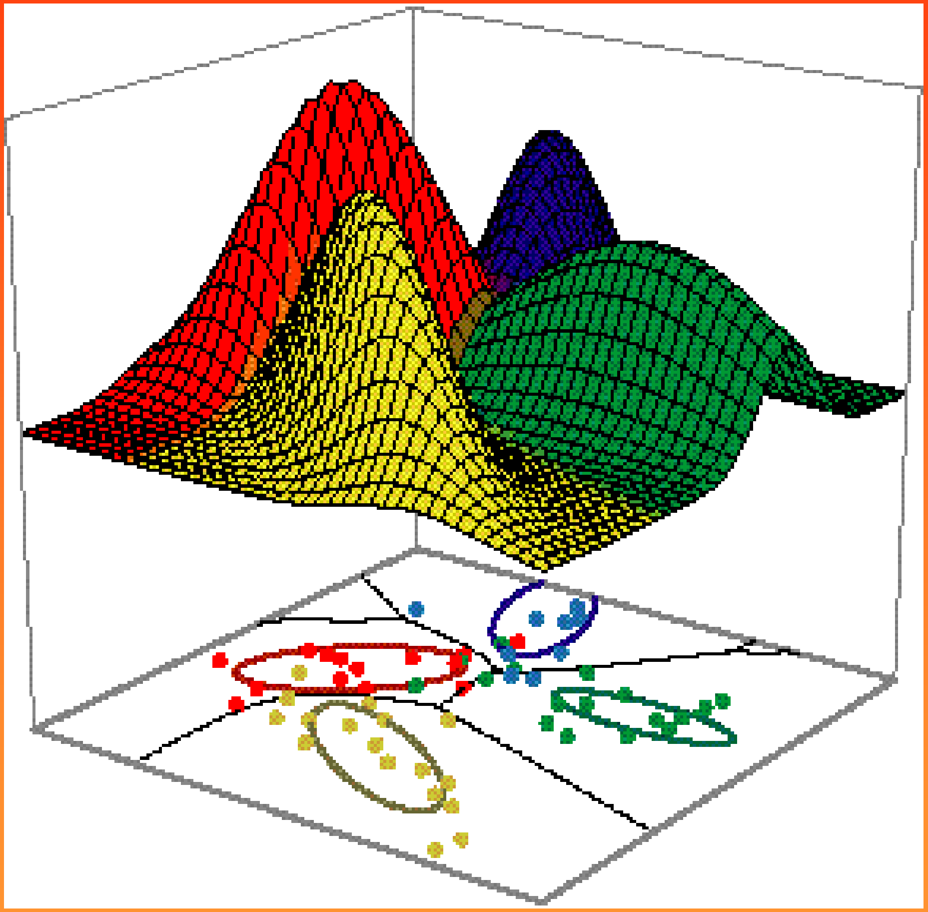


Pattern Classification

All materials in these slides were taken from
Pattern Classification (2nd ed) by R. O. Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2000
with the permission of the authors and the publisher

Non-Parametric Classification (Sections 4.4-4.5)

- Kn –Nearest Neighbor Estimation
- The Nearest-Neighbor Rule

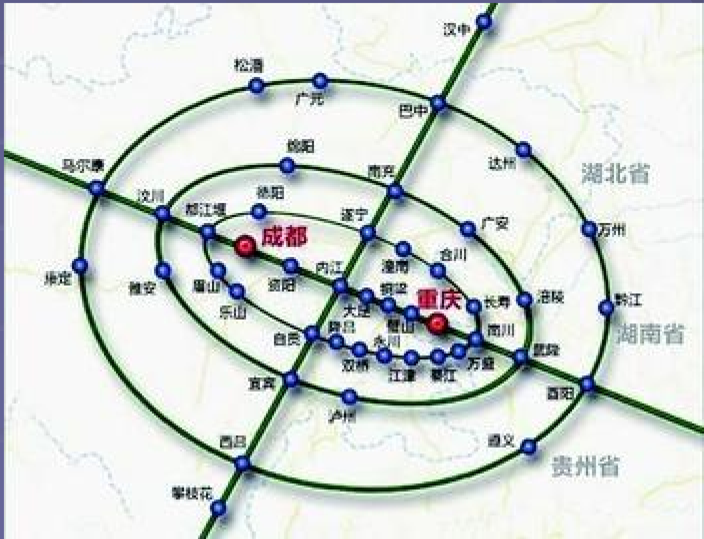
4.4 K_n - Nearest neighbor estimation

- Goal: a solution for the problem of the unknown "best" window function
  - Let the cell volume be a function of the training data (k_n)
  - Center a cell about x and let it grow until it captures k_n samples
    ($k_n = f(n)$)
  - k_n are called the k_n nearest-neighbors of x

2 possibilities can occur:

- Density is high near *x*; therefore the cell will be small which provides a good resolution
- Density is low; therefore the cell will grow large and stop until higher density regions are reached

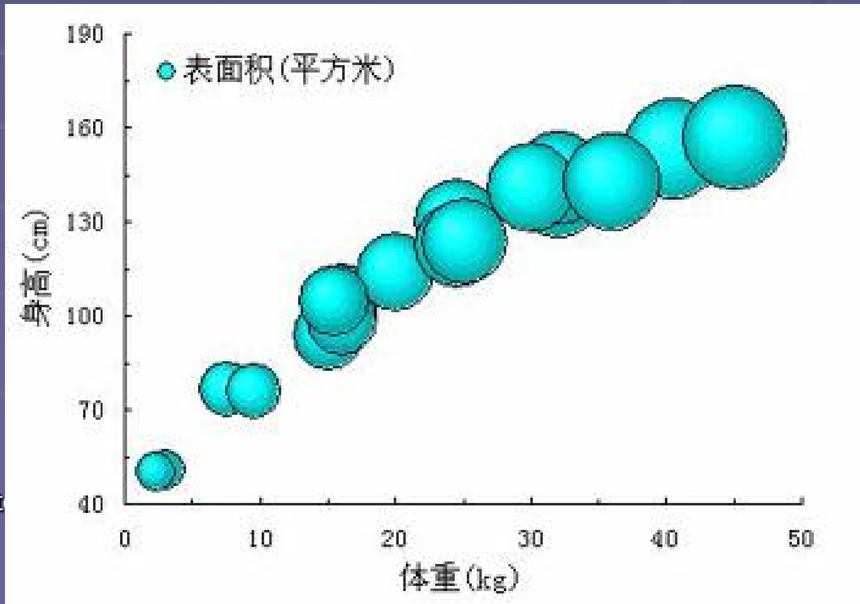
We can obtain a family of estimates by setting $k_n=k_1\sqrt{n}$ and choosing different values for k_1

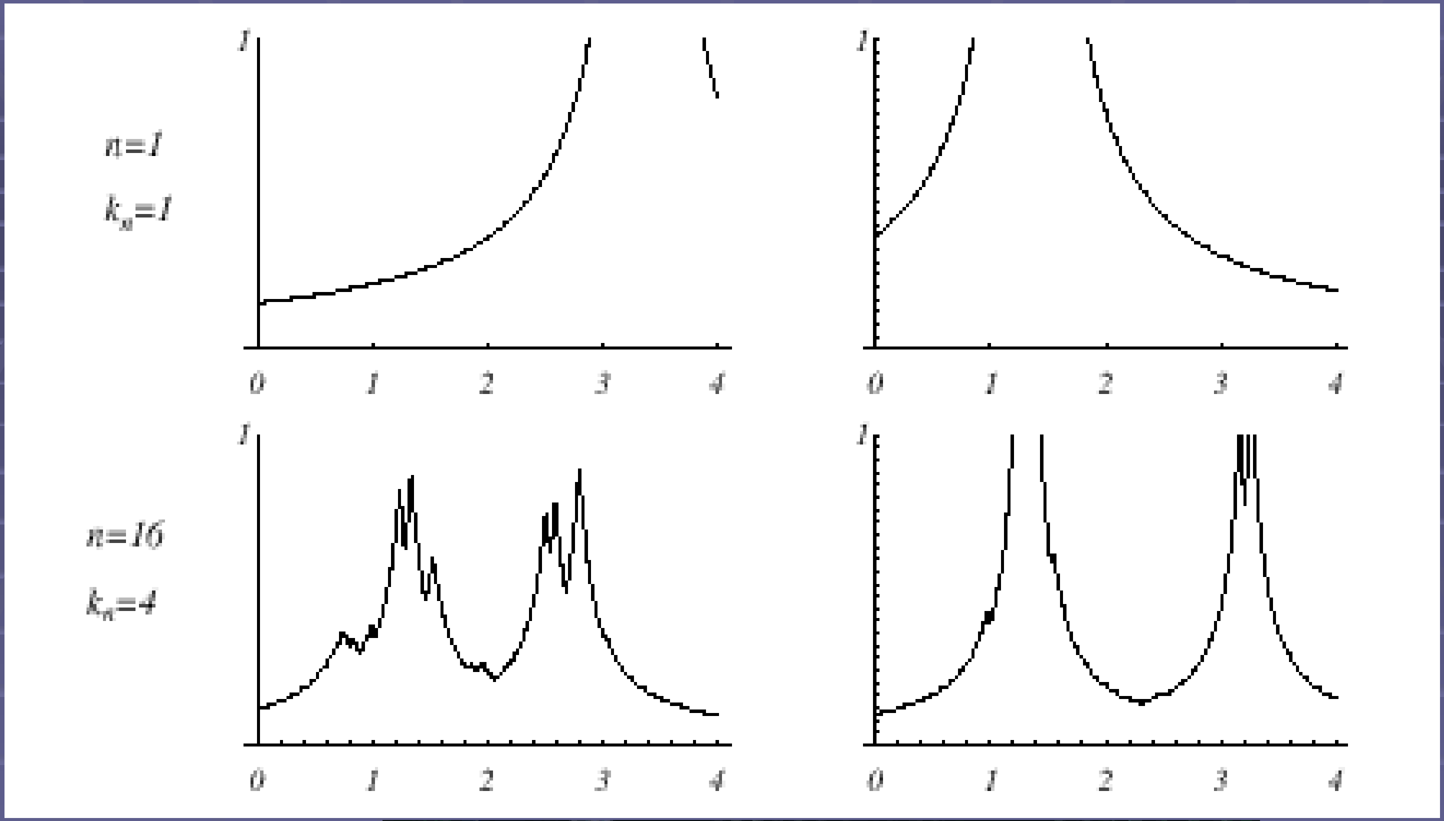


Illustration

For $k_n = \sqrt{n} = 1$; the estimate becomes:

$$P_n(x) = k_n / n / V_n = 1 / V_1 = 1 / 2|x-x_1|$$




n=1
k_n=1
n=16
k_n=4
1
0 1 2 3 4

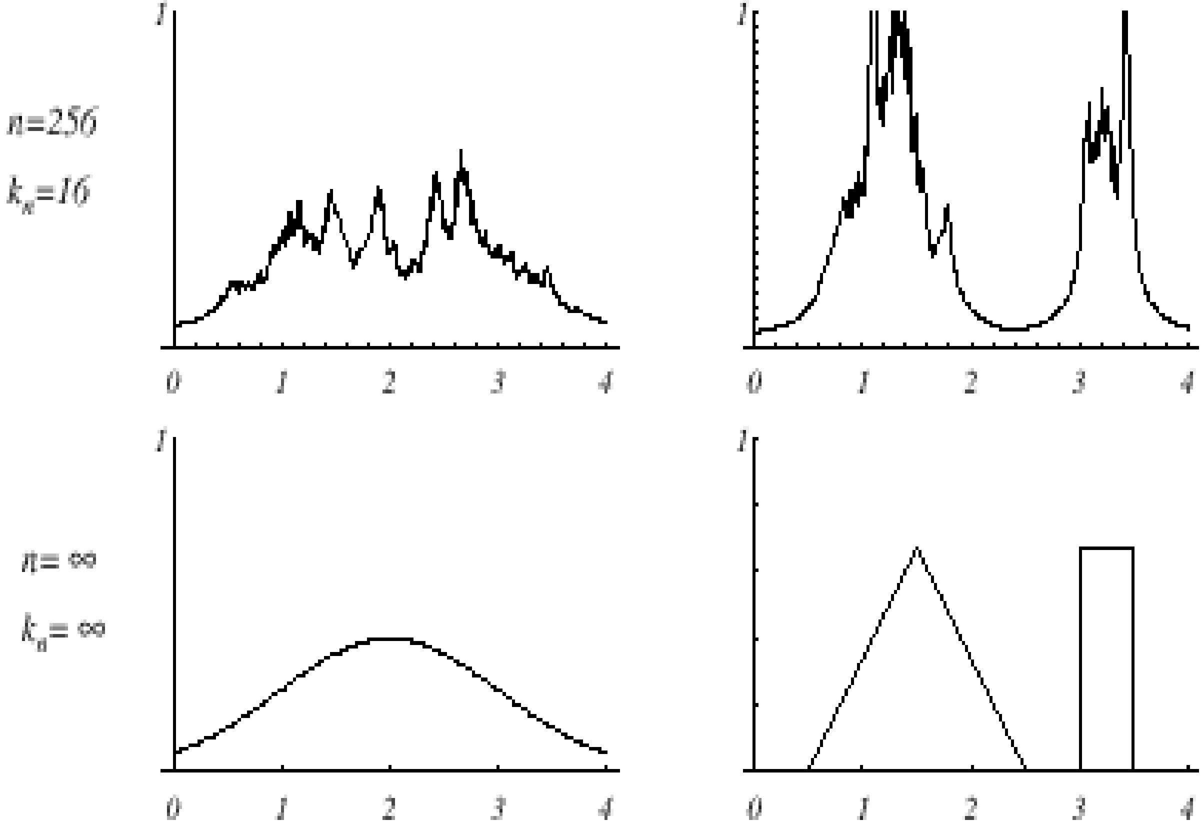
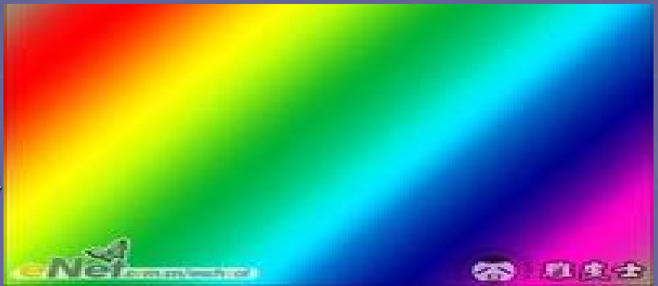


FIGURE 4.12. Several *k*-nearest-neighbor estimates of two unidimensional densities: a Gaussian and a bimodal distribution. Notice how the finite *n* estimates can be quite "spiky." From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*.



- Estimation of a-posteriori probabilities
  - Goal: estimate $P(\omega_i \mid x)$ from a set of n labeled samples
    - Let’s place a cell of volume V around x and capture k samples
    - k_i samples amongst k turned out to be labeled by ω_i then:

$$p_n(x, \omega_i) = k_i /(nV)$$

An estimate for $p_n(\omega_i \mid x)$ is:

$$p_n(\omega_i \mid x) = \frac{p_n(x, \omega_i)}{\sum_{j=1}^{j=c} p_n(x, \omega_j)} = \frac{k_i}{k}$$

- k_i/k is the fraction of the samples within the cell that are labeled as ω_i

- For minimum error rate, the most frequently represented category within the cell is selected

- If k is large and the cell sufficiently small, the performance will approach the best value.

4.5 The Nearest-Neighbor Rule

- The nearest –neighbor rule
  - Let $D_n = \{x_1, x_2, ..., x_n\}$ be a set of n labeled prototypes
  - Let $x' \in D_n$ be the closest prototype to a test point x then the nearest-neighbor rule for classifying x is to assign it the label associated with x'
  - The nearest-neighbor rule leads to an error rate greater than the minimum possible value of the Bayes rate
  - If the number of prototypes is large (unlimited), the error rate of the nearest-neighbor classifier is never worse than twice the Bayes rate (it can be demonstrated!)
  - If $n \to \infty$, it is always possible to find x' sufficiently close so that:

$$P(\omega_i \mid x') \cong P(\omega_i \mid x)$$

$$P(\omega_m \mid x) = \max_i P(\omega_i \mid x)$$

$$P^*(e \mid x) = 1 - P(\omega_m \mid x)$$

- If $P(\omega_m \mid x) \cong 1$, then the nearest neighbor selection is almost always the same as the Bayes selection
- Convergence of the Nearest Neighbor

$$P^*(e \mid x) = 1 - P(\omega_m \mid x)$$

$$P^*(e) = \int P^*(e \mid x) p(x) dx$$

$$P(e) = \lim_{n \to \infty} P_n(e)$$

$$P^*(e) \le P(e) \le P^*(e)\left(2 - \frac{c}{c-1} P^*(e)\right)$$

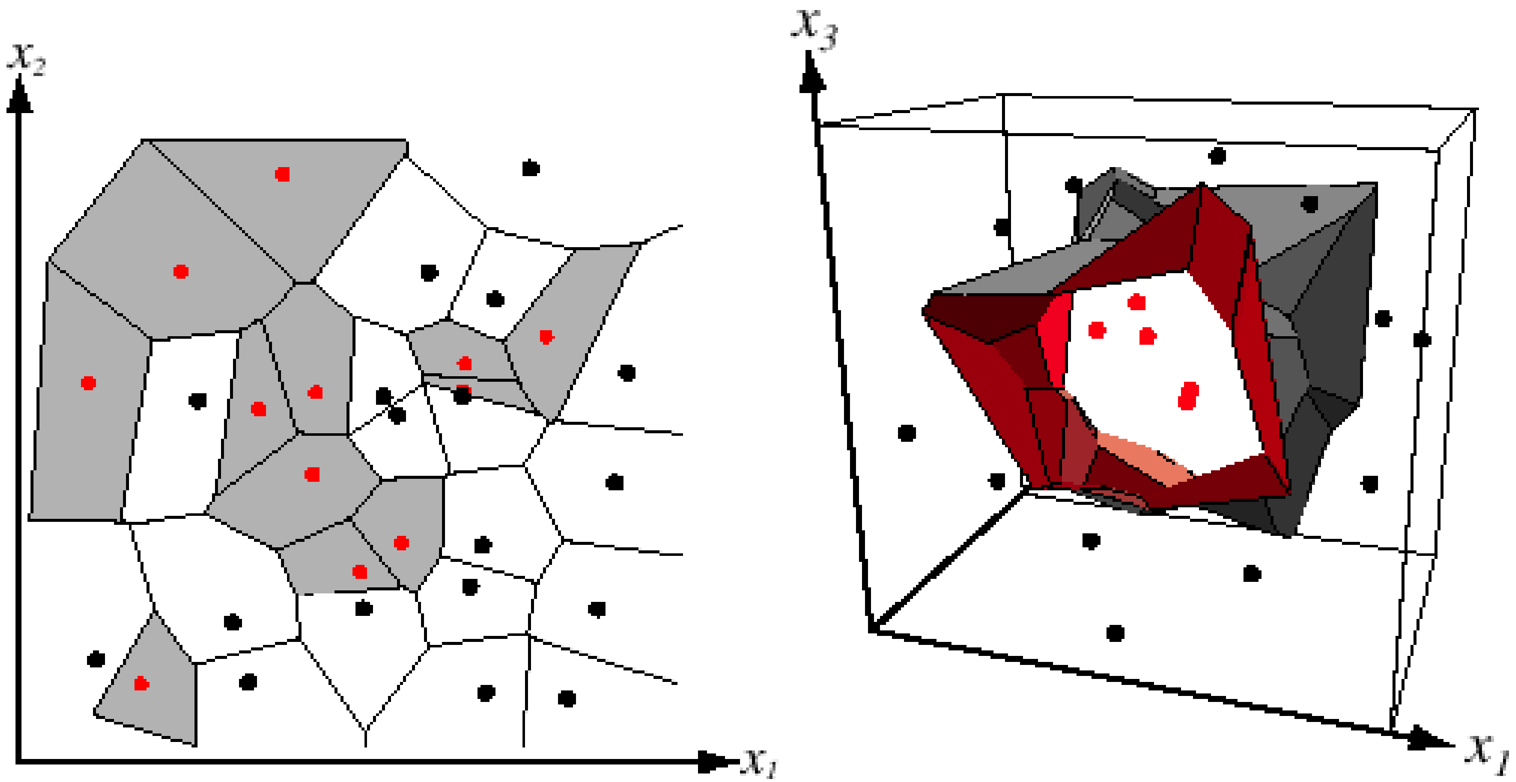


FIGURE 4.13. In two dimensions, the nearest-neighbor algorithm leads to a partitioning of the input space into Voronoi cells, each labeled by the category of the training point it contains. In three dimensions, the cells are three-dimensional, and the decision boundary resembles the surface of a crystal. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- The k – nearest-neighbor rule
  - Goal: Classify x by assigning it the label most frequently represented among the k nearest samples and use a voting scheme

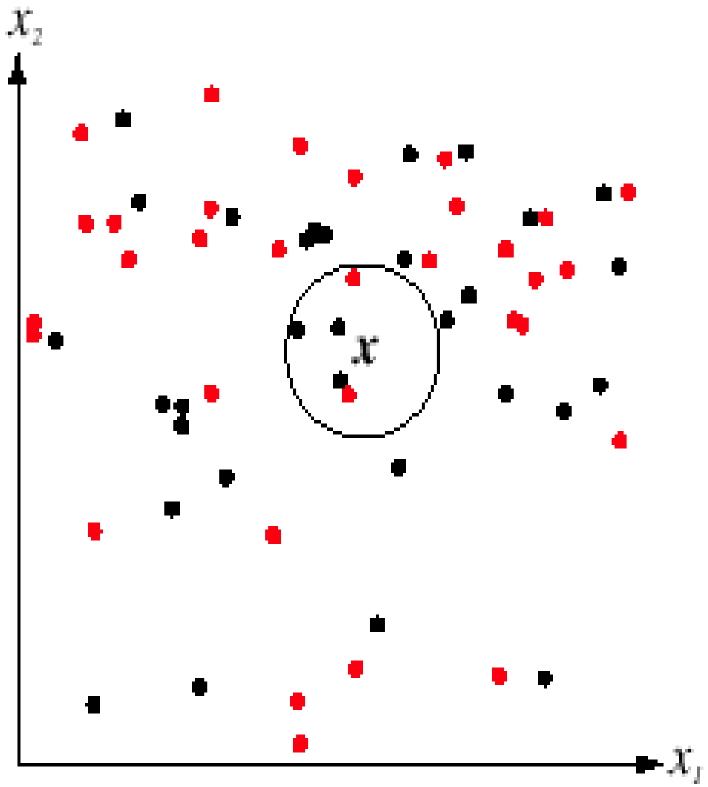


FIGURE 4.15. The k-nearest-neighbor query starts at the test point $\mathbf{x}$ and grows a spherical region until it encloses k training samples, and it labels the test point by a majority vote of these samples. In this $k = 5$ case, the test point $\mathbf{x}$ would be labeled the category of the black points. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*.

Example:

$k = 3$ (odd value) and $x = (0.10, 0.25)^t$

Prototypes	Labels
(0.15, 0.35)	ω_1
(0.10, 0.28)	ω_2
(0.09, 0.30)	ω_5
(0.12, 0.20)	ω_2

Closest vectors to x with their labels are:

$\{(0.10, 0.28, \omega_2); (0.12, 0.20, \omega_2); (0.15, 0.35, \omega_1)\}$

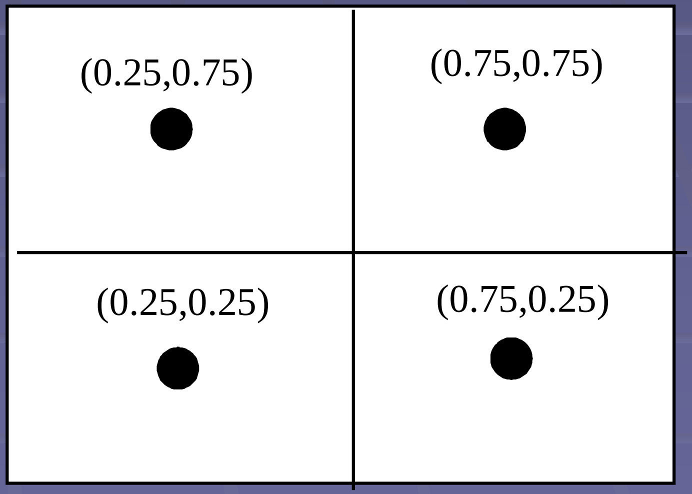
One voting scheme assigns the label ω_2 to x since ω_2 is the most frequently represented

- Reducing the computational complex in nearest-neighbor search
  - Computing partial distance

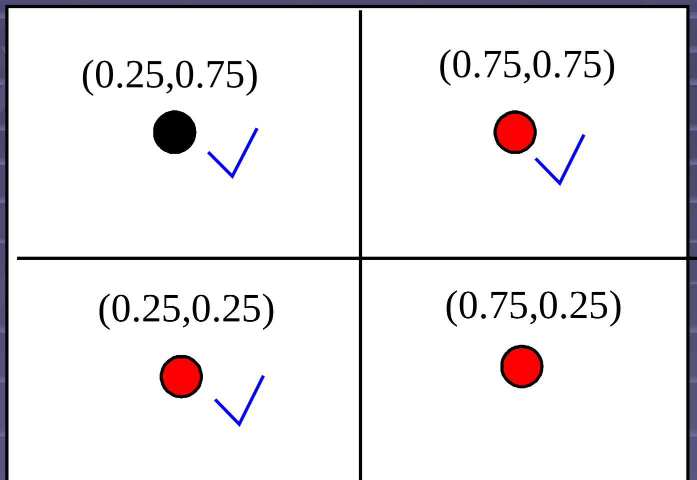
$$D_r(a,b) = \left(\sum_{k=1}^{r}(a_k - b_k)^2\right)^{1/2}$$

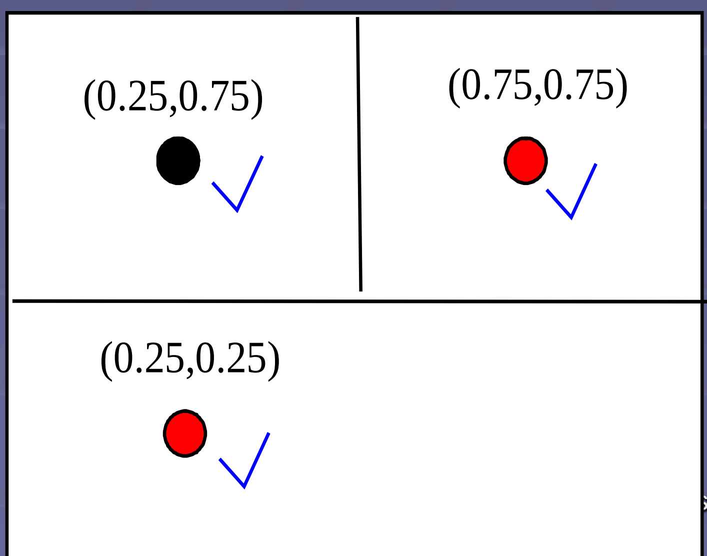
  - Create some form of search tree, for example

$$p(x) \sim U\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix}\right)$$



- Nearest-Neighbor Editing





4.6 Metrics

- Properties of Metrics
  - Nonnegativity:D(a,b)>=0
  - Reflexivity:D(a,b)=0 if and only if a=b
  - Symmetry:D(a,b)=D(b,a)
  - *Euclidean* formula for distance in d dimensions

$$D(a,b) = \left(\sum_{k=1}^{d} (a_k - b_k)^2\right)^{1/2}$$

Minkowski metric for distance in d dimensions

$$L_k(a,b) = \left(\sum_{k=1}^{d} (a_k - b_k)^k\right)^{1/k}$$