机器学习课程　第7次作业

黄昊　　*20204205*

选择题目7.1　7.4　7.5　7.7　7.8

7.1

- 色泽

$$P(\text{青绿}|\text{好瓜}) = \frac{3}{8}$$
$$P(\text{乌黑}|\text{好瓜}) = \frac{1}{2}$$
$$P(\text{浅白}|\text{好瓜}) = \frac{1}{8}$$
$$P(\text{青绿}|\text{坏瓜}) = \frac{1}{3}$$
$$P(\text{乌黑}|\text{坏瓜}) = \frac{2}{9}$$
$$P(\text{浅白}|\text{坏瓜}) = \frac{4}{9}$$

- 根蒂

$$P(\text{蜷缩}|\text{好瓜}) = \frac{5}{8}$$
$$P(\text{稍蜷}|\text{好瓜}) = \frac{3}{8}$$
$$P(\text{硬挺}|\text{好瓜}) = 0$$
$$P(\text{蜷缩}|\text{坏瓜}) = \frac{1}{3}$$
$$P(\text{稍蜷}|\text{坏瓜}) = \frac{4}{9}$$
$$P(\text{硬挺}|\text{坏瓜}) = \frac{2}{9}$$

- 敲声

$$P(\text{浊响}|\text{好瓜}) = \frac{3}{4}$$
$$P(\text{沉闷}|\text{好瓜}) = \frac{1}{4}$$
$$P(\text{清脆}|\text{好瓜}) = 0$$
$$P(\text{浊响}|\text{坏瓜}) = \frac{4}{9}$$
$$P(\text{沉闷}|\text{坏瓜}) = \frac{1}{3}$$
$$P(\text{清脆}|\text{坏瓜}) = \frac{2}{9}$$

7.4　概率取对数，连乘转化成连加即可。

7.5　对于最小化分类错误率的贝叶斯最优分类器，有：

$$\begin{aligned}
h^*(x) &= arg\max_{c\in y} P(c|x) \\
&= arg\max_{c\in y} P(x|c)P(c) \\
&= arg\max_{c\in y} \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp(-\frac{1}{2}(x-\mu_c)^T\Sigma^{-1}(x-\mu_c))P(c) \\
&= arg\max_{c\in y} \ln(\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp(-\frac{1}{2}(x-\mu_c)^T\Sigma^{-1}(x-\mu_c))P(c)) \\
&= arg\max_{c\in y} -\frac{1}{2}(x-\mu_c)^T\Sigma^{-1}(x-\mu_c)+\ln P(c) \\
&= arg\max_{c\in y} x^T\Sigma^{-1}\mu_c-\frac{1}{2}\mu_c^T\Sigma^{-1}\mu_c+\ln P(c)
\end{aligned}$$

对于二分类问题，贝叶斯决策边界可以表示为：

$$g(x) = x^T\Sigma^{-1}(\mu_1-\mu_0)-\frac{1}{2}(\mu_1+\mu_0)^T\Sigma^{-1}(\mu_1-\mu_0)+\ln(\frac{P(1)}{P(0)})$$

对于线性判别分析，投影界面$w=(\Sigma_0+\Sigma_1)^{-1}(\mu_1-\mu_0)$，两类别在投影面连线中点为：$w=\frac{1}{2}\Sigma^{-1}(\mu_1-\mu_0)w$，则LDA的决策边界为$g(x)=x^T\Sigma^{-1}(\mu_1-\mu_0)-\frac{1}{2}(\mu_1+\mu_0)^T\Sigma^{-1}(\mu_1-\mu_0)$，此时仅相差$\ln\frac{P(1)}{P(0)}$项，由贝叶斯学派的同等无知原则，令$P(1)=P(0)$，即可消去此项，从而得证。

7.7　规定问题为**二分类**问题，样本有d个属性，在第i个属性上的取值种类个数为n_i个，则最好的情况为：每个样本有d个属性，每个属性都满足30个的最低要求，则最少个数为$2\times 30\times \max_{1\le i\le d}\{n_i\}=60\max_{1\le i\le d}\{n_i\}$个。最坏的情况为：从第一个属性开始，选取的样本满足要求后，在第二个属性上的取值完全相同，第三个及之后的属性以此类推。不难发现最坏的样本个数为：

$$\begin{aligned}
&2\times 30\times n_1+2\times 30\times(n_2-1)+2\times 30\times(n_3-1)+\cdots+2\times 30\times(n_d-1) \\
&=60\sum_{i=1}^{d}(n_i-1)+60 \\
&=60\sum_{i=1}^{d}n_i+60(1-d)
\end{aligned}$$

7.8　对于同父结构：

$$\begin{aligned}
P(x_3,x_4) &= \sum_{x_1}P(x_1,x_3,x_4) \\
&= \sum_{x_1}P(x_1)P(x_3|x_1)P(x_4|x_1) \\
&\neq P(x_3)P(x_4)
\end{aligned}$$

因此x_3,x_4关于x_1边际独立不成立。

对于顺序结构：

$$P(x,y,z)=P(z)P(x|z)P(y|x)$$

给定x时：

$$\begin{aligned}
P(y,z|x) &= \frac{P(z)P(x|z)P(y|x)}{P(x)} \\
&= \frac{P(x,z)P(y|x)}{P(x)} \\
&= P(z|x)P(y|x)
\end{aligned}$$

即顺序结构中y,z关于x条件独立。

x取值未知时：

$$\begin{aligned} P(y,z) &= \sum_x P(x,y,z) \\ &= \sum_x P(z)P(x|z)P(y|x) \\ &\neq P(y)P(z) \end{aligned}$$

所以y,z关于x边际独立不成立。