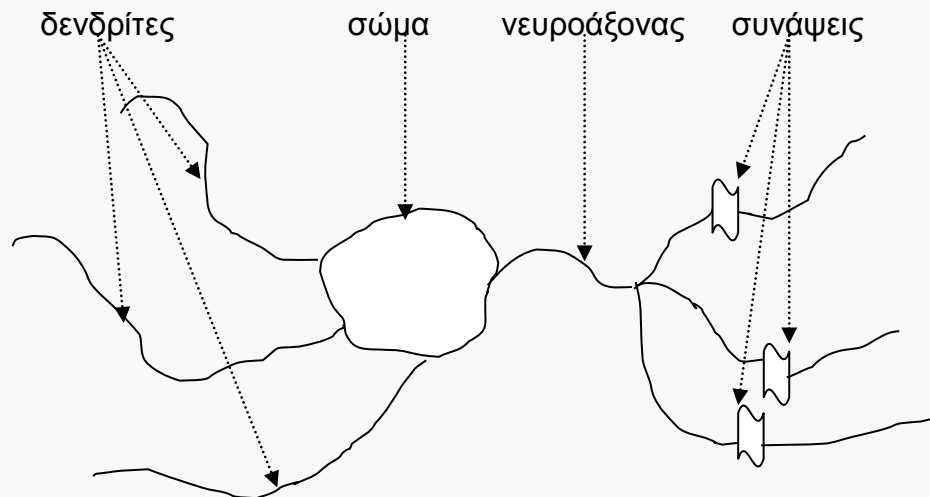


Βασικές αρχές εκπαίδευσης ΤΝΔ: το perceptron

Βιολογικός Νευρώνας

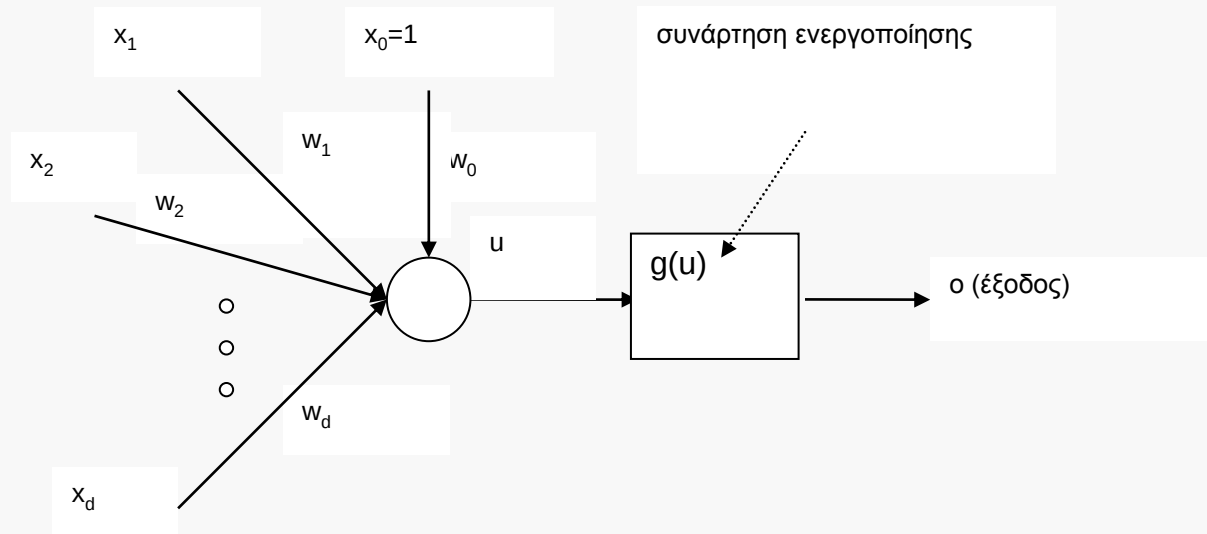
- **Δενδρίτες**, που αποτελούν τις γραμμές εισόδου των ερεθισμάτων (βιολογικών σημάτων)
- **Σώμα**, στο οποίο γίνεται η συσσώρευση των ερεθισμάτων και ο καθορισμός της διέγερσης του νευρώνα.
- **Νευροάξονας**, που αποτελεί τη γραμμή εξόδου του νευρώνα.
- **Σύναψη**, που είναι το σημείο διασύνδεσης μεταξύ δύο νευρώνων.
- Έχει παρατηρηθεί ότι το σήμα που εξέρχεται από το νευροάξονα ενός νευρώνα και εισέρχεται στο δενδρίτη του άλλου νευρώνα **διαμορφώνεται** κατά ένα ποσοστό που σχετίζεται με την **ισχύ της σύναψης** που ονομάζεται **συναπτικό δυναμικό**.



Συναπτικό Δυναμικό

- Το συναπτικό δυναμικό μπορεί να ενισχύει (θετικό) ή να καταστέλλει (αρνητικό) το σήμα εξόδου.
- Η γνώση μας είναι 'αποθηκευμένη' στις τιμές των συναπτικών δυναμικών.
- **Μάθηση στα βιολογικά συστήματα είναι η μεταβολή των συναπτικών δυναμικών.**
- Όσο περισσότερο χρησιμοποιείται μια σύναψη τόσο ενισχύεται το δυναμικό της.

Τεχνητός Νευρώνας (neuron)



- d είσοδοι,
- σήμα εισόδου x_i ($i=1, \dots, d$)
- βάρη εισόδων w_i , ($i=1, \dots, d$)
- πόλωση $w_0 \rightarrow$ τιμή βάρους μιας σύνδεσης που η είσοδός της είναι μόνιμα στην τιμή 1

Τεχνητός Νευρώνας (neuron)

- Ο υπολογισμός σε **δύο στάδια**:
 - υπολογισμός της **συνολικής εισόδου (ενεργοποίηση)**:

$$u(x) = \sum_{i=1}^d w_i x_i + w_0$$

- υπολογισμός της εξόδου $o(x)$ του νευρώνα περνώντας την συνολική είσοδο $u(x)$ από μια **συνάρτηση ενεργοποίησης (activation function)**

$$o(x) = g(u)$$

- **νευρώνας εσωτερικού γινομένου**

$$u(x) = w^T x + w_0$$

Τεχνητός νευρώνας

- Εναλλακτική διατύπωση:
 - Διάνυσμα βαρών: $w=(w_1, w_2, \dots, w_d)^T$
 - Εκτεταμένο (extended) διάνυσμα βαρών:
 - $w_e=(w_0, w_1, w_2, \dots, w_d)^T$
 - Εκτεταμένο (extended) διάνυσμα εισόδου:
 - $x_e=(1, x_1, x_2, \dots, x_d)^T$
- $u(x)=w_e^T x_e \iff u(x)=w^T x + w_0$

Συναρτήσεις ενεργοποίησης

- **Βηματική Συνάρτηση (ή συνάρτηση κατωφλίου):**
 - Η συνάρτηση ενεργοποίησης στο βιολογικό νευρώνα
 - Χαρακτηρίζεται από δύο τιμές a και b .
 - Αν $x < 0$ τότε $g(x) = a$ και εάν $x > 0$ τότε $g(x) = b$. Συνήθως χρησιμοποιούνται οι τιμές $a = 0$ και $b = 1$ είτε $a = -1$ και $b = 1$.
 - Για $x = 0$ υπάρχει ασυνέχεια και μπορούμε κατά σύμβαση να αναθέσουμε είτε την τιμή a είτε την τιμή b .
 - **Η βηματική συνάρτηση έχει το μειονέκτημα ότι η παράγωγός της είναι μηδέν.**
 - Δεδομένου ότι μάθηση στα ΤΝΔ είναι η μεταβολή των τιμών των βαρών και μεταβολή σχετίζεται με την παράγωγο, η βηματική συνάρτηση δεν θεωρείται βολική ως συνάρτηση ενεργοποίησης των νευρώνων στα ΤΝΔ

Συναρτήσεις ενεργοποίησης

- Σιγμοειδείς συναρτήσεις
 - Έχουν μορφή τελικού σίγμα
 - Αποτελούν συνεχείς και παραγωγίσιμες προσεγγίσεις της βηματικής.
 - Στο όριο που η κλίση γίνεται πολύ μεγάλη, η σιγμοειδής γίνεται βηματική.
 - Δύο βασικοί τύποι:

1) Λογιστική:

$$\sigma(x) = 1 / (1 + \exp(-ax))$$

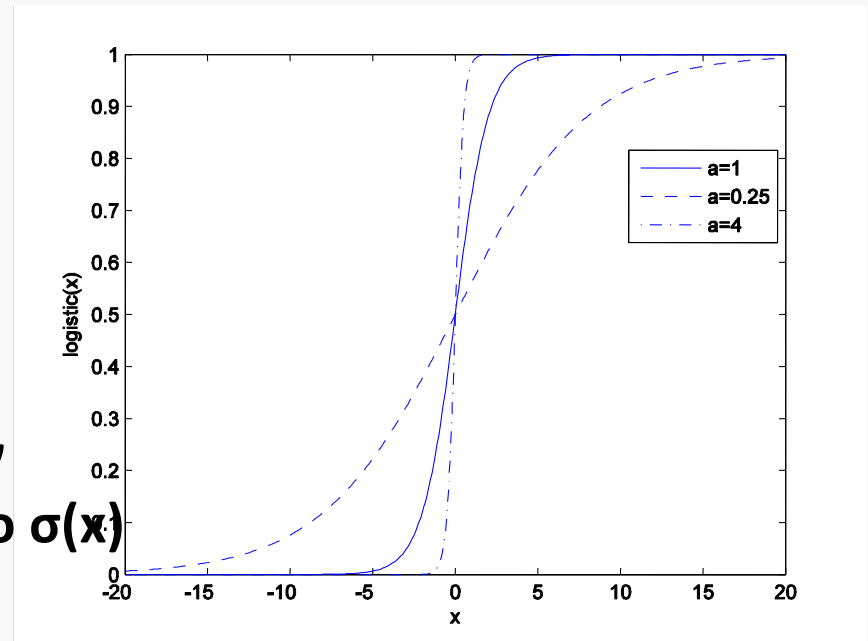
(a: κλίση, συνήθως $a=1$)

δίνει τιμές στο $(0,1)$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

$$\sigma''(x) = \sigma(x)(1 - \sigma(x))(1 - 2\sigma(x))$$

Μπορούμε να υπολογίσουμε την παράγωγο $\sigma'(x)$ ξέροντας μόνο το $\sigma(x)$ χωρίς να χρειάζεται η τιμή του x .



Συναρτήσεις ενεργοποίησης

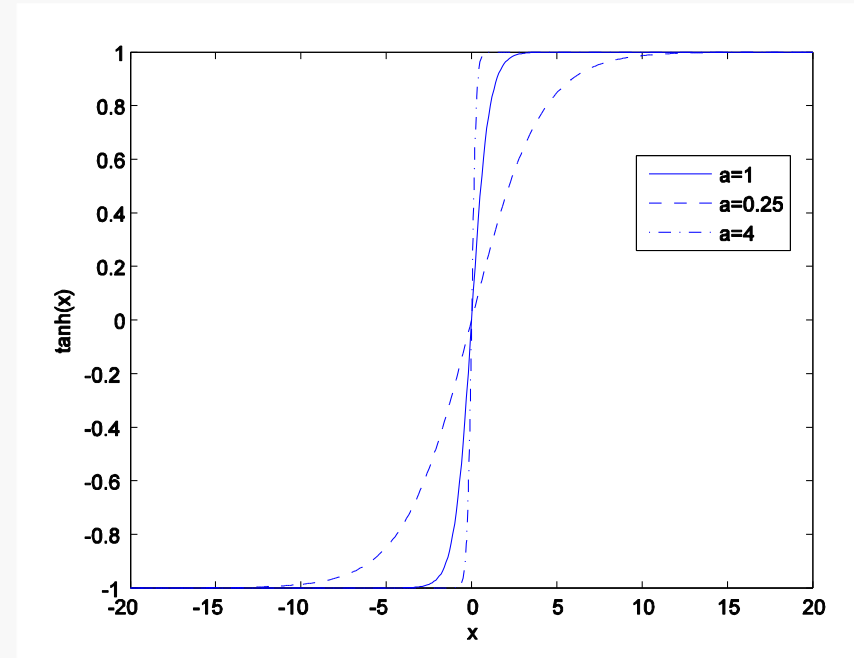
2) Υπερβολική εφαπτομένη:

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

(a: κλίση, συνήθως $a=1$)

δίνει τιμές στο $(-1,1)$

$$\tanh'(x) = 1 - \tanh^2(x)$$



- Γραμμική συνάρτηση

$$g(x)=x, g'(x)=1$$

δίνει τιμές στο $\mathbb{R}$

Ερώτηση

Εάν σε κάποιο σημείο η λογιστική συνάρτηση έχει τιμή 0.6, ποια η τιμή της πρώτης παραγώγου της σε αυτό το σημείο; Ποια η τιμή της δεύτερης παραγώγου στο ίδιο σημείο;

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) = 0.24$$

$$\sigma''(x) = \sigma(x)(1 - \sigma(x))(1 - 2\sigma(x)) = -0.048$$

Ερώτηση

Με βάση τη λογιστική συνάρτηση $\sigma(x)$, να ορίσετε μια σιγμοειδή συνάρτηση ενεργοποίησης η οποία να δίνει τιμές στο διάστημα (a, b) , όπου a, b πραγματικοί αριθμοί.

$$g_1(x) = (b - a) * \sigma(x) \text{ (τιμές στο } (0, b - a))$$

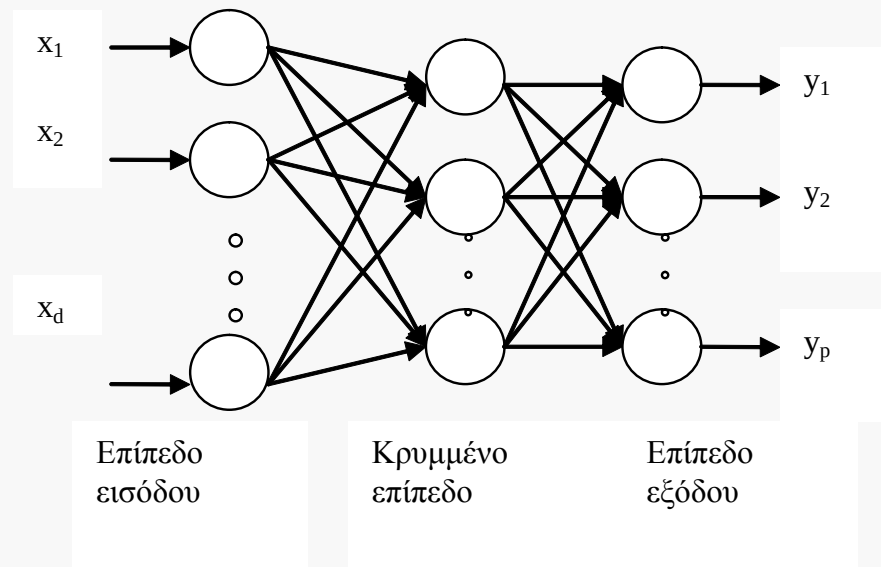
$$g_2(x) = a + (b - a) * \sigma(x) \text{ (τιμές στο } (a, b))$$

Από τον τεχνητό νευρώνα στο ΤΝΔ

- Διασυνδέοντας τεχνητούς νευρώνες ορίζουμε τεχνητά νευρωνικά δίκτυα.
- Η αρχιτεκτονική (τρόπος διασύνδεσης) και ο τύπος των νευρώνων (π.χ. το είδος της συνάρτησης ενεργοποίησης που χρησιμοποιούν) καθορίζονται κάθε φορά από τον τύπο του προβλήματος μάθησης που θέλουμε να αντιμετωπίσουμε.

ΤΝΔ Πρόσθιας Τροφοδότησης (feedforward)

- Οι υπολογισμοί σε μία κατεύθυνση: από την είσοδο προς την έξοδο
- Υλοποιούν στατικές απεικονίσεις εισόδου-εξόδου



Ασκηση:

Να ορίσετε την αρχιτεκτονική ενός ΤΝΔ με δύο εισόδους x_1 , x_2 και μία έξοδο ο το οποίο να υλοποιεί τη συνάρτηση:

$$o(x_1, x_2) = 3\sigma(3x_1 - 2x_2 + 7) - \sigma(-5x_1 + x_2 - 4) + 6$$

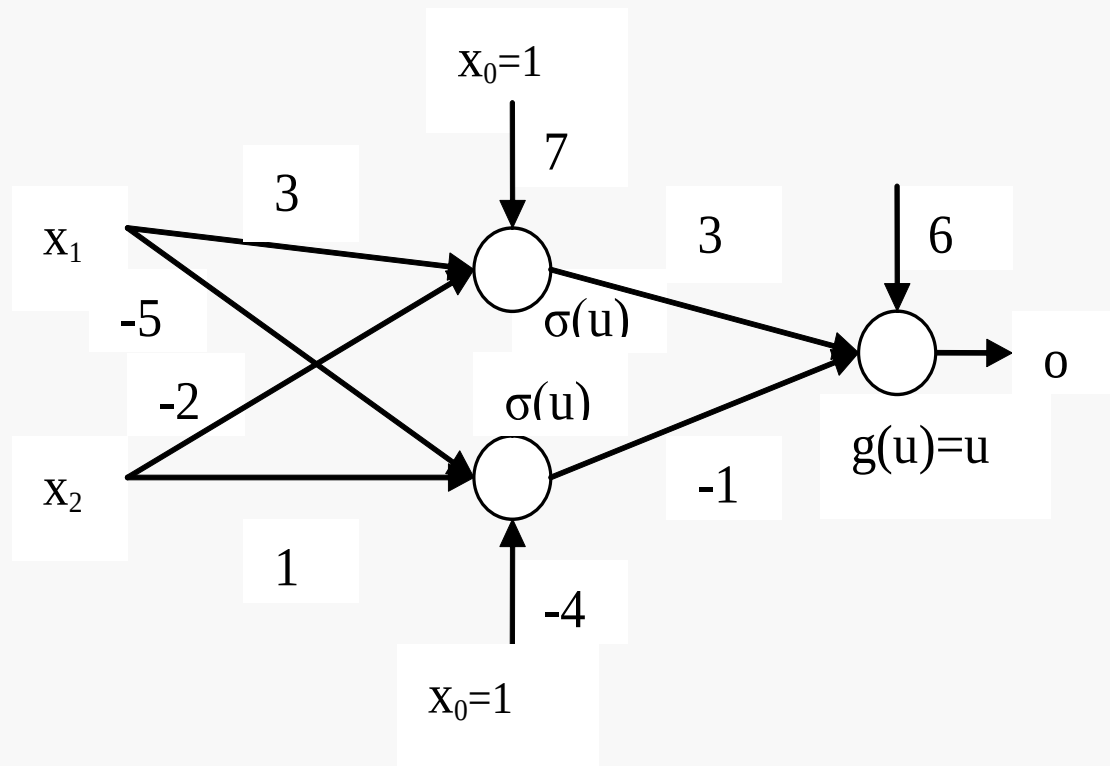
Σε κάθε λογιστική συνάρτηση $\sigma(u)$ αντιστοιχίζουμε ένα νευρώνα με λογιστική συνάρτηση ενεργοποίησης.

Το όρισμα κάθε λογιστικής είναι συνάρτηση των εισόδων του νευρώνα και μας παρέχει τις τιμές των βαρών εισόδου στο νευρώνα αυτόν.

Οι έξοδοι των δύο σιγμοειδών νευρώνων πολλαπλασιάζονται με συντελεστές βάρους και αθροίζονται για να δώσουν την έξοδο του ΤΝΔ. Άρα στην έξοδο έχουμε ένα νευρώνα με γραμμική συνάρτηση ενεργοποίησης.

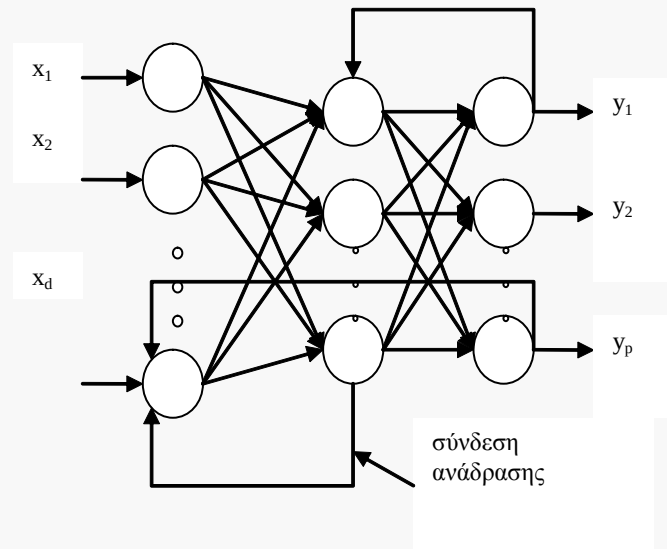
Για όλους τους νευρώνες οι σταθερές που προστίθενται αποτελούν τις πολώσεις των νευρώνων

$$o(x_1, x_2) = 3\sigma(3x_1 - 2x_2 + 7) - \sigma(-5x_1 + x_2 - 4) + 6$$



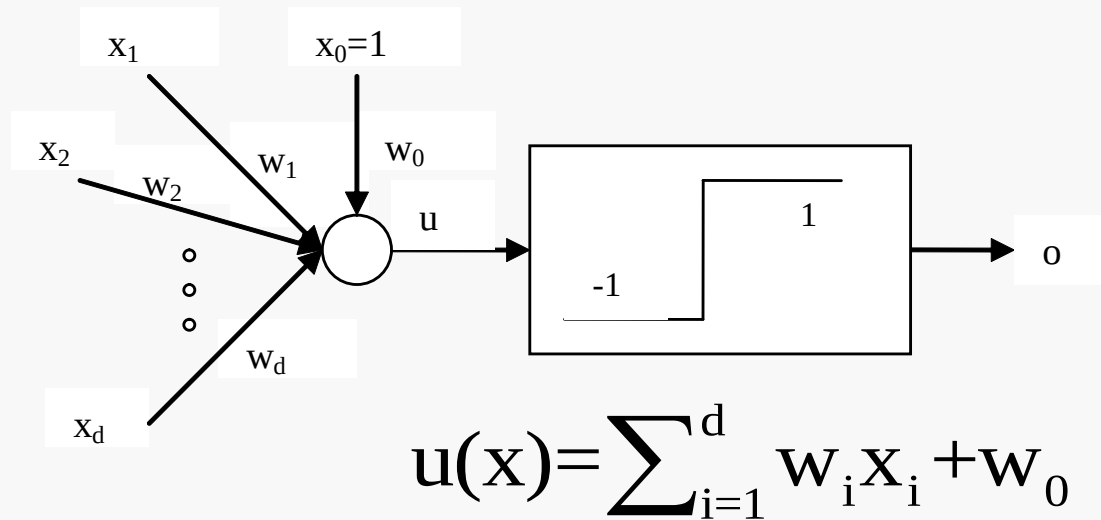
Επαναληπτικά ΤΝΔ (recurrent)

- Συνδέσεις ανάδρασης από την έξοδο προς την είσοδο $\rightarrow$ δυναμικά συστήματα (εξελισσονται στο χρόνο)



Το perceptron

- Το αρχαιότερο και απλούστερο νευρωνικό δίκτυο
- Ένας μόνο νευρώνας εσωτερικού γινομένου με **βηματική** συνάρτηση ενεργοποίησης



$$o(x) = 1 \text{ εάν } u(x) > 0 \text{ και } o(x) = -1 \text{ εάν } u(x) < 0.$$

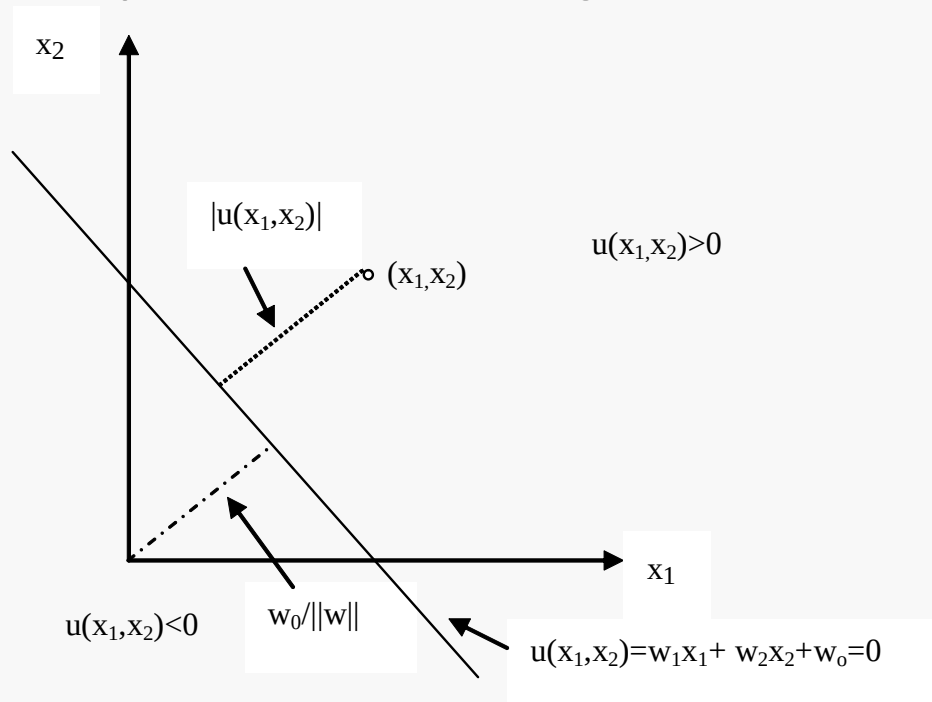
Τι δυνατότητες έχει το perceptron;

Εξίσωση υπερεπιπέδου

- Τα σημεία $x=(x_1, \dots, x_d)^T$ που ικανοποιούν την εξίσωση:

$$\sum_{i=1}^d w_i x_i + w_0 = 0$$

ορίζουν ένα **υπερεπίπεδο** $H(w)$ στον R^d



$$u(x) = \sum_{i=1}^d w_i x_i + w_0$$

Υπερεπίπεδο στον R^2
(ευθεία γραμμή).

Εξίσωση υπερεπιπέδου

- Για $x=(x_1, \dots, x_d)^T$ η απόλυτη τιμή $|u(x)|$ δηλώνει την απόσταση του σημείου x από το υπερεπίπεδο $H(w)$.
- Το υπερεπίπεδο χωρίζει το χώρο R^d σε δύο ημιχώρους.
- Για όλα τα σημεία x που ανήκουν στον ένα ημιχώρο ισχύει $u(x) > 0$ (**θετικός ημιχώρος**)
- Για όλα τα σημεία x που ανήκουν στον άλλο ημιχώρο ισχύει $u(x) < 0$ (**αρνητικός ημιχώρος**).

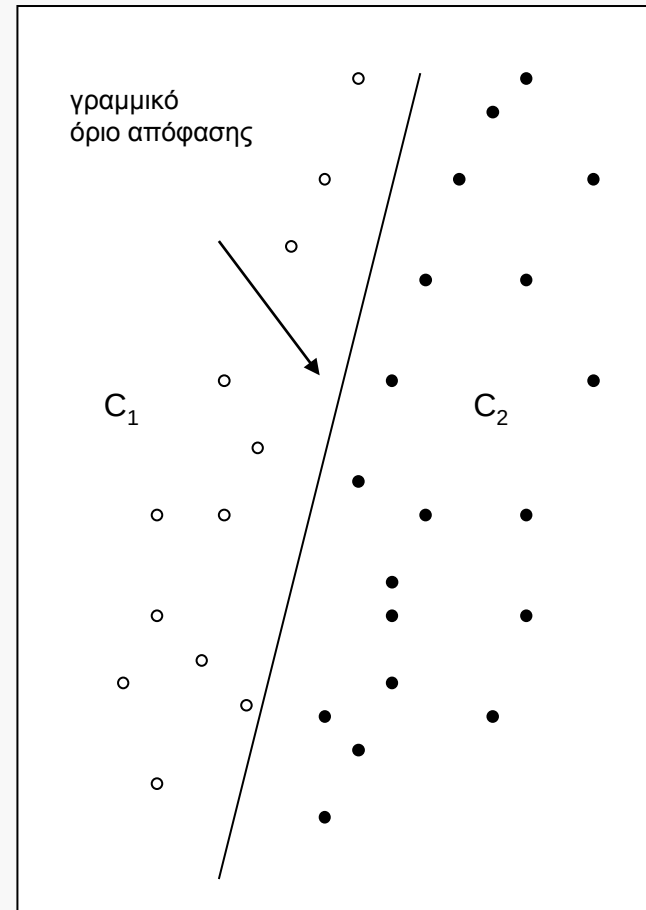
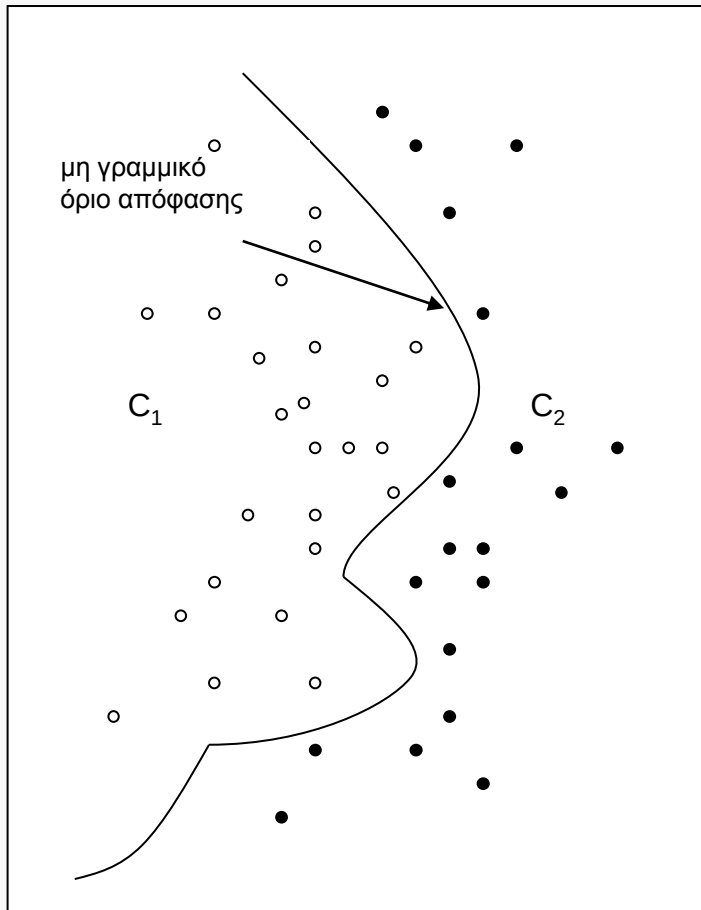
Γεωμετρική Ερμηνεία για Perceptron

- **Ενα perceptron με διάνυσμα βαρών w ορίζει ένα υπερεπίπεδο $H(w)$**
- Δοθέντος ενός διανύσματος εισόδου x υπολογίζει το $u(x)$ και παρέχει έξοδο
 - $o(x)=1$ εάν το x ανήκει στο θετικό ημιχώρο ($u(x)>0$)
 - $o(x)=-1$ εάν το x ανήκει στον αρνητικό ημιχώρο ($u(x)<0$).
- Η έξοδος $o(x)$ δηλώνει τον ημιχώρο στον οποίο βρίσκεται το x σε σχέση με το υπερεπίπεδο $H(w)$.
- Δεδομένου ότι εκπαίδευση στα ΤΝΔ είναι η ρύθμιση των βαρών, **η εκπαίδευση του perceptron ανάγεται στον καθορισμό του υπερεπιπέδου ώστε να λύνεται κάποιο πρόβλημα.**

Γραμμική Διαχωρισιμότητα

- Ένα σύνολο δεδομένων δύο κατηγοριών ονομάζεται **γραμμικά διαχωρίσιμο**, εάν υπάρχει υπερεπίπεδο που διαχωρίζει τις κατηγορίες. Δηλαδή αφήνει όλα τα πρότυπα της μιας κατηγορίας στο θετικό ημιχώρο και όλα τα πρότυπα της άλλης κατηγορίας στον αρνητικό ημιχώρο.
- Αν δεν ικανοποιείται η παραπάνω ιδιότητα το σύνολο δεδομένων ονομάζεται **μη γραμμικά διαχωρίσιμο**.
- Το διαχωριστικό υπερεπίπεδο εάν υπάρχει **δεν** είναι μοναδικό.

Γραμμική Διαχωρισιμότητα



Γραμμική Διαχωρισιμότητα

- Ένα σύνολο δεδομένων δύο κατηγοριών ονομάζεται διαχωρίζεται από την ευθεία $y=5x+4$. Ποιο είναι το perceptron που λύνει αυτό το πρόβλημα;
- Τα δεδομένα είναι στον R^2 (δύο είσοδοι: x_1 , x_2)
- $x_2 = 5 x_1 + 4 \rightarrow 5 x_1 - x_2 + 4 = 0$
- $w_1 = 5, w_2 = -1, w_0 = 4$

Εκπαίδευση Perceptron

- Το perceptron μπορεί να χρησιμοποιηθεί για προβλήματα **ταξινόμησης δύο κατηγοριών** παρέχοντας έξοδο 1 για δεδομένα της μιας κατηγορίας και έξοδο -1 για δεδομένα της δεύτερης κατηγορίας.
- Η εκπαίδευση του perceptron συνίσταται στην εύρεση του διαχωριστικού υπερεπιπέδου $H(w)$.
- Μάθηση είναι η μεταβολή των βαρών $\rightarrow$ μεταβολή της θέσης του υπερεπιπέδου.

Κωδικοποίηση Κατηγοριών

- $X=\{(x^n, C(x^n)), n=1, \dots, N\}$ σύνολο παραδειγμάτων εκπαίδευσης (διάστασης d) δύο κατηγοριών
- $x^n=(x_{n1}, \dots, x_{nd})^T$
- $C(x^n)$ δηλώνει την κατηγορία (C_1 ή C_2) του x^n .
- Στα ΤΝΔ η έξοδος είναι αριθμητική
- Το $C(x^n)$ είναι συμβολική πληροφορία
- Χρειάζεται η συμβολική πληροφορία να μετατραπεί σε αριθμητική.
- **Κωδικοποίηση κατηγοριών: καθορισμός τιμών-στόχων t για τις κατηγορίες: $C_1 \rightarrow t=1, C_2 \rightarrow t=-1$**
- $\{(x^n, C(x^n))\} \rightarrow \{(x^n, t^n)\}$, όπου $t^n=1$ εάν $C(x^n)=C_1$ και όπου $t^n=-1$ εάν $C(x^n)=C_2$.

Εκπαίδευση perceptron με διόρθωση σφάλματος

- $X = \{(x^n, t^n), n=1, \dots, N\}$ σύνολο παραδειγμάτων εκπαίδευσης (διάστασης d) δύο κατηγοριών
- $x^n = (x_{n1}, \dots, x_{nd})^T$ $t^n = 1$ ή -1
- **Στόχος της εκπαίδευσης: εύρεση των βαρών** ώστε
$$o(x^n) = t^n, \text{ για κάθε } n=1, \dots, N$$
- Εκπαίδευση: μεταβολή των βαρών (δηλ. του διαχωριστικού υπερεπιπέδου) ώστε να διορθώνονται τα σφάλματα.
- Αρχικά επιλέγεται ένα τυχαίο διάνυσμα βαρών δηλαδή μια τυχαία θέση του διαχωριστικού υπερεπιπέδου.
- Με αυτή την τυχαία επιλογή, κάποια από τα δεδομένα x^n θα ταξινομούνται σωστά ($o(x^n) = t^n$) και κάποια λάθος ($o(x^n) \neq t^n$).
- Σε κάθε επανάληψη του αλγορίθμου εκπαίδευσης επιλέγεται ένα παράδειγμα εκπαίδευσης (x^n, t^n) , εφαρμόζεται ως είσοδος στο perceptron και υπολογίζεται η έξοδος o .

Εκπαίδευση perceptron με διόρθωση σφάλματος

- **Σωστή ταξινόμηση** του x^n : δεν υπάρχει κάποιος λόγος να αλλάξουμε τις τιμές των βαρών.
- **Λάθος ταξινόμηση**: το x^n βρίσκεται από τη λάθος πλευρά σε σχέση με το διαχωριστικό υπερεπίπεδο.
 - Για το λόγο αυτό μεταβάλλουμε τα βάρη, δηλ. μετακινούμε το διαχωριστικό υπερεπίπεδο, στην **κατεύθυνση διόρθωσης του σφάλματος**.
 - Ο στόχος είναι είτε το δεδομένο να βρεθεί από την άλλη πλευρά σε σχέση με το υπερεπίπεδο είτε να παραμείνει σε λάθος πλευρά αλλά να μειωθεί η απόστασή του από το υπερεπίπεδο (έτσι ώστε σε επόμενη επανάληψη να τοποθετηθεί στη σωστή πλευρά).
- Αποδεικνύεται ότι **εάν το σύνολο εκπαίδευσης είναι γραμμικά διαχωρίσιμο**, η επαναληπτική διαδικασία θα εντοπίσει ένα σωστό διαχωριστικό υπερεπίπεδο σε πεπερασμένο αριθμό επαναλήψεων και **θα τερματίσει**.

Αλγόριθμος Εκπαίδευσης perceptron με διόρθωση σφάλματος

- Έστω $w_e(k)$ το εκτεταμένο διάνυσμα βαρών του perceptron και $(x(k), t(k))$ το παράδειγμα εκπαίδευσης που χρησιμοποιείται για την εκπαίδευση του perceptron κατά την επανάληψη k .
- Αρχικοποίηση: Θέτουμε $k=0$, δίνουμε στα βάρη $w_i(0)$ κάποιες τυχαία επιλεγμένες αρχικές τιμές (συνήθως στο $(-1, 1)$) και καθορίζουμε την τιμή του **ρυθμού μάθησης η** ($\eta > 0$).
- Ο ρυθμός μάθησης καθορίζει το πόσο μεγάλες ή μικρές θα είναι οι μεταβολές στις τιμές των βαρών (τυπική τιμή $\eta=0.1$)

Αλγόριθμος Εκπαίδευσης perceptron με διόρθωση σφάλματος

- Σε κάθε επανάληψη k εκτελούμε τα ακόλουθα βήματα:
 - Επιλογή κάποιου παραδείγματος εκπαίδευσης $(x(k), t(k))$, εφαρμογή του $x_e(k)$ ως είσοδο στο perceptron και υπολογισμός της εξόδου $o(k) = g(w_e^T(k)x_e(k))$.
- 2. Ενημέρωση του διανύσματος βαρών του perceptron σύμφωνα με τον κανόνα διόρθωσης σφάλματος:
 - $w_e(k+1) = w_e(k) + n[t(k) - o(k)]x_e(k)$ (διανυσματική μορφή)
ή
 - $w_i(k+1) = w_i(k) + n[t(k) - o(k)]x_i(k)$ ($i=1, \dots, d$),
 - $w_0(k+1) = w_0(k) + n[t(k) - o(k)]$
- 3. $k := k+1$

Αλγόριθμος Εκπαίδευσης perceptron με διόρθωση σφάλματος

- Αποδεικνύεται ότι **εάν το σύνολο εκπαίδευσης είναι γραμμικά διαχωρίσιμο** και όλα τα παραδείγματα χρησιμοποιούνται για την εκπαίδευση, **ο αλγόριθμος σε πεπερασμένο αριθμό βημάτων θα καταλήξει σε τιμές βαρών που θα ταξινομούν σωστά όλα τα παραδείγματα εκπαίδευσης ανεξάρτητα από την τιμή του ρυθμού μάθησης (που πρέπει πάντα να είναι θετική).**

Γιατί διορθώνονται τα σφάλματα;

- Έστω $o(k)=-1$ και $t(k)=1 \rightarrow t(k) - o(k) = 2$

Ισχύει ότι $u(k)<0$ και επιθυμούμε $u(k+1)>u(k)$.

$$\begin{aligned} u(k+1) &= w_e^T(k+1) x_e(k) = [w_e^T(k) + 2 \eta x_e^T(k)] x_e(k) \\ &= u(k) + 2 \eta x_e^T(k) x_e(k) > u(k) \end{aligned}$$

δεδομένου ότι $x_e^T(k) x_e(k) = \|x_e(k)\|^2 > 0$ και $\eta > 0$.

- Έστω $o(k)=1$ και $t(k)=-1 \rightarrow t(k) - o(k) = -2$

Ισχύει ότι $u(k)>0$ και επιθυμούμε $u(k+1) < u(k)$.

$$\begin{aligned} u(k+1) &= w_e^T(k+1) x_e(k) = [w_e^T(k) - 2 \eta x_e^T(k)] x_e(k) \\ &= u(k) - 2 \eta x_e^T(k) x_e(k) < u(k) \end{aligned}$$

δεδομένου ότι $x_e^T(k) x_e(k) = \|x_e(k)\|^2 > 0$ και $\eta > 0$.

Αλγόριθμος Εκπαίδευσης perceptron με διόρθωση σφάλματος

- Σε ότι αφορά την **επιλογή του παραδείγματος** εκπαίδευσης που χρησιμοποιείται σε κάποια επανάληψη, αυτή μπορεί να γίνεται τυχαία, αλλά συνήθως χρησιμοποιούμε τα παραδείγματα **διαδοχικά το ένα μετά το άλλο** (έτσι εξασφαλίζουμε ότι όλα χρησιμοποιούνται για εκπαίδευση ίσο αριθμό φορές).
- Όταν ολοκληρωθεί **ένας κύκλος εκπαίδευσης** κατά τον οποίο όλα τα παραδείγματα έχουν διαδοχικά χρησιμοποιηθεί μια φορά το καθένα για την ενημέρωση των βαρών ενός ΤΝΔ λέμε ότι έχει συμπληρωθεί μια **εποχή (epoch)**.

Αλγόριθμος Εκπαίδευσης perceptron με διόρθωση σφάλματος

- **Τερματισμός του αλγορίθμου:** όταν όλα τα παραδείγματα εκπαίδευσης έχουν δοκιμαστεί (**εποχή**) και δεν προκύπτει κανένα σφάλμα ταξινόμησης.
- Αυτό θα συμβεί μόνο στην περίπτωση που το σύνολο εκπαίδευσης είναι γραμμικά διαχωρίσιμο.
- Στην αντίθετη περίπτωση ο αλγόριθμος δεν τερματίζει διότι το υπερεπίπεδο δεν επαρκεί ως επιφάνεια διαχωρισμού.
- Δεν μπορούμε γνωρίζουμε εκ των προτέρων εάν ένα σύνολο παραδειγμάτων είναι γραμμικά διαχωρίσιμο.
- Για τον τερματισμό του αλγορίθμου διόρθωσης σφάλματος μετράμε τον αριθμό των σφαλμάτων σε κάθε εποχή. Εάν αυτός δεν μηδενιστεί για κάποιο μέγιστο αριθμό εποχών σταματάμε.

Αλγόριθμος Εκπαίδευσης perceptron με διόρθωση σφάλματος

- Το perceptron δεν μπορεί να λύσει το πρόβλημα XOR.

