PART4 贝叶斯决策

2020-05-25

河北师范大学 软件学院

要点：

1. 贝叶斯决策规则；
2. 正态分布概率密度函数形式、估计；
3. 朴素贝叶斯分类模型的学习与使用。

基本内容：

1. 掌握连续特征空间、离散特征空间两种情况下，基于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，基于最小风险的贝叶斯分类决策规则；
3. 掌握正态分布概率密度函数：

   (1)一维、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？

   (2)对于多类别分类：若多维连续特征空间，各个特征类条件独立，正态分布的类条件概率密度函数具体形式？具有什么特点？

4. 已知原始**d**维特征空间的随机向量**x**的期望向量、协方差矩阵，该随机向量经线性变换后的期望、协方差矩阵？

若 $\begin{cases} \text{随机向量}\boldsymbol{X}=[X_1,\dots,X_d]^T & \boldsymbol{\mu}_X,\Sigma_X \\ \text{变换矩阵}\boldsymbol{A}=[\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_k] & d\text{行}\times k\text{列}; \\ \text{线性变换} \quad \boldsymbol{Y}=\boldsymbol{A}^T\boldsymbol{X} & \end{cases}$

则 $\begin{cases} \boldsymbol{\mu}_Y=\boldsymbol{A}^T\boldsymbol{\mu}_X \\ \Sigma_Y=\boldsymbol{A}^T\Sigma_X\boldsymbol{A} \end{cases}$

特别地：$\begin{cases} y=\boldsymbol{a}^T\boldsymbol{X} \\ \mu_y=\boldsymbol{a}^T\boldsymbol{\mu}_X \\ \sigma_y^2=\boldsymbol{a}^T\Sigma_X\boldsymbol{a} \end{cases}$

5. **(1)**一元、多元正态分布的概率密度函数的参数的最大似然估计结果？

参数μ、σ^2的最大似然估计

$$\begin{cases} \mu^* = \frac{1}{N}\sum_{i=1}^{N} x_i = \overline{x} \\ \left(\sigma^2\right)^* = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu^*\right)^2 \end{cases}$$

最大似然参数估计

$$\begin{cases} \textbf{期望} \quad \mu^* = \frac{1}{N}\sum_{i=1}^{N} x_i = \overline{x} \\ \textbf{协方差矩阵} \quad \Sigma^* = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu^*\right)\left(x_i - \mu^*\right)^T \end{cases}$$

(2)特征独立时，多元正态分布概率密度函数的参数的最大似然估计结果？

随机变量$X^{(i)}$的边缘密度

$$p_{X^{(i)}}(x^{(i)}; \mu^{(i)}, \sigma^{(i)2}) = \frac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp\left[-\frac{1}{2}\left(\frac{x^{(i)} - \mu^{(i)}}{\sigma^{(i)}}\right)^2\right]$$

随机向量X的概率密度函数：$p_X(x; \mu, \Sigma) = \prod_{i=1}^{d} p_{X^{(i)}}(x^{(i)}; \mu^{(i)}, \sigma^{(i)2})$

边缘密度$p_{X^{(i)}}(x^{(i)}; \mu^{(i)}, \sigma^{(i)2})$最大似然参数估计

$$\begin{cases} \textbf{期望} \quad \left(\mu^{(i)}\right)^* = \frac{1}{N}\sum_{k=1}^{N} x_k^{(i)} \\ \textbf{方差} \quad \left(\sigma^{(i)2}\right)^* = \frac{1}{N}\sum_{k=1}^{N}\left(x_k^{(i)} - \left(\mu^{(i)}\right)^*\right)^2 \end{cases}$$

$$\Rightarrow \begin{cases} \textbf{期望向量} \quad \mu^* = \left[\left(\mu^{(1)}\right)^*, \left(\mu^{(2)}\right)^*, \ldots, \left(\mu^{(d)}\right)^*\right]^T \\ \textbf{协方差矩阵} \quad \Sigma^* = diag\left(\left(\sigma^{(1)2}\right)^*, \left(\sigma^{(2)2}\right)^*, \ldots, \left(\sigma^{(d)2}\right)^*\right) \end{cases}$$

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**$\{(x_j, y_j), j=1,\ldots,N\}$，

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

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

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

练习：

1. 对于二维连续特征空间的两类别分类问题，训练样本集$\boldsymbol{X}=\{(\boldsymbol{x}_i,\mathrm{y_i}),i=1,\dots,N\}$，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。

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

(2) 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。

参考答案： 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：

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

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

(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。

参考答案：

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

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

$i=1,2$

(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$\mathrm{p}\left(x^{(i)}|\omega_1\right)$及$\mathrm{p}\left(x^{(i)}|\omega_2\right)$的具体表达式，$i=1,2$

参考答案：

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

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

(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$\mathrm{P}(\omega_2)$。

参考答案：

$$\hat{P}(\omega_\mathrm{i})=\frac{\boldsymbol{N}_i}{\boldsymbol{N}},\qquad i=1,2$$

(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$\boldsymbol{x}=\left[\boldsymbol{x}^{(1)},\ \boldsymbol{x}^{(2)}\right]^{\mathrm{T}}\in\boldsymbol{R}^2$分类。

参考答案：

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$\boldsymbol{X}=\{(\boldsymbol{x}_i,\mathrm{y_i}),i=1,\dots,N\}$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. 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 $\boldsymbol{P}(Y=-1)=\frac{1+4}{2\times1+5}=\frac{5}{7}$ $\boldsymbol{P}(Y=1)=\frac{1+1}{2\times1+5}=\frac{2}{7}$

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

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

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

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

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

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

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

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

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

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

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

5. 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.

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$X^{(1)}\in\{1,2\}$，$X^{(,2)}\in\{S,M,L\}$；类别标记$\boldsymbol{Y}\in\{\boldsymbol{-1}，\boldsymbol{1}\}$.

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

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

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