EE 6885 Statistical Pattern Recognition

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

Lecture 11 (10/17/05)

- Reading
  - Distance Metrics
    - DHS Chap. 4.6
  - Linear Discriminant Functions
    - DHS Chap. 5.1-5.4
- Midterm Exam
  - Oct. 24th 2005 Monday 1pm-2:30pm (90mins)
    - Open books/notes, no computer
- Review Class
  - Oct. 21st Friday 4pm. Location TBA

k_n-Nearest-Neighbor

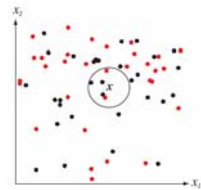
- For classification, estimate $p_n(\omega_i \mid x)$ for each class ω_i

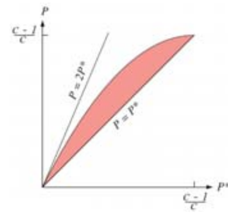
$$p_n(\omega_i \mid x) = \frac{p_n(x,\omega_i)}{\sum_{j=1}^{c} p_n(x,\omega_j)} = \frac{k_i}{k}$$

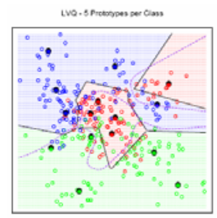
- Performance bound of 1-nearest neighbor (Cover & Hart '67)

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

- Combine K-NN with clustering
  - K-Means, LVQ, GMM
  - Reduce complexity
    - When K increases, complexity?
  - Smooth decision boundaries





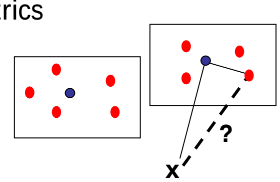


Distance Metrics

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

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

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



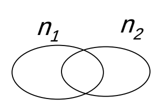
useful in indexing

- 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

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



Discriminant Functions Revisited

define discriminant function $g_i(x)$ for class ω_i

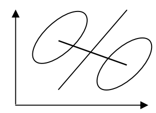
map x to class ω_i if $g_i(x) \geq g_j(x)\ \forall j \neq i$

e.g., $g_i(x) = \ln P(x \mid \omega_i) + \ln P(\omega_i)$ **MAP classifier**

- Gaussian Case: $P(x \mid \omega_i) = N(\mu_i, \Sigma_i)$

$$P(\mathbf{x} \mid w_i) = \frac{1}{(2\pi)^{d/2}\sqrt{|\boldsymbol{\Sigma}_i|}} \exp\left(\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right)$$

- Case I: $\Sigma_i = \Sigma$



$g_i(x) = w_i^t x + w_{i0}$ a hyperplane with *bias* w_{i0}

- Case II: $\Sigma_i =$ arbitrary

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

$$g_i(x) = x^t W_i x + w_i^t x + w_{i0}$$

Decision boundaries may be Hyperplane, hypersphere, hyperellipsoid, hyperparaboloids, hyperhyperboloids

Discriminant Functions (Chap. 5)

- Directly define discriminant functions
  - Without assuming parameter distribution functions for $P(x \mid \omega_i)$
  - Easy to derive useful classifiers
- Linear Functions

$g(\mathbf{x}) = \mathbf{w}^t\mathbf{x} + w_0$,**w**: weight vector, w_0 : bias

- Two-Category Case

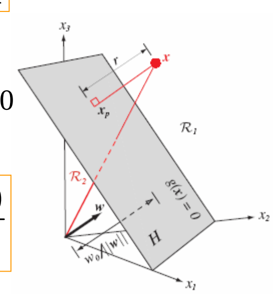
map $\mathbf{x}$ to class ω_1 if $g(\mathbf{x})>0$, otherwise class ω_2

Decision surface H: $g(\mathbf{x}) = 0$

$$\mathbf{x} = \mathbf{x}_p + r \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

$\mathbf{x}_p$: projection of $\mathbf{x}$ onto H, $g(\mathbf{x}_p) = 0$

r : distance from $\mathbf{x}$ to H



$$g(\mathbf{x}) = g\left(\mathbf{x}_p + r \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|}\right) = r\mathbf{w}^t \frac{\mathbf{w}}{\|\mathbf{w}\|} = r\|\mathbf{w}\| \quad \Rightarrow \quad r = \frac{g(\mathbf{x})}{\|\mathbf{w}\|}$$

Multi-category Case

c categories: $\omega_1, \omega_2, \ldots, \omega_c$

- Approaches

number of classifiers needed?

  - Use two-class discriminant for each class

    $\Rightarrow$ $\mathbf{x}$ belongs to class ω_i or not?

  - Use two-class discriminant for each pair of classes

    $\Rightarrow$ $\mathbf{x}$ belongs to class ω_i or ω_j?

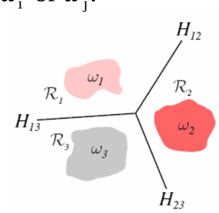
- General Approach

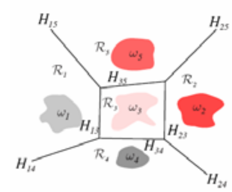
one function for each class $g_i(\mathbf{x}) = \mathbf{w}_i{}^t \mathbf{x} + w_{i0}$

map x to class ω_i if $g_i(x) \geq g_j(x)\ \forall j \neq i$

decision boundary H_{ij}: $g_i(\mathbf{x}) = g_j(\mathbf{x})$

$$\boxed{H_{ij}:\ (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (w_{i0} - w_{j0}) = 0}$$





- Each decision regions is convex and singly connected.
- Good for monomodal distributions

Method for searching decision boundaries

$g_i(\mathbf{x}) = \mathbf{w}_i{}^t \mathbf{x} + w_{i0}$ $\Rightarrow$ find weight $\boldsymbol{\omega}$ and bias ω_o

- Augmented Vector

$$\mathbf{y} = \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} \quad \mathbf{a}_i = \begin{bmatrix} w_{i0} \\ \mathbf{w}_i \end{bmatrix} = \begin{bmatrix} w_{i0} \\ w_{i1} \\ \vdots \\ w_{id} \end{bmatrix} \quad \Rightarrow g_i(\mathbf{x}) = g_i(\mathbf{y}) = \mathbf{a}_i \mathbf{y}$$

- Decision Boundary

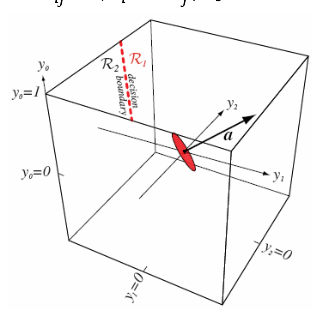
$$H_{ij}:\ (\mathbf{w}_i - \mathbf{w}_j)^t \mathbf{x} + (w_{i0} - w_{j0}) = 0 \quad \Rightarrow H_{ij}:\ (\mathbf{a}_i - \mathbf{a}_j)^t \mathbf{y} = 0$$

- 2-category case

$$H:\ (\mathbf{w})^t \mathbf{x} + w_{i0} = 0$$

$$\Downarrow$$

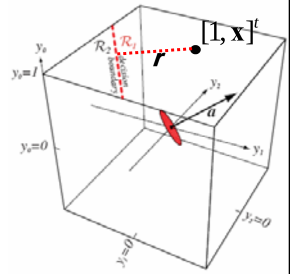
$$\boxed{H:\ \mathbf{a}^t \mathbf{y} = 0}$$



- A hyperplane in augmented y space, with normal vector **a**

Search Method for Linear Discriminant

all sample points reside in the $y_1 = 1$ subspace



distance from $\mathbf{x}$ to boundary in $\mathbf{x}$ space: $r = \frac{g(\mathbf{x})}{\|\mathbf{w}\|}$

distance from $\mathbf{x}$ to boundary in $\mathbf{y}$ space:

$r' = |\mathbf{a}^t\mathbf{y}| / \|\mathbf{a}\| \le r$ i.e., r' and r same signs, r' lower bound for r

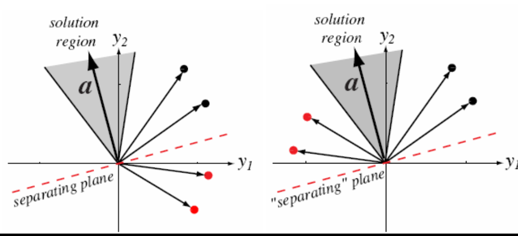
- Design Objective for finding **a**
  - Find **a** that correctly classify each sample data

$\forall \mathbf{y}_i$ in class ω_1, $\mathbf{a}^t\mathbf{y}_i > 0$ $\forall \mathbf{y}_i$ in class ω_2, $\mathbf{a}^t\mathbf{y}_i < 0$

- Normalization $\forall \mathbf{y}_i$ in class ω_2, $\mathbf{y}_i \leftarrow -(\mathbf{y}_i)$
- New Design Objective

$\forall \mathbf{y}_i$ in class ω_1 or ω_2, $\mathbf{a}^t\mathbf{y}_i > 0$

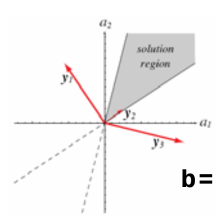
- Solution region
  - Intersection of positive sides of all hyperplanes



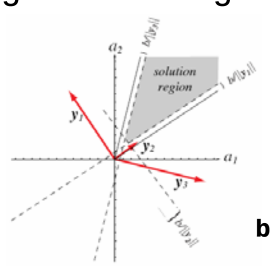
Searching Linear Discriminant Solutions

- Stricter criterion: Solution region with margin

$\forall \mathbf{y}_i$ in class ω_1 or ω_2, $\mathbf{a}^t\mathbf{y}_i > b$



b=0



b>0

- Search Approaches
  - Gradient decent methods to find a solution in the solution region
  - Maximize margin
  - Mapping to high-dimensional space