PART4 贝叶斯决策

基本内容：

1. 掌握连续特征空间、离散特征空间两种情况下，最小错误率的贝叶斯分类决策规则；

2. 掌握连续特征空间、离散特征空间两种情况下，最小风险的贝叶斯分类决策规则；

3. **(1)**一维或多维连续特征空间，正态分布的类条件概率密度函数具体的形式？

一维：

$$p_X(x|\omega_j)=\frac{1}{(2\pi)^{1/2}\sigma_j}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2\right]$$

多维：

$$p_X(\boldsymbol{x}|\omega_j)=\frac{1}{(2\pi)^{d/2}\left|\Sigma_j\right|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T\Sigma_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)\right]$$

(2)多维连续特征空间，若各个特征类条件独立，正态分布的类条件概率密度函数具有什么特点？

$$p_X(\boldsymbol{x}|\omega_j)=\frac{1}{(2\pi)^{d/2}\left|\Sigma_j\right|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T\Sigma_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)\right]$$

$$=\frac{1}{(2\pi)^{d/2}\prod_{i=1}^{d}\sigma_{ji}}\exp\left[-\frac{1}{2}\sum_{i=1}^{d}\left(\frac{x_i-\mu_{ji}}{\sigma_{ji}}\right)^2\right]$$

$$=\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\sigma_{ji}}\exp\left[-\frac{1}{2}\left(\frac{x_i-\mu_{ji}}{\sigma_{ji}}\right)^2\right]=\prod_{i=1}^{d}\boldsymbol{p}_{X_i}(x_i|\omega_j)$$

$\boldsymbol{j}=1,\ldots,C$

4. 已知原始**d**维特征空间的随机向量**x**的期望向量$\boldsymbol{\mu}_X$、协方差矩阵$\boldsymbol{\Sigma}_X$，该随机向量经矩阵**A**线性变换后得**Y**，并且$\boldsymbol{Y}=\boldsymbol{A}^T\boldsymbol{X}$， 则其期望$\boldsymbol{\mu}_Y$、协方差矩阵$\boldsymbol{\Sigma}_Y$=？

$$\boldsymbol{\mu}_Y=\boldsymbol{A}^T\boldsymbol{\mu}_Y,\qquad \boldsymbol{\Sigma}_Y=\boldsymbol{A}^T\boldsymbol{\Sigma}_X\boldsymbol{A}$$

特别地，若$y=\boldsymbol{a}^T\boldsymbol{X}\in R$,则有

$$\mu_y=\boldsymbol{a}^T\boldsymbol{\mu}_Y,\qquad \sigma_y^2=\boldsymbol{a}^T\boldsymbol{\Sigma}_X\boldsymbol{a}$$

5. 一维或多维连续特征空间，各类条件概率密度函数为正态分布时，相应密度函数参数的最大似然估计结果？

6. 给定连续特征空间/离散特征空间，已知类别标记的训练样本集，能够基于该样本集，实现两步法的(朴素)贝叶斯决策。

基于样本的两步贝叶斯决策

[1] **利用有限规模训练样本**$\left\{\left(x_j, y_j\right), j=1,\ldots,N\right\}$,

设计贝叶斯分类器. 估计$\begin{cases} \text{先验概率} & P(\omega_i) \\ \text{条件概率密度} & p(x|\omega_i)\text{或条件概率}P(x|\omega_i) \end{cases}$

$i=1,2,\ldots,c$

[2] $\begin{cases} \text{利用估计的}\hat{P}(\omega_i)\text{、}\hat{p}(x|\omega_i)\text{或}\hat{P}(x|\omega_i) \\ \text{对未知样本}x\text{进行判决} \end{cases}$

练习题

1. 对于<u>二维连续特征空间</u>的两类别分类问题，训练样本集$\boldsymbol{X}=\{(\boldsymbol{x}_i, \mathrm{y}_i), i=1,\ldots,N\}$按照$p(\boldsymbol{x})=\mathrm{P}(\omega_1)\mathrm{p}(\boldsymbol{x}|\omega_1)+\mathrm{P}(\omega_2)\mathrm{p}(\boldsymbol{x}|\omega_2)$独立抽取，其中$\boldsymbol{x}_i=\left[x_i^{(1)}, x_i^{(2)}\right]^{\mathrm{T}} \in R^2$，$\boldsymbol{y}_i \in \{\omega_1, \omega_2\}$. 并且，训练集$\boldsymbol{X}$内：对应第一类$\omega_1$及第二类$\omega_2$的样本集分别为$\boldsymbol{X}_1$及$\boldsymbol{X}_2$ ，两类样本数目分别N_1及N_2.

请按要求完成如下工作：

(1) 若两个类别的类条件概率密度函数$\mathrm{p}(\boldsymbol{x}|\omega_1)$、$\mathrm{p}(\boldsymbol{x}|\omega_2)$均为正态分布，并且各自期望向量分别为$\boldsymbol{\mu}_1$，$\boldsymbol{\mu}_2$，协方差矩阵分别为$\Sigma_1$、$\Sigma_2$，写出$\mathrm{p}(\boldsymbol{x}|\omega_1)$、$\mathrm{p}(\boldsymbol{x}|\omega_2)$的具体表达式。

(2) 若基于样本集$\boldsymbol{X}_1$及$\boldsymbol{X}_2$，采用最大似然法估计上述概率密度函数的参数$\boldsymbol{\mu}_1$，$\boldsymbol{\mu}_2$及Σ_1、Σ_2，请直接写出参数的估计结果。

(3) 对于每个类别，若二维连续特征空间两种特征相互独立，基于样本集$\boldsymbol{X}_1$及$\boldsymbol{X}_2$，请直接写出上述概率密度函数的参数$\boldsymbol{\mu}_1$，$\boldsymbol{\mu}_2$及Σ_1、Σ_2的最大似然估计结果。

(4) 若$\boldsymbol{x}=\left[x^{(1)}, x^{(2)}\right]^{\mathrm{T}} \in R^2$，请基于问题(3)的估计结果，分别写出$\mathrm{p}\left(\boldsymbol{x}^{(2)}|\omega_1\right)$及

$\mathrm{p}\left(\boldsymbol{x}^{(1)}|\omega_2\right)$的具体表达式。

(5)请采用上述给定的样本集$\boldsymbol{X}=\boldsymbol{X}_1\cup\boldsymbol{X}_2$，分别估计$\mathrm{P}(\omega_1)$以及$\mathrm{P}(\omega_2)$。

(6)请基于上述估计结果，采用最小错误率的**朴素贝叶斯分类模型**对特征空间任意观测样本$\boldsymbol{x}=\left[x^{(1)}，x^{(2)}\right]^{\mathrm{T}}\in R^2$分类。

答案：

(1) $\mathrm{p}(\boldsymbol{x}|\omega_{\mathrm{i}})=\frac{1}{2\pi|\Sigma_i|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}{\Sigma_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right]\quad \mathrm{i}=1,2$

(2) 以ω_{i}类的样本集$\boldsymbol{X}_i$估计$\mathrm{p}(\boldsymbol{x}|\omega_{\mathrm{i}})$的参数$\boldsymbol{\mu}_i$以及$\Sigma_i$，$i=1,2$

其最大似然估计结果为：

$$\widehat{\boldsymbol{\mu}}_i=\frac{1}{N_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}\boldsymbol{x}$$

$$\widehat{\Sigma}_i=\frac{1}{N_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}(\boldsymbol{x}-\widehat{\boldsymbol{\mu}}_i)(\boldsymbol{x}-\widehat{\boldsymbol{\mu}}_i)^{\boldsymbol{T}}$$

(3)

$$\widehat{\boldsymbol{\mu}}_i=\left[{\widehat{\boldsymbol{\mu}}_i}^{(1)}\ {\widehat{\boldsymbol{\mu}}_i}^{(2)}\right]^T=\left[\frac{1}{N_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}\boldsymbol{x}^{(1)}\quad \frac{1}{N_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}\boldsymbol{x}^{(2)}\right]^T$$

$$\widehat{\Sigma}_i=\begin{bmatrix}{\widehat{\sigma}_{\mathrm{i}}}^{(1)2} & 0\\ 0 & {\widehat{\sigma}_{\mathrm{i}}}^{(2)2}\end{bmatrix}=\begin{bmatrix}\frac{1}{N_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}\left[\boldsymbol{x}^{(1)}-{\widehat{\boldsymbol{\mu}}_i}^{(1)}\right]^{\boldsymbol{2}} & 0\\ 0 & \frac{1}{N_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}\left[\boldsymbol{x}^{(2)}-{\widehat{\boldsymbol{\mu}}_i}^{(2)}\right]^{\boldsymbol{2}}\end{bmatrix}$$

$i=1,2$

(4)

$$\mathrm{p}\left(\boldsymbol{x}^{(i)}|\omega_1\right)=\frac{1}{\sqrt{2\pi}{\widehat{\sigma}_1}^{(\mathrm{i})}}\exp\left[-\frac{1}{2}\left(\frac{\boldsymbol{x}^{(i)}-{\widehat{\boldsymbol{\mu}}_1}^{(i)}}{{\widehat{\sigma}_1}^{(\mathrm{i})}}\right)^2\right]\quad \mathrm{i}=1,2$$

$$\mathrm{p}\left(\boldsymbol{x}^{(i)}|\omega_2\right)=\frac{1}{\sqrt{2\pi}{\widehat{\sigma}_2}^{(\mathrm{i})}}\exp\left[-\frac{1}{2}\left(\frac{\boldsymbol{x}^{(i)}-{\widehat{\boldsymbol{\mu}}_2}^{(i)}}{{\widehat{\sigma}_2}^{(\mathrm{i})}}\right)^2\right]\quad \mathrm{i}=1,2$$

(5)

$\widehat{P}(\omega_1)=\frac{\boldsymbol{N}_1}{\boldsymbol{N}}$，$\widehat{P}(\omega_2)=\frac{\boldsymbol{N}_2}{\boldsymbol{N}}$

（6）对于任意观测样本$\boldsymbol{x}=\left[x^{(1)}，\ x^{(2)}\right]^{\mathrm{T}}\in R^2$

首先，计算：

$\widehat{\mathrm{P}}(\omega_1)\mathrm{p}\left(\boldsymbol{x}^{(1)}|\omega_1\right)\mathrm{p}\left(\boldsymbol{x}^{(2)}|\omega_1\right)$，　以及$\widehat{\mathrm{P}}(\omega_2)\mathrm{p}\left(\boldsymbol{x}^{(1)}|\omega_2\right)\mathrm{p}\left(\boldsymbol{x}^{(2)}|\omega_2\right)$

若$\widehat{\mathrm{P}}(\omega_1)\mathrm{p}\left(\boldsymbol{x}^{(1)}|\omega_1\right)\mathrm{p}\left(\boldsymbol{x}^{(2)}|\omega_1\right)>\widehat{\mathrm{P}}(\omega_2)\mathrm{p}\left(\boldsymbol{x}^{(1)}|\omega_2\right)\mathrm{p}\left(\boldsymbol{x}^{(2)}|\omega_2\right)$, 则将观测样本$x$判断为$\omega_1$类；

若$\widehat{\mathrm{P}}(\omega_2)\mathrm{p}\left(\boldsymbol{x}^{(1)}|\omega_2\right)\mathrm{p}\left(\boldsymbol{x}^{(2)}|\omega_2\right)>\widehat{\mathrm{P}}(\omega_1)\mathrm{p}\left(\boldsymbol{x}^{(1)}|\omega_1\right)\mathrm{p}\left(\boldsymbol{x}^{(2)}|\omega_1\right)$, 则将观测样本$x$判断为$\omega_2$类；

否则，拒绝决策x的类别，或结合具体问题，进行决策。

2. 对于二维连续特征空间的两类别分类问题，训练样本集$\boldsymbol{X}=\{(\boldsymbol{x}_i,\mathrm{y_i}),i=1,\ldots,N\}$按照$p(\boldsymbol{x})=\mathrm{P}(\omega_1)\mathrm{p}(\boldsymbol{x}|\omega_1)+\mathrm{P}(\omega_2)\mathrm{p}(\boldsymbol{x}|\omega_2)$独立抽取，其中$\boldsymbol{x}_i=\left[x_i^{(1)}，x_i^{(2)}\right]^{\mathrm{T}}\in R^2$，$\boldsymbol{y}_i\in\{\omega_1,\omega_2\}$. 并且，训练集$\boldsymbol{X}$中：来自第一类$\omega_1$及第二类$\omega_2$的样本集分别为$\boldsymbol{X}_1$及$\boldsymbol{X}_2$ ，两类样本数目分别N_1及N_2. 请按照要求完成如下工作：

(1) 基于训练样本集$\boldsymbol{X}$设计一个最小错误率高斯朴素贝叶斯分类模型；

(2) 基于上述模型，写出面向特征空间任意观测样本$\boldsymbol{x}=\left[x^{(1)}，\ x^{(2)}\right]^{\mathrm{T}}\in R^2$分类的决策规则。

3. 针对二维连续特征空间的两类别分类问题，训练样本集$\boldsymbol{X}=\{(\boldsymbol{x}_i,\mathrm{y_i}),i=1,\ldots,N\}$按照$p(\boldsymbol{x})=\mathrm{P}(\omega_1)\mathrm{p}(\boldsymbol{x}|\omega_1)+\mathrm{P}(\omega_2)\mathrm{p}(\boldsymbol{x}|\omega_2)$独立抽取，其中$\boldsymbol{x}_i=\left[x_i^{(1)}，x_i^{(2)}\right]^{\mathrm{T}}\in R^2$，$\boldsymbol{y}_i\in\{\omega_1,\omega_2\}$. 并且，训练集$\boldsymbol{X}$中：来自第一类$\omega_1$及第二类$\omega_2$的样本集分别为$\boldsymbol{X}_1$及$\boldsymbol{X}_2$ ，两类样本数目分别为N_1及N_2. 请完成如下工作：

(1) 若两个类别的先验概率$\mathrm{P}(\omega_1)$、$\mathrm{P}(\omega_2)$以及各类的$\mathrm{p}(\boldsymbol{x}|\omega_1)$、$\mathrm{p}(\boldsymbol{x}|\omega_2)$已知，对

于特征空间任意观测样本$\boldsymbol{x}=\left[x^{(1)}，x^{(2)}\right]^{\mathrm{T}}\in R^2$，如何采用最小错误率的贝叶斯分类规则预测该样本的类别？

(2)若两个类别的类条件概率密度函数$\mathrm{p}(\boldsymbol{x}|\omega_1)$、$\mathrm{p}(\boldsymbol{x}|\omega_2)$均为正态分布，各自期望向量分别为$\boldsymbol{\mu}_1$，$\boldsymbol{\mu}_2$，协方差矩阵分别为$\sum_1$、$\sum_2$，写出$\mathrm{p}(\boldsymbol{x}|\omega_1)$、$\mathrm{p}(\boldsymbol{x}|\omega_2)$的具体函数形式；

(3)如果上述贝叶斯分类模型为**高斯朴素贝叶斯分类模型**，直接写出基于上述训练集、采用最大似然法估计$\boldsymbol{\mu}_1$，$\boldsymbol{\mu}_2$以及$\sum_1$、$\sum_2$的估计结果.

答：

(1)任意观测样本$\boldsymbol{x}=\left[x^{(1)}，x^{(2)}\right]^{\mathrm{T}}\in R^2$

首先，计算：

$P(\omega_1)\mathrm{p}(\boldsymbol{x}|\omega_1)$，　以及$\mathrm{P}(\omega_2)\mathrm{p}(\boldsymbol{x}|\omega_2)$

若$P(\omega_1)\mathrm{p}(\boldsymbol{x}|\omega_1)>\mathrm{P}(\omega_2)\mathrm{p}(\boldsymbol{x}|\omega_2)$, 则将观测样本$x$判断为$\omega_1$类；

若$\mathrm{P}(\omega_2)\mathrm{p}(\boldsymbol{x}|\omega_2)>P(\omega_1)\mathrm{p}(\boldsymbol{x}|\omega_1)$, 则将观测样本$x$判断为$\omega_2$类；

否则，拒绝决策x的类别，或结合具体问题，进行决策。

(2)　$\mathrm{p}(\boldsymbol{x}|\omega_\mathrm{i})=\frac{1}{2\pi|\sum_i|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}{\sum_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right]$　$\mathrm{i}=1,2$

(3)　以ω_i类的样本集$\boldsymbol{X}_i$估计$\mathrm{p}(\boldsymbol{x}|\omega_\mathrm{i})$的参数$\boldsymbol{\mu}_i$以及$\sum_i$，$i=1,2$

其最大似然估计结果为：

$$\widehat{\boldsymbol{\mu}}_i=\frac{1}{N_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}\boldsymbol{x}$$

$$\widehat{\sum}_i=\begin{bmatrix}\widehat{\sigma}_\mathrm{i}^{(1)2} & 0\\ 0 & \widehat{\sigma}_\mathrm{i}^{(2)2}\end{bmatrix}=\begin{bmatrix}\frac{1}{N_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}\left[\boldsymbol{x}^{(1)}-\widehat{\boldsymbol{\mu}}_i^{(1)}\right]^2 & 0\\ 0 & \frac{1}{N_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}\left[\boldsymbol{x}^{(2)}-\widehat{\boldsymbol{\mu}}_i^{(2)}\right]^2\end{bmatrix}$$

$i=1,2$

3. 类似地，将问题 1 及 2 情况扩展至四维连续特征空间的鸢尾花数据集的分类问题。进行求解。

4.下表所示为来自二维离散特征空间关于两种类别的训练样本.其中：样本特征描述为$\boldsymbol{X}=\left[X^{(1)}，X^{(2)}\right]^{\mathrm{T}}$，并且$X^{(1)}\in\{1,2\}$，$X^{(,2)}\in\{S,M,L\}$；类别标记$\boldsymbol{Y}\in$

$\{-\mathbf{1}，\mathbf{1}\}$.

	1	2	3	4	5
特征$\mathbf{X}^{(1)}$	1	1	2	2	1
特征$\mathbf{X}^{(2)}$	S	M	S	M	L
Y	-1	-1	-1	-1	1

请结合该表，采用 LAPLACE 平滑方式，估计朴素贝叶斯分类模型的有关概率信息，并进行给定观测样本的类别决策.

(1) 两个类别的先验概率$P(Y=-1),P(Y=1)$；

(2) $P\left(X^{(1)}=2|Y=-1\right)$, $P\left(X^{(1)}=2|Y=1\right)$

(3) $P\left(X^{(2)}=\mathrm{L}|Y=-1\right)$, $P\left(X^{(2)}=\mathrm{L}|Y=1\right)$

(4) $P\left(X^{(1)}=2，X^{(2)}=\mathrm{L}|Y=-1\right)$

(5) $P\left(X^{(1)}=2，X^{(2)}=\mathrm{L}|Y=1\right)$

(6) 对观测样本$\boldsymbol{X}=[2，L]^{\mathrm{T}}$进行类别决策.

解： (1)先验概率 $\boldsymbol{P}(Y=-1)=\frac{1+4}{2\times1+5}=\frac{5}{7}$ $\quad\boldsymbol{P}(Y=1)=\frac{1+1}{2\times1+5}=\frac{2}{7}$

$$P(Y=-1)=\frac{\lambda+\sum_{i=1}^{N}\delta(Y=-1)}{C\lambda+N}=\frac{1+4}{2+5}$$

$$P(Y=1)=\frac{\lambda+\sum_{i=1}^{N}\delta(Y=1)}{C\lambda+N}=\frac{1+1}{2+5}$$

所以：先验概率 $\boldsymbol{P}(Y=-1)=\frac{1+4}{2\times1+5}=\frac{5}{7}$ $\quad\boldsymbol{P}(Y=1)=\frac{1+1}{2\times1+5}=\frac{2}{7}$

(2)由于$\boldsymbol{P}\left(X^{(1)}=2\,|\,Y=-1\right)=\dfrac{\lambda+\sum_{i=1}^{N}\delta\left(X^{(1)}=2\text{并且}Y=-1\right)}{\lambda S_{(1)}+\sum_{i=1}^{N}\delta(Y=-1)}=\frac{1+2}{2+4}=\frac{1}{2}$

同理：$\boldsymbol{P}\left(X^{(1)}=2\,|\,Y=1\right)=\frac{1+0}{2\times1+1}=\frac{1}{3}$

$(3)\boldsymbol{P}\left(X^{(2)}=L \mid Y=-1\right)=\frac{1+0}{3\times1+4}=\frac{1}{7} \qquad \boldsymbol{P}\left(X^{(2)}=L \mid Y=1\right)=\frac{1+1}{3\times1+1}=\frac{1}{2}$

$(4)\boldsymbol{P}\left(X^{(1)}=2, X^{(2)}=L \mid Y=-1\right)=\boldsymbol{P}\left(X^{(1)}=2 \mid Y=-1\right)\boldsymbol{P}\left(X^{(2)}=L \mid Y=-1\right)=\frac{1}{2}\times\frac{1}{7}=\frac{1}{14}$

$(5)\boldsymbol{P}\left(X^{(1)}=2, X^{(2)}=L \mid Y=1\right)=\boldsymbol{P}\left(X^{(1)}=2 \mid Y=1\right)\boldsymbol{P}\left(X^{(2)}=L \mid Y=1\right)=\frac{1}{3}\times\frac{1}{2}=\frac{1}{6}$

$(6)\boldsymbol{P}(Y=-1)\boldsymbol{P}\left(X^{(1)}=2, X^{(2)}=L \mid Y=-1\right)=\frac{5}{7}\times\frac{1}{14}=\frac{5}{98}$

$\boldsymbol{P}(Y=1)\boldsymbol{P}\left(X^{(1)}=2, X^{(2)}=L \mid Y=1\right)=\frac{2}{7}\times\frac{1}{6}=\frac{2}{42}$

所以$\boldsymbol{P}(Y=1)\boldsymbol{P}\left(X^{(1)}=2, X^{(2)}=L \mid Y=1\right)<\boldsymbol{P}(Y=-1)\boldsymbol{P}\left(X^{(1)}=2, X^{(2)}=L \mid Y=-1\right)$

观测样本$\boldsymbol{X}=\left[X^{(1)}, X^{(2)}\right]^T=[2, L]^T$ 决策结果为$Y=-1$

6.下表所示为来自二维离散特征空间关于两种类别的训练样本. 其中：样本特征描述为$\boldsymbol{X}=\left[X^{(1)}，X^{(2)}\right]^{\mathrm{T}}$，并且$X^{(1)}\in\{1,2,3\}$，$X^{(,2)}\in\{S,M,L\}$；类别标记$\boldsymbol{Y}\in\{\boldsymbol{-1}，\boldsymbol{1}\}$. 待决策样本为$\left(X^{(1)}=1，X^{(2)}=\mathrm{L}\right)$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$X^{(1)}$	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
$X^{(2)}$	S	M	M	S	S	S	M	M	L	L	L	M	M	L	L
Y	−1	−1	1	1	−1	−1	−1	1	1	1	1	1	1	1	−1

请基于该表，采用 LAPLACE 平滑方式，完成朴素贝叶斯分类模型如下内容：

(1)估计两个类别的先验概率$P(Y=-1), P(Y=1)$；

(2)估计$P\left(X^{(1)}=1|Y=-1\right)$, $P\left(X^{(1)}=1|Y=1\right)$

(3)估计$P\left(X^{(2)}=\mathrm{L}|Y=-1\right)$, $P\left(X^{(2)}=\mathrm{L}|Y=1\right)$；

(4)估计$P\left(X^{(1)}=1，X^{(2)}=\mathrm{L}|Y=-1\right)$；

(5)估计$P\left(X^{(1)}=1，X^{(2)}=\mathrm{L}|Y=1\right)$；

(6) 预测样本$\left(X^{(1)}=1，X^{(2)}=\mathrm{L}\right)$的类别.

解： (1)先验概率 $\boldsymbol{P}(Y=-1)=\frac{1+6}{2\times1+15}=\frac{7}{17}$ $\boldsymbol{P}(Y=1)=\frac{1+9}{2\times1+15}=\frac{10}{17}$

(2) $\boldsymbol{P}\left(X^{(1)}=1 \mid Y=-1\right)=\frac{1+3}{3+6}=\frac{4}{9}$ $\boldsymbol{P}\left(X^{(1)}=1 \mid Y=1\right)=\frac{1+2}{3\times1+9}=\frac{1}{4}$

(3) $\boldsymbol{P}\left(X^{(2)}=L \mid Y=-1\right)=\frac{1+1}{3\times1+6}=\frac{2}{9}$ $\boldsymbol{P}\left(X^{(2)}=L \mid Y=1\right)=\frac{1+4}{3\times1+9}=\frac{5}{12}$

(4)朴素贝叶斯分类模型中，各特征满足类条件独立，因此

$$\boldsymbol{P}\left(X^{(1)}=1, X^{(2)}=L \mid Y=-1\right)=\boldsymbol{P}\left(X^{(1)}=1 \mid Y=-1\right)\boldsymbol{P}\left(X^{(2)}=L \mid Y=-1\right)=\frac{4}{9}\times\frac{2}{9}=\frac{8}{81}$$

(5)类似于*(4)*,

$$\boldsymbol{P}\left(X^{(1)}=1, X^{(2)}=L \mid Y=1\right)=\boldsymbol{P}\left(X^{(1)}=1 \mid Y=1\right)\boldsymbol{P}\left(X^{(2)}=L \mid Y=1\right)=\frac{1}{4}\times\frac{5}{12}=\frac{5}{48}$$

(6)由于$\begin{cases}\boldsymbol{P}(Y=-1)\boldsymbol{P}\left(X^{(1)}=1, X^{(2)}=L \mid Y=-1\right)=\frac{7}{17}\times\frac{8}{81}\\ \boldsymbol{P}(Y=1)\boldsymbol{P}\left(X^{(1)}=1, X^{(2)}=L \mid Y=1\right)=\frac{10}{17}\times\frac{5}{48}\end{cases}$

并且 $\boldsymbol{P}(Y=-1)\boldsymbol{P}\left(X^{(1)}=1, X^{(2)}=L \mid Y=-1\right)<\boldsymbol{P}(Y=1)\boldsymbol{P}\left(X^{(1)}=1, X^{(2)}=L \mid Y=1\right)$

所以 观测样本$\boldsymbol{X}=\left[X^{(1)}, X^{(2)}\right]^{T}=[1, L]^{T}$ 决策结果为$Y=-1$

7. 给定三个类别分类的训练集。其中：样本特征描述为$\boldsymbol{X}=\left[X^{(1)}，X^{(2)}\right]^{\mathrm{T}}$，并且$X^{(1)}\in\{1,2,3\}$，$X^{(,2)}\in\{S,M,L,XL\}$；类别标记$\boldsymbol{Y}\in\{\mathbf{1},\mathbf{2},\mathbf{3}\}$.

	1	2	3	4	5	6	7	8	9
特征$\mathbf{X}^{(1)}$	1	1	1	2	1	2	2	1	2
特征$\mathbf{X}^{(2)}$	S	M	XL	XL	S	S	M	L	L
Y	2	2	1	3	2	2	2	1	1

完成如下小题：

(1) 三个类别的先验概率$P(Y=1), P(Y=2)$，$P(Y=3)$；

(2) $P\left(X^{(1)}=2|Y=1\right)$, $P\left(X^{(1)}=2|Y=2\right)$, $P\left(X^{(1)}=2|Y=3\right)$

(3) $P\left(X^{(2)}=\mathrm{L}|Y=1\right)$, $P\left(X^{(2)}=\mathrm{L}|Y=2\right)$, $P\left(X^{(2)}=\mathrm{L}|Y=3\right)$

(4) $P\left(X^{(1)}=2，X^{(2)}=\mathrm{L}|Y=1\right)$

(5) $P\left(X^{(1)}=2，X^{(2)}=\mathrm{L}|Y=2\right)$

(6) $P\left(X^{(1)}=2，X^{(2)}=\mathrm{L}|Y=3\right)$

(7) 对观测样本$\boldsymbol{X}=[2，L]^{\mathrm{T}}$进行类别决策.

8. 下表所示为来自二维离散特征空间关于两种类别的训练样本. 其中：样本特征描述为$\boldsymbol{X}=\left[X^{(1)}，X^{(2)}\right]^{\mathrm{T}}$，并且$X^{(1)}\in\{1,2\}$，$X^{(2)}\in\{S,M,L\}$；类别标记$\boldsymbol{Y}\in\{\mathbf{-1}，\mathbf{1}\}$.

	1	2	3	4	5	6	7	8	9
$X^{(1)}$	1	1	1	1	1	2	2	2	2
$X^{(2)}$	S	M	M	S	S	S	M	M	L
Y	-1	-1	1	1	-1	-1	-1	1	1

请基于该表，采用 LAPLACE 平滑方式，估计朴素贝叶斯分类模型的如下信息：

(1) 两个类别的先验概率$P(Y=-1), P(Y=1)$；

(2)$P\left(X^{(1)}=1|Y=-1\right)$, $P\left(X^{(1)}=1|Y=1\right)$

(3)$P\left(X^{(2)}=\mathrm{L}|Y=-1\right)$, $P\left(X^{(2)}=\mathrm{L}|Y=1\right)$

(4)$P\left(X^{(1)}=1，X^{(2)}=\mathrm{L} \mid Y=-1\right)$

(5)$P\left(X^{(1)}=1，X^{(2)}=\mathrm{L} \mid Y=1\right)$

（6）对观测样本$\boldsymbol{X}=[1，L]^{\mathrm{T}}$进行类别决策