❖ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ
❖(PATTERN RECOGNITION)

Σέργιος Θεοδωρίδης

Κωνσταντίνος Κουτρούμπας

Version 2

❖ Υπερεπίπεδα απόφασης

➢ Τετραγωνικοί όροι: $\underline{x}^{T} \Sigma_i^{-1} \underline{x}$

Αν ΟΛΟΙ $\Sigma_i = \Sigma$ (είναι ίδιοι) οι τετραγωνικοί όροι δεν επηρεάζουν, αφού δεν εμπλέκονται στις συγκρίσεις. Τότε μπορούμε να γράψουμε ισοδύναμα:

$$\boxed{\begin{aligned} g_i(\underline{x}) &= \underline{w}_i^{T} \underline{x} + w_{io} \\ \underline{w}_i &= \Sigma^{-1} \underline{\mu}_i \\ w_{i0} &= \ln P(\omega_i) - \frac{1}{2} \underline{\mu}^{T}{}_{i} \Sigma^{-1} \underline{\mu}_i \end{aligned}}$$

Οι συναρτήσεις διάκρισης είναι ΓΡΑΜΜΙΚΕΣ

- Έστω επιπλέον ότι:
  - $\Sigma = \sigma^2 I.$ Τότε

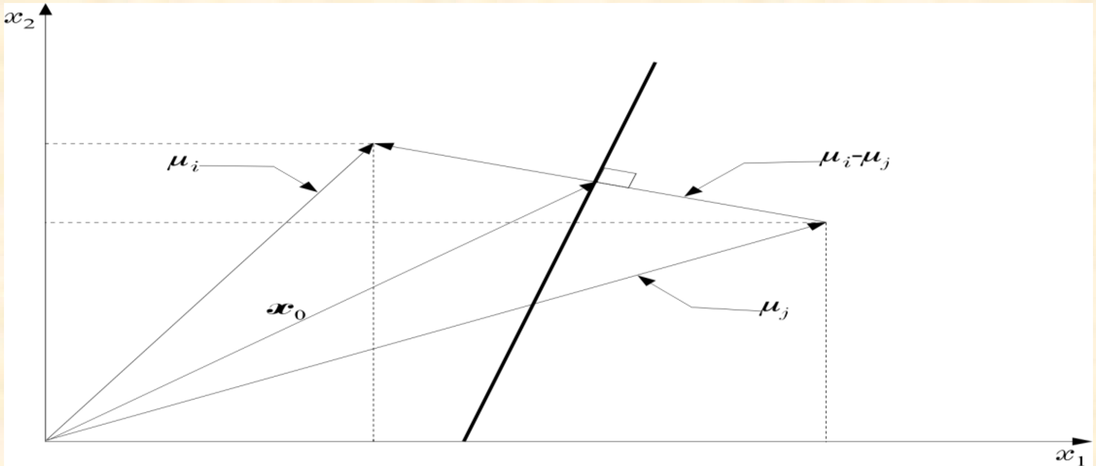
$$g_i(\underline{x}) = \frac{1}{\sigma^2} \underline{\mu}_i^T \underline{x} + w_{i0}$$

  - $$g_{ij}(\underline{x}) = g_i(\underline{x}) - g_j(\underline{x}) = 0$$

$$= \underline{w}^T (\underline{x} - \underline{x}_o)$$

  - $$\underline{w} = \underline{\mu}_i - \underline{\mu}_j,$$

  - $$\underline{x}_o = \frac{1}{2}(\underline{\mu}_i + \underline{\mu}_j) - \sigma^2 \ln \frac{P(\omega_i)}{P(\omega_j)} \frac{\underline{\mu}_i - \underline{\mu}_j}{\left\| \underline{\mu}_i - \underline{\mu}_j \right\|^2}$$



- Μη διαγώνιος: $\Sigma \neq \sigma^2 I$

- $$g_{ij}(\underline{x}) = \underline{w}^T(\underline{x} - \underline{x}_0) = 0$$

- $$\underline{w} = \Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)$$

- $$\underline{x}_0 = \frac{1}{2}(\underline{\mu}_i + \underline{\mu}_j) - \ell n\left(\frac{P(\omega_i)}{P(\omega_j)}\right)\frac{\underline{\mu}_i - \underline{\mu}_j}{\left\|\underline{\mu}_i - \underline{\mu}_j\right\|^2_{\Sigma^{-1}}}$$

$$\left\|\underline{x}\right\|_{\Sigma^{-1}} \equiv (\underline{x}^T \Sigma^{-1} \underline{x})^{\frac{1}{2}}$$

- Υπερεπίπεδο απόφασης
  - όχι κάθετο στο $\underline{\mu}_i - \underline{\mu}_j$
  - κάθετο στο $\Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)$

❖ Ταξινομητές ελάχιστης απόστασης

- $P(\omega_i) = \dfrac{1}{M}$ ισοπίθανες

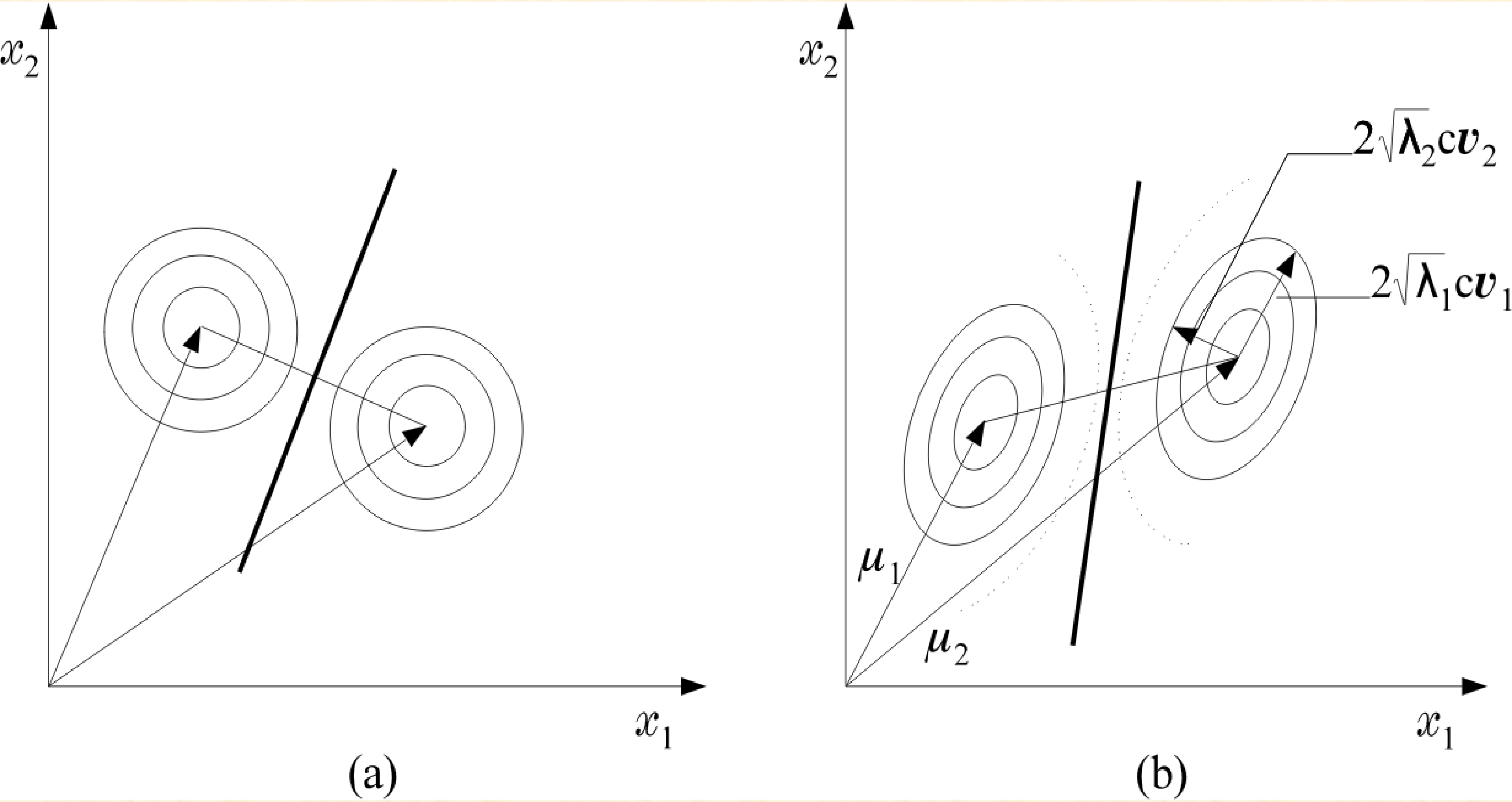
- $g_i(\underline{x}) = -\dfrac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma^{-1}(\underline{x} - \underline{\mu}_i)$

- $\Sigma = \sigma^2 I :$ Καταχώρησε $\underline{x} \rightarrow \omega_i :$

  Ευκλείδεια απόσταση: $d_E \equiv \left\| \underline{x} - \underline{\mu}_i \right\|$
  μικρότερη

- $\Sigma \neq \sigma^2 I :$ Καταχώρησε $\underline{x} \rightarrow \omega_i :$

  Mahalanobis απόσταση: $d_m = ((\underline{x} - \underline{\mu}_i)^T \Sigma^{-1}(\underline{x} - \underline{\mu}_i))^{\frac{1}{2}}$
  μικρότερη



x_2
x_1
(a)
x_2
$2\sqrt{\lambda_2}c\boldsymbol{v}_2$
$2\sqrt{\lambda_1}c\boldsymbol{v}_1$
μ_1
μ_2
x_1
(b)

❖ Παράδειγμα:

Δοθέντων των $\omega_1,\ \omega_2$: $P(\omega_1) = P(\omega_2)$ και $p(\underline{x}|\omega_1) = N(\underline{\mu}_1,\ \Sigma)$,

$$p(\underline{x}|\omega_2) = N(\underline{\mu}_2,\ \Sigma),\ \underline{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \ \underline{\mu}_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix},\ \ \Sigma = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$$

κατηγοριοποίησε το διάνυσμα $\underline{x} = \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix}$ με χρήση Bayesian ταξινομητή:

- $\Sigma^{-1} = \begin{bmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{bmatrix}$
- Υπόλ. Mahalanobis απόστ. d_m από $\mu_1,\ \mu_2$: $d^2_{m,1} = \begin{bmatrix} 1.0, & 2.2 \end{bmatrix} \Sigma^{-1} \begin{bmatrix} 1.0 \\ 2.2 \end{bmatrix} = 2.952,\ d^2_{m,2} = \begin{bmatrix} -2.0, & -0.8 \end{bmatrix} \Sigma^{-1} \begin{bmatrix} -2.0 \\ -0.8 \end{bmatrix} = 3.672$
- Καταχώρησε $\underline{x} \rightarrow \omega_1$. Παρατηρείστε ότι $d_{E,2} < d_{E,1}$

ΕΚΤΙΜΗΣΗ ΑΓΝΩΣΤΩΝ ΣΥΑΡΤΗΣΕΩΝ ΠΥΚΝΟΤΗΤΑΣ ΠΙΘΑΝΟΤΗΤΑΣ

❖ Μέγιστη πιθανοφάνεια (Maximum Likelihood)

(*) Εστω $\underline{x}_1,\ \underline{x}_2,\ldots,\ \underline{x}_N$ γνωστά και ανεξάρτητα

(*) Έστω $p(\underline{x})$ γνωστής μορφής με άγνωστο διάνυσμα

παραμέτρων $\underline{\theta}$: $p(\underline{x}) \equiv p(\underline{x};\underline{\theta})$

(*) $X = \{\underline{x}_1, \underline{x}_2, \ldots \underline{x}_N\}$

(*) $p(X;\underline{\theta}) \equiv p(\underline{x}_1, \underline{x}_2, \ldots \underline{x}_N; \underline{\theta})$

$= \prod_{k=1}^{N} p(\underline{x}_k;\underline{\theta})$

που είναι γνωστή ως η πιθανοφάνεια (Likelihood)

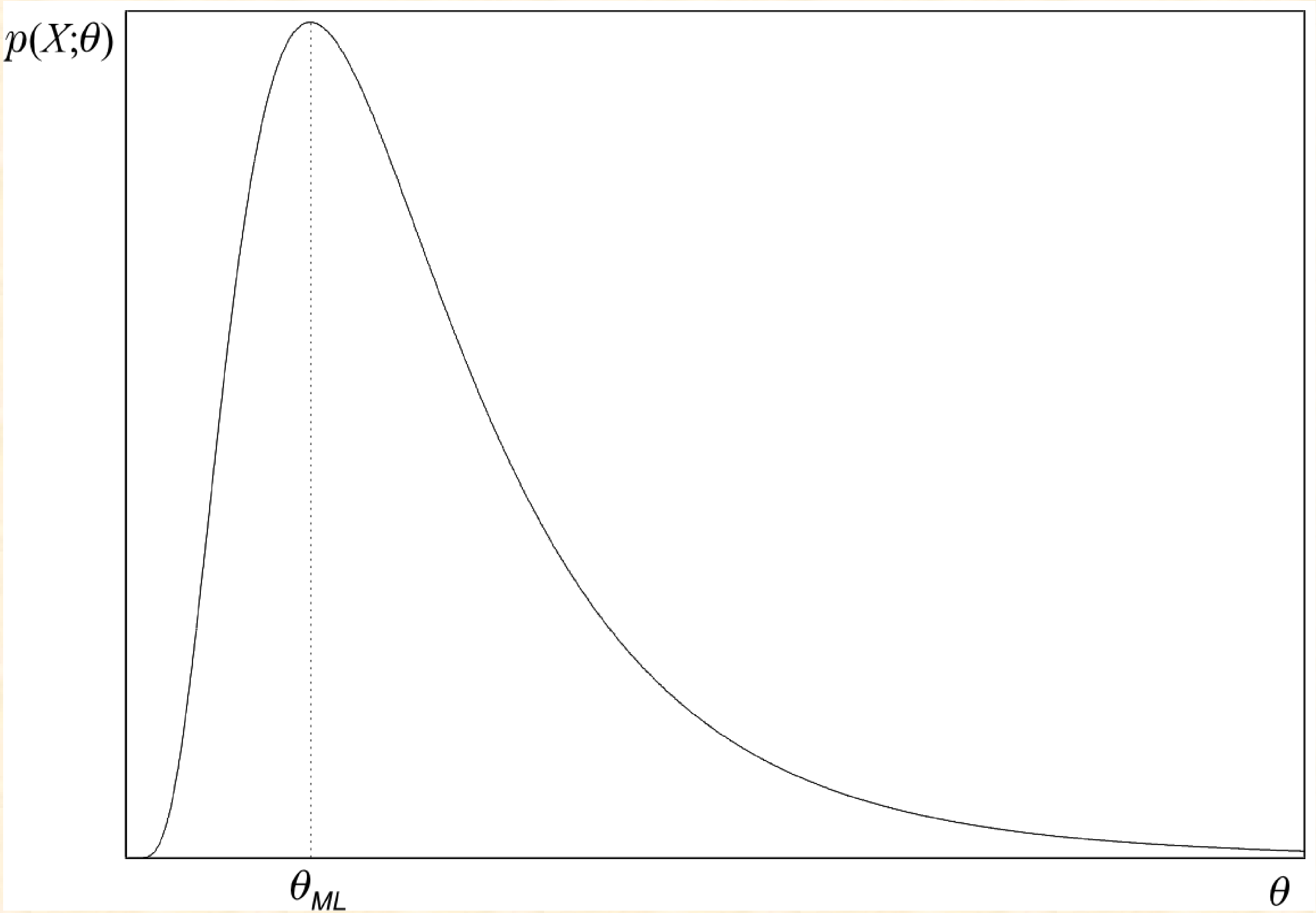
του $\underline{\theta}$ ως προς το X

Η μέθοδος:

- $\boxed{\hat{\underline{\theta}}_{\mathrm{ML}} : \arg\max_{\underline{\theta}} \prod_{k=1}^{N} p(\underline{x}_k;\underline{\theta})}$

- $L(\underline{\theta}) \equiv \ln p(X;\underline{\theta}) = \sum_{k=1}^{N} \ln p(\underline{x}_k;\underline{\theta})$

- $\hat{\underline{\theta}}_{ML} : \dfrac{\partial L(\underline{\theta})}{\partial(\underline{\theta})} = \sum_{k=1}^{N} \dfrac{1}{p(\underline{x}_{k;}\underline{\theta})} \dfrac{\partial p(\underline{x}_{k;}\underline{\theta})}{\partial(\underline{\theta})} = \underline{0}$



$p(X;\theta)$
θ_{ML}
θ

Αν, πράγματι, υπάρχει ένα $\underline{\theta}_0$ έτσι ώστε

$p(\underline{x}) = p(\underline{x};\underline{\theta}_0)$, τότε

$$\lim_{N\to\infty} E[\underline{\theta}_{ML}] = \underline{\theta}_0$$

$$\lim_{N\to\infty} E\left\|\underline{\hat{\theta}}_{ML} - \underline{\theta}_0\right\|^2 = 0$$

- Ασυμπτωτικά αμερόληπτος (unbiased) εκτιμητής:
  Η μέση τιμή της τυχαίας μεταβλητής $\underline{\hat{\theta}}_{ML}$ είναι η πραγματική τιμή της άγνωστης παραμέτρου

- Ασυμπτωτικά συνεπής (consistent) εκτιμητής:
  Η διασπορά της ML εκτίμησης τείνει στο μηδέν.

❖ Παράδειγμα:

$$p(\underline{x}):\ N(\underline{\mu},\Sigma):\ \underline{\mu}\ \text{άγνωστο},\quad \underline{x}_1,\underline{x}_2,\ldots,\underline{x}_N \quad p(\underline{x}_k)\equiv p(\underline{x}_k;\underline{\mu})$$

$$L(\underline{\mu})=\ln\prod_{k=1}^{N}p(\underline{x}_k;\underline{\mu})=C-\frac{1}{2}\sum_{k=1}^{N}(\underline{x}_k-\underline{\mu})^T\Sigma^{-1}(\underline{x}_k-\underline{\mu})$$

$$p(\underline{x}_k;\underline{\mu})=\frac{1}{(2\pi)^{\frac{l}{2}}\left|\Sigma\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\underline{x}_k-\underline{\mu})^T\Sigma^{-1}(\underline{x}_k-\underline{\mu})\right)$$

$$\frac{\partial L(\underline{\mu})}{\partial(\underline{\mu})}\equiv\begin{bmatrix}\frac{\partial L}{\partial\mu_1}\\ .\\ .\\ .\\ \frac{\partial L}{\partial\mu_l}\end{bmatrix}=\sum_{k=1}^{N}\Sigma^{-1}(\underline{x}_k-\underline{\mu})=\underline{0}\Rightarrow\underline{\mu}_{ML}=\frac{1}{N}\sum_{k=1}^{N}\underline{x}_k$$

Υπενθύμιση: Αν $A=A^T\Rightarrow\dfrac{\partial(\underline{\alpha}^TA\underline{\alpha})}{\partial\underline{\alpha}}=2A\underline{\alpha}$

- Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας (Maximum Aposteriori Probability Estimation)
  - Στην ML μέθοδο, το $\underline{\theta}$ λογιζόταν ως παράμετρος
  - Εδώ θα θεωρήσουμε το $\underline{\theta}$ ως τυχαίο διάνυσμα που περιγράφεται από (υποτίθεται γνωστή) pdf $p(\underline{\theta})$.
  - Δοθέντος

$$X = \{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N\}$$

Υπολόγισε το μέγιστο της

$$p(\underline{\theta}|X)$$

  - From Bayes theorem

$$p(\underline{\theta})p(X|\underline{\theta}) = p(X)p(\underline{\theta}|X) \text{ or}$$

$$p(\underline{\theta}|X) = \frac{p(\underline{\theta})p(X|\underline{\theta})}{p(X)}$$

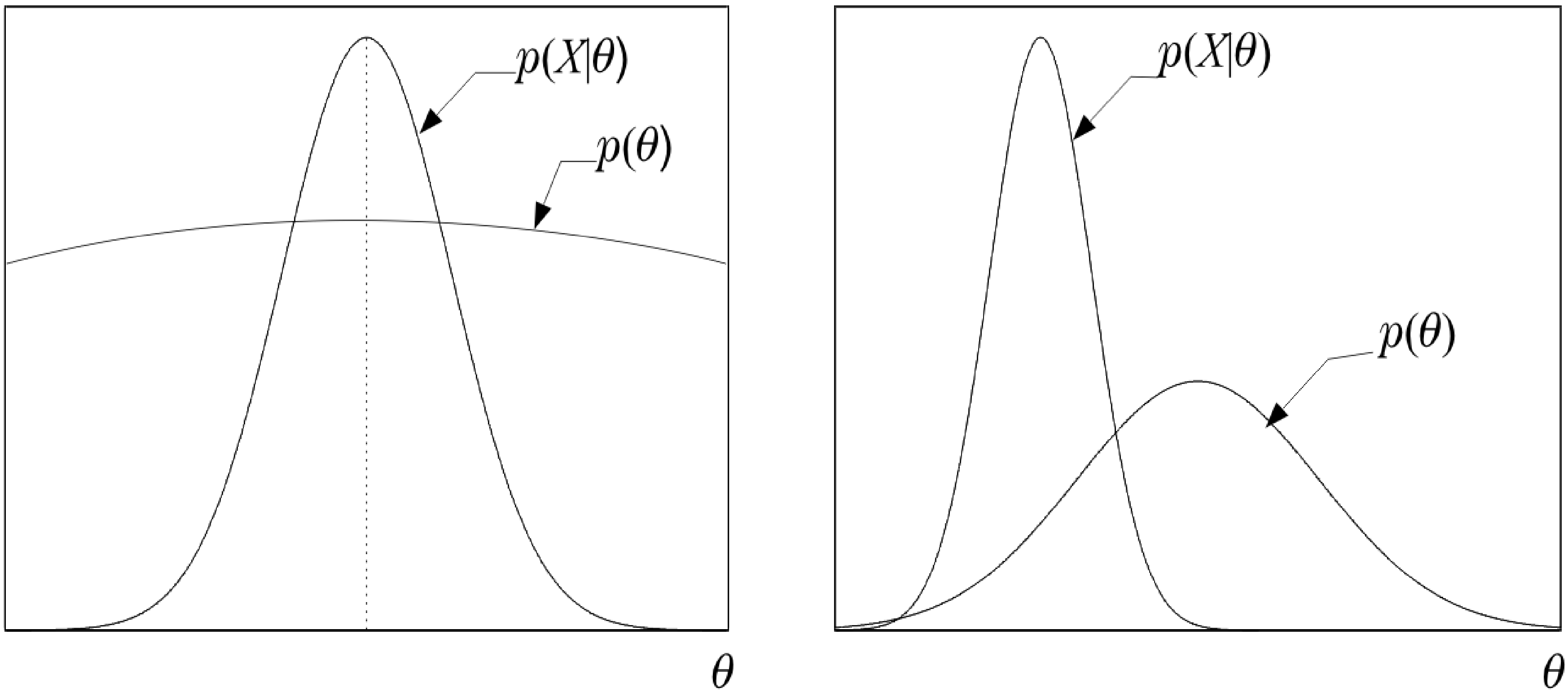
- Η μέθοδος:

$$\hat{\underline{\theta}}_{MAP} = \arg\max_{\underline{\theta}} p(\underline{\theta} \mid X)\ \acute{\eta}$$

$$\boxed{\hat{\underline{\theta}}_{MAP} : \frac{\partial}{\partial \underline{\theta}} \left(p(\underline{\theta})\, p(X \mid \underline{\theta}) \right)}$$

Αν η $p(\underline{\theta})$ είναι ομοιόμορφη ή αρκετά ευρεία:

$$\hat{\underline{\theta}}_{MAP} \cong \underline{\theta}_{ML}$$



(a) (b)

❖ Παράδειγμα:

$$p(\underline{x}): \quad N(\underline{\mu}, \sigma^2 \mathrm{I}), \quad \underline{\mu} \;\; \acute{\alpha}\gamma\nu\omega\sigma\tau o, \quad X = \{\underline{x}_1, \dots, \underline{x}_N\}$$

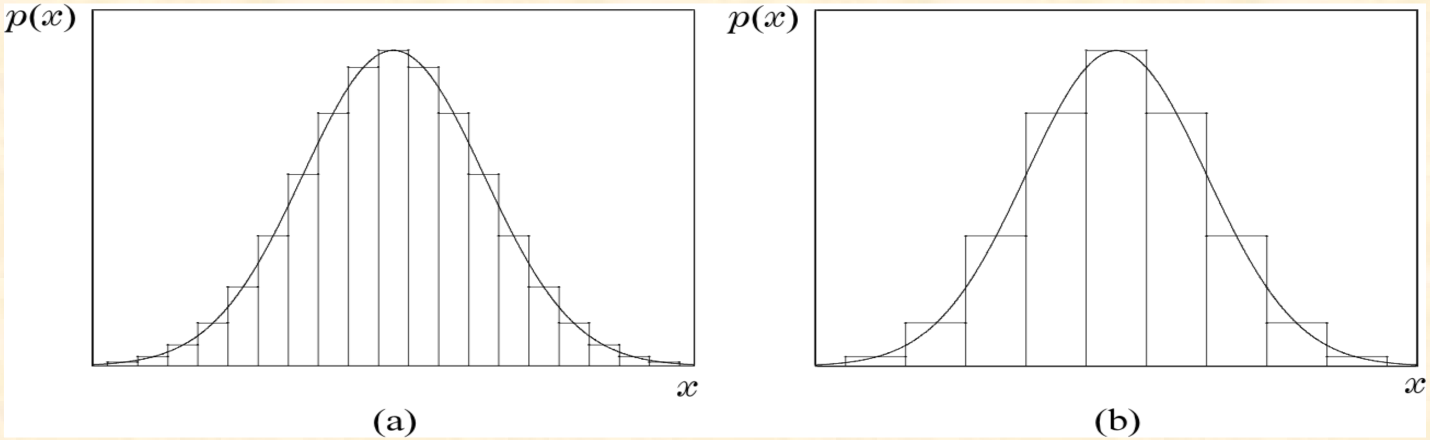
$$p(\underline{\mu}) = \frac{1}{(2\pi)^{\frac{l}{2}} \sigma_\mu^l} \exp\left(-\frac{\left\|\underline{\mu} - \underline{\mu}_0\right\|^2}{2\sigma_\mu^2}\right)$$

$$\underline{\hat{\theta}}_{MAP}: \quad \frac{\partial}{\partial \underline{\mu}} \ln\left(\prod_{k=1}^{N} p(\underline{x}_k | \underline{\mu}) p(\underline{\mu})\right) = \underline{0} \quad \text{ή} \quad \sum_{k=1}^{N} \frac{1}{\sigma^2}(\underline{x}_k - \underline{\hat{\mu}}) - \frac{1}{\sigma_\mu^2}(\underline{\hat{\mu}} - \underline{\mu}_0) = \underline{0} \Rightarrow$$

$$\underline{\hat{\mu}}_{MAP} = \frac{\underline{\mu}_0 + \frac{\sigma_\mu^2}{\sigma^2} \sum_{k=1}^{N} \underline{x}_k}{1 + \frac{\sigma_\mu^2}{\sigma^2} N} \quad \textit{For} \quad \frac{\sigma_\mu^2}{\sigma^2} >> 1, \text{ ή για N} \to \infty$$

$$\underline{\hat{\mu}}_{\mathrm{MAP}} \cong \underline{\hat{\mu}}_{ML} = \frac{1}{N} \sum_{k=1}^{N} \underline{x}_k$$

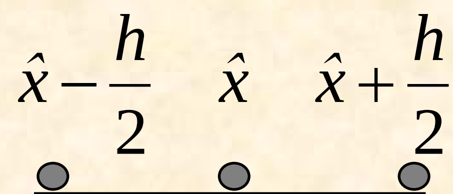
❖ Μη παραμετρική εκτίμηση



- $$P \approx \frac{k_N}{N}$$

  k_N στο h

  N συνολικά



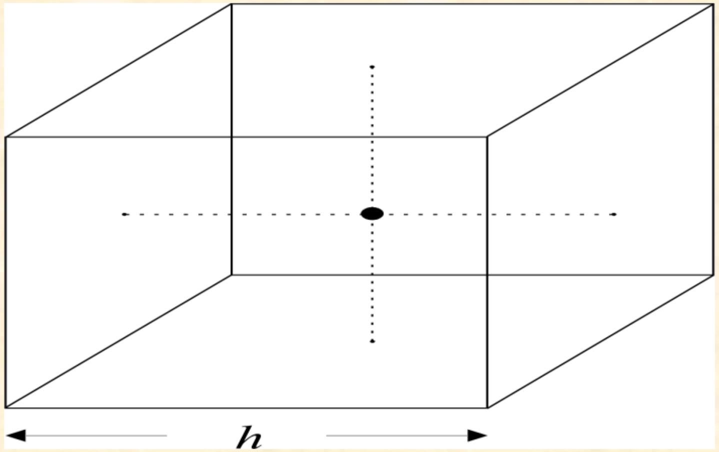
- $$\hat{p}(x) \equiv \hat{p}(\hat{x}) = \frac{1}{h}\;\frac{k_N}{N},\quad |x\text{-}\hat{x}| \leq \frac{h}{2}$$

- Αν $p(x)$ συνεχής, $\hat{p}(x) \to p(x)$ καθώς $N \to \infty$, αν

$$h_N \to 0, \qquad k_N \to \infty, \qquad \frac{k_N}{N} \to 0$$

❖ Παράθυρα Parzen

- ➢ Διαίρεση του πολυδιάστατου χώρου σε υπερκύβους



➢ Ορίζουμε

$$\phi(\underline{x}_i) = \begin{cases} 1 & |x_{ij}| \leq 1/2 \\ 0 & \text{διαφορετικά} \end{cases}$$

- Δηλαδή, είναι 1 μέσα σε υπερκύβο μοναδιαίας πλευράς με κέντρο το 0

- $$\boxed{\hat{p}(\underline{x}) = \frac{1}{h^l}\left(\frac{1}{N}\sum_{i=1}^{N}\varphi\left(\frac{\underline{x}_i - \underline{x}}{h}\right)\right)}$$

- $$\frac{1}{\text{όγκος}} * \frac{1}{N} * \text{ αριθμός σημείων εντός}$$

ένας υπερκύβος πλευράς h κεντραρισμένος στο $\underline{x}$

- Το πρόβλημα: $p(\underline{x})$ συνεχής

  $\phi(.)$ ασυνεχής

- Παράθυρα Parzen-Πυρήνες-συναρτήσεις δυναμικού

  $\phi(\underline{x})$ είναι ομαλή

$$\phi(\underline{x}) \geq 0, \quad \int_{\underline{x}} \phi(\underline{x})\, d\underline{x} = 1$$

➢ Μέση τιμή

$$E[\hat{p}(\underline{x})]=\frac{1}{h^l}\Big(\frac{1}{N}\sum_{i=1}^{N}E[\varphi(\frac{\underline{x}_i-\underline{x}}{h})]\Big)=\int_{\underline{x}'}\frac{1}{h^l}\varphi(\frac{\underline{x}'-\underline{x}}{h})p(\underline{x}')d\underline{x}'$$

- $h\to 0,\ \frac{1}{h^l}\to\infty$
- $h\to 0$ το εύρος της $\phi(\frac{\underline{x}'-\underline{x}}{h})\to 0$
- $\int\frac{1}{h^l}\varphi(\frac{\underline{x}'-\underline{x}}{h})d\underline{x}=1$
- $h\to 0\ \ \frac{1}{h^l}\varphi(\frac{\underline{x}}{h})\to\delta(\underline{x})$

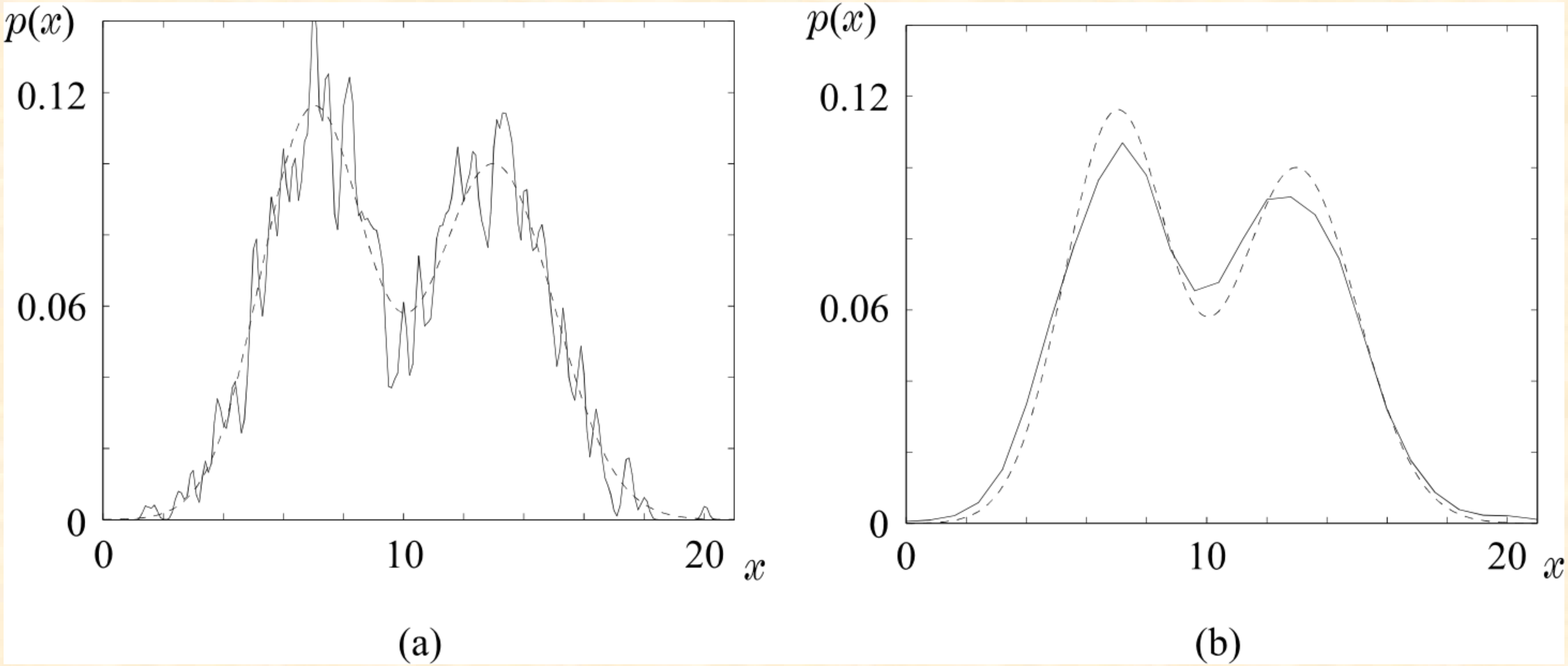
$$E[\hat{p}(x)]=\int_{\underline{x}'}\delta(\underline{x}'-\underline{x})p(\underline{x}')d\underline{x}'=p(\underline{x})$$

Συνεπώς $\boxed{\text{απροκατάληπτος στο όριο}}$

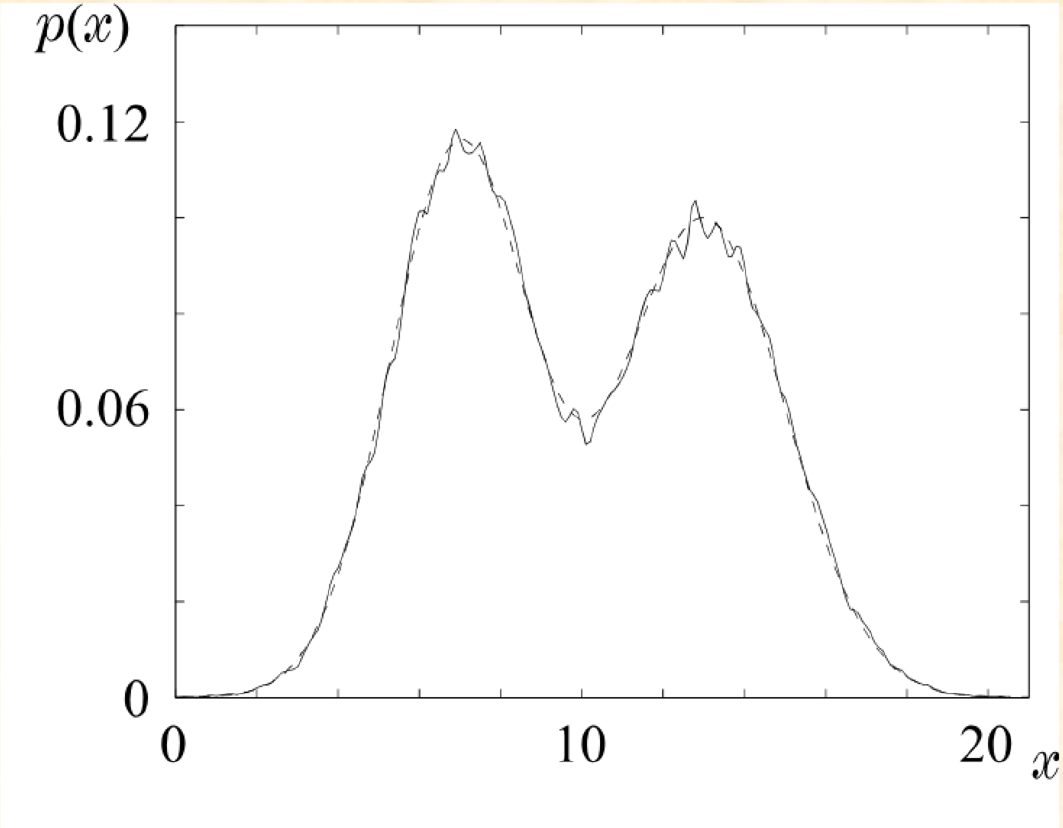
➢ Διασπορά

- Όσο μικρότερο το h, τόσο μεγαλύτερη η διασπορά

h=0.1, N=1000 h=0.8, N=1000



h=0.1, N=10000



(b)

➢Όσο μεγαλύτερο το N, τόσο καλύτερη η ακρίβεια

- Αν
  - $h \to 0$
  - $N \to \infty$
  - $h_N \to \infty$

Ασυμπτωτικά απροκατάληπτος

❖ ΚΑΤΑΡΑ ΤΗΣ ΔΙΑΣΤΑΤΙΚΌΤΗΤΑΣ (CURSE OF DIMENSIONALITY)

- Σε όλες τις μεθόδους, μέχρι τώρα, είδαμε ότι όσο μεγαλύτερος είναι ο αριθμός των σημείων, N, τόσο καλύτερη είναι η προκύπτουσα εκτίμηση.

- Αν στο μονοδιάστατο χώρο ένα διάστημα, που περιέχει N σημεία, is αρκετό (για καλή εκτίμηση), στον δισδιάστατο χώρο το αντίστοιχο τετράγωνο θα απαιτεί N^2 και στον ℓ-διάστατο χώρο ο ℓ-διάστατος κύβος θα απαιτεί N^ℓ σημεία.

- Η εκθετικά αύξηση του αριθμού των αναγκαίων σημείων είναι γνωστή ως κατάρα της διαστατικότητας (curse of dimensionality). Πρόκειται για σημαντικό πρόβλημαπου αντιμετωπίζει κανείς σε χώρους υψηλής διάστασης.