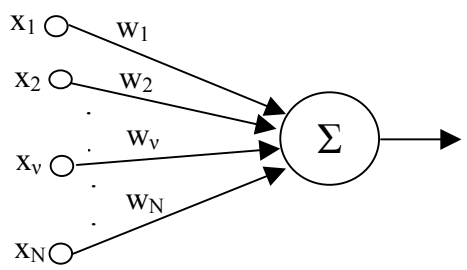
4.3. Γραμμικοί ταξινομητές

Γραμμικός ταξινομητής είναι ένα σύστημα ταξινόμησης που χρησιμοποιεί γραμμικές διακριτικές συναρτήσεις. Οι ταξινομητές αυτοί αναπαρίστανται συχνά με ομάδες κόμβων εντός των οποίων υπολογίζονται οι ποσότητες $\mathbf{w}^T \cdot \mathbf{x}$ όπου **w** ο πίνακας των συντελεστών των γραμμικών διακριτικών συναρτήσεων και **x** ο πίνακας προτύπου που θεωρείται ως είσοδος στον κόμβο (Σχήμα 4.3-1).



Σχήμα 4.3-1

Όπως φαίνεται στο σχήμα οι συντελεστές w_ν σημειώνονται επί των γραμμών που οδηγούν από τις εισόδους x_ν στον κόμβο και ονομάζονται συνάψεις. Ένας τέτοιος κόμβος ονομάζεται νευρώνας λόγω μιας ομοιότητας του με τα ομώνυμα βιολογικά κύτταρα του ανθρώπινου εγκεφάλου. Συστήματα διασυνδεδεμένων νευρώνων συνθέτουν τα λεγόμενα *τεχνητά νευρωνικά δίκτυα* (*ΤΝΔ*), (*Artificial Neural Networks*). Διάφοροι τύποι νευρωνικών δικτύων επιτελούν χρήσιμες εργασίες στην αναγνώριση προτύπων. Προς το παρόν θα ασχοληθούμε με νευρωνικά δίκτυα που περιγράφουν γραμμικούς ταξινομητές. Ένα τέτοιο γραμμικό νευρωνικό δίκτυο ονομάζεται "*perceptron*"

4.4. Εκπαίδευση γραμμικών ταξινομητών δυο κλάσεων

Η εκπαίδευση γραμμικών ταξινομητών ή γραμμικών νευρωνικών δικτύων Perceptron είναι μια επαναληπτική διαδικασία διόρθωσης σφάλματος. Για την απλούστερη μαθηματική διατύπωση του μηχανισμού διόρθωσης θα θεωρήσουμε τους επαυξημένους πίνακες των προτύπων. Αν δηλαδή ο χώρος των πινάκων των Προτύπων έχει N διαστάσεις (E^N) οι σχέσεις θα γραφούν στον χώρο E^{N+1} που έχει N+1 διαστάσεις. Η διαδικασία εν συντομία έχει ως εξής:

Αν A και B οι δυο κλάσεις επιλέγουμε τυχαία ένα πρότυπο με πίνακα $\mathbf{x}^A$ από την A τάξη και ένα πρότυπο με πίνακα $\mathbf{x}^B$ από την B τάξη. Ακολούθως φέρουμε το μεσοκάθετο υπερ-επιπέδου στο ευθύγραμμο τμήμα με άκρα τα άκρα των $\vec{\mathbf{x}}^A$ και $\vec{\mathbf{x}}^B$. Το μεσοκάθετο υπερ-επιπέδου είναι το ενδιάμεσο σημείο αν $\mathbf{x}^A, \mathbf{x}^B \in E^1$, η μεσοκάθετος ευθεία αν $\mathbf{x}^A, \mathbf{x}^B \in E^2$, το μεσοκάθετο υπερ-επιπέδου αν $\mathbf{x}^A, \mathbf{x}^B \in E^3$, και η μεσοκάθετος υπερ-επιπέδου αν $\mathbf{x}^A, \mathbf{x}^B \in E^N, N > 3$.

Αν η εξίσωση του μεσοκάθετου υπερ-επιπέδου στο E^N είναι

(4.4.1) $\mathbf{w}^T \cdot \mathbf{x} + w_{N+1} = 0$

μπορεί να γραφεί στο E^{N+1} ως

(4.4.2) $\mathbf{W}^T \cdot \tilde{\mathbf{x}} = 0$ οπου $\mathbf{W} = [\mathbf{w}, w_{N+1}]^T$ και $\tilde{\mathbf{x}} = [\mathbf{x}, 1]$

Θεωρούμε τώρα την ποσότητα $D(\mathbf{X}) = \mathbf{W}^T \cdot \mathbf{X}$ όπου $\mathbf{X} = [\mathbf{x}, x_{N+1}]^T \in E^{N+1}$. Η σχέση $D(\mathbf{X}) = 0$ περιγράφει ένα υπερεπίπεδο που διέρχεται από το κέντρο των αξόνων του E^{N+1} και τέμνει το επίπεδο των επαυξημένων πινάκων των προτύπων στο μεσοκάθετο υπερεπίπεδο τους E^N σύμφωνα με την (4.2.13). Αν ισχύει $D(\tilde{\mathbf{x}}_A) > 0$, τότε θα ισχύει $D(\tilde{\mathbf{x}}_B) < 0$ και αντίστροφα.

Η βασική ιδέα της εκμάθησης είναι η διόρθωση του υπερεπιπέδου $D(\mathbf{X}) = 0$ του E^{N+1}, κάθε φορά που ταξινομείται εσφαλμένα κάποιο πρότυπο με πίνακα $\tilde{\mathbf{x}}$. Δηλαδή, οι πίνακες των πρότυπων της κλάσης A πρέπει να

καθιστούν το πρόσημο της D(**X**) ίδιο με αυτό της ποσότητας $D(\tilde{\mathbf{x}}_A)$. Αντίστοιχα πρέπει να συμβαίνει με τους πίνακες των προτύπων της κλάσης Β. Όταν δεν συμβαίνει αυτό, σημαίνει ότι το υπερεπιπεδο πρέπει να διορθωθεί ώστε να αφήνει το πρότυπο προς την πλευρά των ορθώς ταξινομημένων προτύπων. Η διόρθωση αυτή γίνεται σύμφωνα με την σχέση

(4.4.3) $\mathbf{W}_{t+1} = \mathbf{W}_t \pm \rho \cdot \tilde{\mathbf{x}}$

Η ποσότητα ρ λέγεται παράμετρος εκμάθησης και παίρνει τιμές στο διάστημα (0,1) και t είναι μετρητής επανάληψης της διαδικασίας εκμάθησης. Το πρόσημο στην (35) είναι αυτό της κλάσης του προτύπου όπως αυτό προσδιορίσθηκε από την ποσότητα $D(\tilde{\mathbf{x}}_A)$. Τα παραπάνω παρουσιάζονται ακολούθως βηματικά.

<u>Βήμα 1ο :</u> Επιλέγουμε τυχαία ένα πρότυπο από την κλάση Α με πίνακα $\mathbf{x}^A$ και ένα από την κλάση Β με πίνακα $\mathbf{x}^B$, με $\mathbf{x}^A, \mathbf{x}^B \in E^N$.

<u>Βήμα 2ο :</u> Φέρουμε το μεσοκάθετο υπερεπίπεδο στο ευθύγραμμο τμήμα με άκρα τα σημεία $\mathbf{x}^A$, $\mathbf{x}^B$ σύμφωνα με τις σχέσεις (4.2.13).

<u>Βήμα 3ο :</u> Ορίζουμε ένα μετρητή επανάληψης t=0,1,2,3,... και του αποδίδουμε αρχικά την τιμή μηδέν. Θεωρούμε την ποσότητα

(4.4.4) $D_t(\mathbf{X})=\mathbf{W}^T\mathbf{X}$ όπου $\mathbf{W}=[\mathbf{w}, w_{N+1}]^T$, $\mathbf{X} \in E^{N+1}$.

<u>Βήμα 4ο :</u> Υπολογίζουμε την ποσότητα $D_t(\tilde{\mathbf{x}}^A)$ και βρίσκουμε το πρόσημο κάθε κλάσης που θα είναι

	Πρόσημο κλάσης Α	Πρόσημο κλάσης Β
$D_0(\tilde{\mathbf{x}}^A) > 0$	(+)	(-)
$D_0(\tilde{\mathbf{x}}^A) < 0$	(-)	(+)

<u>Βήμα 5ο:</u> Επιλέγουμε ένα τυχαίο πρότυπο Π με πίνακα $\tilde{\mathbf{x}}$, από το σύνολο εκπαίδευσης και υπολογίζουμε την ποσότητα $D_t(\tilde{\mathbf{x}})$. Αν το πρόσημο της είναι ίδιο με το πρόσημο της κλάσης του Π το πρότυπο έχει ταξινομηθεί σωστά και $\mathbf{W}_{t+1}=\mathbf{W}_t$, αλλιώς ο $\mathbf{W}_{t+1}$ διορθώνεται σύμφωνα με την σχέση

(4.4.5) $\mathbf{W}_{t+1} = \mathbf{W}_t \pm \rho \cdot \tilde{\mathbf{x}}$ (37)

με πρόσημο το πρόσημο της κλάσης του Π.

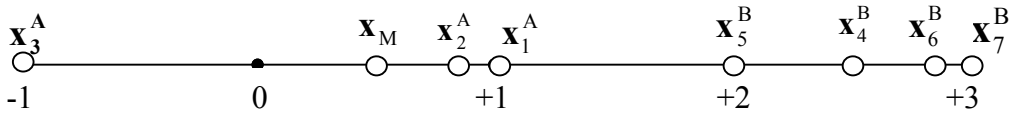
Βήμα 6ο: Αυξάνεται η τιμή του t κατά ένα και η διαδικασία επαναλαμβάνεται από το βήμα 5 έως ότου όλα τα πρότυπα να ταξινομηθούν σωστά.

ΠΑΡΑΔΕΙΓΜΑ

Έστω τα πρότυπα με ένα χαρακτηριστικό ($\mathbf{x} = [x1]$) που χωρίζονται σε δύο κλάσεις Α, Β και με τιμές

$$\mathbf{x}_1^A = [1], \mathbf{x}_2^A = [0.8], \mathbf{x}_3^A = [-1]$$
$$\mathbf{x}_4^B = [2.5], \mathbf{x}_5^B = [2], \mathbf{x}_6^B = [2.8], \mathbf{x}_7^B = [3]$$

Οι πίνακες των προτύπων ανήκουν στο χώρο E^1 και απεικονίζονται στην ευθεία των πραγματικών αριθμών όπως φαίνεται στο Σχ. 4.4-1.



Σχήμα 4.4-1

Επιλέγουμε τυχαία ένα πρότυπο από κάθε κλάση, έστω τα $\mathbf{x}_3^A$ και $\mathbf{x}_5^B$. Το ενδιάμεσο σημείο $\mathbf{x}_M$ των $\mathbf{x}_3^A$ και $\mathbf{x}_5^B$, θα ικανοποιεί τις σχέσεις (4.2.13) και θα ισχύει

$$\mathbf{w} \cdot \mathbf{x}_m + w_2 = 0, \quad \mathbf{w} = [w_1] \in E^1 \text{ και } w_1, w_2 \in R$$
$$\mathbf{w} = \mathbf{x}_3^A - \mathbf{x}_5^B = [-1] - [2] = [-3] \Rightarrow w_1 = -3$$
(4.4.6) $$w_2 = -\frac{1}{2}(\mathbf{x}_3^{A^T} \cdot \mathbf{x}_3^A - \mathbf{x}_5^{B^T} \cdot \mathbf{x}_5^B) = -\frac{1}{2}([-1][-1] - [2][2]) = \frac{3}{2}$$
$$\text{'Αρα } -3 \cdot x_M + \frac{3}{2} = 0 \Rightarrow x_m = \frac{1}{2}$$

Στον επαυξημένο χώρο E^2 η σχέση

(4.4.7) $-3 \cdot x_M + \frac{3}{2} = 0$

γράφεται

(4.4.8) $\mathbf{W}^T \cdot \tilde{\mathbf{x}}_M = 0$ με $\mathbf{W} = [-3, \frac{3}{2}]$ και $\tilde{\mathbf{x}}_M = [\mathbf{x}_M, 1]$

Η σχέση

(4.4.9) $\mathbf{W}^T \cdot \mathbf{X} = 0$ με $\mathbf{X} = [x_1, x_2] \in E^2$

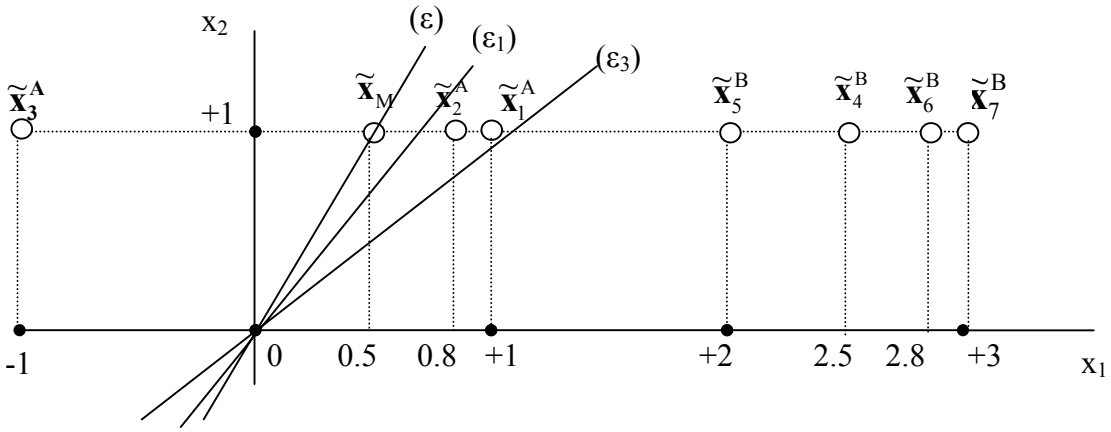
περιγράφει την ευθεία (ε) με εξίσωση :

$$\mathbf{W}^T \cdot \mathbf{X} = 0 \Rightarrow \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \cdot [x_1, x_2] = 0 \Rightarrow$$

(4.4.10) $w_1 \cdot x_1 + w_2 \cdot x_2 = 0 \Rightarrow -3 \cdot x_1 + \frac{3}{2} x_2 = 0 \Rightarrow$

$(\varepsilon) \rightarrow x_2 - 2x_1 = 0$

Στο Σχ.4.4-2 φαίνονται τα άκρα των επαυξημένων διανυσμάτων των προτύπων και η ευθεία (ε).



Σχήμα 4.4-2

Θεωρούμε την ποσότητα

(4.4.11) $D(\mathbf{X}) = \mathbf{W}^T \cdot \mathbf{X} = \begin{bmatrix} -3, & \frac{3}{2} \end{bmatrix} \begin{bmatrix} x1 \\ x2 \end{bmatrix}$

Ισχύει

(4.4.12) $$D(\tilde{\mathbf{x}}_5^B) = \left[-3, \frac{3}{2}\right]\begin{bmatrix}2\\1\end{bmatrix} = (-3)\cdot 2 + \left[\frac{3}{2}\right]\cdot 1 = -4.5 < 0.$$

Θα πρέπει ως εκ τούτου οι επαυξημένοι πίνακες των προτύπων της κλάσης B να καθιστούν την $D(\mathbf{X}) < 0$ και αυτοί της κλάσης A το $D(\mathbf{X}) > 0$. Εάν αυτό δεν ισχύει η ποσότης $D(\mathbf{X})$ πρέπει να διορθωθεί. Προς τούτο ελέγχουμε την ποσότητα $D(\mathbf{X})$ για όλα τα πρότυπα με τυχαία σειρά και διορθώνουμε την $D(\mathbf{X})$ όταν χρειαστεί. Συγκεκριμένα για ρ = 0.4 έχουμε τα ακόλουθα βήματα:

t	$\tilde{\mathbf{x}} = \begin{bmatrix}x_1\\1\end{bmatrix}$	$\mathbf{W}_t = \begin{bmatrix}w_1\\w_2\end{bmatrix}$	$D(\tilde{\mathbf{x}}) = \mathbf{W}^T \cdot \tilde{\mathbf{x}}$	Πρόσημο κλάσης του $\tilde{\mathbf{x}}$	Διόρθωση $\mathbf{W}_{t+1} = \mathbf{W}_t + p \cdot \tilde{\mathbf{x}}$
0	$\tilde{\mathbf{x}}_1^A = \begin{bmatrix}1\\1\end{bmatrix}$	$\mathbf{W}_0 = \begin{bmatrix}-3\\1.5\end{bmatrix}$	-3·1 + 1.5 = -1.5 < 0	+	$\mathbf{W}_1 = \begin{bmatrix}-3\\1.5\end{bmatrix} + 0.4\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}-2.6\\1.9\end{bmatrix}$
1	$\tilde{\mathbf{x}}_4^B = \begin{bmatrix}2.5\\1\end{bmatrix}$	$\mathbf{W}_1 = \begin{bmatrix}-2.6\\1.9\end{bmatrix}$	-2.6·2.5+1.9 = -4.6 < 0	-	Καμία διόρθωση $\mathbf{W}_2 = \mathbf{W}_1$
2	$\tilde{\mathbf{x}}_2^A = \begin{bmatrix}0.8\\1\end{bmatrix}$	$\mathbf{W}_2 = \begin{bmatrix}-2.6\\1.9\end{bmatrix}$	-2.6·0.8 +1.9= – 1.7 < 0	+	$\mathbf{W}_3 = \begin{bmatrix}-2.6\\1.9\end{bmatrix} + 0.4\begin{bmatrix}0.8\\1\end{bmatrix} = \begin{bmatrix}-1.8\\1.9\end{bmatrix}$
3	$\tilde{\mathbf{x}}_5^B = \begin{bmatrix}2\\1\end{bmatrix}$	$\mathbf{W}_3 = \begin{bmatrix}-1.8\\1.9\end{bmatrix}$	-1.8·2+1.9= -1.7 < 0	-	Καμία διόρθωση $\mathbf{W}_4 = \mathbf{W}_3$
4	$\tilde{\mathbf{x}}_6^B = \begin{bmatrix}2.8\\1\end{bmatrix}$	$\mathbf{W}_4 = \begin{bmatrix}-1.8\\1.9\end{bmatrix}$	-1.8·2.8+1.9 = -3.14 < 0	-	Καμία διόρθωση $\mathbf{W}_5 = \mathbf{W}_4$
5	$\tilde{\mathbf{x}}_3^A = \begin{bmatrix}-1\\1\end{bmatrix}$	$\mathbf{W}_5 = \begin{bmatrix}-1.8\\1.9\end{bmatrix}$	(-1.8) · (-1) + 1.9 = 3.7 > 0	+	Καμία διόρθωση $\mathbf{W}_6 = \mathbf{W}_5$
6	$\tilde{\mathbf{x}}_7^B = \begin{bmatrix}3\\1\end{bmatrix}$	$\mathbf{W}_6 = \begin{bmatrix}-1.8\\1.9\end{bmatrix}$	(-1.8) ·3 + 1.9 = -3.5 < 0	-	Καμία διόρθωση $\mathbf{W}_7 = \mathbf{W}_6$
7	$\tilde{\mathbf{x}}_2^A = \begin{bmatrix}0.8\\1\end{bmatrix}$	$\mathbf{W}_7 = \begin{bmatrix}-1.8\\1.9\end{bmatrix}$	(-1.8)·0.8 + 1.9 = 0.46 > 0	+	Καμία διόρθωση $\mathbf{W}_8 = \mathbf{W}_7$
8	$\tilde{\mathbf{x}}_4^B = \begin{bmatrix}2.5\\1\end{bmatrix}$	$\mathbf{W}_8 = \begin{bmatrix}-1.8\\1.9\end{bmatrix}$	(-1.8) ·2.5 + 1.9 = -2.6 < 0	-	Καμία διόρθωση $\mathbf{W}_9 = \mathbf{W}_8$
9	$\tilde{\mathbf{x}}_1^A = \begin{bmatrix}1\\1\end{bmatrix}$	$\mathbf{W}_9 = \begin{bmatrix}-1.8\\1.9\end{bmatrix}$	-1.8·1+1.9 = 0.1 > 0	+	Καμία διόρθωση

Κατά τα βήματα για t = 3, 4, ….,9 έχουν προσπελασθεί όλα τα πρότυπα έχουν ταξινομηθεί σωστά. Συνεπώς η εξίσωση

(4.4.13) $$\begin{bmatrix} -1.8 \\ 1.9 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Rightarrow \mathbf{W}^T \cdot \mathbf{X} = 0 \quad \text{με} \quad \mathbf{W} = [-1.8, 1.9]^T$$

ή ισοδύναμα

(4.4.14) $$-1.8 \cdot x_1 + 1.9 \cdot x_2 = 0$$

διαχωρίζει τις δύο κλάσεις. Στο Σχ.4.4-2 φαίνονται οι ευθείες (ε_1), (ε_3) για $\mathbf{W} = \mathbf{W}_1$ και $\mathbf{W} = \mathbf{W}_3$ στις αντίστοιχες φάσεις διόρθωσης.

Ακολούθως παρουσιάζεται η υλοποίηση ενός γραμμικού ταξινομητή με την γλώσσα προγραμματισμού *C*, για την εκμάθηση των τιμών της λογικής διάζευξης (OR).

```
#include<stdio.h>
#include<conio.h>
#define K 4
#define N 2

float x[K][N+1]={ 0.0, 0.0, 1.0,
                  0.0, 1.0, 1.0,
                  1.0, 0.0, 1.0,
                  1.0, 1.0, 1.0
                    };
float _class[K] = { 0.0, 1.0, 1.0, 1.0 };
float w[N+1] = { 1.0, 1.0, -1.5};
float Dx, r = 0.2; int k, n; bool error;

main()
{ //---- Training phase ----------
  do
  {error = false;
   for(k = 0; k < K; k++)
   {
    Dx = x[ k][0]*w[ 0] + x[ k][ 1]*w[ 1] + x[ k][ 2]*w[ 2];
//--------------
    printf("w=(%+.2f %+.2f %+.2f | ",w[0], w[1], w[2]);
    printf("x=(%+.2f, %+.2f, %+.f) | ",x[k][0],x[k][1],x[k][2]);
    printf("D(x)=%+.2f, class:%.2f ",Dx, _class[ k]);
//--------------
    if( Dx >=0 && _class[ k] == 0.0)
      { w[0]-=r*x[k][0]; w[1]-=r*x[k][1]; w[2]-=r*x[k][2]; error = true;
        printf("False\n");
      }
    else if(Dx <= 0 && _class[ k] == 1.0)
      { w[0]+=r*x[k][0]; w[1]+=r*x[k][1]; w[2]+=r*x[k][2]; error = true;
        printf("False\n");
      }
    else printf("True\n");
   }
  }while(error);

   printf("D(x) = %+.2fx1 %+.2fx2 %+.2f\n", w[0], w[1], w[2]);
}
```

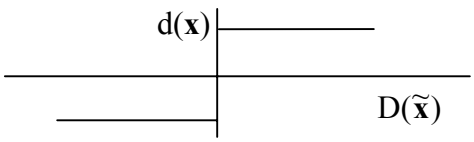
Από την εκτέλεση του προγράμματος προκύπτουν τα ενδιάμεσα στάδια εκπαίδευσης και η ευθεία διαχωρισμού των κλάσεων.

w=(+1.00 +1.00 -1.50 | x=(+0.00, +0.00, +1) | D(x)=-1.50, class:0.00 True
w=(+1.00 +1.00 -1.50 | x=(+0.00, +1.00, +1) | D(x)=-0.50, class:1.00 False
w=(+1.00 +1.20 -1.30 | x=(+1.00, +0.00, +1) | D(x)=-0.30, class:1.00 False
w=(+1.20 +1.20 -1.10 | x=(+1.00, +1.00, +1) | D(x)=+1.30, class:1.00 True
w=(+1.20 +1.20 -1.10 | x=(+0.00, +0.00, +1) | D(x)=-1.10, class:0.00 True
w=(+1.20 +1.20 -1.10 | x=(+0.00, +1.00, +1) | D(x)=+0.10, class:1.00 True
w=(+1.20 +1.20 -1.10 | x=(+1.00, +0.00, +1) | D(x)=+0.10, class:1.00 True
w=(+1.20 +1.20 -1.10 | x=(+1.00, +1.00, +1) | D(x)=+1.30, class:1.00 True
D(x) = +1.20x1 +1.20x2 -1.10

Η απόφαση ταξινόμησης σε μία από τις δύο κλάσεις που διαχωρίζονται από ένα υπερεπίπεδο εξαρτάται τελικώς από το πρόσημο της ποσότητας $D(\tilde{\mathbf{x}})$. Για την περιγραφή με μαθηματικό τρόπο ορίζουμε την συνάρτηση d(x) με την παρακάτω σχέση :

$$(4.4.15) \qquad d(\mathbf{x}) = \begin{cases} +1 \text{ αν } D(\tilde{\mathbf{x}}) = W^T\tilde{x} > 0 \\ -1 \text{ αν } D(\tilde{\mathbf{x}}) = W^T\tilde{x} < 0 \end{cases} \text{ για } D(\tilde{\mathbf{x}}) \in \Re - \{0\}$$

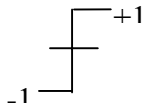
Η τιμή +1 κωδικοποιεί την περίπτωση το πρότυπο να ανήκει στην κλάση



Σχήμα 4.4-3

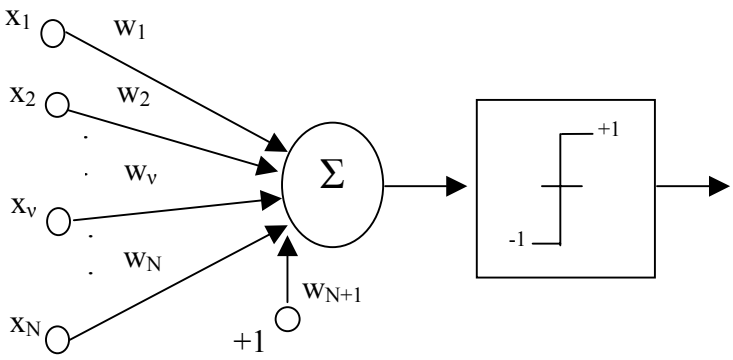
για την οποία $D(\tilde{\mathbf{x}})>0$ και η –1 την περίπτωση το πρότυπο να ανήκει στην κλάση για την οποία $D(\tilde{\mathbf{x}})<\mathbf{0}$. Η γραφική παράσταση των τιμών της d(**x**) σε σχέση με τις τιμές της $D(\tilde{\mathbf{x}})$ δείχνεται στο Σχ. 4.4-3.

Στα διαγράμματα η d(x) θα παριστάνεται όπως στο Σχ. 4.4-4



Σχήμα 4.4-4.

Κατόπιν τούτων μπορούμε να παραστήσουμε διαγραμματικά την ταξινόμηση ενός νέου προτύπου σε μία από τις δύο γραμμικά διαχωρίσιμες κλάσεις A, B, όπως δείχνεται στο Σχ.4.4-5



Σχήμα 4.4-5

Ο τύπος διόρθωσης (4.4.5) γράφεται τώρα $\mathbf{W}_{t+1} = \mathbf{W}_t - \rho d(x)\tilde{\mathbf{x}}$ και αιτιολογείται εάν θεωρήσουμε το πρόβλημα εύρεσης του **W** ως πρόβλημα βελτιστοποίησης (*optimization*) μιας συνάρτησης κόστους K(**W**). Μία τέτοια συνάρτηση είναι η

(4.4.16) $\quad \mathrm{K}(\mathbf{W}) : \Re^{N+1} \to \Re$ με $\mathrm{K}(\mathbf{W}) = \sum_{x \in S} (d(\mathbf{x})\mathbf{W}^{\mathrm{T}}\tilde{\mathbf{x}})$

όπου *S* το σύνολο των διανυσμάτων που ταξινομήθηκαν λάθος. Η συνάρτηση K(**W**) εκφράζει ουσιαστικά το συνολικό σφάλμα ταξινόμησης και είναι κατά τμήματα γραμμική συνάρτηση. Η βέλτιστη λύση του προβλήματος είναι η εύρεση τέτοιου **W** ώστε K(**W**)=**0**. Αν $\mathbf{W}=\mathbf{W}_0$ είναι μία αρχική τιμή, η βέλτιστη τιμή του **W** μπορεί να προσεγγισθεί από τον τύπο

(4.4.17) $$\mathbf{W}_{t+1} = \mathbf{W}_t - \rho \frac{\partial K(\mathbf{W})}{\partial \mathbf{W}} = \mathbf{W}_t - \rho \sum_{x \in S} d(\mathbf{x})\tilde{\mathbf{x}}$$

Ο τύπος αυτός αποτελεί έναν επαναληπτικό αλγόριθμο ελαχιστοποίησης της συνάρτησης κόστους.

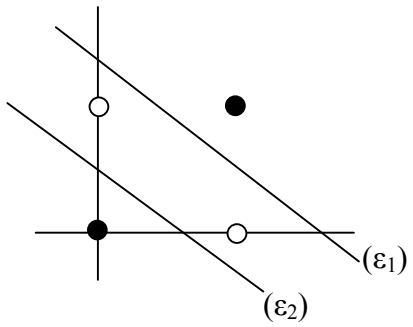
4.5. Γραμμικοί ταξινομητές πολλών επιπέδων

Όταν το πρόβλημα της ταξινόμησης αφορά περισσότερες από δύο κλάσεις ή όταν οι κλάσεις δεν είναι γραμμικά διαχωρίσιμες, είναι δυνατόν να επιτευχθούν λύσεις με κατάλληλους συνδυασμούς γραμμικών ταξινομητών. Μια τέτοια χαρακτηριστική περίπτωση είναι αυτή της λογικής πύλης XOR της οποίας ο πίνακας αληθείας δείχνεται στο Σχ.4.5-1

α	**β**	**α XOR β**
0	0	0
0	1	1
1	0	1
1	1	0

Σχήμα 4.5-1

Σύμφωνα με αυτόν οι συνδυασμοί των τιμών των λογικών μεταβλητών α, β αποτελούν τέσσερα πρότυπα που περιγράφονται από τα διανύσματα στοιχείων συνόλου **Ω**= { (0,0), (0,1), (1,0), (1,1) } και η πράξη α XOR β ορίζει τις κλάσεις A={(0,0),(1,1)} και B={(0,1),(1,0)}. Στο Σχ.4.5-2 φαίνονται τα άκρα των διανυσμάτων στο E^2. Είναι προφανές ότι οι κλάσεις A, B δεν διαχωρίζονται με μια ευθεία. Ο διαχωρισμός των κλάσεων μπορεί να γίνει με δύο ευθείες (Σχ.4.5-2) που ορίζουν μία ζώνη στο εσωτερικό της οποίας βρίσκονται τα πρότυπα της κλάσης B. Η ευθεία (ε_1) μπορεί να προσδιοριστεί από ένα γραμμικό ταξινομητή που θα διαχωρίζει το πρότυπο (1,1) από τα υπόλοιπα. Η ευθεία (ε_2) μπορεί να προσδιορισθεί από ένα γραμμικό ταξινομητή T_2 που θα διαχωρίζει το πρότυπο (0,0) από τα υπόλοιπα.

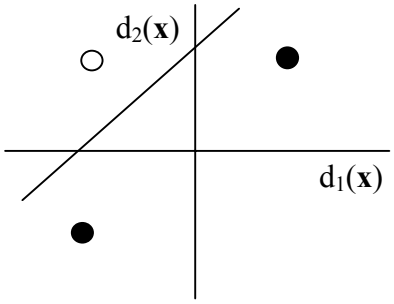


Σχήμα 4.5-2

Οι έξοδοι των T_1, T_2 θα είναι οι τιμές των συναρτήσεων $d_1(\mathbf{x})$, $d_2(\mathbf{x})$ για x $\in \Omega$ θα είναι :

α	β	$d_1(\mathbf{x})$	$d_2(\mathbf{x})$	**ΚΛΑΣΗ**
0	0	-1	-1	A
0	1	-1	+1	B
1	0	-1	+1	B
1	1	+1	+1	A

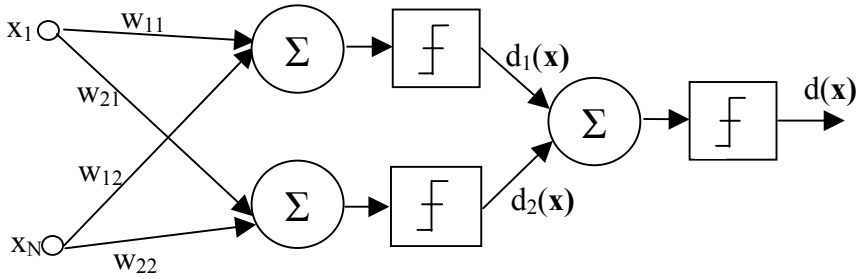
Οι τιμές των $d_1(\mathbf{x})$, $d_2(\mathbf{x})$ προσδιορίζουν τις κλάσεις A και B αποτελώντας ένα νέο σύνολο προτύπου **Φ**={(-1,-1),(-1,+1),(+1,+1)} (η περίπτωση (+1,-1) είναι αδύνατη). Στο Σχ.4.5-3 φαίνεται ο χώρος του **Φ**.



Σχήμα 4.5-3

Η κλάση A προσδιορίζεται από το σημείο (+1,+1) του χώρου **Φ** που διαχωρίζεται γραμμικά με την ευθεία (ε) από τα διανύσματα (-1,-1),(+1,+1) που προσδιορίζουν την κλάση B. Ένας γραμμικός ταξινομητής T μπορεί να

προσδιορίσει την ε. Σ' αυτόν αν $d_T(\mathbf{x})>0 \Rightarrow \mathbf{x}\in A$ και αν $d_T(\mathbf{x})<0 \Rightarrow \mathbf{x}\in B$. Το όλο σύστημα φαίνεται στο Σχ.4.5-4



Σχήμα 4.5-4 *Γραμμικός ταξινομητής-ΝΔ Perceptron δύο επιπέδων.*