

# Υπολογιστική Νοημοσύνη

Ιωάννης Γ. Τσούλος

Τμήμα Πληροφορικής και τηλεπικοινωνιών  
Πανεπιστήμιο Ιωαννίνων

2021

# Περίληψη

- 1 Εισαγωγή
  - Βασικά στοιχεία
  - Συναρτήσεις ενεργοποίησης
- 2 Δίκτυα Perceptron
  - Βασικά στοιχεία
  - Εκτεταμένο παράδειγμα μάθησης
- 3 Τα 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 από 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 είσοδο που δέχονται από νευρώνες εισόδου ή από άλλους νευρώνες επεξεργασίας με μια τιμή συσχετισμένη με αυτούς που ονομάζεται βάρος. Το συνολικό αποτέλεσμα εισάγεται σε μια συνάρτηση που θα ονομάζουμε συνάρτηση ενεργοποίησης και είτε παρουσιάζεται στην έξοδο είτε δίδεται σε κάποιον άλλο νευρώνα επεξεργασίας.

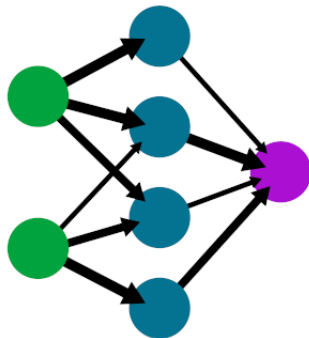
# Σχηματικό Παράδειγμα

A simple neural network

input  
layer

hidden  
layer

output  
layer



## Παράδειγμα εξίσωσης.

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

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

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

## 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.

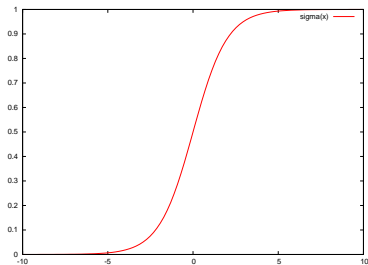
Είναι η συνάρτηση που χρησιμοποιείται πιο συχνά

- 1 Η συνάρτηση αυτή δίνεται από την εξίσωση

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

- 2 Η συνάρτηση είναι στο διάστημα  $[0,1]$
- 3 Παραγωγίζεται εύκολα.

# Σιγμοειδής συνάρτηση - Γράφημα



# Συνάρτηση Υπερβολικής Εφαπτομένης.

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

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

- 1 Ιατρική πληροφορική όπως και διάγνωση παθήσεων
- 2 Οικονομικά θέματα όπως χρηματιστηριακές προβλέψεις, εξακρίβωση πιστοληπτικής ικανότητας κτλ.
- 3 Ανάπτυξη νέων φαρμάκων για ασθένειες.
- 4 Ανίχνευση βλαβών σε μηχανήματα

## Ορισμοί

- 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)[1].
- 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 [2].
- Αποτέλεσαν την βάση για τα δίκτυα πολλών επιπέδων.

- 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.

- Αν  $u(x) > 0$  τότε το πρότυπο πάει στην πρώτη κατηγορία και αν  $u(x) < 0$  στην δεύτερη.

## Αναπροσαρμογή βαρών

Τα βάρη  $w$  αναπροσαρμόζονται σύμφωνα με την εξίσωση:

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

όπου

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

# Βασικός αλγόριθμος

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

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

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



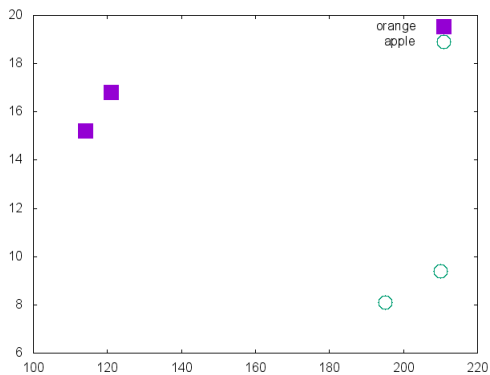
## Παράδειγμα Μήλα και Πορτοκάλια για το δίκτυο Perceptron

- 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



## Αρχικοποίηση βαρών

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  $\eta$  αρχικοποιείται σε  $\eta=0.1$

# Πέρασμα προτύπων

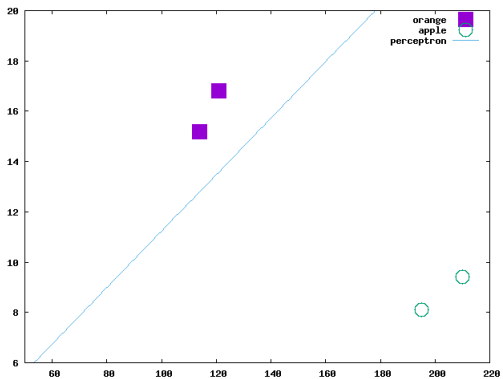
- 1 Πρότυπο  $x=[121,16.8,1]$  ,  $w^T x = 138.8$ , κατηγορία είναι η +1. Επειδή η κατηγορία που παράγεται καθώς και η επιθυμητή είναι η ίδια δεν υπάρχει μεταβολή στα βάρη και επομένως  $w = [1, 1, 1]$ .
- 2 Πρότυπο  $x=[114,15.2,1]$ ,  $w^T x = 140,2$  , κατηγορία είναι η +1. Επειδή η κατηγορία που 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 τον κανόνα δέλτα πρέπει να γίνει διόρθωση βαρών από τον τύπο  $w = w + n(y(x) - o(x))x$ , συνεπώς  $w=[-41,-0.88,0.8]$
- 4 Πρότυπο  $x=[195,8.1,1]$ ,  $w^T x = -8001.328$ , κατηγορία -1. Επειδή η κατηγορία που παράγεται καθώς και η επιθυμητή είναι η ίδια δεν υπάρχει μεταβολή στα βάρη.

## Ευθεία διαχωρισμού

- 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  $x_2 = \frac{-w_1x_1 - b}{w_2} = \frac{1.8 \cdot 100 - 1.6}{16.04} = 11.12$
  - 2 Για  $x_1 = 200$  θα πάρουμε  $x_2 = \frac{-w_1x_1 - b}{w_2} = \frac{1.8 \cdot 200 - 1.6}{16.04} = 22.34$
- 3 Η ευθεία διαχωρισμού είναι η

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

# Ευθεία διαχωρισμού (σχήμα)



Μοιάζουν αρκετά με τα νευρωνικά Perceptron αλλά μπορούν να χρησιμοποιηθούν και για την μάθηση συναρτήσεων όπως και για την μάθηση (υπό προϋποθέσεις) κατηγοριών που δεν είναι μόνον δύο.

- 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 εκπαιδευτεί όταν το μέσο τετραγωνικό σφάλμα πέσει κάτω από ένα όριο ή φτάσουμε σε έναν μέγιστο αριθμό εποχών. Το μέσο τετραγωνικό σφάλμα ορίζεται ως

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

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

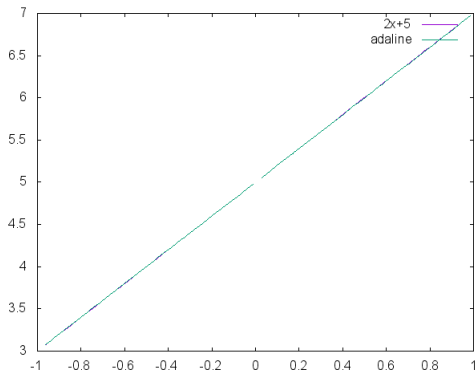


## Ο αλγόριθμος

- 1 Είσοδος τα πρότυπα εισόδου  $x_1, x_2, \dots, x_M$
- 2 Είσοδος οι επιθυμητές έξοδοι  $y_1, y_2, \dots, y_M$
- 3 Ορισμός  $e$  το όριο για το σφάλμα εκπαίδευσης
- 4 Ορισμός  $b$  για την παράμετρο εκπαίδευσης
- 5 Ορισμός μέγιστο αριθμού εποχών MAXEPOCHS
- 6  $k=1$
- 7 Για κάθε πρότυπο  $x_i$ 
  - 1  $u_i = w^T x_i$
  - 2  $w = w + n * (y_i - u_i) x_i$
- 8 Υπολογισμός μέσου τετραγωνικό σφάλματος  $E$
- 9 Αν  $E < e$  τερματισμός
- 10  $k=k+1$
- 11 Αν  $k > \text{MAXEPOCHS}$  τερματισμός

# Παράδειγμα μάθησης συναρτήσης

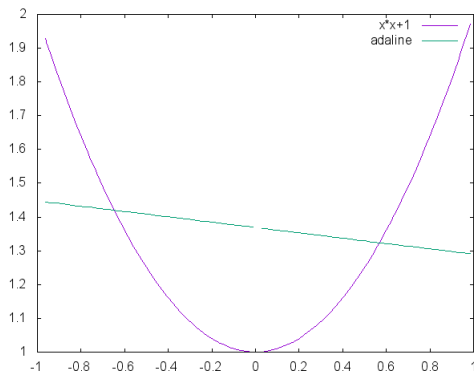
- Σε μια πρώτη προσπάθεια έγινε προσέγγιση της συναρτήσεως  $2x+5$  στο διάστημα  $[-1,1]$



- Είναι προφανές πως το δίκτυο κατάφερε να μάθει με μεγάλη επιτυχία την συνάρτηση.

## Παράδειγμα μάθησης συναρτήσεως(2)

- Προσπάθει για προσέγγιση της συναρτήσεως  $x^2 + 1$  στο διάστημα  $[-1,1]$





- όπου φαίνεται και η αδυναμία του απλού δικτύου Adaline να προσεγγίσει συναρτήσεις πιο σύνθετες από αυτές της  $x^2 + 1$

# Σύνοψη

- Παρουσίαση των δικτύων Perceptron
- Παρουσίαση των δικτύων Adaline

# Βιβλιογραφία I

-  Rosenblatt, Frank (1958), The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6, pp. 386–408.
-  Freund, Y. and Schapire, R. E. 1998. Large margin classification using the perceptron algorithm. In Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT' 98). ACM Press.