第七章 贝叶斯分类器 作业

7.1 解：由表4.3可知：$D_{好瓜}=8$，$D_{坏瓜}=9$.

① 对于色泽：$P(青绿|好瓜)=\frac{D_{好瓜,青绿}}{D_{好瓜}}=\frac{3}{8}$，$P(青绿|坏瓜)=\frac{D_{青绿,坏瓜}}{D_{坏瓜}}=\frac{3}{9}=\frac{1}{3}$

同理：$P(乌黑|好瓜)=\frac{4}{8}=\frac{1}{2}$，$P(乌黑|坏瓜)=\frac{2}{9}$

$P(浅白|好瓜)=\frac{1}{8}$，$P(浅白|坏瓜)=\frac{4}{9}$

② 对于根蒂：$P(蜷缩|好瓜)=\frac{5}{8}$，$P(蜷缩|坏瓜)=\frac{3}{9}=\frac{1}{3}$

$P(稍蜷|好瓜)=\frac{3}{8}$，$P(稍蜷|坏瓜)=\frac{4}{9}$

$P(硬挺|好瓜)=0$，$P(硬挺|坏瓜)=\frac{2}{9}$

③ 对于敲声：$P(浊响|好瓜)=\frac{6}{8}=\frac{3}{4}$，$P(浊响|坏瓜)=\frac{4}{9}$

$P(沉闷|好瓜)=\frac{2}{8}=\frac{1}{4}$，$P(沉闷|坏瓜)=\frac{3}{9}=\frac{1}{3}$

$P(清脆|好瓜)=0$，$P(清脆|坏瓜)=\frac{2}{9}$

7.4 解：为了防止概率连乘 $\prod_{i=1}^{d}P(x_i|C)$ 的结果下溢，可以对 $h_{nb}(\vec{x})$ 的目标函数取对数：即求：$\arg\max_{c\in\mathcal{Y}}\left[\ln P(c)+\sum_{i=1}^{d}\ln P(x_i|c)\right]$，即可。

7.8：证明：在同父结构中，若 x_1 的取值未知，则 $x_3 \perp x_4$ 不成立

证：当 x_1 未知时，$P(x_3,x_4)=\sum_{x_1}P(x_3,x_4,x_1)=\sum_{x_1}P(x_1)\cdot P(x_3|x_1)\cdot P(x_4|x_1)$

$\neq P(x_1)\cdot P(x_2)$

~~故：x_1 在取值未知条件下，$x_3 \perp x_4$ 不成立~~

当 x_1 已知时，$P(x_3,x_4|x_1)=\frac{P(x_1,x_3,x_4)}{P(x_1)}=\frac{P(x_1)\cdot P(x_3|x_1)\cdot P(x_4|x_1)}{P(x_1)}=P(x_3|x_1)\cdot P(x_4|x_1)$

故：同父结构中，x_3，x_4 关于 x_1 条件独立

故 x_1 取值未知条件下，$x_3 \perp x_4$ 不成立

另证：顺序结构中，$y \perp z \mid x$，但 $y \perp z$ 不成立。

在顺序结构中：$p(x,y,z) = p(z)\cdot p(x|z)\cdot p(y|x)$

给定 x 条件下，$p(y,z|x) = \frac{p(z)\cdot p(x|z)\cdot p(y|x)}{p(x)} = \frac{p(x,z)\cdot p(y|x)}{p(x)} = p(z|x)\cdot p(y|x)$

故：y、z 关于 x 条件独立

在 x 未知时，$p(y,z) = \sum_x p(y,z,x) = \sum_x p(z)\cdot p(x|z)\cdot p(y|x) \neq p(y)\cdot p(z)$.

故：$y \perp z$ 不成立

第八章 集成学习作业 2021459 牟正强

8.1：证明：对于 $\sum_{k=0}^{\lfloor \frac{T}{2} \rfloor} \binom{T}{k} (1-\varepsilon)^{k}\cdot \varepsilon^{T-k}$

令 $\frac{T}{2} = (1-\varepsilon-\delta)T$ 得：$\delta = \frac{1}{2}-\varepsilon$

则：$\sum_{k=0}^{\lfloor \frac{T}{2} \rfloor} \binom{T}{k} (1-\varepsilon)^{k}\cdot \varepsilon^{T-k} \le e^{-2\delta^2\cdot T} = e^{-2\cdot(\frac{1}{2}-\varepsilon)^2\cdot T} = e^{-\frac{T}{2}(1-2\varepsilon)^2}$

得证

8.2：证明：对于损失函数 $L(-f(x)H(x))$，其总损失

$LOSS = L(-H(x))\cdot P(f(x)=1|x) + L(H(x))\cdot P(f(x)=-1|x)$，

其中 $H(x) \in \{-1, 1\}$

若使LOSS最小，①若 $p(f(x)=1|x) > p(f(x)=-1|x)$，则

$L(-H(x)) < L(H(x))$

~~由于 L 在 $[-\infty, \delta]$ 上单调递减，则：~~

~~$\forall |H(x)| \le \delta$，有 $-H(x) > H(x)$~~

由于 $H(x)$ 在区间 $[-\infty, \delta]$ $(\delta > 0)$ 单调递减.

则：$H(x) > -H(x)$，则：$H(x) = 1$

②若 $P(f(x)=1|x) < p(f(x)=-1|x)$ 时有 $H(x) = -1$

则：最小化LOSS函数时得到的总体损失函数时已达到贝叶斯最优错误率，

故：L 是0/1损失函数的一致替代函数.

8.4 简述 Gradient boosting 与 Ada Boost 的比较.

(1) 相同之处：① 均属于 Boosting 算法，即以某种方式在每一个基学习器的训练过程中更加关注上一轮训练错误的样本；

② 都可以将弱学习器提升为强学习器.

③ 都是个体学习器之间存在强依赖关系、必须串行生成的序列化方法.

(2) 不同之处：Gradient Boosting：用负梯度来作为上一轮基学习器犯错的衡量指标，依据函数空间梯度下降.

Ada Boost：主要通过增加上一轮中训练错误样本的权重来达到关注预测错误样本的目的.