模式识别课程

正态分布下的贝叶斯分类器

张俊超

中南大学
航空航天学院

贝叶斯分类器-正态分布及其性质

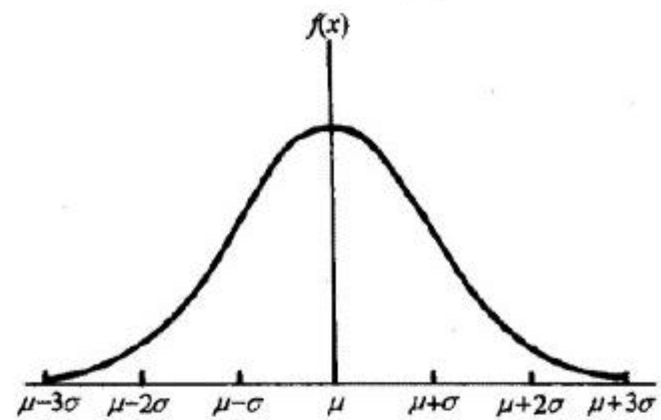
- **单变量正态分布**

单变量正态分布概率密度函数：

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

$$\mu=E\{x\}=\int_{-\infty}^{\infty}xp(x)$$

$$\sigma^2=\int_{-\infty}^{\infty}(x-\mu)^2p(x)$$



贝叶斯分类器-正态分布及其性质

- **多元正态分布**

多元正态分布概率密度函数：

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

标量

$$\mathbf{x} = (x_1, x_2, \ldots, x_d)^T$$

$$\boldsymbol{\mu} = E\{\mathbf{x}\} = (\mu_1, \mu_2, \ldots, \mu_d)^T$$

$$\mathbf{\Sigma} = E\left\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\right\} \in \mathbb{R}^{d\times d}$$

协方差矩阵（对称非负定阵$|\mathbf{\Sigma}| \geq 0$）

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1d} & \sigma_{2d} & \cdots & \sigma_{dd} \end{pmatrix}$$

协方差矩阵计算的是不同维度之间的协方差，而不是不同样本之间的

贝叶斯分类器-正态分布及其性质

➢ **多元正态分布的性质：**

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

① 均值向量和协方差矩阵决定分布

均值向量 $\boldsymbol{\mu}=E\{\mathbf{x}\}=(\mu_1,\mu_2,\ldots,\mu_d)^T$

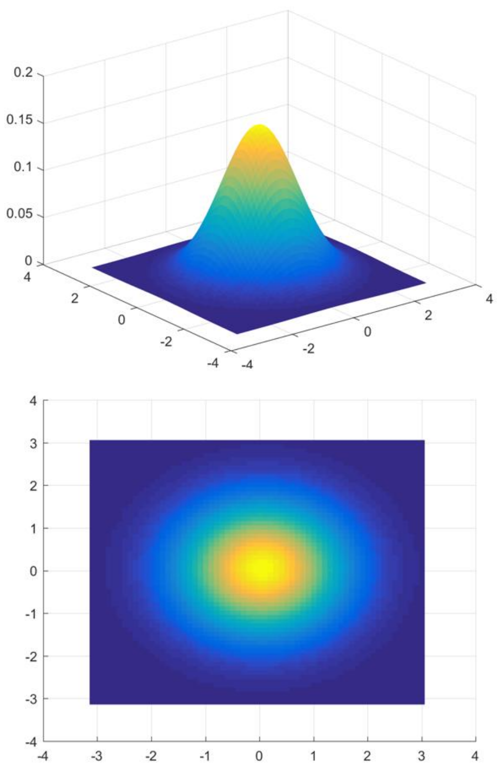
协方差矩阵 $\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d}\\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2d}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{1d} & \sigma_{2d} & \cdots & \sigma_{dd}\end{pmatrix}\in\mathbb{R}^{d\times d}$

$$\boxed{\begin{aligned}\sigma_{ij}&=E\{(x_i-\mu_i)(x_j-\mu_j)\}\\&=E\{x_ix_j\}-E\{x_i\}E\{x_j\}\end{aligned}}$$

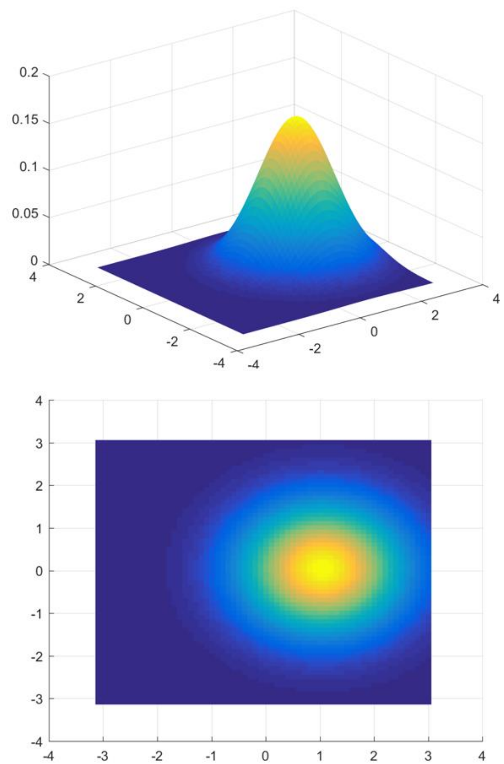
参数总数: $d(d+1)/2+d$

贝叶斯分类器－正态分布及其性质

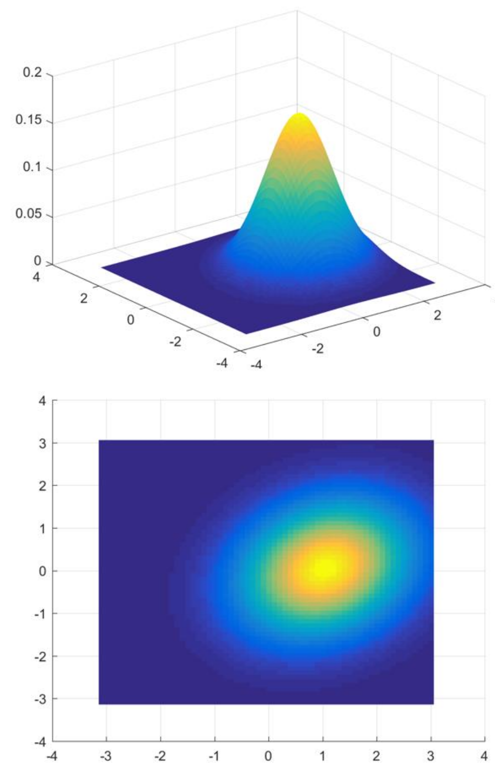
$$\boldsymbol{\mu}=\begin{pmatrix}0\\0\end{pmatrix},\ \boldsymbol{\Sigma}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$$



$$\boldsymbol{\mu}=\begin{pmatrix}0\\1\end{pmatrix},\ \boldsymbol{\Sigma}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$$



$$\boldsymbol{\mu}=\begin{pmatrix}0\\1\end{pmatrix},\ \boldsymbol{\Sigma}=\begin{pmatrix}1&0.2\\0.2&1\end{pmatrix}$$



贝叶斯分类器-正态分布及其性质

```
%% 多元正态分布
mu=[0;
    1];
sigma=[1 0.2;
       0.2 1];
%%
det_sigma = sqrt(det(sigma));
inv_sigma = inv(sigma);
Item = 1/(2*pi*det_sigma);
x=-pi:0.1:pi;
y=-pi:0.1:pi;
[X,Y]=meshgrid(x,y); % 产生网格数据并处理
p = zeros(size(X));
for i = 1:length(x)
    for j = 1:length(y)
        xx = [x(i);y(j)];
        xx = xx-mu;
        p(i,j) = Item*exp(-0.5*xx'*inv_sigma*xx);
    end
end

figure(1),surf(X,Y,p),shading flat
figure(2),surf(X,Y,p),shading flat,view(2)
%% 等高线
figure(3),contour(X,Y,p)
```

贝叶斯分类器-正态分布及其性质

② 等密度点的轨迹为一超椭球面

常数

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

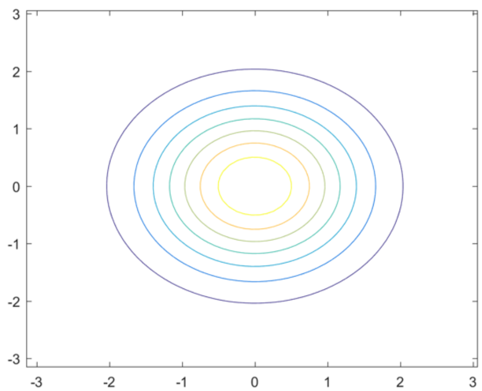
指数项为常数，概率密度p(x)不变，即，等密度点应满足:

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

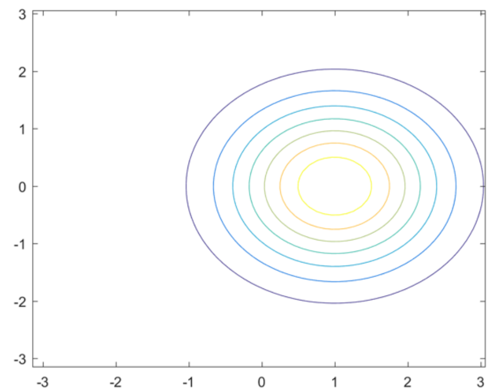
轨迹是一超椭球面(二维：椭圆)，中心由均值向量决定，主轴方向由协方差矩阵的特征向量决定，主轴的长度与协方差矩阵的特征值决定。

贝叶斯分类器-正态分布及其性质

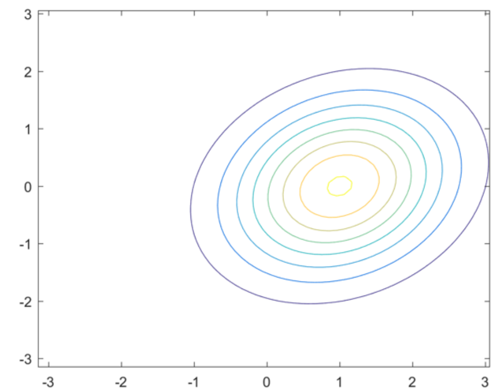
$$\boldsymbol{\mu}=\begin{pmatrix}0\\0\end{pmatrix},\ \boldsymbol{\Sigma}=\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$$



$$\boldsymbol{\mu}=\begin{pmatrix}0\\1\end{pmatrix},\ \boldsymbol{\Sigma}=\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$$

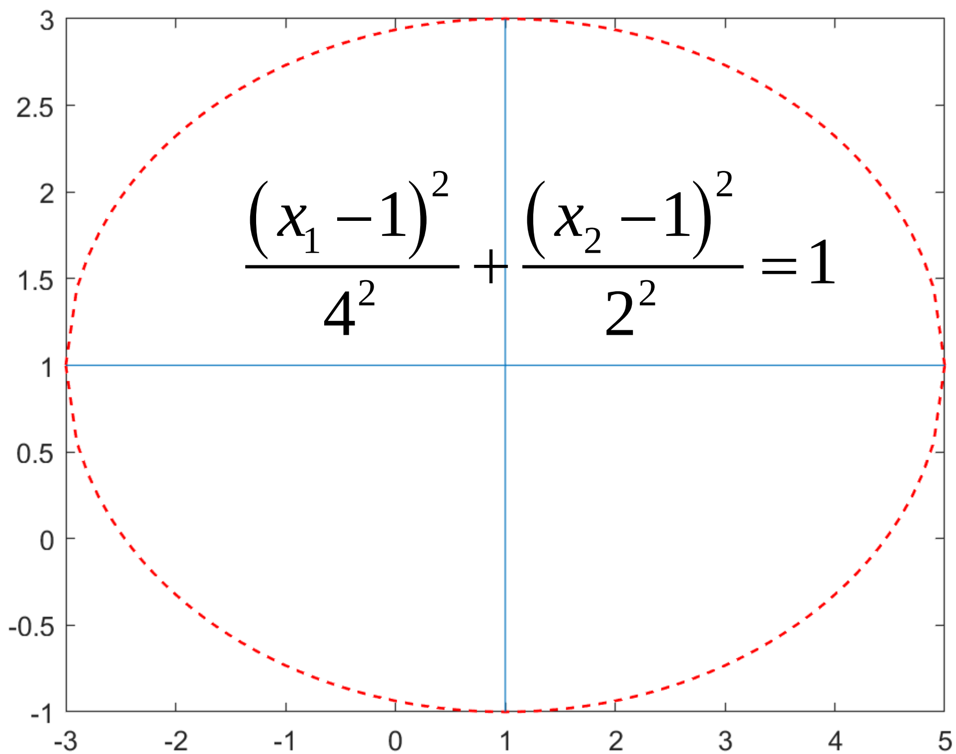


$$\boldsymbol{\mu}=\begin{pmatrix}0\\1\end{pmatrix},\ \boldsymbol{\Sigma}=\begin{pmatrix}1 & 0.2\\0.2 & 1\end{pmatrix}$$



贝叶斯分类器－正态分布及其性质

为什么主轴方向由协方差矩阵的特征向量决定，
主轴的长度与协方差矩阵的特征值决定？

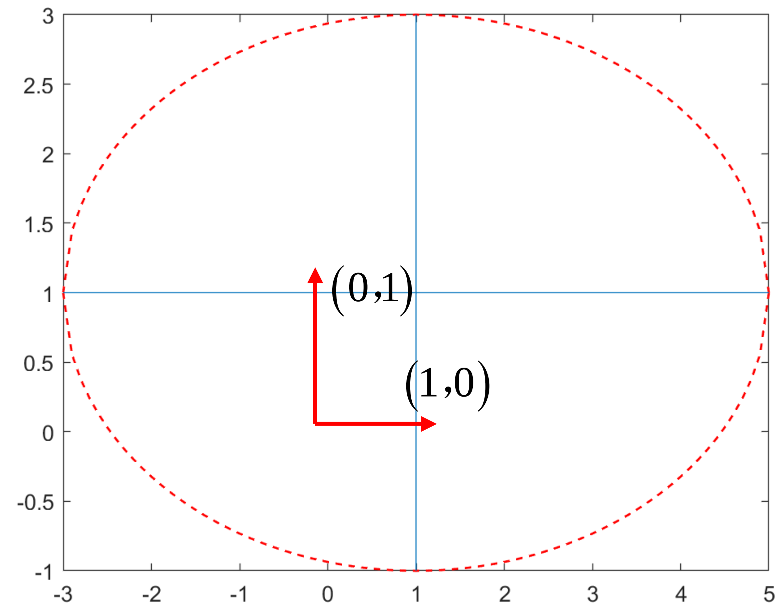


贝叶斯分类器-正态分布及其性质

$$\begin{pmatrix} x_1-1 & x_2-1 \end{pmatrix} \boxed{\begin{pmatrix} \frac{1}{4^2} & 0 \\ 0 & \frac{1}{2^2} \end{pmatrix}} \begin{pmatrix} x_1-1 \\ x_2-1 \end{pmatrix} = 1$$

$\begin{pmatrix} 4^2 & 0 \\ 0 & 2^2 \end{pmatrix}$的特征值$(16,4)$,

特征向量:$\begin{pmatrix} 1 \\ 0 \end{pmatrix},\begin{pmatrix} 0 \\ 1 \end{pmatrix}$



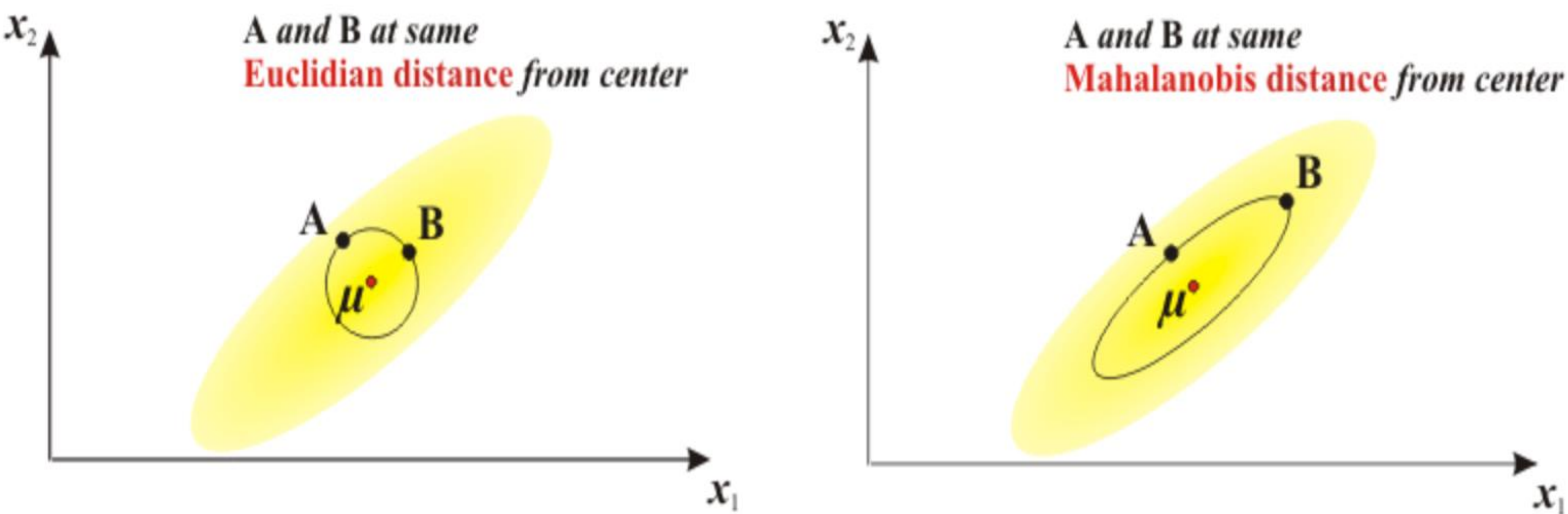
贝叶斯分类器－正态分布及其性质

- 马氏距离(Mahalanobis)
- 欧式距离

$$\gamma^2 = \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{x}_i - \mathbf{x}_j\right)$$

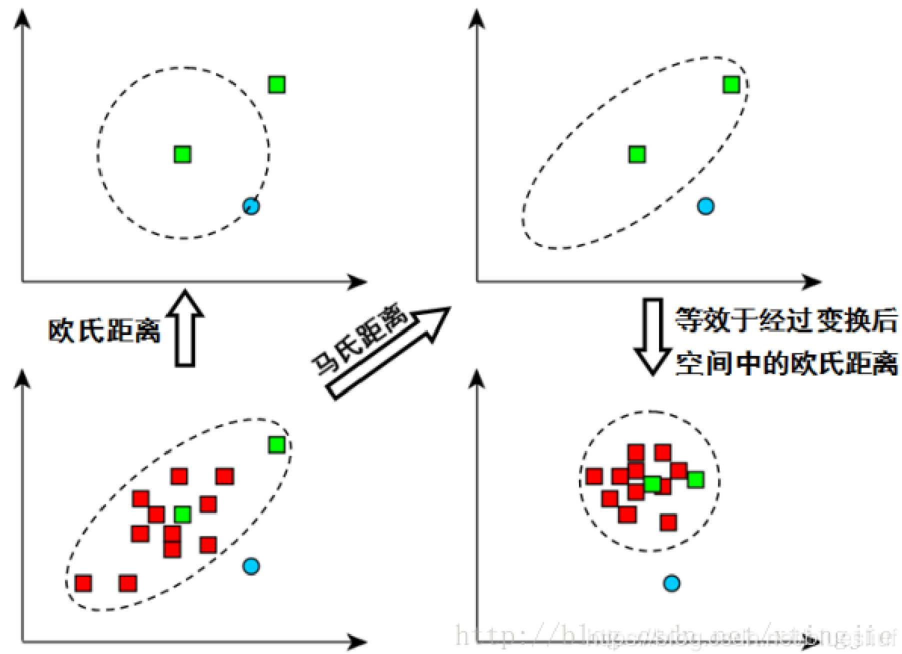
$$d^2 = \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \left(\mathbf{x}_i - \mathbf{x}_j\right)$$

不同程度（单位）之间做了归一化



贝叶斯分类器-正态分布及其性质

- 马氏距离(Mahalanobis) $\gamma^2 = \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{\Sigma}^{-1} \left(\mathbf{x}_i - \mathbf{x}_j\right)$
- 欧式距离 $d^2 = \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \left(\mathbf{x}_i - \mathbf{x}_j\right)$



贝叶斯分类器-补充知识

- 自学
  - 如何理解矩阵的特征值分解？
    https://www.matongxue.com/madocs/228.html
  - 马氏距离VS欧式距离
    https://zhuanlan.zhihu.com/p/46626607
  - 协方差矩阵
    https://zhuanlan.zhihu.com/p/37609917

贝叶斯分类器-正态分布及其性质

③ 不相关等价于独立

独立一定不相关，不相关不一定独立；
在正态分布下，独立与不相关等价。

$$独立: p(x_i x_j) = p(x_i)\, p(x_j)$$

$$不相关: E\{x_i x_j\} = E\{x_i\} \cdot E\{x_j\}$$

※多元随机向量的各分量相互独立，则协方差矩阵是对角阵。

$$\boxed{\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{dd} \end{pmatrix}}$$

④ 多元正态分布的边缘分布和条件分布仍是正态分布

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

二元正态向量 $(x_1, x_2)^T$，有

$$p(x_1) \sim N(\mu_1, \sigma_{11}) \qquad p(x_2) \sim N(\mu_2, \sigma_{22})$$

贝叶斯分类器-正态分布及其性质

⑤ 多元正态随机向量的线性变换仍满足正态分布

$$p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow p(\mathbf{y} = \mathbf{A}\mathbf{x}) \sim N\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T\right), \mathbf{A}\text{是非奇异的}$$

存在某个$\mathbf{A}$使得$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T$是对角阵 $\Rightarrow$ $\mathbf{y}$的各个分量不相关

⑥ 多元正态随机向量的线性组合，是一维的正态随机变量

$$p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow p\left(\mathbf{a}^T\mathbf{x}\right) \sim N\left(\mathbf{a}^T\boldsymbol{\mu}, \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{a}\right)$$

贝叶斯分类器-正态分布及其性质

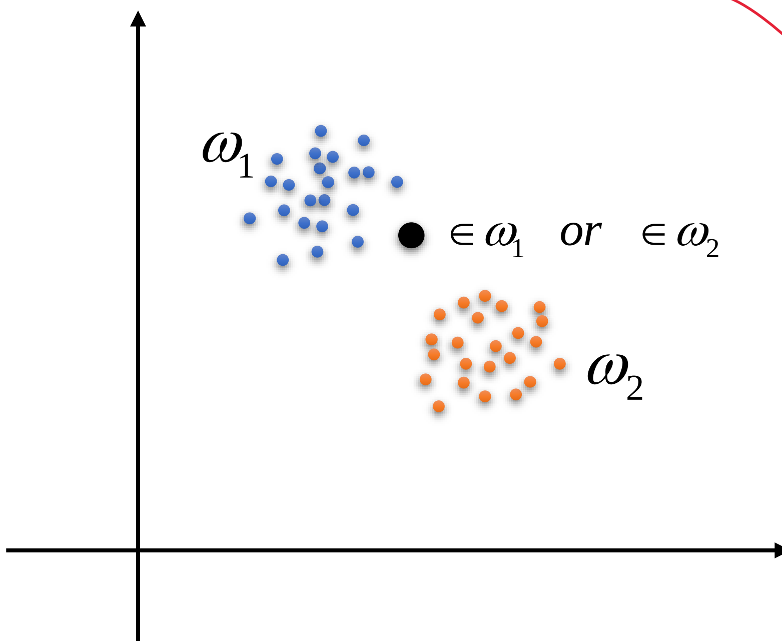
- 小结：
  - 6条性质
  - 正态分布下，不相关等价于独立
  - 马氏距离
  - 线性变换的正态性($\mathbf{A\Sigma A}^T$是对角阵)

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

贝叶斯分类器－正态分布下的最小错误率贝叶斯决策

最小错误率贝叶斯决策：

$$若P(\omega_i \mid \mathbf{x}) = \max_{j=1,2,\ldots,c} P(\mathbf{x}|\omega_j)\boxed{P(\omega_j),}\ 则\mathbf{x} \in \omega_i$$



首先，要估计类条件概率密度和先验概率。

贝叶斯分类器－正态分布下的最小错误率贝叶斯决策

- 讨论类条件概率密度是正态分布

$$g_i(\mathbf{x}) = P(\mathbf{x}|\omega_i)P(\omega_i)$$

$$= P(\omega_i)\left\{\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_i|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\}\right\}$$

取对数，仍用g表示

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

贝叶斯分类器-正态分布下的最小错误率贝叶斯决策

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

$\Downarrow$ 去掉常数项

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

$\Downarrow$ 决策面方程: $g_i(\mathbf{x}) = g_j(\mathbf{x})$

后验概率相等情况

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

$$\Updownarrow$$

$$g_j(\mathbf{x}) = \ln\left(P(\omega_j)\right) - \frac{1}{2}\ln\left(\left|\boldsymbol{\Sigma}_j\right|\right) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)$$

贝叶斯分类器－正态分布下的最小错误率贝叶斯决策

➢ 多元正态分布下的最小错误率**判别函数**为：

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

➢ 多元正态分布下的最小错误率贝叶斯**决策面方程**为：

$$g_i(\mathbf{x}) = g_j(\mathbf{x})$$

$$\Downarrow$$

$$\boxed{-\frac{1}{2}\left[\left(\mathbf{x}-\boldsymbol{\mu}_i\right)^T \boldsymbol{\Sigma}_i^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_i\right) - \left(\mathbf{x}-\boldsymbol{\mu}_j\right)^T \boldsymbol{\Sigma}_i^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_j\right)\right] - \frac{1}{2}\ln\left(\frac{\left|\boldsymbol{\Sigma}_i\right|}{\left|\boldsymbol{\Sigma}_j\right|}\right) + \ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right) = 0}$$

接下来，讨论特殊情况下的判别函数的表达式

贝叶斯分类器－正态分布下的最小错误率贝叶斯决策

- **Case 1**: 每类的协方差矩阵都相等，且类内各特征相互独立，具有相同的方差。各类的先验概率也相等。

椭圆退化为圆

$$\boxed{\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}}, P(\omega_i) = \text{常数}, i = 1, 2, \ldots, c$$

$$\Downarrow$$

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

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

$$\Downarrow$$

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

贝叶斯分类器-正态分布下的最小错误率贝叶斯决策

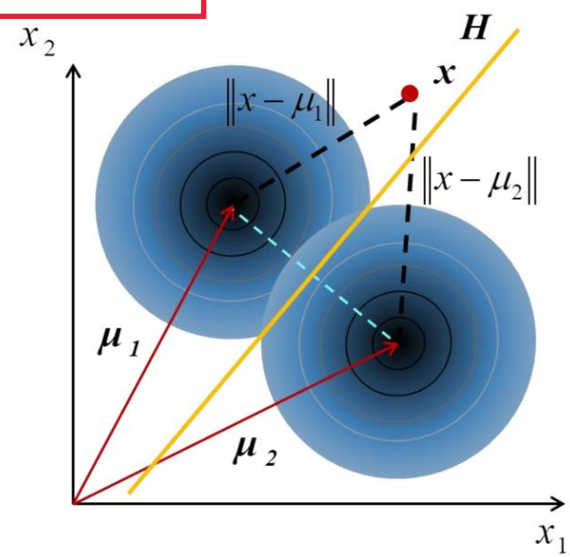
$$\boxed{g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)}$$

$$g_i(\mathbf{x}) = \max_{i=1,2,\ldots,c}\left\{-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)\right\} \Rightarrow \mathbf{x}\in\omega_i$$

$$\boxed{= \min_{i=1,2,\ldots,c}\left\{(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)\right\}}$$

欧式距离下的
最小距离分类器

决策面方程呢?



贝叶斯分类器-正态分布下的最小错误率贝叶斯决策

$$g_i(\mathbf{x}) = -\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)$$

决策面方程

$$(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i) = (\mathbf{x}-\boldsymbol{\mu}_j)^T(\mathbf{x}-\boldsymbol{\mu}_j)$$

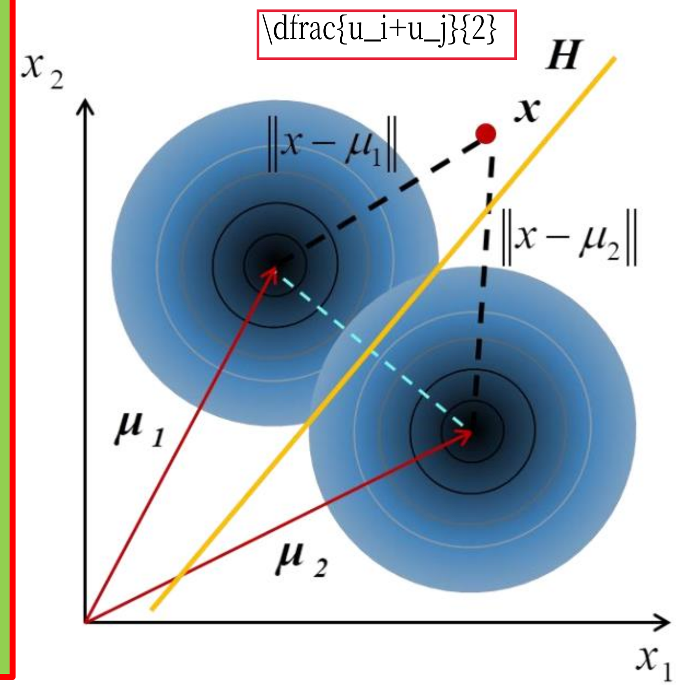
$$\mathbf{x}^T\mathbf{x} - 2\boldsymbol{\mu}_i^T\mathbf{x} + \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i = \mathbf{x}^T\mathbf{x} - 2\boldsymbol{\mu}_j^T\mathbf{x} + \boldsymbol{\mu}_j^T\boldsymbol{\mu}_j$$

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

$$2(\boldsymbol{\mu}_i^T - \boldsymbol{\mu}_j^T)\mathbf{x} = (\boldsymbol{\mu}_i^T - \boldsymbol{\mu}_j^T)(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$$

$$(\boldsymbol{\mu}_i^T - \boldsymbol{\mu}_j^T)\left(\mathbf{x} - \frac{\boldsymbol{\mu}_i + \boldsymbol{\mu}_j}{2}\right) = 0$$

分界面是正交平分均值向量的连线



贝叶斯分类器–正态分布下的最小错误率贝叶斯决策

- **Case 2:** 每类的协方差矩阵都相等，且类内各特征相互独立，具有相同的方差。

没有强调先验概率了

$$\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}, i = 1, 2, \ldots, c$$

$$\Downarrow$$

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

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

$$\Downarrow$$

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

贝叶斯分类器－正态分布下的最小错误率贝叶斯决策

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

$$\Downarrow$$

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

$$\Downarrow$$

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

$$\Downarrow$$

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

线性分类器

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

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

决策面方程呢？

贝叶斯分类器-正态分布下的最小错误率贝叶斯决策

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

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

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

决策面方程

$$(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \mathbf{x} - \frac{1}{2}\left(\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i - \boldsymbol{\mu}_j^T\boldsymbol{\mu}_j\right) + \sigma^2 \cdot \ln\frac{P(\omega_i)}{P(\omega_j)} = 0$$

$$(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) + \boxed{\frac{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2}} \cdot \sigma^2 \cdot \ln\frac{P(\omega_i)}{P(\omega_j)} = 0$$

$$(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T\left(\mathbf{x} - \frac{(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)}{2} + \frac{\sigma^2}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2} \cdot \ln\frac{P(\omega_i)}{P(\omega_j)}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\right) = 0$$

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

$(\mu_i-\mu_j)^T\left[x - \frac{\mu_i+\mu_j}{2} + \frac{(\mu_i-\mu_j)}{\|\mu_i-\mu_j\|^2}\sigma^2 \ln\frac{P(\omega_i)}{P(\omega_j)}\right] = 0$

$(\mu_i-\mu_j)^T \cdot \left[x - \left(\frac{\mu_i+\mu_j}{2} - \frac{\mu_i-\mu_j}{\|\mu_i-\mu_j\|^2} \cdot \sigma^2 \cdot \ln\frac{P(\omega_i)}{P(\omega_j)}\right)\right] = 0$

贝叶斯分类器-正态分布下的最小错误率贝叶斯决策

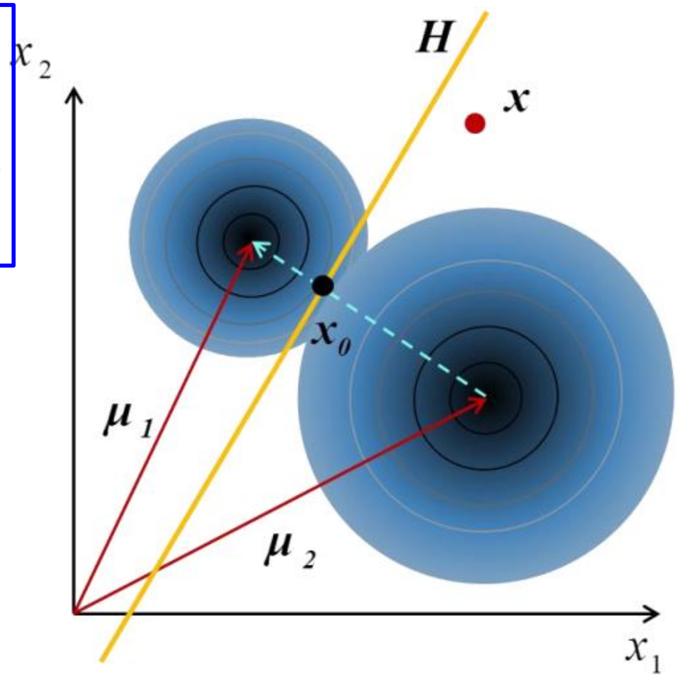
$$\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right)^T\left(\mathbf{x}-\frac{\left(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j\right)}{2}+\frac{\sigma^2}{\left\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right\|^2}\cdot\boxed{\ln\frac{P\left(\omega_i\right)}{P\left(\omega_j\right)}}\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right)\right)=0$$

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

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

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

分界面是正交均值向量的连线，向先验概率小的方向偏移



贝叶斯分类器-正态分布下的最小错误率贝叶斯决策

- **Case 3:** 每类的协方差矩阵都相等，各类的先验概率也相等。

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, P(\omega_i) = \text{常数}, i = 1, 2, \ldots, c$$

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

$$\Downarrow$$

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

$$\Downarrow$$

马式距离下的最小距离分类器

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

贝叶斯分类器－正态分布下的最小错误率贝叶斯决策

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

决策面方程

$$(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) = (\mathbf{x}-\boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)$$

$$\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} - 2\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{x} + \boldsymbol{\mu}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i = \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} - 2\boldsymbol{\mu}_j^T\boldsymbol{\Sigma}^{-1}\mathbf{x} + \boldsymbol{\mu}_j^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j$$

$$\boxed{2(\boldsymbol{\mu}_i^T - \boldsymbol{\mu}_j^T)\boldsymbol{\Sigma}^{-1}\mathbf{x} = \boldsymbol{\mu}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i - \boldsymbol{\mu}_j^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j}$$

$$2(\boldsymbol{\mu}_i^T - \boldsymbol{\mu}_j^T)\boldsymbol{\Sigma}^{-1}\mathbf{x} = (\boldsymbol{\mu}_i^T - \boldsymbol{\mu}_j^T)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$$

那两项是因为标量消掉了

$$(\boldsymbol{\mu}_i^T - \boldsymbol{\mu}_j^T)\boxed{\boldsymbol{\Sigma}^{-1}}\left(\mathbf{x} - \frac{\boldsymbol{\mu}_i + \boldsymbol{\mu}_j}{2}\right) = 0$$

右边：∵有

$u_i^T\Sigma^{-1}u_j = u_j^T\Sigma^{-1}u_i$

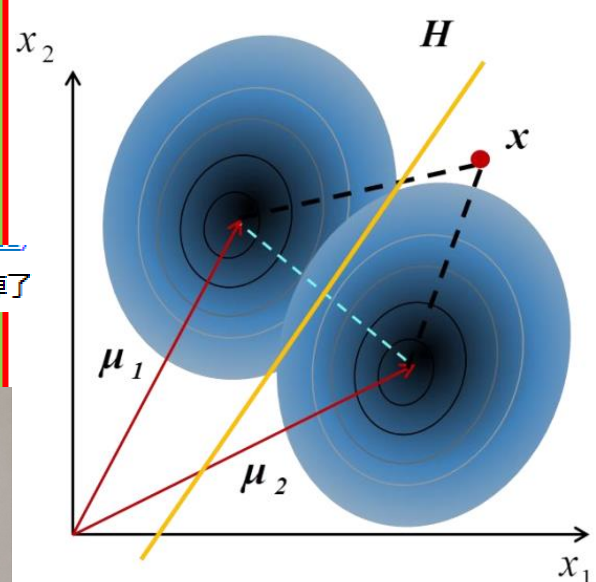
标量

$\Rightarrow$

$u_i^T\Sigma^{-1}u_i - u_j^T\Sigma^{-1}u_j + u_i^T\Sigma^{-1}u_j - u_j^T\Sigma^{-1}u_i$

$= (u_i^T - u_j^T)\Sigma^{-1}(u_i + u_j).$

分界面通过均值向量连线的中点



贝叶斯分类器-正态分布下的最小错误率贝叶斯决策

- **Case 4**: 每类的协方差矩阵都相等。

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, i = 1, 2, \ldots, c$$

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

线性分类器

$$\Downarrow$$

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

$$\Downarrow$$

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

$$\Downarrow$$

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

$$\boxed{\begin{gathered} g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0} \\ \mathbf{w}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i \\ w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln(P(\omega_i)) \end{gathered}}$$

贝叶斯分类器－正态分布下的最小错误率贝叶斯决策

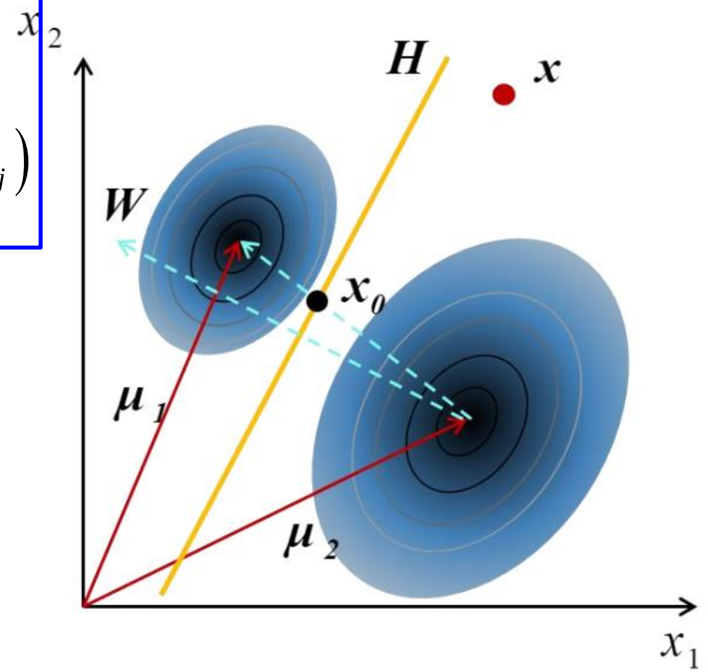
$$\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\frac{\left(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j\right)}{2}+\frac{1}{\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right)}\cdot\ln\frac{P\left(\omega_i\right)}{P\left(\omega_j\right)}\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right)\right)=0$$

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

$$\mathbf{w}=\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right)$$

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

分界面不通过均值向量连线的中点，向先验概率小的方向偏移



贝叶斯分类器－正态分布下的最小错误率贝叶斯决策

- **Case 5:** 每类的协方差矩阵不相等。

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

$$\Downarrow$$

非线性分类器

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

$$\mathbf{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1} \qquad \mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

$$w_{i0} = \ln\left(P(\omega_i)\right) - \frac{1}{2}\ln\left(|\boldsymbol{\Sigma}_i|\right) - \frac{1}{2}\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

决策面可能为超球面、超椭球面、超抛物面、超双曲面或超平面

贝叶斯分类器-正态分布下的最小错误率贝叶斯决策

- 小结：

➢多元正态分布下的最小错误率**判别函数**为：

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

➢多元正态分布下的最小错误率贝叶斯**决策面方程**为：

$$g_i(\mathbf{x}) = g_j(\mathbf{x})$$

$$\Downarrow$$

$$-\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - (\mathbf{x}-\boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\right] - \frac{1}{2}\ln\left(\frac{|\boldsymbol{\Sigma}_i|}{|\boldsymbol{\Sigma}_j|}\right) + \ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right) = 0$$

- **小结：**

$Case1:\boldsymbol{\Sigma}_i=\sigma^2\mathbf{I},P(\omega_i)=$常数$,i=1,2,\ldots,c$

$Case2:\boldsymbol{\Sigma}_i=\sigma^2\mathbf{I},i=1,2,\ldots,c$

$Case3:\boldsymbol{\Sigma}_i=\boldsymbol{\Sigma},P(\omega_i)=$常数$,i=1,2,\ldots,c$

$Case4:\boldsymbol{\Sigma}_i=\boldsymbol{\Sigma},i=1,2,\ldots,c$

$Case5:\boldsymbol{\Sigma}_i\neq\boldsymbol{\Sigma}_j,i,j=1,2,\ldots,c,j\neq i$

分界面是正交平分均值向量的连线

分界面是正交均值向量的连线，向先验概率小的方向偏移

分界面通过均值向量连线的中点

分界面不通过均值向量连线的中点，向先验概率小的方向偏移

非线性分类问题

- 正态分布下的贝叶斯分类器(步骤)

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

根据训练样本(有标签)估计先验概率

根据训练样本(有标签)估计均值向量和协方差矩阵

待识别样本代入判别函数，根据最大后验概率决策进行分类


若干素材取自网络，特此致谢！


谢谢聆听！