7.5

$R(c|\boldsymbol{x})=1-P(c|\boldsymbol{x})$ [推导]：由式7.1和式7.4可得：
$R(c_i|\boldsymbol{x})=1*P(c_1|\boldsymbol{x})+1*P(c_2|\boldsymbol{x})+\ldots+0*P(c_i|\boldsymbol{x})+\ldots+1*P(c_N|\boldsymbol{x})$ 又$\sum_{j=1}^{N}P(c_j|\boldsymbol{x})=1$，则：
$R(c_i|\boldsymbol{x})=1-P(c_i|\boldsymbol{x})$ 此即为式7.5

7.8

$P(c|\boldsymbol{x})=\frac{P(c)P(\boldsymbol{x}|c)}{P(\boldsymbol{x})}$ [解析]：最小化误差，也就是最大化P(c|x)，但由于P(c|x)属于后验概率无法直接计算，由贝叶斯公式可计算出: $P(c|\boldsymbol{x})=\frac{P(c)P(\boldsymbol{x}|c)}{P(\boldsymbol{x})}$ $P(\boldsymbol{x})$可以省略，因为我们比较的时候$P(\boldsymbol{x})$一定是相同的，所以我们就是用历史数据计算出$P(c)$和$P(\boldsymbol{x}|c)$。

1. $P(c)$根据大数定律，当样本量到了一定程度且服从独立同分布，c的出现的频率就是c的概率。
2. $P(\boldsymbol{x}|c)$，因为$\boldsymbol{x}$在这里不对单一元素是个矩阵，涉及n个元素，不太好直接统计分类为c时，$\boldsymbol{x}$的概率，所以我们根据假设独立同分布，对每个$\boldsymbol{x}$的每个特征分别求概率 $P(\boldsymbol{x}|c)=P(x_1|c)*P(x_2|c)*P(x_3|c)\ldots*P(x_n|c)$ 这个式子就可以很方便的通过历史数据去统计了,比如特征n，就是在分类为c时特征n出现的概率，在数据集中应该是用1显示。 但是当某一概率为0时会导致整个式子概率为0，所以采用拉普拉斯修正

当样本属性独依赖时，也就是除了c多加一个依赖条件，式子变成了 $\prod_{i=1}^{n}P(x_i|c,p_i)$ p_i是x_i所依赖的属性

当样本属性相关性未知时,我们采用贝叶斯网的算法，对相关性进行评估，以找出一个最佳的分类模型。

当遇到不完整的训练样本时，可通过使用EM算法对模型参数进行评估来解决。

7.17-7.18

$P_{(\boldsymbol{x}_i|c)}\in[0,1]$ $p_{(\boldsymbol{x}_i|c)}$ [解析]：式(7.17)所得$P_{(\boldsymbol{x}_i|c)}\in[0,1]$为条件概率，但式(7.18)所得$p_{(\boldsymbol{x}_i|c)}$为条件概率密度而非概率，其值并不在局限于区间[0,1]之内。

7.23

$P(c|\boldsymbol{x})\propto\sum_{\substack{i=1\\|D_{x_i}|\geq m'}}^{d}P(c,x_i)\prod_{j=1}^{d}P(x_j|c,x_i)$ [推导]：

$$\begin{aligned}P(c|\boldsymbol{x})&=\frac{P(\boldsymbol{x},c)}{P(\boldsymbol{x})}\\&=\frac{P\left(x_1,x_2,\ldots,x_d,c\right)}{P(\boldsymbol{x})}\\&=\frac{P\left(x_1,x_2,\ldots,x_d|c\right)P(c)}{P(\boldsymbol{x})}\\&=\frac{P\left(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_d|c,x_i\right)P\left(c,x_i\right)}{P(\boldsymbol{x})}\end{aligned}$$

$$\begin{aligned}P(c|\boldsymbol{x})&\propto P(c,x_i)P(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_d|c,x_i)\\&=P(c,x_i)\prod_{j=1}^{d}P(x_j|c,x_i)\end{aligned}$$

$P(c|\boldsymbol{x})\propto\sum_{\substack{i=1\\|D_{x_i}|\geq m'}}^{d}P(c_i|\boldsymbol{x}_i)\prod_{j=1}^{d}P(c_i|\boldsymbol{x}_i)$ 此即为式7.23，由于式(7.24)和式(7.25)的使用到了$|D_{c,x_i}|$与$|D_{c,x_i,x_j}|$，若$|D_{x_i}|$集合中样本数量过少，则$|D_{c,x_i}|$与$|D_{c,x_i,x_j}|$将会更小，因此在式(7.23)中要求$|D_{x_i}|$集合中样本数量不少于m'。

附录

勘误

7.3节例子计算有笔误，第152页的第9个等式应为$P_{(凹陷|是)}=\frac{5}{8}$。截止到目前（第28次印刷），西瓜书官方勘误修订仅在第8次印刷时修正了第3个等式$P_{(蜷缩|是)}$，但第9个等式仍未修正（若修正第84页的西瓜数据集3.0也可，但亦未修正）。

sklearn调包

```
 import numpy as np
 X = np.array([[-1, -1], [-2, -1], [-3, -2], [1, 1], [2, 1], [3, 2]])
 Y = np.array([1, 1, 1, 2, 2, 2])
from sklearn.naive_bayes import GaussianNB
 clf = GaussianNB()
clf.fit(X, Y)
GaussianNB(priors=None, var_smoothing=1e-09)
print(clf.predict([[-0.8, -1]]))
```

参数:

priors : array-like, shape (n_classes,) Prior probabilities of the classes. If specified the priors are not adjusted according to the data.

var_smoothing : float, optional (default=1e-9) Portion of the largest variance of all features that is added to variances for calculation stability.

贝叶斯应用

1. 中文分词 分词后，得分的假设是基于两词之间是独立的，后词的出现与前词无关
2. 统计机器翻译 统计机器翻译因为其简单，无需手动添加规则，迅速成为了机器翻译的事实标准。
3. 贝叶斯图像识别 首先是视觉系统提取图形的边角特征，然后使用这些特征自底向上地激活高层的抽象概念，然后使用一个自顶向下的验证来比较到底哪个概念最佳地解释了观察到的图像。