Αναγνώριση Προτύπων

Ταξινόμηση

Τεχνητά νευρωνικά δίκτυα

Intelligent Systems
& Software Engineering
Labgroup

Ανδρέας Λ. Συμεωνίδης

Αν. Καθηγητής

Τμήμα Ηλεκτρολόγων Μηχ/κών
&
Μηχ/κών Υπολογιστών, Α.Π.Θ.

Email: asymeon@eng.auth.gr

ARISTOTLE
UNIVERSITY OF
THESSALONIKI

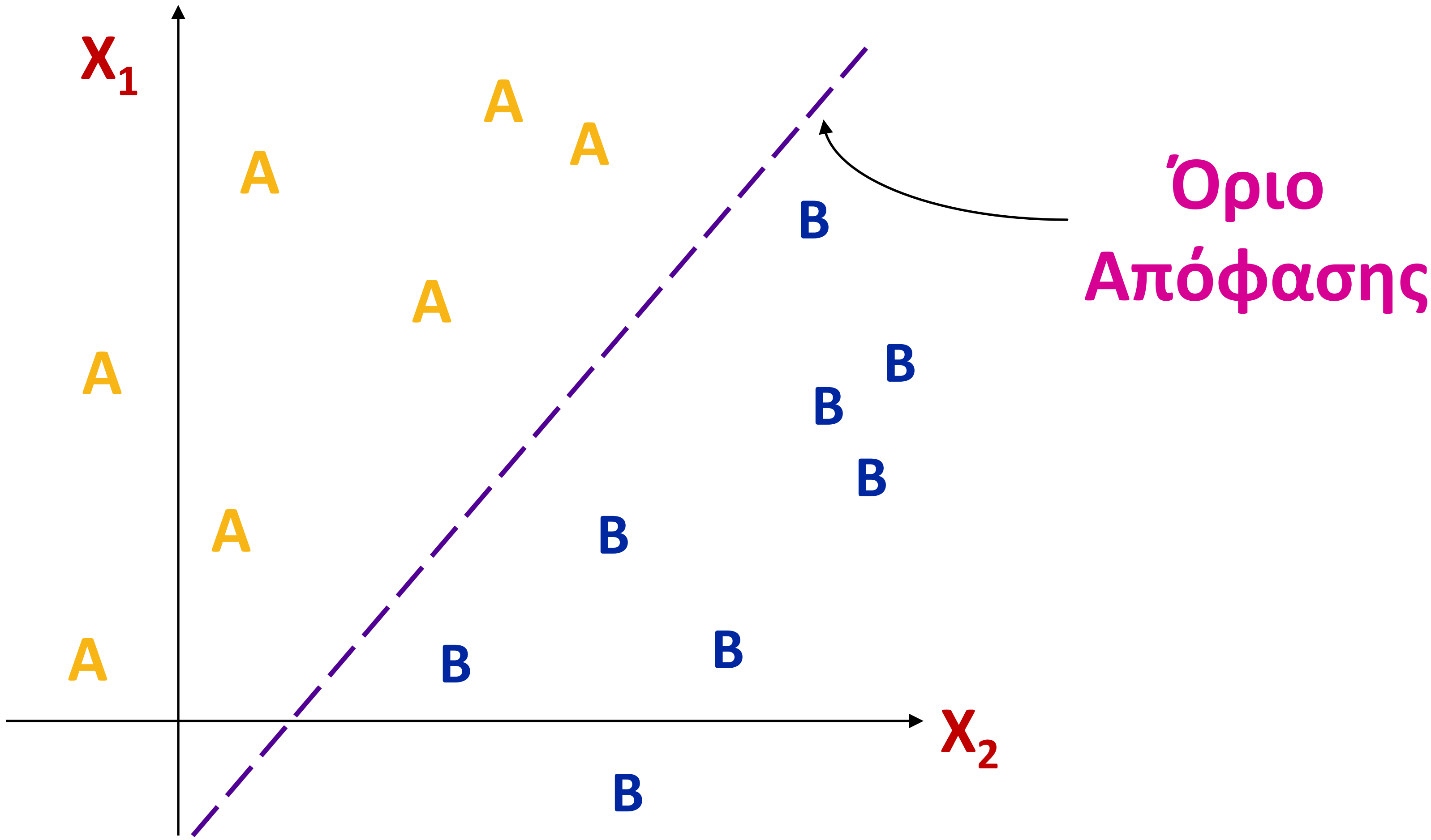
Διάρθρωση διάλεξης

- Τεχνητά νευρωνικά δίκτυα
- Perceptrons

Τεχνητά Νευρωνικά Δίκτυα

- Γραμμικός διαχωρισμός
- Τεχνητά νευρωνικά δίκτυα
- Perceptrons – Μέθοδος Ελαχίστων Τετραγώνων

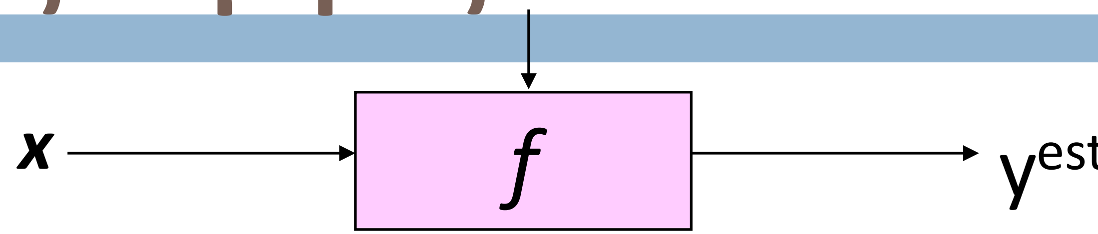
Γραμμικά διαχωρίσιμα προβλήματα



Όρια απόφασης

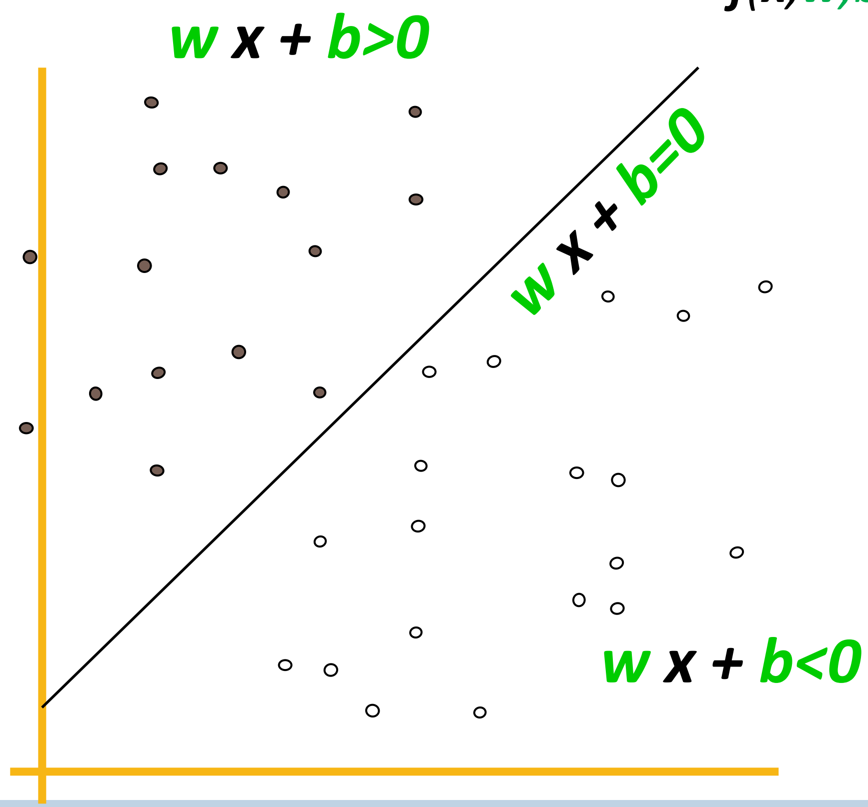
- Σε απλές περιπτώσεις χωρίζει τον χώρο των μεταβλητών ορίζοντας ένα υπερεπίπεδο που το διασχίζει.
- Γνωστό ως όριο απόφασης.
- Συνάρτηση διαχωρισμού: επιστρέφει διαφορετικές τιμές στις διαφορετικές πλευρές του υπερεπιπέδου.
- Κατά συνέπεια, επιλύει προβλήματα γραμμικής ταξινόμησης.

Γραμμικοί ταξινομητές



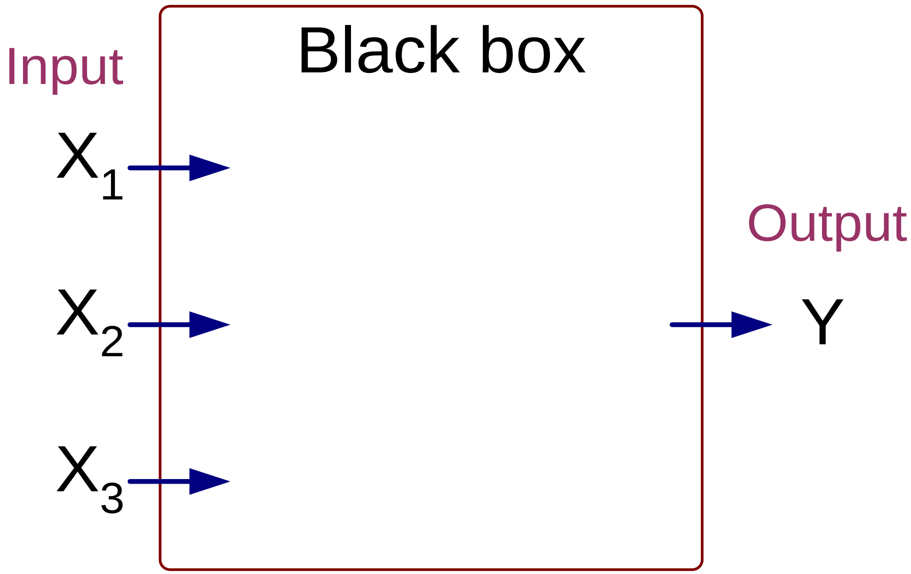
$f(\boldsymbol{x},\boldsymbol{w},b) = sign(\boldsymbol{w}\,\boldsymbol{x} + b)$

- ● : Δηλώνει +1
- ○ : Δηλώνει -1



Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks - ANN)

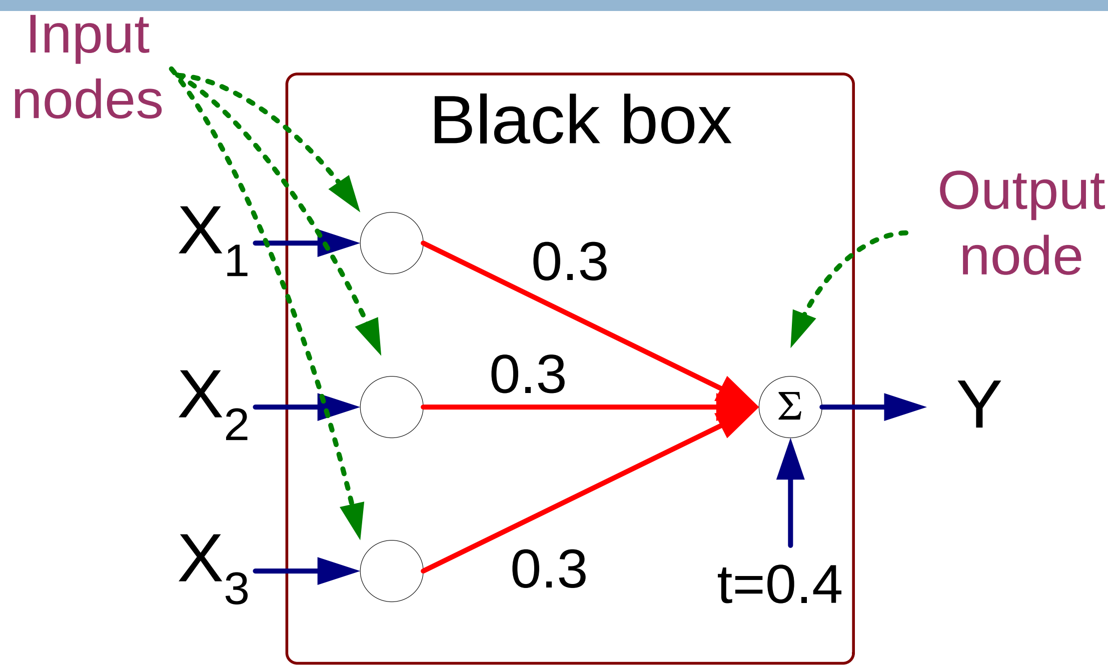
X_1	X_2	X_3	Y
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	0
0	1	0	0
0	1	1	1
0	0	0	0



Η έξοδος Y είναι 1 εάν τουλάχιστον 2 από τις 3 εισόδους είναι ίσες με 1.

Τεχνητά Νευρωνικά Δίκτυα - I

X_1	X_2	X_3	Y
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	0
0	1	0	0
0	1	1	1
0	0	0	0

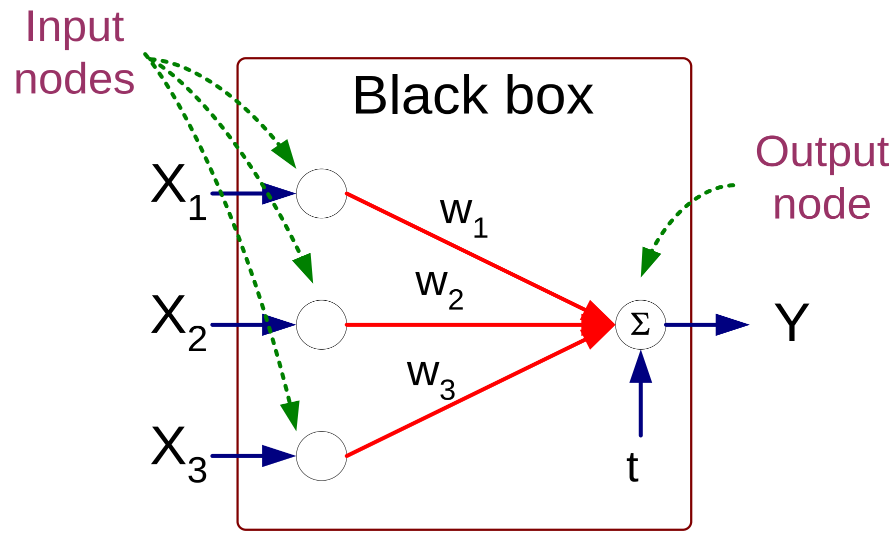


$$Y = I(0.3X_1 + 0.3X_2 + 0.3X_3 - 0.4 > 0)$$

$$\text{όπου } I(z) = \begin{cases} 1 & \text{εάν } z \text{ αληθές} \\ 0 & \text{αλλιώς} \end{cases}$$

Τεχνητά Νευρωνικά Δίκτυα - ΙΙ

- Το μοντέλο είναι μια σύνθεση από συνδεδεμένους κόμβους και συνδέσμους με βάρη
- Ο κόμβος εξόδου αθροίζει κάθε μια από τις τιμές εισόδου του, ανάλογα με τα βάρη των συνδέσμων
- Σύγκριση του κόμβου εξόδου με ένα κατώφλι t

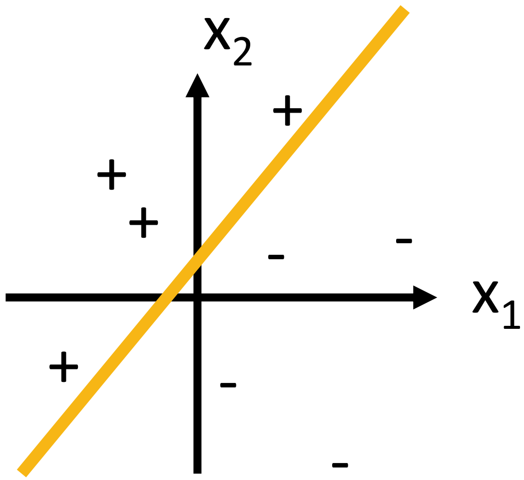


Μοντέλο **Perceptron**

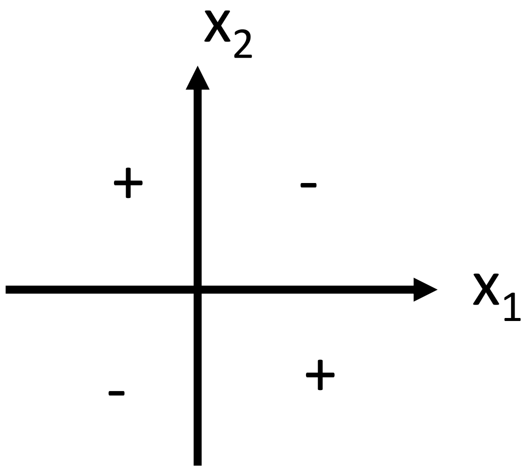
$$Y = I\left(\sum_i w_i X_i - t\right) \text{ ή}$$

$$Y = sign\left(\sum_i w_i X_i - t\right)$$

Το υπερεπίπεδο απόφασης του Perceptron



Γραμμικά διαχωρίσιμο



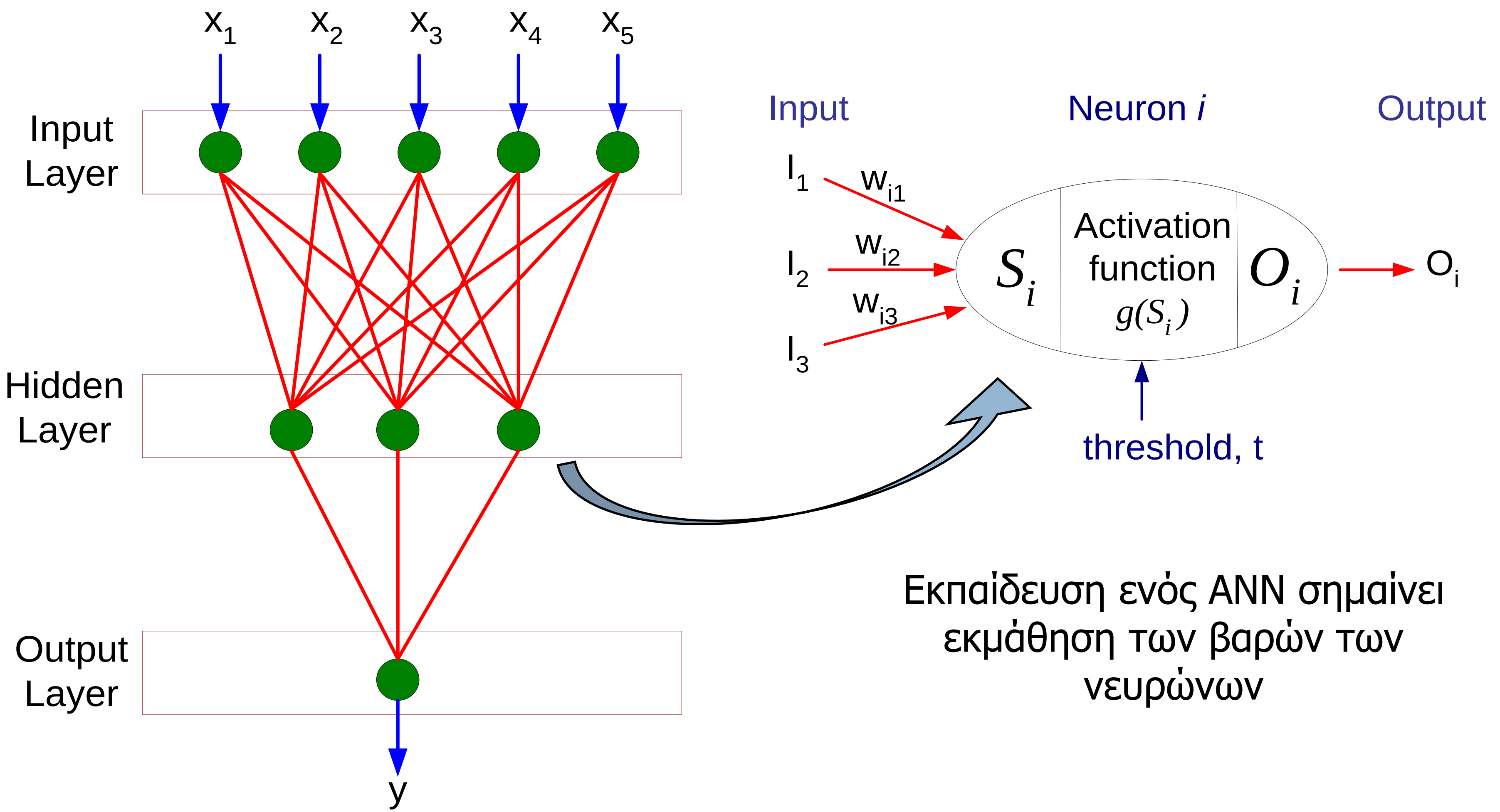
Μη γραμμικά διαχωρίσιμο

- Το Perceptron μπορεί να αναπαραστήσει μερικές πολύ χρήσιμες συναρτήσεις (π.χ. AND(x_1,x_2))
- Συναρτήσεις που δεν είναι γραμμικά διαχωρίσιμες (π.χ. XOR(x_1,x_2)) δεν μπορούν να αναπαρασταθούν

Τμηματοποίηση του υπερεπιπέδου

- Ένα επιπλέον επίπεδο μοντελοποιεί κυρτά τμήματα
  - “Μια περιοχή χωρίς «βαθουλώματα»”
  - Το Perceptron μοντελοποιεί, αλλά δεν μπορεί να μάθει
  - Μια Sigmoid συνάρτηση μαθαίνει κυρτά τμήματα
  - Δυο επίπεδα συναθροίζουν κυρτά τμήματα
- Στη θεωρία, τα περισσότερα επίπεδα δεν προσθέτουν κάτι
- Στην πράξη, τα περισσότερα επίπεδα προσθέτουν!

Γενική δομή ενός ANN



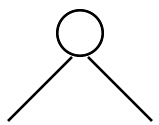
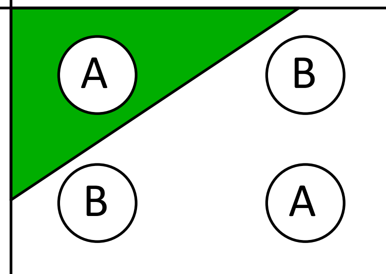
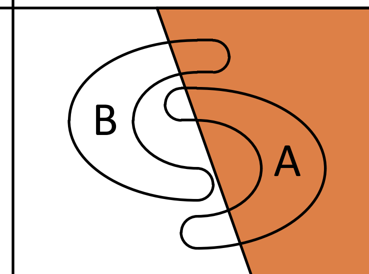
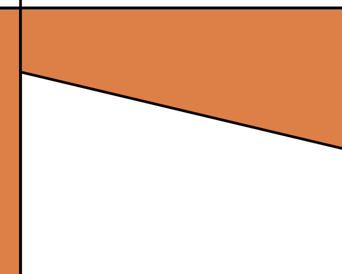
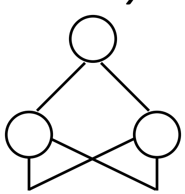
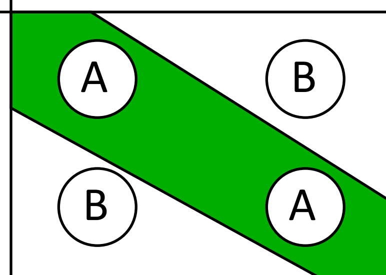
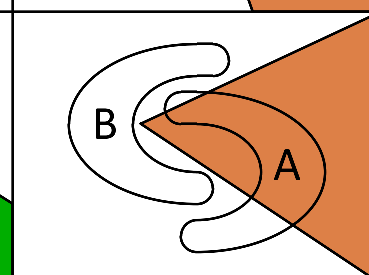
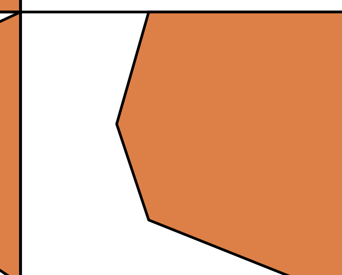
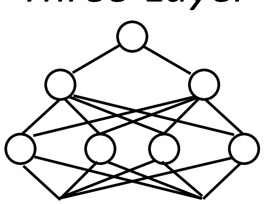
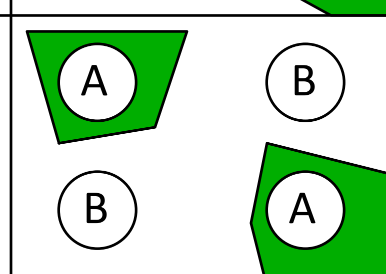
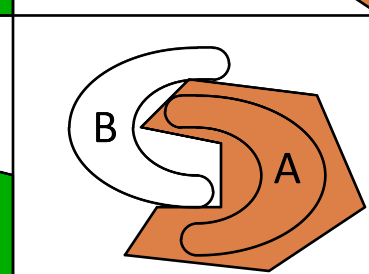
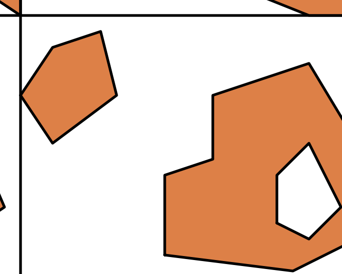
Εκπαίδευση ενός ANN σημαίνει εκμάθηση των βαρών των νευρώνων

Τυπικές συναρτήσεις ενεργοποίησης

- Αυστηρού κατωφλίου (σύμφωνα με το βιολογικό παράδειγμα)
  - Είτε ενεργοποιείται είτε όχι
- Sigmoid συναρτήσεις ('S'-shaped καμπύλες)
  - Η logistic συνάρτηση
  - Η υπερβολική εφαπτομένη (συμμετρική)
    - Και οι δυο έχουν απλές διαφορικές
    - Μόνο το σχήμα μετράει

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

Διάφοροι τύποι μη γραμμικών προβλημάτων

Δομή	*Τύποι περιοχών αποφάσεων*	*Exclusive-OR πρόβλημα*	*Κλάσεις με μικτές περιοχές*	*Πιο γενικά σχήματα περιοχών*
Single-Layer 	*Half Plane Bounded By Hyperplane*			
Two-Layer 	*Convex Open Or Closed Regions*			
Three-Layer 	Arbitrary (Complexity Limited by No. of Nodes)			

Perceptrons

- Εκπαίδευση σε προβλήματα 2 κλάσεων
- Μέθοδος Ελαχίστων Τετραγώνων
- Προβλήματα πολλαπλών κλάσεων

Εκπαίδευση Perceptron

- Θεωρήστε ότι δημιουργείτε διανύσματα που προκύπτουν από δυο κλάσεις C_1 and C_2, έστω:
  - $x_1(1), x_1(2), x_1(3), \ldots$. ανήκουν στο C_1, και
  - $x_2(1), x_2(2), x_2(3), \ldots$. ανήκουν στο C_2
- Επίσης, θεωρήστε ότι επανορίζουμε τα διανύσματα **x**(n) και **w**(n) ώστε να συμπεριλαμβάνουν μια πόλωση (bias), έστω:
  - $\mathbf{x}(n)=[+1,x_1(n),\ldots,x_m(n)]^T$; και
  - $\mathbf{w}(n)=[b(n),w_1(n),\ldots,w_m(n)]^T \quad \mathbf{x}, \mathbf{w} \in \Re^m$

Εκπαίδευση Perceptron (συν.)

- Τότε θα πρέπει να υπάρχει ένα διάνυσμα βαρών **w** τέτοιο ώστε:
  - $\mathbf{w}^T\mathbf{x} > 0$, όταν το **x** ανήκει στην κλάση C_1; και
  - $\mathbf{w}^T\mathbf{x} \leq 0$, όταν το **x** ανήκει στην κλάση C_2;

- Επιλέγουμε αυθαίρετα σε ποια κλάση ανήκει το $\mathbf{w}^T\mathbf{x}=0$ (έστω στην κλάση C_2)
- Ο αλγόριθμος εκπαίδευσης των βαρών του perceptron μπορεί να οριστεί ως εξής:

Εκπαίδευση Perceptron (συν.)

1. Εάν το n^th^ στοιχείο του σετ εκπαίδευσης, **x**(n) είναι σωστά ταξινομημένο από το τρέχον σετ βαρών, το **w**(n) δεν χρειάζεται διόρθωση. Ισχύει δηλαδή:
   - **w**(n+1)=**w**(n),

     εάν $\mathbf{w}^T\mathbf{x}(n) > 0$ και το **x**(n) ανήκει στην κλάση C_1;
   - **w**(n+1)=**w**(n),

     εάν $\mathbf{w}^T\mathbf{x}(n) \leq 0$ και το **x**(n) ανήκει στην κλάση C_2;

Εκπαίδευση Perceptron (συν.)

2. Αλλιώς, το διάνυσμα βαρών προσαρμόζεται σύμφωνα με τον παρακάτω κανόνα:
   - $\mathbf{w}(n+1)=\mathbf{w}(n) - \eta(n)\mathbf{x}(n)$,

     εάν $\mathbf{w}^T\mathbf{x}(n) > 0$ και $\mathbf{x}(n)$ ανήκει στην κλάση C_2;
   - $\mathbf{w}(n+1)=\mathbf{w}(n)+ \eta(n)\mathbf{x}(n)$**,**

     εάν $\mathbf{w}^T\mathbf{x}(n) \leq 0$ και $\mathbf{x}(n)$ ανήκει στην κλάση C_1;

- Η παράμετρος $\eta(n)$ ονομάζεται *ρυθμός εκπαίδευσης (learning rate)* και ελέγχει το ποσοτό προσαρμογής του διανύσματος βαρών κατά την *n* επανάληψη.

Εκπαίδευση Perceptron (συν.)

- Εάν θεωρήσουμε ότι η *επιθυμητή απόκριση* (desired response) δίνεται από:

$$d(n) = \begin{cases} +1 & \text{if} \quad x(n) \quad \text{belongs} \quad \text{to} \quad \text{class} \quad C_1 \\ -1 & \text{if} \quad x(n) \quad \text{belongs} \quad \text{to} \quad \text{class} \quad C_2 \end{cases}$$

τότε μπορούμε να ξαναγράψουμε τον κανόνα προσαρμογής (εκπαίδευσης) στη μορφή ενός κανόνα *διόρθωσης σφάλματος* (error-correction):

- $\mathbf{w}(n+1)=\mathbf{w}(n)+ \eta[d(n)-y(n)]\mathbf{x}(n)$

  όπου $e(n)= \eta[d(n)-y(n)]$ είναι το *σήμα σφάλματος*

- Ο ρυθμός εκπαίδευσης είναι μια θετική σταθερά στο $0 < \eta \leq 1$.

Εκπαίδευση Perceptron (συν.)

- Όταν δίνουμε τιμή στο η μέσα στο εύρος (0,1] πρέπει να έχουμε στο μυαλό μας δυο αντικρουόμενες απαιτήσεις:
  - Η συνάθροιση (*Averaging*) των προηγούμενων εισόδων, η οποία οδηγεί και σε σταθερές εκτιμήσεις βαρών, απαιτεί μικρή τιμή για το η
  - Η γρήγορη σύγκλιση (*Fast adaptation*) σε σχέση με τις πραγματικές αλλαγές στις κατανομές της διαδικασίας που είναι υπεύθυνη για τη δημιουργία του διανύσματος εισόδου **x**, απαιτεί μεγάλη τιμή για το η

Εκπαίδευση με τη μέθοδο Ελαχίστων τετραγώνων (LSF)

- Εάν οι κλάσεις είναι γραμμικά διαχωρίσιμες, τότε το perceptron δίνει έξοδο ± 1
- Εάν οι κλάσεις είναι **<u>ΔΕΝ</u>** γραμμικά διαχωρίσιμες, τότε πρέπει να υπολογίσουμε τα βάρη $w_1, w_2, \ldots, w_0$ ώστε η διαφορά ανάμεσα:
  - Στην πραγματική έξοδο του ταξινομητή $w^T x$ και
  - Τις επιθυμητές εξόδους, π.χ.

$$+1, \text{ αν } x \in \omega_1$$

$$-1, \text{ αν } x \in \omega_2$$

να είναι όσο το δυνατόν ΜΙΚΡΟΤΕΡΗ.

Εκπαίδευση με τη μέθοδο Ελαχίστων τετραγώνων (LSF) - II

- ΜΙΚΡΗ, ως προς τη λογική του μέσου τετραγωνικού σφάλματος, σημαίνει επιλογή κατάλληλου, ώστε η συνάρτηση κόστους:
    - $J(w) \equiv E[(y - w^T x)^2]$ γίνεται ελάχιστη
    - $\hat{w} = \arg\min_{w} J(w)$
    - y είναι η αντίστοιχη επιθυμητή απόκριση

Εκπαίδευση με τη μέθοδο Ελαχίστων τετραγώνων (LSF) - III

- Ελαχιστοποίηση του $J(w)$ ως προς w:

$$\frac{\partial J(w)}{\partial w} = \frac{\partial}{\partial w} E[(y - w^T x)^2] = 0 = 2E[x(y - x^T w)] \Rightarrow$$

$$E[xx^T]w = E[xy] \Rightarrow \hat{w} = \boxed{R_x^{-1} E[xy]}$$

- Όπου R_x είναι ο πίνακας αυτοσυσχέτισης

$$R_x \equiv E[xx^T] = \begin{bmatrix} E[x_1 x_1] & E[x_1 x_2] \ldots & E[x_1 x_l] \\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots \\ E[x_l x_1] & E[x_l x_2] \ldots & E[x_l x_l] \end{bmatrix}$$

- και $E[xy] = \begin{bmatrix} E[x_1 y] \\ \cdots \\ E[x_l y] \end{bmatrix}$ το διάνυσμα ετεροσυσχέτισης (crosscorrelation)

Γενίκευση για προβλήματα πολλαπλών κλάσεων

- Στόχος είναι ο υπολογισμός M γραμμικών συναρτήσεων διαχωρισμού: $g_i(x) = w_i^T x$

  σύμφωνα με το MSE.

- Υιοθετήστε ως επιθυμητές αποκρίσεις y_i:

$$y_i = 1 \quad \text{if} \quad \underline{x} \in c_i$$
$$y_i = 0 \quad \text{otherwise}$$

- Έστω: $y = [y_1, y_2, \ldots, y_M]^T$

- Και ο πίνακας: $W = [w_1, w_2, \ldots, w_M]$

Γενίκευση για προβλήματα πολλαπλών κλάσεων

- Στόχος είναι ο υπολογισμός του W:

$$\hat{W} = \arg\min_{W} E[\|y - W^T x\|^2] = \arg\min_{W} E\left[\sum_{i=1}^{M} (y_i - w_i{}^T \cdot x)^2\right]$$

- Το παραπάνω είναι ισοδύναμο με έναν αριθμό M από MSE προβλήματα ελαχιστοποίησης.
- Μεθοδολογία:
  - Σχεδιάστε κάθε w_i ώστε η επιθυμητή έξοδος είναι 1 για $x \in c_i$ και 0 για κάθε άλλη κλάση
  - Σημείωση: Το κριτήριο MSE ανήκει σε μια πιο γενική ομάδα συναρτήσεων κόστους με την εξής βασική ιδιότητα:
    - Η τιμή του $g_i(x)$ είναι εκτίμηση, με τη λογική του MSE, της $P(c_i \mid x)$ εκ των υστέρων πιθανότητας, δεδομένου ότι οι επιθυμητές έξοδοι που χρησιμοποιούνται κατά τη διάρκεια της εκπαίδευσης είναι $y_i = 1, x \in c_i$ και 0 διαφορετικά.

Πηγές

- Introduction to Data Mining, Tan, Steinbach, Kumar.
- Pattern recognition, Theodoridis & Koutroumbas.