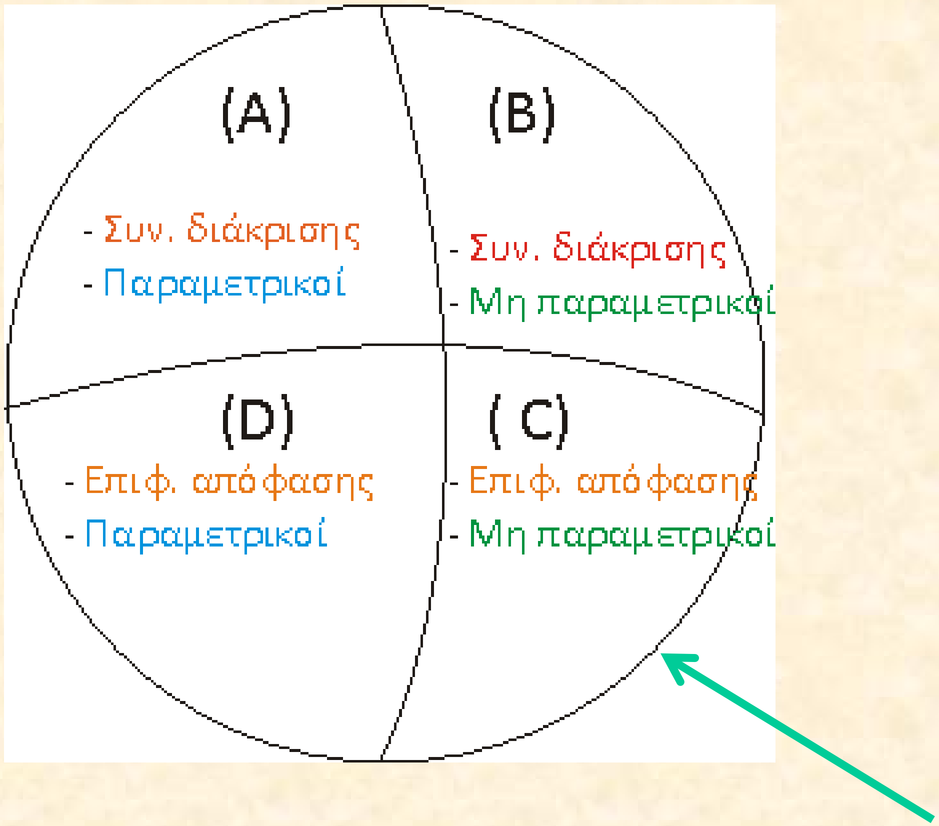
ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ (PATTERN RECOGNITION)

Σέργιος Θεοδωρίδης

Κωνσταντίνος Κουτρούμπας

Version 2


(A)
- Συν. διάκρισης
- Παραμετρικοί
(B)
- Συν. διάκρισης
- Μη παραμετρικοί
(D)
- Επιφ. απόφασης
- Παραμετρικοί
(C)
- Επιφ. απόφασης
- Μη παραμετρικοί

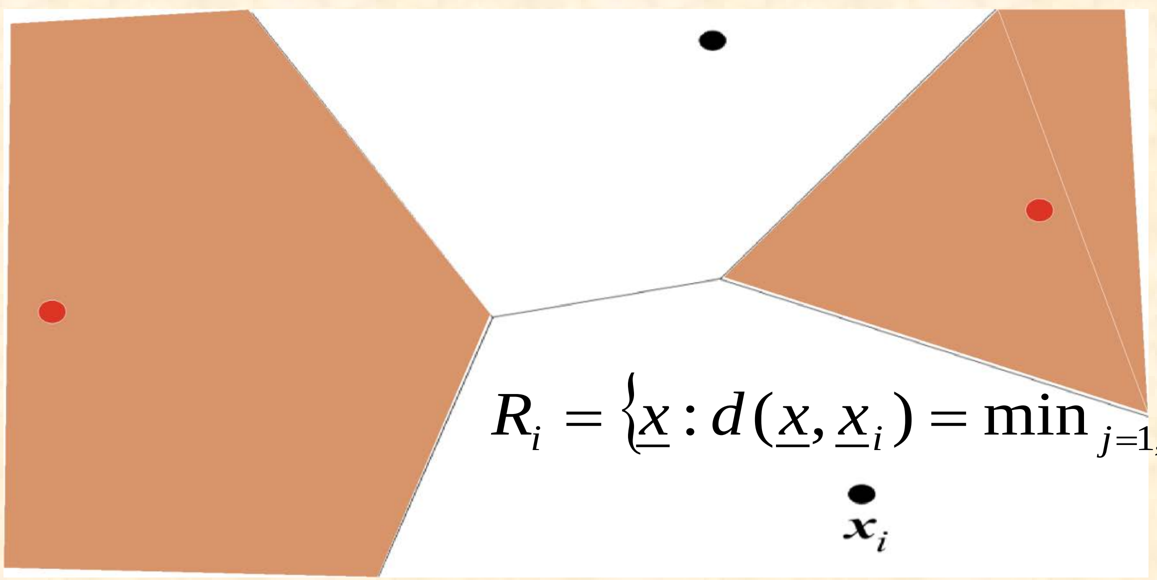
(C) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΕΠΙΦΑΝΕΙΕΣ ΑΠΟΦΑΣΗΣ

Ταξιν. κ-πλησιέστερων γειτόνων

Υπενθύμιση: *X* είναι το σύνολο δεδομένων που περιέχει τα διαθέσιμα δεδομένα από όλες τις κλάσεις.

X_j είναι το υποσύνολο του *X* που περιέχει τα δεδομένα από την κλάση ω_j, $X=X_1\cup\ldots\cup X_M$

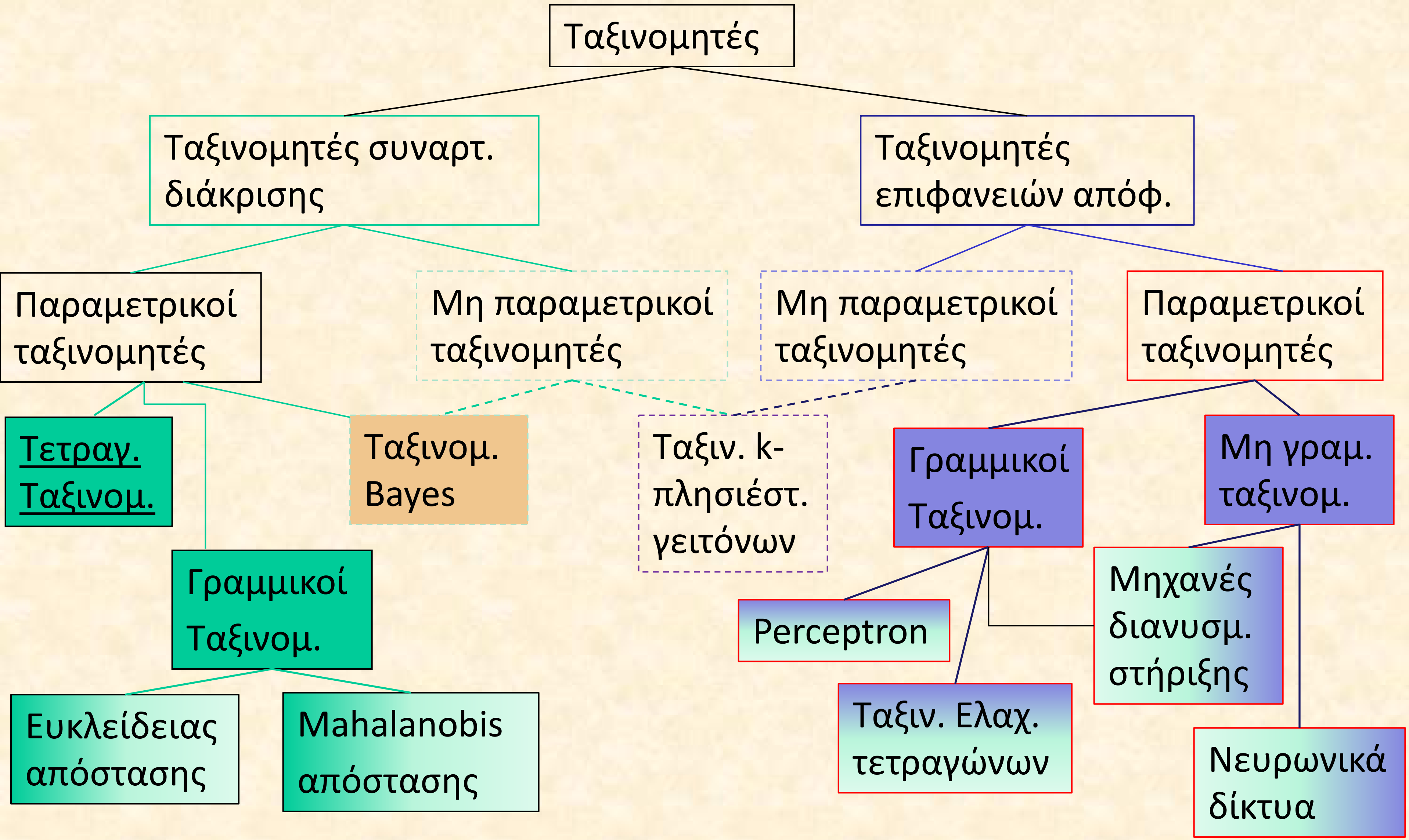
- Στην περίπτωση αυτή οι επιφάνειες απόφασης ορίζονται έμμεσα μέσω των σημείων του *X*.
- Για παράδειγμα ο ταξινομητής πλησιέστερου γείτονα διαχωρίζει το χώρο των χαρακτηριστικών με βάση την ψηφοθέτηση Voronoi (Voronoi tesselation)



$$R_i=\{\underline{x}:d(\underline{x},\underline{x}_i)=\min_{j=1,\ldots,N} d(\underline{x},\underline{x}_j)\}$$

Για Ευκλείδεια απόσταση

“ΧΑΡΤΟΓΡΑΦΗΣΗ” ΤΟΥ ΧΩΡΟΥ ΤΩΝ ΤΑΞΙΝΟΜΗΤΩΝ



"ΧΑΡΤΟΓΡΑΦΗΣΗ" ΤΟΥ ΧΩΡΟΥ ΤΩΝ ΤΑΞΙΝΟΜΗΤΩΝ

Υπενθ.: X είναι το σύνολο των δεδομένων σημείων όλων των κλάσεων

X_j είναι το υποσύνολο του X που περιέχει τα διανύσματα της κλάσης ω_j,

$X = X_1 \cup \ldots \cup X_M$

Παραμετρικοί ταξινομητές με βάση τις συναρτήσεις διάκρισης

Ταξινομ. Bayes

- $g_j(x) = f\left(P(\omega_j)\, p(x|\omega_j)\right)$
- Υπόθ.: $p(x|\omega_j) \approx \hat{p}(x|\omega_j;\, \theta_j)$
- Εκτίμ. θ_j με βάση το X_j

  (ML, EM): $X_j \rightarrow \hat{\theta}_j$
- $x \rightarrow g_j(x) = f\left(P(\omega_j)\hat{p}(x|\omega_j;\, \theta_j)\right)$

Τετραγωνικός ταξινομητής

- $g_j(x) = -(x-\mu_i)^T \Sigma_j^{-1} (x-\mu_i)$
- Υπόθ.: $N(\mu_j, \Sigma_j)$
- Εκτίμ. μ_j, Σ_j με βάση το X_j

  (ML): $X_j \rightarrow \hat{\mu}_j, \hat{\Sigma}_j$
- $x \rightarrow g_j(x) = (x-\hat{\mu}_j)^T \hat{\Sigma}_j (x-\hat{\mu}_j)$

Γραμμικός Ευκλείδειος ταξινομητής

- $g_j(x) = -(x-\mu_i)^T (x-\mu_i)$
- Υπόθ.: $N(\mu_j, \sigma^2 I)$
- Εκτίμ. μ_j με βάση το X_j

  (ML): $X_j \rightarrow \hat{\mu}_j$
- $x \rightarrow g_j(x) = (x-\mu_i)^T (x-\mu_i)$

Γραμμικός Mahalanobis ταξινομητής

- $g_j(x) = -(x-\mu_i)^T \Sigma_j^{-1} (x-\mu_i)$
- Υπόθ.: $N(\mu_j, \Sigma)$
- Εκτίμ. μ_j, Σ_j με βάση το X_j

  (ML): $X_j \rightarrow \hat{\mu}_j, \hat{\Sigma}_j$
- Εκτίμ. Σ με βάση τα $\hat{\Sigma}_j$

  (π.χ. $\hat{\Sigma} = \frac{1}{M} \sum_{j=1}^{M} \hat{\Sigma}_j$)
- $x \rightarrow g_j(x) = (x-\hat{\mu}_j)^T \hat{\Sigma} (x-\hat{\mu}_j)$

“ΧΑΡΤΟΓΡΑΦΗΣΗ” ΤΟΥ ΧΩΡΟΥ ΤΩΝ ΤΑΞΙΝΟΜΗΤΩΝ

Υπενθ.: X είναι το σύνολο των δεδομένων σημείων όλων των κλάσεων

X_j είναι το υποσύνολο του X που περιέχει τα διανύσματα της κλάσης ω_j,

$X = X_1 \cup \ldots \cup X_M$

Μη παραμετρικοί ταξινομητές με βάση τις συναρτήσεις διάκρισης

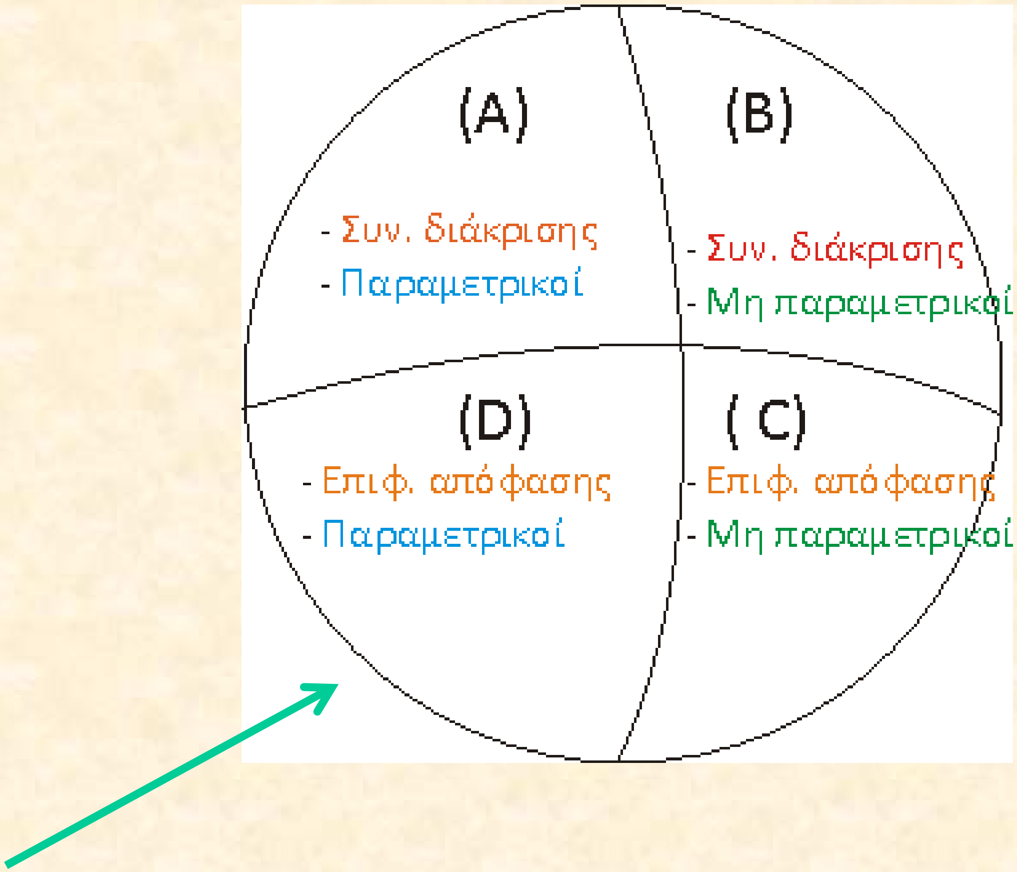
Ταξινομητής Bayes

- $g_j(x) = f(P(\omega_j)\, p(x|\omega_j))$
- $x_i \rightarrow$ *Εκτίμ.* $p(x|\omega_j) \approx \acute{p}(x|\omega_j;\, X_j)$

  (*Παρ. Parzen, k-NN εκτ. πυκν. πιθ.*)
- $x \rightarrow g_j(x) = f(P(\omega_j)\acute{p}(x|\omega_j;\, X_j))$

Ταξινομητής *k*-πλησιέστερων γειτόνων

$- x \rightarrow g_j(x_i) = k^j$


(A)
- Συν. διάκρισης
- Παραμετρικοί
(B)
- Συν. διάκρισης
- Μη παραμετρικοί
(D)
- Επιφ. απόφασης
- Παραμετρικοί
(C)
- Επιφ. απόφασης
- Μη παραμετρικοί

ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μερικά προκαταρτικά:

- **Στην πράξη** έχουμε στη διάθεσή μας ένα σύνολο δεδομένων (σύν. εκπαίδευσης)

$$X = \{(x_i, d_i), x_i \in R^l, d_i \in \{1,2,\ldots,M\}, i = 1,\ldots N\}$$

όπου

***x**$_i$* είναι η *l*-διάστατη αναπαράσταση του *i*-στού από τις *N* οντότητες (διάνυσμα εκπαίδευσης)

d$_i$ είναι η ετικέτα της κλάσης όπου ανήκει το ***x**$_i$* (1 για την ω_1, 2 για την ω_2,…).

- Εστιάζουμε κυρίως στην περίπτωση των δύο κλάσεων.
- Δεν υιοθετούμε καμία υπόθεση σχετικά με τις pdfs που μοντελοποιούν τις διάφορες κλάσεις.
- Αναζητούμε την επιφάνεια (γραμμική ή μη γραμμική) που επιτυγχάνει τον “βέλτιστο” διαχωρισμό των (διανυσμάτων δεδομένων των) κλάσεων.

ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Επιφάνεια απόφασης (*C*): Περιγράφεται από εξίσωση της μορφής $h(\mathbf{x})=0$, ή $h(\mathbf{x};\mathbf{w})=0$, όπου ***w*** είναι το διάνυσμα των παραμέτρων που ορίζουν την επιφάνεια (*C*).

Παραδείγματα:

1. Αν η (*C*) είναι **καμπύλη 1ου βαθμού** (υπερεπίπεδο – γραμμικός διαχωρισμός), τότε

$$h(x;w) = w_1 x_1 + \ldots + w_l x_l + w_0 = \sum_{k=1}^{l} w_k x_k + w_0 = w^T x + w_0 = 0$$

όπου $\mathbf{w}=[w_1,\ldots,w_l]^T$, $\mathbf{x}=[x_1,\ldots,x_l]^T$. Οι παράμετροι είναι το διάνυσμα ***w*** και το w_0.

2. Αν η (*C*) είναι **καμπύλη 2ου βαθμού** (π.χ. Υπερέλλειψη – μη γραμμικός διαχωρισμός), τότε

$$h(x;w) = \sum_{k=1}^{l}\sum_{q=k}^{l} w_{kq} x_k x_q + \sum_{k=1}^{l} w_k x_k + w_0 = 0$$

όπου $\mathbf{w}=[w_0, w_1,\ldots,w_l, w_{11},\ldots,w_{1l}, w_{22},\ldots,w_{2l},\ldots,w_{ll}]^T$

Σημείωση: Σε αρκετούς μη γραμμικούς ταξινομητές (π.χ. στα νευρωνικά δίκτυα) η μορφή της (*C*) δεν μπορεί να εκφραστεί άμεσα (explicitly).

ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Υπόθεση: Αν δεν ορίζεται διαφορετικά, θεωρούμε την περίπτωση των δύο κλάσεων, δηλ., ω_1 (+1) και ω_2 (-1).

Ορισμός προβλήματος: Δοθέντος ενός συνόλου δεδομένων *X* προσδιόρισε μία επιφάνεια που επιτυγχάνει το “βέλτιστο” δυνατό διαχωρισμό των (διανυσμάτων των) δύο κλάσεων.

<u>Στρατηγική αντιμετώπισης του προβλήματος:</u>

1. **Υιοθέτησε** μια συγκεκριμένη (“άμεση” ή “έμμεση”) παραμετρική μορφή για την επιφάνεια $h(\boldsymbol{x};\boldsymbol{w})=0$.
2. **Όρισε** κατάλληλη συνάρτηση (συνάρτηση κόστους - cost function) του $\boldsymbol{w}$, $J(\boldsymbol{w})$, η οποία περιλαμβάνει επίσης τα διανύσματα του *X*, έτσι ώστε **τα βέλτιστά της (ελάχιστα ή μέγιστα) να αντιστοιχούν στις καλύτερες δυνατές επιφάνειες για το υπό μελέτη πρόβλημα**.
3. **Βελτιστοποίησε** την $J(\boldsymbol{w})$ ως προς το $\boldsymbol{w}$. **<u>Η θέση του βέλτιστου αυτής ορίζει την επιφάνεια απόφασης</u>**.

<u>Σημαντική παρατήρηση:</u> Το νόημα της φράσης “βέλτιστη δυνατή καμπύλη” διαφέρει για διαφορετικές επιλογές της $J(\boldsymbol{w})$.

ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

Έστω *J*(***w***) συνεχής συνάρτηση του ***w***.

Πρόβλημα (P1): Προσδιόρισε τη θέση ***w**** όπου η συνάρτηση *J*(***w***) λαμβάνει την ελάχιστη τιμή της.

Μια απλή μέθοδος για την επίλυση του **(P1)** είναι αυτή της οξύτερης καθόδου (gradient descent - GD).

-**Αρχικοποίησε *w***=***w***(0)

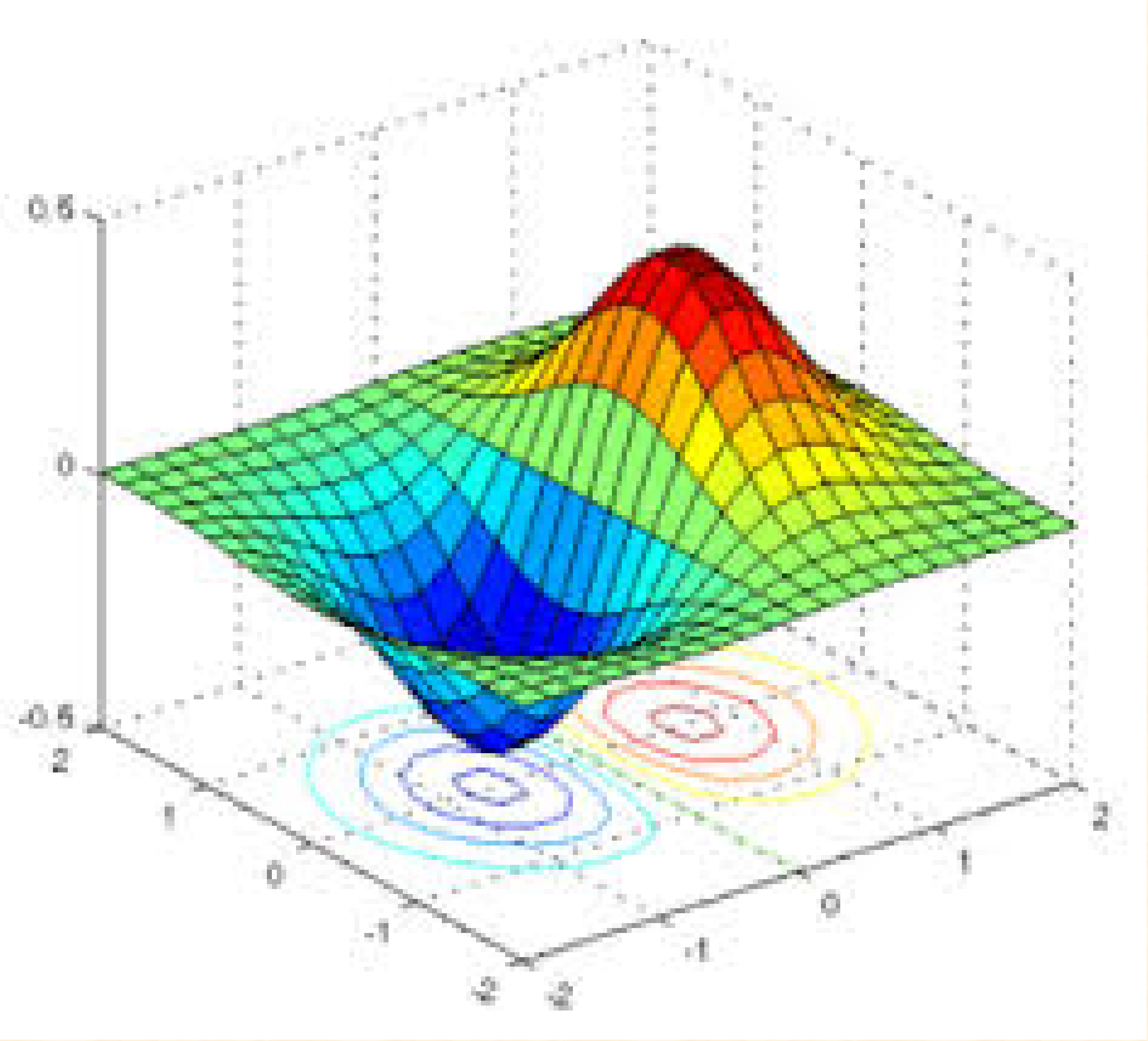
-t=0

-**Επανέλαβε**

- $w(t+1) = w(t) - \mu \frac{\partial J(w)}{\partial w}\Big|_{w=w(t)}$
- t=t+1

-**Έως ότου** επιτευχθεί σύγκλιση

ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ



ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

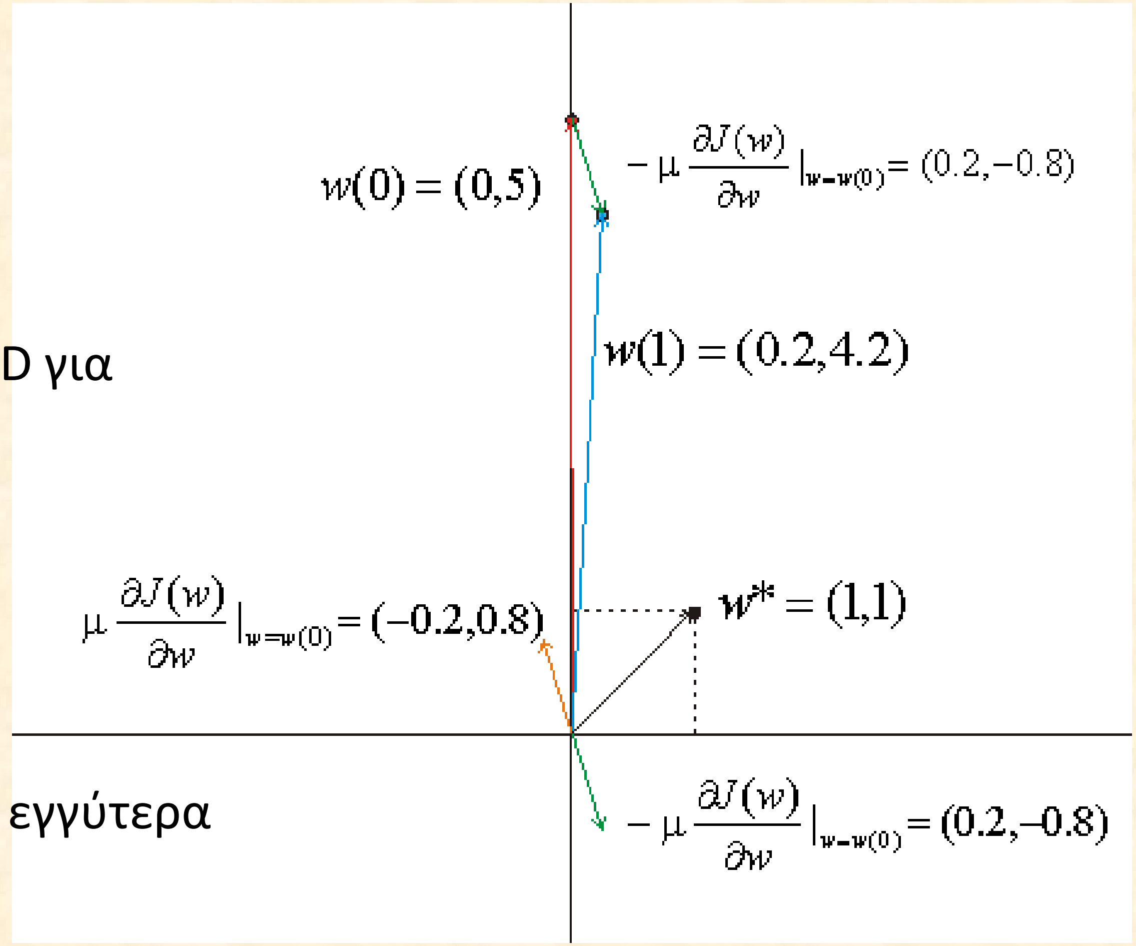
-**Ένα παράδειγμα:** Έστω $\boldsymbol{w}=[w_1, w_2]^T$ και $J(\boldsymbol{w})=(w_1-1)^2+(w_2-1)^2$. Είναι σαφές ότι η $J(\boldsymbol{w})$ λαμβάνει την ελάχιστη τιμή της στο $\boldsymbol{w}^*=[1, 1]^T$.

-Είναι $\dfrac{\partial J(w)}{\partial w} = \begin{bmatrix} 2w_1 - 2 \\ 2w_2 - 2 \end{bmatrix}$

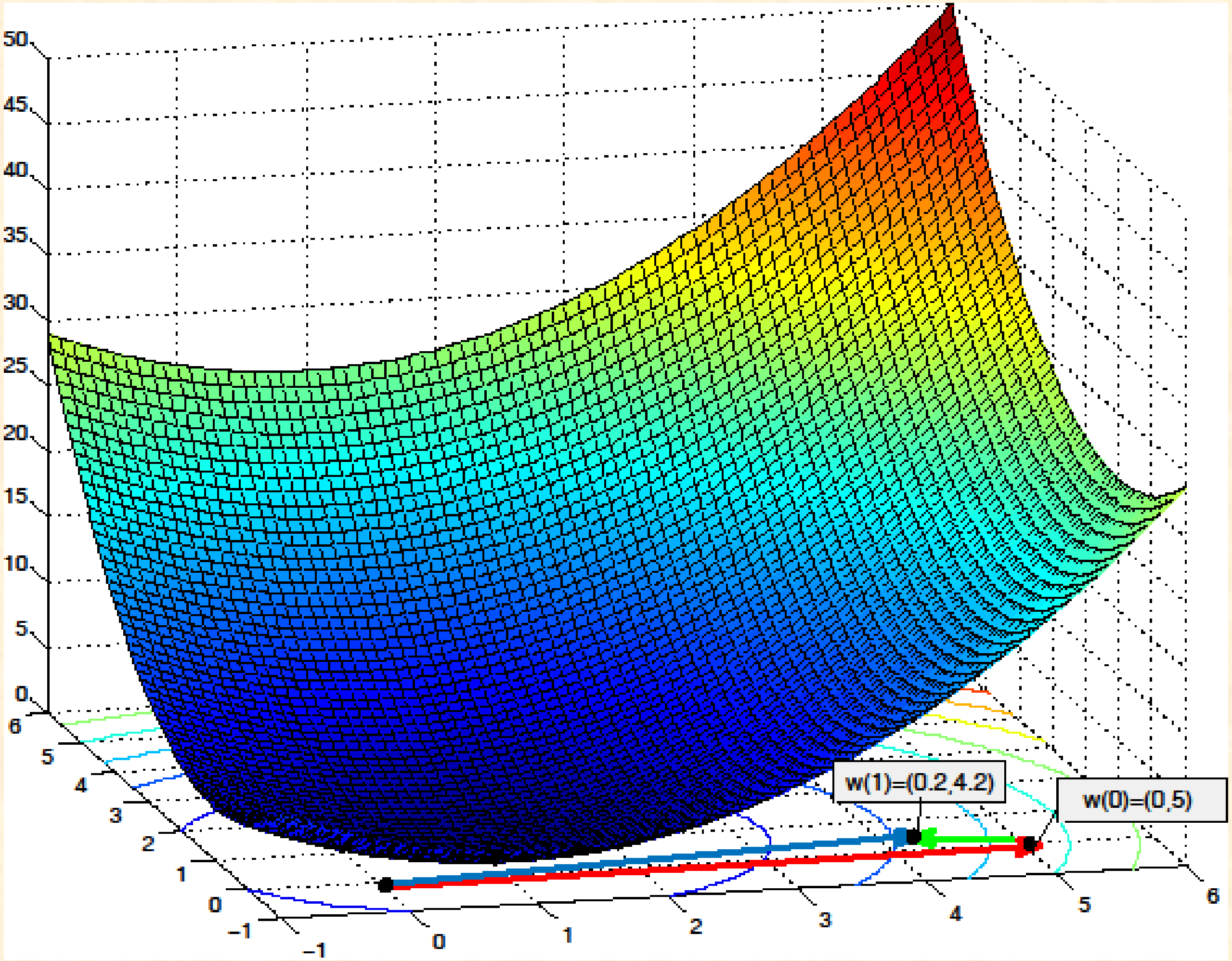
-Εφαρμόζοντας τον αλγόριθμο GD για $\boldsymbol{w}(0)=[0, 5]^T$, και $\mu=0.1$, έχουμε

$$w(1) = \begin{bmatrix} 0 \\ 5 \end{bmatrix} - 0.1 \begin{bmatrix} -2 \\ 8 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 4.2 \end{bmatrix}$$

-Βλέπουμε ότι το $\boldsymbol{w}(1)$ βρίσκεται εγγύτερα στο $\boldsymbol{w}^*$.



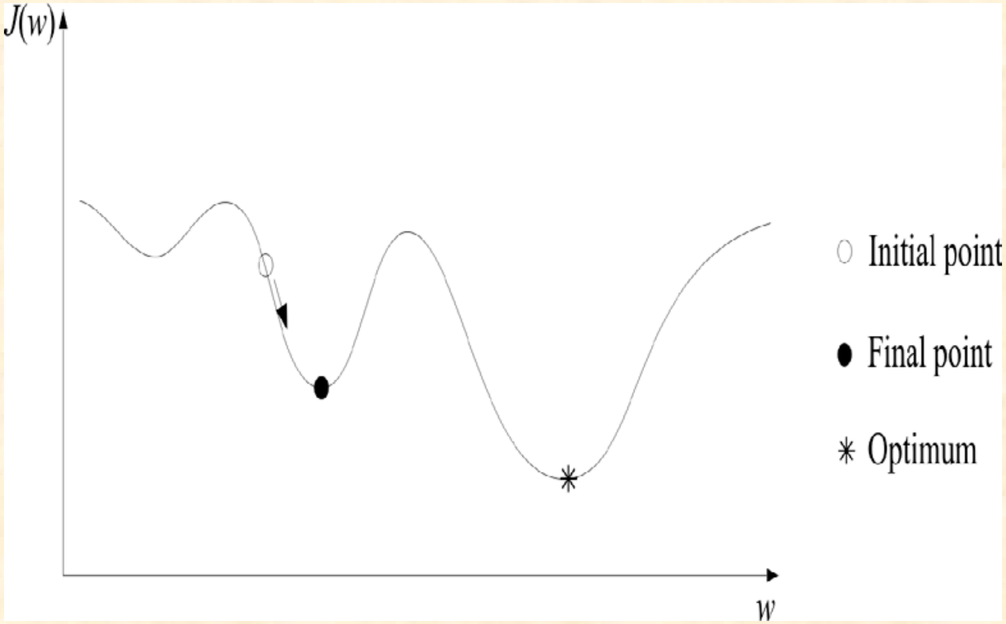
ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ



ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

Σχόλια για τον αλγόριθμο GD:

- Η τιμή του μ πρέπει να είναι όχι πολύ μεγάλη, ώστε να αποφεύγονται ταλαντώσεις γύρω από το ελάχιστο της συνάρτησης και όχι πολύ μικρή, ώστε να αποφεύγονται άσκοπες καθυστερήσεις στη σύγκλιση.
- Αν η $J(\boldsymbol{w})$ έχει περισσότερα από ένα ελάχιστα, ο αλγόριθμος GD θα συγκλίνει (γενικά) σε εκείνο που βρίσκεται εγγύτερα στο $\boldsymbol{w}(0)$.
- Αν ο αλγόριθμος παγιδευτεί σε ένα τοπικό ελάχιστο *που αντιστοιχεί σε μια "κακή" λύση*, ο μόνος τρόπος για να το αποφύγουμε είναι να επανεκκινήσουμε τον αλγόριθμο από μια διαφορετική αρχική θέση.
- Μπορεί να αποδειχθεί ότι, κάτω από κάποιες συνθήκες, ο αλγόριθμος συγκλίνει ασυμπτωτικά σε ένα τοπικό ελάχιστο της $J(\boldsymbol{w})$.



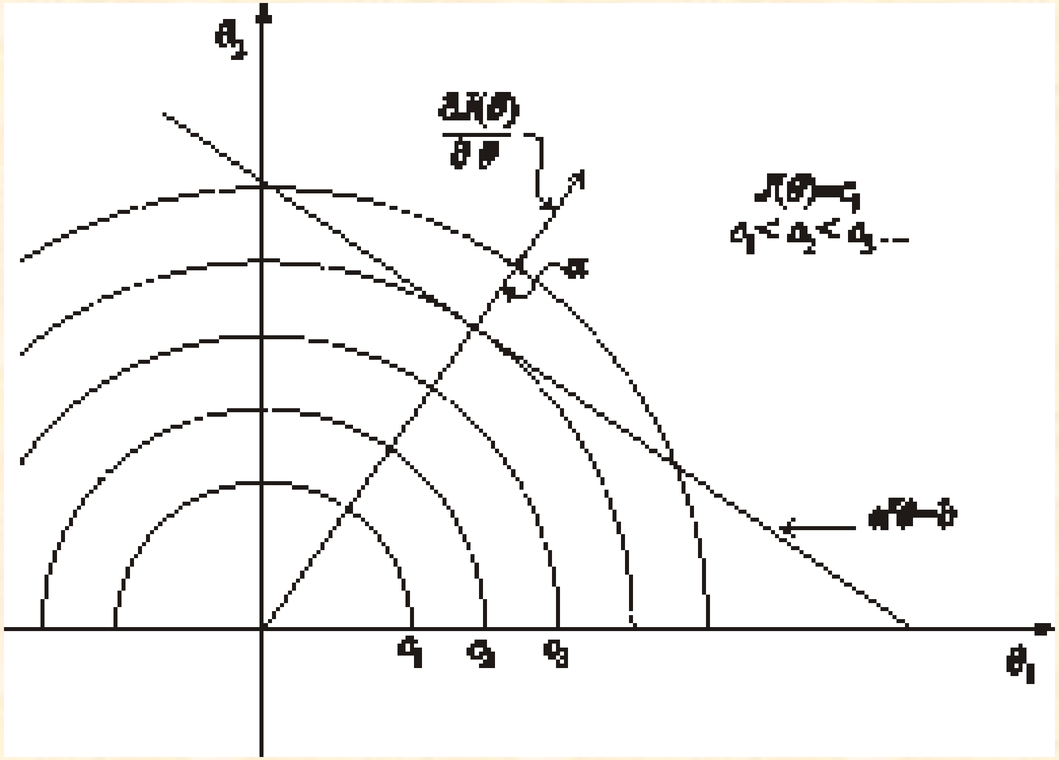
ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

Έστω $J(\boldsymbol{w})$ μια συνεχής συνάρτηση του $\boldsymbol{w}$.

Πρόβλημα (P2): Προσδιόρισε τη θέση $\boldsymbol{w}^*$ όπου η συνάρτηση $J(\boldsymbol{w})$ λαμβάνει την ελάχιστη τιμή της, υπό την προϋπόθεση ότι το $\boldsymbol{w}$ ικανοποιεί μερικούς περιορισμούς ισότητας (equality constraints).

Για γραμμικούς περιορισμούς ισότητας, το πρόβλημα διατυπώνεται ως εξής

- Ελαχιστοποίησε την $J(\boldsymbol{w})$
- Υπό τους περιορισμούς $A\boldsymbol{w}=\boldsymbol{b}$, όπου A ένας mxl πίνακας και $\boldsymbol{b}$ ένα m-διάστατο διάνυσμα.



Λύση: Πολ/στές Lagrange

Ελαχιστοποίησε την

- $L(\boldsymbol{w})=J(\boldsymbol{w})+\boldsymbol{\lambda}^T(A\boldsymbol{w}-\boldsymbol{b})$
- $\boldsymbol{\lambda}$ είναι ένα m-διάστατο διάνυσμα που εκτιμάται μέσω των περιορισμών $A\boldsymbol{w}=\boldsymbol{b}$

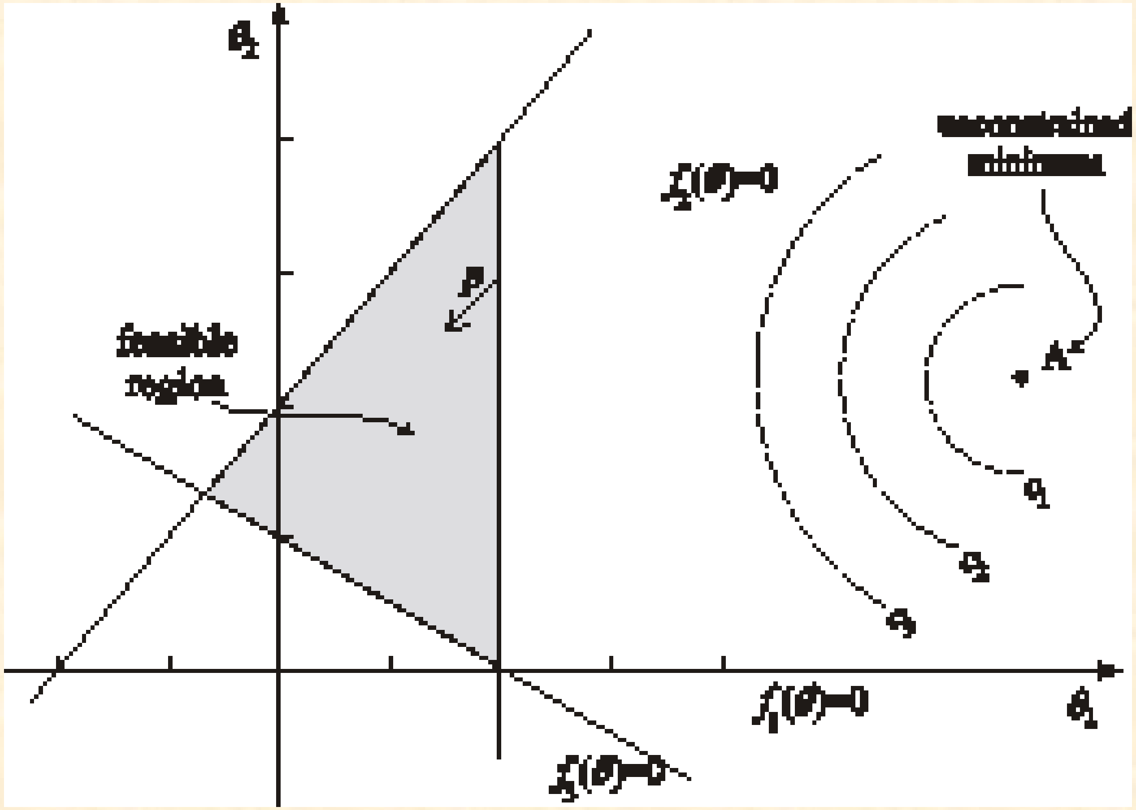
ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

Έστω *J*(***w***) μια συνεχής συνάρτηση του ***w***.

Πρόβλημα (P3): Προσδιόρισε τη θέση ***w**** όπου η συνάρτηση *J*(***w***) λαμβάνει την ελάχιστη τιμή της, υπό τις την προϋπόθεση ότι το ***w*** ικανοποιεί μερικούς περιορισμούς ανισότητας.

Για γραμμικούς περιορισμούς ανισότητας, το πρόβλημα διατυπώνεται ως εξής

- Ελαχιστοποίησε την *J*(***w***)
- Υπό τους περιορισμούς A***w***≥***b***, όπου A ένας *m*x*l* πίνακας και ***b*** ένα *m*-διάστατο διάνυσμα.



ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

-Στην περίπτωση αυτή η επιφάνεια είναι ένα υπερεπίπεδο (*H*) της μορφής

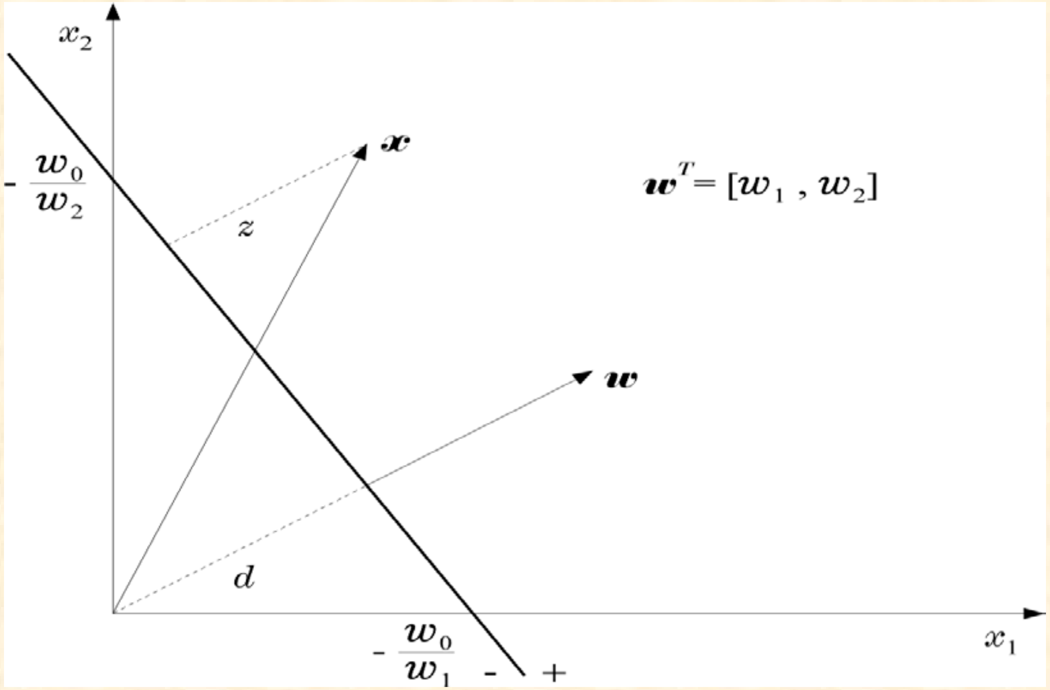
$$(H): h(x;w) = w_1 x_1 + \dots + w_l x_l + w_0 = \sum_{k=1}^{l} w_k x_k + w_0 = w^T x + w_0 = 0$$

όπου $\boldsymbol{w}=[w_1,\dots,w_l]^T$, $\boldsymbol{x}=[x_1,\dots,x_l]^T$

- Το (*H*) **ορίζεται πλήρως** από τα $\boldsymbol{w}$ και w_0.

- Μερικά στοιχεία από τη Γεωμετρία:
    - Το διάνυσμα $\boldsymbol{w}$ είναι κάθετο στο (*H*).
    - Το $\boldsymbol{w}$ ορίζει τον προσανατολισμό (orientation) του (*H*).
    - Όλα τα υπερεπίπεδα που είναι παράλληλα με το (*H*) έχουν τον ίδιο προσανατολισμό με αυτό.
    - Το $\boldsymbol{w}$ είναι κάθετο σε όλα τα υπερεπίπεδα που είναι παράλληλα με το (*H*).
    - Το w_0 είναι η παράμετρος που ταυτοποιεί μοναδικά ένα συγκεκριμένο υπερεπίπεδο, ανάμεσα στα υπερεπίπεδα ίδιου προσανατολισμού.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ



$$d = \frac{|w_0|}{\sqrt{w_1^2 + w_2^2}}, \quad z = \frac{|g(\underline{x})|}{\sqrt{w_1^2 + w_2^2}} \qquad g(x) = w_1 x_1 + w_2 x_2 + w_0$$

Γενικά

$$g(x) = w_1 x_1 + \cdots + w_l x_l + w_0$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

ΜΙΑ ΣΥΜΒΑΣΗ: Προκειμένου να καταστήσουμε πιο συμπαγή το συμβολισμό

$$(H): w^T x + w_0 = 0$$

ορίζουμε

$\boldsymbol{w}^* = [\boldsymbol{w}^T, w_0]^T$ and $\boldsymbol{x}^* = [\boldsymbol{x}^T\ 1]^T$.

’Ετσι έχουμε

$$(H): w^{*T} x^* = 0$$

ΣΗΜΕΙΩΣΗ: Στη συνέχεια, εκτός αν ορίζεται διαφορετικά, θα γράφουμε ***w*** και ***x***, εννοώντας τα ***w****** και ***x******, αντίστοιχα.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

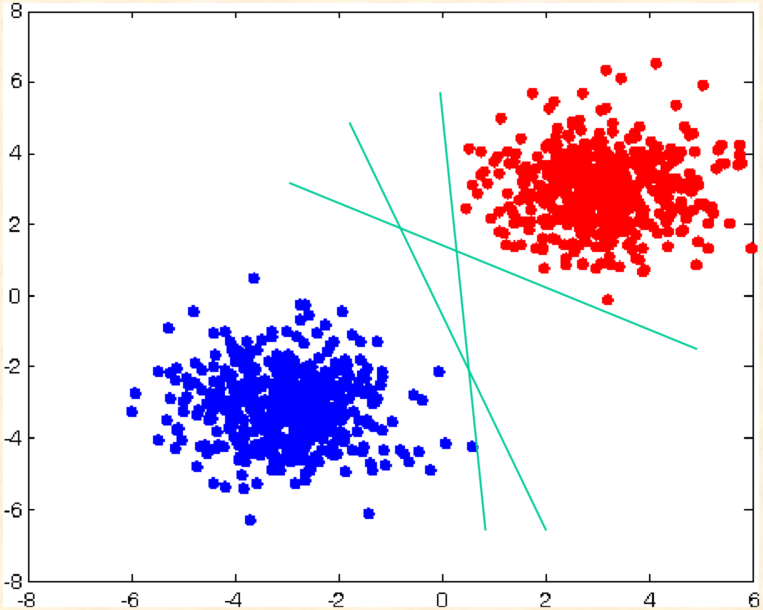
Ο αλγόριθμος perceptron

Υπόθεση: (Τα διανύσματα των) δύο κλάσεων ω_1 και ω_2 είναι γραμμικώς διαχωρίσιμα.

Έτσι, υπάρχει ένα υπερεπίπεδο (H^*): $\boldsymbol{w}^{*T}\boldsymbol{x}=0$, ώστε

$$w^{*T}\, x > 0,\ \ \forall x \in \omega_1$$
$$w^{*T}\, x < 0,\ \ \forall x \in \omega_2$$

Το κριτήριο στο perceptron: Προσδιόρισε ένα υπερεπίπεδο το οποίο διαχωρίζει τέλεια τα διανύσματα από τις δύο κλάσεις.



ΣΗΜΕΙΩΣΗ: Αν οι κλάσεις είναι γραμμικώς διαχωρίσιμες, υπάρχουν άπειρα υπερεπίπεδα που ικανοποιούν το κριτήριο perceptron.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

Η συνάρτηση κόστους:

$$J(\underline{w}) = \sum_{\underline{x} \in Y(w)} (\delta_x w^T x)$$

όπου

- Y(***w***) το σύνολο των λανθασμένα ταξινομημένων διανυσμάτων από το ***w*** και

- $$\begin{aligned} \delta_x &= -1 \text{ if } x \in \omega_1 \\ \delta_x &= +1 \text{ if } x \in \omega_2 \end{aligned}$$

Σημειώσεις:

- Είναι $J(\mathbf{w}) \geq 0$.

Πράγματι, αν $\mathbf{x} \in Y(\mathbf{w})$ και $\mathbf{x} \in \omega_1$ τότε $\mathbf{w}^T\mathbf{x}<0$ και $\delta_{\mathbf{x}}=-1$. Επομένως $\delta_{\mathbf{x}}\, \mathbf{w}^T\mathbf{x}>0$.

Επίσης, αν $\mathbf{x} \in Y(\mathbf{w})$ και $\mathbf{x} \in \omega_2$ τότε $\mathbf{w}^T\mathbf{x}>0$ και $\delta_{\mathbf{x}}=+1$. Επομένως $\delta_{\mathbf{x}}\, \mathbf{w}^T\mathbf{x}>0$.

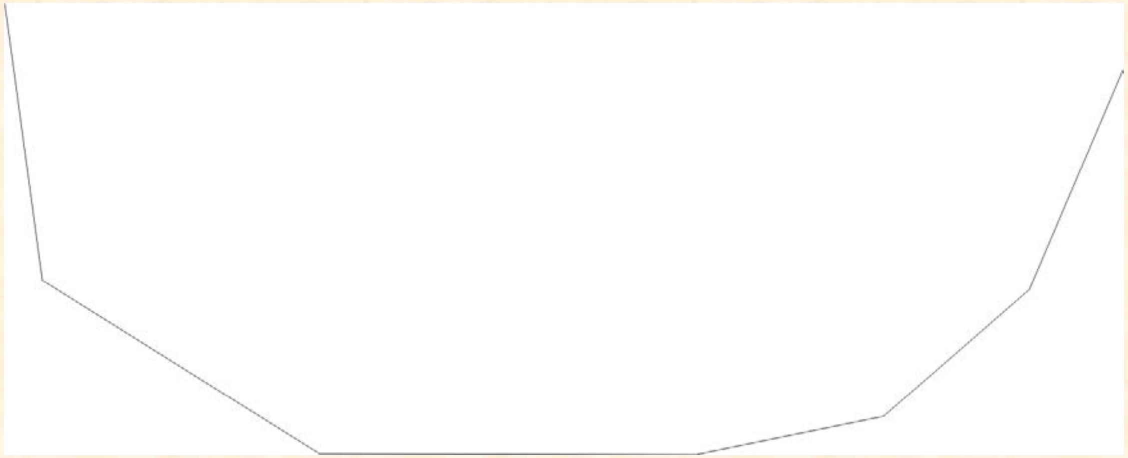
- Είναι $J(\mathbf{w})=0$ μόνον όταν $Y=\emptyset$, πράγμα που σημαίνει ότι προσδιορίστηκε **μία λύση που ικανοποιεί το κριτήριο perceptron**.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

Η συνάρτηση κόστους:

Η J(***w***) είναι τμηματικά γραμμική (piecewise linear) (γιατί;)



Η ελαχιστοποίηση της *J*(***w***) επιτυγχάνεται μέσω διαδικασίας που ακολουθεί τη φιλοσοφία της GD.

Όπου είναι έγκυρο, ορίζουμε $$\frac{\partial J(w)}{\partial w} = \frac{\partial}{\partial w}\Big(\sum_{x \in Y(w)} \delta_x w^T x\Big) = \sum_{x \in Y(w)} \delta_x x$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

-**Αρχικοποίησε *w*=*w***(0)

-t=0

-**Επανέλαβε**

-Προσδιόρισε τα λανθασμένα ταξινομημένα διανύσματα του *X* από το ***w***(*t*) και συσσώρευσέ τα στο $Y(\boldsymbol{w}_t)$.

$$w(t+1) = w(t) - \rho_t \sum_{x \in Y(w_t)} \delta_x x$$

- t=t+1

-**Έως ότου** επιτευχθεί σύγκλιση

Σημειώσεις:

•Πρόκειται για αλγόριθμο επεξεργασίας κατά δέσμες (batch algorithm), όπου η ενημέρωση των παραμέτρων σε κάθε επανάληψη πραγματοποιείται μετά την επεξεργασία όλων των διανυσμάτων του *X*.

•Ο αλγόριθμος perceptron συγκλίνει μόνον όταν οι κλάσεις είναι γραμμικώς διαχωρίσιμες. Διαφορετικά, αποτυγχάνει να συγκλίνει.

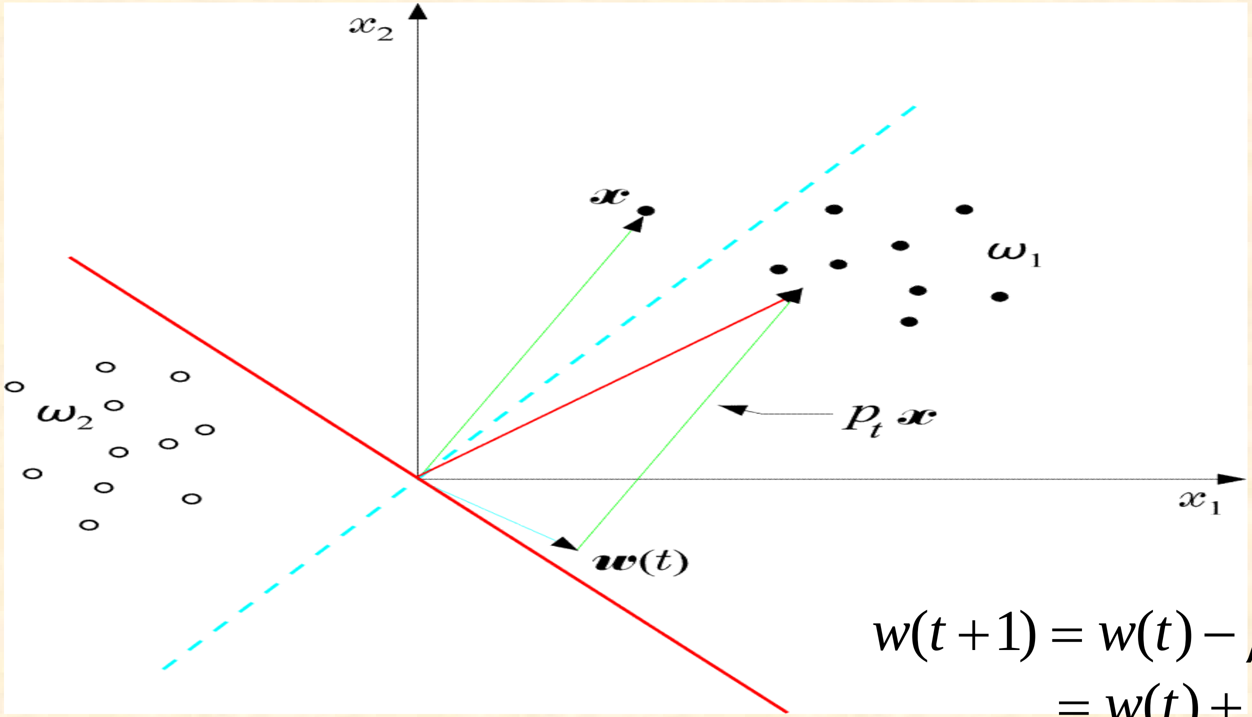
•Αν οι κλάσεις είναι γραμμικώς διαχωρίσιμες, ο αλγόριθμος συγκλίνει σε πεπερασμένο αριθμό βημάτων αν

$$\lim_{t\to\infty} \sum_{k=0}^{t} \rho_k \to \infty, \quad \lim_{t\to\infty} \sum_{k=0}^{t} {\rho_k}^2 < +\infty \quad \pi.\chi. \;\; \rho_t = c/t$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

Παράδειγμα 1:



$$\begin{aligned} w(t+1) &= w(t) - \rho_t \delta_x x \\ &= w(t) + \rho_t x \quad (\delta_x = -1) \end{aligned}$$

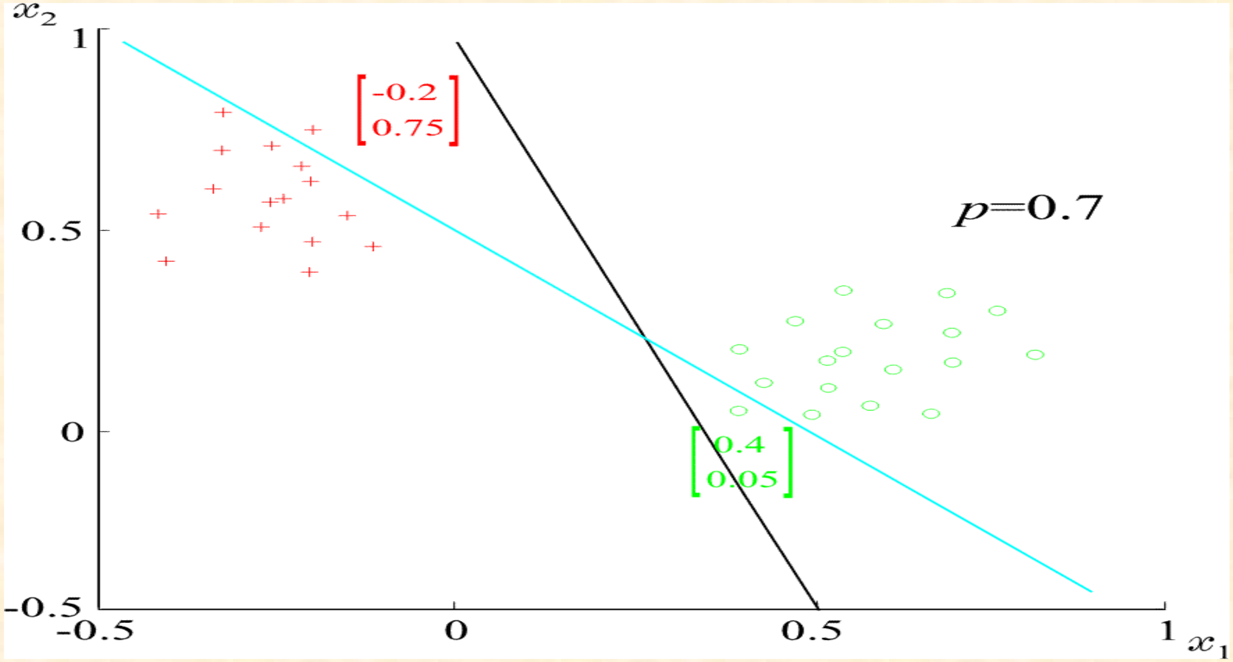
ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

Παράδειγμα 2: Σε κάποια επανάληψη *t* ο αλγόριθμος perceptron δίνει

$w_1 = 1,\ w_2 = 1,\ w_0 = -0.5$ που αντιστοιχεί στην ευθεία $x_1 + x_2 - 0.5 = 0$

Το υπερεπίπεδο της επόμενης επανάληψης είναι



$$\underline{w}(t+1) = \begin{bmatrix} 1 \\ 1 \\ -0.5 \end{bmatrix} - 0.7(-1)\begin{bmatrix} 0.4 \\ 0.05 \\ 1 \end{bmatrix} - 0.7(+1)\begin{bmatrix} -0.2 \\ 0.75 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.42 \\ 0.51 \\ -0.5 \end{bmatrix}$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Παραλλαγές του αλγόριθμου perceptron:

-Ο αλγόριθμος “τσέπης” (pocket algorithm).

- Κρατά την καλύτερη λύση που βρέθηκε κατά την εκτέλεση του αλγόριθμου perceptron για συγκεκριμένο αριθμό επαναλήψεων.
- Κατάλληλος για προβλήματα μη γραμμικώς διαχωρίσιμων κλάσεων.

-Ο on-line αλγόριθμος perceptron:

- Η ενημέρωση των παραμέτρων λαμβάνει χώρα μετά την μεμονωμένη επεξεργασία κάθε διανύσματος του *X*.

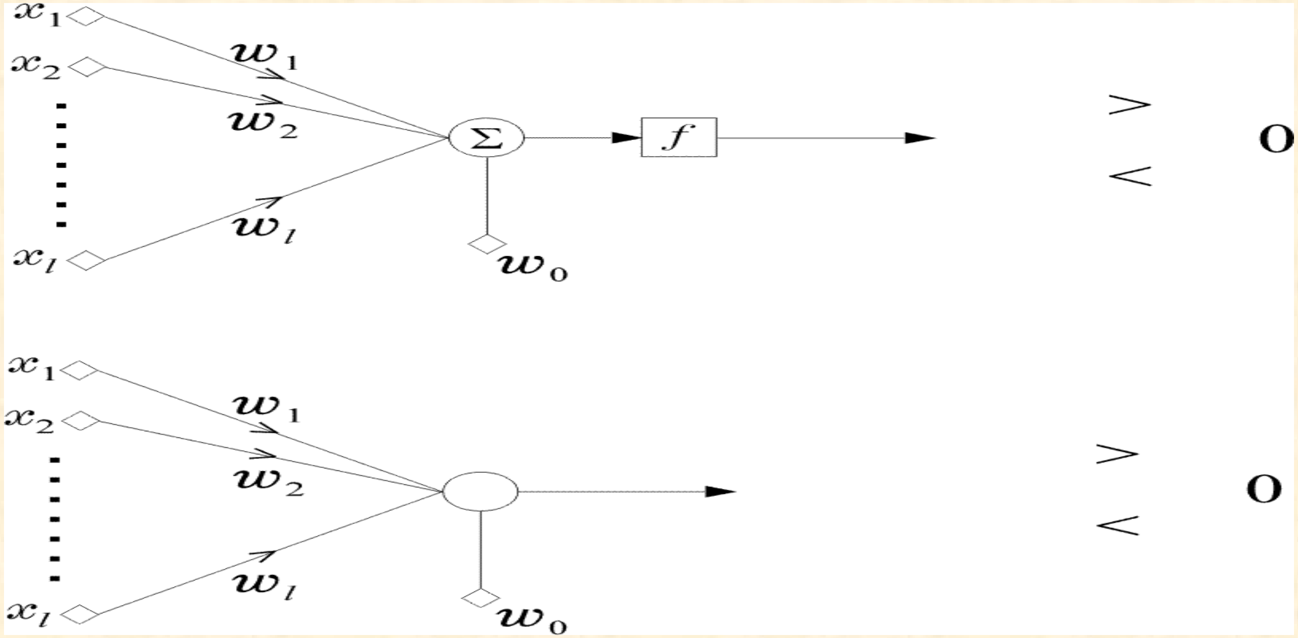
$$w(t+1) = w(t) + \rho x_{(t)} \,, \quad \begin{matrix} w^T(t)x_{(t)} \leq 0 \\ x_{(t)} \in \omega_1 \end{matrix}$$

$$w(t+1) = w(t) - \rho x_{(t)} , \quad \begin{matrix} w^T(t)x_{(t)} \geq 0 \\ x_{(t)} \in \omega_2 \end{matrix}$$

$$w(t+1) = w(t) \quad \text{otherwise}$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Η δομή perceptron



w_i's synapses or synaptic weights
w_0 threshold

- Η δομή αυτή καλείται perceptron ή νευρώνας (neuron)
- Είναι μια μηχανή που μπορεί να "μάθει" (δηλ. να προσδιορίσει τις τιμές των εμπλεκόμενων παραμέτρων) από τα διανύσματα εκπαίδευσης μέσω του αλγόριθμου perceptron.