第二章 贝叶斯决策论

一、填空题

1. 已知样本x的具体数值，其从属于某个类别的概率是（后验概率）；已知样本x属于某个类别，则x取某个具体数值的概率称为（类条件概率）。
2. 常见的3种分类准则是(最大后验概率分类准则),(最小错误概率分类准则)，（最小风险分类准则）。
3. 常见的两种分类器设计准则是（最小平均错误概率准则)、(最小平均风险准则)
4. 面向 d 维数据的线性分类器决策面是 d 维空间中的一个$(d-1)$维的（超平面）。

二、判断题

5. 在分类问题中，类条件概率是概率质量，后验概率是概率密度。(×)
6. 有些问题中，无需了解先验概率也可以直接估计后验概率。(√)
7. 最大后验概率分类准则和最小错误概率分类准则在任何情况下给出的分类结果都是一样的。(√)
8. 最小错误概率分类准则给出的分类结果总是正确的。(×)

三、选择题

9. 假设样本$\boldsymbol{x} \in \Re^3$，服从多元正态分布$\mathcal{N}(\boldsymbol{\mu}, \Sigma)$，则类条件概率密度函数的有效参数数量为：(B)

A. 12 个　　B. 9 个　　C. 6 个　　D. 不确定

10. 对于一维样本二分类问题，两类ω_1, ω_2均服从正态分布，假设按照最小平均错误概率准则设计的分类器为 C1，分类阈值为 T1，按照最小平均风险准则设计的分类器为 C2，T2，如果风险系数$\lambda_{12} > \lambda_{21}$，$\lambda_{11} = \lambda_{22}$则以下论断正确的是：(B)

A. T1 比 T2 更靠近ω_1类的均值　　B. T2 比 T1 更靠近ω_1类的均值

C. T1 与 T2 相同　　D. 无法判断

四、简答题

11. 如果对于某个数据集的概率分布模型，最大似然估计和最大后验概率估计给

出结果完全相同，试解释可能造成这一结果的原因（可画图说明）

最大似然估计的结果是似然性函数的最大值解，记为$\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, p(X;\boldsymbol{\theta})$；最大后验概率估计的结果是后验概率的最大值解，等价于求$\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, p(X;\boldsymbol{\theta})\,p(\boldsymbol{\theta})$。如果两种方法的解完全相同，说明似然函数$p(X;\boldsymbol{\theta})$与参数$\boldsymbol{\theta}$的先验概率分布可能存在相同的最大解。（图略）

12. 写出 K 个高斯成分的 GMM 模型的对数似然函数的数学形式

包含 K 个高斯成分的 GMM 模型的对数似然函数数学形式如下：

$$\sum_{i=1}^{N}\ln\left(\sum_{j=1}^{K}p(\boldsymbol{x}_i;\boldsymbol{\mu}_j,\Sigma_j)P_j\right)$$

其中$\boldsymbol{x}_i$为第i个样本，$\boldsymbol{\mu}_j,\Sigma_j,P_j$分别为第$j$个高斯成分的均值向量、协方差矩阵与比例系数。

13. 请写出 EM 算法的 M 步骤中，各高斯成分的均值向量$\boldsymbol{\mu}_j$、协方差矩阵Σ_j和比例系数P_j在 t 时刻的迭代更新公式。

答：

(1) 第 j 个高斯成分的均值向量$\boldsymbol{\mu}_j$的更新公式为：

$$\hat{\boldsymbol{\mu}}_j=\frac{\sum_{i=1}^{N}\gamma_{ji}^{t}\,\boldsymbol{x}_i}{\sum_{i=1}^{N}\gamma_{ji}^{t}}\Longrightarrow\boldsymbol{\mu}_j^{t+1}$$

(2) 第 j 个高斯成分的协方差矩阵Σ_j的更新公式为：

$$\hat{\Sigma}_j=\frac{\sum_{i=1}^{N}\gamma_{ji}^{t}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_j\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_j\right)^{T}}{\sum_{i=1}^{N}\gamma_{ji}^{t}}\Longrightarrow\Sigma_j^{t+1}$$

(3) 第 j 个高斯成分的均值向量P_j的更新公式为：

$$\hat{P}_j=\frac{\sum_{i=1}^{N}\gamma_{ji}^{t}}{N}\Longrightarrow P_j^{t+1}$$

其中，$\gamma_{ji}^{t}=P(j|x_i;\Theta^t)$，表示 t 时刻样本$\boldsymbol{x}_t$从属于第 j 个高斯成分的概率。

五、计算（画图）题

14. 假设一个学校里有 60%的男生和 40%的女生。女生点外卖的人数和吃食堂的人数相等，所有男生都吃外卖。一个人在校门口看到了一个拿着外卖的学生，那么这个学生时女生的概率是多少？

设男生为ω_1类，女生为ω_2类，样本的取值为吃外卖$x=$‘外卖’与吃食堂$x=$′食堂′两个离散取值。则有：

$$P(\omega_1)=0.6, P(\omega_2)=0.4$$
$$P\left(x='外卖'\middle|\omega_1\right)=1, P\left(x='食堂'\middle|\omega_1\right)=0$$
$$P\left(x='外卖'\middle|\omega_2\right)=0.5, P\left(x='食堂'\middle|\omega_2\right)=0.5$$

求：$P(\omega_2|x='外卖')$，根据贝叶斯公式有：

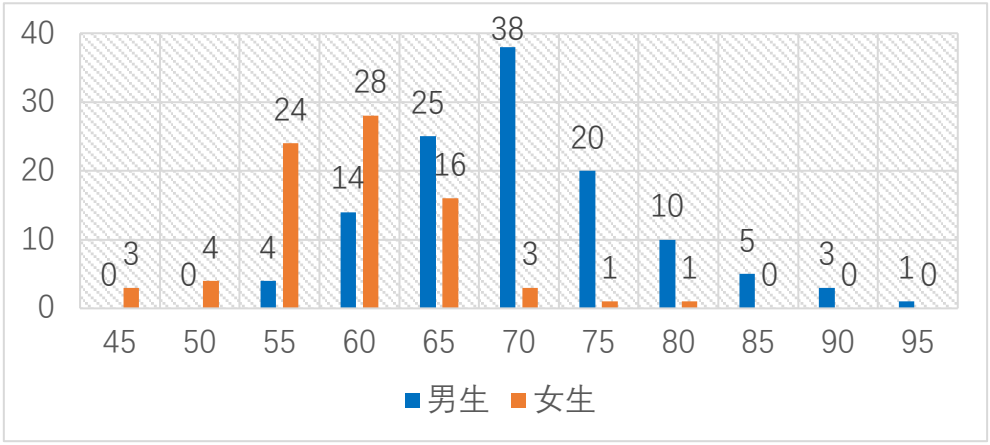
$$P\left(\omega_2\middle|x='外卖'\right)=\frac{P\left(x='外卖'\middle|\omega_2\right)P(\omega_2)}{P\left(x='外卖'\right)}$$

进一步根据全概率公式有：

$$P\left(\omega_2\middle|x='外卖'\right)=\frac{P\left(x='外卖'\middle|\omega_2\right)P(\omega_2)}{P\left(x='外卖'\middle|\omega_1\right)P(\omega_1)+P\left(x='外卖'\middle|\omega_2\right)P(\omega_2)}$$
$$=\frac{0.5\times0.4}{1\times0.6+0.5\times0.4}$$
$$=0.25$$

答：该生为女生的概率为 25%

15. 一个班级中共有学生 200 人，其中男生 120 人，女生 80 人，他们的体重分布如下表所示：



（1）请计算男生的先验概率、女生体重处于 55 公斤区间段的类条件概率、男学生体重为 55 公斤的类条件概率密度函数估计值、任意学生体重处于 55 公斤区间段的概率、某个体重为 55 公斤的学生属于女生类别的后验概率，并写出计算过程。

（2）请根据最大后验概率预测某个体重为 55 公斤的学生的性别，并写出计算过程。

解

(1)

-男生的先验概率：$P(\omega_1)=\frac{120}{200}=\frac{3}{5}$

-女生体重处于 55 公斤区间段的类条件概率：$P(x\in\Omega_{55}|\omega_2)=\frac{24}{80}=\frac{3}{10}$

-男学生体重为 55 公斤的类条件概率密度函数：$p(x=55|\omega_1)=\frac{4}{120}\times\frac{1}{5}=\frac{1}{150}$

-任意学生体重处于 55 公斤区间段的概率：$P(x\in\Omega_{55})=\frac{28}{200}=\frac{7}{50}$

-某个体重为 55 公斤的学生属于女生类别的后验概率：

$$P(\omega_2)=\frac{80}{200}=\frac{2}{5}$$

$$P(\omega_2|55)=\frac{P(55|\omega_2)P(\omega_2)}{P(55)}=\frac{\frac{3}{10}\times\frac{2}{5}}{\frac{7}{50}}=\frac{6}{7}$$

(2)

-某个体重为 55 公斤的学生属于男生类别的后验概率：

$$P(\omega_1|55)=\frac{P(55|\omega_1)P(\omega_1)}{p(55)}=\frac{\frac{4}{120}\times\frac{3}{5}}{\frac{7}{50}}=\frac{1}{7}$$

-由于$P(\omega_2|55)>P(\omega_1|55)$，根据最大后验概率分类准则预测该学生为女生。

16. 数据如上题：设某学生体重为 55 公斤，问题：

（1）将该生预测为男生的错误概率为多少？

（2）将该生预测为女生的错误概率为多少？

（3）根据最小错误概率分类准则，预测结果为什么？

17. 数据如上题：设男生为ω_1类，女生为ω_2类，给出预测学生性别的风险矩阵为：

$$\Lambda=\begin{bmatrix}\lambda_{11} & \lambda_{12}\\ \lambda_{21} & \lambda_{22}\end{bmatrix}=\begin{bmatrix}0 & 2\\ 1 & 0\end{bmatrix}$$

（1）预测该生为男生的风险为多少？

（2）预测该生为女生的风险为多少？

（3）根据最小风险分类准则，预测结果为什么？

18. 已知两个类别ω_1和ω_2，其先验概率相等，两类的类条件概率服从正态分布，有$\mathrm{p}(\boldsymbol{x}|w_1)=\mathcal{N}\left(\mu_1,\sum\right)$和$\mathrm{p}(\boldsymbol{x}|w_2)=\mathcal{N}\left(\mu_2,\sum\right)$，且$\mu_1=\begin{bmatrix}0\\0\end{bmatrix}$，$\mu_2=\begin{bmatrix}3\\3\end{bmatrix}$，$\sum=$

$\begin{bmatrix}1.1 & 0.3\\ 0.3 & 1.9\end{bmatrix}$，试采用贝叶斯决策对样本$\boldsymbol{x}=\begin{bmatrix}1.0\\ 2.2\end{bmatrix}$进行分类。

解：

根据贝叶斯定理

$$P(\omega|\boldsymbol{x})=\frac{p(\boldsymbol{x}|\omega)P(\omega)}{p(\boldsymbol{x})}$$

其中的$p(\boldsymbol{x})$对每个类别相同，因此直接比较$p(\boldsymbol{x}|\omega)P(\omega)$

令判别函数为

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

由于两类的类条件概率服从正态分布且先验概率相等

则

$$g_i(x)=\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_i})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu_i})\right)$$

$$g_1(x)=\exp\left(-\frac{1}{2}\begin{bmatrix}1.0-0\\ 2.2-0\end{bmatrix}^T\begin{bmatrix}1.1 & 0.3\\ 0.3 & 1.9\end{bmatrix}^{-1}\begin{bmatrix}1.0-0\\ 2.2-0\end{bmatrix}\right)=e^{-1.476}=\ 0.159$$

$$g_2(x)=\exp\left(-\frac{1}{2}\begin{bmatrix}1.0-3\\ 2.2-3\end{bmatrix}^T\begin{bmatrix}1.1 & 0.3\\ 0.3 & 1.9\end{bmatrix}^{-1}\begin{bmatrix}1.0-3\\ 2.2-3\end{bmatrix}\right)=e^{-1.836}=\ 0.229$$

由于$g_2(x)>g_1(x)$,因此样本$\boldsymbol{x}$属ω_2类别

19. 两个学生玩猜硬币的游戏；共猜了 30 次，其中正面 17 次，反面 13 次。假设一枚硬币扔出正面的概率为θ，单次扔硬币的结果服从伯努利分布：

$$P(x;\theta)=\theta^x(1-\theta)^{1-x},x=0\ or\ 1$$

其中，$x=1$表示扔出正面，$x=0$表示扔出反面；

（1）根据题意，写出似然性函数的数学表达式。

（2）写出概率模型最大似然估计的最优化问题标准形式。

（3）采用最大似然估计求取参数θ的值?

解：

（1）$\mathcal{L}_{ML}(X;\theta)=P(X;\theta)=\prod_{i=1}^{30}P(x_i;\theta)=\theta^{17}(1-\theta)^{13}$

（2）$\hat{\theta}_{ML}=\underset{\theta}{\operatorname{argmax}}\,\mathcal{L}_{ML}(X;\theta)$

$$\hat{\theta}_{ML}=\underset{\theta}{\operatorname{argmax}}\prod_{i=1}^{30}P(x_i,\theta)=\underset{\theta}{\operatorname{argmax}}\,\theta^{17}(1-\theta)^{13}$$

令目标函数$\mathcal{L}_{ML}(X;\theta)$对$\theta$求导数为 0,得：

$$17\theta^{16}(1-\theta)^{13}-13\theta^{17}(1-\theta)^{12}=0$$

$$\Longrightarrow\hat{\theta}_{ML}=0.567$$

20. 问题描述如上题所示

（1）假设θ的分布服从$\mathcal{N}(0.4,0.2^2)$，写出最大后验概率估计？

（2）根据最大后验概率估计推导出θ的值？

21. 已知 iris 数据中的 setosa 类的 10 个样本的部分特征数据如下，请依据下述数据估计 setosa 类的特征均值向量与协方差矩阵，写出计算过程。

序号	花萼长度/cm	花萼宽度/cm	类别
1	5.1	3.5	setosa
2	4.9	3	setosa
3	4.7	3.2	setosa
4	4.6	3.1	setosa
5	5	3.6	setosa
6	5.4	3.9	setosa
7	4.6	3.4	setosa
8	5	3.4	setosa
9	4.4	2.9	setosa
10	4.9	3.1	setosa

22. 数据如上题所述，分别采用直方图法（bin 边长为 0.4cm），KDE 法（$\delta^2=1$）和 KNN 法（K=3）估计花萼长度=4.8cm，花萼宽度=3.2cm 处的概率密度函数估计值。

解：

(1) 直方图：bin 的边长：$h=0.4$，样本特征维度：$d=2$。设关注点$\boldsymbol{x}=[4.8,3.2]$为中心的 bin 的范围为：

$$B_1(\boldsymbol{x})=[4.6,5.0]$$
$$B_2(\boldsymbol{x})=[3.0,3.4]$$

落入此区域中有 6 个样本

$$\hat{p}(x)\approx\frac{1}{h^d}\frac{k_i}{N}=\frac{1}{0.4^2}\frac{6}{10}=3.75$$

(2) KDE 方法：$(\delta^2=1);d=2$

采用高斯核函数：

$$\mathcal{K}(\boldsymbol{x}_i,\boldsymbol{x})=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_i)^2}{2\sigma^2}}$$

$$\hat{p}(x)=\frac{1}{\sigma^2}\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{K}(\boldsymbol{x}_i,x)\right)=3.675$$

(3) KNN 方法：（k=3）

使用欧式距离做最近邻依据，则距离估计样本$\boldsymbol{x}=[4.8,3.2]$最近的 3 个样本分别为：$(4.7,3.2),(4.6,3.1),(4.9,3.1)$

$$h=\max_{j,l}\{|x_l^j-x_l|\},\ j=1,\dots,K;l=1,\dots,d\Longrightarrow h=0.2$$

$$\hat{p}(\boldsymbol{x})=\frac{1}{h_k^d}\frac{K}{N}=\frac{1}{0.2^2}\frac{3}{10}=7.5$$

23. 数据如上题所述，利用朴素贝叶斯分类器+直方图法估计花萼长度=4.8cm，花萼宽度=3.2cm 的样本的类条件概率密度函数值。

解：$d=2,h=0.4$,类条件概率密度函数值：

$$P(\boldsymbol{x}|\omega)=\prod_{i=1}^{d}P(x_i|\omega)$$

$$p\left(x_i|x_i\in B_{i,j},\omega\right)\approx\frac{N_{i,j}}{N}\frac{1}{A\left(B_{i,j}\right)}$$

$$P(x_1|\omega)=p\left(x_1|x_1\in B_{1,1},\omega\right)=\frac{7}{10}\times\frac{1}{0.4}=\frac{7}{4}$$

$$P(x_2|\omega)=p\left(x_2|x_2\in B_{2,1},\omega\right)=\frac{6}{10}\times\frac{1}{0.4}=\frac{6}{4}$$

$$P(\boldsymbol{x}|\omega)=\prod_{i=1}^{d}P(x_i|\omega)=\frac{7}{4}\times\frac{6}{4}=2.625$$