

第二章 贝叶斯决策理论

2010. 09. 29

多元正态分布

□ 概率密度函数

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})),$$

其中:

$\mathbf{x} = [x_1, x_2, \dots, x_d]^T$ 是 d 维列向量 (T 表示向量的转置);

$\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_d]^T$ 是 d 维均值向量;

Σ 是 $d \times d$ 协方差矩阵;

Σ^{-1} 是 Σ 的逆矩阵, $|\Sigma|$ 是 Σ 的行列式。

多元正态分布

□ 均值 $\boldsymbol{\mu}$ 和协方差矩阵 Σ

$$\boldsymbol{\mu} = E[\mathbf{x}] = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x};$$

$$\Sigma = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] = \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T p(\mathbf{x}) d\mathbf{x};$$

$$\mu_i = E[x_i] = \int x_i p(x_i) dx_i$$

$$\sigma_{ij}^2 = E[(x_i - \mu_i)(x_j - \mu_j)^T]$$

$$= \iint (x_i - \mu_i)(x_j - \mu_j)^T p(x_i, x_j) dx_i dx_j.$$

多元正态分布的性质

1. 参数 $\boldsymbol{\mu}$ 和 Σ 对分布具有决定性

$$p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \Sigma);$$

- 多元正态分布由 $d+d(d+1)/2$ 个参数完全确定。

2. 分布的离散程度

- 由参数 $|\Sigma|^{1/2}$ 决定, 与单变量时由标准差 σ 决定相一致:

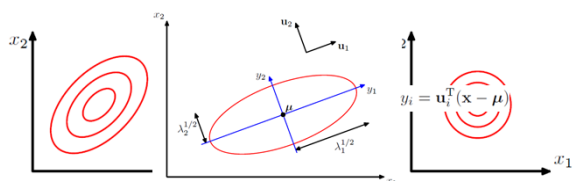
$$|\Sigma|^{1/2} = \prod_{j=1}^d \lambda_j^{1/2}.$$

多元正态分布的性质

3. 等密度点的轨迹为超椭球面

$$(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \text{常数};$$

- 超椭球面的主轴方向由协方差矩阵的特征向量决定, 各主轴的长度与相应的特征值成正比。



多元正态分布的性质

3. 等密度点的轨迹为超椭球面

- 马氏距离(Mahalanobis distance): 随机向量 $\mathbf{x}$ 偏离均值向量 $\boldsymbol{\mu}$ 的距离

$$r = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})};$$

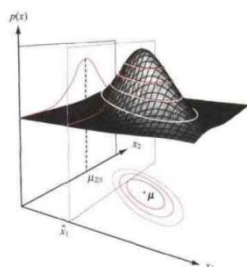
- 确定未知样本集和已知样本集的相似性: 考虑各特性间的联系; 与尺度无关;
- 衡量两个服从同一分布且其协方差矩阵为 Σ 的随机变量的差异程度。

多元正态分布的性质

4. 边缘分布和条件分布的正态性

- 多元正态分布的**边缘分布**和**条件分布**仍然是正态分布;

$$p(x_i) \sim N(\mu_i, \sigma_{ii}^2)$$



多元正态分布的性质

5. 不相关性等价于独立性

- x_i 和 x_j 相互独立: $p(x_i, x_j) = p(x_i)p(x_j)$;
- x_i 和 x_j 不相关: $E[x_i x_j] = E[x_i] \cdot E[x_j]$;
- 如多元正态分布的任意两个分量互不相关, 则它们一定独立。
- 如多元正态随机向量 $\mathbf{x}$ 的协方差阵 Σ 是对角阵, 则 $\mathbf{x}$ 各分量之间是相互独立的正态分布随机变量。

$$\Rightarrow \Sigma = \begin{bmatrix} \sigma_{11}^2 & 0 & \dots & 0 \\ 0 & \sigma_{22}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_{nn}^2 \end{bmatrix}$$

$$\Rightarrow p(\mathbf{x}) = \prod_{i=1}^n p(x_i)$$

多元正态分布的性质

6. 线性变换的正态性

- 多元正态分布的随机向量的**线性变换**仍然是多元正态分布的随机向量, 即

$$p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \Sigma)$$

$$\mathbf{y} = \mathbf{A}^T \mathbf{x}, \mathbf{A} \in \mathbb{R}^{d \times k}$$

$$\Rightarrow p(\mathbf{y}) \sim N(\mathbf{A}^T \boldsymbol{\mu}, \mathbf{A}^T \Sigma \mathbf{A});$$

- 存在一个线性变换 $\mathbf{A}$, 使变换后的协方差矩阵 $\mathbf{A}^T \Sigma \mathbf{A}$ 为对角矩阵; 即在某一新坐标系中, 可以做到使各分量之间相互独立。

多元正态分布的性质

6. 线性变换的正态性

- 白化变换**

$$\mathbf{A}_w = \Phi \Lambda^{-1/2} \Rightarrow p(\mathbf{y}) \sim N(\mathbf{A}_w^T \boldsymbol{\mu}, \mathbf{I});$$

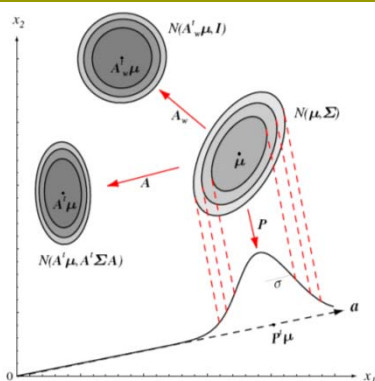
其中, 矩阵 Φ 的列向量是 Σ 的正交特征向量, 矩阵 Λ 由 Σ 相应的特征根构成的对角矩阵。

- 线性组合的正态性**

- 当 $k=1$ 时(即 $\mathbf{A}$ 是 d 维向量 $\mathbf{a}$), 则 $\mathbf{y}=\mathbf{a}^T \mathbf{x}$ 是一个标量, 是 $\mathbf{x}$ 的线性组合

$$p(y) \sim N(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \Sigma \mathbf{a});$$

多元正态分布的性质



多元正态分布的性质

- 参数 $\boldsymbol{\mu}$ 和 Σ 完全决定分布;
- 等概率密度点的轨迹为超椭球面;
- 不相关性等价于独立性;
- 边缘分布和条件分布的正态性;
- 线性变换的正态性;
 - 白化变换;
 - 线性组合的正态性。

正态分布&最小错误率Bayes决策

□ 各模式类的特征向量服从多元正态分布:

$$p(\mathbf{x} | \omega_i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad i = 1, \dots, c;$$

□ 最小错误率 Bayes 决策规则和判别函数:

$$\omega_i = \arg \max_{\omega_j} \{p(\mathbf{x} | \omega_j)P(\omega_j)\},$$

$$g_i(\mathbf{x}) = \ln p(\mathbf{x} | \omega_i) + \ln P(\omega_i);$$

采用对数形式的判别函数:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| - \frac{d}{2} \ln 2\pi + \ln P(\omega_i).$$

正态分布&最小错误率Bayes决策

□ 决策面方程

$$g_i(\mathbf{x}) = g_j(\mathbf{x}), \quad \text{即}$$

$$-\frac{1}{2}[(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - (\mathbf{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{x} - \boldsymbol{\mu}_j)]$$

$$-\frac{1}{2} \ln \frac{|\boldsymbol{\Sigma}_i|}{|\boldsymbol{\Sigma}_j|} + \ln \frac{P(\omega_i)}{P(\omega_j)} = 0$$

正态分布&最小错误率Bayes决策

情况一: $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}, \quad i = 1, \dots, c$

■ 即有 $|\boldsymbol{\Sigma}_i| = \sigma^{2d}, \quad \boldsymbol{\Sigma}_i^{-1} = \frac{1}{\sigma^2} \mathbf{I};$

■ 判别函数:

$$g_i(\mathbf{x}) = -\frac{(\mathbf{x} - \boldsymbol{\mu}_i)^T (\mathbf{x} - \boldsymbol{\mu}_i)}{2\sigma^2} - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^{2d} + \ln P(\omega_i) \Rightarrow$$

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

其中, 欧式距离平方 $\|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = (\mathbf{x} - \boldsymbol{\mu}_i)^T (\mathbf{x} - \boldsymbol{\mu}_i) = \sum_{j=1}^d (x_j - \mu_{ij})^2.$

正态分布&最小错误率Bayes决策

□ 最小距离分类器

■ 条件: $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}, P(\omega_i) = 1/c, \quad i = 1, \dots, c.$

■ 判别函数:

$$g_i(\mathbf{x}) = -(\mathbf{x} - \boldsymbol{\mu}_i)^T (\mathbf{x} - \boldsymbol{\mu}_i) = -\|\mathbf{x} - \boldsymbol{\mu}_i\|^2.$$

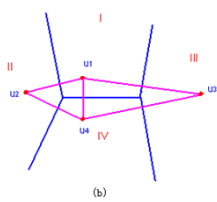
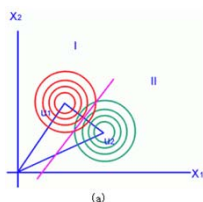
■ 判别规则: 每个样本以它到每类样本均值的欧式距离平方的最小值确定其分类, 即

如果 $\|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \min_{j=1, \dots, c} \|\mathbf{x} - \boldsymbol{\mu}_j\|^2$, 则 $\mathbf{x} \in \omega_i$.

正态分布&最小错误率Bayes决策

□ 最小距离分类器

■ 可看作模板匹配: 每个类有一个典型样本(即均值向量), 称为模板; 而待分类样本 $\mathbf{x}$ 只需按欧式距离计算与哪个模板最相似(欧式距离最短)即可作决定。



正态分布&最小错误率Bayes决策

□ 线性分类器

■ 条件: $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}$, 各先验概率关系未知;

■ 判别函数:

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

$$= -\frac{1}{2\sigma^2} (\mathbf{x}^T \mathbf{x} - 2\boldsymbol{\mu}_i^T \mathbf{x} + \boldsymbol{\mu}_i^T \boldsymbol{\mu}_i) + \ln P(\omega_i);$$

$\therefore \mathbf{x}^T \mathbf{x}$ 与类别号 i 无关

$$\therefore g_i(\mathbf{x}) = -\frac{1}{2\sigma^2} (-2\boldsymbol{\mu}_i^T \mathbf{x} + \boldsymbol{\mu}_i^T \boldsymbol{\mu}_i) + \ln P(\omega_i) = \mathbf{w}_i^T \mathbf{x} + \omega_{i0},$$

其中, $\mathbf{w}_i^T = \frac{1}{\sigma^2} \boldsymbol{\mu}_i^T, \quad \omega_{i0} = -\frac{1}{2\sigma^2} \boldsymbol{\mu}_i^T \boldsymbol{\mu}_i + \ln P(\omega_i).$

正态分布&最小错误率Bayes决策

□ 线性分类器

- 判别函数为线性函数或决策面为超平面;
- 决策面方程:

$$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0,$$

其中: $\mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$,

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\sigma^2}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j).$$

正态分布&最小错误率Bayes决策

Case 1: $\Sigma_i = \sigma^2 \mathbf{I}$

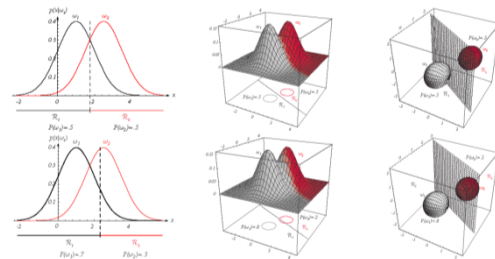


Figure 5: If the covariance matrices of two distributions are equal and proportional to the identity matrix, then the distributions are spherical in d dimensions, and the boundary is a generalized hyperplane of $d - 1$ dimensions, perpendicular to the line separating the means. The decision boundary shifts as the priors are changed.

正态分布&最小错误率Bayes决策

情况二: $\Sigma_i = \Sigma$, $i=1, \dots, c$

- 几何上, 具有同样概率密度函数的点的轨迹是同样大小和形状的超椭球面, 中心由类均值 $\boldsymbol{\mu}_i$ 决定。
- 判别函数:

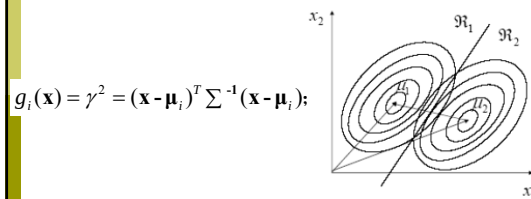
$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i) - \frac{d}{2} \ln 2\pi$$

$$\Rightarrow g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) + \ln P(\omega_i);$$

正态分布&最小错误率Bayes决策

□ 马式距离分类器

- 条件: $\Sigma_i = \Sigma$, $P(\omega_i) = 1/c$, $i=1, \dots, c$.
- 判别函数: $\mathbf{x}$ 到 $\boldsymbol{\mu}_i$ 的马式距离的平方



正态分布且 $P(\omega_1) = P(\omega_2)$, $\Sigma_1 = \Sigma_2$ 时的决策面

正态分布&最小错误率Bayes决策

□ 线性分类器

- 条件: $\Sigma_i = \Sigma$, 各先验概率关系未知;
- 判别函数:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}^T \Sigma^{-1} \mathbf{x} - \boldsymbol{\mu}_i^T \Sigma^{-1} \mathbf{x} - \mathbf{x}^T \Sigma^{-1} \boldsymbol{\mu}_i + \boldsymbol{\mu}_i^T \Sigma^{-1} \boldsymbol{\mu}_i) + \ln P(\omega_i)$$

$$\Rightarrow g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + \omega_{i0}$$

其中, $\mathbf{w}_i = \Sigma^{-1} \boldsymbol{\mu}_i$,

$$\omega_{i0} = -\frac{1}{2} \boldsymbol{\mu}_i^T \Sigma^{-1} \boldsymbol{\mu}_i + \ln P(\omega_i).$$

正态分布&最小错误率Bayes决策

□ 线性分类器

- 决策面方程

$$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0,$$

其中: $\mathbf{w} = \Sigma^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$,

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{1}{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \Sigma^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)} \ln \frac{P(\omega_i)}{P(\omega_j)} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j).$$

正态分布&最小错误率Bayes决策

Case 2: $\Sigma_i = \Sigma$

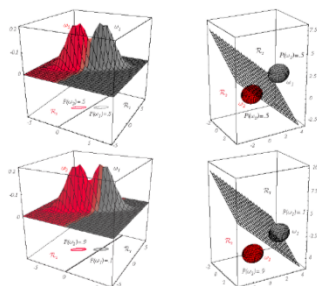


Figure 6: Probability densities with equal but asymmetric Gaussian distributions. The decision hyperplanes are not necessarily perpendicular to the line connecting the means.

正态分布&最小错误率Bayes决策

情况三：各类的协方差矩阵不相等

■ 判决函数：

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

$$= \mathbf{x}^T \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + \omega_{i0}$$

其中, $\mathbf{W}_i = -\frac{1}{2} \boldsymbol{\Sigma}_i^{-1}$ ($d \times d$ 矩阵),

$\mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i$ (d 维列向量),

$$\omega_{i0} = -\frac{1}{2} \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i).$$

正态分布&最小错误率Bayes决策

情况三：各类的协方差矩阵不相等

■ 决策面方程

□ 决策面为二次超曲面：超球面、超椭球面、超抛物面、超双曲面，也可能是超平面。

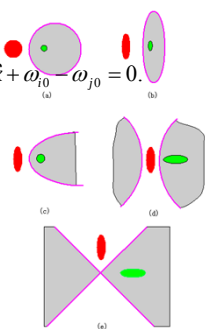
(a) 是两个超球体等密度分布-圆；

(b) 是(a)在 x_2 轴方向有扩展-椭圆；

(c) 表示决策面为抛物面；

(d) 与(c) 差别在于均值点相互关系不同；

(e) 中出现了对称性情况，双曲线退化成直线；



例：二维正态分布数据的决策域

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

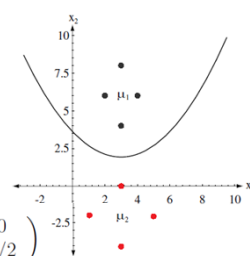
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}; \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}; \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_1^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \boldsymbol{\Sigma}_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$x_2 = 3.514 - 1.125x_1 + 0.1875x_1^2$$

a parabola with vertex at $\begin{pmatrix} 3 \\ 1.83 \end{pmatrix}$



总结

□ 贝叶斯决策理论是统计模式识别的重要理论基础。

■ 理论上讲，贝叶斯决策方法是最优的（在最小错误率或最小风险意义上）；

■ 贝叶斯决策所需条件最多。实际应用中，需要首先得到先验概率和类条件概率密度。