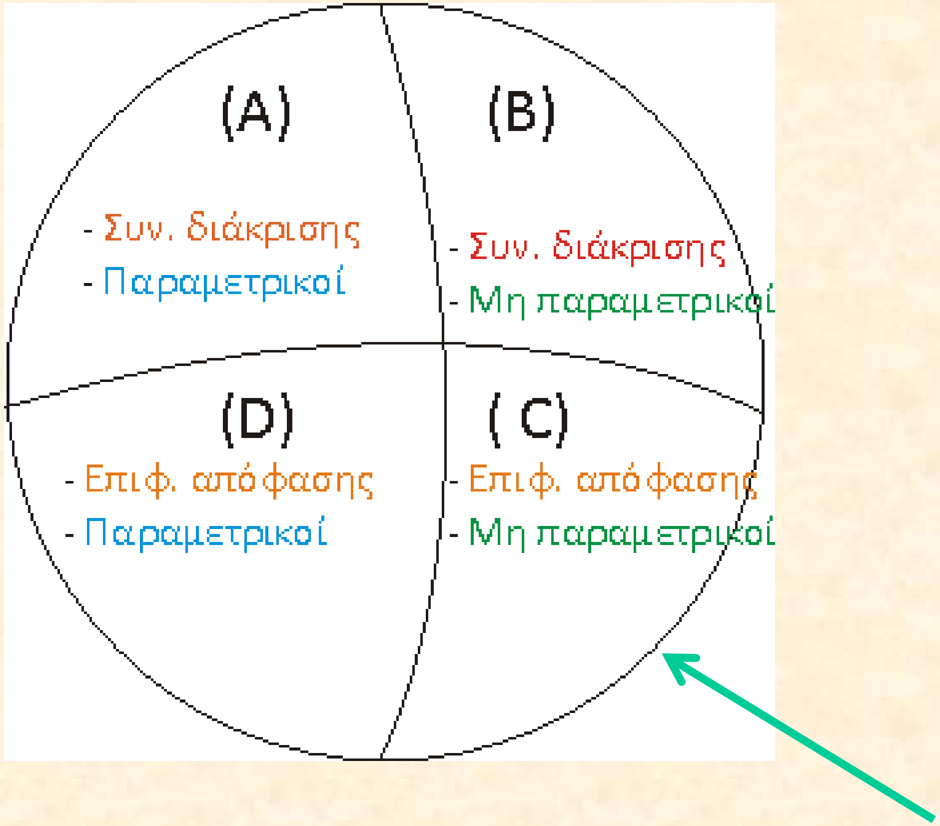
❖ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ
❖(PATTERN RECOGNITION)

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

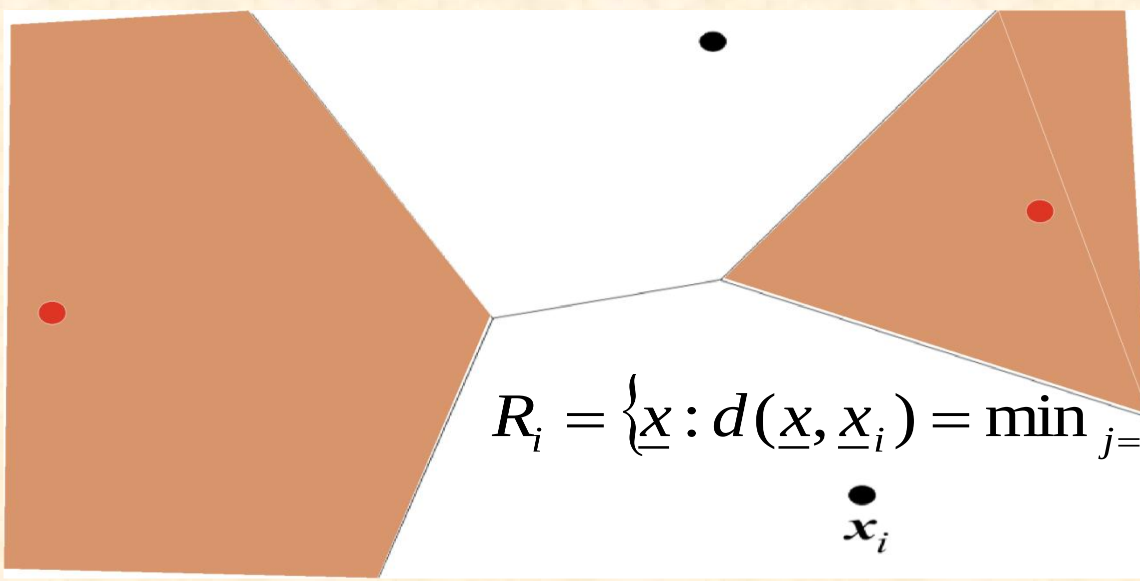
Ταξιν. k-πλησιέστερων γειτόνων

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

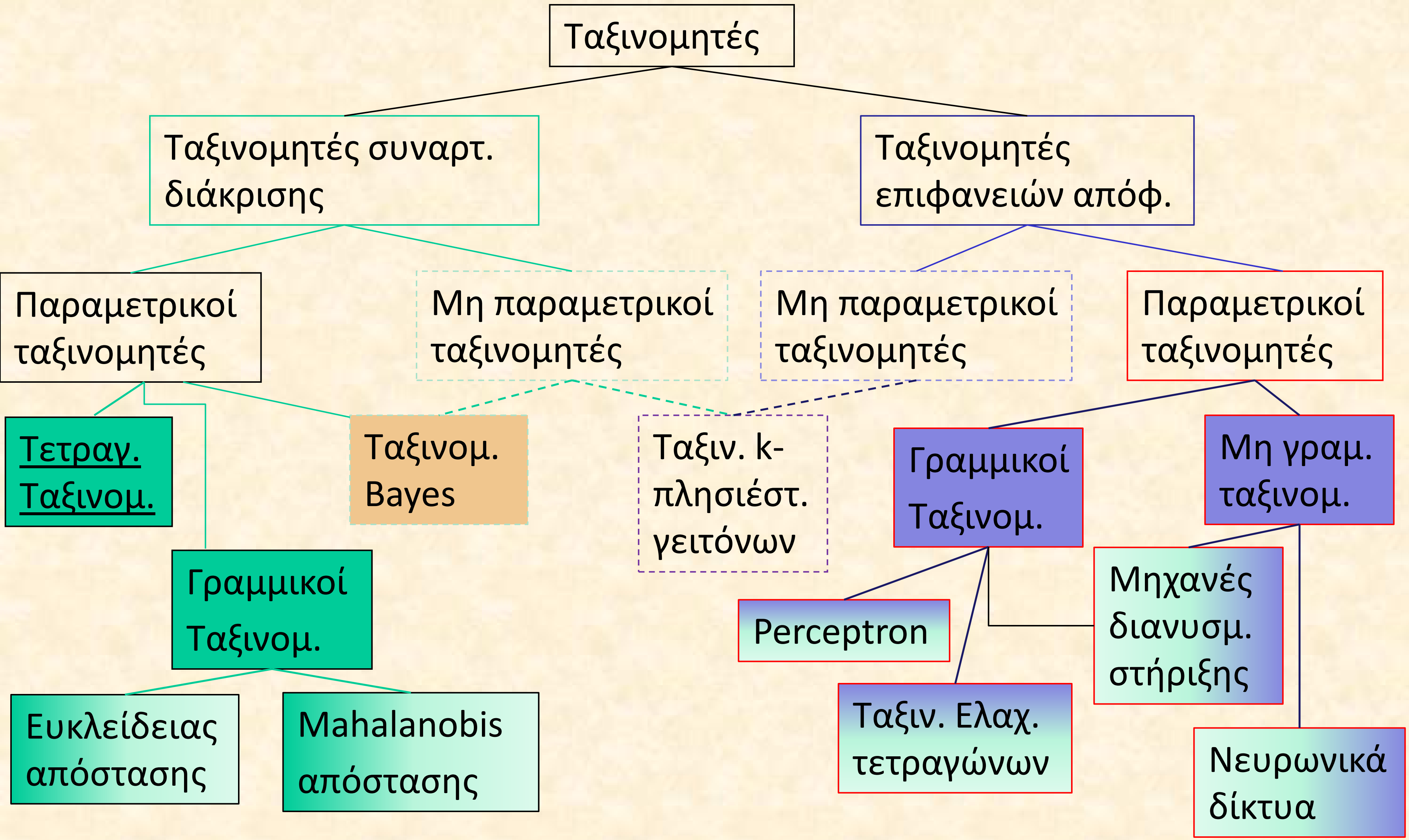
•Στην περίπτωση αυτή οι επιφάνειες απόφασης ορίζονται έμμεσα μέσω των σημείων του X.

•Για παράδειγμα ο ταξινομητής πλησιέστερου γείτονα διαχωρίζει το χώρο των χαρακτηριστικών με βάση την ψηφοθέτηση Voronoi (Voronoi tesselation)



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

"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,

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

Ταξινομητές με βάση τις συναρτήσεις διάκρισης

Ταξινομ. Bayes

$$-g_j(x) = f(P(\omega_j)p(x|\omega_j))$$
$$-\text{Εκτιμ. } p(x|\omega_j) \approx \hat{p}(x|\omega_j;\vartheta_j)$$
$$-\text{Εκτιμ. } \vartheta_j, \ \textit{με βάση το } X_j$$
$$(ML,\ EM): X_j \to \hat{\vartheta}_j$$
$$-x_i \to p(x_i|\omega_j) \approx \hat{p}(x_i|\omega_j;\hat{\vartheta}_j)$$

Τετραγωνικός ταξινομητής

$$-g_j(x) = (x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)$$
$$-\text{Υπόθεση}: \ N(\mu_j, \Sigma_j)$$
$$-\text{Εκτιμ. } \mu_j, \ \Sigma_j, \ \textit{με βάση το } X_j$$
$$(ML): X_j \to \hat{\mu}_j, \hat{\Sigma}_j$$
$$-x_i \to g_j(x_i) = (x_i-\hat{\mu}_j)^T \hat{\Sigma}_j^{-1}(x_i-\hat{\mu}_j)$$

Γραμμικός Ευκλείδειος ταξινομητής

$$-g_j(x) = (x-\mu_j)^T(x-\mu_j)$$
$$-\text{Υπόθεση}: \ N(\mu_j, I)$$
$$-\text{Εκτίμ. } \mu_j, \ \textit{με βάση το } X_j$$
$$(ML): X_j \to \hat{\mu}_j$$
$$-x_i \to g_j(x_i) = (x_i-\hat{\mu}_j)^T(x_i-\hat{\mu}_j)$$

Γραμμικός Mahalanobis ταξινομητής

$$-g_j(x) = (x-\mu_j)^T \Sigma^{-1}(x-\mu_j)$$
$$-\text{Υπόθεση}: \ N(\mu_j, \Sigma)$$
$$-\text{Εκτίμ. } \mu_j, \ \Sigma, \ \textit{με βάση το } X_j$$
$$(ML): X_j \to \hat{\mu}_j, \hat{\Sigma}$$
$$-x_i \to g_j(x_i) = (x_i-\hat{\mu}_j)^T \hat{\Sigma}^{-1}(x_i-\hat{\mu}_j)$$

“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,

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

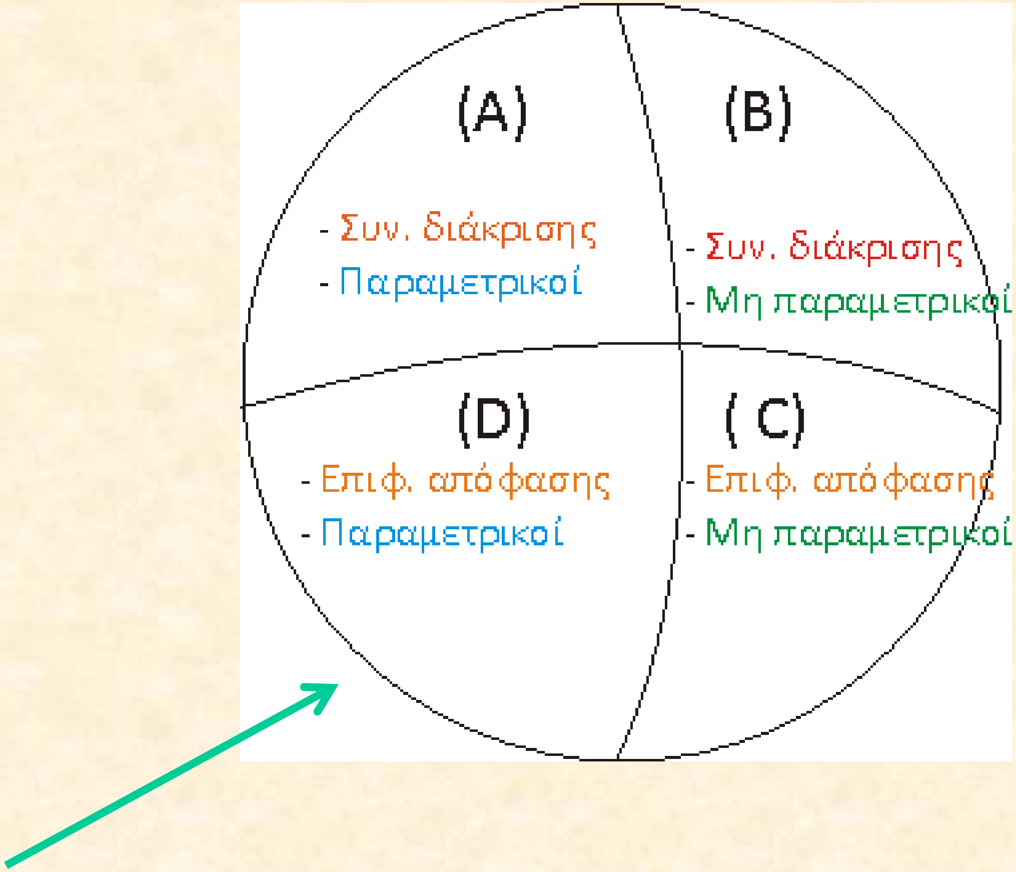
Μη παραμετρικοί ταξινομητές με βάση τις συναρτήσεις διάκρισης

Ταξινομητής Bayes

$$-g_j(x)=f(P(\omega_j)p(x|\omega_j))$$
$$-x_i \rightarrow p(x_i|\omega_j)\approx\hat{p}(x_i|\omega_j;X_j)$$
$$(\pi\alpha\rho\acute{\alpha}\theta\upsilon\rho\alpha\ \text{Parzen},$$
$$\varepsilon\kappa\tau\acute{\iota}\mu.\ \pi\upsilon\kappa\nu.\ \beta\acute{\alpha}\sigma\varepsilon\iota\ \tau\omega\nu\ k-\pi\lambda\eta\sigma.\ \gamma\varepsilon\iota\tau.$$

Ταξινομητής k-πλησιέστερων γειτόνων

$$-x_i \rightarrow g_j(x_i)=k_i^j$$


(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

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

d_i είναι η ετικέτα της κλάσης όπου ανήκει το $\boldsymbol{x}_i$ (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 $\mathbf{w}$ 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 $\mathbf{w}$ 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 $\mathbf{w}=[w_0, w_1,\ldots,w_l, w_{11},\ldots,w_{1l},w_{22},\ldots,w_{2l},\ldots,w_{ll}]^{\mathrm{T}}$

<u>Σημείωση:</u> 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.

Στρατηγική αντιμετώπισης του προβλήματος:

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 $\boldsymbol{w}$. **Η θέση του βέλτιστου αυτής ορίζει την επιφάνεια απόφασης**.

Σημαντική παρατήρηση: 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)

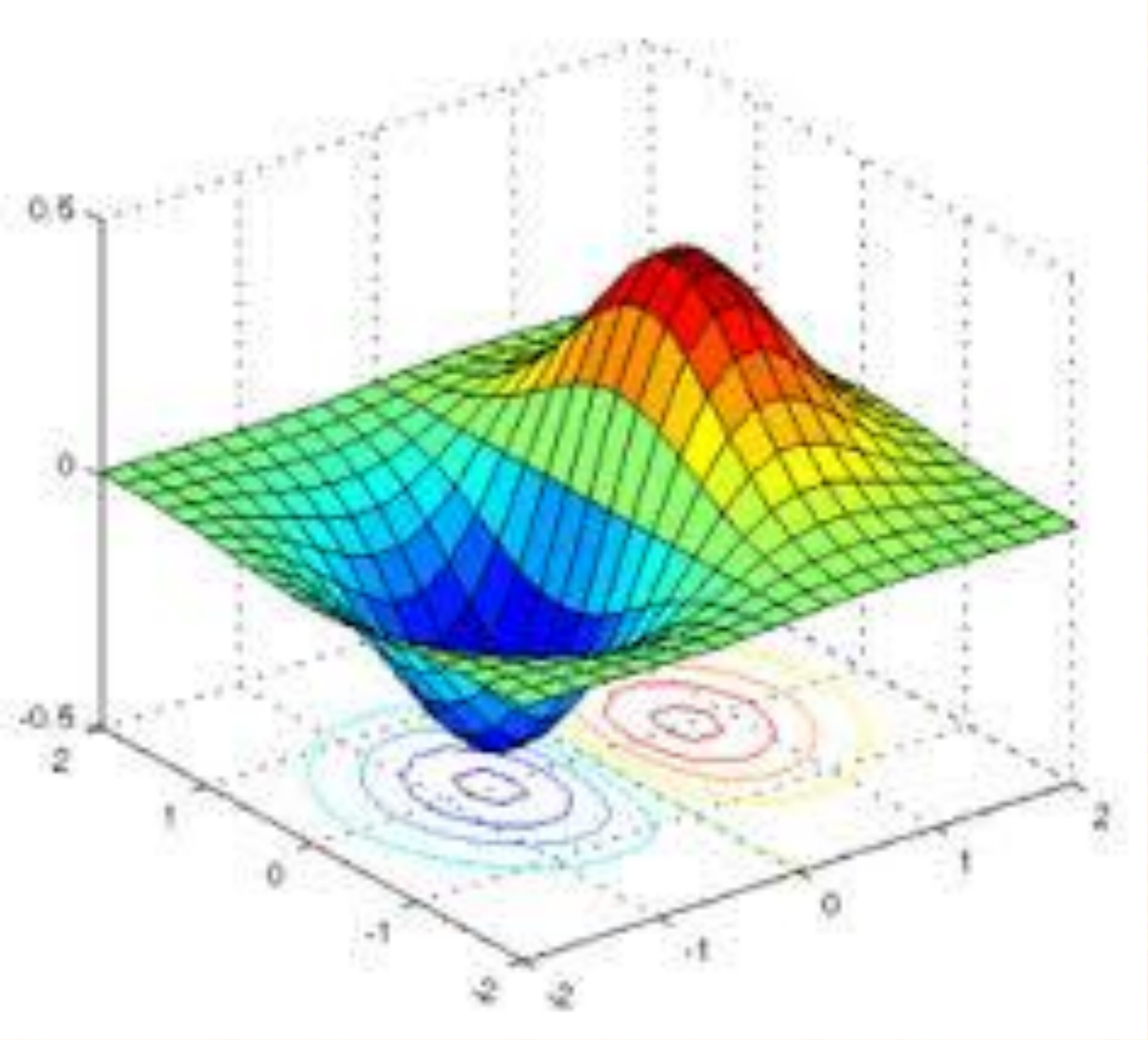
-t=0

-**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

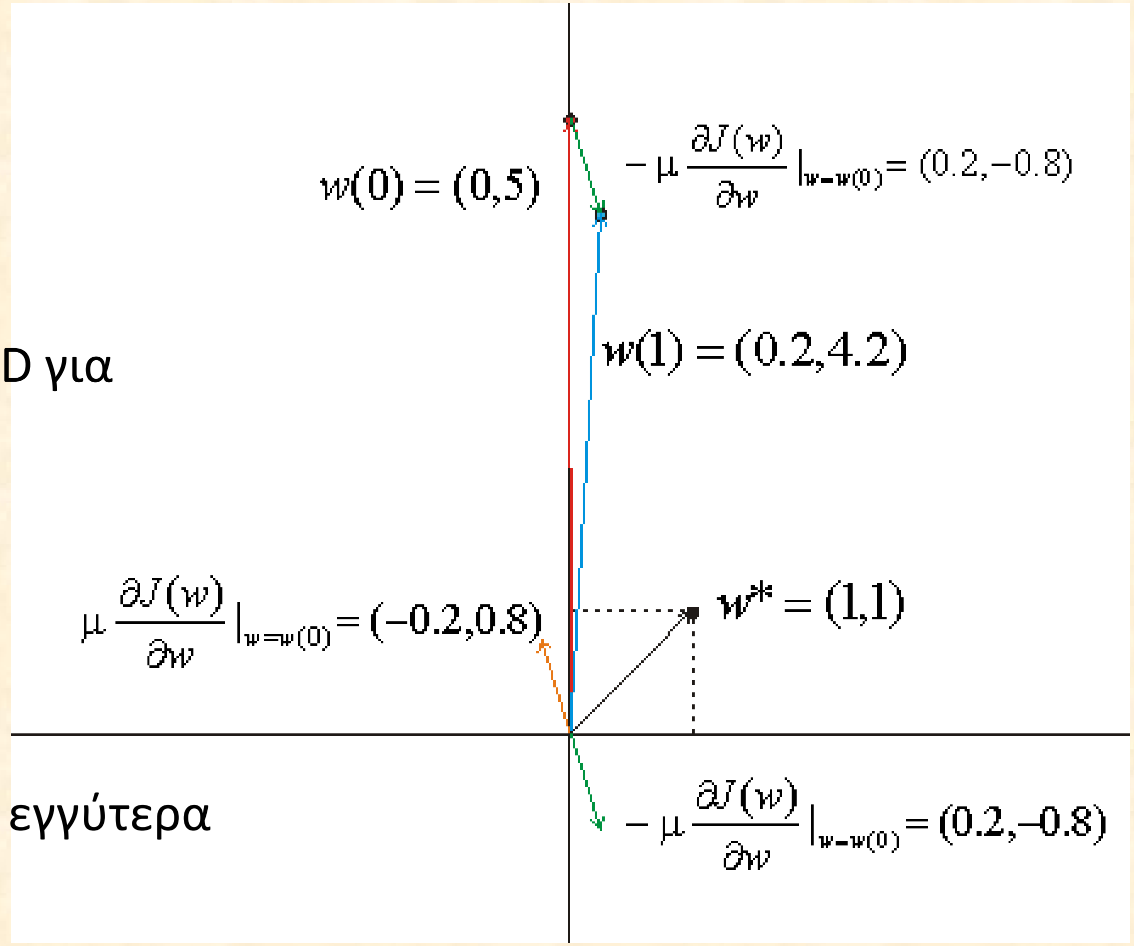
-**Ένα παράδειγμα:** Έστω $\boldsymbol{w}=[w_1, w_2]^T$ και $J(\boldsymbol{w})=(w_1-1)^2+(w_1-1)^2$. Είναι σαφές ότι η $J(\boldsymbol{w})$ λαμβάνει την ελάχιστη τιμή της στο $\boldsymbol{w}^*=[1, 1]^T$.

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

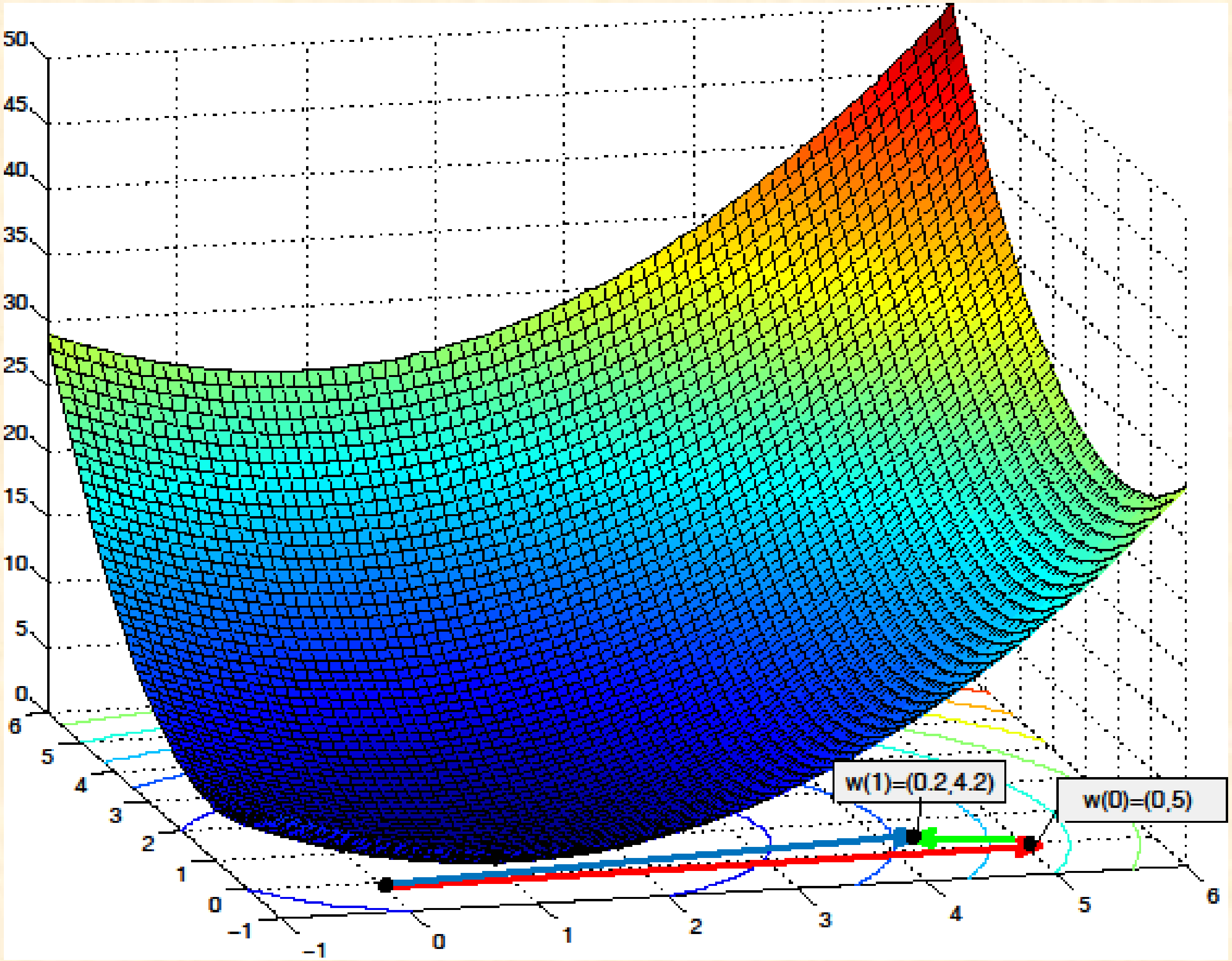
-Εφαρμόζοντας τον αλγόριθμο GD για $\boldsymbol{w}(0)=[0, 5]^T$, και $\mu=0.1$, έχουμε

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

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



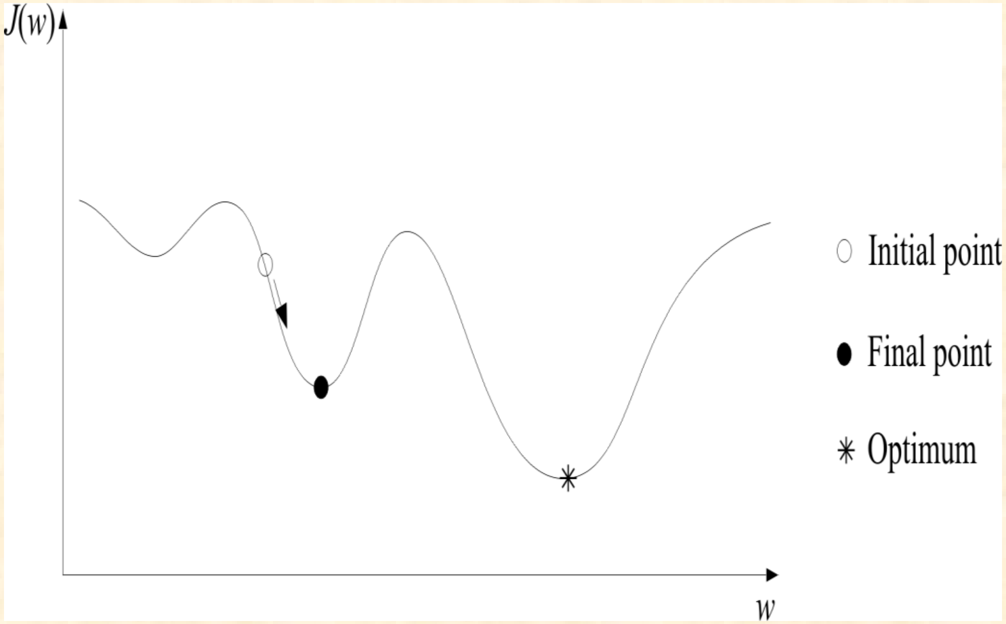
ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ



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 $J(\boldsymbol{w})$.



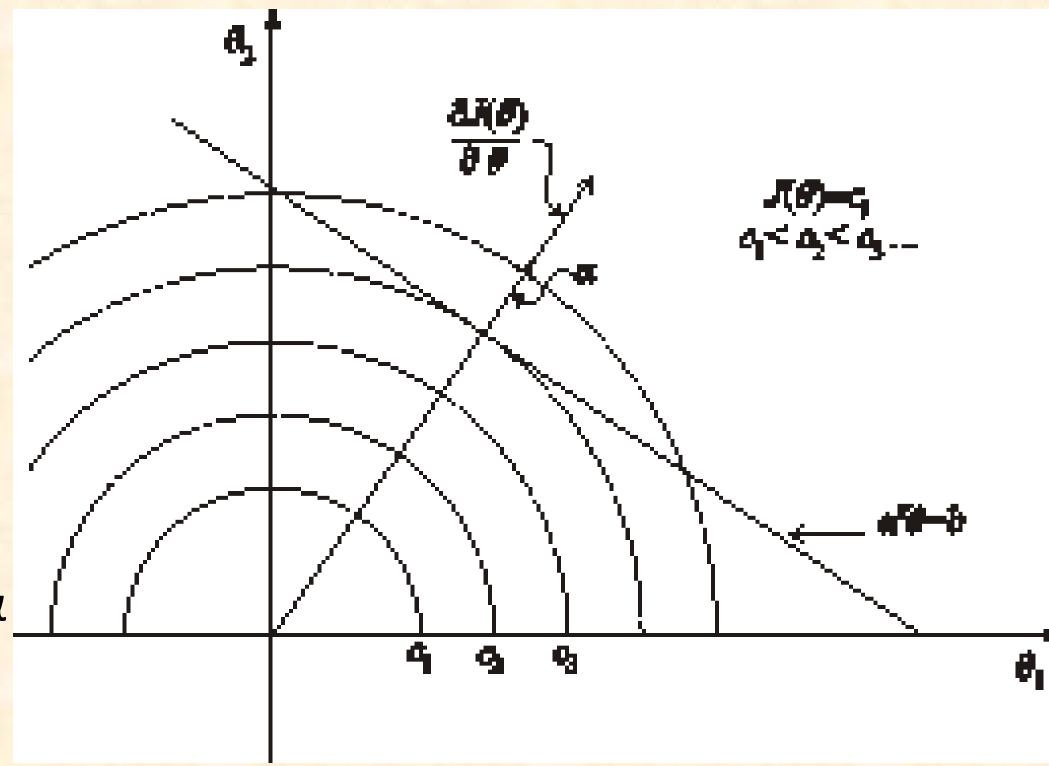
ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

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 $A\boldsymbol{w}=\boldsymbol{b}$

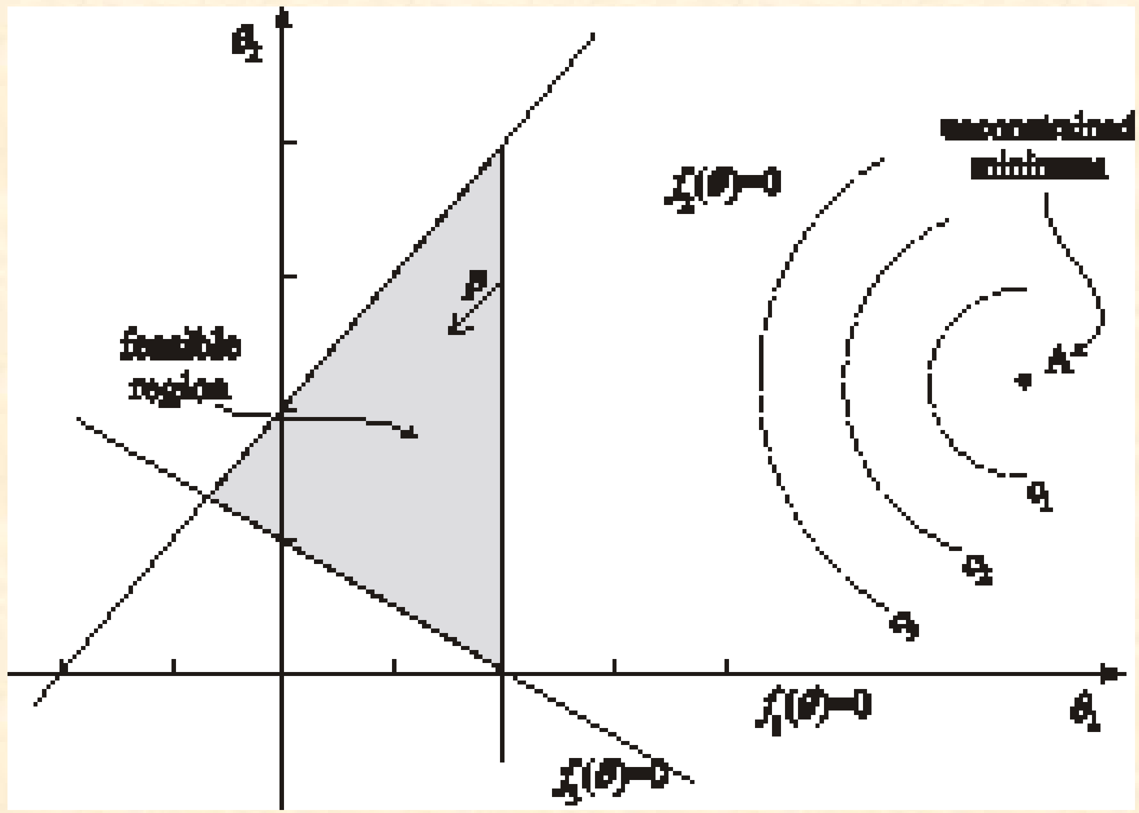
ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

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

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

Για γραμμικούς περιορισμούς ανισότητας, το πρόβλημα διατυπώνεται ως εξής

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



ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

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

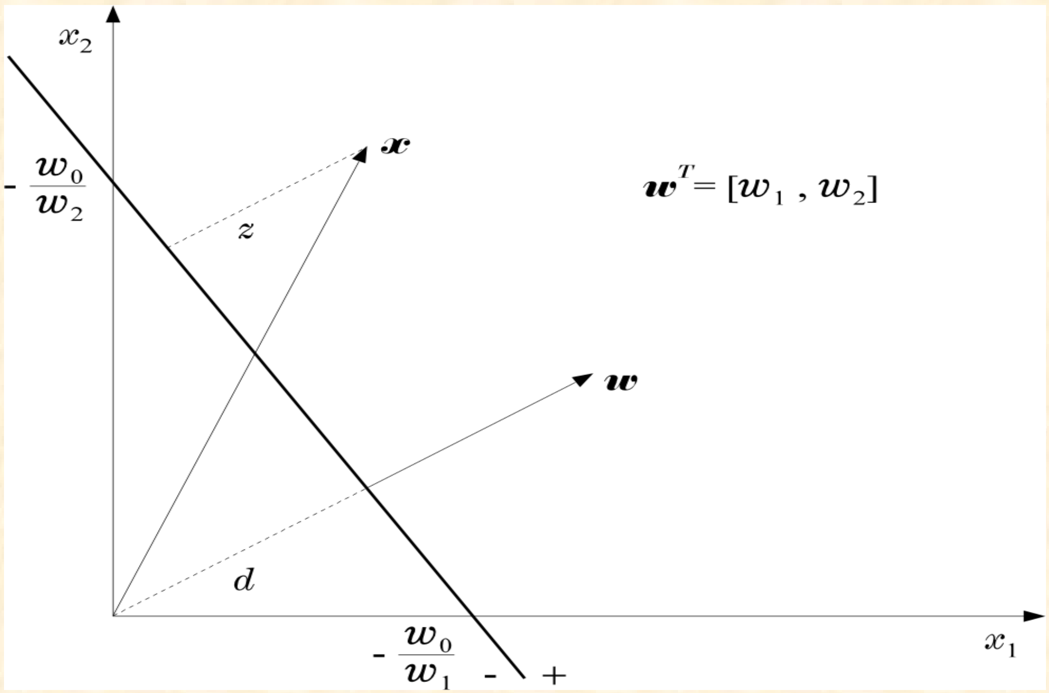
$$(H): h(x;w) = w_1\, x_1 + \ldots + w_l\, x_l + w_0 = \sum_{k=1}^{l} w_k x_k + w_0 = w^T x \,+\, w_0 = 0$$

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

- 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 από υπερεπίπεδα ίδιου προσανατολισμού.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ



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