

Εισαγωγή στα νευρωνικά δίκτυα

Ιωάννης Γ. Τσούλος

Τμήμα Πληροφορικής και τηλεπικοινωνιών
Πανεπιστήμιο Ιωαννίνων

2025

Περίληψη παρουσίασης

- 1 Εισαγωγικές έννοιες
 - Βασικά στοιχεία
 - Συναρτήσεις ενεργοποίησης
- 2 Νευρωνικά δίκτυα Perceptron
 - Βασικά στοιχεία
 - Εκτεταμένο παράδειγμα μάθησης
- 3 Τα νευρωνικά δίκτυα Adaline

Ορισμός

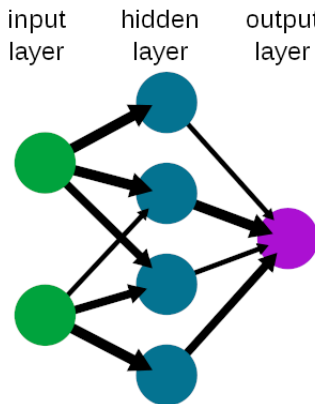
- 1 Νευρωνικό δίκτυο ονομάζεται ένα κύκλωμα διασυνδεδεμένων μονάδων επεξεργασίας που ονομάζουμε **Νευρώνες**.
- 2 Στους υπολογιστές είναι ένα υπολογιστικό μοντέλο που χρησιμοποιείται για την επίλυση κάποιου υπολογιστικών προβλημάτων.
- 3 Το νευρωνικό δίκτυο ονομάζεται δίκτυο καθώς αποτελείται απο υπολογιστικούς κόμβους που συνδέονται μεταξύ τους.
- 4 Κάθε υπολογιστικός κόμβος δέχεται ένα σύνολο αριθμητικών εισόδων (από άλλους νευρώνες είτε από κάποια άλλη είσοδος), εκτελεί έναν υπολογισμό με βάση αυτές τις εισόδους και παράγει μία έξοδο.

Είδη νευρώνων

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

Σχηματικό παράδειγμα

A simple neural network



Παράδειγμα εξίσωσης

- 1 Αν θέλουμε να εκφράσουμε την έξοδο ενός νευρώνα επεξεργασίας k θα μπορούσαμε να χρησιμοποιήσουμε την παρακάτω εξίσωση:

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

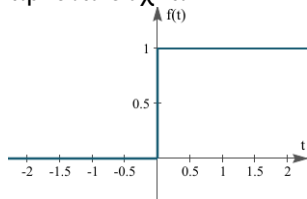
- 2 x_i είναι η i είσοδος από τις d που έχει το πρόβλημα
- 3 w_i είναι το βάρος της διασύνδεσης με την i είσοδο
- 4 θ_k είναι η πόλωση για τον νευρώνα k . Η τιμή της πόλωσης συνήθως είναι ανεξάρτητη από το πρόβλημα.
- 5 Η συνάρτηση $f(x)$ είναι η συνάρτηση ενεργοποίησης και στην βιβλιογραφία χρησιμοποιείται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

Βηματική συνάρτηση

Είναι η πιο απλή συνάρτηση ενεργοποίησης και δίνεται από τον τύπο

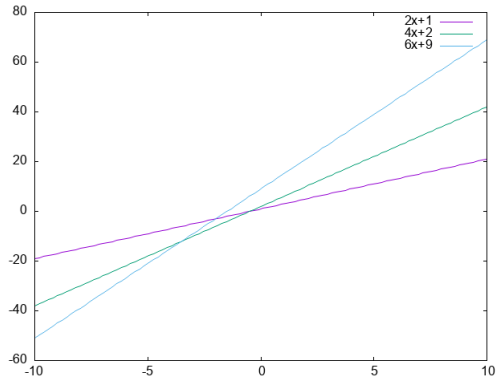
$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Επειδή δεν είναι παραγωγίσιμη συνάρτηση δεν χρησιμοποιείται αρκετά συχνά.



Γραμμική συνάρτηση

• $y = ax + b$



Σιγμοειδής συνάρτηση

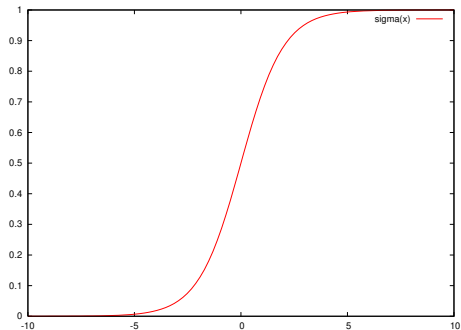
Είναι η συνάρτηση που χρησιμοποιείται πιο συχνά

- 1 Η συνάρτηση αυτή δίνεται από την εξίσωση

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

- 2 Οι τιμές της συνάρτησης είναι στο διάστημα $[0, 1]$
- 3 Παραγωγίζεται εύκολα.

Σιγμοειδής συνάρτηση - γράφημα



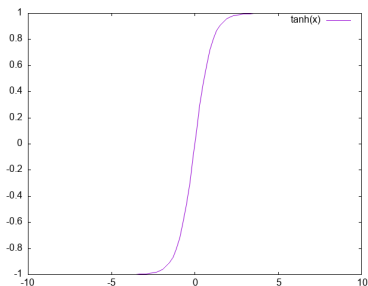
Σιγμοειδής συνάρτηση - Παράγωγοι

- Πρώτη παράγωγος: $\sigma'(x) = \sigma(x) \times (1 - \sigma(x))$
- Δεύτερη παράγωγος: $\sigma''(x) = \sigma(x) \times (1 - \sigma(x)) \times (1 - 2\sigma(x))$
- Οι υπόλοιπες παράγωγοι βγαίνουν πάντα σε σχέση με την $\sigma(x)$.
- Μπορούμε να χρησιμοποιήσουμε μεθόδους βελτιστοποίησης μεγάλης ακρίβειας μιας και παραγωγίζεται εύκολα.

Συνάρτηση υπερβολικής εφαπτομένης

- Μια ακόμα συνάρτηση που χρησιμοποιείται συχνά είναι η υπερβολική εφαπτομένη που ορίζεται ως

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

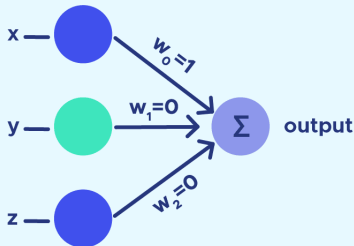


- Γράφημα:

Εισαγωγή

- Είναι η απλούστερη μορφή ΤΝΔ στα οποία υπάρχει μόνον ένας νευρώνας επεξεργασίας και έχει εφαρμογή κυρίως σε προβλήματα δύο κατηγοριών.
- Αναπτύχθηκε το 1957 στο Αεροναυτικό Εργαστήριο του Κορνέλλ (Cornell Aeronautical Laboratory) από τον Φρανκ Ρόζενμπλαττ (Frank Rosenblatt).
- Έμοιαζε πολλά υποσχόμενο δίκτυο στην αρχή.
- Δουλεύει μόνο σε δεδομένα τα οποία είναι γραμμικώς διαχωρίσιμα.
- Με τον καιρό χάθηκε το ενδιαφέρον για αυτά.
- Το ενδιαφέρον αυξήθηκε για αυτά πάλι το 1998 με την παρουσίαση ενός αλγορίθμου οποίος χρησιμοποιεί μία μέθοδο ψηφοφορίας .
- Αποτέλεσαν την βάση για τα δίκτυα πολλών επιπέδων.

Σχήμα δικτύου



- Σε αυτή την περίπτωση γίνεται ο υπολογισμός

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

όπου x είναι το διάνυσμα εισόδου, w είναι τα βάρη του νευρώνα και b η πόλωση.

- Ο πίνακας w ονομάζεται πίνακας παραμέτρων ή πίνακας βαρών.
- Σε αρκετές περιπτώσεις η πόλωση (bias) μπορεί να είναι 0
- Αν $u(x) > 0$ τότε το πρότυπο πάει στην πρώτη κατηγορία και αν $u(x) < 0$ στην δεύτερη.

Εκπαίδευση δικτύου

Τα βάρη w αναπροσαρμόζονται σύμφωνα με την εξίσωση:

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

όπου

- 1 $y(x)$ είναι η πραγματική έξοδος για το πρότυπο x (+1 και -1)
- 2 $o(x)$ είναι +1 αν $u(x) > 0$ και -1 αν $u(x) < 0$
- 3 n είναι ο ρυθμός μάθησης με $n \in (0, 1)$

Αλγόριθμος εκπαίδευσης

- ❶ Αρχικοποίηση των βαρών $w_i^{(0)}$, $i = 1..d + 1$ (+1 για την πόλωση).
- ❷ Η αρχικοποίηση μπορεί να σημαίνει είτε μηδενισμός των βαρών είτε αρχικοποίηση τους σε πολύ μικρές τυχαίες τιμές.
- ❸ Καθορισμός n
- ❹ Καθορισμός EMAX, μέγιστου αριθμού επαναλήψεων.
- ❺ $E = 0$ (αριθμός επαναλήψεων)
- ❻ Για κάθε πρότυπο $x = [1, x_1, x_2, \dots, x_d]$ στο σύνολο εκπαίδευσης (το 1 είναι για την πόλωση)
 - ❶ Υπολογισμός της ποσότητας $u(x) = w^T x$
 - ❷ Αν $u(x) < 0$, $o(x) = 1$ αλλιώς $o(x) = -1$
 - ❸ Ενημέρωση βαρών:

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

- ❹ $k = k + 1$
- ❼ $E = E + 1$
- ❶ Αν δεν υπάρχει μεταβολή βαρών από γενιά σε γενιά η $E > \text{EMAX}$ τότε ο αλγόριθμος τερματίζει.

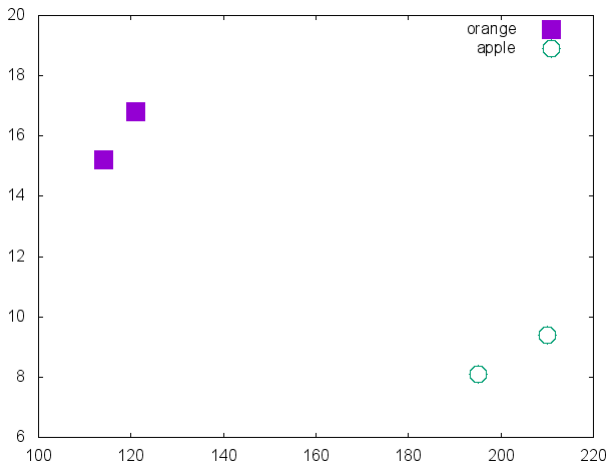
Παράδειγμα Μήλα και Πορτοκάλια για το δίκτυο Περσεπτρον

- Για την περιγραφή της διάκρισης μήλα και πορτοκάλια γίνεται χρήση των χαρακτηριστικών Βάρος (gr) και μέγεθος (cm).
- Πορτοκάλια +1 κατηγορία
- Μήλα -1 κατηγορία

Πίνακας φρούτων

Βάρος	Μέγεθος	Φρούτο	Κατηγορία
121	16.8	Πορτοκάλι	+1
114	15.2	Πορτοκάλι	+1
210	9.4	Μήλο	-1
195	8.1	Μήλο	-1

Σχηματική κατανομή φρούτων



- Τα δεδομένα φαίνεται να είναι γραμμικός διαχωρίσιμα.

Για το παραπάνω παράδειγμα χρειάζονται τρεις τιμές βαρών (2 για τα χαρακτηριστικά και 1 για την πόλωση). Ας υποθέσουμε πως ο πίνακας βαρών αρχικοποιείται αρχικά ως

$$w = [1, 1, 1]$$

Ο ρυθμός μάθησης n αρχικοποιείται σε $n = 0.1$

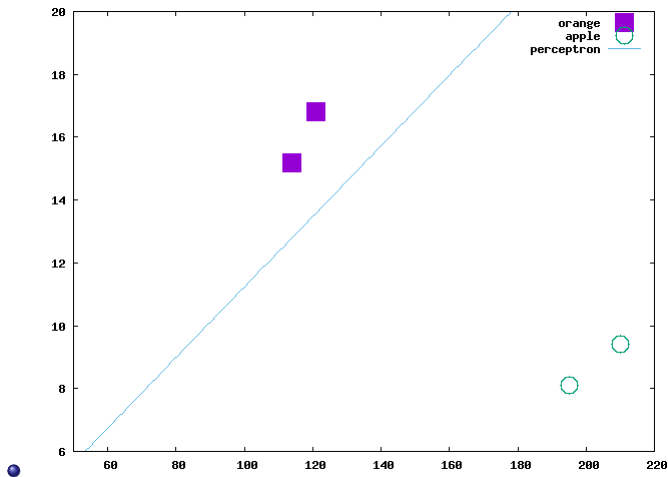
Πέρασμα τιμών από το δίκτυο

- 1 Πρότυπο $x = [121, 16.8, 1]$, $w^T x = 138.8$, κατηγορία είναι η +1. Επειδή η κατηγορία που παράγεται καθώς και η επιθυμητή είναι η ίδια δεν υπάρχει μεταβολή στα βάρη και επομένως $w = [1, 1, 1]$.
- 2 Πρότυπο $x = [114, 15.2, 1]$, $w^T x = 140, 2$, κατηγορία είναι η +1. Επειδή η κατηγορία που παράγεται καθώς και η επιθυμητή είναι η ίδια δεν υπάρχει μεταβολή στα βάρη και επομένως $w = [1, 1, 1]$.
- 3 Πρότυπο $x = [210, 9.4, 1]$, $w^T x = 220.4$, κατηγορία είναι η +1. Αυτό όμως έρχεται σε αντίθεση με την πραγματική κατηγορία που είναι η -1. Επομένως με διόρθωση βαρών $w = [-41, -0.88, 0.8]$
- 4 Πρότυπο $x = [195, 8.1, 1]$, $w^T x = -80001.328$, κατηγορία -1. Επειδή η κατηγορία που παράγεται καθώς και η επιθυμητή είναι η ίδια δεν υπάρχει μεταβολή στα βάρη.

- 1 Μετά από αρκετά περάσματα παράγεται ο ακόλουθος πίνακας βαρών $w = [-1.8, 16.04, 1.6]$
- 2 Για να βρούμε την ευθεία διαχωρισμού $w_1x_1 + w_2x_2 + b = 0$ θα έχουμε:
 - 1 Για $x_1 = 100$ θα πάρουμε
$$x_2 = \frac{-w_1x_1 - b}{w_2} = \frac{1.8*100 - 1.6}{16.04} = 11.12$$
 - 2 Για $x_1 = 200$ θα πάρουμε
$$x_2 = \frac{-w_1x_1 - b}{w_2} = \frac{1.8*200 - 1.6}{16.04} = 22.34$$
- 3 Η ευθεία διαχωρισμού είναι η

$$y(x) = (22.34 - 11.12)/100 * (x - 100) + 11.12$$

Ευθεία διαχωρισμού (σχήμα)



Ορισμοί

Μοιάζουν με τα νευρωνικά Perceptron αλλά μπορούν να χρησιμοποιηθούν και για την μάθηση συναρτήσεων.

- 1 Διαθέτουν έναν μόνον νευρώνα
- 2 Έχουν και πάλι ένα συντελεστή μάθησης
- 3 Η έξοδος δεν είναι απαραίτητα $+1, -1$
- 4 Τα πρότυπα παρουσιάζονται επαναληπτικά στο δίκτυο (Εποχή)
- 5 Το δίκτυο πάει να εκπαιδεύεται όταν το μέσο τετραγωνικό σφάλμα πέσει κάτω από ένα όριο ή φτάσουμε σε έναν μέγιστο αριθμό εποχών και ορίζεται ως

$$E = \frac{1}{M} \sum_{i=1}^M (w^T x_i - y_i)^2$$

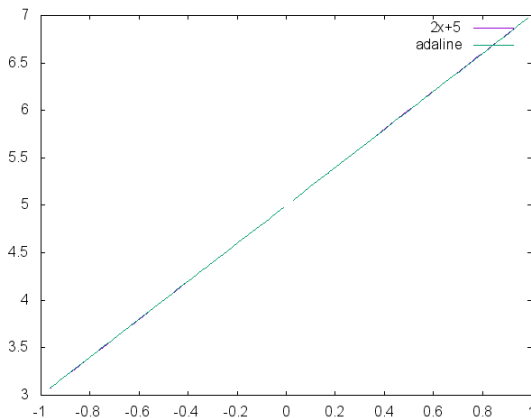
όπου x_i είναι το i πρότυπο εκπαίδευσής και y_i η επιθυμητή έξοδος.

Αλγόριθμος εκπαίδευσης

- ❶ Είσοδος τα πρότυπα εισόδου $x_1, x_2, \dots, x_M$
- ❷ Είσοδος οι επιθυμητές έξοδοι $y_1, y_2, \dots, y_M$
- ❸ Ορισμός e το όριο για το σφάλμα εκπαίδευσης
- ❹ Ορισμός b για την παράμετρο εκπαίδευσης
- ❺ Ορισμός μέγιστο αριθμού εποχών MAXEPOCHS
- ❻ $k = 1$
- ❼ Για κάθε πρότυπο x_i
 - ❶ $u_i = w^T x_i$
 - ❷ $w = w + n * (y_i - u_i) x_i$
- ❽ Υπολογισμός μέσου τετραγωνικό σφάλματος E
- ❾ Αν $E < e$ τερματισμός
- ❿ $k = k + 1$
- ⓫ Αν $k > \text{MAXEPOCHS}$ τερματισμός

Μάθηση γραμμικής συνάρτησης

- Σε μια πρώτη προσπάθεια έγινε προσέγγιση της συναρτήσεως $2x + 5$ στο διάστημα $[-1, 1]$



Μάθησης τετραγωνικής συνάρτησης

- Προσπάθεια για προσέγγιση της συνάρτησεως $x^2 + 1$ στο διάστημα $[-1, 1]$

