北京交通大学

图像处理与机器学习

Digital Image Processing and Machine Learning

主讲人：黄琳琳

电子信息工程学院

第五章　贝叶斯决策

- 基本概念
- 贝叶斯决策
- 判别函数
- 概率密度估计

基本概念

- 举例：手写字符识别 （Handwritten Character Recognition）

$$\boxed{y_k(\vec{x})=y_k(\vec{x},\vec{w})}$$

贝叶斯决策

判别函数，特征空间，决策面

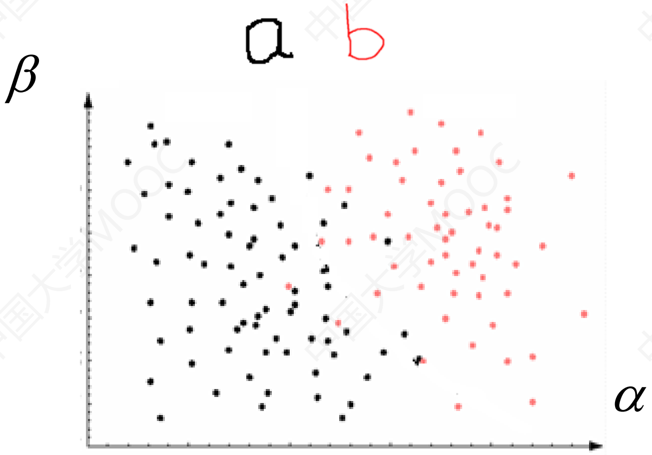
判别函数

✓ 判别函数 (Discriminant Function， DF)

-- 表征模式（pattern）属于每一类的广义似然度

✓ 特征空间(Feature space)

-- 表征模式（pattern）的空间



特征空间

判别函数

- ✓ 决策面（Decision surface，boundary）

  -- 特征空间中两类判别函数**相等的点**的集合

$$y_k(\vec{x}) = p(\omega_k \mid \vec{x}) \qquad g(\vec{x}) = y_1(\vec{x}) - y_2(\vec{x}) = 0$$

$$g(\vec{x}) = p(\omega_1 \mid \vec{x}) - p(\omega_2 \mid \vec{x})$$

$$g(\vec{x}) = p(\vec{x} \mid \omega_1) p(\omega_1) - p(\vec{x} \mid \omega_2) p(\omega_2)$$

判别函数

决策面： 特征空间中两类判别函数**相等的点**的集合

$$g(\vec{x}) = p(\vec{x}|\omega_1)p(\omega_1) - p(\vec{x}|\omega_2)p(\omega_2)$$

$$g(\vec{x}) = \ln[p(\vec{x}|\omega_1)p(\omega_1)] - \ln[p(\vec{x}|\omega_2)p(\omega_2)]$$

$$g(\vec{x}) = \ln p(\vec{x}|\omega_1) + \ln p(\omega_1) - \ln p(\vec{x}|\omega_2) - \ln p(\omega_2)$$

$$g(\vec{x}) = \ln p(\vec{x}|\omega_1) - \ln p(\vec{x}|\omega_2) + \ln p(\omega_1) - \ln p(\omega_2)$$

$$g(\vec{x}) = \ln \frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)} + \ln \frac{p(\omega_1)}{p(\omega_2)}$$

判别函数

- ✓ 正态分布概率
- ✓ 二维特征空间

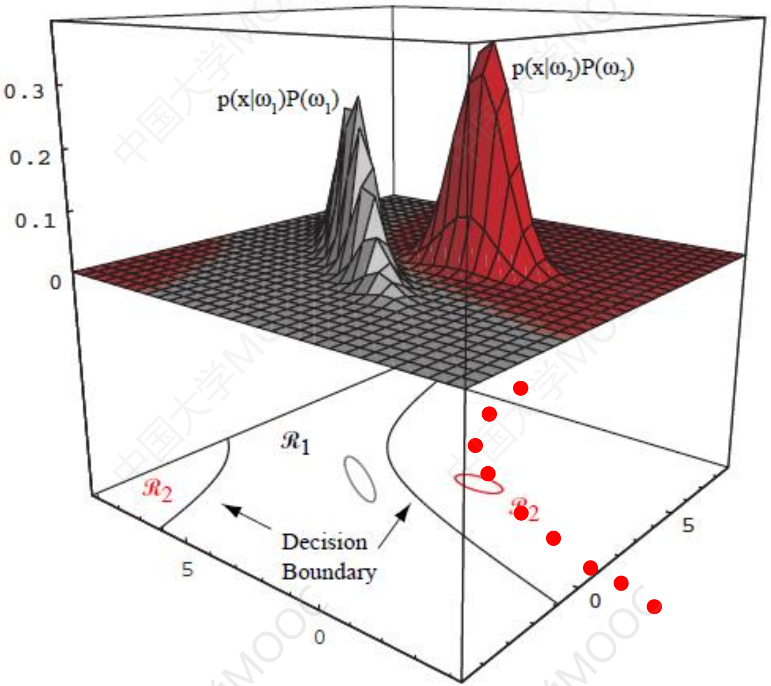
$$g(\vec{x}) = p(\vec{x} \mid \omega_1) p(\omega_1) - p(\vec{x} \mid \omega_2) p(\omega_2)$$

$$= 0$$

Decision boundary

有判别函数就可以分类了，
为什么还来求决策面？

加深对特征空间的理解



判别函数

- 高斯概率密度下的判别函数

$$p(\omega_i \mid \vec{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2} (\vec{x} - \mu_i) \Sigma_i^{-1} (\vec{x} - \mu_i) \right]$$

$$\vec{\mu} = E[\vec{x}] = \int \vec{x} p(\vec{x}) d\vec{x}$$

$$\Sigma = E\left[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^t \right] = \int (\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^t p(\vec{x}) d\vec{x}$$

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd} \end{bmatrix}$$

$$\sigma_{ij} = E\left[(x_i - \mu_i)(x_j - \mu_j)^t \right]$$

如果 σ_i 和 σ_j 互不相关，则 $\sigma_{ij}=0$

判别函数

◆ 高斯密度下的判别函数

$$y_k(\vec{x}) = p(\omega_k \mid \vec{x}) \qquad p(\omega_i \mid \vec{x}) = \frac{1}{(2\pi)^{d/2} \mid \Sigma_i \mid^{1/2}} \exp\left[-\frac{1}{2}(\vec{x}-\mu_i)\Sigma_i^{-1}(\vec{x}-\mu_i) \right]$$

$$y_k(\vec{x}) = \ln p(\omega_k \mid \vec{x})$$

$$y_k(\vec{x}) = -\frac{1}{2}(\vec{x}-\vec{\mu}_k)^t \Sigma_k^{-1}(\vec{x}-\vec{\mu}_k) \cancel{-\frac{d}{2}\ln 2\pi} - \frac{1}{2}\ln \mid \Sigma_k \mid + \ln p(\omega_k)$$

$$\boxed{y_k(\vec{x}) = -\frac{1}{2}(\vec{x}-\vec{\mu}_k)^t \Sigma_k^{-1}(\vec{x}-\vec{\mu}_k) - \frac{1}{2}\ln \mid \Sigma_k \mid + \ln p(\omega_k)}$$

判别函数

- 高斯密度下的判别函数

$$y_k(\vec{x}) = -\frac{1}{2}(\vec{x}-\vec{\mu}_k)^t \boxed{\Sigma_k^{-1}} (\vec{x}-\vec{\mu}_k) - \frac{1}{2}\ln|\Sigma_k| + \ln p(\omega_k)$$

- Case 1 $\boxed{\Sigma_k = \sigma^2 I}$
- Case 2 $\boxed{\Sigma_k = \Sigma}$
- Case 3 $\boxed{\Sigma_k = arbitrary}$

判别函数

$$y_k(\vec{x}) = -\frac{1}{2}(\vec{x}-\vec{\mu}_k)^t \Sigma_k^{-1}(\vec{x}-\vec{\mu}_k) - \frac{1}{2}\ln|\Sigma_k| + \ln p(\omega_k)$$

- Case 1 $\boxed{\Sigma_k = \sigma^2 I}$
  - ✓ 特征独立,相同方差，对角矩阵

$$y_k(\vec{x}) = -\frac{1}{2}(\vec{x}-\vec{\mu}_k)^t \cancel{\Sigma_k^{-1}}(\vec{x}-\vec{\mu}_k) - \cancel{\frac{1}{2}\ln|\Sigma_k|} + \ln p(\omega_k)$$

$$y_k(\vec{x}) = -\frac{\|\vec{x}-\vec{\mu}_k\|^2}{2\sigma^2} + \ln p(\omega_k)$$

判别函数

- Case 1 $\boxed{\Sigma_k = \sigma^2 I}$

$$y_k(\vec{x}) = -\frac{\|\vec{x}-\vec{\mu}_k\|^2}{2\sigma^2} + \ln p(\omega_k)$$

$$y_k(\vec{x}) = \frac{1}{\sigma^2}\vec{\mu}_k{}^t\vec{x} - \vec{\mu}_k{}^t\vec{\mu}_k + \ln p(\omega_k)$$

$$\vec{w}_k = \frac{1}{\sigma^2}\vec{\mu}_k, w_{k0} = -\vec{\mu}_k{}^t\vec{\mu}_k + \ln p(\omega_k)$$

$$\boxed{y_k(\vec{x}) = \vec{w}_k^t\vec{x} + w_{k0}}$$

- ✓ 判别函数是线性函数
- ✓ 对应分类器称为**线性**分类器

Linear Discriminant Function (LDF)

判别函数

- Case 1 $\boxed{\Sigma_k = \sigma^2 I}$ $y_k(\vec{x}) = \vec{w}_k^t \vec{x} + w_{k0}$

✓ 决策面（判别函数相等的点构成） $y_k(\vec{x}) = y_i(\vec{x})$ *or* $y_k(\vec{x}) - y_i(\vec{x}) = 0$

$$(\vec{w}_k^t \vec{x} + w_{k0}) - (\vec{w}_i^t \vec{x} + w_{i0}) = 0 \qquad \vec{w}_k = \frac{1}{\sigma^2}\vec{\mu}_k \qquad w_{k0} = -\vec{\mu}_k^t \vec{\mu}_k + \ln p(\omega_k)$$

$$\downarrow$$

$$(\vec{w}_k^t - \vec{w}_i^t)\vec{x} + (w_{k0} - w_{i0}) = 0$$

$$\downarrow$$

$$\vec{w}^t(\vec{x} - \vec{x}_0) = 0 \qquad \vec{w} = \vec{\mu}_k - \vec{\mu}_i \qquad \vec{x}_0 = \frac{1}{2}(\vec{\mu}_k + \vec{\mu}_i) - \frac{\sigma^2}{\|\vec{\mu}_k - \vec{\mu}_i\|^2} \ln \frac{p(\omega_k)}{p(\omega_i)} (\vec{\mu}_k - \vec{\mu}_i)$$

判别函数

- Case 1 $\boxed{\Sigma_k = \sigma^2 I}$ $y_k(\vec{x}) = \vec{w}_k^t \vec{x} + w_{k0}$

  - 决策面（判别函数相等的点构成）

$$\vec{w}^t(\vec{x} - \vec{x}_0) = 0 \qquad \vec{w} = \vec{\mu}_k - \vec{\mu}_i \qquad \vec{x}_0 = \frac{1}{2}(\vec{\mu}_k + \vec{\mu}_i) - \frac{\sigma^2}{\left\|\vec{\mu}_k - \vec{\mu}_i\right\|^2} \ln \frac{p(\omega_k)}{p(\omega_i)} (\vec{\mu}_k - \vec{\mu}_i)$$

$\vec{w}$ 为决策面的法向量，决定决策面的方向；

$\vec{x}_0$ 为决策面距离原点的距离，决定决策面的位置；

判别函数

- Case 1 $\boxed{\Sigma_k = \sigma^2 I}$ $y_k(\vec{x}) = \vec{w}_k^t \vec{x} + w_{k0}$

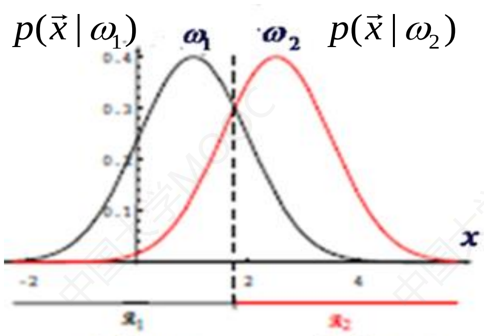
$$\vec{w}^t(\vec{x} - \vec{x}_0) = 0 \quad \vec{w} = \vec{\mu}_k - \vec{\mu}_i$$

✓ 先验概率相等的情况 $p(\omega_1)=p(\omega_2)$

$$\vec{x}_0 = \frac{1}{2}(\vec{\mu}_k + \vec{\mu}_i) - \cancel{\frac{\sigma^2}{\|\vec{\mu}_k - \vec{\mu}_i\|^2} \ln \frac{p(\omega_k)}{p(\omega_i)}(\vec{\mu}_k - \vec{\mu}_i)}$$

$$\vec{x}_0 = \frac{1}{2}(\vec{\mu}_k + \vec{\mu}_i) \qquad x_0 = \frac{1}{2}(1.5 + 2.5) = 2$$

决策面为两类均值的等分面



$p(\omega_1)=0.5$ $p(\omega_2)=0.5$

$\mu_1=1.5$ $\mu_2=2.5$

判别函数

- Case 1 $\boxed{\Sigma_k = \sigma^2 I}$ $y_k(\vec{x}) = \vec{w}_k^t \vec{x} + w_{k0}$ $\vec{w}^t(\vec{x} - \vec{x}_0) = 0$ $\vec{w} = \vec{\mu}_k - \vec{\mu}_i$

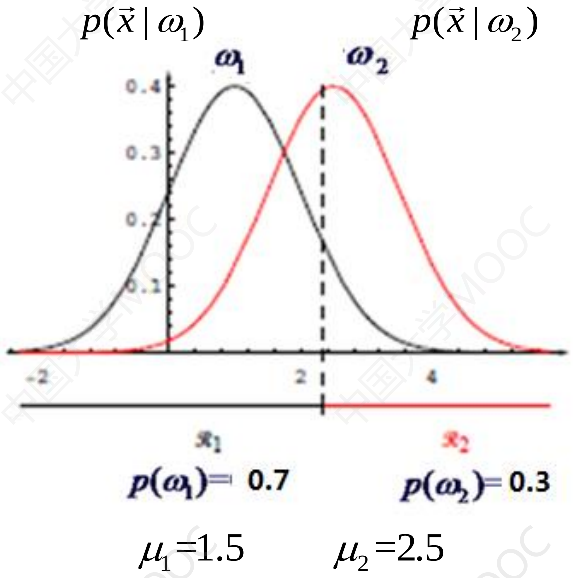
✓ 先验概率不相等的情况 $p(\omega_1) \neq p(\omega_2)$

$$\vec{x}_0 = \frac{1}{2}(\vec{\mu}_k + \vec{\mu}_i) - \frac{\sigma^2}{\|\vec{\mu}_k - \vec{\mu}_i\|^2} \ln \frac{p(\omega_k)}{p(\omega_i)} (\vec{\mu}_k - \vec{\mu}_i)$$

$$p(\omega_1)=0.7, p(\omega_2) = 0.3$$

$$x_0 = \frac{1}{2}(1.5+2.5) - \sigma^2 \ln \frac{0.7}{0.3}(1.5-2.5)$$

$$x_0 = \frac{1}{2}(1.5+2.5) + \sigma^2 \ln \frac{7}{3}$$



决策面移向先验概率小的类别

判别函数

- 高斯密度下的判别函数

$$y_k(\vec{x}) = -\frac{1}{2}(\vec{x}-\vec{\mu}_k)^t \Sigma_k^{-1}(\vec{x}-\vec{\mu}_k) - \frac{1}{2}\ln|\Sigma_k| + \ln p(\omega_k)$$

- Case 1 $\boxed{\Sigma_k = \sigma^2 I}$

$$y_k(\vec{x}) = \vec{w}_k^t \vec{x} + w_{k0}$$ 判别函数是线性函数

$$y_k(\vec{x}) = -\frac{\|\vec{x}-\vec{\mu}_k\|^2}{2\sigma^2} + \ln p(\omega_k)$$

- 计算与类中心的距离
- 最小距离分类器

谢 谢

本课程所引用的一些素材为主讲老师多年的教学积累，来源于多种媒体及同事和同行的交流，难以一一注明出处，特此说明并表示感谢！