# 第三、四章作业答案

1. 说明最大后验概率判决准则为什么可以被成为最小错误概率判决准则。

最大后验概率判决：

$$p(\boldsymbol{X}|\omega_1)P(\omega_1) > p(\boldsymbol{X}|\omega_2)P(\omega_2) \Rightarrow \boldsymbol{X} \in \omega_1$$
$$p(\boldsymbol{X}|\omega_1)P(\omega_1) < p(\boldsymbol{X}|\omega_2)P(\omega_2) \Rightarrow \boldsymbol{X} \in \omega_2$$

因此，决策面为

$$p(\boldsymbol{X}|\omega_1)P(\omega_1) = p(\boldsymbol{X}|\omega_2)P(\omega_2)$$

在课本图 3.2 中对应于图(c)的情形。

而错误概率：

$$P(e) = P(x\text{在}\Omega_2\text{中}|\omega_1) + P(x\text{在}\Omega_1\text{中}|\omega_2) = \int_{\Omega_2} p(x|\omega_1)P(\omega_1)\mathrm{d}x + \int_{\Omega_1} p(x|\omega_2)P(\omega_2)\mathrm{d}x$$

即由阈值 t 所确定的阴影部分的面积，显然阴影部分的面积在 $p(x|\omega_1)P(\omega_1) = p(x|\omega_2)$

$P(\omega_2)$ 时最小，所以最大后验概率判决准则也是最小错误概率判决准则。

2. 已知两个一维模式类别的类概率密度函数为

$$p(x|\omega_1) = \begin{cases} x, & 0 \le x < 1 \\ 2-x, & 1 \le x \le 2 \\ 0, & else \end{cases}$$

$$p(x|\omega_2) = \begin{cases} x-1, & 1 \le x < 2 \\ 3-x, & 2 \le x \le 3 \\ 0, & else \end{cases}$$

先验概率分别为 $p(\omega_1) = 0.4, p(\omega_2) = 0.6$。试求最大后验概率判决函数以及总的分类错误概率 $P(e)$。

$$p(\boldsymbol{X}|\omega_1)P(\omega_1) > p(\boldsymbol{X}|\omega_2)P(\omega_2) \Rightarrow \boldsymbol{X} \in \omega_1$$
$$p(\boldsymbol{X}|\omega_1)P(\omega_1) > p(\boldsymbol{X}|\omega_2)P(\omega_2) \Rightarrow \boldsymbol{X} \in \omega_2$$

$$p(x|\omega_1)P(\omega_1) = \begin{cases} 0.4x, & 0 \le x < 1 \\ 0.8-0.4x, & 1 \le x \le 2 \\ 0, & else \end{cases}$$

$$p(x|\omega_2)P(\omega_2) = \begin{cases} 0.6x-0.6, & 1 \le x < 2 \\ 1.8-0.6x, & 2 \le x \le 3 \\ 0, & else \end{cases}$$

$0 < x < 1:$
$p(x|\omega_1)P(\omega_1) = 0.4x$
$p(x|\omega_2)P(\omega_2) = 0$
$p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2) \Rightarrow x \in \omega_1$
$1 \le x < 2:$
$p(x|\omega_1)P(\omega_1) = 0.8 - 0.4x$
$p(x|\omega_2)P(\omega_2) = 0.6x - 0.6$
$s.t. p(x|\omega_1)P(\omega_1) = p(x|\omega_2)P(\omega_2) \Rightarrow x = 1.4$

$1 \le x < 1.4:$
$p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2) \Rightarrow x \in \omega_1$
$1.4 < x < 2:$
$p(x|\omega_1)P(\omega_1) < p(x|\omega_2)P(\omega_2) \Rightarrow x \in \omega_2$
$2 \le x < 3:$
$p(x|\omega_1)P(\omega_1) = 0$
$p(x|\omega_2)P(\omega_2) = 1.8 - 0.6x$
$p(x|\omega_1)P(\omega_1) < p(x|\omega_2)P(\omega_2) \Rightarrow x \in \omega_2$

综上，

$$\begin{cases} \omega_1, & 0 < x < 1.4 \\ \omega_2, & 1.4 < x < 3 \\ \text{无法判断}, & otherwise \end{cases}$$

分类错误概率：

$$P(e) = \int_0^{1.4} p(x|\omega_2)P(\omega_2)\mathrm{d}x + \int_{1.4}^{3} p(x|\omega_1)P(\omega_1)\mathrm{d}x$$
$$= \int_1^{1.4} (0.6x - 0.6)\mathrm{d}x + \int_{1.4}^{2} (0.8 - 0.4x)\mathrm{d}x = 0.12$$

3. 从似然比形式的角度说明最大后验概率判决准则同最小风险判决准则之间的联系与区别。

注意：似然比为相应两个类别的类条件概率密度函数之比的形式。

即：$l_{12} = \dfrac{p(\boldsymbol{X}|\omega_1)}{p(\boldsymbol{X}|\omega_2)}$

似然比形式的判决规则：

$l_{12} > \theta_{12} \Rightarrow \boldsymbol{X} \in \omega_1$
$l_{12} < \theta_{12} \Rightarrow \boldsymbol{X} \in \omega_2$

最大后验概率判决规则中：$\theta_{12}=\dfrac{P(\omega_2)}{P(\omega_1)}$

最小风险判决规则中：$\theta_{12}=\dfrac{[L(\alpha_1\mid\omega_2)-L(\alpha_2\mid\omega_2)]P(\omega_2)}{[L(\alpha_2\mid\omega_1)-L(\alpha_1\mid\omega_1)]P(\omega_1)}$

两者在似然比表现形式上面完全一样，但是判决阈值的计算不同。若令最小风险判决准则中判决正确的风险为 0，判决错误的代价为 1，则二者的判决阈值计算也是一样的。所以最大后验概率判决准则是 0-1 风险判决准则。

4. 在图像识别中，假定有灌木丛和坦克两种类型，它们的先验概率分别是 0.8 和 0.2，损失函数如下表所示，其中 $\omega_1$ 和 $\omega_2$ 分别表示灌木丛和坦克，$\alpha_1$ 和 $\alpha_2$ 表示判决为灌木丛和坦克，$\alpha_3$ 表示拒绝判决。

| | $\omega_1$ | $\omega_2$ |
|---|---|---|
| $\alpha_1$ | 0.5 | 6 |
| $\alpha_2$ | 2 | 1 |
| $\alpha_3$ | 1.5 | 1.5 |

现在做了三次实验，从类概率密度函数曲线上查得三个样本 $\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3$ 的类概率密度值如下：

$$
\begin{aligned}
&\boldsymbol{X}_1: p(\boldsymbol{X}_1\mid\omega_1)=0.1,\ p(\boldsymbol{X}_1\mid\omega_2)=0.7\\
&\boldsymbol{X}_2: p(\boldsymbol{X}_2\mid\omega_1)=0.3,\ p(\boldsymbol{X}_2\mid\omega_2)=0.45\\
&\boldsymbol{X}_3: p(\boldsymbol{X}_3\mid\omega_1)=0.6,\ p(\boldsymbol{X}_3\mid\omega_2)=0.5
\end{aligned}
$$

(1) 试用贝叶斯最小误判概率准则判决三个样本各属于哪一个类型。

(2) 假定只考虑前两种判决，试用贝叶斯最小风险判决准则判决三个样本各属于哪一个类型。

(3) 把拒绝判决考虑在内，重新考核三次实验的结果。

(1) 贝叶斯最小误判准则：

$$
\begin{aligned}
&p(\boldsymbol{X}\mid\omega_1)P(\omega_1)>p(\boldsymbol{X}\mid\omega_2)P(\omega_2)\Rightarrow\boldsymbol{X}\in\omega_1\\
&p(\boldsymbol{X}\mid\omega_1)P(\omega_1)>p(\boldsymbol{X}\mid\omega_2)P(\omega_2)\Rightarrow\boldsymbol{X}\in\omega_2
\end{aligned}
$$

第一个样本：

$$p(\boldsymbol{X}_1 | \omega_1)P(\omega_1) = 0.1 \times 0.8 = 0.08$$
$$p(\boldsymbol{X}_1 | \omega_2)P(\omega_2) = 0.14$$
$$\boldsymbol{X}_1 \in \omega_2$$

第二个样本：

$$p(\boldsymbol{X}_2 | \omega_1)P(\omega_1) = 0.24$$
$$p(\boldsymbol{X}_2 | \omega_2)P(\omega_2) = 0.09$$
$$\boldsymbol{X}_2 \in \omega_1$$

第三个样本：

$$p(\boldsymbol{X}_3 | \omega_1)P(\omega_1) = 0.48$$
$$p(\boldsymbol{X}_3 | \omega_2)P(\omega_2) = 0.1$$
$$\boldsymbol{X}_3 \in \omega_1$$

(2) 考虑前两种判决，贝叶斯最小风险判决准则

$$l_{12} > \theta_{12} \Rightarrow x \in \omega_1$$
$$l_{12} < \theta_{12} \Rightarrow x \in \omega_1$$
$$l_{12} = \frac{p(x | \omega_1)}{p(x | \omega_2)}$$
$$\theta_{12} = \frac{[L(\alpha_1 | \omega_2) - L(\alpha_2 | \omega_2)]P(\omega_2)}{[L(\alpha_2 | \omega_1) - L(\alpha_1 | \omega_1)]P(\omega_1)} = \frac{5}{6}$$

第一个样本：

$$l_{12} = \frac{1}{7} < \theta_{12} = \frac{5}{6}$$
$$\boldsymbol{X}_1 \in \omega_2$$

第二个样本：

$$l_{12} = \frac{2}{3} < \theta_{12} = \frac{5}{6}$$
$$\boldsymbol{X}_2 \in \omega_2$$

第三个样本：

$$l_{12} = 1.2 > \theta_{12} = \frac{5}{6}$$
$$\boldsymbol{X}_3 \in \omega_1$$

(3) 考虑拒绝判决

注意：书中的推导是在假设所有的正确判决损失均为 0，所有的错误判决损失均为 $\lambda_F$，所

有的拒绝判决损失均为$\lambda_R$的情况下进行的。而本题正确判决损失不为 0，错误判决损失也不都相同。

$$R(\alpha_{N+1} \mid \boldsymbol{X}) < R(\alpha_i \mid \boldsymbol{X}), \forall i = 1, 2, \cdots, N \Rightarrow \boldsymbol{X} \in \alpha_{N+1}$$

$$R(\alpha_i \mid \boldsymbol{X}) = \sum_{j=1}^{N} L(\alpha_i \mid \omega_j) P(\omega_j \mid \boldsymbol{X}) = \sum_{j=1}^{N} L(\alpha_i \mid \omega_j) \frac{p(\boldsymbol{X} \mid \omega_j) P(\omega_j)}{\sum_{j=1}^{N} p(\boldsymbol{X} \mid \omega_j) P(\omega_j)}, i = 1, 2, 3, \cdots, N+1$$

第一个样本：

$$R(\alpha_1 \mid \boldsymbol{X}_1) = 4$$
$$R(\alpha_2 \mid \boldsymbol{X}_1) = \frac{15}{11} \approx 1.36$$
$$R(\alpha_3 \mid \boldsymbol{X}_1) = 1.5$$
$$\boldsymbol{X}_1 \in \omega_2$$

第二个样本：

$$R(\alpha_1 \mid \boldsymbol{X}_2) = 2$$
$$R(\alpha_2 \mid \boldsymbol{X}_2) = \frac{19}{11} \approx 1.73$$
$$R(\alpha_3 \mid \boldsymbol{X}_2) = 1.5$$

拒绝判决

第三个样本：

$$R(\alpha_1 \mid \boldsymbol{X}_3) = \frac{42}{29} \approx 1.45$$
$$R(\alpha_2 \mid \boldsymbol{X}_3) = \frac{53}{29} \approx 1.83$$
$$R(\alpha_3 \mid \boldsymbol{X}_3) = 1.5$$
$$\boldsymbol{X}_3 \in \omega_1$$

5. 简要分析贝叶斯统计判决准则的缺陷以及 Neyman-Peason 判决准则的基本思想。

贝叶斯风险判决准则的缺陷：需要已知先验概率和类条件概率密度且二者不变。在最小风险判决情况需要对误判风险做出恰当的定义。

Neyman-Peason 判决规则基本思想：在保持一类分类错误概率不变的情况下，使另外一类分类错误概率最小。

6. 简要分析最小最大判决准则的基本思想。它的错分概率是最小的吗？为什么？

最小最大判决准则的基本思想：在最坏的情况下取得最好的结果，在其他情况下并不追求性能的最佳化。在最坏的情况下，它给出最好的结果，而在其他情况下给出的结果既不是最差

也不是最好。具体来讲就是使最大的总平均风险最小化而给出的分类判决。

显然，它的错分概率并不是最小的，因为它只获得最坏的情况下的最好的结果。采取的是一种过于保守的判决方法，所以它的分类器性能在大多数情况下相比于贝叶斯分类器是下降的。

7. 二维空间中的两类样本均服从正态分布，其参数分别为：

均值向量：$\boldsymbol{\mu}_1=(1,0)^T,\boldsymbol{\mu}_2=(-1,0)^T$

协方差矩阵：$\boldsymbol{\Sigma}_1=\begin{pmatrix}1&0\\0&1\end{pmatrix},\boldsymbol{\Sigma}_2=\begin{pmatrix}2&0\\0&2\end{pmatrix}$

且两类的先验概率相等，试证明其基于最小错误率判决准则的决策分界面方程为一圆，并求其方程。

$$d=2,\left|\boldsymbol{\Sigma}_1\right|=1,\left|\boldsymbol{\Sigma}_2\right|=4,\boldsymbol{\Sigma}_1^{-1}=\begin{pmatrix}1&0\\0&1\end{pmatrix},\boldsymbol{\Sigma}_2^{-1}=\begin{pmatrix}1/2&0\\0&1/2\end{pmatrix}$$

$$p(\boldsymbol{X}\mid\omega_1)=\frac{1}{2\pi}\exp\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_1)\}$$

$$p(\boldsymbol{X}\mid\omega_2)=\frac{1}{4\pi}\exp\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_2)\}$$

由 $p(\boldsymbol{X}\mid\omega_1)P(\omega_1)=p(\boldsymbol{X}\mid\omega_2)P(\omega_2)$ 得

$$\frac{1}{2\pi}\exp\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_1)^T\Sigma_1^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_1)\}=\frac{1}{4\pi}\exp\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_2)\}$$

令 $\boldsymbol{X}=\begin{pmatrix}x_1\\x_2\end{pmatrix}$，并对上式左右同时取自然对数

$$\ln 2-\frac{1}{2}[(x_1-1)^2+x_2^2]=-\frac{1}{4}(x_1+1)^2-\frac{1}{4}x_2^2$$

整理得到

$$(x_1-3)^2+x_2^2=8+4\ln 2$$

8. 参数估计和非参数估计有何区别？试述最大似然估计和 Parzen 窗估计的基本原理。

参数估计：适用于类条件概率密度的函数形式已知，但是相关参数未知的情况。此时相应的估计问题转换为一个参数估计问题求解。

非参数估计：适用于类条件概率密度函数的形式和相关参数都未知的情况。此时需要根据给定的样本同时确定类条件概率密度的函数形式和相关参数。

最大似然估计的基本思想：对于有监督的样本集，按照所属类别（类别数为 S）将样本集中

所有样本划分为 S 个子集。另假设所有这些样本按照类条件概率密度 $p(\boldsymbol{X}|\omega_i)$ 从总体中独立抽取，这样，若 $p(\boldsymbol{X}|\omega_i)$ 的函数形式已知而仅是参数未知（参数集合为 $\theta_i$ ）只要能用观测样本确定出 $\theta_i$ ,即可以求得 $p(\boldsymbol{X}|\omega_i)$ 的估计。

Parzen 窗估计法的基本思想：把包含估计点 $\boldsymbol{X}$ 在内的区域 $R_n$ 选为训练样本数 n 的函数。用变换以后的窗函数来表征特征空间中围绕 $\boldsymbol{X}$ 的超立方体区域 $R_n$ ，得到第 n 次操作以后样本落入 $R_n$ 中的个数与 $R_n$ 的体积之比即为所求概率密度的估计式，此时只需要窗函数满足 $\phi(u) \geq 0, \quad \int_{E_d} \phi(u)du = 1$ 。