模式识别导论第一次作业

杨登天-202028015926089-微电子研究所

1. 请描述最小错误率贝叶斯决策的计算步骤（包含已知条件以及求解任务）；给出一种两类情形下的最小错误率贝叶斯决策规则。

解答：

任务一：在已知条件和求解任务的情况下，最小错误率贝叶斯决策的计算步骤描述。

步骤一、求解观测样本若分类到各种可能类别的后验概率。

步骤二、求解观测样本分类到各种可能类别的错误率，划分到某一类别的错误率和其后验概率之和为 1，所以划分到某一类别的错误率就可以通过相应的后验概率求解。

步骤三、横向对比样本分类到各种可能类别的错误率，并找到最小错误率。

步骤四、根据最小错误率确定样本的最佳类别。

任务二：给出一种两类情形下的最小错误率贝叶斯决策规则。

现假定如下变量，类别$\omega_i(i=1,2)$，特征矢量$\mathrm{X}=[x_1,x_2\dots x_n]$，其中特征矢量是观测样本，问题讨论的就是观测样本的归属。已知条件包括先验概率$P(\omega_i)$，其中$i=1,2$，并且先验概率显然满足$\sum_{i=1}^{2}P(\omega_i)=1$；已知条件还包括概率密度函数$P(X|\omega_i)$，其中$i=1,2$。接下来按照上述步骤依次展开求解步骤。

步骤一、观测样本所属类别仅有两类，所以可以求出样本分别归属到两类的后验概率$P(\omega_i|X)$。

$$P(\omega_i|X)=\frac{P(\omega_i,X)}{P(X)}=\frac{P(X|\omega_i)P(\omega_i)}{P(X)}=\frac{P(X|\omega_i)P(\omega_i)}{\sum_{i=1}^{2}P(X|\omega_i)P(\omega_i)},i\epsilon\{1,2\}$$

由于先验概率$P(\omega_i)$和概率密度函数$P(X|\omega_i)$均已知，所以样本归属到两类的后验概率均可以求解。

步骤二、求解观测样本分类到各种可能类别的错误率$P(error|X)$。

$$P(error|X)=\begin{cases}1-P(\omega_1|X)=P(\omega_2|X) & if\ X\epsilon\omega_1\\ 1-P(\omega_2|X)=P(\omega_1|X) & if\ X\epsilon\omega_2\end{cases}$$

步骤三、横向对比样本分类到各种可能类别的错误率，并找到最小错误率。

$$\min[P(error|X)]=\min\{P(\omega_2|X),P(\omega_1|X)\}=\begin{cases}P(\omega_2|X) & if\ P(\omega_2|X)<P(\omega_1|X)\\ P(\omega_1|X) & if\ P(\omega_2|X)>P(\omega_1|X)\end{cases}$$

步骤四、根据最小错误率确定样本的最佳类别。

$$\omega_{better}=\begin{cases}\omega_1 & if\ \min[P(error|X)]=P(\omega_2|X)\\ \omega_2 & if\ \min[P(error|X)]=P(\omega_1|X)\end{cases}$$

解答完毕

2. 请描述最小风险贝叶斯决策的计算步骤（包含已知条件以及求解任务）；给出一种两类情形下的最小风险贝叶斯决策规则。

解答:

任务一：在已知条件和求解任务的情况下，最小风险贝叶斯决策的计算步骤描述。

步骤一、求解观测样本若分类到各种可能类别的后验概率。

步骤二、通过观测样本分类到各种可能类别的后验概率结合决策计算风险。

步骤三、横向对比各种决策引起的风险，并找到最小风险。

步骤四、根据最小风险确定样本的最佳决策和类别。

任务二：给出一种两类情形下的最小风险贝叶斯决策规则。

现假定如下变量，类别$\omega_i(i=1,2)$，特征矢量$\mathrm{X}=[x_1,x_2\dots x_n]$，其中特征矢量是观测样本，问题讨论的就是观测样本的归属。已知条件包括先验概率$P(\omega_i)$，其中$i=1,2$，并且先验概率显然满足$\sum_{i=1}^{2}P(\omega_i)=1$；已知条件还包括概率密度函数$P(X|\omega_i)$，其中$i=1,2$；已知条件包括决策空间内若干的决策$\alpha_j$，其中$j=1,2\dots a$；已知条件还包括损失函数$\lambda(\alpha_j|\omega_i)$，表示当类别为$\omega_i$时所采取的决策$\alpha_j$所引起的损失，并简要表示为$\lambda_{ji}$。接下来按照上述步骤依次展开求解步骤。

步骤一、求解观测样本若分类到各种可能类别的后验概率。观测样本所属类别仅有两类，所以可以求出样本分别归属到两类的后验概率$P(\omega_i|X)$。

$$P(\omega_i|X)=\frac{P(\omega_i,X)}{P(X)}=\frac{P(X|\omega_i)P(\omega_i)}{P(X)}=\frac{P(X|\omega_i)P(\omega_i)}{\sum_{i=1}^{2}P(X|\omega_i)P(\omega_i)},i\epsilon\{1,2\}$$

由于先验概率$P(\omega_i)$和概率密度函数$P(X|\omega_i)$均已知，所以样本归属到两类的后验概率均可以求解。

步骤二、通过观测样本分类到各种可能类别的后验概率结合决策计算风险$R(\alpha_j|X)$，考虑无拒识的情况（也就是决策空间内仅有两种决策，α_j，其中$j=1,2$）。

$$R(\alpha_j|X)=\begin{cases}\lambda_{11}P(\omega_1|X)+\lambda_{12}P(\omega_2|X) & if\ \alpha_j=\alpha_1\\ \lambda_{21}P(\omega_1|X)+\lambda_{22}P(\omega_2|X) & if\ \alpha_j=\alpha_2\end{cases}$$

步骤三、横向对比各种决策引起的风险，并找到最小风险。

$$\begin{aligned}\min[R(\alpha_j|X)]&=\min\{R(\alpha_1|X),R(\alpha_2|X)\}\\&=\begin{cases}\lambda_{11}P(\omega_1|X)+\lambda_{12}P(\omega_2|X) & if\ R(\alpha_1|X)<R(\alpha_2|X)\\ \lambda_{21}P(\omega_1|X)+\lambda_{22}P(\omega_2|X) & if\ R(\alpha_1|X)>R(\alpha_2|X)\end{cases}\end{aligned}$$

步骤四、根据最小风险确定样本的最佳决策和类别。

$$\omega_{better}=\begin{cases}\omega_1 & if\ \min[R(\alpha_j|X)]=R(\alpha_1|X)\\ \omega_2 & if\ \min[R(\alpha_j|X)]=R(\alpha_2|X)\end{cases}$$

解答完毕

3. 对于 c 类问题，假定各类条件概率密度函数均为多元正态分布。在最小错误率贝叶斯决策的框架下，请写出其判别函数；请分别指出在什么情况下可以获得最小距离分类器，在什么情况下可以得到线性判别函数。

解答：

任务一：c 类问题，假定各类条件概率密度函数均为多元正态分布。在最小错误率贝叶斯决策的框架下，请写出其判别函数。

先进行变量说明，对于c分类($\omega_i(i=1,2\dots c)$)问题，某一变量为$\mathrm{X}=[x_1,x_2\dots x_n]$，且知道各先验概率$P(\omega_i)$，其中$i=1,2\dots\mathrm{c}$，并且先验概率显然满足$\sum_{i=1}^{c}P(\omega_i)=1$。并且各类别条件概率密度函数均为多元正态分布，$P(X|\omega_i)\sim N(\mu_i,\Sigma_i)$。

那么针对最小错误率，将样本归属到某一类别的后验概率可以作为判决函数，

$$P(\omega_i|X)=\frac{P(\omega_i,X)}{P(X)}=\frac{P(X|\omega_i)P(\omega_i)}{P(X)}=\frac{P(X|\omega_i)P(\omega_i)}{\sum_{i=1}^{2}P(X|\omega_i)P(\omega_i)},i\epsilon\{1,2\dots c\}$$

但是考虑到所有类别的后验概率分母相同，因此所有类别的分子可以作为判决函数，考虑到正态分布，因此对其取自然对数得到判别函数$g_i(X)$。

$$g_i(X)=\ln\big(P(X|\omega_i)\big)+\ln\big(P(\omega_i)\big)$$
$$=-\frac{1}{2}(X-\mu_i)^T{\Sigma_i}^{-1}(X-\mu_i)-\frac{d}{2}\ln(2\pi)-\frac{1}{2}\ln(|\Sigma_i|)+\ln\big(P(\omega_i)\big)\qquad(i=1,2\dots c)$$

任务二：何种情况下可以获得最小距离分类器。

当$\Sigma_i=\sigma^2I$，$i=1,2\dots\mathrm{c}$并且先验概率$P(\omega_i)=P(\omega_j)$。此时判别函数$g_i(X)$可以简化为

$$g_i(X)=-\frac{1}{2\sigma^2}(X-\mu_i)^T(X-\mu_i)=-\frac{1}{2\sigma^2}{\|X-\mu_i\|_2}^2$$

所以对于样本的分类只需要计算X到各类的均值向量的欧氏距离平方，并将其归类为最小距离平方的那一类。此即为最小距离分类器。

任务三：何种情况下可以获得线性判别函数。

第一种情况：当$\Sigma_i=\sigma^2I$，$i=1,2\dots\mathrm{c}$并且先验概率$P(\omega_i)\neq P(\omega_j)$。此时判别函数$g_i(X)$可以简化为

$$g_i(X)=-\frac{1}{2\sigma^2}(X-\mu_i)^T(X-\mu_i)+\ln\big(P(\omega_i)\big)=-\frac{1}{2\sigma^2}(X^TX-2{\mu_i}^TX+{\mu_i}^T\mu_i)+\ln\big(P(\omega_i)\big)$$

进一步可以简化为

$$g_i(X)=\left(\frac{1}{\sigma^2}{\mu_i}^T\right)X+\left[-\frac{1}{2\sigma^2}{\mu_i}^T\mu_i+\ln\big(P(\omega_i)\big)\right]={w_i}^TX+\omega_{i0}$$

其中$w_i=\frac{1}{\sigma^2}\mu_i$，$\omega_{i0}=-\frac{1}{2\sigma^2}{\mu_i}^T\mu_i+\ln\big(P(\omega_i)\big)$

上述判别函数即为线性判别函数。

第二种情况：当$\Sigma_i=\Sigma$，$i=1,2\dots\mathrm{c}$并且先验概率$P(\omega_i)\neq P(\omega_j)$。此时判别函数$g_i(X)$可以简化为

$$g_i(X) = -\frac{1}{2}(X-\mu_i)^T\Sigma^{-1}(X-\mu_i) + \ln\big(P(\omega_i)\big)$$

$$= -\frac{1}{2}(\mathrm{X}^T\Sigma^{-1}X - 2\mu_i{}^T\Sigma^{-1}X + \mu_i{}^T\Sigma^{-1}\mu_i) + \ln\big(P(\omega_i)\big)$$

进一步简化得到

$$g_i(X) = (\mu_i{}^T\Sigma^{-1})X + \left[\ln\big(P(\omega_i)\big) - \frac{1}{2}\mu_i{}^T\Sigma^{-1}\mu_i\right] = w_i{}^TX + \omega_{i0}$$

其中$w_i = \Sigma^{-1}\mu_i$，$\omega_{i0} = \ln\big(P(\omega_i)\big) - \frac{1}{2}\mu_i{}^T\Sigma^{-1}\mu_i$

上述判别函数即为线性判别函数。

解答完毕

4. 针对概率密度函数参数估计问题，请描述最大似然估计的计算步骤（包含已知条件以及求解任务）。

解答:

已知条件包括样本集D包含n个独立样本，即$D = \{x_1, x_2 \dots x_n\}$，且样本的类概率密度模型已知，但是相关参数$\theta(\theta = \{\theta_1, \theta_2 \dots \theta_m\})$未知，则类概率密度模型可以表示为$P(D|\theta)$。

求解任务则是通过概率密度函数模型和样本集求解最有可能的参数θ。

步骤一、表示获得n个独立样本的似然函数$l(\theta)$

$$l(\theta) = P(D|\theta) = P(x_1, x_2 \dots x_n|\theta) = \prod_{i=}^{n} P(x_i|\theta)$$

或者也可以对其取自然对数，化简成更为简洁的形式——对数似然函数。

$$H(\theta) = \sum_{i=1}^{n} \ln\,(P(x_i|\theta))$$

步骤二、为了获得最有可能的参数θ。那么需要满足$l(\theta)$或者$H(\theta)$最大的情况时的θ，将此时的θ表示为$\hat{\theta}$

$$\max\big(l(\theta)\big) = l\big(\hat{\theta}\big)\ or\ \max\big(H(\theta)\big) = H\big(\hat{\theta}\big)$$

对似然函数或者对数似然函数求偏导数即可求出有可能的$\hat{\theta}$。

$$\begin{cases} \dfrac{\partial \sum_{i=1}^{n} \ln\,(P(x_i|\theta))}{\partial\theta_1} = 0 \\ \dfrac{\partial \sum_{i=1}^{n} \ln\,(P(x_i|\theta))}{\partial\theta_2} = 0 \\ \dots \\ \dfrac{\partial \sum_{i=1}^{n} \ln\,(P(x_i|\theta))}{\partial\theta_m} = 0 \end{cases} \text{or} \begin{cases} \dfrac{\partial \prod_{i=}^{n} P(x_i|\theta)}{\partial\theta_1} = 0 \\ \dfrac{\partial \prod_{i=}^{n} P(x_i|\theta)}{\partial\theta_2} = 0 \\ \dots \\ \dfrac{\partial \prod_{i=}^{n} P(x_i|\theta)}{\partial\theta_m} = 0 \end{cases}$$

步骤三、严格意义上偏导数为 0，未必就获得最大似然函数，因此需要进一步验证，比如对其取二阶偏导数验证其凹凸性。验证结束后，得到能够使得$l(\theta)$或者$H(\theta)$最大的θ，也就是$\hat{\theta}$即为类概率模型的参数。

解答完毕

5. 针对样本的类条件概率密度函数估计问题，请描述贝叶斯估计的计算步骤（包含已知条件以及求解任务）。

解答:

已知条件包括样本集D包含n个独立样本，即$D=\{x_1,x_2 \dots x_n\}$，且样本的类概率密度模型未知，相关参数$\theta(\theta=\{\theta_1,\theta_2 \dots \theta_m\})$未知，则类概率密度函数模型可以表示为$P(D|\theta)$。

求解任务则是通过样本集使得后验概率密度函数$P(\theta|D)$最大。

步骤一、表达出后验概率密度函数关于先验概率的函数

$$P(\theta|D)=\frac{P(\theta,D)}{P(D)}=\frac{P(D|\theta)P(\theta)}{P(D)}$$

而其中$P(D)$则是由参数空间Θ内所有可能的模型共同决定，因此

$$P(D)=\int P(D|\theta)P(\theta)\,d\theta,\text{积分空间为 }\Theta=Constant$$

并且考虑到样本集D包含的样本都是独立的，所以有

$$P(D|\theta)=P(x_1,x_2 \dots x_n|\theta)=\prod_{i=}^{n}P(x_i|\theta)$$

那么后验概率密度函数可以进一步表达为

$$P(\theta|D)=\frac{\prod_{i=}^{n}P(x_i|\theta)P(\theta)}{\int P(D|\theta)P(\theta)\,d\theta, area\ is\ \Theta}=\frac{\prod_{i=}^{n}P(x_i|\theta)P(\theta)}{Constant}$$

步骤二、通过后验概率密度函数求解θ的平均估计量$\hat{\theta}$（其中$P(\theta)$视具体实例而定，有可能为均匀分布，也有可能是正态分布，甚至别的分布模型；$P(x_i|\theta)$也同样视具体情况而定）

$$\hat{\theta}=\int \theta P(D|\theta)\,d\theta,\text{积分空间为 }\Theta$$

步骤三、通过求解得到的平均估计量$\hat{\theta}$将其代入可能的概率密度模型，表达出概率密度函数$P(x|\theta)$。

步骤四、通过概率密度函数和后验概率密度函数得到后验数据分布$P(x|D)$

$$P(x|D)=\int P(x|\theta)P(\theta|D)\,d\theta,\text{积分空间为 }\Theta$$

解答完毕

6. 请指出最大似然估计和贝叶斯估计的不同之处。

解答:

一、最大似然估计认为数据集是基于某个确定但具备未知参数的概率密度模型独立抽样产生；但是贝叶斯估计认为数据集是基于无穷多个未知的概率密度模型以非均匀或者均匀的方式独立抽样产生(均匀或者非均匀的表达来自于$P(\theta)$)——基于概率密度模型参与数目。

二、最大似然估计对概率密度模型未知参数的估计是基于后验的，也就是通过确定模型模拟数据集的产生，从参数空间中寻找最有可能产生当前数据集的参数；而贝叶斯估计对未知参数的估计是基于先验的，以数据集计算所有可能产生当前数据集的概率密度模型及其贡献的分布模式。——基于参数计算的先验和后验。

三、最大似然估计由于预先假定概率密度模型，而致其概率密度模型在数据集样本增加时显得更加不稳定；但贝叶斯估计在参数估计上随着数据集增加而显得更具鲁棒性。——基于参数随数据样本增加的稳定性。

四、最大似然估计的任务是根据观测数据估计其在参数空间的取值，而贝叶斯估计根据观测数据对参数的分布进行估计。——基于任务不同。