ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ (PATTERN RECOGNITION)

Σέργιος Θεοδωρίδης

Κωνσταντίνος Κουτρούμπας

Version 2

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας (Maximum Aposteriori Probability (MAP) Estimation)

- Στην ML μέθοδο, το $\underline{\vartheta}$ λογιζόταν ως παράμετρος
- Εδώ θα θεωρήσουμε το $\underline{\vartheta}$ ως τυχαίο διάνυσμα που περιγράφεται από (υποτίθεται γνωστή) pdf $p(\underline{\vartheta})$.
- Δοθέντος

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

Υπολόγισε το μέγιστο της

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

- Από το Θεώρημα του Bayes έχουμε

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

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

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας (MAP)

- Η μέθοδος:

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

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

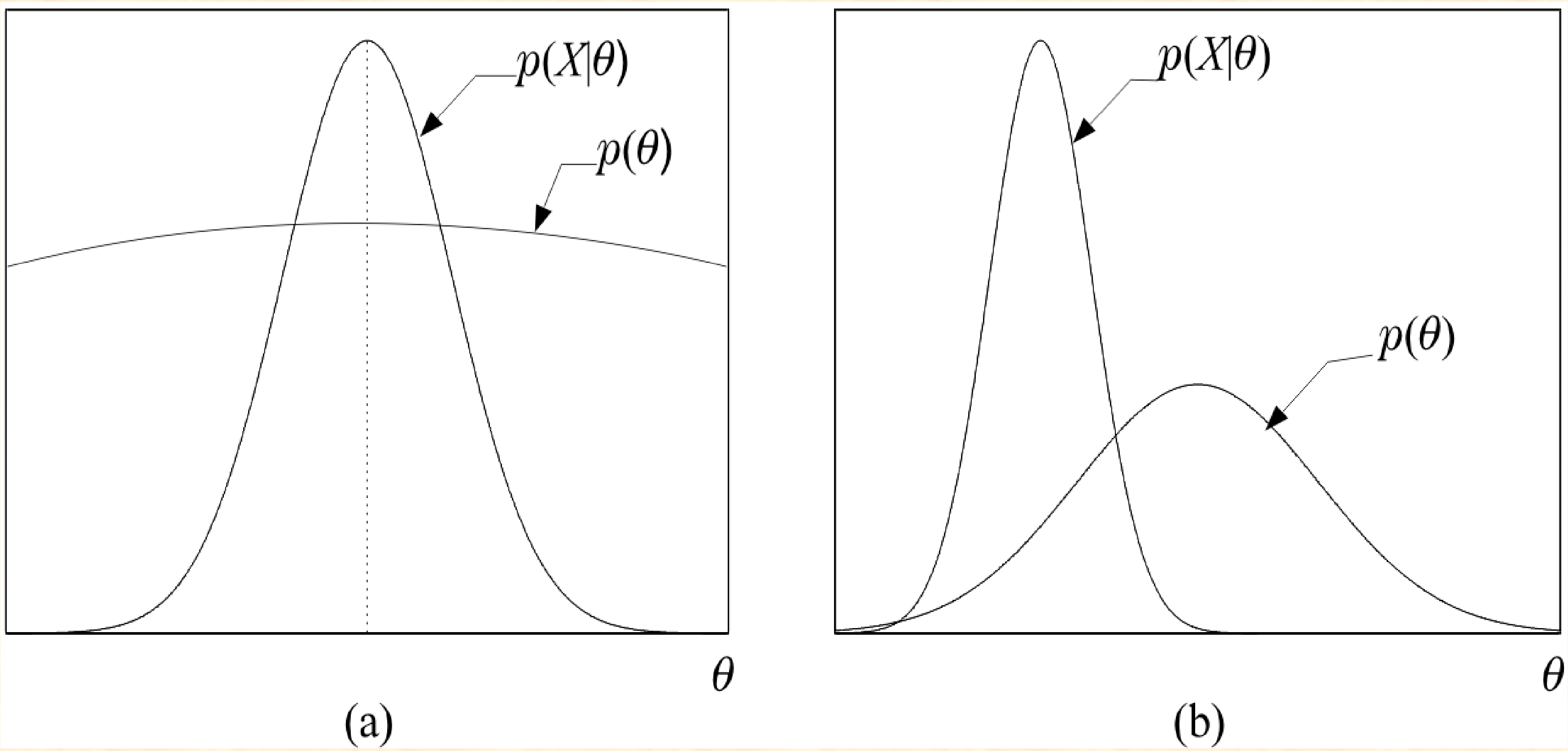
Αν η $p(\underline{\theta})$ είναι ομοιόμορφη ή αρκετά ευρεία:

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

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

<u>Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας (MAP)</u>



(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Εκτίμηση μέγιστης εκ των υστέρων πιθανότητας (MAP) - Παράδειγμα

$$p(\underline{x}|\underline{\mu}) \sim N(\underline{\mu}, \Sigma), \quad \underline{\mu} \text{ άγνωστο}, \quad X = \{\underline{x}_1, \ldots, \underline{x}_N\} \quad \Sigma = \sigma^2 I$$

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

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

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

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

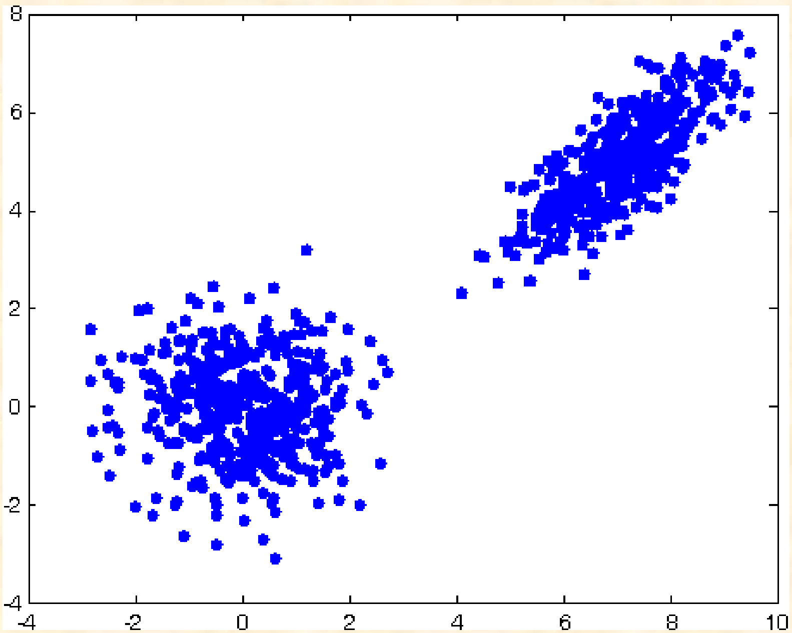
(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος αναμονής-μεγιστοποίησης (expectation-maximization – EM) (μοντέλα μίξης - mixture models)

Μοντέλο μίξης: Σταθμισμένο άθροισμα pdfs γνωστής παραμετρικής μορφής

$$p(x) = \sum_{k=1}^{K} P_k\, p(x\,|\,k), \qquad \sum_{k=1}^{K} P_k = 1, \qquad \int_{-\infty}^{+\infty} p(x\,|\,k) = 1$$



❖Η μέθοδος ML δεν μπορεί να αξιοποιηθεί εδώ εξαιτίας των *ετικετών k*, που είναι επίσης άγνωστες. **Λύση:** Ο αλγόριθμος **EM**.

Συμβολισμοί:

➢$p(\boldsymbol{x}|k) = p(\boldsymbol{x}|k;\boldsymbol{\vartheta}_k)$

➢$\boldsymbol{\Theta}=(\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_\kappa)$, $\boldsymbol{P}=[P_1,\ldots,P_k]^T$, $\boldsymbol{\Xi}=(\boldsymbol{\Theta},\boldsymbol{P})$.

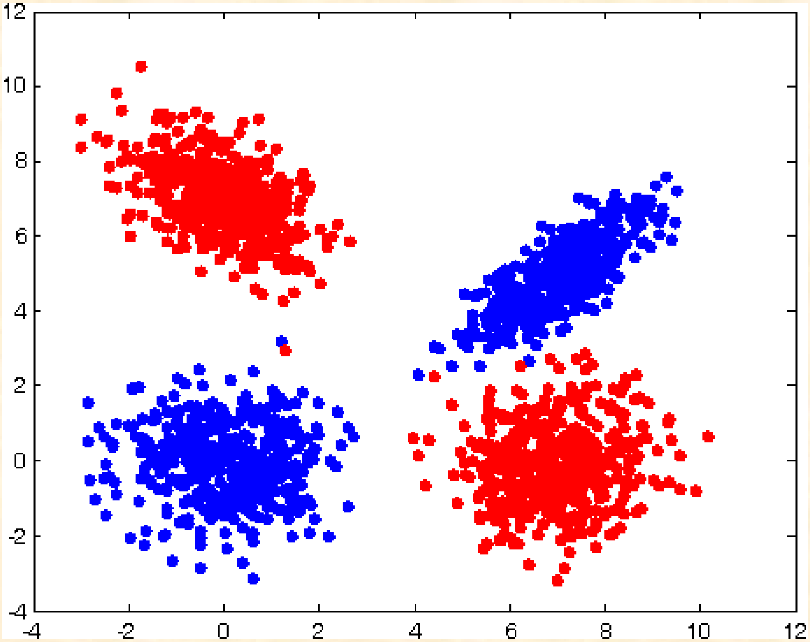
➢$X=\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\}$: μη πλήρες (incomplete) (παρατηρούμενο) σύνολο δεδομένων.

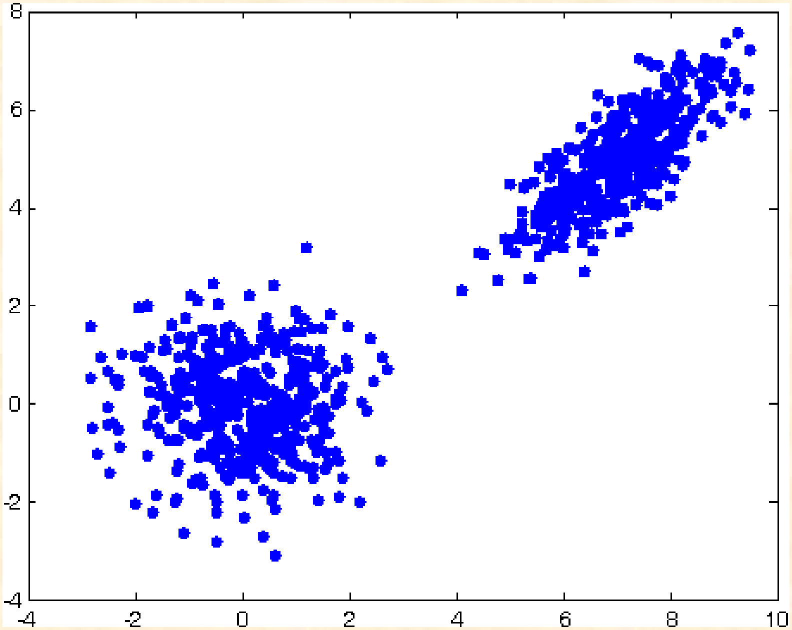
➢$X^c=\{(\boldsymbol{x}_1,k_1),\ldots,(\boldsymbol{x}_N,k_N)\}$: πλήρες (complete) σύνολο δεδομένων

(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος EM (μοντέλα μίξης)





(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Η μέθοδος ΕΜ (μοντέλα μίξης)

Στόχος: Εκτίμησε τα ***Θ*** και ***P*** μέσω της μεγιστοποίησης της αναμενόμενης τιμής (expectation) της συνάρτησης log-likelihood **conditioned** στα παρατηρούμενα δεδομένα.

$$\ln p(X;\Theta,P)=\sum_{n=1}^{N}\sum_{j=1}^{m}P(j\mid \boldsymbol{x}_n;\boldsymbol{\theta}_j)\ln p(\boldsymbol{x}_n, j;\boldsymbol{\theta}_j)=$$

$$\sum_{n=1}^{N}\sum_{j=1}^{m}P(j\mid \boldsymbol{x}_n;\boldsymbol{\theta}_j)\ln\left(p(\boldsymbol{x}_n\mid j;\boldsymbol{\theta}_j)P_j\right)$$

Μίξη κανονικών κατανομών.

$$p(\boldsymbol{x})=\sum_{j=1}^{m}P_j\, p(\boldsymbol{x}\mid j;\boldsymbol{\mu}_j,\Sigma_j),$$

$$p(\boldsymbol{x}\mid j;\boldsymbol{\mu}_j,\Sigma_j)=\frac{1}{(2\pi)^{l/2}\,|\Sigma_j|^{1/2}}\exp\left(-0.5\cdot(\boldsymbol{x}-\boldsymbol{\mu}_j)^T\Sigma_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)\right)$$

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η περίπτωση του ταξινομητή Bayes

Ο αλγόριθμος **EM** για **μίξεις κανονικών κατανομών**.

- Αρχικοποιήσε $\boldsymbol{\mu}_j=\boldsymbol{\mu}_j^{(0)}$, $\Sigma_j=\Sigma_j^{(0)}$, $\boldsymbol{P}=\boldsymbol{P}^{(0)}$
- t=0
- Επανέλαβε

Expectation step

$$P\left(j \mid \boldsymbol{x}_n;\Theta^{(t)},P^{(t)}\right)=\frac{p\left(\boldsymbol{x}_n \mid j;\boldsymbol{\theta}_k^{(t)}\right)P_j^{(t)}}{\sum_{q=1}^{m} p\left(\boldsymbol{x}_n \mid q;\boldsymbol{\theta}_q^{(t)}\right)P_q^{(t)}}\equiv\gamma_{jn}^{(t)}$$

Maximization step

- $t=t+1$

$$\boldsymbol{\mu}_j^{(t+1)}=\sum_{n=1}^{N}\gamma_{jn}^{(t)}\boldsymbol{x}_n \Big/ \sum_{n=1}^{N}\gamma_{jn}^{(t)}$$

$$\Sigma_j^{(t+1)}=\sum_{n=1}^{N}\gamma_{jn}^{(t)}(\boldsymbol{x}_n-\boldsymbol{\mu}_j^{(t+1)})(\boldsymbol{x}_n-\boldsymbol{\mu}_j^{(t+1)})^T \Big/ \sum_{n=1}^{N}\gamma_{jn}^{(t)}$$

$$P_j^{(t+1)}=\frac{1}{N}\sum_{n=1}^{N}\gamma_{jn}^{(t)}$$

- Έως ότου ικανοποιηθεί ένα κατάλληλο κριτήριο τερματισμού.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Ανασκόπηση του (προσεγγιστικού) ταξινομητή Bayes

- Available data: $X = X_1 \cup X_2 \cup \ldots X_M$. Each X_j corresponds to class ω_j.

Σχεδιασμός ταξινομητή

- **Υιοθέτησε** ένα μοντέλο pdf για κάθε ω_j, με άγνωστες παραμέτρους $\boldsymbol{\vartheta}_j$.
- **Εφάρμοσε** την ML (ή τις MAP, EM) μεθόδους ***M*** **φορές** (μία για κάθε κλάση) για την εκτίμηση των $\boldsymbol{\vartheta}_j$'s, με βάση τα αντίστοιχα σύνολα X_j,

$$\hat{p}(x \mid \omega_j) \equiv p(x \mid \omega_j; \hat{\vartheta}_j)$$

-**Προσέγγισε** τις $P(\omega_j)$ ως εξής

$$\hat{P}(\omega_j) = N_i / N$$

-**Όρισε**

$$g_j(x) = \hat{P}(\omega_j)\hat{p}(x \mid \omega_j)$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, $j=1,\ldots,M$.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x}) = \max_{j=1,\ldots,M} g_j(\boldsymbol{x})$.

Ένα σημαντικό ζήτημα – Η κατάρα της διαστατικότητας (Curse of dimensionality)

- Σε όλες τις μεθόδους που εξετάσαμε μέχρι στιγμής είδαμε ότι όσο μεγαλύτερος είναι ο αριθμός των σημείων, N, τόσο καλύτερες είναι και οι προκύπτουσες εκτιμήσεις.

- Αν στο μονοδιάστατο χώρο για ένα διάστημα μήκους L, N σημεία είναι αρκετά για να δώσουν μια καλή εκτίμηση των εμπλεκόμενων παραμέτρων, στο διδιάστατο χώρο για μια επιφάνεια διαστάσεων LxL απαιτούνται N^2 σημεία για να πάρουμε καλές εκτιμήσεις και, γενικά, στον l-διάστατο χώρο για έναν υπερκύβο διαστάσεων L^l απαιτούνται Nl σημέια.

- Η εκθετική αύξηση του αριθμού των σημείων που απαιτούνται για μια καλή εκτίμηση των εμπλεκόμενων παραμέτρων με τη διάσταση του χώρου είναι γνωστή ως κατάρα της διαστατικότητας (curse of dimensionality). Πρόκειται για ένα σημαντικό πρόβλημα που απαντάται σε προβλήματα που απεικονίζονται σε χώρους υψηλής διαστατικότητας.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Ο απλοϊκός ταξινομητής Bayes (naive Bayes classifier)

- Διαθέσιμα δεδομένα: $X=X_1 \cup X_2 \cup \ldots X_M$. Κάθε X_j αντιστοιχεί στην κλάση ω_j. Έστω $X_j=X_{j1} \times \ldots \times X_{jl}$

-**Υπόθεση:** Όλα τα **χαρακτηριστικά (features)** είναι μεταξύ τους **στατιστικώς ανεξάρτητα**. Έτσι

$$p(x \mid \omega_j) = \prod_{k=1}^{l} p(x_k \mid \omega_j), \qquad j = 1,\ldots M$$

-Ως εκ τούτου, μπορούμε για κάθε κλάση να εργαστούμε για κάθε $p(\boldsymbol{x}_k|\omega_j)$ ανεξάρτητα από τις υπόλοιπες $p(\boldsymbol{x}_q|\omega_j)$, q=1,…,l, q ≠k.

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Ο απλοϊκός ταξινομητής Bayes (naive Bayes classifier)

Σχεδιασμός ταξινομητή

- **Υιοθέτησε** ένα μοντέλο pdf για κάθε ω_j, με άγνωστες παραμέτρους $\boldsymbol{\vartheta}_j$.
- **Εφάρμοσε** την ML μέθοδο ***M l* φορές** (***l*** φορές για κάθε κλάση) για την εκτίμηση των $\boldsymbol{\vartheta}_{jk}$'s των μονοδιάστατων $p(\boldsymbol{x}_k|\omega_j)$, με βάση τα αντίστοιχα X_{jk}

-**Προσέγγισε** τις $P(\omega_j)$ ως ακολούθως

$$\hat{P}(\omega_j) = N_i / N$$

-**Όρισε**

$$g_j(x) = \hat{P}(\omega_j)\hat{p}(x \mid \omega_j) = \hat{P}(\omega_j)\prod_{k=1}^{l} \hat{p}(x_k \mid \omega_j)$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, j=1,…,M.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x})=\max_{j=1,\ldots,M} g_j(\boldsymbol{x})$.

(A) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Γραμμικοί ταξινομητές – Ο ταξινομητής ελάχιστης Ευκλείδειας απόστασης

- Διαθέσιμα δεδομένα: $X=X_1 \cup X_2 \cup \ldots X_M$. Κάθε X_j αντιστοιχεί στην κλάση ω_j.

Σχεδιασμός ταξινομητή

- **Εκτίμησε** το **μέσο διάνυσμα** , m_j, κάθε κλάσης ω_j, με βάση το X_j , j=1,…,N_j.

(συνήθως, η μέθοδος ML χρησιμοποιείται για κάθε κλάση, με την **υπόθεση** ότι κάθε κλάση μοντελοποιείται από μια κανονική κατανομή **αγνώστου μέσου διανύσματος** και το m_j τίθεται ίσο με την προκύπτουσα εκτίμηση του μέσου διανύσματος)

-**Όρισε**

$$g_j(\boldsymbol{x})=-||\boldsymbol{x}-\boldsymbol{m}_j||^2$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, j=1,…,M.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x})=\max_{j=1,\ldots,M} g_j(\boldsymbol{x})$.

ΣΗΜ: Γενικά, ο ταξινομητής ελάχιστης Ευκλείδειας απόστασης δεν είναι βέλτιστος ως προς το κριτήριο της πιθανότητας λάθους (κάτω από ποιές συνθήκες ο ταξινομητής αυτός θα μπορούσε να είναι βέλτιστος;)

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Γραμμικοί ταξινομητές – Ο ταξινομητής ελάχιστης Mahalanobis απόστασης

- Διαθέσιμα δεδομένα: $X=X_1 \cup X_2 \cup \ldots X_M$. Κάθε X_j αντιστοιχεί στην κλάση ω_j.

Σχεδιασμός ταξινομητή

- **Εκτίμησε** το **μέσο διάνυσμα** , m_j, και το μητρώο συνδιασποράς, S_j, κάθε κλάσης ω_j, με βάση το X_j , j=1,…,N_j.

(συνήθως, η μέθοδος ML χρησιμοποιείται για κάθε κλάση, με την **υπόθεση** ότι κάθε κλάση μοντελοποιείται από μια κανονική κατανομή **αγνώστου μέσου διανύσματος** και **μητρώου συνδιασποράς** και τα m_j και S_j, τίθενται ίσα με τις προκύπτουσες εκτιμήσεις (το κοινό μητρώο συνδιασποράς μπορεί να τεθεί ίσο με το μέσο των εκτιμήσεων για τα S_j)

-**Όρισε**

$$g_j(\boldsymbol{x})=-(\boldsymbol{x}-\boldsymbol{m}_j)^T S^{-1}(\boldsymbol{x}-\boldsymbol{m}_j)$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, j=1,…,M.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x})=\max_{j=1,\ldots,M} g_j(\boldsymbol{x})$.

ΣΗΜ: Γενικά, ο ταξινομητής ελάχιστης Mahalanobis απόστασης δεν είναι βέλτιστος ως προς το κριτήριο της πιθανότητας λάθους (κάτω από ποιές συνθήκες ο ταξινομητής αυτός θα μπορούσε να είναι βέλτιστος;)

(Α) ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Τετραγωνικοί ταξινομητές

- Διαθέσιμα δεδομένα: $X=X_1 \cup X_2 \cup \ldots X_M$. Κάθε X_j αντιστοιχεί στην κλάση ω_j.

Σχεδιασμός ταξινομητή

- **Εκτίμησε** το **μέσο διάνυσμα** , m_j, και το μητρώο συνδιασποράς, S_j, κάθε κλάσης ω_j, με βάση το X_j , j=1,…,N_j.

(συνήθως, η μέθοδος ML χρησιμοποιείται για κάθε κλάση, με την **υπόθεση** ότι κάθε κλάση μοντελοποιείται από μια κανονική κατανομή **αγνώστου μέσου διανύσματος** και **μητρώου συνδιασποράς** και τα m_j και S_j, τίθενται ίσα με τις προκύπτουσες εκτιμήσεις

-**Όρισε**

$$g_j(\boldsymbol{x})=-(\boldsymbol{x}-\boldsymbol{m}_j)^T S_j^{-1}(\boldsymbol{x}-\boldsymbol{m}_j)$$

Χρήση ταξινομητή

-Για δεδομένο $\boldsymbol{x}$,

-Υπολόγισε τις τιμές των $g_j(\boldsymbol{x})$, j=1,…,M.

-Καταχώρησε το $\boldsymbol{x}$ στην κλάση k με $g_k(\boldsymbol{x})=\max_{j=1,\ldots,M} g_j(\boldsymbol{x})$.

ΣΗΜ: Γενικά, ο τετραγωνικός ταξινομητής δεν είναι βέλτιστος ως προς το κριτήριο της πιθανότητας λάθους (κάτω από ποιές συνθήκες ο ταξινομητής αυτός θα μπορούσε να είναι βέλτιστος;)