第二章 贝叶斯决策理论

2009.10.13

多元正态分布的性质

1. 参数 μ 和 Σ 完全决定分布;
2. 等概率密度点的轨迹为超椭球面;
3. 不相关性等价于独立性;
4. 边缘分布和条件分布的正态性;
5. 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

- 最小错误率Bayes决策的判决规则和判决函数
- 各模式类的观测特征向量服从多元正态分布

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

- 采用对数形式的判别函数

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

  - 判别函数中与类别 i 无关的项，对于类别的决策没有影响，可以忽略。

正态分布&最小错误率Bayes决策

- 决策面方程

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

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

$$-\frac{1}{2}\ln\frac{|\sum_i|}{|\sum_j|} + \ln\frac{P(\omega_i)}{P(\omega_j)} = 0$$

正态分布&最小错误率Bayes决策

情况一: $\Sigma_i = \sigma^2\mathbf{I},\ i=1,\ldots,c$

- $\mathbf{I}$ 是 $d\times d$ 维单位矩阵，即

$$\sum_i = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \sigma^2 & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix},\ \left|\sum_i\right| = \sigma^{2d},\ \sum_i^{-1} = \frac{1}{\sigma^2}\mathbf{I};$$

- 代入判决函数即有

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

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

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

情况一: $\Sigma_i = \sigma^2\mathbf{I},\ i=1,\ldots,c$

a) **最小距离分类器**: 各先验概率相等，则判决函数可简化为

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

- 判决规则: 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

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

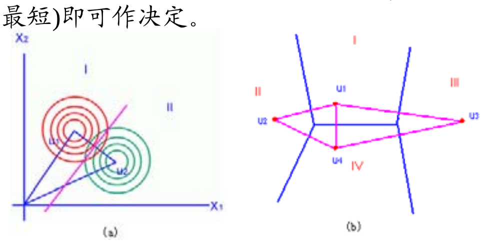
- 条件: 正态分布 $\Sigma_i = \sigma^2\mathbf{I}$, $P(\omega_i)=1/c$。

正态分布&最小错误率Bayes决策

情况一：$\Sigma_i=\sigma^2 I$, $i=1,\ldots,c$

a) **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 **x** 只要按欧氏距离计算与哪个模板最相似(欧氏距离最短)即可作决定。



正态分布&最小错误率Bayes决策

情况一：$\Sigma_i=\sigma^2 I$, $i=1,\ldots,c$

b) **线性分类器**：（各先验概率关系未知）

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

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

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

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

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

正态分布&最小错误率Bayes决策

情况一：$\Sigma_i=\sigma^2 I$, $i=1,\ldots,c$

b) **线性分类器**：判别函数为线性函数或者决策面为超平面的分类器。

- 决策面方程： $\mathbf{w}^T(\mathbf{x}-\mathbf{x}_0)=0,$

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

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

  - 决策面为一个过$\mathbf{x}_0$的超平面，法线方向为($\mu_i-\mu_j$)。当$P(\omega_i)=P(\omega_j)$，该超平面过($\mu_i+\mu_j$)/2点；在二维情况下，即过μ_i与μ_j连线的垂直平分线。当$P(\omega_i)\neq P(\omega_j)$，该超平面的位置要向远离先验概率大的方向偏，但超平面方向不变。

正态分布&最小错误率Bayes决策

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

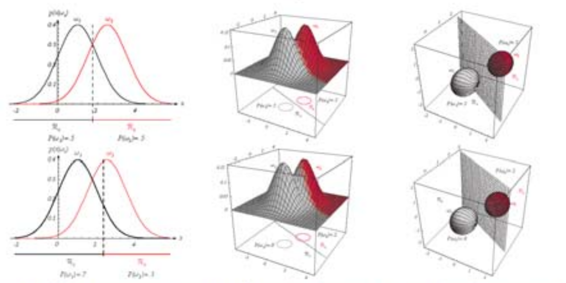


Figure 5: If the covariance matrices of two distributions are equal and proportional to the identity matrix, then the distributions are spherical in d dimensions, and the boundary is a generalized hyperplane of $d-1$ dimensions, perpendicular to the line separating the means. The decision boundary shifts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

正态分布&最小错误率Bayes决策

情况二：$\Sigma_i=\Sigma$，$i=1,\ldots,c$，即各类协方差矩阵相等

- 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μ_i决定。
- Σ与 i 无关，

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

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

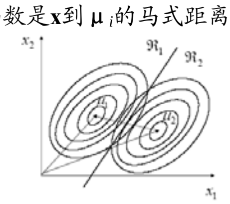
正态分布&最小错误率Bayes决策

情况二：$\Sigma_i=\Sigma$，$i=1,\ldots,c$，即各类协方差矩阵相等

a) **马式距离分类器**：各先验概率相等，则判决函数可简化为

$$g_i(\mathbf{x})=\gamma^2=(\mathbf{x}-\boldsymbol{\mu}_i)^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}_i);$$

- 判决函数是**x**到μ_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

情况二：$\Sigma_i=\Sigma$，$i=1,\ldots,c$，即各类协方差矩阵相等

b) **各先验概率关系未知**

- 忽略与 i 无关的项

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

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

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

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

- 决策方程也是线性方程，决策面是超平面。

正态分布&最小错误率Bayes决策

情况二：$\Sigma_i=\Sigma$，$i=1,\ldots,c$，即各类协方差矩阵相等

b) **线性分类器**（各先验概率关系未知）

- 决策面方程

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

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

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

- 决策面为一个过$\mathbf{x}_0$的超平面。当$P(\omega_i)=P(\omega_j)$，该超平面过$(\mu_i+\mu_j)/2$点；当$P(\omega_i)\neq P(\omega_j)$，该超平面朝远离先验概率大的方向移动。一般情况下，该超平面不与两均值向量的连线正交。

正态分布&最小错误率Bayes决策

Case 2: $\Sigma_i=\Sigma$

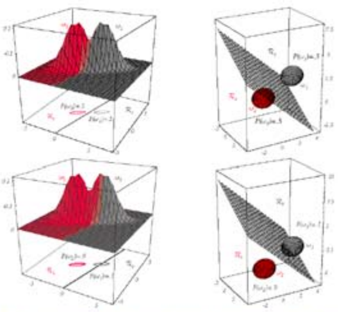


Figure 6: Probability densities with equal but asymmetric Gaussian distributions. The decision hyperplanes are not necessarily perpendicular to the line connecting the means.

正态分布&最小错误率Bayes决策

- 线性分类器小结
  - 在多元正态分布的条件下，基于最小错误率贝叶斯决策只要能做到各类别的协方差矩阵是一样的，那么无论先验概率是否相等，都可以用线性分界面实现。
  - 最小（欧氏）距离分类器则要求各正态分布的协方差矩阵为单位阵，且各类别的先验概率相等。

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

- 判决函数为 $\mathbf{x}$ 的二次型

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

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

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

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

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

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

$$x^T(\mathbf{W}_i-\mathbf{W}_j)x+(\mathbf{w}_i-\mathbf{w}_j)^Tx+\omega_{i0}-\omega_{j0}=0.$$

  - 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。

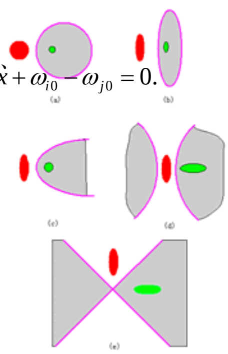
(a)是两个超球体等密度分布-圆;

(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(e) 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;



讨论

- 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
    - 方法一：先估计概率密度，后求解决策规则
    - 方法二：若已知或可假设概率密度为某种形式（比如正态分布），可先求出判决函数形式，再从样本估计其中的参数。
    - 方法三：直接选择或假设某种判决函数形式，用样本确定其参数。