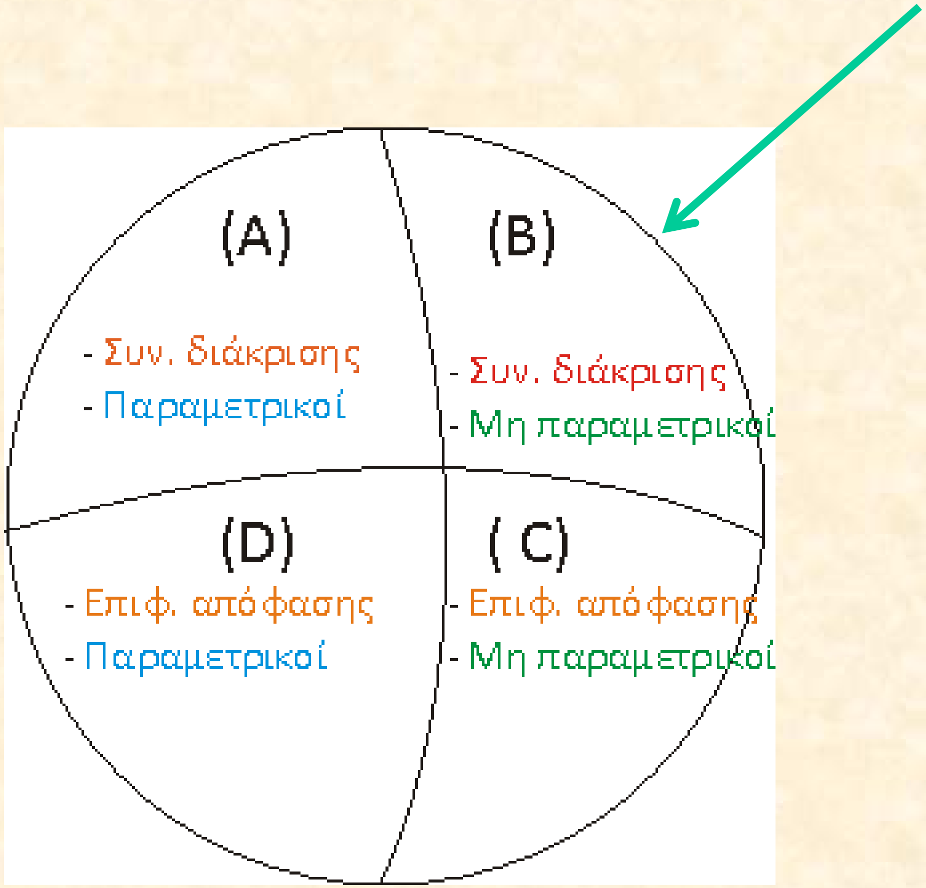
❖ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ
❖(PATTERN RECOGNITION)

Σέργιος Θεοδωρίδης

Κωνσταντίνος Κουτρούμπας


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

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση

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

-Στο πλαίσιο αυτό, η τιμή κάθε pdf $p(\boldsymbol{x}|\omega_j)$ που εμπλέκεται στον **ταξινομητή Bayes** για δεδομένο $\boldsymbol{x}$ εκτιμάται βασιζόμενη **άμεσα** στα σημεία του X_j.

Δύο βασικές μέθοδοι της κατηγορίας αυτής

-**Parzen windows**

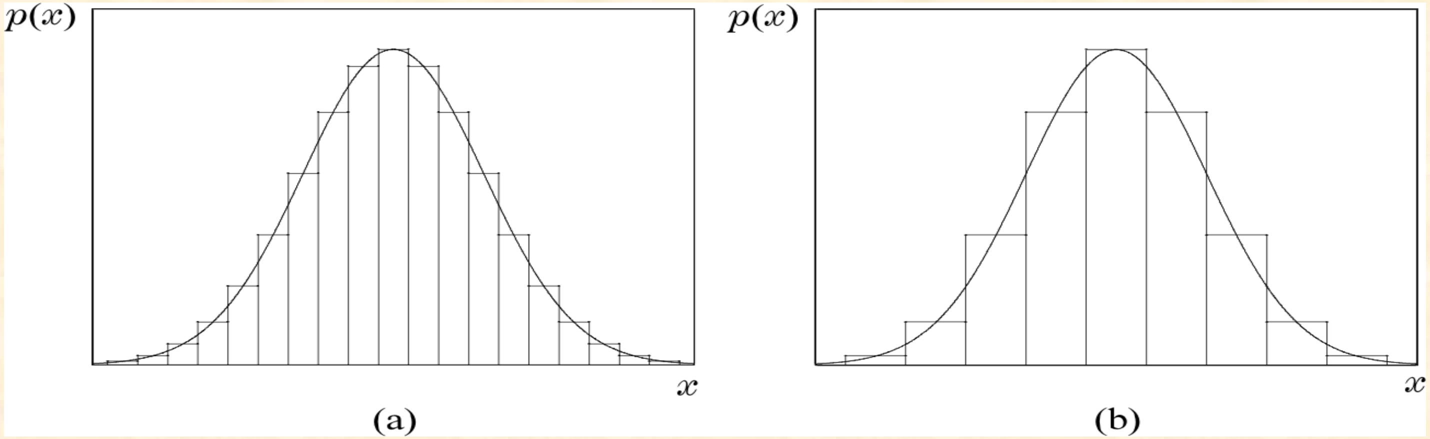
-**Εκτίμηση πυκνότητας με βάση του k-πλησιέστερους γείτονες (k-nearest neighbor (NN) density estimation)**

Το πρόβλημα:

- Έστω $Y=\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\}$ δεδομένο σύνολο διανυσμάτων που έχουν προκύψει από άγνωστη pdf, $p(\boldsymbol{x})$.

- **Στόχος:** Εκτίμησε την τιμή της $p(\boldsymbol{x})$ σε δεδομένο σημείο $\boldsymbol{x}$ με βάση το σύνολο Y.

Μη παραμετρική εκτίμηση – Υπενθύμιση βασικών εννοιών



- $P \approx \dfrac{k_N}{N}$ — k_N στο h — N συνολικά

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

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

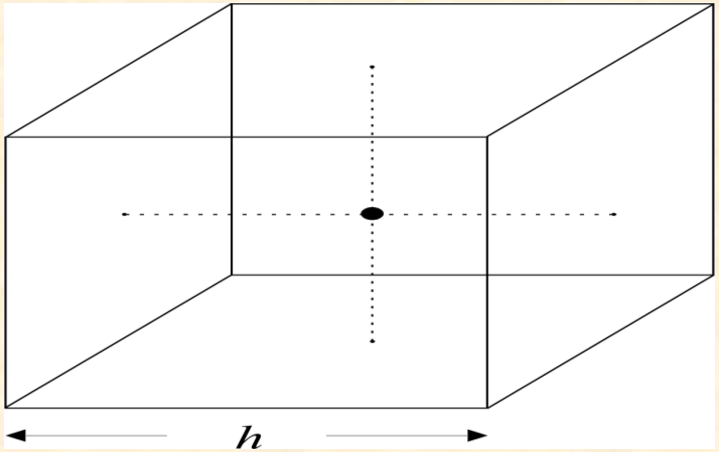
- Αν $p(x)$ συνεχής, $\hat{p}(x) \to p(x)$ καθώς $N \to \infty$, αν

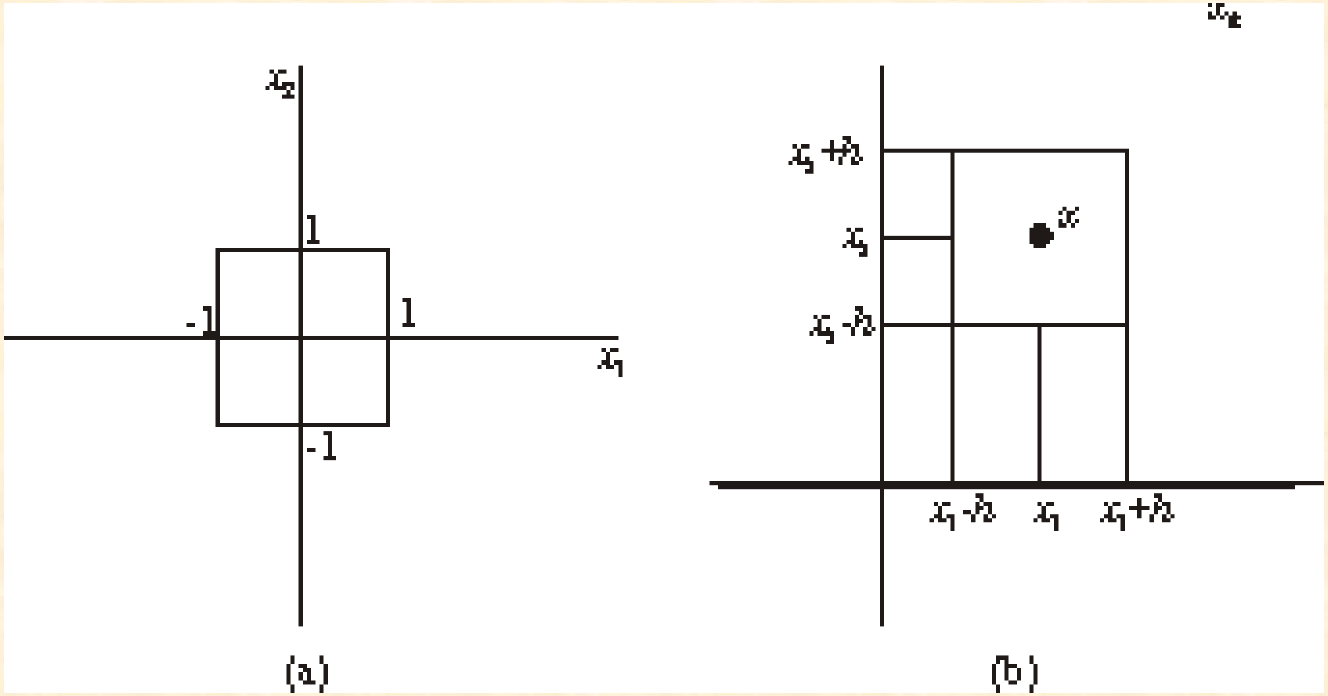
$$h_N \to 0, \qquad k_N \to \infty, \qquad \frac{k_N}{N} \to 0$$

(B) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

H Bayesian περίπτωση – Parzen windows

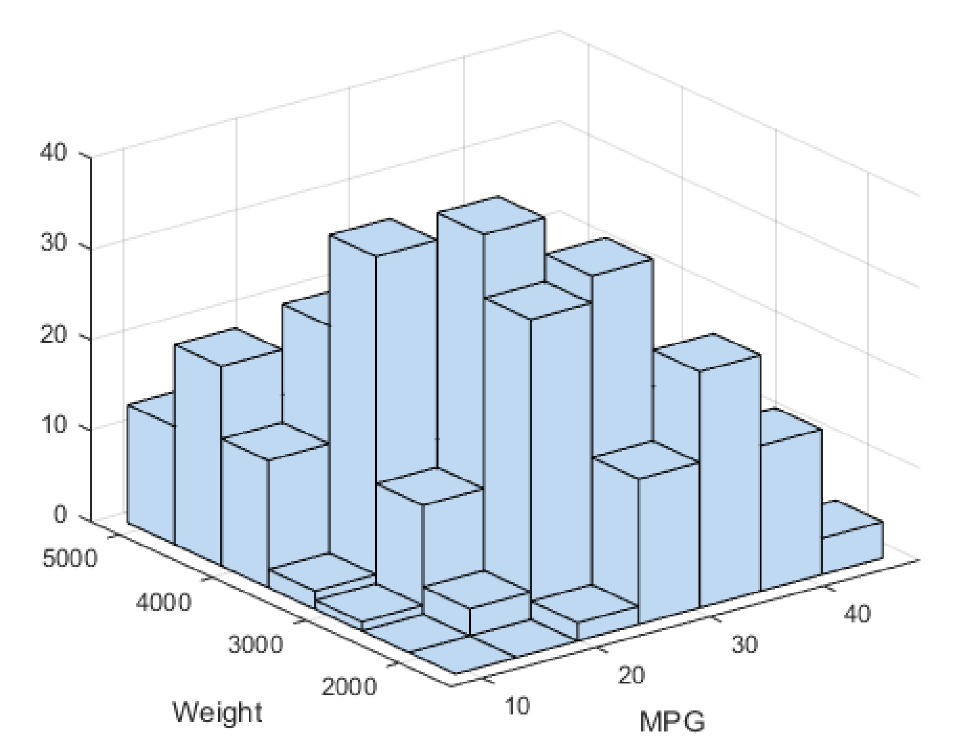
- Διαίρεση του πολυδιάστατου χώρου σε υπερκύβους





(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

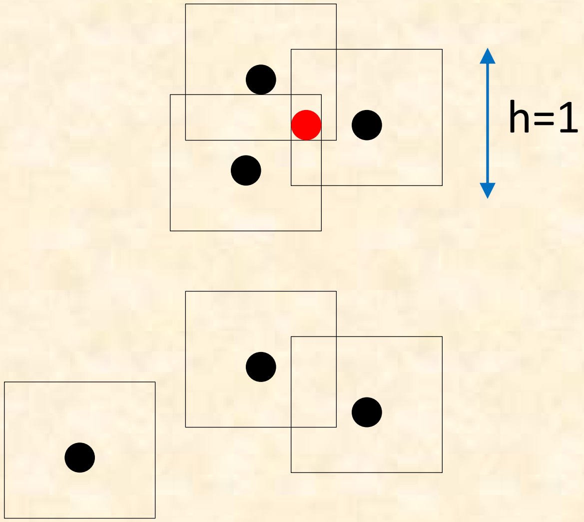
Η Bayesian περίπτωση – Parzen windows



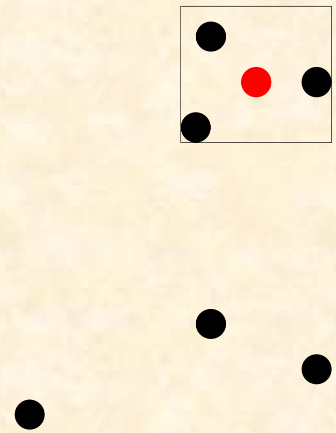
(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Parzen windows

Παράδειγμα: Έστω τα ακόλουθα **N=6** σημεία δεδομένων (μαύρες κουκκίδες), τα οποία προέρχονται από μία άγνωστη κατανομή $p(\boldsymbol{x})$. Για **h=1**, να εκτιμήσετε την τιμή της συν. πυκν. πιθ. για την **κόκκινη** κουκκίδα.



ή



- **Κεντράρουμε υπερκύβο** πλευράς h σε κάθε σημείο δεδομένων

- Μετράμε σε πόσους υπερκύβους ανήκει το υπό μελέτη σημείο (κόκκινη κουκίδα)

→ k=3→ P = k/N = 3/6→p(x) = $(1/h^2)$*P =0.5

- **Κεντράρουμε υπερκύβο** πλευράς h στο υπό μελέτη σημείο (κόκκινη κουκκίδα)

- Μετράμε πόσα σημεία δεδομένων ανήκουν στον υπερκύβο

→ k=3→ P = k/N = 3/6→p(x) = $(1/h^2)$*P =0.5

➢ Ορίζουμε

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

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

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

- (1/όγκος)*(1/N)* πλήθος σημείων εντός υπερκύβου πλευράς *h* κεντραρισμένου στο $\underline{x}$.

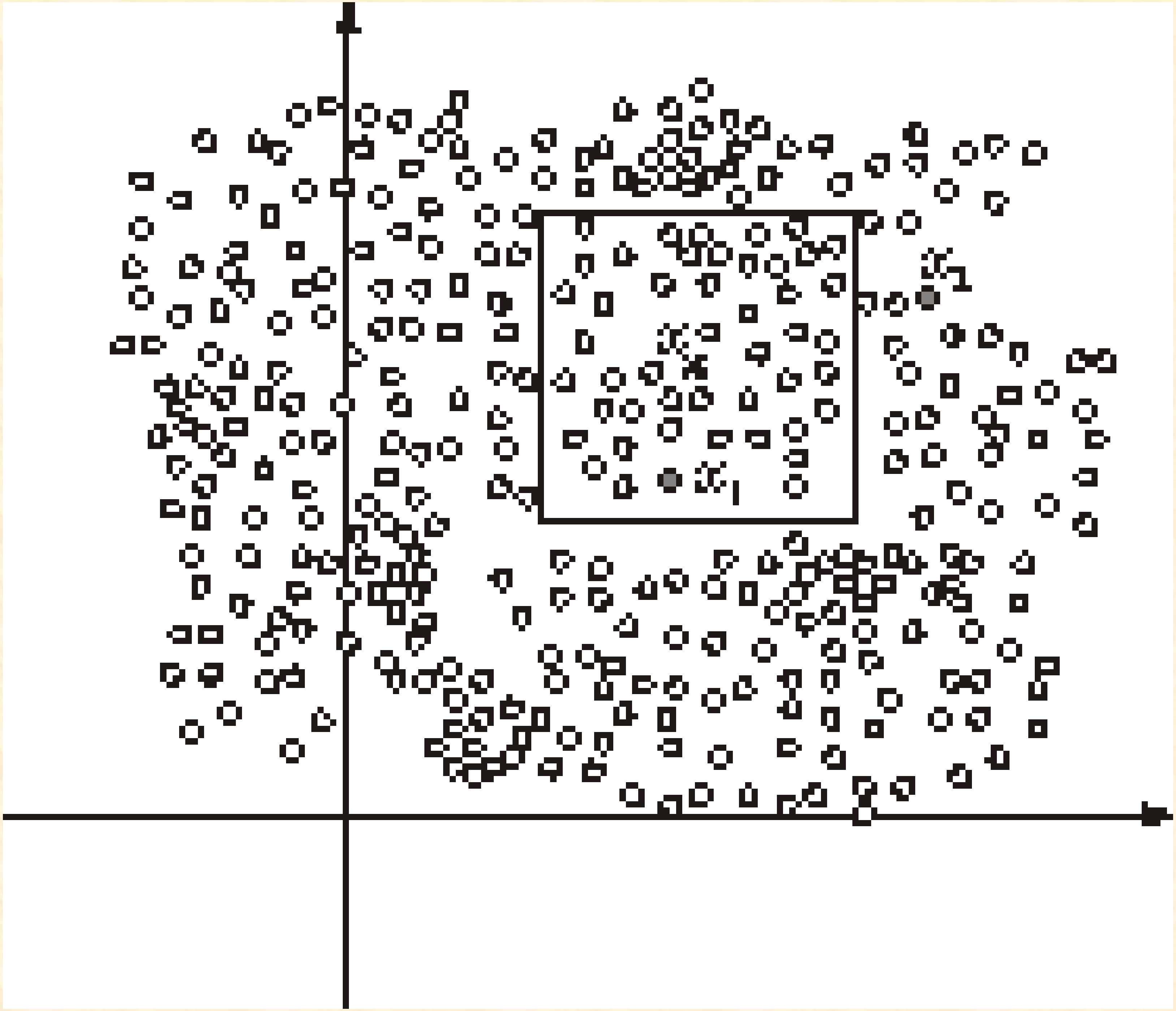
- Το πρόβλημα: $p(\underline{x})$ συνεχής

  $\phi(.)$ ασυνεχής

- Παράθυρα Parzen-Πυρήνες-συναρτήσεις δυναμικού

$$\phi(\underline{x}) \text{ είναι ομαλή}$$

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



(B) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Parzen windows

Αν *φ(x)=N(0,I)* τότε

$$\hat{p}(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{(2\pi)^{\frac{l}{2}} h^l} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{x}_i)^T(\boldsymbol{x}-\boldsymbol{x}_i)}{2h^2}\right)$$

Παράδειγμα: Εκτίμηση της τιμής κανονικής κατανομής *N*(0,1) στο σημείο *x*=1.

Παράγουμε *n* σημεία από την κατανομή και εκτιμούμε την *p*(1) (=***0,2420***) για δεδομένα *h*.

	h=0.01	h=0.05	h=0.1	h=0.5	h=0.8
n=100	0.1635	0.4750	0.1455	0.2305	0.2464
n=1000	0.1776	0.2428	0.2629	0.2270	0.2211
n=10000	0.2662	0.2415	0.2437	0.2365	0.2317
n=100000	0.2543	0.2424	0.2403	0.2378	0.2304
n=1000000	0.2413	0.2415	0.2406	0.2394	0.2296

➢ Μέση τιμή

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

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

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

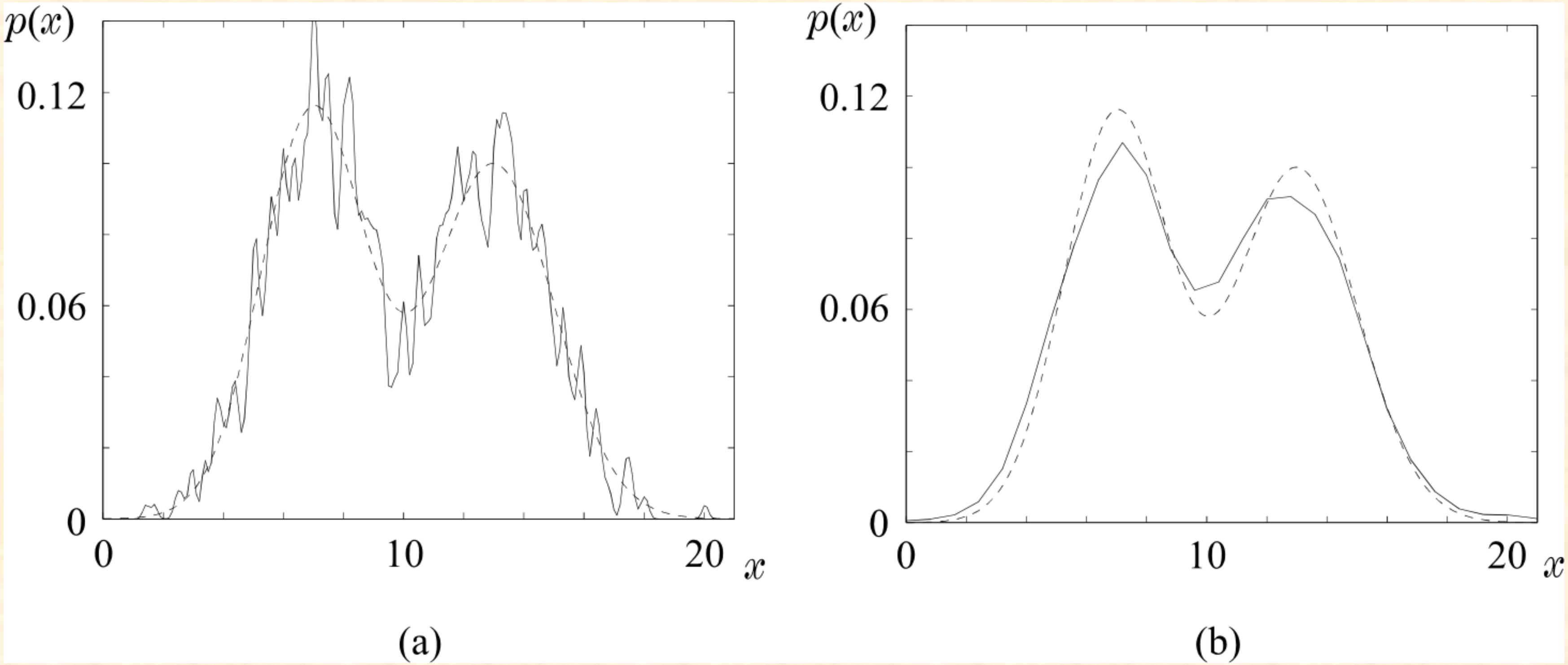
Συνεπώς αμερόληπτος (unbiased) στο όριο

(B) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Parzen windows

- Διασπορά
  - Όσο μικρότερο το *h*, τόσο μεγαλύτερη η διασπορά

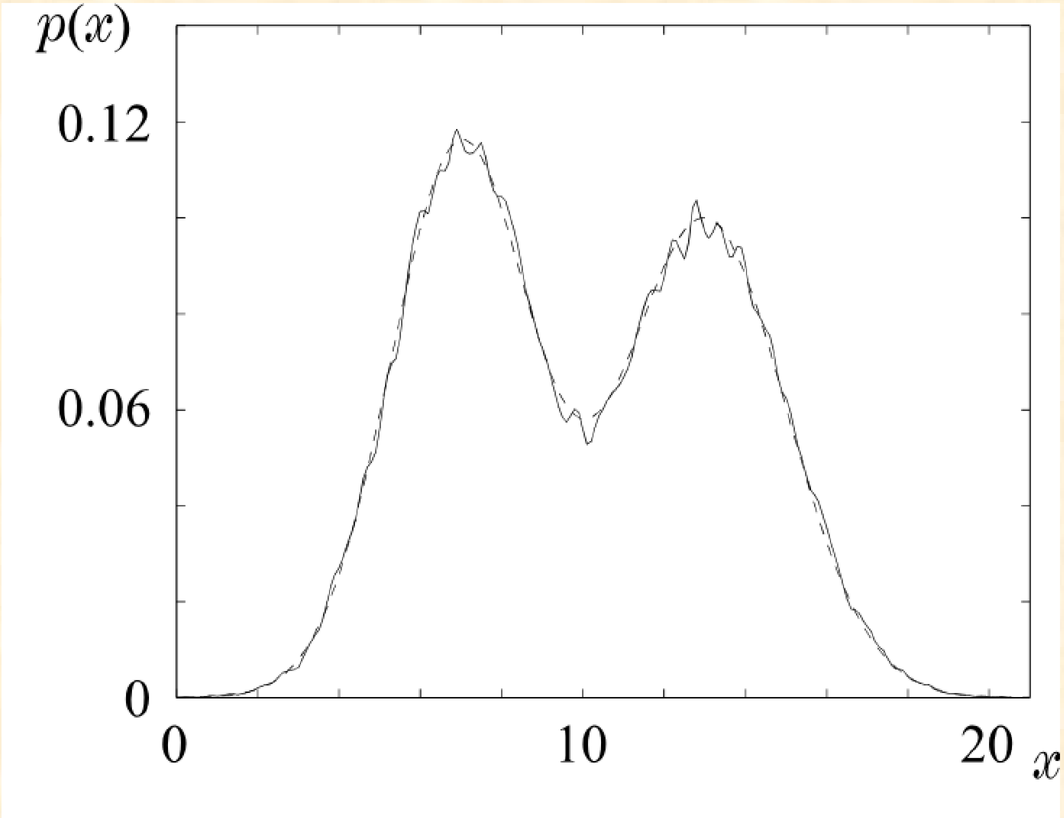
h=0.1, N=1000 *h=0.8, N=1000*



(B) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Parzen windows

h=0.1, N=10000

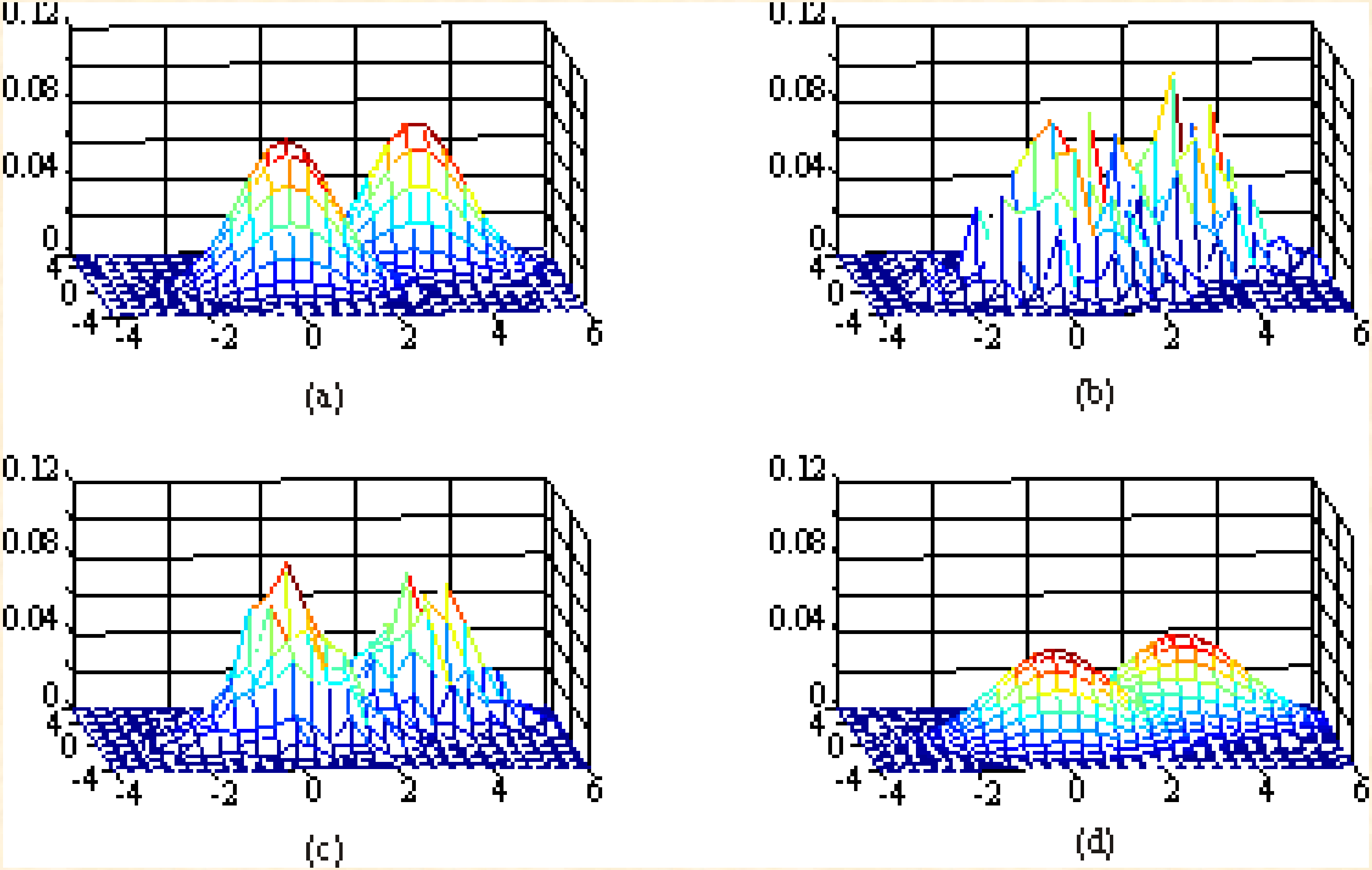


(b)

- Όσο μεγαλύτερο το *N,* τόσο καλύτερη η ακρίβεια

(B) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

H Bayesian περίπτωση – Parzen windows



(a) αρχ. κατανομή, (b) h=0.05, N=1000, (c) h=0.05, N=20000, (d) h=0.8, N=20000

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Parzen windows

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

$\boxed{\text{Ασυμπτωτικά απροκατάληπτος}}$

(B) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Parzen windows

- Έστω $Y=\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$ δεδομένο σύνολο διανυσμάτων που έχουν προκύψει από άγνωστη pdf, $p(\mathbf{x})$.

- **Στόχος:** Εκτίμησε την τιμή της $p(\mathbf{x})$ σε δεδομένο σημείο $\mathbf{x}$ με βάση το σύνολο Y.

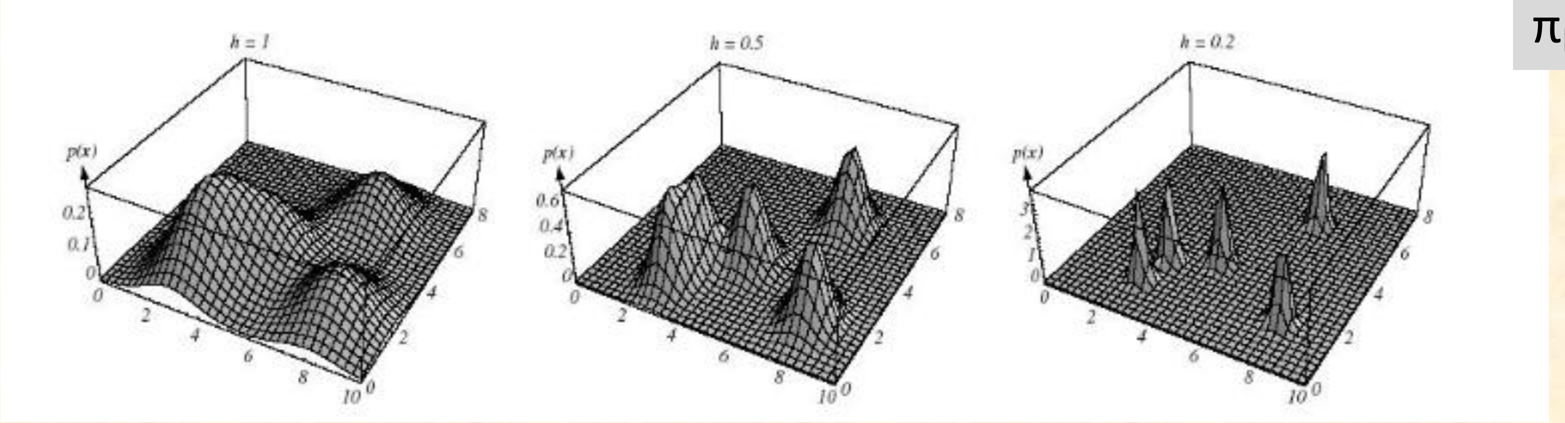
- Η **ιδέα**: Προσέγγισε την $p(\mathbf{x})$ με μια μίξη συναρτήσεων πυρήνων κεντραρισμένων στα $\mathbf{x}_i$s του Y , οι οποίες μειώνονται πολύ γρήγορα γύρω από αυτά.

Μια συνηθισμένη συνάρτηση πυρήνων: Η Gaussian. Έτσι

$$p(x)=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{(2\pi)^{l/2}h^l}\exp\left(-\frac{(x-x_i)^T(x-x_i)}{2h^2}\right)$$

όπου h (η τυπική απόκλιση) καθορίζεται από το χρήστη.

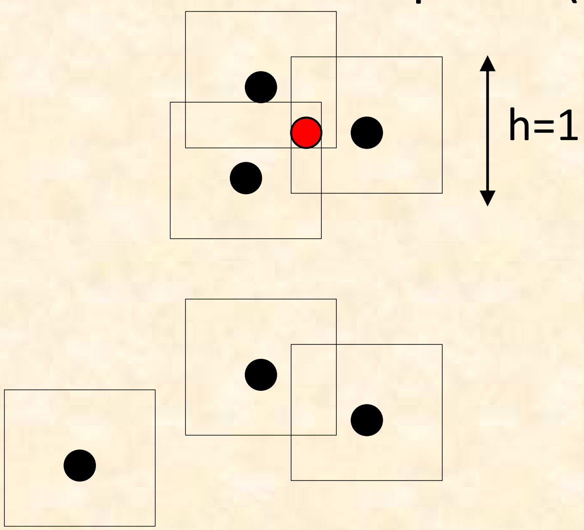
Συγκρίνετε τα Parzen windows με τη μίξη των Gaussians που συζητήθηκε προηγουμένως

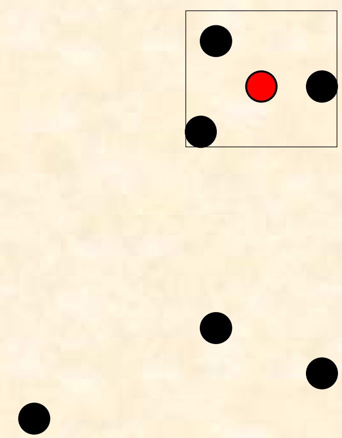


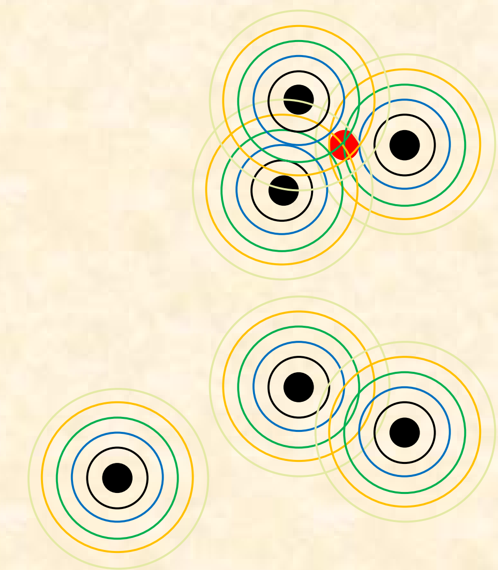
(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Parzen windows

Παράδειγμα: Έστω τα ακόλουθα ***N*=6** σημεία δεδομένων (μαύρες κουκκίδες), τα οποία προέρχονται από μία άγνωστη κατανομή $p(\boldsymbol{x})$. Να εκτιμήσετε την τιμή της συν. πυκν. πιθ. για την **κόκκινη** κουκκίδα, θεωρώντας ότι η $p(\boldsymbol{x})$ είναι το σταθμισμένο άθροισμα N κανονικών κατανομών $N(\boldsymbol{x}_i,h^2)$, i=1,…,6.







- **Κεντράρουμε** μία κανονική κατανομή $N(\boldsymbol{x}_i,h^2)$ σε κάθε σημείο δεδομένων
- Υπολογίζουμε την τιμή κάθε κατανομής στο υπό μελέτη σημείο (κόκκινη κουκίδα)
- Υπολογίζουμε το μέσο όρο των παραπάνω τιμών.

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Parzen windows

Θεωρείστε ένα πρόβλημα c κλάσεων με εμπλεκόμενες κλάσεις τις $\omega_1,\ldots,\omega_c$.

Υπάρχουν N_i διαθέσιμα σημεία από κάθε κλάση ω_i, $i=1,\ldots,c$, με $N_1+\ldots+N_c=N$.

<u>Πρόβλημα:</u> Ταξινομείστε δεδομένο σημείο $\boldsymbol{x}$ σε μία από τις παραπάνω κλάσεις, κανόντας χρήση της τεχνικής των Parzen windows.

-Υπολόγισε τις **συναρτήσεις διάκρισης**, $j=1,\ldots,c$

$$g_j(x) = (2\pi)^{l/2} h^l P(\omega_j) p(x|\omega_j) = \frac{P(\omega_j)}{N_j} \sum_{i=1}^{N_j} \exp\left(-\frac{(x-x_i)^T (x-x_i)}{2h^2}\right)$$

- Καταχώρησε το $\boldsymbol{x}$ στην κλάση ω_q για την οποία $g_q(\boldsymbol{x})=\max_{j=1,\ldots,c} g_j(\boldsymbol{x})$.

-Αν επιπλέον προσεγγίσουμε $P(\omega_j) \approx N_j/N$, $j=1,\ldots,c$ και επανορίσουμε την $g_i(\boldsymbol{x})$ ως

$$g'_j(x) = N(2\pi)^{l/2} h^l P(\omega_j) p(x|\omega_j) = \sum_{i=1}^{N_j} \exp\left(-\frac{(x-x_i)^T (x-x_i)}{2h^2}\right)$$

έχουμε:

καταχώρησε το $\boldsymbol{x}$ στην κλάση ω_q για την οποία

$g_q(\boldsymbol{x})=\max_{j=1,\ldots,c} g_j(\boldsymbol{x}) \Leftrightarrow g'_q(\boldsymbol{x})=\min_{j=1,\ldots,c} g'_j(\boldsymbol{x})$

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Εκτίμηση πυκνότητας *k*-NN

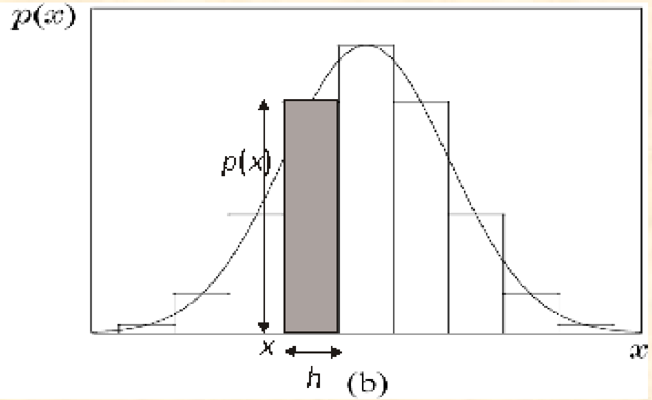
- Έστω $Y=\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$ δεδομένο σύνολο σημείων τα οποία προέρχονται από άγνωστη pdf, $p(\mathbf{x})$.
- **Στόχος:** Έκτίμησε την τιμή της $p(\mathbf{x})$ σε δεδομένο σημείο $\mathbf{x}$, με βάση το σύνολο Y.
- Η **ιδέα:**
- Διάλεξε μια τιμή για το k.
- Για δεδομένο $\mathbf{x}$, στο οποίο θέλουμε να εκτιμήσουμε την $p(\mathbf{x})$, κάνε τα ακόλουθα
    - Ανάμεσα στα σημεία του Y, προσδιόρισε τα k εγγύτερα στο $\mathbf{x}$ σημεία
    - Προσδιόρισε τον όγκο $V(\mathbf{x})$ της υπερσφαίρας με κέντρο το $\mathbf{x}$ και ακτίνα ίση με την απόσταση του x από το μακρύτερο από τα k παραπάνω σημεία.
    - Προσέγγισε την $p(\mathbf{x})$ ως

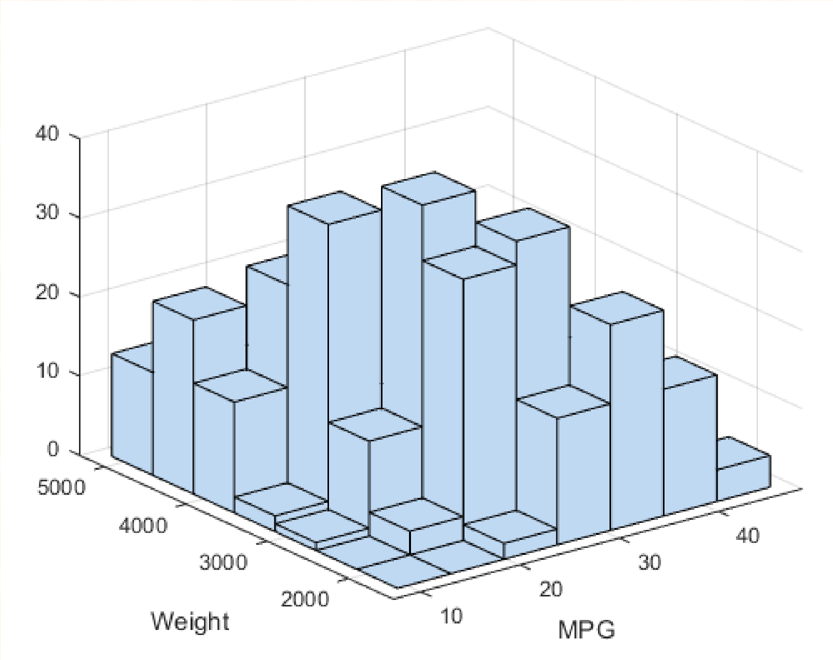
$$p(x) \approx \frac{k}{NV(x)}$$

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Εκτίμηση πυκνότητας *k*-NN

$$p(x) \approx \frac{k}{NV(x)} = \frac{\frac{k}{N}}{V(x)} \approx \frac{P}{V(x)} \Leftrightarrow P \approx p(x)V(x)$$





$$P(x < X < x+h) = \int_{-\infty}^{+\infty} p(y)dy \approx p(x)\cdot h$$

$$P(x < X < x+h) \approx k_n / N$$

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Εκτίμηση πυκνότητας *k*-NN

Θεωρείστε ένα πρόβλημα c κλάσεων, με εμπλεκόμενες κλάσεις $\omega_1,\ldots,\omega_c$.

Υπάρχουν N_i διαθέσιμα σημεία από κάθε κλάση ω_i, $i=1,\ldots,c$, με $N_1+\ldots+N_c=N$.

<u>Πρόβλημα:</u> Ταξινόμησε δεδομένο σημείο $\boldsymbol{x}$ σε μια από τις παραπάνω κλάσεις, χρησιμοποιώντας εκτιμήσεις πυκνότητας με βάση τους k-NN.

-Προσέγγισε $p(\boldsymbol{x}|\omega_j) \approx k/ N_j V_j(\boldsymbol{x})$, $j=1,\ldots,c$.

-Υπολόγισε τις **συναρτήσεις διάκρισης**, $j=1,\ldots,c$

$$g_j(x) = \frac{1}{k} P(\omega_j) p(x|\omega_j) = \frac{1}{k} P(\omega_j) \frac{k}{N_j V_j(x)} = \frac{P(\omega_j)}{N_j V_j(x)}$$

- Καταχώρησε το $\boldsymbol{x}$ στην κλάση ω_q για την οποία $g_q(\boldsymbol{x})=\max_{j=1,\ldots,c} g_j(\boldsymbol{x})$.

-Αν επιπλέον προσεγγίσουμε $P(\omega_j) \approx N_j/N$, $j=1,\ldots,c$ και επανορίσουμε την $g_i(\boldsymbol{x})$ as

$$g'_j(x) = \frac{N}{k} P(\omega_j) p(x|\omega_j) = \frac{N}{k} \frac{N_j}{N} \frac{k}{N_j V_j(x)} = \frac{1}{V_j(x)}$$

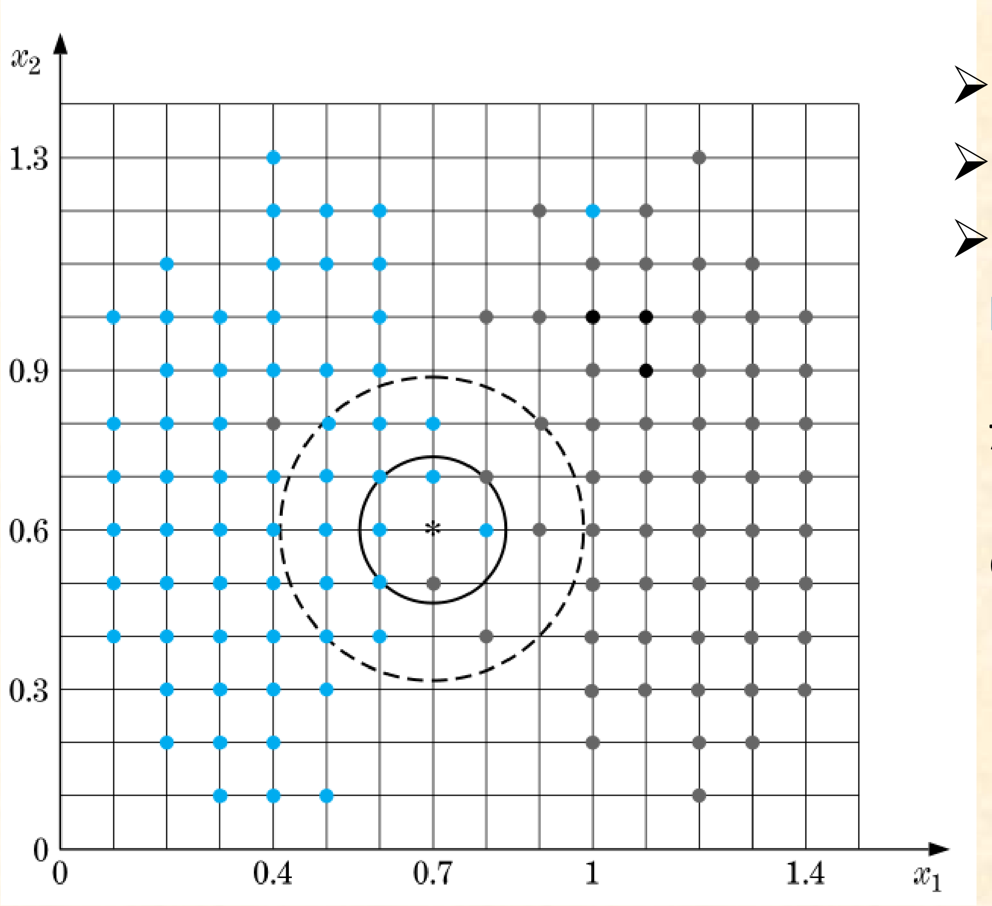
έχουμε: καταχώρησε το $\boldsymbol{x}$ στη κλάση ω_q για την οποία

$g_q(\boldsymbol{x})=\max_{j=1,\ldots,c} g_j(\boldsymbol{x}) \Leftrightarrow V_q(\boldsymbol{x})=\min_{j=1,\ldots,c} V_j(\boldsymbol{x})$

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Η Bayesian περίπτωση – Εκτίμηση πυκνότητας *k*-NN

-Ένα παράδειγμα ταξινόμησης



- ***k*=5**
- **Ισοπίθανες** κλάσεις
- Το σημείο "*" ταξινομείται στη **μπλε κλάση** αφού η επιφάνεια του κύκλου με κέντρο αυτό που περιέχει 5 σημεία από την μπλε κλάσης είναι μικρότερη από την αντίστοιχη για την **μαύρη κλάση**.

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Ο ταξινομητής των k πλησιέστερων γειτόνων (k-nearest neighbor classifier (k-NN))

- Έστω

$$X = \{(x_i, d_i), x_i \in R^l, d_i \in \{1,2,\ldots,M\}, i = 1,\ldots N\}$$

δεδομένο σύνολο που περιέχει διανύσματα από *M* κλάσεις, ω_1,…, ω_M.

- **Στόχος:** Καταχώρησε δεδομένο διάνυσμα ***x*** σε μια από τις παραπάνω κλάσεις, βασιζόμενος στο σύνολο *X*.

- Η **ιδέα**:

- Διάλεξε μία (περιττή) τιμή για το *k*.

- Για δεδομένο ***x***

  - Προσδιόρισε τους *k* πλησιέστερους γείτονες του από το σύνολο *X*.
  - Ανάμεσα στα *k* σημεία έστω k_j εκείνα τα οποία προέρχονται από την *j*-στη κλάση, $j=1,\ldots,M$ (εδώ μπορούμε να ορίσουμε $g_j(\mathbf{x})=k_j$).
  - Καταχώρησε το ***x*** στην κλάση ω_q για την οποία $k_q=\max_{j=1,\ldots,M} k_j$.

Ένα μειονέκτημα: Αυξημένη υπολογιστική πολυπλοκότητα ($O(N)$) για κάθε διάνυσμα

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Ο ταξινομητής των k πλησιέστερων γειτόνων (k-nearest neighbor classifier (k-NN))

Μερικά θεωρητικά αποτελέσματα

Έστω P_B η πιθανότητα λάθους του ταξινομητή Bayes.

-Για τον ταξινομητή του 1-πλησιέστερου γείτονα (1-Nearest neighbor ή, απλά, NN) είναι

$$P_B \le P_{NN} \le P_B\left(2 - \frac{M}{M-1} P_B\right) \le 2P_B$$

-Για τον ταξινομητή k-πλησιέστερων γειτόνων είναι

$$P_B \le P_{kNN} \le P_B + \sqrt{\frac{2P_{NN}}{k}}$$

$$\boxed{k \to \infty,\ P_{kNN} \to P_B}$$

Συσχέτιση του ταξινομητή *kNN* με τον ταξινομητή που βασίζεται στην εκτίμηση της πυκνότητας με βάση τους *k*-πλησιέστερους γείτονες: Είναι ισοδύναμοι για *k*=1 (Γιατί;)

(Β) ΜΗ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ

Ο ταξινομητής των k πλησιέστερων γειτόνων (k-nearest neighbor classifier (k-NN))

<u>Μια γεωμετρική άποψη για τον ταξινομητή του πλησιέστερου γείτονα -NN classifier</u>

Ο **ταξινομητής πλησιέστερου γείτονα** **διαμερίζει** το χώρο R^l σε ***N* περιοχές** (τόσες, όσες είναι και τα στοιχεία του συνόλου *X*) έτσι ώστε

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