机器学习

Machine learning

Bayesian Learning

练习题答案

授课人：周晓飞
zhouxiaofei@iie.ac.cn

题目 1：

已知 $P(\omega_1)=0.2$， $P(\omega_2)=0.8$，

$P(x=\text{阴天}\mid\omega_1)=0.6$， $P(x=\text{晴天}\mid\omega_1)=0.4$，

$P(x=\text{阴天}\mid\omega_2)=0.1$， $P(x=\text{晴天}\mid\omega_2)=0.9$

已知 $x=\text{阴天}$，求 x 所属类别。

解：利用贝叶斯公式，有：

$$
\begin{aligned}
P(\omega_1 \mid x=\text{阴天}) &= \frac{p(x=\text{阴天} \mid \omega_1)P(\omega_1)}{p(x=\text{阴天})} \\
&= \frac{p(x=\text{阴天} \mid \omega_1)P(\omega_1)}{p(x=\text{阴天} \mid \omega_1)P(\omega_1) + p(x=\text{阴天} \mid \omega_2)P(\omega_2)} \\
&= \frac{0.6\times 0.2}{0.6\times 0.2 + 0.1\times 0.8} = 0.6
\end{aligned}
$$

$$
\begin{aligned}
P(\omega_2 \mid x = \text{阴天}) &= \frac{p(x = \text{阴天} \mid \omega_2)P(\omega_2)}{p(x = \text{阴天})} \\
&= \frac{p(x = \text{阴天} \mid \omega_2)P(\omega_2)}{p(x = \text{阴天} \mid \omega_1)P(\omega_1) + p(x = \text{阴天} \mid \omega_2)P(\omega_2)} \\
&= \frac{0.1 \times 0.8}{0.6 \times 0.2 + 0.1 \times 0.8} = 0.4
\end{aligned}
$$

$$\therefore x \in \omega_1$$

题目 2：一种疾病的判别：正常为 ω_1 ，不正常为 ω_2 ，已知：

$$P(\omega_1)=0.9, P(\omega_2)=0.1$$

现对某人进行检查，结果为 x，假设已知了：

$$P(x\,|\,\omega_1)=0.2, P(x\,|\,\omega_2)=0.4$$

风险代价矩阵为：风险的正负值

$$L=\begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}=\begin{bmatrix} 0 & 6 \\ 1 & 0 \end{bmatrix}$$

（1）用最小错误率贝叶斯决策进行判别。

（2）用最小风险贝叶斯决策进行判别。

解（1）：

$$P(\omega_1 \mid x) \propto P(\omega_1)P(x \mid \omega_1)$$

$$P(\omega_2 \mid x) \propto P(\omega_2)P(x \mid \omega_2)$$

由于

$$\frac{P(\omega_1 \mid x)}{P(\omega_2 \mid x)} = \frac{P(\omega_1)P(x \mid \omega_1)}{P(\omega_2)P(x \mid \omega_2)} = \frac{9}{2} \quad > 1$$

根据贝叶斯最小错误率判决准则，$x \in \omega_1$。

解（2）：

将 x 判为第 j 类的风险为：

$$r_j(x)=\sum_{i=1}^{2}L_{ij}P(x\,|\,\omega_i)P(\omega_i),\ j=1,2$$

$$\begin{aligned}
r_1(x)-r_2(x)&=L_{11}P(x\,|\,\omega_1)P(\omega_1)+L_{21}P(x\,|\,\omega_2)P(\omega_2)\\
&-L_{12}P(x\,|\,\omega_1)P(\omega_1)-L_{22}P(x\,|\,\omega_2)P(\omega_2)\\
&=P(x\,|\,\omega_1)P(\omega_1)\boxed{(L_{11}-L_{12})}+P(x\,|\,\omega_2)P(\omega_2)(L_{21}-L_{22})
\end{aligned}$$

$$\boxed{-(L_{12}-L_{11})}$$

因为

$$\frac{P(x|\omega_2)P(\omega_2)(L_{21}-L_{22})}{P(x|\omega_1)P(\omega_1)(L_{12}-L_{11})}=\frac{1}{27}<1$$

所以

$$r_1(x)<r_2(x)$$

根据贝叶斯最小风险决策可知

$x\in\omega_1$。

题目 3：以下为标注数据以及对应的特征，其中，A，B，C 为两类特征，Y 为类别标签，利用朴素贝叶斯分类器求 A=0，B=1，C=1 时，Y 的分类标签。

A	1	0	0	1	0	1	0	0	1	1	0
B	0	1	1	0	1	0	0	1	0	1	1
C	0	0	1	0	1	1	0	1	0	0	1
Y	1	0	1	1	0	0	1	0	1	1	?

解：

$P(A=0\,|\,Y=0)=\frac{3}{4}$， $P(A=0\,|\,Y=1)=\frac{1}{3}$

$P(B=1\,|\,Y=0)=\frac{3}{4}$， $P(B=1\,|\,Y=1)=\frac{1}{3}$

$P(C=1\,|\,Y=0)=\frac{3}{4}$， $P(C=1\,|\,Y=1)=\frac{1}{6}$

$P(Y=0)=\frac{2}{5}$， $P(Y=1)=\frac{3}{5}$

由贝叶斯公式得

$$P(Y=0 \mid A=0,B=1,C=1)=\frac{P(A=0,B=1,C=1 \mid Y=0)P(Y=0)}{P(A=0,B=1,C=1)}$$

$$=\frac{P(A=0 \mid Y=0)P(B=1 \mid Y=0)P(C=1 \mid Y=0)P(Y=0)}{P(A=0,B=1,C=1)}$$

$$=\frac{\frac{3}{4}\times\frac{3}{4}\times\frac{3}{4}\times\frac{2}{5}}{P(A=0,B=1,C=1)}$$

$$=\frac{\frac{27}{160}}{P(A=0,B=1,C=1)}$$

同理

$$P(Y=1|A=0,B=1,C=1)=\frac{P(A=0,B=1,C=1|Y=1)P(Y=1)}{P(A=0,B=1,C=1)}$$

$$=\frac{P(A=0|Y=1)P(B=1|Y=1)P(C=1|Y=1)P(Y=1)}{P(A=0,B=1,C=1)}$$

$$=\frac{\frac{1}{3}\times\frac{1}{3}\times\frac{1}{6}\times\frac{3}{5}}{P(A=0,B=1,C=1)}$$

$$=\frac{\frac{1}{90}}{P(A=0,B=1,C=1)}$$

$$\because P(Y=0 \mid A=0, B=1, C=1) > P(Y=1 \mid A=0, B=1, C=1)$$

$$\therefore Y=0$$

题目 4：两类三维分类问题中，每一类的特征向量为正态分布，协方差矩阵均为

$$\Sigma = \begin{bmatrix} 0.3 & 0.1 & 0.1 \\ 0.1 & 0.3 & -0.1 \\ 0.1 & -0.1 & 0.3 \end{bmatrix}$$

均值向量分别为$[0,0,0]^T$和$[0.5,0.5,0.5]^T$，两类先验概率相等。

写出相应的类别相似性函数、决策面的方程。

解：

多维正态分布的概率密度函数为：

$$p(x\,|\,\omega_i) = \frac{1}{(2\pi)^{l/2}\,|\,\Sigma_i\,|^{1/2}} \exp(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i))$$

判别函数为：

$$g_i(x) = \ln p(x\,|\,\omega_i) + \ln p(\omega_i)$$

代入得：

$$g_i(x) = -\frac{1}{2}x^T\Sigma^{-1}x + \frac{1}{2}x^T\Sigma^{-1}\mu_i - \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + \frac{1}{2}\mu_i^T\Sigma^{-1}x + c_1$$

其中，

$$c_1 = \ln p(\omega_i) - \frac{l}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma|$$

进一步化简得：

$$g_i(x) = \mu_i^T \Sigma^{-1} x - \frac{1}{2}\mu_i^T \Sigma^{-1} \mu_i + c_2$$

其中，

$$c_2 = c_1 - \frac{1}{2} x^T \Sigma^{-1} x$$

，是一个与类别无关的常量。

最终，类别相似性函数：

$$g_i(x) = w_i^T x + w_{i0}$$

$$w_i = \Sigma^{-1} \mu_i$$

$$w_{i0} = -\frac{1}{2} \mu_i^T \Sigma^{-1} \mu_i$$

决策超平面方程：

$$g_1(x)-g_2(x)=0$$

将题中所给的条件代入：

$$g_1(x)=0$$

$$g_2(x)=w_2^T x+w_{20}\text{，}\quad w_2=[0,2.5,2.5]^T \quad w_{20}=-1.25$$

决策面方程：

$$g_2(x)=0$$

也可以推导决策平面方程为：

$$(\mu_1-\mu_2)^T\Sigma^{-1}(x-\frac{1}{2}(\mu_1+\mu_2))=0$$

然后代入题目的值。

推导过程：

$$\mu_1^T\Sigma^{-1}x-\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1+\mu_2^T\Sigma^{-1}x-\frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2=0$$

$$\Rightarrow(\mu_1-\mu_2)^T\Sigma^{-1}x-\frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1+\mu_2)=0$$

其中，

$$\frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1+\mu_2)=\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1+\frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2$$

$$+\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_2-\frac{1}{2}\mu_2^T\Sigma^{-1}\mu_1$$

$$=\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1+\frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2$$

最终，决策平面方程为：

$$(\mu_1-\mu_2)^T\Sigma^{-1}(x-\frac{1}{2}(\mu_1+\mu_2))=0$$

————————————————————————————————

题目 5：假设一维样本服从 $N(\theta,\sigma^2)$，均值 θ 未知，即

$$p(x\,|\,\theta)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\theta)^2}{2\sigma^2})$$

给定训练样本 D={$x_1, x_2 \ldots x_N$}，用贝叶斯学习的思想估计均值 θ。

解： 假设θ的先验分布服从$N(\theta_0,\sigma_0{}^2)$，则后验分布为

$$p(\theta\,|\,D)=C_1p(D\,|\,\theta)p(\theta)=C_1p(\theta)\prod_{i=1}^{N}p(x_i\,|\,\theta)$$

$$=C_2\exp(-\frac{(\theta-\theta_0)^2}{2\sigma_0{}^2})\prod_{i=1}^{N}\exp(-\frac{(x_i-\theta)^2}{2\sigma^2})$$

$$=C_2\exp\{-\frac{1}{2}[(\sum_{i=1}^{N}\frac{(x_i-\theta)^2}{\sigma^2})+\frac{(\theta-\theta_0)^2}{\sigma_0{}^2}]\}$$

$$=C_3\exp\{-\frac{1}{2}[(\frac{N}{\sigma^2}+\frac{1}{\sigma_0{}^2})\theta^2-2(\frac{\sum_{i=1}^{N}x_i}{\sigma^2}+\frac{\theta_0}{\sigma_0{}^2})\theta]\}$$

假设θ的后验概率密度函数为正态分布$N(\theta_N, {\sigma_N}^2)$，则

$$p(\theta \mid D) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left(-\frac{(\theta - \theta_N)^2}{2{\sigma_N}^2}\right)$$

$$= C \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{{\sigma_N}^2}\right)\theta^2 - 2\left(\frac{\theta_N}{{\sigma_N}^2}\right)\theta\right]\right\}$$

对比上述两式可得：

$$\frac{1}{{\sigma_N}^2} = \frac{N}{\sigma^2} + \frac{1}{{\sigma_0}^2}$$

$$\frac{\theta_N}{{\sigma_N}^2} = \frac{\sum_{i=1}^{N} x_i}{\sigma^2} + \frac{\theta_0}{{\sigma_0}^2} = \frac{N\hat{m}_N}{\sigma^2} + \frac{\theta_0}{{\sigma_0}^2}$$

解方程组得：

$$\sigma_N{}^2 = \frac{\sigma^2 \sigma_0{}^2}{N\sigma_0{}^2 + \sigma^2}$$

$$\theta_N = \frac{N\sigma_0{}^2}{N\sigma_0{}^2 + \sigma^2} \hat{m}_N + \frac{\sigma^2}{N\sigma_0{}^2 + \sigma^2} \theta_0$$

题目 6．基于朴素 Bayes 的文本分类

思路：

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

(1) $p(x_i|c_j)$和 $p(c_j)$均由训练语料中统计；

（2）文档的各个词的分布相互独立。

任务描述：

一个文档 D：$x_1, x_2, \ldots, x_d,$

x_i 是第 i 个位置出现的词，$x_i \in \{t_1, t_2, \ldots t_k, \ldots t_v\}$，$t_k$ 是词典中第 k 个值。

$$c = \underset{c \in C}{\arg\max}\, p(c|D) = \underset{c \in C}{\arg\max}\, p(c)\, p(D|c) = \underset{c \in C}{\arg\max}\, p(c) \prod_{i=1}^{d} p(x_i|c)$$

x_i 是一个多值变量，

$$p(D|c) = \prod_{i=1}^{d} p(x_i|c) = \prod_{k=1}^{V} p(t_k|c)^{Nct_k}$$

（可见，文档服从多项式分布）

朴素贝叶斯分类器:

$$c = \underset{c \in C}{\arg\max}\, p(c) \prod_{k=1}^{V} p(t_k \mid c)^{Nt_k}$$

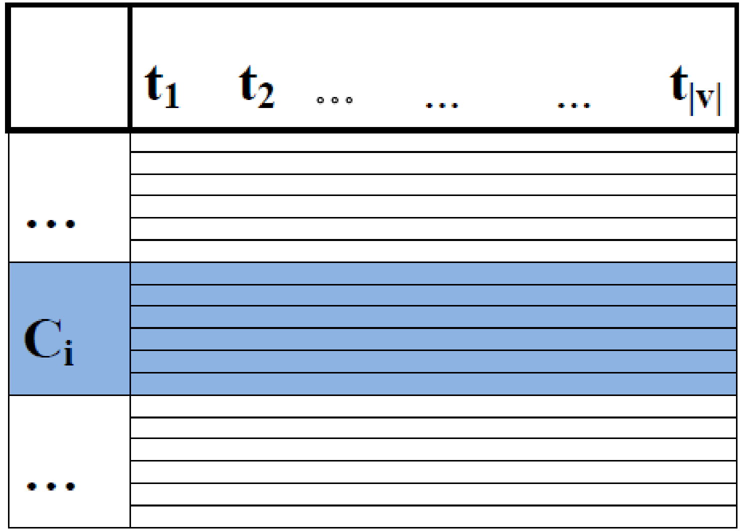
或者

$$c = \underset{c \in C}{\arg\max}\, p(c) \prod_{k=1}^{V} p(t_k \mid c)^{TFt_k}$$

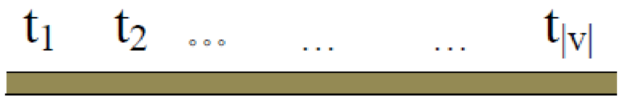
（编程技巧）考虑到概率连乘可能会导致浮点数下界溢出，可将上式取对数：

问题：如何估计 $\hat{p}(c)$ 和 $\hat{p}(t_k \mid c)$？

训练数据 D：　总数|D|



测试文档 Doc：



最大似然估计结论：

$$\hat{p}(c) = \frac{N_c}{N}$$

$$\hat{p}(t_k \mid c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}}$$

$$= \frac{N_{ct}}{\sum_{t'} (N_{ct'})} = \frac{(N_{ct} / N_c)}{\sum_{t'} (N_{ct'} / N_c)} = \frac{T_{ct}}{\sum_{t'} T_{ct'}}$$

其中，N_c 是 c 类文档数目，N 是文档总数目，T_{ct} 是 t 在类别 c 中的词频，V 是词典集合。

拉普拉斯平滑方法：

对于 $\hat{p}(t_k \mid c)$ 可能会出现零概率导致连乘积为零，采取加一平滑：

$$\hat{p}(t_k \mid c) = \frac{T_{ct} + 1}{\sum_{t' \in V}(T_{ct'} + 1)} = \frac{T_{ct} + 1}{\sum_{t' \in V} T_{ct'} + B}$$

其中， $B = |V|$ 为词典大小。

思考： 如果只考虑文档中该词是否出现，如何估计 $\hat{p}(t_k \mid c)$？

解：

估计不考虑词频，只考虑文档中该词是否出现。

$$\hat{p}(t_k \mid c) = \frac{N_{ct}}{N_c}$$

其中， N_{ct} 是类别 c 中出现词 t 的文档数。

对应的加一平滑： $$\hat{p}(t_k \mid c) = \frac{N_{ct} + 1}{N_c + 2}$$

练习题目：分别使用多项式朴素贝叶斯分类器和贝努利朴素贝叶斯分类器对下列测试文本进行分类(对 $\hat{p}(t_k \mid c)$ 采用加一平滑)：

	文档 ID	文档内容	c=China?
训练集	1	Taipei Taiwan	Yes
	2	Macao Taiwan Shanghai	Yes
	3	Japan Sapporo	No
	4	Sapporo Osaka Taiwan	No
测试集	5	Taiwan Taiwan Sapporo	?

解：

（1） 多项式朴素贝叶斯分类器

$$\hat{p}(c) = \hat{p}(\overline{c}) = \frac{1}{2}$$

$$\hat{p}(Taiwan \mid c) = \frac{T_{cTaiwan} + 1}{\sum_{t' \in V} T_{ct'} + B} = \frac{2+1}{5+7} = \frac{1}{4}$$

$$\hat{p}(Taiwan \mid \overline{c}) = \frac{T_{\overline{c}Taiwan} + 1}{\sum_{t' \in V} T_{\overline{c}t'} + B} = \frac{1+1}{5+7} = \frac{1}{6}$$

$$\hat{p}(Sapporo \mid c) = \frac{T_{cSapporo} + 1}{\sum_{t' \in V} T_{ct'} + B} = \frac{0+1}{5+7} = \frac{1}{12}$$

$$\hat{p}(Sapporo \mid \overline{c}) = \frac{T_{\overline{c}Sapporo} + 1}{\sum_{t' \in V} T_{\overline{c}t'} + B} = \frac{2+1}{5+7} = \frac{1}{4}$$

$$\hat{p}(c \mid d) = a \times \frac{1}{2} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{12} = \frac{a}{384}$$

$$\hat{p}(\overline{c} \mid d) = a \times \frac{1}{2} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{4} = \frac{a}{288}$$

$$\therefore d \in \overline{c}$$

（2） 贝努利朴素贝叶斯分类器

$$\hat{p}(c)=\hat{p}(\overline{c})=\frac{1}{2}$$

$$\hat{p}(Taiwan\,|\,c)=\frac{N_{cTaiwan}+1}{N_c+2}=\frac{2+1}{2+2}=\frac{3}{4}$$

$$\hat{p}(Taiwan\,|\,\overline{c})=\frac{N_{\overline{c}Taiwan}+1}{N_c+2}=\frac{1+1}{2+2}=\frac{1}{2}$$

$$\hat{p}(Sapporo\,|\,c)=\frac{N_{cSapporo}+1}{N_c+2}=\frac{0+1}{2+2}=\frac{1}{4}$$

$$\hat{p}(Sapporo \mid \overline{c}) = \frac{N_{\overline{c}Sapporo} + 1}{N_c + 2} = \frac{2+1}{2+2} = \frac{3}{4}$$

$$\hat{p}(c \mid d) = a \times \frac{1}{2} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} = \frac{9a}{128}$$

$$\hat{p}(\overline{c} \mid d) = a \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{3}{4} = \frac{3a}{32}$$

$$\therefore d \in \overline{c}$$