模式识别的理论与方法

Pattern Recognition

裴继红

Chapter 2 (Part 3): Bayesian Decision Theory

(Sections 2-6)

正态分布的判决函数

一元正态密度

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

这里:

μ 为 x 的均值 (或称：期望值)

σ^2 为方差

d 维多元正态分布函数

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

其中，

$x=(x_1, x_2, \ldots, x_d)^t$ （t 表示转置）为特征矢量

$\mu=(\mu_1, \mu_2, \ldots, \mu_d)^t$ 为均值矢量

$\Sigma=d\times d$ 是协方差矩阵

$|\Sigma|$ 和 Σ^{-1} 分别代表 Σ 的 行列式和逆矩阵

正态分布判别函数

- 根据贝叶斯决策理论，最小错误率（minimum error-rate）分类可以通过判别函数完成
- $$g_i(x) = ln\ P(x \mid \omega_i) + ln\ P(\omega_i)$$
- 在 $g_i(x)$为多元正态情况下

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

正态分布：$\Sigma_i = \sigma^2 I$ 时

- 在 $\Sigma_i = \sigma^2 I$ 时（I 表示单位矩阵，各个特征维上的方差相同），不考虑判别式中的常量，判别式简化为

$$g_i(x) = \frac{1}{\sigma^2} w_i^t x + w_{i0}$$

是线性判别函数，其中 $w_i = \mu_i$

$$w_{i0} = -\frac{1}{2\sigma^2}\mu_i^t \mu_i + \ln P(\omega_i)$$

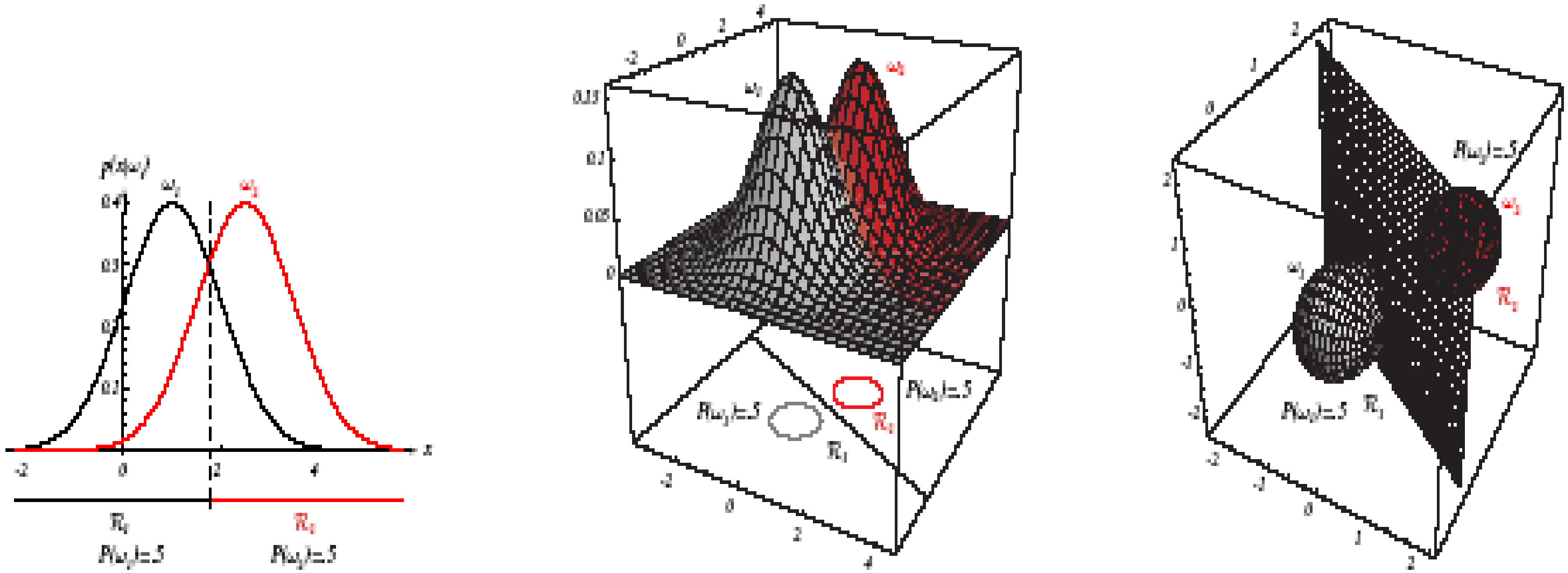
称 w_{i0} 为第 i 类的偏置

线性分类机

- 使用线性判别函数的分类机称为“线性机”

- 线性机的决策面是下式定义的一个超平面

$$g_i(x) = g_j(x)$$



- 若两种高斯分布的协方差相等，且与单位矩阵成比例，则其分布是 d 维空间的超高斯球型。其判别边界是一般化的d−1维的超平面，该超平面垂直于被分割开的两个中心的连线。
- 图中给出了 1 维、2 维和 3 维例子，同时标出了判别函数 $p(\mathbf{x}|\omega_i)$，以及边界 $P(\omega_1)=P(\omega_2)$。在 3 维图的情况中，将R_1和 R_2分开的网格平面是判别边界。

判别超平面

- 判别超平面是将区域 R_i 和 R_j 分开的超平面

$$w^t(x - x_0) = 0 \qquad\qquad w = (\mu_i - \mu_j)$$

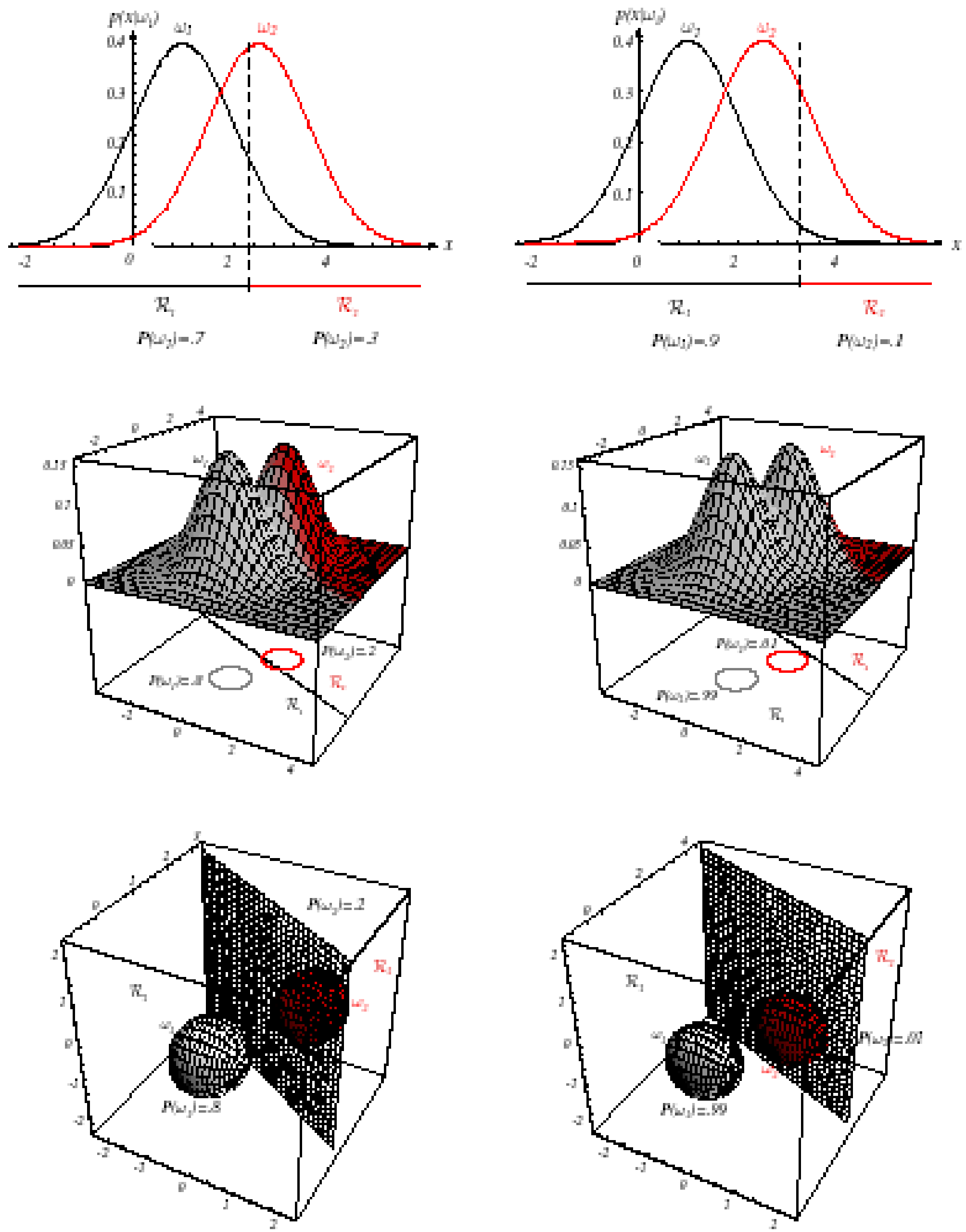
$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \sigma^2 \cdot \frac{\mu_i - \mu_j}{\|\mu_i - \mu_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)}$$

该平面总是与两类类均值的连线垂直

$$if \quad P(\omega_i) = P(\omega_j) \quad \text{then} \quad x_0 = \frac{1}{2}(\mu_i + \mu_j)$$

随着先验概率的改变，决策边界移动；

若各类之间的先验概率差异充分大，则决策面将不会处在这些 1维、2 维和 3 维球状高斯分布的均值点的中间。



正态分布：$\Sigma_i = \Sigma$ 时

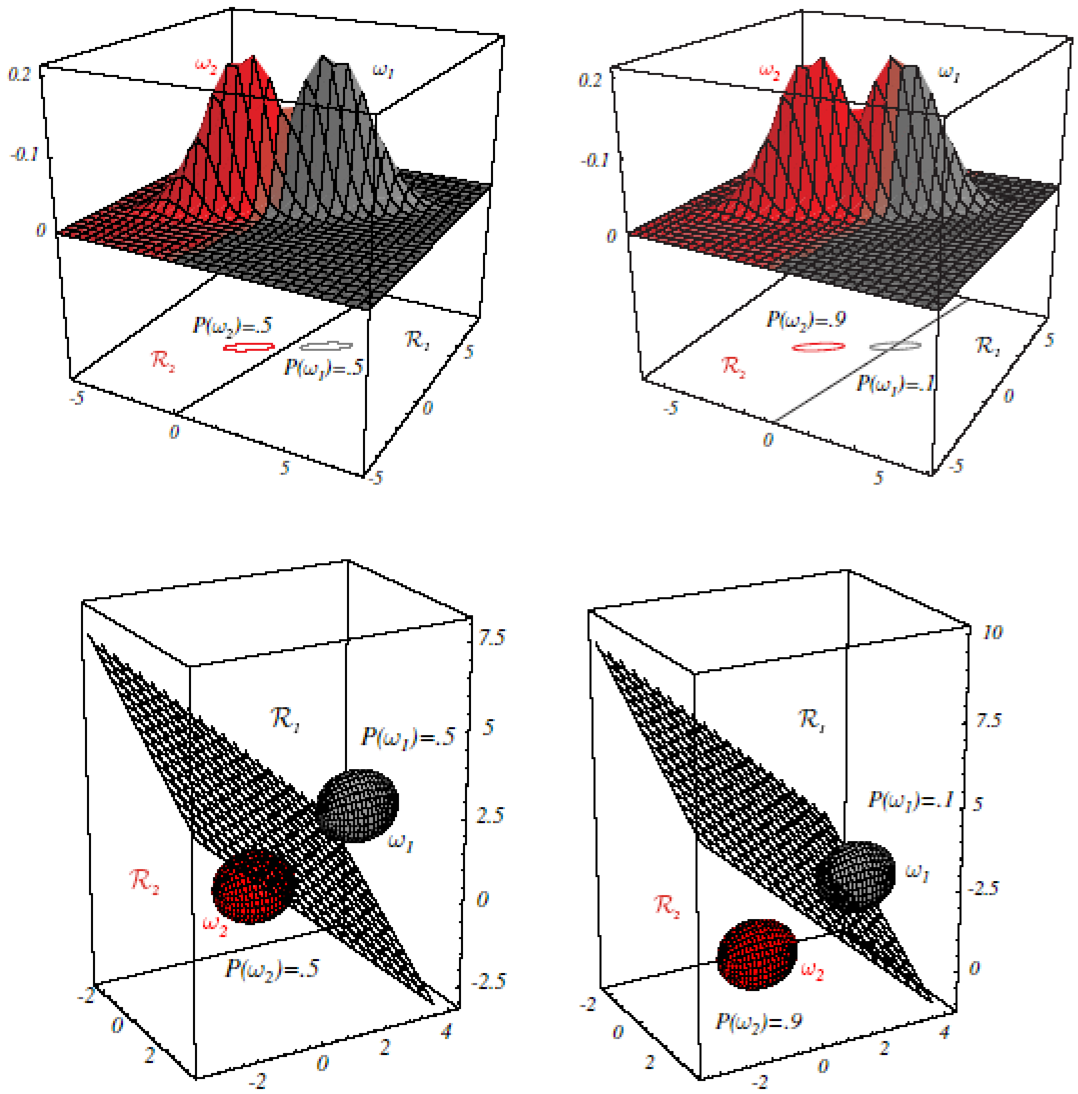
- 在 $\Sigma_i = \Sigma$ 的情况下，即所有类的协方差矩阵都相等，但不同方向的方差可以不相等的条件下，将区域 R_i 和 R_j 分开的超平面的位置为

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\mu_i - \mu_j}{(\mu_i - \mu_j)^t \Sigma^{-1} (\mu_i - \mu_j)} \cdot \ln \frac{P(\omega_i)}{P(\omega_j)}$$

将区域 R_i 和 R_j 分开的超平面一般情况下与两类均值点的连线不正交

- 相等但各向异性的高斯分布的概率密度（在 2 维时为曲面，在 3 维时为椭球面）和决策区域 。

- 决策超平面一般与均值的连线不垂直。



正态分布：Σ_i = 任意 时

- 此时，对每一类来说，协方差矩阵都不同。

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

$where:$

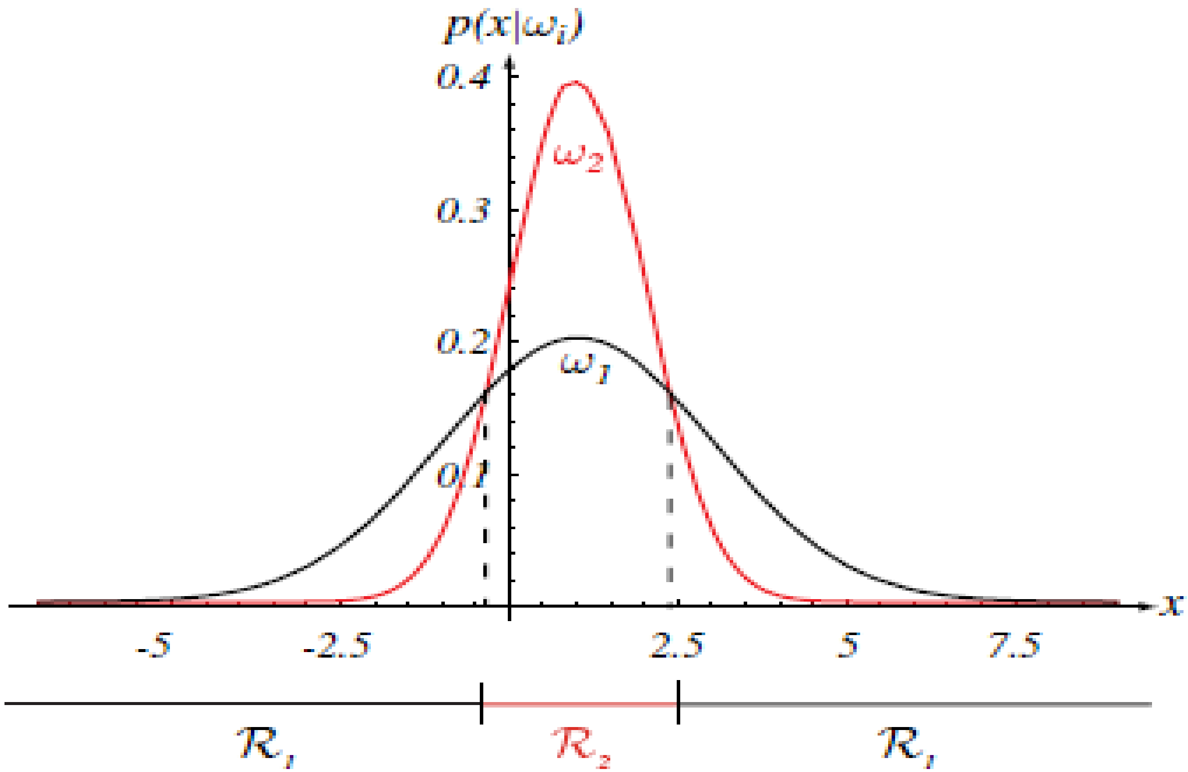
$$\mathrm{W_i} = -\frac{1}{2}\Sigma_i^{-1}$$

$$\mathrm{w_i} = \Sigma_i^{-1}\mu_i$$

$$\mathrm{w}_{i0} = -\frac{1}{2}\mu_i^t\Sigma_i^{-1}\mu_i - \frac{1}{2}\ln\left|\Sigma_i\right| + \ln P(\omega_i)$$

  – 对应的判决面是超二次曲面：**超平面、超椭球面、超抛物面、超双曲面**

- 在方差不相等的情况下，还有可能产生非单连通决策区域



二维情况下的判别面

- 类的任意高斯分布产生了一般的超双曲判决边界面
- 反之，给定一个超二次曲面，总可以找到两个高斯分布，其贝叶斯决策边界是该超二次曲面。
- 图中的轮廓线示出了某个常数的概率密度的分布