΔ καθώς αποτελούν τη πρώτη μορφή δικτύου που μπορούσε να περιγραφεί αλγοριθμικά. Χρησιμοποιούνται για την ταξινόμηση γραμμικά διαχωρίσιμων προτύπων, αρκεί αυτά να ανήκουν σε δύο μόνο κατηγορίες. Η σπουδαιότητα τους δεν είναι μόνο ιστορική καθώς συνεχίζουν να έχουν πρακτική χρησιμότητα. (Haykin, 2010) Υπάρχουν αρκετές παραλλαγές με την πιο απλή να είναι αυτή του στοιχειώδους Perceptron (elementary Perceptron). Κοινό χαρακτηριστικό σε όλες είναι πως περιλαμβάνουν μόνο έναν νευρώνα. Εφόσον δεν υπάρχουν παραπάνω από ένας νευρώνας ώστε να συνδεθούν μεταξύ τους, ο όρος δίκτυο χρησιμοποιείται καταχρηστικά.

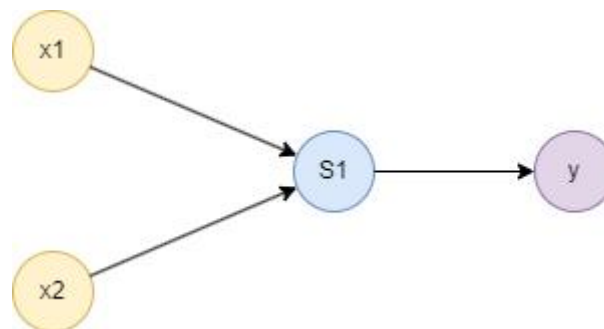
Τα δίκτυα αυτά έχουν αναπτυχθεί βάσει τριών (3) ερωτημάτων. Το πρώτο από αυτά είναι το πως γίνεται αισθητή η ανίχνευση πληροφοριών για τον φυσικό κόσμο, από το βιολογικό σύστημα. Το δεύτερο ερώτημα, αφορά την μορφή με την οποία αποθηκεύονται ή απομνημονεύονται οι πληροφορίες. Το τρίτο και τελευταίο ερώτημα, αφορά τον τρόπο με τον οποίο οι πληροφορίες που έχουν αποθηκευτεί επηρεάζουν την αναγνώριση και τη συμπεριφορά. Στόχος των δικτύων αυτών είναι η απεικόνιση μερικών θεμελιωδών ιδιοτήτων των ευφυών συστημάτων χωρίς όμως να είναι απαραίτητο να εμβαθύνουν σε ειδικές συνθήκες που μπορεί να αφορούν συγκεκριμένους βιολογικούς οργανισμούς, οι οποίες συχνά μπορεί να είναι και άγνωστες. (Rosenblatt, 1958)

Έχει αποδειχθεί πως το αποτέλεσμα του Perceptron είναι το ίδιο, ανεξάρτητα από την συνάρτηση ενεργοποίησης που θα χρησιμοποιηθεί, αν και η πιο σύνηθες συνάρτηση για αυτό το δίκτυο είναι η βηματική. (Haykin, 2010)· (Βλαχάβας et al., 2006) Τα δίκτυα Perceptron αποτελούν την πιο απλή μορφή δικτύου πρόσθιας τροφοδότησης χρησιμοποιώντας μάθηση με επίβλεψη, η οποία σύμφωνα με τους Βλαχάβα et al. (2006) είναι καθοδηγούμενη από το σφάλμα (error driven). Ζητούμενο αυτού του δικτύου είναι η αυτόματη εκμάθηση των παραμέτρων του συστήματος - δηλαδή των βαρών - με στόχο την επίτευξη του επιθυμητού στόχου. Ο κλασικός κανόνας εκπαίδευσης Perceptron είναι

γνωστός και ως κανόνας σταθερής αύξησης (fixed increment rule). Πρακτικά, τα πρότυπα παρουσιάζονται επαναληπτικά με κυκλική σειρά στο δίκτυο. Σε περίπτωση που υπάρχει σφάλμα ταξινόμησης - δηλαδή η πραγματική έξοδος είναι διαφορετική από την επιθυμητή, μόνο τότε γίνεται η εκπαίδευση των βαρών. Για τον λόγο αυτό, ο αλγόριθμος ανήκει σε μία πιο γενική οικογένεια αλγορίθμων, γνωστή ως σχήματα ανταμοιβής και τιμωρίας (reward and punishment). Αν η ταξινόμηση είναι σωστή, παρέχεται ανταμοιβή και δε χρειάζεται να γίνει καμία περαιτέρω ενέργεια. Σε αντίθετη περίπτωση, επιβάλλεται τιμωρία και απαιτείται εκπαίδευση των βαρών. Με το που γίνει η εκπαίδευση, έχει αποδειχθεί πως το πρότυπο αυτό, την επόμενη φορά που θα χρησιμοποιηθεί είτε θα ταξινομηθεί σωστά είτε θα πλησιάζει ακόμη περισσότερο στην σωστή ταξινόμηση του.

Βέβαια, αν το πρόβλημα δεν είναι γραμμικά διαχωρίσιμο τότε ο αλγόριθμος Perceptron δεν συγκλίνει ποτέ. Αυτό είναι ένα σοβαρό μειονέκτημα το οποίο αποτέλεσε και το βασικό σημείο κριτικής εναντίον του. Πολλά προβλήματα στον πραγματικό κόσμο είναι μη γραμμικά διαχωρίσιμα, επομένως δεν μπορεί να χρησιμοποιηθεί αυτό το μοντέλο για αυτά. (Διαμαντάρας, 2007)· (Θεοδωρίδης και Κουτρούμπας, 2012)

Στο παρακάτω σχήμα μπορούμε να δούμε μία απεικόνιση του δικτύου Perceptron με δύο (2) νευρώνες στο στρώμα εισόδου, έναν (1) νευρώνα επεξεργασίας και έναν (1) νευρώνα στο στρώμα εξόδου.



Εικόνα 8 Perceptron

Προκειμένου να γίνει υλοποίηση ενός δικτύου Perceptron θα πρέπει αρχικά να γίνει αποθήκευση των διαθέσιμων προτύπων σε συνδυασμό με τις αντίστοιχες επιθυμητές εξόδους τους. Αξίζει να σημειωθεί σε αυτό το σημείο, πως τυχόν λάθη στα πρότυπα ή στις επιθυμητές εξόδους θα οδηγήσει σε μη αξιόπιστα αποτελέσματα του δικτύου. Κάτι τέτοιο μπορεί να προκύψει από λανθασμένη εισαγωγή δεδομένων, ακόμη και από αστοχία υλικού - για παράδειγμα από κάποιον αισθητήρα. Κάποιες φορές μπορεί να υπάρχουν και ακραίες

τιμές, το οποίο δυσκολεύει τη λειτουργία των ΤΝΔ. Το δεύτερο βήμα είναι η αρχικοποίηση του διανύσματος βαρών με σκοπό την κατάλληλη εκπαίδευση του δικτύου. Θα πρέπει όμως πρώτα να είναι γνωστό, το απαραίτητο πλήθος των τιμών των βαρών, το οποίο προκύπτει από τον τύπο:

$$size(w) = (d + 1)_5$$

Τύπος 1 Υπολογισμός πλήθους βαρών για ένα Perceptron δίκτυο

Υπάρχουν αρκετοί τρόποι αρχικοποίησης των βαρών, με τον πιο απλό αυτόν της εκχώρησης σε όλες τις τιμές, μία τιμή είτε ίση με το μηδέν (0) είτε με την μονάδα (1). Εναλλακτικά, μπορεί να γίνει αρχικοποίηση των τιμών σε τυχαίες τιμές στο κλειστό διάστημα $[0,1]$. Έπειτα, γίνεται αρχικοποίηση του ρυθμού μάθησης του δικτύου σε μία μικρή θετική δεκαδική τιμή, συνήθως στο κλειστό διάστημα $[0,1]$. Συχνά γίνεται χρήση της τιμής 0.1. (Τσούλος, 2019) Στη συνέχεια εκτελείται μία επαναληπτική διαδικασία για κάθε διαθέσιμο πρότυπο. Έτσι, για κάθε πρότυπο γίνεται ο υπολογισμός της εξόδου του δικτύου χρησιμοποιώντας τις διαθέσιμες τιμές των βαρών. Ο υπολογισμός αυτός προκύπτει από την συνάρτηση ενεργοποίησης που έχει επιλεγεί για χρήση σε αυτό το δίκτυο. Ως είσοδος της συνάρτησης αυτής είναι ο υπολογισμός του αθροίσματος των εισόδων πολλαπλασιασμένων με τα αντίστοιχα συναπτικά βάρη. Η εκπαίδευση των βαρών γίνεται μόνο σε περίπτωση που το αποτέλεσμα είναι διαφορετικό από το επιθυμητό. Η διαδικασία αυτή τερματίζει είτε μετά από μία ολόκληρη εποχή που δεν έχει χρειαστεί εκπαίδευση των βαρών είτε μετά από κάποιο πεπερασμένο αριθμό επαναλήψεων - εφόσον αυτός έχει οριστεί στην αρχή της υλοποίησης. (Βλ. Ψευδοκώδικα 1 στο Παράρτημα Β)

Η ενημέρωση των βαρών, γίνεται με την πρόσθεση ή την αφαίρεση ενός ποσοστού από το πρότυπο με βάση τον παρακάτω μαθηματικό τύπο:

⁵ Το πλήθος των βαρών ισούται με το πλήθος των χαρακτηριστικών των προτύπων που έχουμε στη διάθεση μας, αυξημένο κατά 1 - για το bias που αναφέραμε στο υποκεφάλαιο 1.2 πως είναι απαραίτητο.

$$w^{(k)} = w^{(k-1)} + \beta(d^p - y^p)x^p,$$

Τύπος 2 Ενημέρωση βαρών του δικτύου Perceptron

2.2 Δίκτυα Adaline

Ο όρος Adaline είναι ένα ακρώνυμο των λέξεων ADaptive LINear Element και μεταφράζεται ως Αυτοπροσαρμόζόμενο Γραμμικό Στοιχείο. Ο όρος αυτός εισήχθη από τον Windrow. Η διαφορά σε σχέση με τον κλασικό νευρώνα των McCulloch και Pitts είναι ότι η έξοδος αυτού του δικτύου αλλά και οι στόχοι του προαιρετικά, μπορούν να πάρουν συνεχείς αντί για διακριτές τιμές. Η χρήση όμως των συνεχών τιμών δημιουργεί μία επιπλέον πρόκληση. Για παράδειγμα, αν επιθυμούμε να χρησιμοποιήσουμε το δίκτυο Adaline για να διαχωρίσουμε δύο κατηγορίες, θα πρέπει να αποδώσουμε τις συνεχείς τιμές στις κατηγορίες. Αν και συνήθως αυτό επιτυγχάνεται με τη διάκριση των θετικών και αρνητικών τιμών. Μοιάζουν αρκετά με τα δίκτυα Perceptron καθώς διαθέτουν μόνο έναν νευρώνα, αλλά μπορούν να χρησιμοποιηθούν επιπλέον για τη μάθηση συναρτήσεων καθώς και για τη μάθηση περισσότερων από δύο κατηγοριών υπό προϋποθέσεις. (Τσούλος, 2019)

Τα δίκτυα Adaline χρησιμοποιούν τον αναδρομικό on-line αλγόριθμο εκπαίδευσης που είναι γνωστός ως LMS και προκύπτει από τη θεωρία στοχαστικών προσεγγίσεων. Υπάρχουν και άλλες ονομασίες για αυτόν, όπως κανόνας δέλτα (delta rule) ή κανόνας ADALINE ή ακόμη και κανόνας Widrow-Hoff από τα ονόματα αυτών που τον εισήγαγαν. Για πολύ καιρό αντιμετωπιζόταν ως μία “στιγμιαία” προσέγγιση της μεθόδου Gradient Descent που θα αναλυθεί στο υποκεφάλαιο 2.5. (Haykin, 2010) Σε αυτόν τον αλγόριθμο, η ελαχιστοποίηση αφορά τη μείωση της τετραγωνικής απόστασης μεταξύ των διανυσμάτων της πραγματικής και της επιθυμητής εξόδου.

⁶ Η μεταβλητή k αντιπροσωπεύει τον τρέχον αριθμό επανάληψης. Επομένως ο πίνακας βαρών ενημερώνεται σύμφωνα με τιμές των βαρών που είχε κατά την προηγούμενη κατάσταση $w^{(k-1)}$, προσθέτοντας τον ρυθμό μάθησης β πολλαπλασιασμένο με το γινόμενο ανάμεσα στην διαφορά που προκύπτει μεταξύ της επιθυμητής (d^p) και της πραγματικής (y^p) εξόδου του τρέχον προτύπου, και τις τιμές που έχει το ίδιο το πρότυπο (x^p).

Έχει παρατηρηθεί βέβαια, πως με την χρήση αυτού του δικτύου υπάρχει περίπτωση να μην μπορέσουμε να διαχωρίσουμε γραμμικά διαχωρίσιμα πρότυπα δύο (2) κατηγοριών, κάτι το οποίο κάνει με απόλυτη επιτυχία το δίκτυο Perceptron. Αλλά σε αντίθετη περίπτωση, το δίκτυο μπορεί να βρει ένα σχετικά καλό διάνυσμα βαρών ώστε να κατηγοριοποιήσει τα πρότυπα με αρκετά μεγάλο ποσοστό επιτυχίας. (Διαμαντάρας, 2007)

Προκειμένου να γίνει υλοποίηση ενός δικτύου Adaline, θα πρέπει αρχικά να γίνει αποθήκευση των διαθέσιμων προτύπων σε συνδυασμό με τις αντίστοιχες επιθυμητές εξόδους τους. Και σε αυτή την περίπτωση - όπως και στα δίκτυα Perceptron, μπορεί να προκύψουν μη αξιόπιστα αποτελέσματα αν υπάρχουν τυχόν λάθη στα πρότυπα ή στις επιθυμητές εξόδους. Το δεύτερο βήμα είναι η αρχικοποίηση του διανύσματος βαρών έτσι ώστε να γίνει η κατάλληλη εκπαίδευση του δικτύου καθώς και του ρυθμού μάθησης. Οι τρόποι αρχικοποίησης των τιμών των βαρών, είναι και αυτοί ίδιοι με αυτούς κατά την υλοποίηση ενός δικτύου Perceptron. Θα πρέπει πρώτα να είναι γνωστό το απαραίτητο πλήθος των τιμών, το οποίο ισούται με το πλήθος των χαρακτηριστικών με βάση τα διαθέσιμα πρότυπα, αυξημένο κατά 1 - για το bias. (Βλ. Τύπος 1) Ο πιο απλός τρόπος αρχικοποίησης είναι η εκχώρηση σε όλες τις τιμές, μία τιμή είτε ίση με το μηδέν (0) είτε με την μονάδα (1). Εναλλακτικά, μπορεί να γίνει αρχικοποίηση των τιμών σε τυχαίες τιμές στο κλειστό διάστημα $[0,1]$. Η αρχικοποίηση του ρυθμού μάθησης είναι και αυτή ίδια με αυτή του Perceptron, δηλαδή σε μία μικρή θετική δεκαδική τιμή, συνήθως στο κλειστό διάστημα $[0,1]$. Συχνά γίνεται χρήση της τιμής 0.1. (Τσούλος, 2019) Στην περίπτωση των Adaline δικτύων, είναι απαραίτητος ο ορισμός και μίας τιμής ως κατώφλι προκειμένου να τερματίσει ο αλγόριθμος σε περίπτωση που το σφάλμα εκπαίδευσης έχει χαμηλότερη τιμή από αυτή. Στη συνέχεια εκτελείται μία επαναληπτική διαδικασία για κάθε διαθέσιμο πρότυπο. Έτσι, για κάθε πρότυπο γίνεται υπολογισμός της εξόδου του δικτύου χρησιμοποιώντας τις διαθέσιμες τιμές των βαρών. Ο υπολογισμός αυτός - σε αντίθεση με το Perceptron, δε προκύπτει από την συνάρτηση ενεργοποίησης, αλλά από το γινόμενο ανάμεσα στην είσοδο και τα συναπτικά βάρη. Σε κάθε περίπτωση γίνεται εκπαίδευση των βαρών και με αυτόν τον τρόπο, ενσωματώνεται και η έννοια του χρόνου. Στο τέλος κάθε εποχής, θα πρέπει να γίνει υπολογισμός του σφάλματος εκπαίδευσης και συνήθως υπολογίζεται με βάση το τετραγωνικό σφάλμα που προκύπτει από τον παρακάτω τύπο:

$$E = \sum_{i=1}^M (y^p - d^p)^2 \quad 7$$

Τύπος 3 Τετραγωνικό σφάλμα

Η διαδικασία αυτή τερματίζει, όταν πληροί μία από τις ακόλουθες συνθήκες. Είτε όταν για μία ολόκληρη εποχή, δεν υπάρχει μεταβολή στις τιμές των βαρών παρά την εκπαίδευση τους, είτε όταν το σφάλμα εκπαίδευσης έχει τιμή μικρότερη από την τιμή κατωφλίου που έχει οριστεί. Εναλλακτικά, η διαδικασία μπορεί να τερματίσει μετά από κάποιο πεπερασμένο αριθμό επαναλήψεων - εφόσον αυτός έχει οριστεί στην αρχή της υλοποίησης. Η ενημέρωση των βαρών γίνεται με την χρήση του ίδιου τύπου που χρησιμοποιείται και σε ένα δίκτυο Perceptron. (Βλ. Τύπο 2 για την ενημέρωση των βαρών) (Βλ. Ψευδοκώδικα 2 στο Παράρτημα Β)

2.3 Δίκτυα MLP

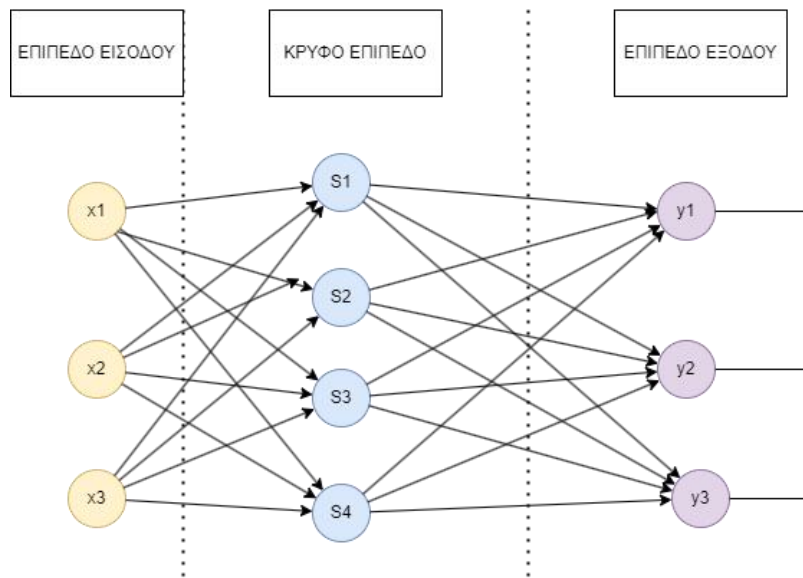
Όπως αναφέρθηκε σε προηγούμενα υποκεφάλαια, το δίκτυο Perceptron μπορεί να χρησιμοποιηθεί για τον διαχωρισμό προτύπων μόνο δύο κατηγοριών εφόσον αυτά είναι γραμμικά διαχωρίσιμα, ενώ το δίκτυο Adaline μπορεί να είναι αποδοτικό σε πρότυπα δύο κατηγοριών ακόμη και αν αυτά είναι σχεδόν γραμμικά διαχωρίσιμα. Εξέλιξη τους λοιπόν, είναι τα δίκτυα MLP, τα οποία αποτελούνται από περισσότερους από έναν νευρώνες, αίροντας τους περιορισμούς που έχουν αποδειχθεί για το Perceptron. Με τη χρήση της βηματικής συνάρτησης είτε με την απλή είτε με τη διπολική της μορφή, τα MLP δίκτυα έχουν τη δυνατότητα να υλοποιήσουν συναρτήσεις που δεν έχει τη δυνατότητα να υλοποιήσει κάποιο από τα δίκτυα Perceptron και Adaline. Ωστόσο, προτιμάται η σιγμοειδής συνάρτηση έναντι της βηματικής στα κρυφά στρώματα, καθώς αποδεικνύεται πως με την χρήση της, τα δίκτυα αυτά μπορούν να προσεγγίσουν οποιαδήποτε ομαλή συνάρτηση, όσο κοντά επιθυμούμε. Αυτός είναι και ο λόγος που αυτά τα δίκτυα καλούνται και Καθολικοί Προσεγγιστές (Universal Approximators). Αξιοσημείωτο είναι το γεγονός πως αρκεί να χρησιμοποιηθούν μόνο δύο στρώματα νευρώνων πέρα από το στρώμα εισόδου. Ένα από τα πιο βασικά

⁷ Το τετραγωνικό σφάλμα (E) προκύπτει από το άθροισμα των τετραγώνων των διαφορών μεταξύ της επιθυμητής (d^p) και της πραγματικής (y^p) εξόδου για κάθε ένα από τα M πρότυπα.

προβλήματα που δεν μπορεί να λύσει το δίκτυο Perceptron αλλά μπορεί το δίκτυο MLP είναι η υλοποίηση της πύλης XOR (eXclusive OR, αποκλειστική διάζευξη). (Haykin, 2010)· (Διαμαντάρας, 2007)· (Τσούλος, 2019)

Μπορούμε να πούμε πως η θεωρητική ανάλυση αυτών των δικτύων είναι αρκετά δύσκολη υπόθεση. Αυτό οφείλεται λόγω των κρυφών στρωμάτων που υπάρχουν σε αυτά καθώς και της μεγάλης διασυνδεσιμότητας τους. Στη γενική του μορφή το δίκτυο είναι πλήρως συνδεδεμένο (fully connected). Στα δίκτυα MLP μπορούμε να αναγνωρίσουμε δύο είδη σημάτων. Ένα από αυτά είναι τα λειτουργικά σήματα ή αλλιώς σήματα εισόδου. Όπως λέει και το όνομα τους, είναι σήματα που φτάνουν στην είσοδο του δικτύου, διαδίδονται σε όλο το δίκτυο και εν τέλει φτάνουν στην έξοδο του δικτύου σαν σήμα εξόδου. Το δεύτερο είδος σημάτων είναι τα σήματα σφάλματος (error signal) τα οποία προέρχονται από έναν νευρώνα εξόδου και διαδίδονται προς τα πίσω διαμέσου του δικτύου. Έτσι, κάθε νευρώνας εξόδου σχεδιάζεται με στόχο την εκτέλεση δύο (2) υπολογισμών. Ο πρώτος υπολογισμός αφορά το λειτουργικό σήμα που εμφανίζεται στην έξοδο κάθε νευρώνα, το οποίο εκφράζεται ως μία συνάρτηση του σήματος εισόδου και των συναπτικών βαρών που σχετίζονται με αυτόν τον νευρώνα. Ο δεύτερος υπολογισμός γίνεται μέσω του νευρώνα εξόδου, και είναι αυτός της εκτίμησης του διανύσματος κλίσης ο οποίος χρειάζεται για την κατάσταση κατά την οποία το δίκτυο εξελίσσεται προς τα πίσω. Οι κρυφοί νευρώνες, δρουν ως ανιχνευτές χαρακτηριστικών (feature detectors) παίζοντας κρίσιμο ρόλο στη λειτουργία των MLP δικτύων. Καθώς προχωράει η διαδικασία μάθησης, οι κρυφοί νευρώνες είναι αυτοί που σταδιακά “ανακαλύπτουν” τα ξεχωριστά χαρακτηριστικά των προτύπων, κάτι το οποίο διαφοροποιεί τα MLP δίκτυα από το δίκτυο Perceptron του Rosenblatt. (Haykin, 2010)

Μπορούμε να παρατηρήσουμε στο παρακάτω σχήμα ένα δίκτυο MLP με τρεις (3) νευρώνες στο στρώμα εισόδου, ένα (1) κρυμμένο στρώμα με τέσσερις (4) νευρώνες επεξεργασίας και ένα στρώμα εξόδου με τρεις (3) νευρώνες.



Εικόνα 9MLP

Προκειμένου να γίνει υλοποίηση ενός δικτύου MLP, θα πρέπει αρχικά να γίνει αποθήκευση των διαθέσιμων προτύπων σε συνδυασμό με τις αντίστοιχες επιθυμητές εξόδους τους. Και σε αυτή την περίπτωση - όπως στα δίκτυα Perceptron και Adaline, μπορεί να προκύψουν μη αξιόπιστα αποτελέσματα αν υπάρχουν τυχόν λάθη στα πρότυπα ή στις επιθυμητές εξόδους. Το δεύτερο βήμα είναι η αρχικοποίηση του πλήθους των κόμβων που θα χρησιμοποιηθούν στο δίκτυο. Το πλήθος αυτό, δεν προκύπτει από κάποιο μαθηματικό τύπο, και χρειάζεται πειραματισμό. (Τσούλος, Γλαβάς & Γαβριλής, 2006) Επόμενο βήμα, είναι η αρχικοποίηση του διανύσματος βαρών που θα χρησιμοποιηθεί με σκοπό την κατάλληλη εκπαίδευση του δικτύου καθώς και του ρυθμού μάθησης. Οι τρόποι αρχικοποίησης των τιμών των βαρών, είναι και αυτοί ίδιοι με αυτούς κατά την υλοποίηση των δικτύων Perceptron και Adaline. Θα πρέπει πρώτα να είναι γνωστό το απαραίτητο πλήθος των τιμών, το οποίο προκύπτει από τον παρακάτω τύπο:

$$size(w) = (d + 2) \cdot H_s$$

Τύπος 4 Υπολογισμός πλήθους βαρών για ένα MLP δίκτυο

⁸ Το πλήθος των βαρών προκύπτει από το γινόμενο του πλήθους των χαρακτηριστικών (d) αυξημένο κατά 2, με το πλήθος των νευρώνων (H) των κρυφών στρωμάτων

Ο πιο απλός τρόπος αρχικοποίησης είναι η εκχώρηση σε όλες τις τιμές, μία τιμή είτε ίση με το μηδέν (0) είτε με την μονάδα (1). Εναλλακτικά, μπορεί να γίνει αρχικοποίηση των τιμών σε τυχαίες τιμές στο κλειστό διάστημα [0,1]. Η αρχικοποίηση του ρυθμού μάθησης είναι και αυτή ίδια με αυτή των δικτύων Perceptron και Adaline, δηλαδή σε μία μικρή θετική δεκαδική τιμή, συνήθως στο κλειστό διάστημα [0,1]. Συχνά γίνεται χρήση της τιμής 0.1. (Τσούλος, *Τεχνητά Νευρωνικά Δίκτυα (μέρος δεύτερο)*, χ.χ.) Όπως στα Adaline δίκτυα, έτσι και στα MLP, απαιτείται ο ορισμός και μίας τιμής ως κατώφλι προκειμένου να τερματίσει ο αλγόριθμος σε περίπτωση που το σφάλμα εκπαίδευσης έχει χαμηλότερη τιμή από αυτή. Στη συνέχεια γίνεται μία επαναληπτική διαδικασία για κάθε διαθέσιμο πρότυπο. Έτσι, για κάθε πρότυπο γίνεται υπολογισμός της εξόδου του δικτύου χρησιμοποιώντας τις διαθέσιμες τιμές των βαρών. Σε αυτή την περίπτωση δικτύου γίνεται ο υπολογισμός της εξόδου με τη χρήση κατάλληλων συναρτήσεων ενεργοποίησης μέσω της προώθησης του προτύπου από το ένα στρώμα του δικτύου στο αμέσως επόμενο. Ο πιο απλός μαθηματικός τρόπος αναπαράστασης της εξόδου ενός MLP δικτύου είναι ο παρακάτω:

$$N(\vec{x}, \vec{w}) = \sum_{i=1}^H w_{(d+2)i-(d+1)} \sigma\left(\sum_{j=1}^d x_j w_{(d+2)i-(d+1)+j} + w_{(d+2)i}\right)$$

Τύπος 5 Υπολογισμός εξόδου ενός MLP δικτύου⁹ (Τσούλος, Τζάλλας & Τσαλικάκης, 2018)

Εφόσον είναι γνωστή η έξοδος του δικτύου για το πρότυπο, το επόμενο βήμα είναι ο υπολογισμός του σήματος σφάλματος με σκοπό την εκπαίδευση των βαρών. Η πιο διαδεδομένη μέθοδος για τον υπολογισμό αυτό, είναι η Back Propagation. Θα γίνει ανάλυση της, στο υποκεφάλαιο 2.4.

Η διαδικασία αυτή τερματίζει, όταν πληροί μία από τις ακόλουθες συνθήκες. Είτε όταν για μία ολόκληρη εποχή, δεν υπάρχει μεταβολή στις τιμές των βαρών παρά την εκπαίδευση τους, είτε όταν το σφάλμα εκπαίδευσης έχει τιμή μικρότερη από την τιμή κατωφλίου που έχουμε ορίσει. Εναλλακτικά, η διαδικασία μπορεί να τερματίσει μετά από κάποιο πεπερασμένο αριθμό επαναλήψεων - εφόσον αυτός έχει οριστεί στην αρχή της υλοποίησης. (Βλ. Ψευδοκώδικα 3 στο Παράρτημα Β)

⁹ Όπου H είναι το πλήθος των νευρώνων των κρυφών στρωμάτων και ο αριθμός d αφορά το πλήθος των χαρακτηριστικών. Τα βάρη συμβολίζονται με \vec{w} και τα πρότυπα με \vec{x} . Ο συμβολισμός $w_{(d+2)i}$ αφορά την πόλωση κάθε μονάδας επεξεργασίας. (Τσούλος, *Τεχνητά Νευρωνικά Δίκτυα MLP*, χ.χ.)

2.4 Μέθοδος Back Propagation

Η εκπαίδευση ενός δικτύου MLP είναι η διαδικασία ρύθμισης των συναπτικών βαρών του έτσι ώστε να ικανοποιείται κάποιο κριτήριο καταλληλότητας. Άλλωστε, αυτός είναι και ο στόχος της εκπαίδευσης σε οποιοδήποτε νευρωνικό δίκτυο όπως για παράδειγμα σε ένα απλό Perceptron. Αυτό που κάνει την εκπαίδευση ενός MLP δικτύου πολύ πιο ενδιαφέρουσα είναι η ιδιότητα του καθολικού προσεγγιστή, σύμφωνα με την οποία - όπως αναφέρθηκε στο υποκεφάλαιο 2.3, αν έχουμε το κατάλληλο σε μέγεθος δίκτυο μπορούμε να το εκπαιδύσουμε να μάθει οποιαδήποτε συνάρτηση εμείς επιθυμούμε. Κυριότερος εκπρόσωπος των αλγορίθμων εκπαίδευσης MLP είναι αυτός της Οπισθοδιάδοσης Σφάλματος (Back Propagation). Η χρήση του όρου αυτού εμφανίζεται στο προσκήνιο μετά από το 1985, όταν ο όρος απέκτησε ευρεία απήχηση χάρη στο σημαντικό βιβλίο με τίτλο “Parallel Distributed Processing”. Ο αλγόριθμος αυτός, βασίζεται στο γενικευμένο κανόνα Δέλτα (generalized Delta rule), ο οποίος επιτρέπει τον καθορισμό του ποσοστού του συνολικού σφάλματος που αντιστοιχεί στα βάρη του κάθε νευρώνα, ακόμη και εκείνων που ανήκουν στα κρυφά στρώματα, για τους οποίους η επιθυμητή έξοδος δεν είναι γνωστή. Αξίζει να σημειωθεί πως δε θα μπορούσε να βασιστεί στον απλό κανόνα Δέλτα, διότι αυτός δεν είναι σε θέση να λειτουργήσει σε κρυφά στρώματα.

Βασικό χαρακτηριστικό της μεθόδου αυτής είναι η ύπαρξη στόχων, όπως ακριβώς και στο απλό δίκτυο Perceptron. Ιδανικό σενάριο θεωρούμε την περίπτωση που υπάρχει πλήρη ταύτιση ανάμεσα στις πραγματικές και επιθυμητές εξόδους. Ωστόσο, αυτό μπορεί να μην είναι απολύτως εφικτό, και για αυτόν τον λόγο επιζητούμε τη βέλτιστη προσέγγιση της επιθυμητής κατάστασης χρησιμοποιώντας ένα κριτήριο κόστους. Κλασικό παράδειγμα κριτηρίου κόστους αποτελεί το τετραγωνικό σφάλμα, το οποίο είδαμε να χρησιμοποιείται και στα δίκτυα Adaline. Για αυτόν τον λόγο θεωρείται πως ο Back Propagation περιλαμβάνει τον αλγόριθμο LMS σαν ειδική περίπτωση. Η κεντρική ιδέα λοιπόν της μεθόδου Back Propagation, είναι ο καθορισμός και η ελαχιστοποίηση του σφάλματος σε κάθε νευρώνα για κάθε στρώμα του ΤΝΔ από το στρώμα εξόδου προς το στρώμα εισόδου. Ο καθορισμός όπως είδαμε, γίνεται με την χρήση του τετραγωνικού σφάλματος. Η ελαχιστοποίηση γίνεται με βάση τον υπολογισμό της παραγώγου του σφάλματος. Αφού γίνουν οι παραπάνω υπολογισμοί,

προκειμένου να γίνει η εκπαίδευση των βαρών, θα πρέπει να γίνει χρήση μίας ακόμη μεθόδου που ονομάζεται Gradient Descent. Θα παρουσιαστεί αναλυτικά στο υποκεφάλαιο 2.5. Αξίζει να αναφερθεί βέβαια πως ο αλγόριθμος αυτός έχει αρκετά προβλήματα όπως η αργή σύγκλιση αλλά και το ότι δεν υπάρχει εγγύηση γενίκευσης. (Soltanian, Ahmadizar, Tab, & Tsoulos, 2013)· (Βλαχάβας et al., 2006)· (Διαμαντάρας, 2007)· (Δούνιας και Καραμπότσης, 2023)

2.5 Μέθοδος Gradient Descent

Η μέθοδος Gradient Descent ή αλλιώς μέθοδος κατάβασης δυναμικού, παίρνει και το όνομα “Γενικευμένος Κανόνας Δέλτα” σύμφωνα με τον Διαμαντάρα (2007). Είναι μία από τις πιο δημοφιλείς μεθόδους για την εκτέλεση βελτιστοποίησης (Sebastian, χ.χ.) και είναι μακράν ο πιο συνηθισμένος τρόπος για τη βελτιστοποίηση ΤΝΔ. Είναι μια κλασική μέθοδος εύρεσης της ελάχιστης τιμής μίας συνάρτησης κόστους κάποιου πλήθους μεταβλητών. Μπορεί επίσης να αποδοθεί ως επικλινής κάθοδος ή επικλινής κατάδυση. (Θεοδωρίδης και Κουτρούμπας, 2012) Αν επιθυμούμε την εύρεση του μέγιστου σημείου της συνάρτησης, τότε η ίδια μέθοδος μπορεί να χρησιμοποιηθεί πάλι με μόνη διαφορά το πρόσημο της κατεύθυνσης αναζήτησης του βέλτιστου σημείου. Στη περίπτωση αυτή, η μέθοδος καλείται ανάβαση δυναμικού (Gradient Ascent). (Διαμαντάρας, 2007)

Υπάρχουν τρεις παραλλαγές αυτής της μεθόδου, οι οποίες διαφέρουν ανάλογα με τον όγκο των δεδομένων προς χρήση για τον υπολογισμό της κλίσης της αντικειμενικής συνάρτησης. Η επιλογή που πρέπει να γίνει, εξαρτάται από την ακρίβεια ενημέρωσης των παραμέτρων και του χρόνου κατά τον οποίο, θα πραγματοποιηθεί μία ενημέρωση.

Σε αυτή την πτυχιακή εργασία, θα γίνει χρήση της μεθόδου Stochastic Gradient Descent κατά την οποία ο υπολογισμός γίνεται σε κάθε πέρασμα κάποιου προτύπου. Έτσι, κατά την εκπαίδευση βαρών σε ένα δίκτυο MLP με τη χρήση των μεθόδων Back Propagation και Gradient Descent (Βλ. Ψευδοκώδικα 4 στο Παράρτημα Β) γίνεται χρήση του παρακάτω μαθηματικού τύπου:

$$w = w - \beta \frac{\partial(E)}{\partial w} \quad 10$$

Τύπος 6 Ενημέρωση των βαρών με χρήση Gradient Descent

2.6 Εφαρμογές νευρωνικών δικτύων

Τα ΤΝΔ βρίσκουν πληθώρα εφαρμογών σε διαφορετικούς τομείς της καθημερινότητας μας. Στον ιατρικό τομέα, αξιοποιούνται για τη διάγνωση παθήσεων, την ανάλυση καρκινικών κυττάρων και την ανάπτυξη νέων φαρμάκων, επιταχύνοντας σημαντικά τη χρονοβόρα διαδικασία ελέγχου των δεδομένων ιατρικών εξετάσεων. Ωστόσο, η αναγκαιότητα των ιατρών παραμένει αναλλοίωτη. Επιπλέον, χρησιμοποιούνται για τη μείωση του χρόνου νοσηλείας αλλά και του νοσοκομειακού κόστους.

Στον στρατιωτικό τομέα, εφαρμόζονται για την κατηγοριοποίηση εικόνων προερχόμενων από συσκευές radar ή sonar με σκοπό την ανίχνευση και την παρακολούθηση αντικειμένων στον αέρα, στη γη, ακόμη και στο νερό. Επιπλέον, χρησιμοποιούνται στην πλοήγηση όπλων, στην ανίχνευση στόχων, στην ψηφιακή επεξεργασία σημάτων, στην εξαγωγή χαρακτηριστικών αλλά και στην αναγνώριση σημάτων, ακόμη και εικόνων.

Στον τραπεζικό τομέα, ενσωματώνονται για τον έλεγχο της γνησιότητας των υπογραφών αλλά και των τραπεζογραμματίων. Επίσης, χρησιμοποιούνται σε συστήματα αξιολόγησης αιτήσεων δανειοδότησης αλλά και ανάγνωσης παραστατικών όπως είναι οι επιταγές.

Στον τεχνολογικό τομέα, εφαρμόζονται τεχνικές για την αναγνώριση ήχου - ακόμη και φωνής, εικόνας καθώς και γραπτού κειμένου είτε αυτό είναι χειρόγραφο είτε τυπωμένο. Στην ηλεκτρονική, συμβάλλουν στην πρόβλεψη ακολουθίας κωδικών, στη μορφοποίηση ολοκληρωμένων κυκλωμάτων, στον έλεγχο διεργασιών, στη μηχανική όραση, ακόμη και στη σύνθεση φωνής. Στις τηλεπικοινωνίες, υιοθετούνται για τη συμπίεση εικόνας καθώς και δεδομένων, για τη μετάφραση σε πραγματικό χρόνο, ακόμη και σε συστήματα επεξεργασίας πληρωμών. Στην ασφάλεια εφαρμόζονται για τον εντοπισμό κίνησης και την ταύτιση δακτυλικών αποτυπωμάτων.

¹⁰ Τα βάρη ενημερώνονται αφαιρώντας το ρυθμό μάθησης β πολλαπλασιασμένο με την παράγωγο της συνάρτησης σφάλματος ως προς το βάρος.

Τέλος, στον τομέα της πρόβλεψης, ενσωματώνονται εφαρμογές πρόβλεψης ισοτιμίας νομισμάτων, τιμών μετοχών, πωλήσεων, αλλά και χρηματιστηριακές προβλέψεις. Στη γεωργία γίνεται πρόβλεψη παραγωγής κυρίως με χρήση δορυφορικών εικόνων, ενώ στη μετεωρολογία, γίνεται πρόβλεψη του καιρού. Επίσης μπορούμε να συμπεριλάβουμε την αυτόματη πλοήγηση στον χώρο καθώς απαιτεί πρόβλεψη της κίνησης στον χώρο μέσω δυναμικών εμποδίων. Μπορούμε να αναφέρουμε επιπλέον τη λήψη αποφάσεων, διότι περιλαμβάνει την πρόβλεψη των συνεπειών διάφορων επιλογών. Τέλος, αξίζει να αναφερθεί η πρόβλεψη οικονομικών μεγεθών όπως η πτώχευση και οι πωλήσεις.

Άλλες κατηγορίες προβλημάτων στα οποία έχουν εφαρμογή τα ΤΝΔ είναι η προσέγγιση συνάρτησης, η εξόρυξη δεδομένων, ο εντοπισμός φυσικών πόρων όπως το πετρέλαιο και το φυσικό αέριο, η ρομποτική, η διάγνωση βλαβών αλλά και ο χρονοπρογραμματισμός. Επιπλέον, εφαρμόζονται με σκοπό την επεξεργασία σήματος, τον έλεγχο και τη λύση συνηθισμένων και μερικών διαφορικών εξισώσεων. Τα τελευταία χρόνια τα ΤΝΔ χρησιμοποιούνται σε συστήματα ελέγχου που βασίζονται στην ασαφή λογική (neufuzzy systems) με κύριο ρόλο τον υπολογισμό της συνάρτησης συγγένειας.

Πρέπει να τονιστεί ότι αν και ένα εκπαιδευόμενο ΤΝΔ μπορεί να αναγνωρίσει δεδομένα τα οποία δεν έχει δει ποτέ του, αυτό δεν συμβαίνει στην περίπτωση που τα δεδομένα δεν ανήκουν στην ίδια κατηγορία προβλημάτων για την οποία έχει εκπαιδευτεί. Δεν υπάρχει κάποιο ΤΝΔ με σκοπό την αντιμετώπιση ετερογενών προβλημάτων. Υπάρχουν όμως προγραμματιστικά περιβάλλοντα τα οποία επιτρέπουν τη δόμηση ΤΝΔ για διάφορους σκοπούς. Αυτό είναι πολύ σημαντικό καθώς στο ερώτημα “Ποιο είδος ΤΝΔ είναι κατάλληλο για δεδομένο πρόβλημα” η απάντηση δεν είναι ούτε εύκολη, ούτε μοναδική. Κάθε πρόβλημα είναι συνήθως ειδική περίπτωση και η αντιμετώπισή του με ΤΝΔ απαιτεί μελέτη και πειραματισμό. (Βλαχάβας et al., 2006)· (Λυκοθανάσης και Κουτσημητρόπουλος, 2021)· (Ντούνης, 2024)· (Τσούλος, Λαγαρής & Λύκας, 2002)· (Τσούλος, *Τεχνητά Νευρωνικά Δίκτυα*, χ.χ.)

3. Γενετικοί αλγόριθμοι

3.1 Ιστορική Αναδρομή

Η συνεχώς αυξανόμενη ανάγκη για παραγωγή λογισμικού με στόχο την αποδοτική εκμετάλλευση των τεράστιων δυνατοτήτων του υλικού καθώς και οι περιορισμένες δυνατότητες των κλασσικών μεθόδων αναζήτησης και βελτιστοποίησης, ήταν η βασική αιτία έτσι ώστε να αναπτυχθούν μέθοδοι επίλυσης προβλημάτων βασισμένων στις αρχές της Γενετικής εξέλιξης και της κληρονομικότητας. Το 1950 έγινε η πρώτη εμφάνιση των γενετικών αλγορίθμων, καθώς κάποιοι βιολόγοι επιστήμονες αποφάσισαν να χρησιμοποιήσουν υπολογιστές στην προσπάθεια τους να προσομοιώσουν πολύπλοκα βιολογικά συστήματα.

Στις αρχές του 1970, ο μηχανισμός της φυσικής επιλογής προσέλκυσε ιδιαίτερα τον John Holland, ο οποίος φαντάστηκε πως κάποιες λειτουργίες της φύσης, θα μπορούσαν να χρησιμοποιηθούν αποτελεσματικά σε αλγορίθμους για υπολογιστές, έτσι ώστε να προκύψουν αποδοτικές τεχνικές με σκοπό την επίλυση δύσκολων προβλημάτων. Τέτοια προβλήματα αφορούν τον εντοπισμό σταθερών μοριακών διαμορφώσεων, την πρωτεϊνική αναδίπλωση (protein folding), την προσαρμογή δεδομένων σε θεωρητικά και παραμετρικά μοντέλα αλλά και τη σχεδίαση VLSI κυκλωμάτων. Αποτέλεσμα αυτού, ήταν η συστηματική ανάπτυξη μιας τεχνικής αναζήτησης και βελτιστοποίησης, αυτής των γενετικών αλγορίθμων από τον ίδιο τον Holland, αλλά και από τους συναδέλφους του και τους μαθητές του, στο πανεπιστήμιο του Michigan το 1975. Την ίδια χρονολογία εκδόθηκε και το βιβλίο “Adaptation in Natural and Artificial Systems”, από το Πανεπιστήμιο του Michigan Press, το οποίο περιγράφει τις βασικές έννοιες των γενετικών αλγορίθμων. (Μακρόπουλος και Ευστρατιάδης, 2018) Η αρχική ορολογία που εισήγαγε ο Holland ήταν γενετικό σχέδιο, ωστόσο ο όρος γενετικός αλγόριθμος υιοθετήθηκε αργότερα για να τονίσει τον κεντρικό ρόλο του υπολογισμού στον καθορισμό και την υλοποίηση των σχεδίων.

Οι γενετικοί αλγόριθμοι είναι στοχαστικοί εξελικτικοί αλγόριθμοι και παρόλο που δεν υπάρχει εγγύηση για την εύρεση της βέλτιστης λύσης, είναι ικανοί να βρουν

μία αρκετά καλή λύση σε εύλογο χρονικό διάστημα. Θεωρούνται μέθοδος καθολικής βελτιστοποίησης και μάλιστα είναι μία από τις περισσότερο ευέλικτες μεθόδους. Παρόλο που ανήκουν στην κατηγορία των στοχαστικών αλγορίθμων, υπάρχει αρκετά μεγάλη διαφορά με τους αλγορίθμους στους οποίους εφαρμόζονται τυχαίες μέθοδοι αναζήτησης και βελτιστοποίησης. Αυτό συμβαίνει διότι οι γενετικοί αλγόριθμοι είναι σε θέση να συνδυάζουν στοιχεία και από άμεσες αναζητήσεις. Για τον λόγο αυτό, θεωρούνται πιο εύρωστοι από τις υπάρχουσες μεθόδους.

Ο βασικός μηχανισμός των γενετικών αλγορίθμων είναι εμπνευσμένος από τη θεωρία της εξέλιξης των ειδών (evolution of species) που εισήγαγε ο Δαρβίνος με το βιβλίο “Η καταγωγή των ειδών” το 1958. Η θεωρία αυτή προκάλεσε έντονες αντιδράσεις όταν δημοσιεύτηκε, καθώς προσέφερε μία νέα επιστημονική προοπτική για την προέλευση των ειδών. Οι αντιδράσεις αυτές, συνεχίζουν να υπάρχουν μέχρι και σήμερα. Στην ουσία, οι γενετικοί αλγόριθμοι λειτουργούν με έναν τρόπο που μιμείται τη βιολογική εξέλιξη.

Ένα χαρακτηριστικό παράδειγμα χρήσης των γενετικών αλγορίθμων αναπτύχθηκε το 1993 με σκοπό τη μάθηση Boolean εννοιών που αναπαρίστανται από ένα διαζευκτικό (disjunctive) σύνολο προτασιακών κανόνων. Το σύστημα αυτό ονομάζεται GABIL και χρησιμοποιήθηκαν τεχνητά προβλήματα μάθησης προκειμένου να εξεταστεί η ακρίβεια του. Εφαρμόστηκε όμως και σε πραγματικό πρόβλημα, τη διάγνωση καρκίνου του μαστού. (Βλαχάβας et al., 2006)· (Καμπουρλάζος και Παπακώστας, 2015)· (Λυκοθανάσης, 2001)· (Λυκοθανάσης, 2022-2023)· (Τσούλος, Γενετικοί Αλγόριθμοι, χ.χ.)

3.2 Δομή και Λειτουργία των Γενετικών Αλγορίθμων

Ένας πληθυσμός από υποψήφιες λύσεις - γνωστές και ως χρωμοσώματα περιέχουν χαρακτηριστικά - γνωστά και ως γονίδια. Η αρχικοποίηση του πληθυσμού εξαρτάται από διάφορους παράγοντες όπως είναι το ίδιο το πρόβλημα προς επίλυση καθώς και η επιλεγόμενη μέθοδος κωδικοποίησης που θα δούμε στο υποκεφάλαιο 3.2. Τα χρωμοσώματα αυτά, υπόκειται σε διαδικασίες αναπαραγωγής και μετάλλαξης, με στόχο την εύρεση της καλύτερης λύσης. Το περιβάλλον είναι σε θέση να καθορίσει πόσους απογόνους μπορεί να αφήσει κάθε μέλος του πληθυσμού. Η βελτίωση των συνθηκών διαβίωσης αυξάνει τις πιθανότητες να εμφανιστούν απόγονοι που κληρονομούν τα

χαρακτηριστικά των γονέων τους έτσι ώστε ένα χαρακτηριστικό να επικρατήσει και σε επόμενες γενιές. Η μετάλλαξη συμβαίνει σε πολύ αραιά χρονικά διαστήματα και προκαλείται είτε από γενετικούς είτε από περιβαλλοντολογικούς παράγοντες. Η κύρια διαφορά των γενετικών αλγορίθμων με την Δαρβινική θεωρία είναι πως στον αλγόριθμο πρέπει να υπάρχει κάποιος τερματισμός. Τις περισσότερες φορές εξαρτάται από το ίδιο το πρόβλημα, αλλά συνήθως επιτυγχάνεται μετά από κάποιο αριθμό προκαθορισμένων επαναλήψεων/εποχών ή μετά από κάποιο συγκεκριμένο ποσοστό βελτίωσης είτε για το καλύτερο άτομο είτε για ολόκληρο τον πληθυσμό, συγκριτικά με τις προηγούμενες γενιές. (Τσούλος, Γενετικοί Αλγόριθμοι. Εισαγωγικές έννοιες, 2008)

3.3 Μέθοδοι κωδικοποίησης

Στην κλασσική προσέγγιση των γενετικών αλγορίθμων, κάθε χρωμόσωμα αναπαρίσταται με μία συμβολοσειρά ενός πεπερασμένου αλφαβήτου. Προκειμένου να αποφασίσουμε ποια κωδικοποίηση θα χρησιμοποιηθεί στον γενετικό αλγόριθμο, πρέπει να υπολογίσουμε έναν κρίσιμο παράγοντα. Αυτός είναι το να μην αλλοιωθεί η αρχική διαθέσιμη πληροφορία.

Η πρώτη κωδικοποίηση που χρησιμοποιήθηκε ήταν η δυαδική (binary encoding) κατά την οποία χρησιμοποιείται το δυαδικό αλφάβητο. Σε αυτή την περίπτωση, οι συμβολοσειρές ονομάζονται και δυαδικές συμβολοσειρές (bit-strings). (Καμπουρλάζος και Παπακώστας, 2015) Έχουν αναπτυχθεί αρκετές κωδικοποιήσεις, οι οποίες διαμορφώνονται ανάλογα με το κάθε πρόβλημα προς εξέταση. Καμία κωδικοποίηση δε γίνεται να είναι αποτελεσματική για όλα τα προβλήματα, ενώ ενδέχεται κάποιο πρόβλημα να είναι αποτελεσματικό με περισσότερες από μία. (Λυκοθανάσης, 2001)

Σε αυτή την πτυχιακή εργασία, γίνεται χρήση της άμεσης κωδικοποίησης δεκαδικών αριθμών, μία από τις πιο κοινές προς χρήση λύσεις. Στην κωδικοποίηση αυτή, κάθε γονίδιο αποτελείται από κάποιον δεκαδικό αριθμό - όπως λέει και το όνομα της. Ανάλογα με το πρόβλημα, μπορούν να δοθούν και κάποια όρια στις τιμές που μπορεί να έχει κάθε γονίδιο. Η ακρίβεια των τιμών, εξαρτάται από την απεικόνιση των δεκαδικών αριθμών που κάνει ο εκάστοτε υπολογιστής και από το πλήθος των δεκαδικών ψηφίων που μας ικανοποιεί για το πρόβλημα. Αυτό όμως σημαίνει πως μπορεί να προκύψουν και προβλήματα μεταφερσιμότητας. (Τσούλος, Γενετικοί Αλγόριθμοι. Κωδικοποίηση χρωμοσωμάτων, 2008)

3.4 Συνάρτηση καταλληλότητας

Η συνάρτηση καταλληλότητας λειτουργεί ως το κριτήριο για την αξιολόγηση των χρωμοσωμάτων. Ως είσοδος της συνάρτησης είναι ένα χρωμόσωμα και ως έξοδος, ένας αριθμός που υποδηλώνει το βαθμό καταλληλότητας του. Συνήθως ο αριθμός αυτός ανήκει στο σύνολο των πραγματικών αριθμών στο κλειστό διάστημα $[0,1]$ όπου η τιμή 1 υποδηλώνει ότι το χρωμόσωμα ικανοποιεί όλες τις προϋποθέσεις του προβλήματος και αποτελεί αποδεκτή λύση, ενώ η τιμή 0 υποδηλώνει μη αποδεκτή λύση.

Ο γενικός κανόνας για την κατασκευή της συνάρτησης είναι πως πρέπει να αντικατοπτρίζει την αξία του χρωμοσώματος. Ωστόσο υπάρχουν περιπτώσεις στις οποίες δεν έχει τόσο νόημα η αξία αυτή, αλλά χρειαζόμαστε μία τιμή για το πόσο κοντά βρισκόμαστε σε μία επιθυμητή λύση. Υπάρχουν δύο (2) επιλογές για τη χρήση της συνάρτησης καταλληλότητας. Η ακριβής και η προσεγγιστική συνάρτηση καταλληλότητας (approximate fitness function). Το κρίσιμο θέμα είναι η επιθυμητή ακρίβεια της συνάρτησης αυτής και το υπολογιστικό κόστος που θεωρείται αποδεκτό για μία συνάρτηση που ενδεχομένως δίνει άριστα αποτελέσματα. Για παράδειγμα, σε ένα πρόβλημα μπορεί να υπάρχουν διαθέσιμες δύο συναρτήσεις καταλληλότητας, μία ακριβής και μία προσεγγιστική. Αν η ακριβής συνάρτηση έχει πολύ μεγαλύτερο χρόνο υπολογισμού από την προσεγγιστική, ενδεχομένως να είναι αποδοτικότερη η προσεγγιστική συνάρτηση καθώς η διαδικασία αξιολόγησης εκτελείται πάρα πολλές φορές και έτσι επιτρέπει την πραγματοποίηση πολύ μεγαλύτερης ποσότητας υπολογισμών στο ίδιο χρονικό διάστημα, συγκριτικά με την ακριβή συνάρτηση καταλληλότητας. (Βλαχάβας et al., 2006)

Σε αυτή την πτυχιακή εργασία, επειδή ο γενετικός αλγόριθμος θα χρησιμοποιηθεί με σκοπό να αρχικοποιήσει τα βάρη ενός MLP δικτύου, ως συνάρτηση καταλληλότητας θα θεωρήσουμε το μέσο τετραγωνικό σφάλμα των εξόδων των προτύπων, που προκύπτει από τον παρακάτω τύπο:

$$MSE = \frac{1}{M} \sum_{i=1}^M (y^p - d^p)^2$$

11

Τύπος 7 Μέσο Τετραγωνικό Σφάλμα

3.5 Γενετικοί τελεστές

Οι ζωντανοί οργανισμοί εξελίσσονται χρησιμοποιώντας δύο (2) βασικούς τελεστές ή αλλιώς μηχανισμούς. Ο πρώτος αναφέρεται ως φυσική επιλογή (natural selection) και καθορίζει ποια άτομα του πληθυσμού θα επιβιώσουν με σκοπό την αναπαραγωγή έτσι ώστε να μεταφέρουν τα γονίδια (genes) τους στην επόμενη γενιά. Ο δεύτερος τελεστής αναφέρεται ως διασταύρωση (crossover) και σκοπός του είναι η εξασφάλιση της ανάμειξης των γονιδίων των γονέων που έχουν επιλεγεί, προκειμένου να δημιουργηθούν νέοι απόγονοι οι οποίοι θα αποτελούνται από συνδυαστικά χαρακτηριστικά των γονέων τους.

Η διαδικασία επιλογής των χρωμοσωμάτων - γονέων σχετίζεται με την απόδοση πιθανοτήτων επιλογής προς αναπαραγωγή στα μέλη ενός πληθυσμού. Κατά τη διαδικασία αυτή, κάποια χρωμοσώματα με υψηλή τιμή στη συνάρτηση καταλληλότητας ενδέχεται να επιλεγούν προς αναπαραγωγή περισσότερες από μία φορές, ενώ κάποια άλλα με χαμηλή καταλληλότητα ενδέχεται να μην επιλεγούν καθόλου. Ωστόσο, υπολογιστικός πειραματισμός έχει δείξει πως η επιλογή των γονέων δεν πρέπει να βασίζεται αποκλειστικά στα καλύτερα χρωμοσώματα διότι κάτι τέτοιο μπορεί να αποκλείσει τη βέλτιστη λύση. Μία καλή πρακτική είναι να επιβιώνουν και κάποια λιγότερα καλά χρωμοσώματα, ώστε να διασφαλίζεται η μετάδοση της γενετικής τους πληροφορίας.

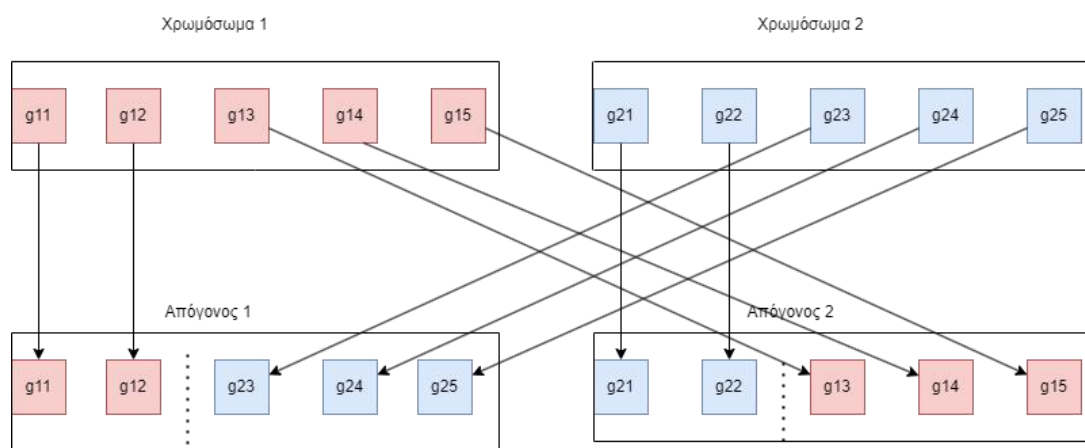
Σε αυτή την πτυχιακή εργασία έχει επιλεγεί ο πιο συνηθισμένος στοχαστικός τελεστής επιλογής, που είναι η επιλογή της ρουλέτας. Σε αυτή τη μέθοδο επιλέγονται χρωμοσώματα και με υψηλή αλλά και με χαμηλή τιμή στη συνάρτηση καταλληλότητας. Ωστόσο, όσο καλύτερη καταλληλότητα έχουν τα χρωμοσώματα, τόσο περισσότερες

¹¹Το μέσο τετραγωνικό σφάλμα (MSE) προκύπτει από τον μέσο όρο των τετραγώνων των διαφορών μεταξύ της επιθυμητής (d^p) και της πραγματικής (y^p) εξόδου για κάθε ένα από τα M πρότυπα.

πιθανότητες έχουν να επιλεγούν, πιθανώς και παραπάνω από μία φορά. Μετά την ολοκλήρωση της παραπάνω διαδικασίας, δημιουργούνται ζευγάρια με τυχαίο τρόπο έτσι ώστε με τη χρήση του τελεστή της διασταύρωσης να προκύψει η νέα γενιά.

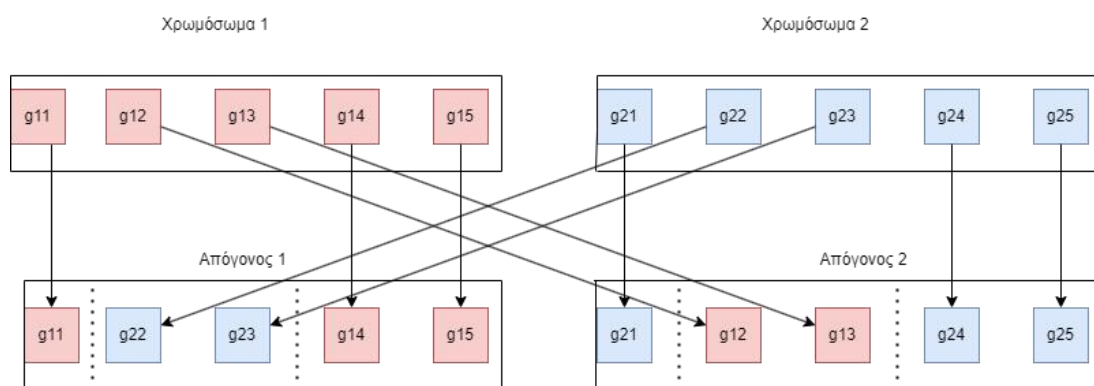
Η διασταύρωση των χρωμοσωμάτων δεν είναι απαραίτητο ότι θα γίνει σε όλα τα ζευγάρια που έχουν δημιουργηθεί από τον τελεστή της φυσικής επιλογής. Η εφαρμογή της εξαρτάται από ένα προκαθορισμένο ποσοστό κατωφλίου κατά το οποίο αν μία τυχαία τιμή είναι μικρότερη από αυτό, τότε εφαρμόζεται αυτός ο τελεστής. Σε αντίθετη περίπτωση, τα γονίδια των γονέων μεταφέρονται αυτούσια στην επόμενη γενιά. Υπάρχουν διάφορες μέθοδοι διασταύρωσης και η επιλογή της καταλληλότερης εξαρτάται τόσο από το είδος του προβλήματος, όσο και το είδος της εφαρμοζόμενης κωδικοποίησης.

Σε αυτή την πτυχιακή εργασία έχει γίνει υλοποίηση δύο διαφορετικών μεθόδων. Η πρώτη είναι αυτή της διασταύρωσης ενός σημείου (single-point crossover) κατά την οποία επιλέγεται μία τυχαία ακέραια τιμή από την τιμή 0 μέχρι την τιμή του πλήθους των γονιδίων που περιέχονται στα χρωμοσώματα. Έτσι, ο πρώτος απόγονος θα κληρονομήσει από την πρώτη θέση μέχρι την τυχαία θέση που έχει επιλεχθεί τα γονίδια του πρώτου χρωμοσώματος, και για τα υπόλοιπα γονίδια θα κληρονομήσει αυτά του δεύτερου χρωμοσώματος που έχουν επιλεχθεί για αναπαραγωγή. Αντίστοιχα, ο δεύτερος απόγονος θα κληρονομήσει από την πρώτη θέση μέχρι την τυχαία θέση που έχει επιλεχθεί τα γονίδια του δεύτερου χρωμοσώματος, και για τα υπόλοιπα γονίδια θα κληρονομήσει αυτά του πρώτου χρωμοσώματος που έχουν επιλεχθεί για αναπαραγωγή.



Εικόνα 10 Single point crossover

Η δεύτερη μέθοδος είναι αυτή της διασταύρωσης δύο σημείων (two-point crossover) κατά την οποία χρειαζόμαστε δύο (2) σημεία διασταύρωσης - όπως λέει και το όνομα της, προκειμένου να τεμαχιστούν τα χρωμοσώματα - γονείς για αμοιβαία ανταλλαγή προκειμένου να έχουμε απογόνους. Η ιδέα πίσω από αυτές τις μεθόδους είναι πως τα τμήματα ενός χρωμοσώματος που συμβάλλουν περισσότερο στην απόδοση του, ενδέχεται να μην είναι γειτονικά. Έτσι, ανταλλάσσονται τμήματα των χρωμοσωμάτων με την προσδοκία βελτίωσης της απόδοσης ενός χρωμοσώματος. Υπολογιστικά πειράματα έχουν επαληθεύσει συχνά αυτή την προσδοκία.



Εικόνα 11 Double point crossover

Ένας ακόμη τελεστής των γενετικών αλγορίθμων είναι αυτός της μετάλλαξης (mutation). Πρόκειται για μία διαδικασία που εμφανίζεται στους ζωντανούς οργανισμούς βάσει της οποίας παρατηρείται μία τυχαία αλλαγή σε κάποια χρωμοσώματα υπό ειδικές συνθήκες. Στους γενετικούς τελεστές η εφαρμογή της μετάλλαξης γίνεται με τυχειότητα τόσο για το χρωμόσωμα όσο και για το γονίδιο στο οποίο θα εφαρμοστεί. Σκοπός αυτού του τελεστή είναι η διασφάλιση της ποικιλομορφίας του πληθυσμού και αυτό επιτυγχάνεται με το να επιτρέπει τον τυχαίο σχηματισμό χρωμοσωμάτων. Έτσι, συμβάλλει στη διεύρυνση της εξερεύνησης του χώρου των λύσεων.

Τέλος, ένας ακόμη τελεστής είναι ο ελιτισμός (elitism), κατά τον οποίο ένας προκαθορισμένος αριθμός χρωμοσωμάτων που έχουν κριθεί ως ‘πλέον κατάλληλα’, θα μεταφερθεί αυτούσιος στον νέο πληθυσμό. Έτσι, εξασφαλίζεται πως κάποια “πλέον

κατάλληλα” χρωμοσώματα θα δε εξαφανιστούν ως αποτέλεσμα της εφαρμογής των τελεστών που αναφέρθηκαν παραπάνω.

Αξίζει να αναφερθεί σε αυτό το σημείο πως προκειμένου να γίνει η υλοποίηση ενός γενετικού αλγορίθμου δεν είναι υποχρεωτική η ύπαρξη κάποιου από τους τελεστές που αναφέρθηκαν, καθώς δεν είναι υποχρεωτική και η σειρά αυτή. (Βλαχάβας et al., 2006)· (Καμπουρλάζος & Παπακώστας, 2015)

3.6 Εφαρμογές γενετικών αλγορίθμων

Οι γενετικοί αλγόριθμοι, όπως και τα ΤΝΔ, αποτελούν έναν εύκολο τρόπο επίλυσης προβλημάτων με μεγάλη δυνατότητα προσαρμογής. Σε αυτό το κεφάλαιο θα αναλύσουμε μερικές από τις πιο αντιπροσωπευτικές εφαρμογές των γενετικών αλγορίθμων ξεκινώντας με την επεξεργασία εικόνων η οποία αποτελεί τη βάση για την μηχανική μάθηση.

Μια σημαντική εφαρμογή των γενετικών αλγορίθμων είναι στον οικονομικό και εμπορικό τομέα όπου χρησιμοποιούνται για την πρόβλεψη οικονομικών μεγεθών, την ανίχνευση απάτης σε ηλεκτρονικές συναλλαγές και την εκτίμηση της αξίας ακινήτων.

Επιπλέον, οι γενετικοί αλγόριθμοι βρίσκουν εφαρμογή στον μαθηματικό τομέα με την εύρεση της μέγιστης τιμής σε αριθμητικές συναρτήσεις, κάτι που δεν είναι εύκολη υπόθεση ιδιαίτερα για συναρτήσεις πολλών μεταβλητών καθώς συχνά εμφανίζουν ασυνέχειες και θόρυβο. Επίσης, βρίσκουν εφαρμογή στην προσέγγιση συναρτήσεων με πιο γνωστή εφαρμογή, αυτή των συστημάτων ταξινόμησης.

Ένα από τα πιο γνωστά προβλήματα που βρίσκουν εφαρμογή οι γενετικοί αλγόριθμοι είναι αυτό του πλανόδιου πωλητή και αφορά την συνδυαστική βελτιστοποίηση. Σε αυτή την περίπτωση, οι γενετικοί αλγόριθμοι μπορούν να δώσουν αρκετές λύσεις κοντά στη βέλτιστη.

Οι γενετικοί αλγόριθμοι μπορούν να χρησιμοποιηθούν και στη σχεδίαση κατασκευών και εξαρτημάτων, όπως γέφυρες ή μηχανολογικά εξαρτήματα με στόχο όχι μόνο την εύρεση μίας λύσης αλλά και τη βελτιστοποίηση της. Οι αλγόριθμοι συχνά δοκιμάζουν συνδυασμούς και ιδέες που ο ανθρώπινος νους δε θα δοκίμαζε ποτέ, δίνοντας ενίοτε αρκετά πρωτότυπα αποτελέσματα.

Επιπλέον, έχουν εφαρμοστεί επιτυχώς σε προβλήματα βελτιστοποίησης όπως η δρομολόγηση καλωδίων (wire routing), ο χρονοπρογραμματισμός (scheduling), ο

προσαρμοστικός έλεγχος (adaptive control), τα παίγνια (game playing), η γνωστική μοντελοποίηση (cognitive modeling), προβλήματα εφοδιαστικής (logistics), προβλήματα βέλτιστου ελέγχου και η βελτιστοποίηση ερωτημάτων σε βάσεις δεδομένων. Κάποιες ακόμη εφαρμογές των γενετικών αλγορίθμων που αξίζει να αναφερθούν είναι η σύνθεση μουσικής, η ρομποτική, η αεροναυπηγική καθώς και η επεξεργασία φυσικής γλώσσας.

Οι γενετικοί αλγόριθμοι βρίσκουν επίσης εφαρμογή σε πολιτικές και οικονομικές αναλύσεις. Επιπλέον έχουν βρει εφαρμογή στη βιομηχανία για βιομηχανικούς ελέγχους, τη ρύθμιση ηλεκτρικού φορτίου, τη ρομποτική καθώς και σε συστήματα οχημάτων. Τέλος, η αναγνώριση προτύπων καθώς και η ασφάλεια αφορούν εφαρμογές των γενετικών αλγορίθμων. (Βλαχάβας et al., 2006)· (Τσούλος, Γενετικοί Αλγόριθμοι. Εισαγωγικές έννοιες, 2008)

4. Μέθοδος - αποτελέσματα

4.1 Τα dataset που χρησιμοποιήθηκαν

Για την υλοποίηση της εφαρμογής, χρησιμοποιήθηκαν 3 διαφορετικά σύνολα δεδομένων τα οποία αντλήθηκαν από τον ιστότοπο <https://archive.ics.uci.edu/> του πανεπιστημίου της Καλιφόρνια στο Ιρβίν. Και τα τρία (3) σύνολα αποτελούν παραδείγματα classification.

Το πρώτο με όνομα ιονόσφαιρα, περιέχει δεδομένα ραντάρ από τη βάση δεδομένων ιονόσφαιρας του Johns Hopkins. Αποτελείται από 351 πρότυπα, το κάθε ένα από τα οποία περιλαμβάνει 34 χαρακτηριστικά. Συνολικά, υπάρχουν 2 διαφορετικές κατηγορίες εξόδου σε αυτό το σύνολο δεδομένων. Στα πειράματα που πραγματοποιήθηκαν χρησιμοποιήθηκαν 175 παραδείγματα για εκπαίδευση και τα υπόλοιπα για δοκιμή.

Το δεύτερο σύνολο δεδομένων, αφορά δεδομένα χημικής ανάλυσης κρασιών με σκοπό τον προσδιορισμό της προέλευσής τους. Αποτελείται από 178 πρότυπα, το κάθε ένα από τα οποία περιλαμβάνει 13 χαρακτηριστικά. Συνολικά, υπάρχουν 3 διαφορετικές κατηγορίες εξόδου σε αυτό το σύνολο δεδομένων. Στα πειράματα που πραγματοποιήθηκαν χρησιμοποιήθηκαν 89 παραδείγματα για εκπαίδευση και τα υπόλοιπα για δοκιμή.

Το τρίτο σύνολο δεδομένων, αφορά λουλούδια και προέρχεται από τον Fisher, το 1936. Αυτό αποτελεί και ένα από τα παλαιότερα δεδομένα που χρησιμοποιήθηκαν σε προβλήματα ταξινόμησης. Αποτελείται από 150 πρότυπα, το κάθε ένα από τα οποία περιλαμβάνει 4 χαρακτηριστικά. Συνολικά, υπάρχουν 3 διαφορετικές κατηγορίες εξόδου σε αυτό το σύνολο δεδομένων. Στα πειράματα που πραγματοποιήθηκαν χρησιμοποιήθηκαν 75 παραδείγματα για εκπαίδευση και τα υπόλοιπα για δοκιμή. (Τσούλος et al., 2018)· (Τσούλος, Γλαβάς & Γαβρίλης, 2006)

4.2 Η προτεινόμενη μέθοδος

Σε αυτή την πτυχιακή εργασία έγινε περιγραφή και ανάλυση τριών (3) πρόσθιας τροφοδότησης δικτύων με στόχο την ταξινόμηση προτύπων χρησιμοποιώντας μάθηση με επίβλεψη. Το πρώτο δίκτυο που παρουσιάστηκε ήταν το Perceptron του Rosenblatt. Δίκτυο, που μπορεί με απόλυτη επιτυχία να διαχωρίσει γραμμικά διαχωρίσιμα δεδομένα που ανήκουν σε μόνο δύο (2) κατηγορίες. Το επόμενο δίκτυο που παρουσιάστηκε, ήταν το Adaline. Μοιάζει πολύ με το δίκτυο Perceptron διότι και τα δύο αποτελούνται από μόνο έναν (1) νευρώνα. Αν και δεν μπορεί να ταξινομήσει με απόλυτη επιτυχία γραμμικά διαχωρίσιμα δεδομένα δύο (2) κατηγοριών όπως το Perceptron, μπορεί να βρει μία πολύ καλή λύση ακόμη και αν τα δεδομένα είναι σχεδόν γραμμικά διαχωρίσιμα.

Με σκοπό την επίτευξη αποδοτικότερων δικτύων και την διεύρυνση των δυνατοτήτων των παραπάνω δικτύων, προτείνεται η χρήση του MLP δικτύου για την ταξινόμηση προτύπων. Ακόμη και με την χρήση της βηματικής συνάρτησης το δίκτυο αυτό έχει περισσότερες δυνατότητες από τα δίκτυα Perceptron και Adaline. Προτείνεται όμως η χρήση της σιγμοειδής συνάρτησης στα κρυφά στρώματα, όπως αναφέρθηκε και στο υποκεφάλαιο 2.3. Προκειμένου να γίνει η εκπαίδευση του δικτύου, προτείνεται η χρήση της μεθόδου Back Propagation ώστε να υπολογίσει το σφάλμα από το στρώμα εξόδου προς το στρώμα εισόδου και η χρήση της μεθόδου Gradient Descent ώστε να ενημερώσει τα συναπτικά βάρη του δικτύου αξιοποιώντας το παραπάνω σφάλμα.

Για ακόμη καλύτερη βελτιστοποίηση του δικτύου MLP προτείνεται η χρήση γενετικού αλγορίθμου με σκοπό την βέλτιστη αρχικοποίηση των βαρών του. Προκειμένου να υπολογιστεί η συνάρτηση καταλληλότητας κάθε χρωμοσώματος του πληθυσμού, μπορεί να γίνει χρήση του υπολογισμού της εξόδου κάθε προτύπου στο MLP δίκτυο.

4.3 Πειραματικά αποτελέσματα

Πραγματοποιήθηκαν είκοσι (20) πειράματα σε κάθε ένα από τα σύνολα δεδομένων που αναφέρθηκαν στο υποκεφάλαιο 4.1. Αρχικά έγινε χρήση του MLP δικτύου με δέκα (10) νευρώνες επεξεργασίας στο κρυφό στρώμα, διακόσιες (200) εποχές και σταθερό ρυθμό μάθησης μηδέν κόμμα μηδέν ένα (0.01). Τα αποτελέσματα καταγράφονται στην πρώτη στήλη κάθε πίνακα. Έπειτα, με σκοπό την βέλτιστη αρχικοποίηση των βαρών του MLP δικτύου έγινε χρήση γενετικού αλγορίθμου με παραμέτρους: τριακόσια (300) χρωμοσώματα ως πληθυσμό, ενενήντα τέσσερα τοις εκατό (94%) διασταύρωση, τέσσερα τοις εκατό (4%) ελιτισμό και δύο τοις εκατό (2%) μετάλλαξη. Στην δεύτερη στήλη καταγράφονται τα αποτελέσματα με τη χρήση της μεθόδου διασταύρωσης single-point ενώ στην τελευταία στήλη, με τη χρήση της μεθόδου διασταύρωσης double-point. Να σημειωθεί πως παρόλο που γίνεται χρήση δεκαπέντε (15) δεκαδικών ψηφίων μετά την υποδιαστολή, εμφανίζονται τα πέντε (5) πρώτα στους παρακάτω πίνακες αποτελεσμάτων, χωρίς καμία στρογγυλοποίηση για λόγους απλότητας.

Τα πειράματα πραγματοποιήθηκαν σε υπολογιστή με επεξεργαστή: 13th Gen Intel(R) Core(TM) i7-13700KF 3.40 GHz, RAM: 32.0 GB, κάρτα γραφικών: GPU AMD Radeon RX 7900 XTX. Το λειτουργικό σύστημα ήταν Windows 11 Pro με έκδοση 22H2, και το περιβάλλον Visual Studio Code με τη χρήση της γλώσσας Java και πιο συγκεκριμένα με την έκδοση java 17.0.8 2023-07-18 LTS.

BP	Single-point crossover	Double-point crossover
7.74869	1.907784	0.89809
13.71115	2.41953	0.39756
17.75755	1.25256	1.29451
16.41501	0.78875	1.76277
13.24487	2.00387	1.34744
16.07421	1.97110	1.48579
13.26669	1.82355	1.62641

16.07693	2.81496	1.94768
6.90382	1.92123	1.01876
17.09095	1.71807	1.75389
14.50936	2.90616	1.11121
5.89010	1.95215	2.36562
14.74056	1.49959	1.62510
4.84435	2.01504	1.94277
12.88190	1.72210	1.18778
4.79301	1.10819	1.92600
16.72229	1.73459	2.43169
11.42903	0.93695	1.15734
8.61992	1.77078	0.54219
10.35135	0.31416	0.80394

Πίνακας 1 Πειραματικά αποτελέσματα για το σύνολο δεδομένων της ιονόσφαιρας

BP	Single-point crossover	Double-point crossover
4.90988	0.62325	0.97439
7.14211	0.60453	0.69211
7.22742	0.68314	0.78524
1.50005	0.60838	0.64656
9.80620	0.86305	0.63467
4.58333	0.96033	0.60510
1.35153	0.79040	0.68748
9.32508	0.71159	0.63920
11.38206	0.67699	0.86342
10.43945	0.90438	0.88029
8.15229	0.74329	1.25042
1.81182	0.98451	0.68706
10.42921	0.60520	0.60710

6.08924	0.93093	0.71819
8.24961	0.62053	0.94795
14.70852	0.64116	0.71236
5.78872	0.60496	1.02384
2.48255	0.62408	0.60776
1.06195	0.70357	0.77189
8.77446	0.69192	0.80280

Πίνακας 2 Πειραματικά αποτελέσματα για το σύνολο δεδομένων των κρασιών

BP	Single-point crossover	Double-point crossover
1.04214	1.32235	1.29045
0.67908	1.35738	1.30974
0.97917	1.38882	1.34568
2.26112	1.31485	1.40487
0.84418	1.34315	1.35725
2.00550	1.35801	1.18248
0.73951	1.22924	1.23436
0.91474	1.41196	1.40055
1.07699	1.37166	1.37418
0.67434	1.35490	1.29205
0.65711	1.47118	1.31311
1.15227	1.44930	1.32181
0.96275	1.19065	1.18826
0.66698	1.39615	1.35641
0.53458	1.16596	1.32631
1.09824	1.19869	1.32055
0.81970	1.43958	1.20944
1.38965	1.01474	1.33450

0.61668	1.46624	1.40728
0.88461	1.37053	1.39802

Πίνακας 3 Πειραματικά αποτελέσματα για το σύνολο δεδομένων των λουλουδιών

5. Συμπεράσματα

Στην παρούσα πτυχιακή έγινε υλοποίηση ενός MLP δικτύου με τη χρήση των μεθόδων Back Propagation και Gradient Descent με σκοπό την εκπαίδευση των παραμέτρων - βαρών του. Έπειτα έγινε χρήση γενετικού αλγορίθμου με σκοπό την βέλτιστη αρχικοποίηση των βαρών του δικτύου, με δύο διαφορετικούς τρόπους διασταύρωσης. Single και Double point crossover.

Σύμφωνα και με τα πειραματικά αποτελέσματα του υποκεφαλαίου 4.3 ο μέσος όρος του σφάλματος ταξινόμησης για το σύνολο δεδομένων της ιονόσφαιρας είναι 12.153587 με τη χρήση του Back Propagation. Αν γίνει χρήση του γενετικού αλγορίθμου με single-point crossover τότε ο μέσος όρος γίνεται 1.7290557 και με τη χρήση double-point crossover γίνεται 1.431327.

Ο μέσος όρος του σφάλματος ταξινόμησης για το σύνολο δεδομένων των κρασιών είναι 6.760774 με τη χρήση Back Propagation. Αν γίνει χρήση του γενετικού αλγορίθμου με single-point crossover τότε ο μέσος όρος γίνεται 0.7288095 και με τη χρήση double-point crossover γίνεται 0.7768915.

Ο μέσος όρος του σφάλματος ταξινόμησης για το σύνολο δεδομένων των λουλουδιών είναι 0.999967 με τη χρήση Back Propagation. Αν γίνει χρήση του γενετικού αλγορίθμου με single-point crossover τότε ο μέσος όρος γίνεται 1.330767 και με τη χρήση double-point crossover γίνεται 1.318365.

Όσο μικρότερη είναι η τιμή του μέσου όρου του σφάλματος ταξινόμησης, τόσο καλύτερα σημαίνει πως εκπαιδεύτηκε το δίκτυο. Έτσι, μπορούμε να παρατηρήσουμε πως η χρήση γενετικού αλγορίθμου με σκοπό την αρχικοποίηση των βαρών ενός MLP δικτύου κάνει αρκετά καλή δουλειά. Σε δεύτερη φάση μπορούμε να παρατηρήσουμε πως υπάρχει μικρή βελτιστοποίηση συγκριτικά με τους δύο διαφορετικούς τρόπους διασταύρωσης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Anderson, D. & McNeill, G. (1992). *Artificial neural networks technology: A DACS State-of-the-Art Report*. New York: Kaman Sciences Corporation.
- Chiarandini, M. (χ.χ.). *Machine Learning: Linear Regression and Neural Networks*. Ανακτήθηκε στις 08/06/2024 από <https://www.imada.sdu.dk/u/rolf/Edu/DM534/E16/DM534-marco.pdf>
- Haykin, S. (2010). *Νευρωνικά δίκτυα και μηχανική μάθηση*. 3η έκδοση. Αθήνα: Εκδόσεις Παπασωτηρίου.
- Rosenblatt, F. (1958). *THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN*. Ανακτήθηκε στις 11/09/2024 από <https://www.ling.upenn.edu/courses/cogs501/Rosenblatt1958.pdf>
- Soltanian, K., Ahmadizar, F., Tab, F.A., & Tsoulos, G.I. (2013). *Artificial Neural Networks Generation using Grammatical Evolution*. Ανακτήθηκε στις 25/09/2024 από https://www.researchgate.net/profile/Khabat-Soltanian/publication/255992665_Artificial_Neural_Networks_Generation_using_Grammatical_Evolution/links/00b7d5215027e73c08000000/Artificial-Neural-Networks-Generation-using-Grammatical-Evolution.pdf
- Βλαχάβας, Ι. (2013). *Τεχνητή Νοημοσύνη: Τεχνητά Νευρωνικά Δίκτυα*. Θεσσαλονίκη. Ανακτήθηκε στις 26/08/2024 από <https://opencourses.auth.gr/modules/document/file.php/OCRS118/%CE%A3%CE%B7%CE%BC%CE%B5%CE%B9%CF%8E%CF%83%CE%B5%CE%B9%CF%82%20%CE%BC%CE%B1%CE%B8%CE%AE%CE%BC%CE%B1%CF%84%CE%BF%CF%82/%CE%9511%20-%20%CE%A4%CE%B5%CF%87%CE%BD%CE%B7%CF%84%CE%AC%20%CE%9D%CE%B5%CF%85%CF%81%CF%89%CE%BD%CE%B9%CE%BA%CE%AC%20%CE%94%CE%AF%CE%BA%CF%84%CF%85%CE%B1.pdf>
- Βλαχάβας, Ι., Κεφάλας, Π., Βασιλειάδης, Ν., Κόκκορας, Φ., & Σακελλαρίου, Η., (2006). *Τεχνητή Νοημοσύνη*. 3η έκδοση. Θεσσαλονίκη: Εκδόσεις Πανεπιστημίου Μακεδονίας.

- Γεωργούλη, Κ. (2015). *Τεχνητή Νοημοσύνη - Μία εισαγωγική προσέγγιση*. Αθήνα: Κάλλιπος Ανακτήθηκε στις 13/07/2024 από <https://repository.kallipos.gr/handle/11419/3381>
- Διαμαντάρας, Κ. (2007). *Τεχνητά Νευρωνικά Δίκτυα*. Αθήνα: Εκδόσεις Κλειδάριθμος.
- Δούνιας, Γ. & Καραμπότσης, Ε. (2023). *Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα*. Ανακτήθηκε στις 15/06/2024 από <https://mde-lab.aegean.gr/images/stories/docs/ANNs-intro.pdf>
- Ηλιάδης, Λ. (χ.χ.). *Νευρωνικά Δίκτυα*. Ανακτήθηκε στις 01/07/2024 από <https://slideplayer.gr/slide/7018453/>
- Θεοδωρίδης, Σ. & Κουτρούμπας, Κ. (2012). *Αναγνώριση προτύπων*. Κύπρος: Εκδόσεις Πασχαλίδη.
- Καμπουρλάζος, Β.Γ. & Παπακώστας, Γ.Α. (2015). *Εισαγωγή στην υπολογιστική νοημοσύνη*. Αθήνα: Κάλλιπος.
- Κίτρου, Κ. (χ.χ.). *Τι είναι μηχανική μάθηση; (Machine Learning)*. Ανακτήθηκε στις 08/06/2024 από
- Λύκας, Α. (χ.χ.). *Βασικές αρχές εκπαίδευσης ΤΝΔ: Το perceptron*. Ανακτήθηκε στις 10/06/2024 από <https://www.cs.uoi.gr/~arly/courses/nn/slides/K2.pdf>
- Λυκοθανάσης, Σ. (χ.χ.). *Υπολογιστική Νοημοσύνη Ι*. Ανακτήθηκε στις 11/09/2024 από <https://eclass.upatras.gr/courses/CEID1041/>
- Λυκοθανάσης, Σ. (2001). *Γενετικοί Αλγόριθμοι και Εφαρμογές*. Τόμος Γ'. Ανακτήθηκε στις 18/09/2024 από <https://psifiakoskosmos.wordpress.com/wp-content/uploads/2009/12/likothanasis1.pdf>
- Λυκοθανάσης, Σ. (2022 - 2023). *Εισαγωγή στους Γενετικούς Αλγόριθμους*. Ανακτήθηκε στις 19/09/2024 από <https://eclass.upatras.gr/modules/document/index.php?course=CEID1104&download=/55e57bea4FY1/633d8fa4j260/633d9587LR5y.pdf>
- Λυκοθανάσης, Σ. & Κουτσομητρόπουλος, Δ. (2021). *Υπολογιστική νοημοσύνη και βαθιά μάθηση*. Εκδόσεις Κάλλιπος Ανακτήθηκε στις 13/07/2024 από <https://repository.kallipos.gr/handle/11419/9117>
- Μακρόπουλος, Χ. & Ευστρατιάδης, Α. (2018). *Γενετικοί Αλγόριθμοι*. Ανακτήθηκε από https://mycourses.ntua.gr/courses/PSTGR1031/document/2017-18/Lecture_GeneticAlgorithms_Hydroinf_2018.pdf

- Ντούνης, Α. (2024). *Βασικές Αρχές Υπολογιστικής Νοημοσύνης*. Εκδόσεις Κάλλιπος Ανακτήθηκε στις 12/07/2024 από <https://repository.kallipos.gr/handle/11419/13341>
- Πούλος, Μ. (2015). *Σημασιολογική Επεξεργασία της Πληροφορίας*. Εκδόσεις Κάλλιπος Ανακτήθηκε στις 12/07/2024 από <https://repository.kallipos.gr/handle/11419/2854>
- Τσίπουρας, Μ. (2015 - 2016). *Τεχνητή Νοημοσύνη*. Ανακτήθηκε στις 15/06/2024 από <https://eclass.teiwm.gr/modules/document/file.php/DNG170/%CE%94%CE%B9%CE%B1%CE%BB%CE%AD%CE%BE%CE%B5%CE%B9%CF%82/%CE%9C%CE%A%CE%B8%CE%B7%CE%BC%CE%B1%2010%20-%20%CE%9D%CE%B5%CF%85%CF%81%CF%8E%CE%BD%CE%B1%CF%82%20Perceptron.pdf>
- Τσούλος, Γ.Ι. (χ.χ.). *Τεχνητά Νευρωνικά Δίκτυα*. Ανακτήθηκε στις 08/06/2024 από <https://www.dit.uoi.gr/e-class/modules/document/file.php/249/%CE%94%CE%99%CE%91%CE%9B%CE%95%CE%9E%CE%95%CE%99%CE%A3/lecture1.pdf>
- Τσούλος, Γ.Ι., (χ.χ.). *Γενετικοί Αλγόριθμοι*.
- Τσούλος, Γ.Ι., (2008). *Γενετικοί Αλγόριθμοι. Εισαγωγικές έννοιες*.
- Τσούλος, Γ.Ι., (2008). *Γενετικοί Αλγόριθμοι. Κωδικοποίηση χρωμοσωμάτων*.
- Τσούλος, Γ.Ι., (2019). *Τεχνητά Νευρωνικά Δίκτυα (μέρος δεύτερο)*. Ανακτήθηκε στις 08/06/2024 από <https://www.dit.uoi.gr/e-class/modules/document/index.php?course=249&openDir=/5c7d2f2cLKRF>
- Τσούλος, Γ.Ι., (χ.χ.). *Τεχνητά Νευρωνικά Δίκτυα MLP*.
- Τσούλος, Γ.Ι., Τζάλλας, Α. & Τσαλικάκης, Δ. (2018). *Evolutionary based weight decaying method for neural network training*. Ανακτήθηκε στις 28/08/2024 από <https://link.springer.com/article/10.1007/s11063-017-9660-0>
- Τσούλος, Γ.Ι., Γλαβάς, Ε. & Γαβρίλης, Δ. (2006). *Neural network construction using grammatical evolution*. Ανακτήθηκε στις 25/09/2024 από https://www.researchgate.net/publication/4215174_Neural_network_construction_using_grammatical_evolution
- Τσούλος, Γ.Ι., Λαγαράς, Ι. & Λύκας, Α. (2002). *Piecewise Neural Networks for Function Approximation, Cast in a Form Suitable for Parallel Computation*. Ανακτήθηκε στις 25/09/2024 από https://www.researchgate.net/publication/221238982_Piecewise_Neural_Networks_for_Function_Approximation_Cast_in_a_Form_Suitable_for_Parallel_Computation

- Τσούλος, Γ.Ι., Γλαβάς, Ε. & Γαβριλής, Δ. (2006). *Neural network construction using grammatical evolution*. Ανακτήθηκε στις 25/09/2024 από https://www.researchgate.net/publication/4215174_Neural_network_construction_using_grammatical_evolution

ΠΑΡΑΡΤΗΜΑ

ΠΑΡΑΡΤΗΜΑ Α: Κώδικας σε Octave για δημιουργία γραφικών παραστάσεων

Στο παρόν παράρτημα, παρατίθενται οι κώδικες που χρησιμοποιήθηκαν σε Octave με σκοπό τη δημιουργία των γραφικών παραστάσεων που εμφανίζονται στο υποκεφάλαιο 1.3 της πτυχιακής εργασίας. Οι κώδικες έχουν δοκιμαστεί και επαληθευτεί στη διεύθυνση: <https://www.mycompiler.io/new/octave>

Κώδικας 1: Δημιουργία γραμμικής συνάρτησης

```
x = -10:0.1:10; % Ορισμός των τιμών του x από -10 έως 10 με βήμα 0.1
y = x;
plot(x, y);
xlabel('x');
ylabel('f(x)');
title('Linear function');
grid on;
```

Κώδικας 2: Δημιουργία βηματικής συνάρτησης 0/1

```
x = -10:0.1:10; % Ορισμός των τιμών του x από -10 έως 10 με βήμα 0.1
y = x >= 0;
plot(x, y);
xlabel('x');
ylabel('f(x)');
title('Step Function');
grid on;
```

Κώδικας 3: Δημιουργία βηματικής συνάρτησης -1/1

```
x = -10:0.1:10;  
%Δημιουργία ενός πίνακα με όνομα y και μέγεθος ίδιο με τον πίνακα x.  
%Αρχικοποίηση όλων των τιμών του πίνακα y με τιμή ίση με τη μονάδα  
y = ones(size(x));  
%Για τις τιμές του x που είναι μικρότερες από το 0, ορίζουμε τιμή -1  
y(x < 0) = -1;  
plot(x, y);  
xlabel('x');  
ylabel('f(x)');  
title('Step Function (-1/1)');  
grid on;
```

Κώδικας 4: Δημιουργία σιγμοειδής συνάρτησης

```
x = -10:0.1:10; % Ορισμός των τιμών του x από -10 έως 10 με βήμα 0.1  
y = 1 ./ (1 + exp(-x));  
plot(x, y);  
xlabel('x');  
ylabel('f(x)');  
title('Sigmoid Function');  
grid on;
```

ΠΑΡΑΡΤΗΜΑ Β: Αλγόριθμοι με την μορφή ψευδοκώδικα

Στο παρόν παράρτημα, παρατίθενται σε μορφή ψευδοκώδικα αλγόριθμοι για τα δίκτυα που αναλύθηκαν σε αυτή την πτυχιακή εργασία.

Ψευδοκώδικας 1: Perceptron

ΕΙΣΟΔΟΙ:

Τα πρότυπα εισόδου με τις αντίστοιχες επιθυμητές εξόδους

ΜΕΘΟΔΟΣ:

1. Υπολογισμός του πλήθους των τιμών που χρειαζόμαστε για τα βάρη
2. Αρχικοποίηση των βαρών που θα χρησιμοποιηθούν.
3. Αρχικοποίηση του ρυθμού μάθησης
4. Αρχικοποίηση του μέγιστου αριθμού επαναλήψεων - προαιρετικό
5. Όσο δεν ικανοποιείται η συνθήκη τερματισμού
 - a) Για κάθε πρότυπο
 - i. Υπολογισμός της εξόδου του δικτύου
 - ii. Αν (πραγματική έξοδος \neq επιθυμητή έξοδος)
 1. Εκπαίδευση των βαρών.

ΤΕΡΜΑΤΙΣΜΟΣ:

Η διαδικασία τερματίζει όταν ισχύει μία από τις ακόλουθες συνθήκες:

1. Αν δεν έχει χρειαστεί εκπαίδευση των βαρών για μία ολόκληρη εποχή ή
2. Αν έχουμε ξεπεράσει τον μέγιστο αριθμό επαναλήψεων - εφόσον αυτός έχει οριστεί.

ΕΞΟΔΟΣ:

Τα ενημερωμένα βάρη

Ψευδοκώδικας 2: Adaline

ΕΙΣΟΔΟΙ:

Τα πρότυπα εισόδου με τις αντίστοιχες επιθυμητές εξόδους

ΜΕΘΟΔΟΣ:

1. Υπολογισμός του πλήθους των τιμών που χρειαζόμαστε για τα βάρη
2. Αρχικοποίηση των βαρών που θα χρησιμοποιηθούν.
3. Αρχικοποίηση του ρυθμού μάθησης
4. Αρχικοποίηση κατωφλίου για το σφάλμα εκπαίδευσης
5. Αρχικοποίηση του μέγιστου αριθμού επαναλήψεων - προαιρετικό
6. Όσο δεν ικανοποιείται η συνθήκη τερματισμού
 - a) Για κάθε πρότυπο
 - i. Υπολογισμός της εξόδου του δικτύου
 - ii. Εκπαίδευση των βαρών.
 - b) Στο τέλος κάθε εποχής, υπολογισμός του σφάλματος εκπαίδευσης

ΤΕΡΜΑΤΙΣΜΟΣ:

Η διαδικασία τερματίζει όταν ισχύει μία από τις ακόλουθες συνθήκες:

1. Αν δεν έχει γίνει μεταβολή των βαρών - παρά την εκπαίδευση τους σε μία ολόκληρη εποχή ή
2. Αν το σφάλμα εκπαίδευσης είναι μικρότερο από το κατώφλι που έχουμε ορίσει ή
3. Αν έχουμε ξεπεράσει τον μέγιστο αριθμό επαναλήψεων - εφόσον αυτός έχει οριστεί.

ΕΞΟΔΟΣ:

Τα ενημερωμένα βάρη

Ψευδοκώδικας 3: MLP με χρήση των μεθόδων Back Propagation και Gradient Descent

ΕΙΣΟΔΟΙ:

Τα πρότυπα εισόδου με τις αντίστοιχες επιθυμητές εξόδους

ΜΕΘΟΔΟΣ:

1. Αρχικοποίηση του πλήθους των κόμβων που θα χρησιμοποιηθούν
2. Υπολογισμός του πλήθους των τιμών που χρειαζόμαστε για τα βάρη
3. Αρχικοποίηση των βαρών που θα χρησιμοποιηθούν
4. Αρχικοποίηση του ρυθμού μάθησης
5. Αρχικοποίηση κατωφλιού για το σφάλμα εκπαίδευσης
6. Αρχικοποίηση του μέγιστου αριθμού επαναλήψεων - προαιρετικό
7. Όσο δεν ικανοποιείται η συνθήκη τερματισμού
 - a) Για κάθε πρότυπο
 - i. Υπολογισμός της εξόδου του δικτύου
 - ii. Υπολογισμός του σφάλματος του νευρωνικού δικτύου
 - iii. Εκπαίδευση βαρών

ΤΕΡΜΑΤΙΣΜΟΣ:

Η διαδικασία τερματίζει όταν ισχύει μία από τις ακόλουθες συνθήκες:

1. Αν το σφάλμα εκπαίδευσης μειωθεί κάτω από το προκαθορισμένο κατώφλι ή
2. Αν έχουμε ξεπεράσει τον μέγιστο αριθμό επαναλήψεων - εφόσον αυτός έχει οριστεί.

ΕΞΟΔΟΣ:

Τα ενημερωμένα βάρη

Ψευδοκώδικας 4: Γενετικός Αλγόριθμος με τη χρήση ρουλέτας

ΜΕΘΟΔΟΣ:

1. Ορισμός ορίων - εφόσον χρειάζονται
2. Αρχικοποίηση του πληθυσμού
3. Όσο δεν ικανοποιείται η συνθήκη τερματισμού
 - a) Υπολογισμός ικανότητας (fitness)
 - b) Κανονικοποίηση των τιμών ικανότητας
 - c) Ταξινόμηση χρωμοσωμάτων
 - d) Υπολογισμός του cumulative sum
 - e) Ελιτισμός
 - f) Επιλογή με τη μέθοδο της ρουλέτας
 - g) Διασταύρωση
 - h) Μετάλλαξη

ΤΕΡΜΑΤΙΣΜΟΣ:

Η διαδικασία τερματίζει όταν ισχύει μία από τις ακόλουθες συνθήκες:

1. Αν επιτευχθεί κάποιο συγκεκριμένο ποσοστό βελτίωσης είτε για το καλύτερο άτομο είτε για ολόκληρο τον πληθυσμό, συγκριτικά με τις προηγούμενες γενιές
2. Αν έχουμε ξεπεράσει τον μέγιστο αριθμό επαναλήψεων - εφόσον αυτός έχει οριστεί.

ΕΞΟΔΟΣ:

Το καλύτερο χρωμόσωμα

