EE 6885 Statistical Pattern Recognition

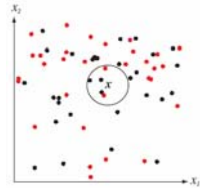
Fall 2005
Prof. Shih-Fu Chang
http://www.ee.columbia.edu/~sfchang

Lecture 10 (10/12/05)

- Reading
  - Nearest Neighbor Estimation, Distance Metrics
    - DHS Chap. 4.4-4.5, 4.6
    - Reference Book HTF Chap. 11.1-11.3
- Midterm Exam
  - Oct. 24th 2005 Monday 1pm-2:30pm (90mins)
    - Open books/notes, no computer

k_n-Nearest-Neighbor

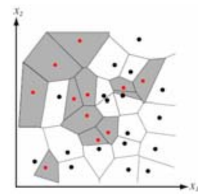
$$p_n(x) \simeq \frac{k_n / n}{V_n}$$



- For classification, estimate $p(x)$ for each class ω_i

$$p_n(x, \omega_i) = \frac{k_i / n}{V}$$

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

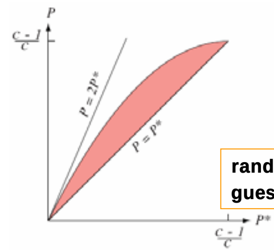


Decision boundary

- Performance bound of 1-nearest neighbor (Cover & Hart '67)

$$P^* \le \lim_{n \to \infty} P_n(e) \le P^*(2 - \frac{c}{c-1} P^*)$$

$$P^*(e \mid x) = 1 - \max_i P(\omega_i \mid x) \quad P^* = \int P^*(e \mid x) p(x) dx$$



random guess

Deriving the error bound ...

Assume n samples : $(x_1, \theta_1), (x_2, \theta_2), \ldots, (x_n, \theta_n)$

Assume x_n' is the nearest neighbor to x

Assume i.i.d.

$$P_n(e \mid x, x_n') = 1 - \sum_{i=1}^{c} P(\theta = \omega_i, \theta_n' = \omega_i \mid x, x_n') = 1 - \sum_{i=1}^{c} P(\omega_i \mid x) P(\omega_i \mid x_n')$$

assume $p(x_n')$ peaks at x

$$\lim_{n \to \infty} P_n(e \mid x) = \lim_{n \to \infty} \int P_n(e \mid x, x_n') p(x_n') dx_n' = \lim_{n \to \infty} \int P_n(e \mid x, x_n') \delta(x_n' - x) dx_n'$$

$$= \int \left[1 - \sum_{i=1}^{c} P(\omega_i \mid x) P(\omega_i \mid x_n') \right] \delta(x_n' - x) dx_n' = 1 - \sum_{i=1}^{c} P^2(\omega_i \mid x)$$

$$P = \lim_{n \to \infty} P_n(e) = \lim_{n \to \infty} \int P_n(e \mid x) p(x) dx = \int [1 - \sum_{i=1}^{c} P^2(\omega_i \mid x)] p(x) dx$$

- We are interested in relation between P & P^* (the min. error prob.)

$$P^* = \int P^*(e \mid x) p(x) dx \qquad P^*(e \mid x) = 1 - \max_i P(\omega_i \mid x) = 1 - P(\omega_m \mid x)$$

Deriving the 1-NN error bound (cont.)

- We are interested in relation between P & P^* (the min. error prob.)

$$P = \int [1 - \sum_{i=1}^{c} P^2(\omega_i \mid x)] p(x) dx$$

$\boxed{\text{Let's fix } P(\omega_m \mid x)\text{, i.e., fix } P^*}$

$$\sum_{i=1}^{c} P^2(\omega_i \mid x) \text{ is minimized when } \quad P(\omega_i \mid x) \text{ are equal } \forall\ i \neq m$$

$$\text{namely } P(\omega_i \mid x) = \begin{cases} P(\omega_m \mid x) & i = m \\ \dfrac{1 - P(\omega_m \mid x)}{c-1} & i \neq m \end{cases} = \begin{cases} 1 - P^*(e \mid x) & i = m \\ \dfrac{P^*(e \mid x)}{c-1} & i \neq m \end{cases}$$

$$\Rightarrow \sum_{i=1}^{c} P^2(\omega_i \mid x) \ \geq (1 - P^*(e \mid x))^2 + \frac{P^{*2}(e \mid x)}{c-1}$$

$$\Rightarrow 1 - \sum_{i=1}^{c} P^2(\omega_i \mid x) \ \leq 2P^*(e \mid x) - \frac{c}{c-1} P^{*2}(e \mid x)$$

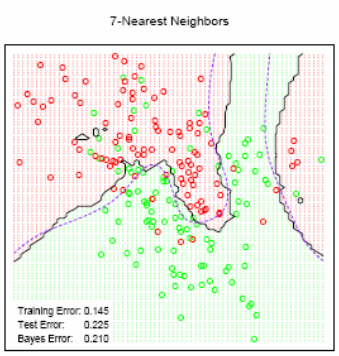
$$\because \int P^{*2}(e \mid x) p(x) dx \geq \left[\int P^*(e \mid x) p(x) dx \right]^2 = P^{*2}$$

$$\Rightarrow P = \int [1 - \sum_{i=1}^{c} P^2(\omega_i \mid x)] p(x) dx \ \leq 2P^* - \frac{c}{c-1} P^{*2}$$

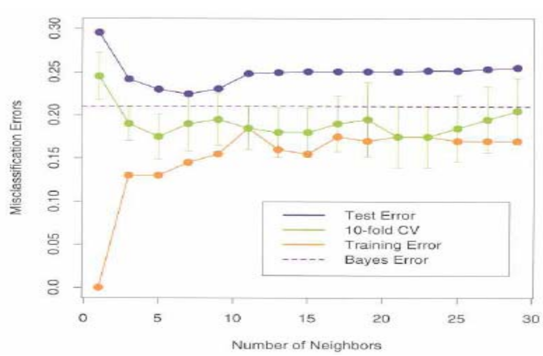
Q.E.D.

K-NN example (Ref. HTF Chap 13)

- **Two Classes, data in each class generated by Gaussian Mixtures**



- **Cross-validation performance**



Reduce Complexity by Clustering

- Training data from each class

3 classes from GMM

- Apply K-Means clustering to each class
- K-means clustering
  - Randomly select K prototypes
  - Map samples to the closest prototype (hard decision)

$x_1, x_2, \ldots, x_N$ *samples*

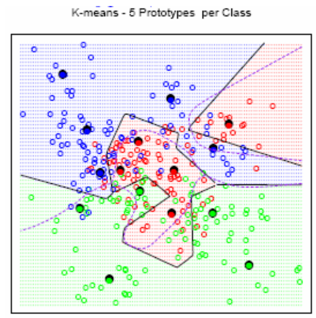
for $i=1,2,\ldots,N,$

$\quad x_i \rightarrow C_k, \text{if } \; Dist(x_i, C_k) < Dist(x_i, C_{k'}), k \neq k'$

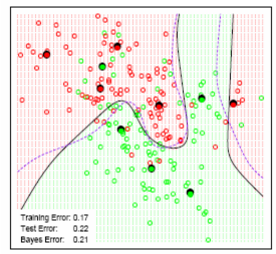
end

  - Re-compute the prototypes

- Use only cluster prototypes in nearest neighbor classification



Comparison with GMM



Learning Vector Quantization (LVQ)

- Learn the prototypes jointly
- Find K prototypes for each class

$m_1(j), m_2(j), \ldots, m_K(j), \;\; j=1,2,\ldots,c$

- Randomly sample data x

find the closest prototype $m_k(j)$

if class label of $x \; = j,$

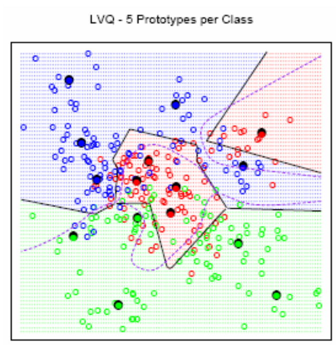
then move prototype $m_k(j)$ closer to x

$$\boxed{m_k(j) \leftarrow m_k(j) + \varepsilon(x - m_k(j))}$$

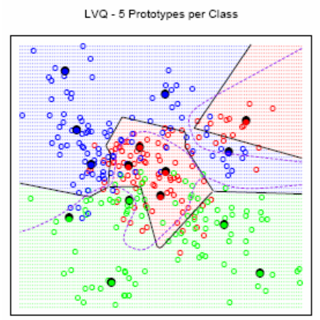
otherwise, move ptotype away from x

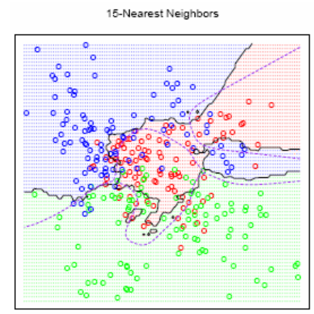
$$\boxed{m_k(j) \leftarrow m_k(j) - \varepsilon(x - m_k(j))}$$

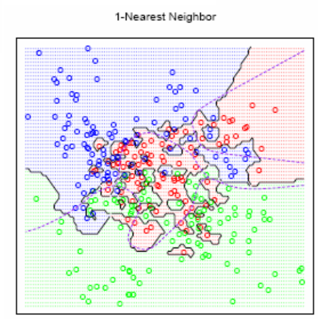
- Repeat the above step, with the learning rate ε decreasing to 0



Comparing LVQ with KNN







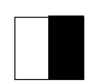
Toy problems for comparison

10-dimensional features in the unit hypercube

$x = \{x_1, x_2, \ldots, x_{10}\}$, x_i uniformly distributed in $[0,1]$

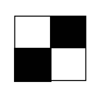
100 training samples, 1000 test samples

- Easy problem class label $Y = I(x_1 > 0.5)$ **hyperplane**



- Difficult problem

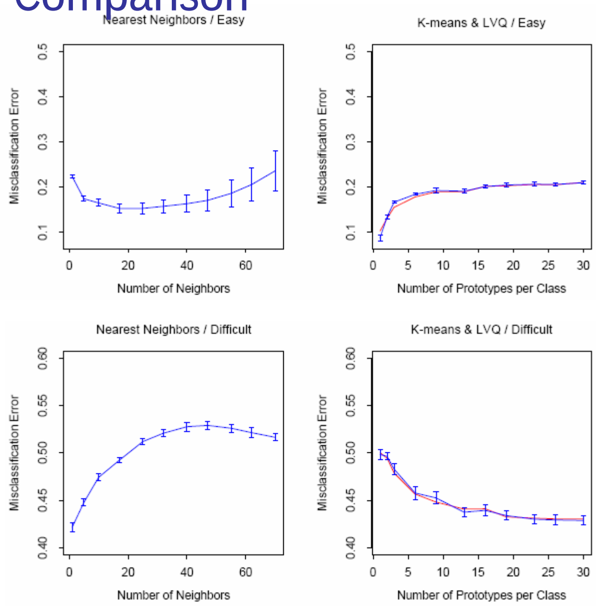
class label $Y = I(sign\left\{\prod_{i=0}^{3}(x_i - 0.5)\right\} > 0)$ **checkerboard**



- What’s the Bayesian Error Rate?

Performance Comparison

- Easy problem
- Difficult problem
- Observations?

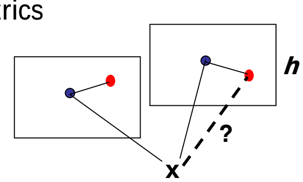


Distance Metrics

- Nearest neighbor rules need distance metrics
- Required properties of a metric

1. non-negativity: $D(a,b) \geq 0$
2. reflexivity: $D(a,b) = 0$ iff $a = b$
3. symmetry: $D(a,b) = D(b,a)$
4. trangular inequality: $D(a,b) + D(b,c) \geq D(c,a)$

$$D(a,b) \geq D(c,a) - D(b,c)$$



useful in indexing

- Minkowski Metric
  - Euclidean
  - Manhattan
  - L_∞

$$L_k(a,b) = \left(\sum_{i=1}^{d} |a_i - b_i|^k\right)^{1/k}$$

- Tanimono Metric
  - sets of elements
  - Point-point distance not useful

$$D_{\text{tanimono}}(S_1, S_2) = \frac{n_1 + n_2 - 2n_{12}}{n_1 + n_2 - n_{12}} = \frac{(n_1 - n_{12}) + (n_2 - n_{12})}{n_1 + n_2 - n_{12}}$$