ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ (PATTERN RECOGNITION)

Σέργιος Θεοδωρίδης

Κωνσταντίνος Κουτρούμπας

Σχεδιαζόντας ταξινομητές: Τα δεδομένα

Στην πράξη η γνώση σχετικά με τη διαδικασία γέννεσης των δεδομένων είναι πολύ **σπάνια γνωστή**. Το μόνο που έχουμε στη διάθεσή μας είναι ένα σύνολο δεδομένων (data set).

$$Y = \{(x_i, d_i), x_i \in R^l, d_i \in \{1,2,\ldots,m\}, i = 1,\ldots D\}$$

όπου

$\boldsymbol{x}_i$ είναι η *l*-διάστατη αναπαράσταση του *i*-στου από τις *N* οντότητες (διάνυσμα δεδομένων – data vector)

d_i είναι η ετικέτα της κλάσης όπου ανήκει το $\boldsymbol{x}_i$ (1 για την ω_1, 2 για την ω_2,…).

Ο βασικός στόχος ενός ταξινομητή

Δοθέντος ενός ***x*** καταχώρησέ το στην πιο κατάλληλη κατηγορία.

Ανάλογα με τον τρόπο που σχεδιάζεται ένας ταξινομητής, έχουμε δύο κύριες κατηγορίες:

- **Παραμετρικοί ταξινομητές (Parametric classifiers)**
- **Μη παραμετρικοί ταξινομητές (Nonparametric classifiers)**

Σχεδιάζοντας ταξινομητές: Η παραμετρική περίπτωση

Στην περίπτωση αυτή:

- Υποθέτουμε ότι η **μορφή** του παραμετρικού μοντέλου παράγει τα διανύσματα δεδομένων είναι γνωστή, ενώ οι (**άγνωστες**) παράμετροί του εκτιμώνται με βάση το διαθέσιμο σύνολο δεδομένων.
- Η ταξινόμηση ενός νέου διανύσματος δεδομένων ***x*** βασίζεται στο μοντέλο που ορίστηκε παραπάνω.

➢**ΣΗΜ:** Η ταξινόμηση του ***x*** εξαρτάται **έμμεσα** από το διαθέσιμο σύνολο δεδομένων.

Σχεδιάζοντας ταξινομητές: Η μη παραμετρική περίπτωση

Στην περίπτωση αυτή:

➢Η ταξινόμηση του ***x*** εξαρτάται **άμεσα** από το διαθέσιμο σύνολο δεδομένων.

Σχεδιάζοντας ταξινομητές: Εκτιμώντας την απόδοσή τους

Ένα σημαντικό ζήτημα: Πώς γνωρίζουμε το **βαθμό αξιοπιστίας** του ταξινομητή που σχεδιάστηκε;

Είναι σαφές ότι, αν χρησιμοποιήσουμε όλα τα διανύσματα δεδομένων του διαθέσιμου συνόλου δεδομένων *Y*, δεν μπορούμε να χρησιμοποιήσουμε τα ίδια δεδομένα για να αξιολογήσουμε αντικειμενικά την απόδοσή του, καθώς ή γνώση που περιέχεται σ’ αυτά έχει ήδη χρησιμοποιηθεί για το σχεδιασμό του ταξινομητή.

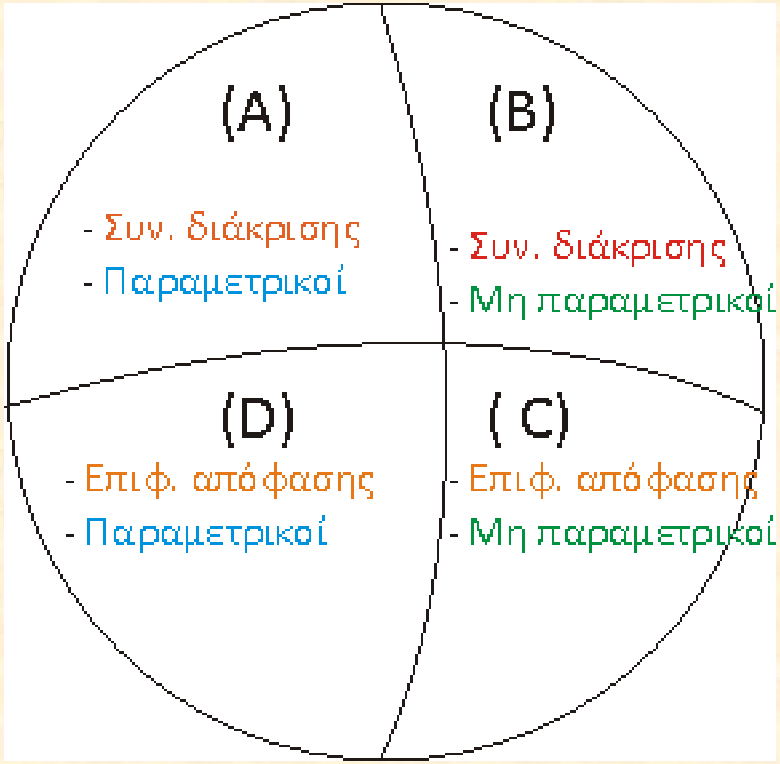
(*) Άλλη σχετική τεχνική είναι το k-fold cross validation.

Μια συνηθισμένη (και απλή) στρατηγική ταξινόμησης(*):

- Διαίρεσε το *Y* σε δύο σύνολα, **X** και **X’,** N και N’ στοιχείων, αντιστοίχως.
- Χρησιμοποίησε το **X** (**σύνολο εκπαίδευσης**) για τη σχεδίαση του ταξινομητή
- Χρησιμοποίησε το **X’** (**σύνολο δοκιμής**) για την αξιολόγηση της απόδοσης του σχεδιασθέντος ταξινομητή
- Αν η απόδοση είναι ικανοποιητική, ο ταξινομητής χρησιμοποιείται στη συνέχεια σε ‘‘πραγματικό περιβάλλον‘‘. Διαφορετικά, θα πρέπει να αναθεωρηθούν κάποια βήματα της διαδικασίας σχεδιασμού.

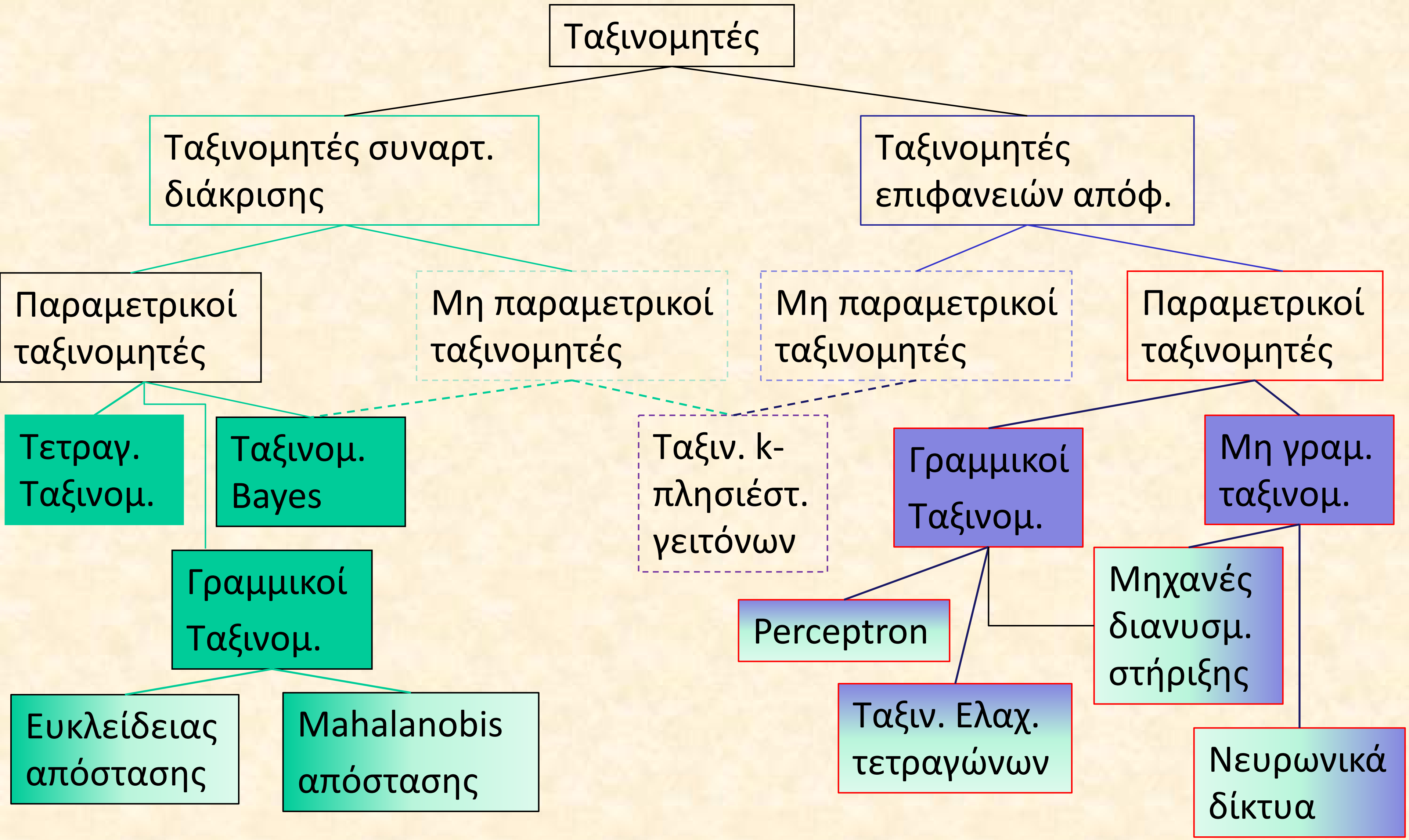
Με βάση ποιο κριτήριο αξιολογείται η απόδοση του ταξινομητή;

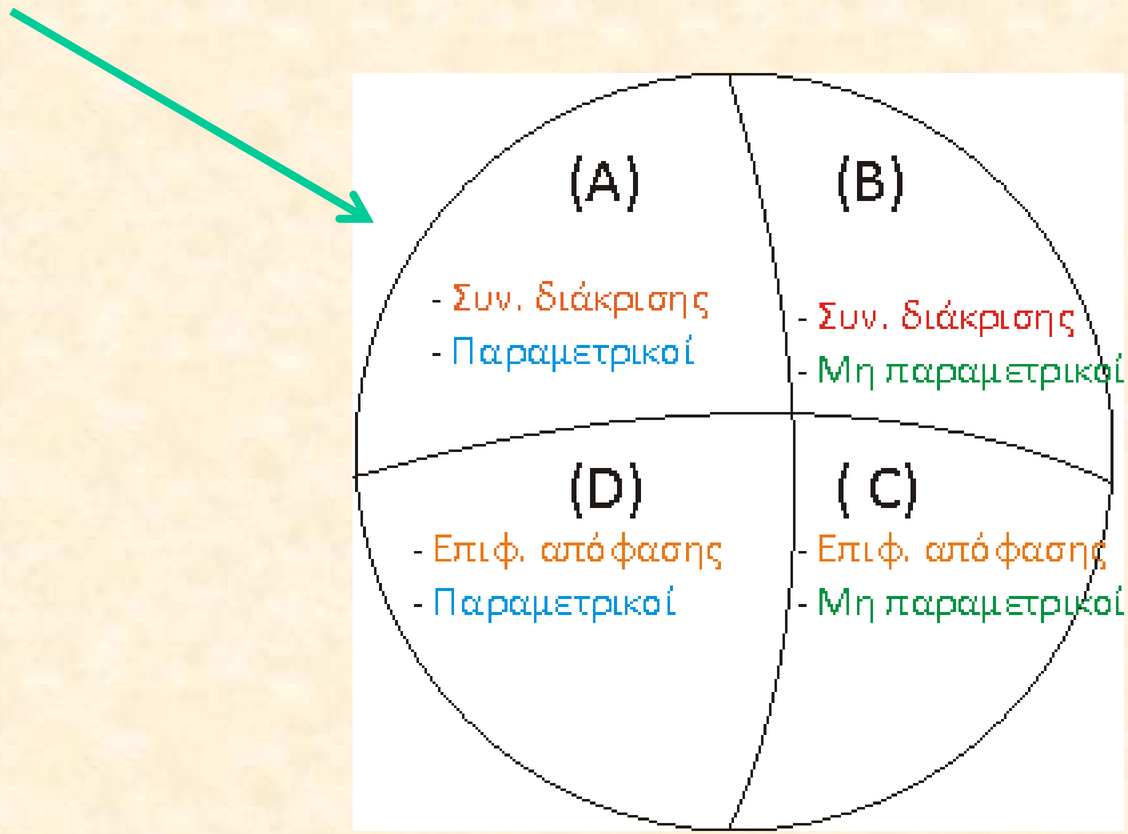
Συνήθως με βάση το ποσοστό των εσφαλμένα ταξινομημένων διανυσμάτων του συνόλου δοκιμής X’ (που είναι μια εκτίμηση της πιθανότητας σφάλματος)



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

“ΧΑΡΤΟΓΡΑΦΗΣΗ” ΤΟΥ ΧΩΡΟΥ ΤΩΝ ΤΑΞΙΝΟΜΗΤΩΝ




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

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Για το **σχεδιασμό** **ΟΛΩΝ των παραμετρικών ταξινομητών**, $y=f(\boldsymbol{x};\boldsymbol{w})$, υπάρχουν δύο κύριες φάσεις
-> Καθορισμός του παραμετρικού μοντέλου
-> Εκτίμηση των παραμέτρων

Ο ταξινομητής Bayes με χρήση παραμετρικών μοντέλων

Ποιες είναι οι παράμετροι που εμπλέκονται στον ταξινομητή Bayes στην περίπτωση αυτή;

- Οι εκ των προτέρων πιθανότητες $P(\omega_j)$, $j=1,\ldots,M$.
- Οι **άγνωστες** παράμετροι $\boldsymbol{\vartheta}_j$ των pdf των κλάσεων $p(\boldsymbol{x}|\omega_j;\boldsymbol{\vartheta}_j)$, j=1,…,$M$.

Παράδειγμα 1: Αν η $p(\boldsymbol{x}|\omega_j)$ **μοντελοποιείται** από μία κανονική κατανομή, $\mathrm{N}(\mu_j,\Sigma_j)$, και οι άγνωστες παράμετροι είναι το μέσο διάνυσμα και το μητρώο συνδιασποράς τότε $\boldsymbol{\vartheta}_j=\{\boldsymbol{\mu}_j, \Sigma_j\}$

Παράδειγμα 2: Αν η $p(\boldsymbol{x}|\omega_j)$ **μοντελοποιείται** από μία κανονική κατανομή, $\mathrm{N}(\mu_j,\Sigma_j)$, και η άγνωστη παράμετρος είναι το μέσο διάνυσμα, τότε $\boldsymbol{\vartheta}_j=\boldsymbol{\mu}_j$.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Στην πράξη

- οι τιμές των παραμέτρων είναι πολύ σπάνια γνωστές.
- Έτσι, χρειάζεται να τις **εκτιμήσουμε**.

Στην πράξη έχουμε στην διάθεσή μας ένα σύνολο δεδομένων (σύνολο 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 το *i*-στό διάνυσμα εκπαίδευσης (training vector)

d_i είναι η ετικέτα της κλάσης όπου ανήκει το $\boldsymbol{x}_i$ (1 για την ω_1, 2 για την ω_2,…).

(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Έστω

$$X_j = \{x_i \in R^l : d_i = j, i = 1, \ldots N_j\}, \qquad j = 1, \ldots, M$$

$$\boxed{N_1 + \ldots + N_M = N}$$

Το σύνολο που περιέχει τα διανύσματα δεδομένων της *j*-στης κλάσης.

<u>Εκτίμηση παραμέτρων</u> (βασίζεται στο σύνολο X)**:**

- Για κάθε $P(\omega_j)$:
  Χρησιμοποίησε την απλή προσέγγιση $P(\omega_j) \approx N_j / N$.
- Για κάθε $p(\boldsymbol{x}|\omega_j)$:
- Υπάρχουν αρκετές μέθοδοι
  - **Εκτίμηση μέγιστης πιθανοφάνειας** (**Maximum Likelihood** (ML) estimation),
  - **Εκτίμησης μέγιστης εκ των υστέρων πιθανότητας** (**Maximum a posteriori** (MAP) estimation),
  - Εκτίμηση μέγιστης εντροπίας (Maximum entropy (ME) estimation),
  - **Expectation-Maximization** (EM) estimation (mixture models) etc

Στη συνέχεια περιγράφουμε τις μεθόδους ML, MAP και (σύντομα) την μέθοδο EM.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

- Έστω $Y=\{ \boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N\}$ ένα σύνολο από ανεξάρτητα διανύσματα δεδομένων
- Έστω $p(\boldsymbol{x})$ μια pdf γνωστής παραμετρικής μορφής αλλά με άγνωστο το διάνυσμα παραμέτρων του $\boldsymbol{\vartheta}$; $p(\boldsymbol{x})=p(\boldsymbol{x};\boldsymbol{\vartheta})$.

  Παραδείγματα:

  Αν η $p(\boldsymbol{x};\boldsymbol{\vartheta})$ είναι κανονική με άγνωστο μέσο διάνυσμα $\boldsymbol{\mu}$, το $\boldsymbol{\vartheta}$ είναι απλά το $\boldsymbol{\mu}$.

  Αν η $p(\boldsymbol{x};\boldsymbol{\vartheta})$ είναι κανονική με άγνωστο μέσο διάνυσμα $\boldsymbol{\mu}$ και άγνωστο μητρώο συνδιασποράς Σ, το $\boldsymbol{\vartheta}$ περιέχει τις συνιστώσες τόσο του $\boldsymbol{\mu}$ όσο και του $\boldsymbol{\Sigma}$.

Το πρόβλημα: Εκτίμησε το $\boldsymbol{\vartheta}$ ώστε η $p(\boldsymbol{x};\boldsymbol{\vartheta})$ να είναι το **καλύτερο δυνατό ταίριασμα** για το σύνολο δεδομένων Y.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

- Έστω η **συνάρτηση πιθανοφάνειας (likelihood function)** του ***ϑ*** ως προς το *Y*:

$$p(Y;\vartheta) = p(x_1,\ldots,x_N;\vartheta) = \prod_{\mathrm{i}=1}^{\mathrm{N}} p(x_i;\vartheta)$$

- ***Στην πράξη***, είναι πιο βολικό να μεγιστοποιήσουμε την συνάρτηση του **λογάριθμου της πιθανοφάνειας** (**log-likelihood function) *L*(*ϑ*)** του ***ϑ*** ως προς το *Y*, η οποία ορίζεται ως

$$L(\vartheta) = \ln p(Y;\vartheta) = \ln\, p(x_1,\ldots,x_N;\vartheta) = \sum_{i=1}^{N} \ln p(x_i;\vartheta)$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

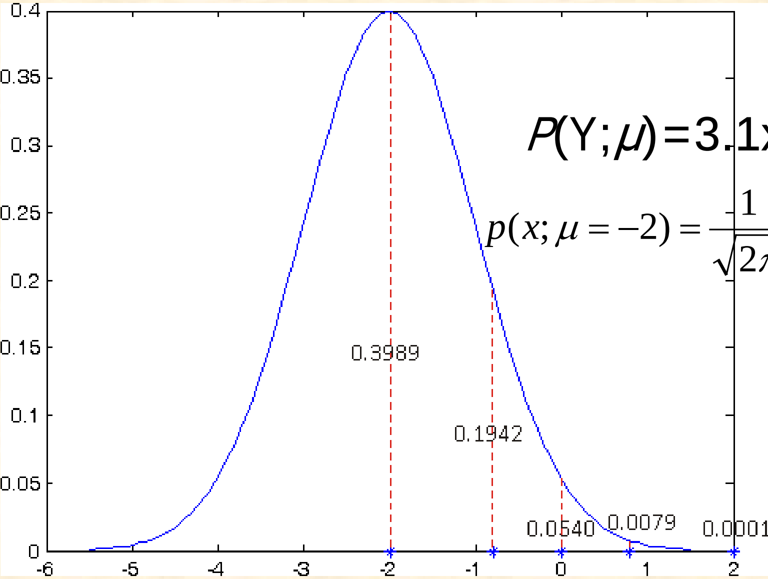
Παράδειγμα:

- *Y*={-2, -1, 0, 1, 2}
- Θεωρείστε το σύνολο των κανονικών κατανομών μοναδιαίας διασποράς, παραμετροποιημένων ως προς το *μ*.
- Η συνάρτηση πιθανοφάνειας του *μ* ως προς το *Y* είναι

$$p(Y;\mu) = p(-2,\text{-}1,0,1,2;\mu) =$$

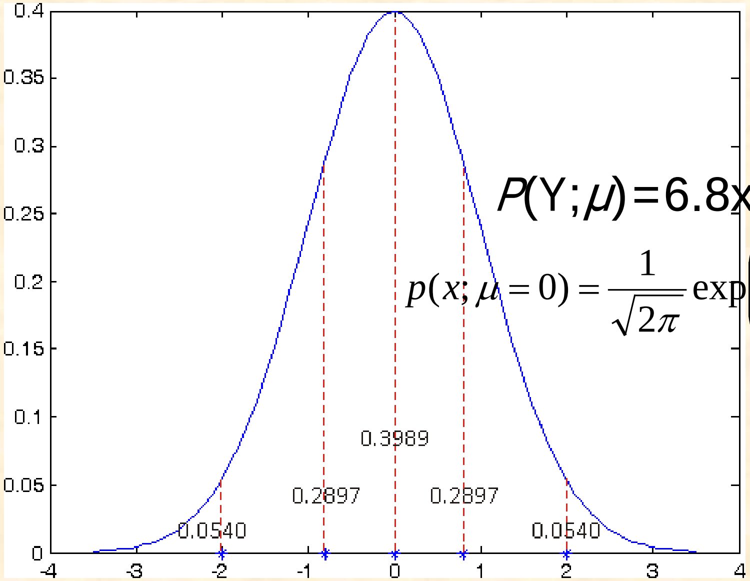
$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(-2-\mu)^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(-1-\mu)^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(0-\mu)^2}{2}\right)$$

$$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(1-\mu)^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(2-\mu)^2}{2}\right)$$



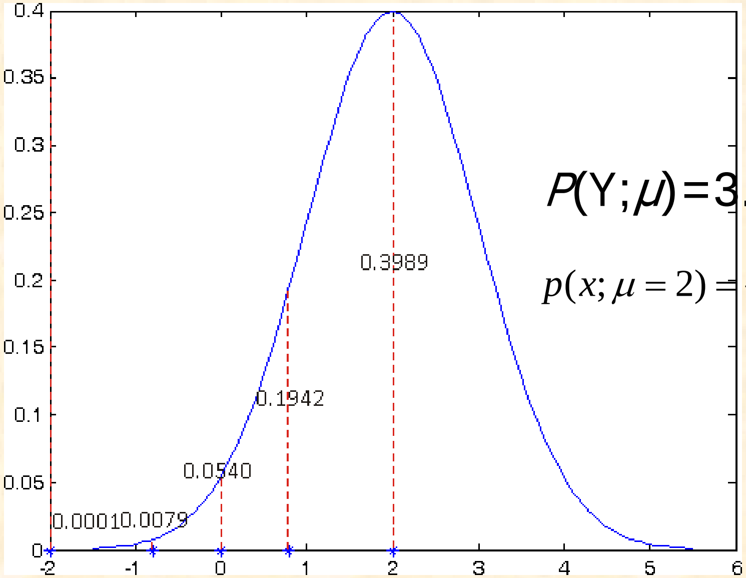
$P(Y;\mu) = 3.1\text{x}10^{-9}$

$$p(x;\mu=-2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x+2)^2}{2}\right)$$



$P(Y;\mu) = 6.8\text{x}10^{-5}$

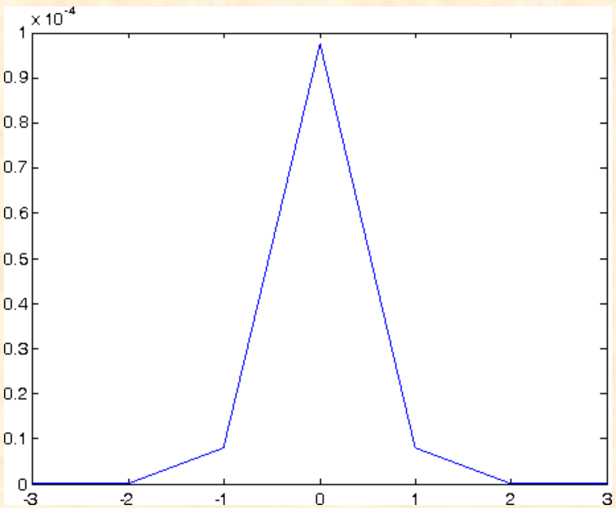
$$p(x;\mu=0) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$



$P(Y;\mu) = 3.1\text{x}10^{-9}$

$$p(x;\mu=2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-2)^2}{2}\right)$$

$P(Y;\mu)$

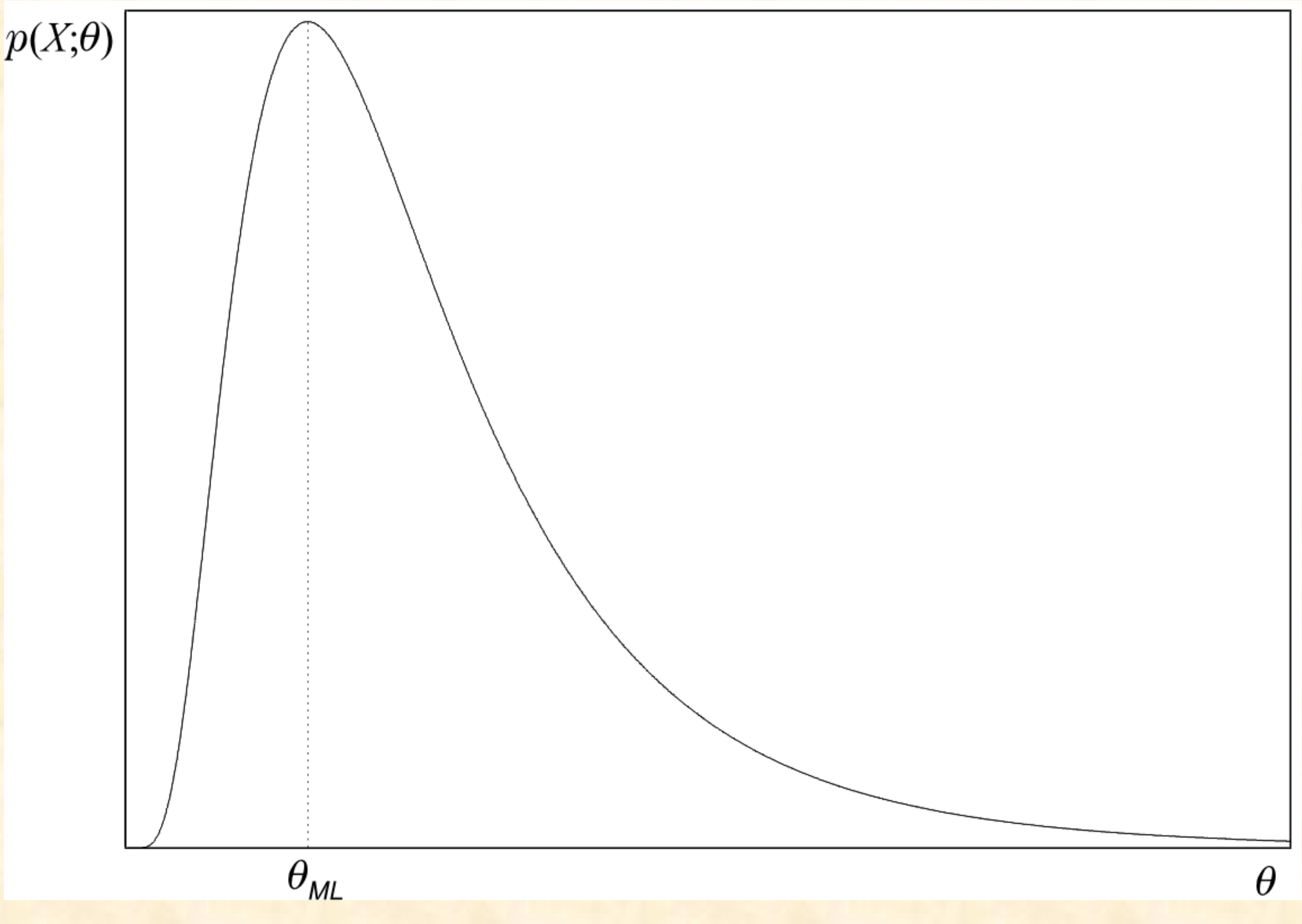


μ

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)



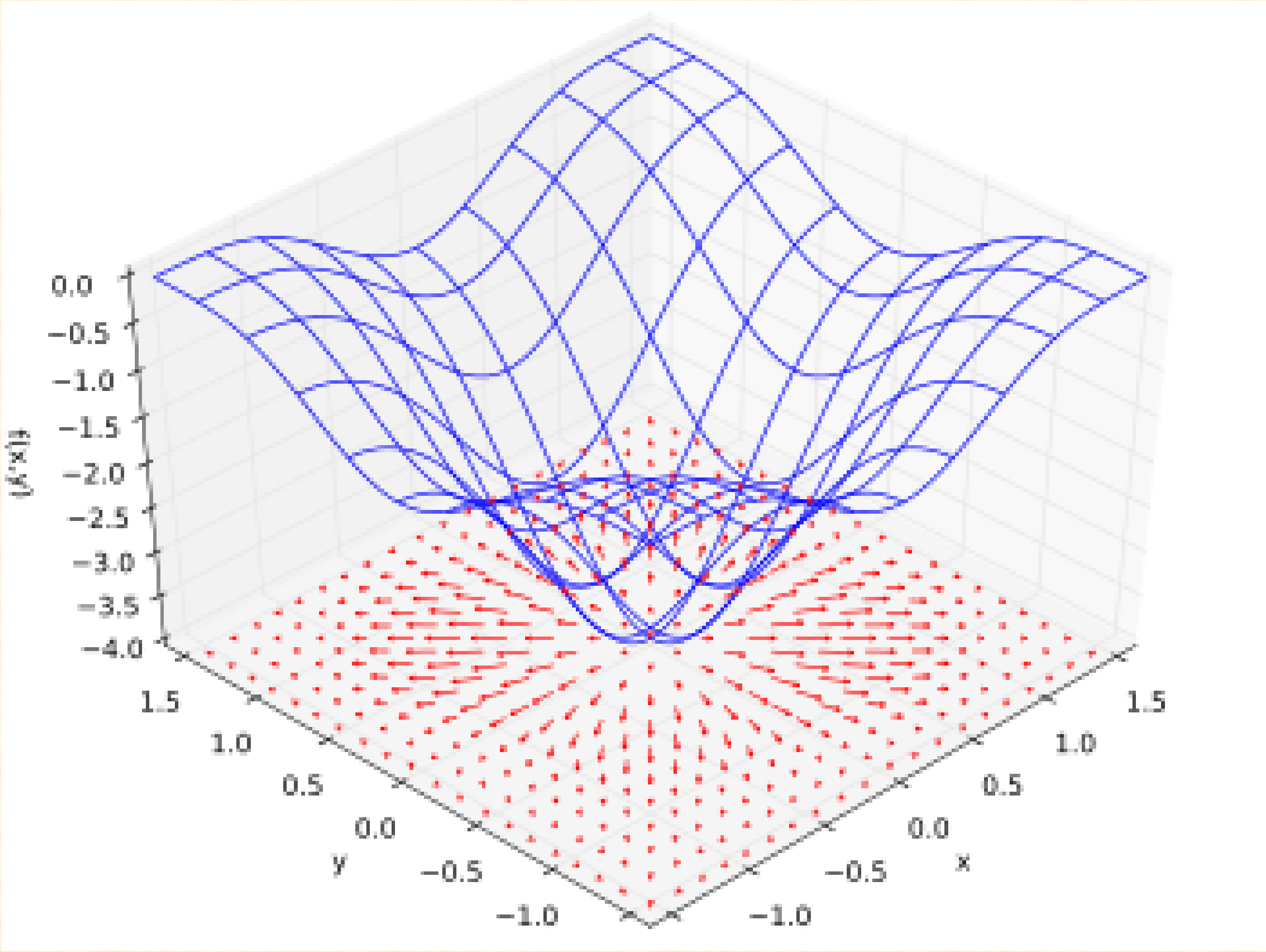
(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας: **Μεγιστοποίησε** την $p(Y;\boldsymbol{\vartheta})$ (ή την $L(\boldsymbol{\vartheta})$) ως προς το $\boldsymbol{\vartheta}$.

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

Μια παρένθεση: Το **gradient** συνάρτησης



Π.χ.: Το gradient της συνάρτησης

$$f(x_1,x_2) = -(\cos^2 x_1 + \cos^2 x_2)^2$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML) - Ένα παράδειγμα

-Έστω *Y* ένα σύνολο *N* διανυσμάτων δεδομένων, $\boldsymbol{x}_i$, i=1,…,*N*.

-Ποια είναι η κανονική κατανομή $p(\boldsymbol{x};\boldsymbol{\vartheta})$ γνωστού μητρώου συνδιασποράς που μοντελοποιεί καλύτερα τα διανύσματα του συνόλου *Y*?

Λύση:

-Το άγνωστο διάνυσμα παραμέτρων στην περίπτωση αυτή είναι το μέσο διάνυσμα $\boldsymbol{\mu}$, δηλ. $\boldsymbol{\vartheta}=\boldsymbol{\mu}$.

-Είναι

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

$$\ln p(x;\mu) = \ln\left(\frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}}\right) - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) = C - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)$$

$$L(\mu) = \sum_{i=1}^{N} \ln p(x_i;\mu) = NC - \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML) - Ένα παράδειγμα

-Θέτοντας την παράγωγο της *L*(**μ**) ως προς το **μ** ίση με **0** έχουμε

$$\frac{\partial L(\mu)}{\partial \mu} = \frac{\partial}{\partial \mu}\left(NC - \sum_{i=1}^{N} \frac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu) \right) = 0$$

$$\sum_{i=1}^{N} \Sigma^{-1}(x_i - \mu) = 0 \Leftrightarrow \sum_{i=1}^{N}(x_i - \mu) = 0 \Leftrightarrow \sum_{i=1}^{N} x_i - N\mu = 0$$

$$\mu_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

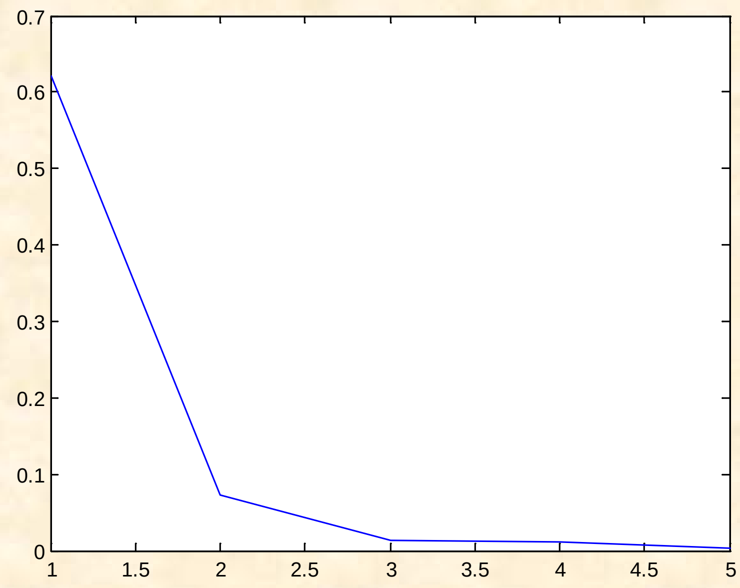
- Έστω N=10 σημεία που παρήχθησαν από τη μονοδιάστατη κανονική κατανομή μηδενικής μέσης τιμής και μοναδιαίας συνδασποράς, $N(0,1)$.

- Ας **προσποιηθούμε** ότι έχουμε στη διάθεσή μας μόνο τα ακόλουθα:
    - Τη γνώση ότι η κατανομή που παρήγαγε τα σημεία είναι κανονική με διασπορά ίση με 1 και άγνωστο μέσο διάνυσμα μ.
    - τα 10 σημεία.

➢Η **ML εκτίμηση** του μέσου διανύσματος, $\hat{\mu}$, της κατανομής είναι 0.6210 (η πραγματική τιμή είναι 0).

Όσο περισσότερα σημεία είναι διαθέσιμα, τόσο πιο ακριβείς εκτιμήσεις θα έχουμε για το μέσο διάνυσμα.

N	Σφάλμα εκτίμησης
10	0.6210 – 0
100	0.0727 – 0
1000	0.0138 – 0
10000	0.0118 – 0
100000	0.0034 – 0



(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

Αν πράγματι υπάρχει $\boldsymbol{\vartheta}_0$ έτσι ώστε $p(\boldsymbol{x})=p(\boldsymbol{x};\boldsymbol{\vartheta}_0)$ τότε

$$\lim_{N\to\infty} E[\vartheta_{ML}] = \vartheta_0$$

$$\lim_{N\to\infty} E\|\vartheta_{ML} - \vartheta_0\|^2 = 0$$

- Ασυμπτωτικά αμερόληπτος (Asymptotically unbiased) εκτιμητής:

Η μέση τιμή του $\boldsymbol{\vartheta}$, $\boldsymbol{\vartheta}_{ML}$, **τείνει προς την πραγματική τιμή** της άγνωστης παραμέτρου ($\boldsymbol{\vartheta}_0$), καθώς $N\to\infty$.

- Ασυμπτωτικά συνεπής (Asymptotically consistent) εκτιμητής:

Η διασπορά της ML εκτίμησης τείνει στο 0, καθώς $N\to\infty$.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

Άσκηση: Η τυχαία μεταβλητή *x* ακολουθεί την κατανομή Erlang

$$p(x;\vartheta) = \vartheta^2 x \exp(-\vartheta x) u(x)$$

με $u(x)=1$ αν $x>0$ και 0, διαφορετικά.

(α) Δεδομένου ενός συνόλου δειγμάτων $x_1,\ldots,x_N$, του x, να δείξετε ότι η εκτίμηση ML του ϑ είναι

$$\vartheta_{ML} = \frac{2N}{\sum_{k=1}^{N} x_k}$$

(β) Δεδομένου ότι η **μέση τιμή** της κατανομής ισούται με **2/ϑ** ποια είναι η ML εκτίμησή της με βάση το παραπάνω σύνολο δειγμάτων;

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

Άσκηση: Η τυχαία μεταβλητή *x* ακολουθεί την κατανομή Rayleigh

$$p(x;\vartheta) = 2\vartheta x \exp(-\vartheta x^2) u(x)$$

με $u(x)=1$ αν $x>0$ και 0, διαφορετικά.

(α) Δεδομένου ενός συνόλου δειγμάτων $x_1,\ldots,x_N$, του *x*, να δείξετε ότι η εκτίμηση ML του ϑ είναι

$$\vartheta_{ML} = \frac{N}{\sum_{k=1}^{N} x_k^{\,2}}$$

(β) Δεδομένου ότι η **μέση τιμή** της κατανομής ισούται με $\sqrt{\pi/4\vartheta}$ ποια είναι η ML εκτίμησή της με βάση το παραπάνω σύνολο δειγμάτων;

(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας (ML)

Άσκηση: Θεωρείστε ένα πρόβλημα ταξινόμησης δύο κλάσεων, ω_1 και ω_2, όπου οι εμπλεκόμενες οντότητες χαρακτηρίζονται από ένα μόνο χαρακτηριστικό. Είναι γνωστό ότι οι τιμές

0.90, 0.95, 1.02, 1.01, 0.99, 1.30, 1.25, 0.87, 0.93, 1.02

ανήκουν στην κατηγορία ω_1, ενώ οι τιμές

1.90, 1.95, 2.02, 2.01, 1.99, 2.30, 2.25, 1.87, 1.93, 2.02, 2.03, 1.91, 2.15, 1.88, 2.07

ανήκουν στην κατηγορία ω_2.

Μοντελοποιώντας με κανονικές κατανομές τις pdfs των δύο κλάσεων για τις οποίες είναι γνωστές οι διασπορές (σ_1^2=0.020 για την ω_1 και σ_2^2=0.017 για την ω_2) αλλά άγνωστες οι μέσες τιμές τους:

(α) Εκτιμήστε τις μέσες τιμές των κατανομών

(β) Διατυπώστε τον (προσεγγιστικό) κανόνα του Bayes για το πρόβλημα.

(γ) Ταξινομείστε τα σημεία 1.4, 1.7, 0.2, 2.5 με βάση τον παραπάνω κανόνα.