ΚΕΣ 03 – Αναγνώριση Προτύπων και Ανάλυση Εικόνας

Εκτίμηση άγνωστων κατανομών πιθανότητας

Τμήμα Επιστήμης και Τεχνολογίας Τηλεπικοινωνιών

Πανεπιστήμιο Πελοποννήσου

☐ Εισαγωγή
☐ Παραμετρικές μέθοδοι
☐ Μη παραμετρικές μέθοδοι
☐ Ο απλοϊκός ταξινομητής Bayes
☐ Ο ταξινομητής k-NN

Περιεχόμενα – Βιβλιογραφία

- Περιεχόμενα Ενότητας
  - Εισαγωγή
  - Παραμετρικές μέθοδοι εκτίμησης κατανομών πιθανότητας
  - Μη παραμετρικές μέθοδοι
  - Ο απλοϊκός ταξινομητής Bayes (Naïve Bayes Classifier)
  - Ο ταξινομητής των k-πλησιέστερων γειτόνων (k-NN Classifier)
- Βιβλιογραφία:
  - Παπαμάρκος [2005]: Κεφάλαιο 7
  - Duda [2004]: Chapter 3
  - Theodoridis [2002]: Chapter 2
  - Bow [2002]: Chapter 2

Εισαγωγή

- Η εφαρμογή της ταξινόμησης σύμφωνα με τον κανόνα του Bayes απαιτεί τη γνώση των κατανομών πιθανότητας $p(\underline{x}|\omega_i)$, $i = 1,2, \ldots, M$
  - Αν οι ανωτέρω κατανομές δεν είναι γνωστές πρέπει να εκτιμηθούν από ένα σύνολο διανυσμάτων $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N\}$ για τα οποία είναι γνωστές οι κλάσεις στις οποίες ανήκουν

- Υπάρχουν δύο βασικές κατηγορίες εκτίμησης κατανομών πιθανότητας:
  - Παραμετρικές. Η μορφή της συνάρτησης κατανομής (π.χ. Gaussian, Rayleigh, Poison) είναι γνωστή και χρειάζεται να εκτιμηθούν οι παράμετροι της
  - Μη παραμετρικές. Δεν υπάρχει καμία γνώση της μορφής της κατανομής. Αυτή απλά εκτιμάται από τα δεδομένα εκπαίδευσης

★ Εισαγωγή
☐ Παραμετρικές μέθοδοι
☐ Μη παραμετρικές μέθοδοι
☐ Ο απλοϊκός ταξινομητής Bayes
☐ Ο ταξινομητής k-NN

Κανόνας του Bayes

- Ο κανόνας ταξινόμησης του Bayes μας λέει:
  - Δεδομένου του $\underline{x}$ ταξινόμησε το σύμφωνα με τον κανόνα:

$$\boxed{P(\omega_i|\underline{x}) > P(\omega_j|\underline{x}) \quad \forall j \neq i}$$

  - όπου $P(\omega_i|\underline{x})$ είναι η εκ των υστέρων (a-posteriori) πιθανότητα το $\underline{x}$ να ανήκει στην κλάση ω_i.

$$P(\omega_i|\underline{x}) = \frac{p(\underline{x}|\omega_i)P(\omega_i)}{p(\underline{x})}$$

$$p(\underline{x}) = \sum_{i=1}^{M} p(\underline{x}|\omega_i)P(\omega_i)$$

Παραμετρικές μέθοδοι

- Σε κάποιες περιπτώσεις η μορφή των κατανομών πιθανότητας είναι γνωστή και χρειάζεται απλά να εκτιμηθούν οι παράμετροι της κατανομής
  - Στην (πολυδιάστατη) κατανομή Gauss Ν($\underline{\mu}$,Σ) οι παράμετροι $\underline{\mu}$, Σ.
  - Στη μονοδιάστατη εκθετική κατανομή οι παράμετροι α, λ.
  - Στη πολυδιάστατη κατανομή Bernouli το διάνυσμα $\underline{\theta}$.
- Στις ανωτέρω περιπτώσεις η εκτίμηση των συναρτήσεων κατανομής πιθανότητας πραγματοποιείται με την εκτίμηση των αγνώστων παραμέτρων
  - Υπάρχουν διάφορες μέθοδοι παραμετρικής εκτίμησης. Οι πιο συχνά χρησιμοποιούμενες αναλύονται στη συνέχεια

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

☑ Εισαγωγή
★ Παραμετρικές μέθοδοι
☐ Μη παραμετρικές μέθοδοι
☐ Ο απλοϊκός ταξινομητής Bayes
☐ Ο ταξινομητής k-NN

Μέθοδος μέγιστης πιθανοφάνειας

- Η μέθοδος της μεγίστης πιθανοφάνειας (ML-Maximum Likelihood) με δεδομένο ένα σύνολο από διανύσματα παρατήρησης $\{x_1,x_2,\ldots,x_N\}$ που ανήκουν στην κλάση ω_i, επιλέγει τις παραμέτρους $\underline{\theta}_i$ της κατανομής $p(\underline{x}|\omega_i)$ ώστε να μεγιστοποιείται η συνολική πιθανότητα των διανυσμάτων παρατήρησης
- Παράδειγμα: Το σχήμα μας δείχνει το ιστόγραμμα των τιμών 1000 παρατηρήσεων για τις οποίες γνωρίζουμε ότι ακολουθούν την κανονική κατανομή (με άγνωστες τιμές για το μ, και σ).
  - Επιλέξτε τις τιμές μ και σ σύμφωνα με τον κανόνα της μέγιστης πιθανοφάνειας
  - Υπόδειξη: Για κανονικές κατανομές Ν(μ,σ²) ισχύουν τα κατωτέρω:

$$p(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$$

$$P(x < (\mu - 2\sigma)) = P(x > (\mu + 2\sigma)) = 0.025$$

$$P(x) = \int_{-\infty}^{x} p(z)dz$$

☑ Εισαγωγή
★ Παραμετρικές μέθοδοι
☐ Μη παραμετρικές μέθοδοι
☐ Ο απλοϊκός ταξινομητής Bayes
☐ Ο ταξινομητής k-NN

Μέθοδος μέγιστης πιθανοφάνειας

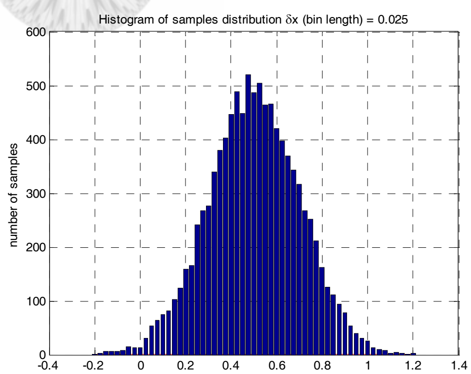
- Η μέθοδος της μεγίστης πιθανοφάνειας (ML-Maximum Likelihood) με δεδομένο ένα σύνολο από διανύσματα παρατήρησης $\{x_1, x_2, \ldots, x_N\}$ που ανήκουν στην κλάση ω_i, επιλέγει τις παραμέτρους $\underline{\theta}_i$ της κατανομής $p(\underline{x}|\omega_i)$ ώστε να μεγιστοποιείται η συνολική πιθανότητα των διανυσμάτων παρατήρησης
- Παράδειγμα: Το σχήμα μας δείχνει το ιστόγραμμα των τιμών 1000 παρατηρήσεων για τις οποίες γνωρίζουμε ότι ακολουθούν την κανονική κατανομή (με άγνωστες τιμές για το μ, και σ).
  - Επιλέξτε τις τιμές μ και σ σύμφωνα με τον κανόνα της μέγιστης πιθανοφάνειας
  - Υπόδειξη: Για κανονικές κατανομές $N(\mu,\sigma^2)$ ισχύουν τα κατωτέρω:

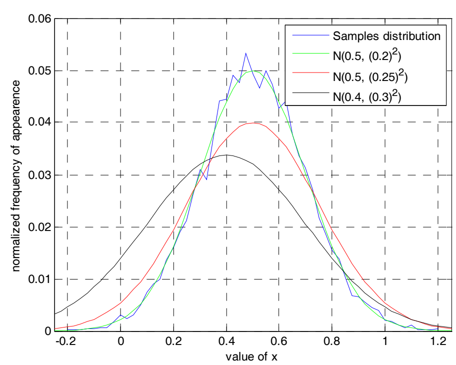
$$p(\mu) = \frac{1}{\sigma\sqrt{2\pi}} \qquad P(x < (\mu - 2\sigma)) = P(x > (\mu + 2\sigma)) = 0.025$$

$$P(x) = \int_{-\infty}^{x} p(z)dz$$

☑ Εισαγωγή
★ Παραμετρικές μέθοδοι
☐ Μη παραμετρικές μέθοδοι
☐ Ο απλοϊκός ταξινομητής Bayes
☐ Ο ταξινομητής k-NN

Μέθοδος μέγιστης πιθανοφάνειας (II)





Μέθοδος μέγιστης πιθανοφάνειας (III)

Έστω ότι τα διανύσματα $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N$
είναι γνωστά και μεταξύ τους ανεξάρτητα
Έστω $p(\underline{x})$ γνωστή υπό την αίρεση ενός
αγνώστου διανύσματος παραμέτρων :
$\underline{\theta}: \ p(\underline{x}) \equiv p(\underline{x};\underline{\theta})$

Έστω $X = \{\underline{x}_1, \underline{x}_2, \ldots \underline{x}_N\}$ το σύνολο των διανυσμάτων παρατήρησης
Ορίζουμε τη συνάρτηση :

$$p(X;\underline{\theta}) \equiv p(\underline{x}_1, \underline{x}_2, \ldots \underline{x}_N;\underline{\theta}) = \prod_{k=1}^{N} p(\underline{x}_k;\underline{\theta})$$

Η συνάρτηση $p(X;\underline{\theta})$ είναι γνωστή ως Πιθανοφάνεια
(Likelihood) του $\underline{\theta}$ ως προς το σύνολο X

Μέθοδος μέγιστης πιθανοφάνειας (IV)

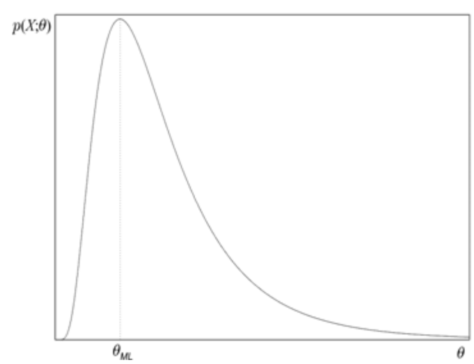
Υπολογισμός των άγνωστων παραμέτρων.
Διάλεξε το διάνυσμα παραμέτρων που μεγιστοποιεί
την πιθανοφάνεια :

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

Η συνάρτηση ln() είναι μια μονοτονική
συνάρτηση.

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

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



Παράδειγμα

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

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

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

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

$$\text{Υπόδειξη: Αν } A=A^T\Rightarrow\frac{\partial(\underline{\alpha}^T A\underline{\alpha})}{\partial\underline{\alpha}}=2A\underline{\alpha}$$

Παράδειγμα (II)

- Έστω οι τιμές παρατήρησης $\{x_1,x_2,\ldots,x_N\}$ που είναι γνωστό ότι προέρχονται από την κατανομή

$$p(x)=\begin{cases}\theta e^{-\theta x} & x\geq 0\\ 0 & x<0\end{cases}$$

- Να εκτιμηθεί η τιμή της παραμέτρου θ με τη μέθοδο της μεγίστης πιθανοφάνειας

Μέθοδος της μέγιστης εκ των υστέρων πιθανότητας

- Η μέθοδος της μεγίστης εκ των υστέρων πιθανότητας (MAP-Maximum Aposteriori Probability) θεωρεί ότι το διάνυσμα παραμέτρων $\underline{\theta}$ είναι ένα τυχαίο διάνυσμα με γνωστή κατανομή πιθανότητας $p(\underline{\theta})$.
- Με δεδομένο ένα σύνολο από διανύσματα παρατήρησης $X = \{x_1, x_2, \ldots, x_N\}$ η εκτίμηση του διανύσματος $\underline{\theta}$ πραγματοποιείται με μεγιστοποίηση της συνάρτησης της εκ των υστέρων πιθανότητας: $p(\underline{\theta}|X)$
- Από τον κανόνα του Bayes έχουμε:

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

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

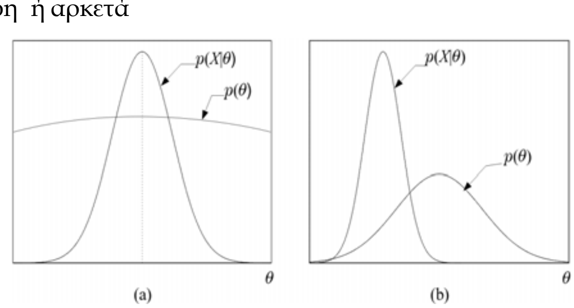
Μέθοδος της μέγιστης εκ των υστέρων πιθανότητας (II)

- Υπολογισμός του διανύσματος παραμέτρων $\underline{\theta}$:

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

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

Αν $p(\underline{\theta})$ είναι ομοιόμορφη ή αρκετά ευρεία τότε $\hat{\underline{\theta}}_{MAP} \cong \underline{\theta}_{ML}$



Παράδειγμα

- Έστω ότι ζητάμε να εκτιμήσουμε το μέσο διάνυσμα $\underline{\mu}$ μιας πολυδιάστατης κανονικής κατανομής N($\underline{\mu}$,σI), για την οποία γνωρίζουμε ότι το διάνυσμα της μέσης τιμής ακολουθεί επίσης κανονική κατανομή N($\underline{\mu}$,σ_μI):

$$p(\underline{x}): N(\underline{\mu},\Sigma),\ \underline{\mu} \text{ unknown},\ X=\{\underline{x}_1,\dots,\underline{x}_N\}$$

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

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

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

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

Μεικτά μοντέλα

- Τα μεικτά μοντέλα (mixture models) προσεγγίζουν μια άγνωστη κατανομή πιθανότητας συνδυάζοντας πολλές κατανομές.
  - Αποδεικνύεται ότι με συνδυασμό κατανομών μπορεί να προσεγγίσουμε οποιαδήποτε άγνωστη κατανομή

$$p(\underline{x})=\sum_{j=1}^{J}p(\underline{x}|j)P_j$$

$$\sum_{j=1}^{M}P_j=1, \int_{\underline{x}}p(\underline{x}|j)d\underline{x}=1$$

  - Εργαζόμενοι όπως και πριν με παραμετρική μοντελοποίηση έχουμε: $p(\underline{x}|j;\underline{\theta})$
  - Ο στόχος είναι η εκτίμηση των $\underline{\theta}$ και $P_1,P_2,\dots,P_J$ δεδομένου του συνόλου $X=\{\underline{x}_1,\underline{x}_2,\dots,\underline{x}_N\}$

Μεικτά μοντέλα (ΙΙ)

- Μπορούμε να εφαρμόσουμε τη μέθοδο μεγίστης πιθανοφάνειας όπως και πριν:

$$\hat{\Theta}_{opt} = \underset{\underline{\theta}, P_i, \ldots, P_j}{\arg\max} \prod_{k=1}^{N} P(\underline{x}_k ; \underline{\theta}, P_i, \ldots, P_j)$$

- Η διαφορά είναι ότι τώρα χρειάζεται να εκτιμηθούν οι εκ των προτέρων πιθανότητες P_j, j = 1,2,…,J για τις επιμέρους κατανομές οι οποίες δεν είναι άμεσα παρατηρήσιμες από το σύνολο X.
  - Το ανωτέρω πρόβλημα είναι ένα χαρακτηριστικό πρόβλημα μη πλήρους συνόλου δεδομένων (incomplete data set)
  - Μια πολύ διαδεδομένη μέθοδος για την επίλυση τέτοιων προβλημάτων είναι ο αλγόριθμος EM (Expectation Maximization)

☑ Εισαγωγή
★ Παραμετρικές μέθοδοι
☐ Μη παραμετρικές μέθοδοι
☐ Ο απλοϊκός ταξινομητής Bayes
☐ Ο ταξινομητής k-NN

Ο αλγόριθμος EM

- Γενική θεώρηση:
  - Έστω $\underline{y}$ το πλήρες σύνολο δεδομένων $\underline{y} \in Y \subseteq R^m$, με $_{y}(\underline{y};\underline{\theta})$, τα οποία δεν είναι άμεσα παρατηρήσιμα
  - Αντίθετα παρατηρήσιμος είναι ο υπόχωρος $\underline{x}$

    $\underline{x} = g(\underline{y}) \in X_{ob} \subseteq R^l, l < m$ with $P_x(\underline{x};\underline{\theta})$,
  - Προφανώς για το μη παρατηρήσιμο μέρος των διανυσμάτων y μπορεί να έχει οποιαδήποτε τιμή
  - Ο αλγόριθμος EM συμπληρώνει τα ελλειπή δεδομένα αναθέτοντας σε αυτά τις πιθανότερες τιμές (αναμενόμενες)
  - Στο προηγούμενο παράδειγμα οι εκτιμήσεις για τις εκ των προτέρων πιθανότητες P_j, $j=1,2,\ldots,J$ βασίζονται στην τρέχουσα εκτίμηση των παραμέτρων $\underline{\theta}$.

Ο αλγόριθμος ΕΜ (ΙΙ)

- Έστω: $Y(\underline{x}) \subseteq Y$ όλα τα $\underline{y} \rightarrow$ που αντιστοιχούν σε ένα συγκεκριμένο $\underline{x}$

$$p_{\underline{x}}(\underline{x};\underline{\theta}) = \int_{Y(\underline{x})} p_{\underline{y}}(\underline{y};\underline{\theta})d\underline{y}$$

  - Αυτό που ζητάμε να υπολογίσουμε είναι:

$$\hat{\theta}_{ML} : \sum_k \frac{\partial \ln(p_{\underline{y}}(\underline{y}_k;\underline{\theta}))}{\partial \underline{\theta}} = \underline{0}$$

  - Αλλά δυστυχώς τα y_k δεν είναι παρατηρήσιμα
  - Ο αλγόριθμος ΕΜ μεγιστοποιεί την αναμενόμενη τιμή για τα $\underline{y}_k$ δεδομένων των $\underline{x}_k$ και της τρέχουσας των αγνώστων παραμέτρων $\underline{\theta}$.

Ο αλγόριθμος ΕΜ (ΙΙΙ)

- Ο αλγόριθμος εκτελείται σε δυο βήματα:
  - Εύρεση της αναμενόμενης τιμής $Q(\theta;\theta(t))$ για τις τιμές των $\underline{y}_k$ δεδομένων των $X = \{x_1, x_2, \ldots, x_N\}$ και της τρέχουσας εκτίμησης για τις παραμέτρους $\underline{\theta}(t)$.
    - E-step: $Q(\underline{\theta};\underline{\theta}(t)) = E[\sum_k \ln(p_{\underline{y}}(\underline{y}_k;\underline{\theta}|X;\underline{\theta}(t))]$
  - Εύρεση της νέας εκτίμησης $\theta(t+1)$ για τις άγνωστες παραμέτρους με μεγιστοποίηση της αναμενόμενης τιμής
    - M-step:

$$\underline{\theta}(t+1) : \frac{\partial Q(\underline{\theta};\underline{\theta}(t))}{\partial \underline{\theta}} = \underline{0}$$

- ☑ Εισαγωγή
- ★ Παραμετρικές μέθοδοι
- ☐ Μη παραμετρικές μέθοδοι
- ☐ Ο απλοϊκός ταξινομητής Bayes
- ☐ Ο ταξινομητής k-NN

Ο αλγόριθμος ΕΜ στην προσέγγιση κατανομών

- Ο αλγόριθμος ΕΜ εφαρμόζεται στην εκτίμηση των αγνώστων κατανομών με χρήση μεικτών μοντέλων:
  - Πλήρη δεδομένα: $(\underline{x}_k, j_k),\ k = 1,2,\ldots,N$
    - j_k η κατανομή από την οποία δημιουργείται το διάνυσμα $\underline{x}_k$.
  - Παρατηρήσιμα είναι μόνο τα $\underline{x}_k$ και όχι οι τιμές j_k $\underline{x}_k,\ k = 1,2,\ldots,N$

    $p(\underline{x}_k, j_k; \underline{\theta}) = p(\underline{x}_k | j_k; \underline{\theta}) P_{jk}$
  - Θεωρώντας ότι τα διανύσματα παρατήρησης xk είναι μεταξύ τους ανεξάρτητα λαμβάνουμε τη λογαριθμική μορφή της συνάρτησης πιθανοφάνειας ως ακολούθως:

$$L(\underline{\theta}) = \sum_{k=1}^{N} \ln(p(\underline{x}_k | j_k; \underline{\theta}) P_{jk})$$

- ☑ Εισαγωγή
- ★ Παραμετρικές μέθοδοι
- ☐ Μη παραμετρικές μέθοδοι
- ☐ Ο απλοϊκός ταξινομητής Bayes
- ☐ Ο ταξινομητής k-NN

Ο αλγόριθμος ΕΜ στην προσέγγιση κατανομών (II)

- Άγνωστες παράμετροι:

$$\underline{\Theta}^T = [\underline{\theta}^T, \underline{P}^T]^T,\ \underline{P} = [P_1, P_2, \ldots, P_j]^T$$

- E-step:

$$Q(\underline{\Theta}; \underline{\Theta}(t)) = E[\sum_{k=1}^{N} \ln(p(\underline{x}_k | j_k; \underline{\theta}) P_{jk})] = \sum_{k=1}^{N} E[\quad] =$$

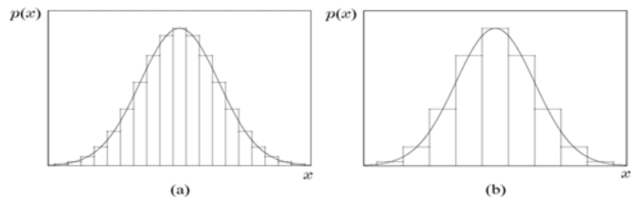
$$\sum_{k=1}^{N} \sum_{j_k=1}^{J} \boxed{P(j_k | \underline{x}_k; \underline{\Theta}(t))} \ln(p(\underline{x}_k | j_k; \underline{\theta}) P_{jk})$$

- M-step: $\dfrac{\partial Q}{\partial \underline{\theta}} = \underline{0} \qquad \dfrac{\partial Q}{\partial P_{jk}} = 0, \qquad j_k = 1,2,\ldots,J$

$$P(j | \underline{x}_k; \Theta(t)) = \frac{p(\underline{x}_k | j; \underline{\Theta}(t)) P_j}{P(\underline{x}_k; \underline{\Theta}(t))} \qquad p(\underline{x}_k; \underline{\Theta}(t)) = \sum_{j=1}^{J} p(\underline{x}_k | j; \underline{\theta}(t)) P_j$$

Μη παραμετρικές μέθοδοι

- Στις μη παραμετρικές μεθόδους η συναρτήσεις πυκνότητας πιθανότητας δεν έχουν μια προδιαγεγραμμένη μορφή.
- Μπορούν να είναι απλά το ιστόγραμμα των εκτιμήσεων πιθανότητας σε ένα εύρος τιμών των διανυσμάτων παρατήρησης



$$P \approx \frac{k_N}{N} \qquad k_N \text{ in } h, \quad N \text{ total}$$

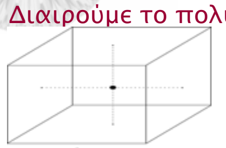
$$\hat{x}-\frac{h}{2} \quad \hat{x} \quad \hat{x}+\frac{h}{2}$$

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

Αν η $p(x)$ είναι συνεχής, $\hat{p}(x) \to p(x)$ όταν $N \to \infty$, εφόσον $h_N \to 0$, $k_N \to \infty$, $\frac{k_N}{N} \to 0$

Μέθοδος παραθύρων Parzen

- Διαιρούμε το πολυδιάστατο χώρο σε υπερκύβους:



Ορίζουμε τη συνάρτηση $\varphi(\underline{x}_i) = \begin{Bmatrix} 1 & |x_{ij}| \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{Bmatrix}$

- Δηλαδή ένα υπερκύβο με πλευρά 1 και κεντραρισμένο στο 0

$$\boxed{\hat{p}(\underline{x}) = \frac{1}{h^l}\left(\frac{1}{N}\sum_{i=1}^{N}\varphi\left(\frac{\underline{x}_i - \underline{x}}{h}\right)\right)} \qquad \frac{1}{\text{όγκος}} * \frac{1}{N} * \text{αριθμός σημείων εντός του κύβου με πλευρά h - και κεντραρισμένο στο } \underline{x}$$

- Υπάρχει ένα πρόβλημα: $p(\underline{x})$ συνεχής
  $\varphi(\cdot)$ ασυνεχής
- Για εξομάλυνση ορίζουμε τα παράθυρα Parzen μέσω κάποιων συναρτήσεων πυρήνα (kernel functions – παράδειγμα N(0,I):

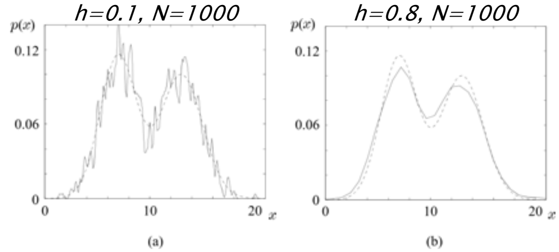
$$\varphi(\underline{x}) \text{ is smooth}$$

$$\varphi(\underline{x}) \geq 0, \; \int_{\underline{x}} \varphi(\underline{x})\, d\underline{x} = 1$$

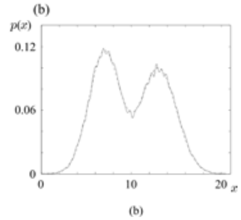
Μέθοδος παραθύρων Parzen (II)

- Μεταβλητότητα:
  - Όσο μικρότερο είναι το h τόσο μεγαλύτερη είναι η διασπορά

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



  - Όσο μεγαλύτερο το N τόσο καλύτερη η ακρίβεια της εκτίμησης



Μέθοδος παραθύρων Parzen (III)

- Εφαρμογή της μεθόδου στη ταξινόμηση με τον κανόνα Bayes:

$$l_{12} \equiv \frac{p(\underline{x}|\omega_1)}{p(\underline{x}|\omega_2)} (><) \frac{P(\omega_2)}{P(\omega_1)} \frac{\lambda_{21} - \lambda_{22}}{\lambda_{12} - \lambda_{11}} \equiv \theta$$

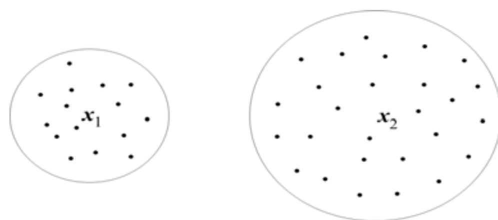
$$\boxed{\frac{\dfrac{1}{N_1 h^l} \sum_{i=1}^{N_1} \varphi\left(\dfrac{\underline{x}_i - \underline{x}}{h}\right)}{\dfrac{1}{N_2 h^l} \sum_{i=1}^{N_2} \varphi\left(\dfrac{\underline{x}_i - \underline{x}}{h}\right)} (><) \theta}$$

Μέθοδος k-NN (II)

- Μέθοδος Parzen
  - Ο όγκος του κάθε υπερκύβου
  - Ο αριθμός των σημείων εντός κάθε υπερκύβου κυμαίνεται

- Μέθοδος k-NN
  - Ο αριθμός των σημείων διατηρείται σταθερός
  - Ο όγκος αυξάνεται

$$\frac{\frac{k}{N_1V_1}}{\frac{k}{N_2V_2}} = \boxed{\frac{N_2V_2}{N_1V_1}(><)\theta}$$



Ο απλοϊκός ταξινομητής Bayes

- Η κατάρα της διάστασης (curse of dimensionality)
  - Σε όλες τις μεθόδους μη παραμετρικής εκτίμησης όσο περισσότερα σημεία N λαμβάνονται σε κάθε υπερκύβο τόσο καλύτερη εκτίμηση έχουμε για τη συνάρτηση πυκνότητας πιθανότητας
  - Αν στη μονοδιάστατη περίπτωση (οι παρατηρήσεις είναι απλά μεμονωμένες τιμές και όχι διανύσματα) απαιτούνται N σημεία σε κάθε διάστημα εύρους h τότε στη πολυδιάστατη περίπτωση (διανύσματα μήκους l) απαιτούνται N^l σημεία για καλή εκτίμηση.
  - Η εκθετική αύξηση των απαιτούμενων σημείων με την αύξηση της διάστασης του διανύσματος ονομάζεται κατάρα της διάστασης

Ο απλοϊκός ταξινομητής Bayes (II)

- Έστω $\underline{x} \in \Re^{\ell}$ και ότι επιθυμούμε να εκτιμήσουμε τις κατανομές $p(\underline{x} \mid \omega_i)$ i = 1, 2, …, M.
- Για αποτελεσματική εκτίμηση χρειάζονται N^{ℓ} σημεία (διανύσματα παρατήρησης).
- Υποθέτοντας ότι όλα τα στοιχεία των διανυσμάτων είναι μεταξύ τους ανεξάρτητα (x_1, x_2 ,…, x_{ℓ} mutually independent). Τότε:

$$p(\underline{x} \mid \omega_i) = \prod_{j=1}^{\ell} p(x_j \mid \omega_i)$$

- Σε αυτή την περίπτωση απαιτούνται N σημεία για κάθε συνάρτηση πυκνότητας πιθανότητας. Επομένως ένας συνολικός αριθμός της τάξης των $N \cdot \ell$ σημείων είναι αρκετός.
- Η ανωτέρω εκτίμηση οδηγεί στο απλοϊκό ταξινομητή Bayes (δηλαδή εφαρμογή της ταξινόμησης Bayes με βάση την απλοϊκή εκτίμηση της κατανομής)

Ο ταξινομητής k-NN

- Έστω ένα διάνυσμα εισόδου $\underline{x}$. Επιλέγουμε τα k πλησιέστερα σε αυτό διανύσματα παρατήρησης (από το σύνολο των N)
- Από τα k αυτά διανύσματα τα k_i ανήκουν στην κατηγορία ω_i.
- Ανέθεσε το $\underline{x}$ σύμφωνα με τον κανόνα: $\underline{x} \rightarrow \omega_i : k_i > k_j \quad \forall i \neq j$
- Στην απλούστερη περίπτωση k=1 !!!
- Για μεγάλο αριθμό δειγμάτων N η προσέγγιση αυτή δεν απέχει πολύ από τον βέλτιστο ταξινομητή (κανόνας Bayes).
- Έστω P_B η πιθανότητα σφάλματος ταξινόμησης με τον κανόνα Bayes και P_{NN} η αντίστοιχη πιθανότητα στον ταξινομητή 1-NN. Ισχύει:

$$P_B \leq P_{NN} \leq P_B(2 - \frac{M}{M-1}P_B) \leq 2P_B \qquad P_B \leq P_{kNN} \leq P_B + \sqrt{\frac{2P_{NN}}{k}}$$