

## 机器学习作业参考答案

## (第 2--5 章)

## 2.4

答: (a)  $S, 4 \leq x \leq 6, 3 \leq y \leq 5$

(b) 对于  $G, \{3 \leq x \leq 8, 2 \leq y \leq 7, 2 \leq x \leq 8, 2 \leq y \leq 5\}$

(c) 能保证减小变型空间的点,  $(7, 4)$ ,

不能保证减小变型空间的点,  $(10, 1)$

(d) 需要提供的训练样本数目最小是 6, 其中 2 个正例, 4 个反例

证明: 为使候选消除算法完全学习到目标概念, 需要满足  $S=G=H=[a, b], [c, d]$

所以  $G$  必须是一个有限区间, 从而至少需要 4 个反例才能构造一个有限区间的假设

又因为  $S=G$

所以  $S$  中必须有点的横坐标在  $a, b$  上, 纵坐标在  $c, d$  上

从而  $S$  至少要有 2 个正例

所以, 需要提供的训练样本数目最小是 6

## 2.7

答: 对于  $A=\{x_i, i=1, 2, \dots\}$  假设存在与  $A$  一致的最特殊假设  $h_1, a < x < b$ , 存在  $a < x_1, b > x_n$

$a=(a+x_1)/2 < (a+x_1)/2, b=(b+x_n)/2 > (b+x_n)/2$ , 所以存在  $h_2$  与  $A$  一致, 所以不存在一个对任何正例集合都一致的最特殊假设。

改正:

假设空间  $H$  中的每个假设形式为  $a < x < b$ , 其中  $a, b$  为任意实常数,  $x$  代表该实例。

例如  $4.5 \leq x \leq 6.1$  这个假设将 4.5 和 6.1 之间的实例划分为正例, 其他为反例。

对于  $A=\{x_i, i=1, 2, \dots\}$   $h_0$  为  $x_1 \leq x \leq x_n$ , 既可与  $A$  一致, 且是对  $A$  一致最特殊假设。

## 3.4

(a) 解: 要画决策树, 需要计算每个候选属性相对于整个样例集合  $S$  的信息增益, 然后选择信息增益最高的一个属性作为树节点上第一个被测试的属性。

$\text{Gain}(S, \text{Sky}) = 0.8113$

$\text{Gain}(S, \text{AirTemp}) = 0.8113$

$\text{Gain}(S, \text{Humidity}) = 0.1226$

$\text{Gain}(S, \text{Wind}) = 0$

Gain(S, Water)= 0.1226

Gain(S, Forecast)= 0.3113

(b) (1) 学习到的决策树只包含一个与训练样例一致的假设，使用变型空间算法得到的变型空间包含了所有与训练样例一致的假设，但变型空间只含各属性合取式的集合，如果目标函数不在假设空间中，即合取连接词不能表示最小的子式时，变型空间将会是空的。

在本例中，学习到的决策树 “ $Sky = Sunny$ ” 与变型空间中的  $G$  集合中的假设  $\langle Sunny, ?, ?, ?, ?, ? \rangle$  等价，“ $Air-Temp = Warm$ ” 与  $G$  中的  $\langle ?, Warm, ?, ?, ?, ? \rangle$  等价。

学习到的决策树是用变型空间算法得到的变型空间是一种包含关系，前者是后者的子集或者说是后者的一个元素，

(2) 在此例子中决策树等价于变型空间的一个成员，但是一般情况的决策树并不一定等价于变型空间中的一个成员，因为决策树的判别有顺序，而假设空间中的元素的各个性质没有顺序

(c)

Gain(S, Sky)= 0.3219

Gain(S, AirTemp)= 0.3219

Gain(S, Humidity)= 0.0200

Gain(S, Wind)= 0.3219

Gain(S, Water)= 0.1710

Gain(S, Forecast)= 0.0200

显然第一个属性应该选择 Sky AirTemp Wind

若第一个属性为 Sky 则：

Gain(Ssunny, AirTemp)= 0

Gain(Ssunny, Humidity)= 0.3113

Gain(Ssunny, Wind)= 0.8113 (最大)

Gain(Ssunny, Water)= 0.1226

Gain(Ssunny, Forecast)= 0.1226

若第一个属性为 AirTemp 则：

Gain(Swarm, Sky)= 0

Gain(Swarm, Humidity)= 0.3113

Gain(Swarm, Wind)= 0.8113 (最大)

Gain(Swarm, Water)= 0.1226

Gain(Swarm, Forecast)= 0.1226

若第一个属性为 Wind 则：

Gain(Sstrony, Sky)= 0.8113 (最大)

Gain(Sstrony, AirTemp)= 0.8113 (最大)

Gain(Sstrony, Humidity)= 0.1226

$$\text{Gain}(\text{Sstrony, Water}) = 0.1226$$

$$\text{Gain}(\text{Sstrony, Forecast}) = 0.3113$$

$$\text{Entropy}(S) = -(3/5)\log_2(3/5) - (2/5)\log_2(2/5) = 0.9710$$

所有六个属性的信息增益为：

$$\begin{aligned}\text{Gain}(S, \text{Sky}) &= \text{Entropy}(S) - 4/5 * (- (1/4)\log_2(1/4) - (3/4)\log_2(3/4)) - 1/5 * \log_2 1 \\ &= 0.9710 - 0.6490 = 0.3220\end{aligned}$$

$$\begin{aligned}\text{Gain}(S, \text{Air-Temp}) &= \text{Entropy}(S) - 4/5 * (- (1/4)\log_2(1/4) - (3/4)\log_2(3/4)) - 1/5 * \log_2 1 \\ &= 0.9710 - 0.6490 = 0.3220\end{aligned}$$

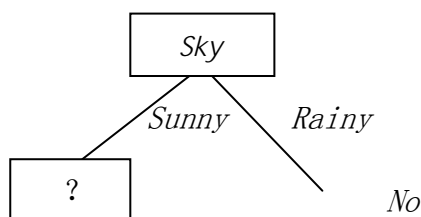
$$\begin{aligned}\text{Gain}(S, \text{Humidity}) &= \text{Entropy}(S) - 2/5 * (- (1/2)\log_2(1/2) * 2) - 3/5 * (- (2/3)\log_2(2/3) - 1/3\log_2(1/3)) \\ &= 0.9710 - 0.9510 = 0.0200\end{aligned}$$

$$\begin{aligned}\text{Gain}(S, \text{Wind}) &= \text{Entropy}(S) - 4/5 * (- (1/4)\log_2(1/4) - 3/4 * \log_2(3/4)) - 1/5 * \log_2 1 \\ &= 0.9710 - 0.6490 = 0.3220\end{aligned}$$

$$\begin{aligned}\text{Gain}(S, \text{Warm}) &= \text{Entropy}(S) - 4/5 * (- (2/4)\log_2(2/4) - 2/4 * \log_2(2/4)) - 1/5 * \log_2 1 \\ &= 0.9710 - 0.8000 = 0.1710\end{aligned}$$

$$\begin{aligned}\text{Gain}(S, \text{Forecast}) &= \text{Entropy}(S) - 2/5 * (- (1/2)\log_2(1/2) * 2) - 3/5 * (- (2/3)\log_2(2/3) - 1/3\log_2(1/3)) \\ &= 0.9710 - 0.9510 = 0.0200\end{aligned}$$

选择 *Sky*, *Air-Temp*, *Wind* 中的任何一个作为根节点的决策属性即可，这里选择 *Sky* 作为根节点的决策属性，建立决策树如下：



计算下一步的信息增益如下：

$$\text{Entropy}(\text{Sunny}) = -(1/4)\log_2(1/4) - (3/4)\log_2(3/4) = 0.8113$$

$$\text{Gain}(\text{Sunny, Air-Temp}) = \text{Entropy}(\text{Sunny}) - (-1/4 * \log_2(1/4) - 3/4 * \log_2(3/4)) = 0$$

$$\begin{aligned}\text{Gain}(\text{Sunny, Humidity}) &= \text{Entropy}(S) - 2/4 * (- (1/2)\log_2(1/2) * 2) - 2/4 * \log_2 1 \\ &= 0.8113 - 0.5000 = 0.3113\end{aligned}$$

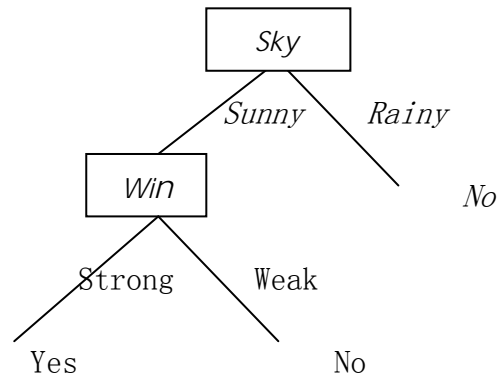
$$\text{Gain}(\text{Sunny, Wind}) = \text{Entropy}(S) - 3/4 * \log_2 1 - 1/4 * \log_2 1 = 0.8113$$

$$\text{Gain}(\text{Sunny, Water}) = \text{Entropy}(S) - 3/4 * (- (1/3)\log_2(1/3) - 2/3 * \log_2(2/3)) - 1/4 * \log_2 1$$

$$=0.8113-0.6887=0.1226$$

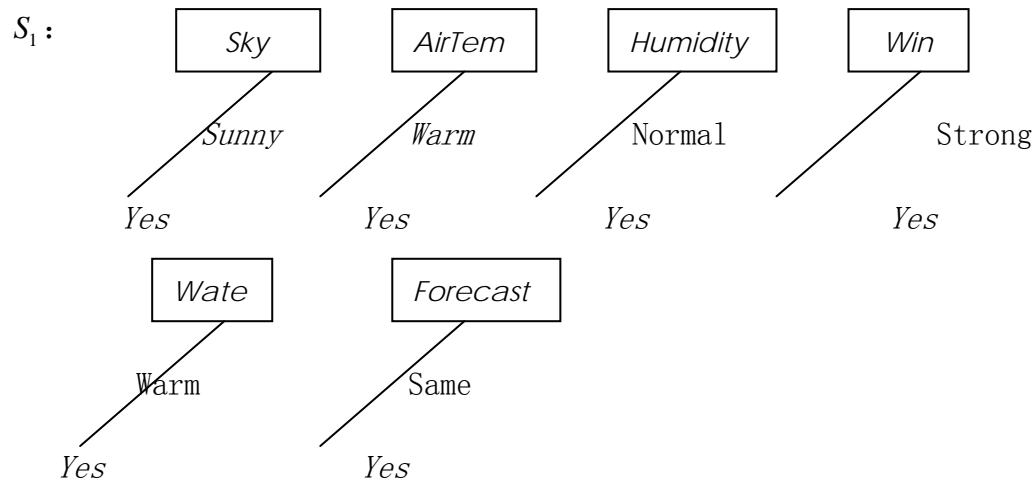
$$\begin{aligned} \text{Gain}(\text{Sunny}, \text{Forecast}) &= \text{Entropy}(S) - 3/4 * (-1/3 * \log_2(1/3) - 2/3 * \log_2(2/3)) - 1/4 * \log_2 1 \\ &=0.8113-0.6887=0.1226 \end{aligned}$$

由于  $\text{Gain}(\text{Sunny}, \text{Wind})$  最大, 选择  $\text{Wind}$  做为新的叶子节点, 建立决策树如下:



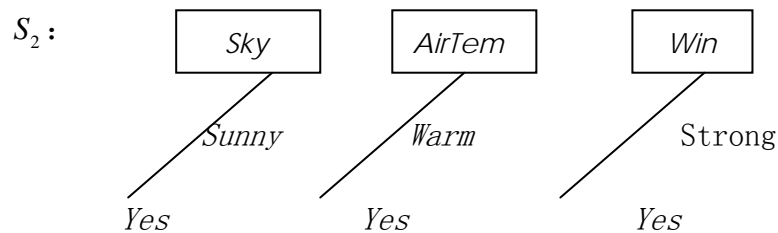
至此, 已用完所有的训练样例, 决策树建立完毕。

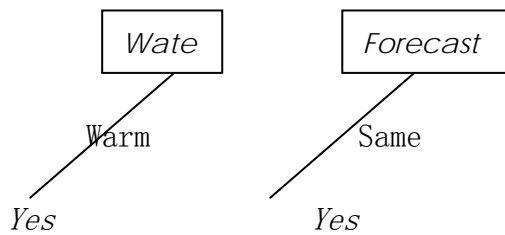
(d) 经过表 2-1 的第一个训练样例后的 S 和 G 如下:



G1: ? 表示对所有例子都接受为正例

当遇到第二个训练样例:  $\langle \text{Sunny}, \text{Warm}, \text{High}, \text{Strong}, \text{Warm}, \text{Same} \rangle$ ,  $\text{EnjoySport} = \text{Yes}$





G2: ? 表示对所有例子都接受为正例

在把候选消除算法应用到决策树假设空间时，预计会遇到如下四种困难：

(1) 在把候选消除算法应用到决策树假设空间时，如果目标函数不在假设空间时，候选消除算法得到的变型空间是空的，或者当遇到含有噪声的数据时，候选消除算法也可能出现空集合，而如果用 ID3 建立决策树则不会出现这种情况。

(2) 如果一个属性的值比较多，一棵决策树将会很宽，

(3) 如何精化  $S$  中的树而不比  $G$  中的树更加一般化是一个困难，反之，如何精化  $G$  中的树而不比  $S$  中的树更加特殊化也是一个困难，另外，要由  $S$  和  $G$  求出中间的合理决策树是十分困难的。其原因都是因为不同形状的决策树可以等价。如果在修改时不进行标准化，那么在构造时就会出现麻烦。

(4) 可能要建立候选决策树的重复信息很多，在选择一棵好的决策树时，计算量会很大，

#### 4.9

不能。存在隐藏单元权值能产生书中建议的隐藏单元编码  $(0.1, 0.2, \dots, 0.8)$ ，例如设定输入端到隐含层的权值分别为  $0.1, 0.2, 0.3, \dots, 0.8$ 。但是不存在输出单元权值能正确解码这样的输入编码。设隐含层到第  $i$  个输出节点的权重为  $\omega_i$ 。以第 1 个输出节点为例，激活函数为 sigmoid 函数。当输入为 00000001 时，隐含层输出为 0.1，要使 output 接近 1， $\omega_1$  必大于零；当输入不是 00000001 时，要使  $\text{output} = \text{sigmoid}(\text{net}_1)$  接近 0， $\text{net}_1$  要小于零，从而  $\omega_1$  必小于零。 $\omega_1$  的取值有矛盾。所以不能通过仅有的一个隐含节点学习恒等函数。

4.10 (证明过程参照教材中相关证明)

$$\text{记 } E_0 = \frac{1}{2} \sum_{d \in D} \sum_{k \in \text{outputs}} (t_{kd} - o_{kd})^2$$

$$\frac{\partial E(\vec{w})}{\partial w_{ji}} = \frac{\partial E_0(\vec{w})}{\partial w_{ji}} + 2\gamma w_{ji} = -\delta_j x_{ji} + 2\gamma w_{ji}$$

$$\text{其中 } \delta_j = \begin{cases} o_j(1-o_j)(t_j-o_j) & j \text{ 为输出单元} \\ o_j(1-o_j) \sum_{k \in \text{outputs}} w_{kj} \delta_k & j \text{ 为隐藏单元} \end{cases}$$

从而我们得到新的权值更新法则:

$$w_{ji} \leftarrow w_{ji} + \eta(\delta_j x_{ji} - 2\gamma w_{ji}) = (1-2\eta\gamma)w_{ji} + \eta\delta_j x_{ji}$$

所以, 新的权值更新法则, 可以通过在标准梯度下降权值更新前, 把每个权值乘以常数  $(1-2\eta\gamma)$

5.4 解: 要保证 95% 双侧置信区间的宽度小于 0.1 则收集的样本数应该满足:

$$2 * 1.96 * \sqrt{\frac{\text{error}_s(h)(1-\text{error}_s(h))}{n}} < 0.1$$

$$\text{即 } n > \left(\frac{1.96}{0.05}\right)^2 [\text{error}_s(h)(1-\text{error}_s(h))]$$

由于假设  $h$  的  $\text{error}_D(h)$  在 0.2 到 0.6 之间。所以, 此时要保证 95% 双侧置信区间的宽度小于 0.1, 则样本数量  $n$  必须满足

$$n \geq \max\left\{\left(\frac{1.96}{0.05}\right)^2 [\text{error}_s(h)(1-\text{error}_s(h))]\right\} = \left(\frac{1.96}{0.05}\right)^2 * \frac{1}{4} = 384.16$$

所以, 要保证 95% 双侧置信区间的宽度小于 0.1, 最少应收集的样例数为 385 个。

5.6 解 从表 5-5 中可以看出  $S_i = D_0 \setminus T_i$ , 并且  $D_0$  是有限数据集, 而 (5.14) 中的  $S$  是从基准分布中抽取的“理想”样本, 我们无法用得到的“理想”样本来计算  $\delta_i$ , 从而无法得到  $\bar{\delta}$ , 所以使用带  $\bar{\delta}$  的置信区间 (5.18) 式, 无法估计 (5.14)。

关于 (5.18) 可以估计 (5.17) 的原因可以从教材中得到 (参考表 5-5)。