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**<sub>n</sub>-Nearest-Neighbor

\[ p_n(x) = \frac{k}{n} \]

- For classification, estimate \( p(x) \) for each class \( \omega_i \)
  \[ p_n(x, \omega_i) = \frac{k}{n} \]
  \[ p_n(\omega_i | x) = \frac{p_n(x, \omega_i)}{\sum_{j=1}^{n} p_n(x, \omega_j)} = \frac{k}{k} \]

- Performance bound of 1-nearest neighbor (Cover & Hart '67)
  \[ P^* \leq \lim_{n \to \infty} P_n(e) \leq P^* (2 - \frac{c}{c-1}) \]
  \[ P^* = \int P^*(e | x) p(x) dx \]

---

**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 \( x \)

\[ P_n(e | x, x'_n) = 1 - \sum_{i=1}^{c} P(\theta = \omega_i, \theta'_n = \omega_i | x, x'_n) = 1 - \sum_{i=1}^{c} P(\omega_i | x) P(\omega_i | x'_n) \]

\[ \lim_{n \to \infty} P_n(e | x) = \lim_{n \to \infty} \int P_n(e | x, x'_n) p(x'_n) dx'_n = \lim_{n \to \infty} \int P_n(e | x, x'_n) \delta(x'_n - x) dx'_n \]

\[ = \int \left[ 1 - \sum_{i=1}^{c} P(\omega_i | x) P(\omega_i | x'_n) \right] \delta(x'_n - x) dx'_n = 1 - \sum_{i=1}^{c} P^2(\omega_i | x) \]

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

- We are interested in relation between \( P \) \\& \( P^* \) (the min. error prob.)
  \[ P^* = \int P^*(e | x) p(x) dx \]
  \[ P^*(e | x) = 1 - \max_i P(\omega_i | x) = 1 - P(\omega_n | x) \]
Deriving the 1-NN error bound (cont.)

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

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

Let's fix \( P(\omega_m \mid x) \), i.e., fix \( P' \)

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

namely \( P(\omega_i \mid x) = \begin{cases} 
\frac{P(\omega_m \mid x)}{C-1} & i = m \\
1 - \frac{P(\omega_m \mid x)}{C-1} & i \neq m
\end{cases} \)

\[
\Rightarrow \sum_{i=1}^{C} P^2(\omega_i \mid x) \geq (1 - P'(e \mid x))^2 + \frac{P^2(e \mid x)}{C-1}
\]

\[
\Rightarrow 1 - \sum_{i=1}^{C} P^2(\omega_i \mid x) \leq 2P'(e \mid x) - \frac{C}{C-1} P^2(e \mid x)
\]

\[
\Rightarrow \int P^2(e \mid x) p(x) dx \geq \left( \int P'(e \mid x) p(x) dx \right)^2 = P'^2
\]

\[
\Rightarrow P = \int [1 - \sum_{i=1}^{C} P^2(\omega_i \mid x)] p(x) dx \leq 2P' - \frac{C}{C-1} P'^2 \quad \text{Q.E.D.}
\]

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

\[ x_1, x_2, ..., x_N \text{ samples} \]
for \( i = 1, 2, ..., N \),
\[ x_i \rightarrow C_k, \text{if } \text{Dist}(x_i, C_k) < \text{Dist}(x_i, C_{k'}) \text{,} \quad k \neq k' \]
- Re-compute the prototypes
- Use only cluster prototypes in nearest neighbor classification

Learning Vector Quantization (LVQ)

- Learn the prototypes jointly
- Find K prototypes for each class
  \( m_1(j), m_2(j), ..., m_k(j) \), \( j = 1, 2, ..., c \)
- 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 \( x \)
    \[ m_k(j) \leftarrow m_k(j) + \varepsilon (x - m_k(j)) \]
  - otherwise, move prototype away from \( x \)
    \[ m_k(j) \leftarrow m_k(j) - \varepsilon (x - m_k(j)) \]
- Repeat the above step, with the learning rate \( \varepsilon \) decreasing to 0
Comparing LVQ with KNN

Toy problems for comparison

10-dimensional features in the unit hypercube

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

100 training samples, 1000 test samples

- Easy problem
  
  class label \( Y = I(x_i > 0.5) \)  

- Difficult problem
  
  class label \( Y = I(\text{sign}\left(\prod_{i=0}^{3} (x_i - 0.5)\right) > 0) \)

- 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. triangular inequality: \( D(a, b) + D(b, c) \geq D(c, a) \)

- Minkowski Metric
  - Euclidean
  - Manhattan
  - \( L_\infty \)

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

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

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