THE UNIVERSITY OF TENNESSEE KNOXVILLE

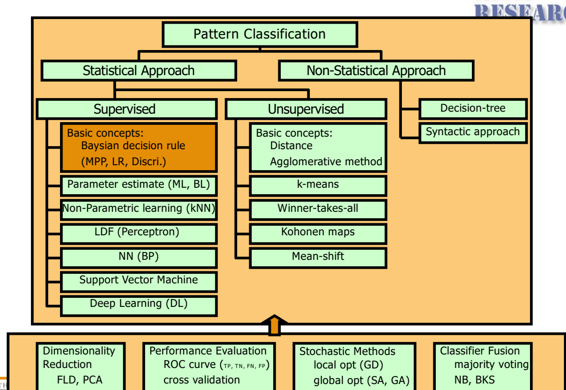
AICIP RESEARCH

ECE471-571 – Pattern Recognition

Lecture 3 – Discriminant Function and Normal Density

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
http://www.eecs.utk.edu/faculty/qi
Email: hqi@utk.edu

AICIP RESEARCH



Bayes Decision Rule

AICIP RESEARCH

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$$

Maximum Posterior Probability

For a given x, if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$, then x belongs to class 1, otherwise, 2.

Likelihood Ratio

If $\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$, then x belongs to class 1, otherwise, 2.

THE UNIVERSITY OF TENNESSEE KNOXVILLE

3

Discrimimant Function

- One way to represent pattern classifier- use discriminant functions $g_i(x)$

$$g_i(x) = P(\omega_i|x)$$
$$g_i(x) = p(x|\omega_i)P(\omega_i)$$
$$g_i(x) = \ln p(x|\omega_i) + \ln P(\omega_i)$$

The classifier will assign a feature vector x to class ω_i if $g_i(x) > g_j(x)$

- For two-class cases,

$$g(x) = g_1(x) - g_2(x) = P(\omega_1 | x) - P(\omega_2 | x)$$

Multivariate Normal Density

$$p(\vec{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1}(\vec{x}-\vec{\mu})\right]$$

$\vec{x}$: d-component column vector

$\vec{\mu}$: d-component mean vector

Σ : d-by-d covariance matrix

$|\Sigma|$: determinant

Σ^{-1} : inverse

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}, \vec{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1d} \\ \vdots & \ddots & \vdots \\ \sigma_{d1} & \cdots & \sigma_{dd} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1d} \\ \vdots & \ddots & \vdots \\ \sigma_{d1} & \cdots & \sigma_d^2 \end{bmatrix}$$

When d = 1, $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$

Discriminant Function for Normal Density

$$p(\vec{x} | w) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1}(\vec{x}-\vec{\mu})\right]$$

$$\begin{aligned} g_i(\vec{x}) &= \ln p(\vec{x} | \omega_i) + \ln P(\omega_i) \\ &= -\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1}(\vec{x}-\vec{\mu}_i) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i) \\ &= -\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1}(\vec{x}-\vec{\mu}_i) - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i) \end{aligned}$$

Case 1: $\Sigma_i=\sigma^2 I$

- The features are statistically independent, and have the same variance
- Geometrically, the samples fall in equal-size hyperspherical clusters
- Decision boundary: hyperplane of d-1 dimension

$$\Sigma = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix}, |\Sigma| = \sigma^{2d}, \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma^2} \end{bmatrix}$$

Linear Discriminant Function and Linear Machine

$\|\vec{x}-\vec{\mu}_i\|$: the Euclidean norm (distance)

$$\|\vec{x}-\vec{\mu}_i\|^2 = (\vec{x}-\vec{\mu}_i)^T(\vec{x}-\vec{\mu}_i)$$

$$g_i(\vec{x}) = -\frac{\|\vec{x}-\vec{\mu}_i\|^2}{2\sigma^2} + \ln P(\omega_i)$$

$$= -\frac{\vec{x}^T\vec{x} - 2\vec{\mu}_i^T\vec{x} + \vec{\mu}_i^T\vec{\mu}_i}{2\sigma^2} + \ln P(\omega_i)$$

$$g_i(\vec{x}) = \frac{\vec{\mu}_i^T}{\sigma^2}\vec{x} - \frac{\vec{\mu}_i^T\vec{\mu}_i}{2\sigma^2} + \ln P(\omega_i)$$

Minimum-Distance Classifier

- When $P(\omega_i)$ are the same for all c classes, the discriminant function is actually measuring the minimum distance from each x to each of the c mean vectors

$$g_i(\vec{x}) = -\frac{\|\vec{x}-\vec{\mu}_i\|^2}{2\sigma^2}$$

Case 2: $\Sigma_i = \Sigma$

AICIP RESEARCH

- The covariance matrices for all the classes are identical but not a scalar of identity matrix.
- Geometrically, the samples fall in hyperellipsoidal
- Decision boundary: hyperplane of d-1 dimension

$$g_i(\vec{x}) = \ln p(\vec{x} \mid \omega_i) + \ln P(\omega_i)$$

$$= -\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}-\vec{\mu}_i) + \ln P(\omega_i)$$

Squared Mahalanobis distance

$$= \vec{\mu}_i^T \left(\Sigma^{-1}\right)^T \vec{x} - \frac{1}{2}\vec{\mu}_i^T \Sigma^{-1} \vec{\mu}_i + \ln P(\omega_i)$$

THE UNIVERSITY OF TENNESSEE

10

Case 3: Σ_i = arbitrary

AICIP RESEARCH

- The covariance matrices are different from each category
- Quadratic classifier
- Decision boundary: hyperquadratic for 2-D Gaussian

$$g_i(\vec{x}) = \ln p(\vec{x} \mid \omega_i) + \ln P(\omega_i)$$

$$= -\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}-\vec{\mu}_i) - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

$$= -\frac{1}{2}\vec{x}^T \Sigma_i^{-1}\vec{x} + \vec{\mu}_i^T \left(\Sigma_i^{-1}\right)^T \vec{x} - \frac{1}{2}\vec{\mu}_i^T \Sigma_i^{-1}\vec{\mu}_i - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

THE UNIVERSITY OF TENNESSEE

11

Case Study

AICIP RESEARCH

a1 3.1 11.70 a	b1 8.3 1.00 b	c1 10.2 6.40 c	u1 5.1 0.4 b
a2 3.0 1.30 a	b2 3.8 0.20 b	c2 9.2 7.90 c	u2 12.9 5.0 c
a3 1.9 0.10 a	b3 3.9 0.60 b	c3 9.6 3.10 c	u3 13.0 0.8 b
a4 3.8 0.04 a	b4 7.8 1.20 b	c4 53.8 2.50 c	u4 2.6 0.1 a
a5 4.1 1.10 a	b5 9.1 0.60 b	c5 15.8 7.60 c	u5 30.0 0.1 o
a6 1.9 0.40 a	b6 15.4 3.60 b		u6 20.5 0.8 o
	b7 7.7 1.60 b		
	b8 6.5 0.40 b		
	b9 5.7 0.40 b		
	b10 13.6 1.60 b		

- Calculate μ
- Calculate Σ
- Derive the discriminant function $g_i(x)$

THE UNIVERSITY OF TENNESSEE

12