第二章 贝叶斯决策论

1． 根据贝叶斯决策理论，1 维特征空间中的平均误差概率可以表示为：

$$P(error)=\int_{-\infty}^{+\infty}P(error,x)\,dx=\int_{-\infty}^{+\infty}P(error|x)\,p(x)\,dx \qquad (1)$$

而对于特定的观察特征 x 两类问题的错误率为：

$$P(error|x)=\min\left[P(w_1|x),P(w_2|x)\right]$$

a) 对于两类问题有：

$$P(w_1|x)+P(w_2|x)=1$$

以及：

$$p(error|x)=\min\left[p(w_1|x),p(w_2|x)\right]\le\frac{1}{2}$$

因此我们可以得到：

$$p(error|x)-2\left[p(error|x)\right]^2\ge 0$$

$$\Rightarrow p(error|x)\le p(error|x)+p(error|x)-2\left[p(error|x)\right]^2$$

$$=2p(error|x)\left[1-p(error|x)\right]$$

$$=2\min\left[p(w_1|x),p(w_2|x)\right]\max\left[p(w_1|x),p(w_2|x)\right]$$

$$=2p(w_1|x)\,p(w_2|x)$$

误差率的上界在 $p(w_1|x)=p(w_2|x)=\dfrac{1}{2}$ 取得。

b) 如果我们令 $p(w_1|x)=p(w_2|x)=\dfrac{1}{2}$，则有：

$$p_a(error)=\int_{-\infty}^{+\infty}a\,p(w_1|x)\,p(w_2|x)\,p(x)\,dx=\frac{a}{4}\int_{-\infty}^{+\infty}p(x)\,dx<\frac{1}{2}\text{（因为 }a<2\text{）}$$

$$p(error)=\int_{-\infty}^{+\infty}\min\left[p(w_1|x),p(w_2|x)\right]p(x)\,dx=\frac{1}{2}\int_{-\infty}^{+\infty}p(x)\,dx=\frac{1}{2}$$

显而易见：$p_a(error)<p(error)$，因此当 $a<2$ 时，无法得到误差率的上界。

c) 因为：

$$p(error|x)\ge p(error|x)-\left[p(error|x)\right]^2$$

$$=p(error|x)\left[1-p(error|x)\right]$$

$$=\min\left[p(w_1|x),p(w_2|x)\right]\max\left[p(w_1|x),p(w_2|x)\right]$$

$$=p(w_1|x)\,p(w_2|x)$$

所以，$p(w_1|x)p(w_2|x)$能过给出误差率的下界。

d) 因为：

$$p_b(error)=\int_{-\infty}^{+\infty} b\, p(w_1|x)\, p(w_2|x)\, p(x)\, dx$$

$$>\int_{-\infty}^{+\infty} p(w_1|x)\, p(w_2|x)\, p(x)\, dx \qquad (b>1)$$

当$b>1$时无法给出误差率的下界。

2. $p(x|w_i)\propto \exp(-|x-a_i|/b_i)$

a) 令：$p(x|w_i)=c_i\exp(-|x-a_i|/b_i)$，则：

$$\int_{-\infty}^{+\infty} p(x|w_i)\,dx=\int_{-\infty}^{+\infty} c_i\exp(-|x-a_i|/b_i)\,dx$$

$$=\int_{-\infty}^{a_i} c_i\exp((x-a_i)/b_i)\,dx+\int_{a_i}^{+\infty} c_i\exp(-(x-a_i)/b_i)\,dx$$

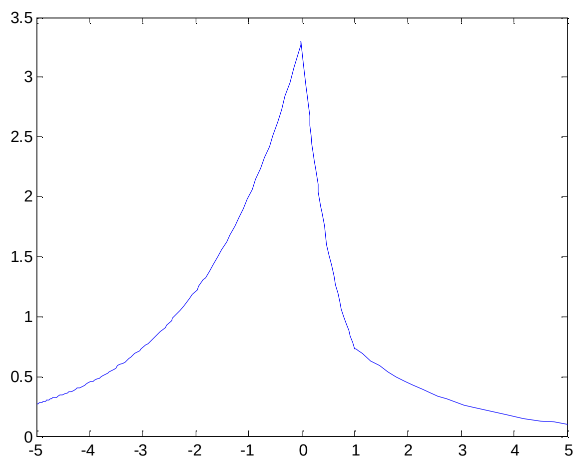
$$=c_ib_i\exp((x-a_i)/b_i)\Big|_{-\infty}^{a_i}-c_ib_i\exp(-(x-a_i)/b_i)\Big|_{a_i}^{+\infty}$$

$$=2c_ib_i=1$$

因此：$c_i=\dfrac{1}{2b_i}$，$p(x|w_i)=\dfrac{1}{2b_i}\exp(-|x-a_i|/b_i)$

b) 似然比$=\dfrac{p(x|w_1)}{p(x|w_2)}=\dfrac{b_2}{b_1}\exp\left(\dfrac{|x-a_2|}{b_2}-\dfrac{|x-a_1|}{b_1}\right)$；

c) $a_1=0,b_1=1,a_2=1,b_2=2$，则：似然比$=2\exp\left(\dfrac{|x-1|}{2}-|x|\right)$



12．$w_{\max}(\mathbf{x})$为类别状态，有$P(w_{\max}|\mathbf{x}) \geq P(w_i|\mathbf{x})$，$i=1,\cdots,c$

a) 对于c类问题，有如下关系成立：

$$\sum_{i=1}^{c} P(w_i|\mathbf{x}) = 1 \tag{5}$$

假设$P(w_{\max}|\mathbf{x}) < \frac{1}{c}$，则$P(w_i|\mathbf{x}) \leq P(w_{\max}|\mathbf{x}) < \frac{1}{c}$，$i=1,\cdots,c$，因此有：

$$\sum_{i=1}^{c} P(w_i|\mathbf{x}) < \sum_{i=1}^{c} \frac{1}{c} = 1 \tag{6}$$

(6)式与(5)式矛盾，因此$P(w_{\max}|\mathbf{x}) \geq \frac{1}{c}$。

b) 根据最小错误率准则：

$$\begin{aligned} P(error) &= \int P(error|\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x} \\ &= \int \sum_{\substack{i=1 \\ w_i \neq w_{\max}}}^{c} P(w_i|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x} \\ &= \int \left(1 - P(w_{\max}|\mathbf{x})\right) p(\mathbf{x})\, d\mathbf{x} \\ &= 1 - \int P(w_{\max}|\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x} \end{aligned}$$

d) 续上式：

$$\begin{aligned} P(error) &= 1 - \int P(w_{\max}|\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x} \\ &\leq 1 - \int \frac{1}{c} p(\mathbf{x})\, d\mathbf{x} = 1 - \frac{1}{c} = \frac{c-1}{c} \end{aligned}$$

e) 当$P(w_1|\mathbf{x}) = P(w_2|\mathbf{x}) = \cdots = P(w_c|\mathbf{x})$时，有$P(w_{\max}|\mathbf{x}) = \frac{1}{c}$，此情况下：

$$P(error) = \frac{c-1}{c}$$

23．三维正态分布$p(\mathbf{x}|w) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，其中：

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix}$$

a) $\mathbf{x}_0 = (0.5, 0, 1)^t$

$$p(\mathbf{x}_0|w) = \frac{1}{(2p)^{3/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x}_0 - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}) \right]$$

$$|\mathbf{\Sigma}| = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{vmatrix} = \begin{vmatrix} 5 & 2 \\ 2 & 5 \end{vmatrix} = 21$$

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5/21 & -2/21 \\ 0 & -2/21 & 5/21 \end{pmatrix}$$

$$(\mathbf{x}_0 - \boldsymbol{\mu})^t \mathbf{\Sigma}^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}) = \left[\begin{pmatrix} 0.5 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}\right]^t \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5/21 & -2/21 \\ 0 & -2/21 & 5/21 \end{pmatrix} \left[\begin{pmatrix} 0.5 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}\right]$$

$$= \begin{pmatrix} -0.5 \\ -8/21 \\ -1/21 \end{pmatrix}^t \begin{pmatrix} -0.5 \\ -8/21 \\ -1/21 \end{pmatrix} = 1.06$$

$$p(\mathbf{x}_0 | w) = \frac{1}{(2p)^{3/2} (21)^{1/2}} e^{-\frac{1}{2} \times 1.06} = 8.16 \times 10^3$$

b) $|\mathbf{\Sigma} - l\,\mathbf{I}| = 0$

计算特征值：

$$\begin{vmatrix} 1-l & 0 & 0 \\ 0 & 5-l & 2 \\ 0 & 2 & 5-l \end{vmatrix} = (1-l)\left[(5-l)^2 - 4\right] = 0$$

$$l = 1, l = 3, l = 7$$

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{pmatrix}$$

计算特征向量：

$$\mathbf{\Sigma e}_1 = l_1 \mathbf{e}_1$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \begin{pmatrix} x_1 \\ 5x_2 + 2x_3 \\ 2x_2 + 5x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$\begin{cases} 5x_2 + 2x_3 = x_2 \\ 2x_2 + 5x_3 = x_3 \end{cases}$，解得：$x_2 = 0, x_3 = 0$，令 $x_1 = 1$，则：

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

同理：

$$\begin{pmatrix} x_1 \\ 5x_2+2x_3 \\ 2x_2+5x_3 \end{pmatrix} = \begin{pmatrix} 3x_1 \\ 3x_2 \\ 3x_3 \end{pmatrix}$$，解得：$x_1=0$，$x_2=-x_3$，令$x_2=1$，则：

$$\mathbf{e}_2=\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$，规格化为：$$\mathbf{e}_2=\begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ 5x_2+2x_3 \\ 2x_2+5x_3 \end{pmatrix} = \begin{pmatrix} 7x_1 \\ 7x_2 \\ 7x_3 \end{pmatrix}$$，解得：$x_1=0$，$x_2=x_3$，令：$x_2=1$，则：

$$\mathbf{e}_3=\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$，规格化为：$$\mathbf{e}_3=\begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

因此：

$$\mathbf{\Phi}=\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$\mathbf{A}_w=\mathbf{\Phi}\mathbf{\Lambda}^{-\frac{1}{2}}=\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{7} \end{pmatrix}$$

$$=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & -1/\sqrt{14} \end{pmatrix}$$

c) $$\mathbf{x}_w=\mathbf{A}_w^t\left(\mathbf{x}_0-\boldsymbol{\mu}\right)$$

$$=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & -1/\sqrt{14} \end{pmatrix}\left[\begin{pmatrix} 0.5 \\ 0 \\ 1 \end{pmatrix}-\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}\right]$$

$$=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & -1/\sqrt{14} \end{pmatrix}\begin{pmatrix} -0.5 \\ -2 \\ -1 \end{pmatrix}=\begin{pmatrix} -0.5 \\ 1/\sqrt{6} \\ -3/\sqrt{14} \end{pmatrix}$$

d) 从$\mathbf{x}_0$到$\boldsymbol{\mu}$的马氏距离平方：

$$r^2=\left(\mathbf{x}_0-\boldsymbol{\mu}\right)^t\mathbf{\Sigma}^{-1}\left(\mathbf{x}_0-\boldsymbol{\mu}\right)=1.06$$

从$\mathbf{x}_w$到$\mathbf{0}$的马氏距离平方：

$$r_w^2 = \mathbf{x}_w^t \mathbf{x}_w = \begin{pmatrix} -0.5 & 1/\sqrt{6} & -3/\sqrt{14} \end{pmatrix} \begin{pmatrix} -0.5 \\ 1/\sqrt{6} \\ -3/\sqrt{14} \end{pmatrix} = 1.06$$

因此：$r = r_w$。

e) $p\left(\mathbf{x}_0 \middle| N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)\right) \sim p\left(\mathbf{x}_0\right) = \dfrac{1}{\left(2p\right)^{\frac{d}{2}} \left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}} \exp\left[-\dfrac{1}{2}\left(\mathbf{x}_0 - \boldsymbol{\mu}\right)^t \boldsymbol{\Sigma}^{-1} \left(\mathbf{x}_0 - \boldsymbol{\mu}\right)\right]$

如果：$\mathbf{x}' = \mathbf{T}^t \mathbf{x}$，则：

$$\boldsymbol{\mu}' = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}'_k = \frac{1}{n}\sum_{k=1}^{n} \mathbf{T}^t \mathbf{x}'_{0k} = \frac{1}{n}\mathbf{T}^t \sum_{k=1}^{n} \mathbf{x}'_{0k} = \mathbf{T}^t \boldsymbol{\mu}$$

$$\boldsymbol{\Sigma}' = \sum_{k=1}^{n} \left(\mathbf{x}'_k - \boldsymbol{\mu}'\right)\left(\mathbf{x}'_k - \boldsymbol{\mu}'\right)^t$$

$$= \sum_{k=1}^{n} \mathbf{T}^t \left(\mathbf{x}_{0k} - \boldsymbol{\mu}\right)\left(\mathbf{x}_{0k} - \boldsymbol{\mu}\right)^t \mathbf{T}$$

$$= \mathbf{T}^t \left[\sum_{k=1}^{n} \left(\mathbf{x}_{0k} - \boldsymbol{\mu}\right)\left(\mathbf{x}_{0k} - \boldsymbol{\mu}\right)^t\right] \mathbf{T}$$

$$= \mathbf{T}^t \boldsymbol{\Sigma} \mathbf{T}$$

因此有：$p\left(\mathbf{T}^t \mathbf{x}_0 \middle| N\left(\mathbf{T}^t \boldsymbol{\mu}, \mathbf{T}^t \boldsymbol{\Sigma} \mathbf{T}\right)\right)$。

f) 因为：$\boldsymbol{\Sigma\Phi} = \boldsymbol{\Phi\Lambda}$，所以：$\boldsymbol{\Sigma} = \boldsymbol{\Phi\Lambda\Phi}^{-1}$，同时 $\boldsymbol{\Phi}$ 为对称矩阵，因此：$\boldsymbol{\Phi}^{-1} = \boldsymbol{\Phi}^t$。

$$\mathbf{A}_w^t \boldsymbol{\Sigma} \mathbf{A}_w = \left(\boldsymbol{\Phi\Lambda}^{-\frac{1}{2}}\right)^t \boldsymbol{\Sigma} \left(\boldsymbol{\Phi\Lambda}^{-\frac{1}{2}}\right)$$

$$= \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{\Phi}^t \boldsymbol{\Phi\Lambda\Phi}^t \boldsymbol{\Phi\Lambda}^{-\frac{1}{2}}$$

$$= \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{\Lambda\Lambda}^{-\frac{1}{2}} = \mathbf{I}$$