

T01 Search and game tree search

17341203 张怡昕

2019 年 12 月 22 日

目录

1	Q1	2
2	Q2	4
3	Q3	6
4	Q4	7
5	Q5	8

1 Q1

执行 ID3 决策树算法

① 题目告诉我们第一步选择 Length 划分:

Length = long 的有 e1, e3, e4, e6, e9, e10, e12, 标签全为 skips, 叶节点.

Length = short 的有 e2, e5, e7, e8, e11, e13 ~ e18, 其标签有 9 个 reads 和 2 个 skips.

$$\text{信息熵 } H_1 = -\left(\frac{9}{11} \log \frac{9}{11} + \frac{2}{11} \log \frac{2}{11}\right) \approx 0.6840$$

下面继续划分 Length = short 的样本.

a) 若划分 Author 属性:

Author = unknown 的有 e2, e7, e8, e11, e18, 其标签 2 个 skips 和 8 个 reads

Author = known 的有 e5, e13, e14, e15, e16, e17, 标签全为 reads

$$\text{信息增益 } IG_{\text{Author}} = H_1 + \left[\frac{5}{11} \left(\frac{2}{5} \log \frac{2}{5} + \frac{3}{5} \log \frac{3}{5} \right) + \frac{6}{11} \times 0 \right] \approx 0.2427$$

b) 若划分 Thread 属性:

Thread = new 的有 e2, e5, e8, e14, e15, e17, e18, 标签全为 reads

Thread = followUp 的有 e7, e11, e13, e16, 标签 2 个 skips 和 2 个 reads

$$\text{信息增益 } IG_{\text{Thread}} = H_1 + \left[\frac{7}{11} \times 0 + \frac{4}{11} \left(\frac{2}{4} \log \frac{2}{4} + \frac{2}{4} \log \frac{2}{4} \right) \right] \approx 0.3204$$

c) 若划分 Where Read 属性:

Where Read = home 的有 e5, e11, e13, e15, e17, 标签 1 个 skips 和 4 个 reads

Where Read = work 的有 e2, e7, e8, e14, e16, e18, 标签 1 个 skips 和 5 个 reads

$$\text{信息增益 } IG_{\text{Where Read}} = H_1 + \left[\frac{5}{11} \left(\frac{1}{5} \log \frac{1}{5} + \frac{4}{5} \log \frac{4}{5} \right) + \frac{6}{11} \left(\frac{1}{6} \log \frac{1}{6} + \frac{5}{6} \log \frac{5}{6} \right) \right] \approx 0.001$$

由于 IG_{Thread} 最大, 故选择 Thread 属性

② 选择 Thread 继续划分:

Thread = new 的有 e2, e5, e8, e14, e15, e17, e18, 标签全为 reads, 叶节点.

Thread = followUp 的有 e7, e11, e13, e16, 标签 2 个 skips 和 2 个 reads.

$$\text{信息熵 } H_2 = -\left(\frac{2}{4} \log \frac{2}{4} + \frac{2}{4} \log \frac{2}{4} \right) = 1$$

下面继续划分 Thread = followUp 的样本

a) 若划分 Author 属性:

Author = unknown 的有 e7, e11, 标签全为 skips

Author = known 的有 e13, e16, 标签全为 reads

$$\text{信息增益 } IG_{\text{Author}} = H_2 + \left[\frac{2}{4} \times 0 + \frac{2}{4} \times 0 \right] = 1$$

b) 若划分 Where Read 属性:

Where Read = home 的有 e11, e13, 标签 1 个 skips 和 1 个 reads

Where Read = work 的有 e7, e16, 标签 1 个 skips 和 1 个 reads

$$\text{信息增益 } IG_{\text{WhereRead}} = H_2 + \left[\frac{2}{4} \left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) + \frac{2}{4} \left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) \right] = 0$$

由于 IG_{Author} 最大, 故选择 Author 属性

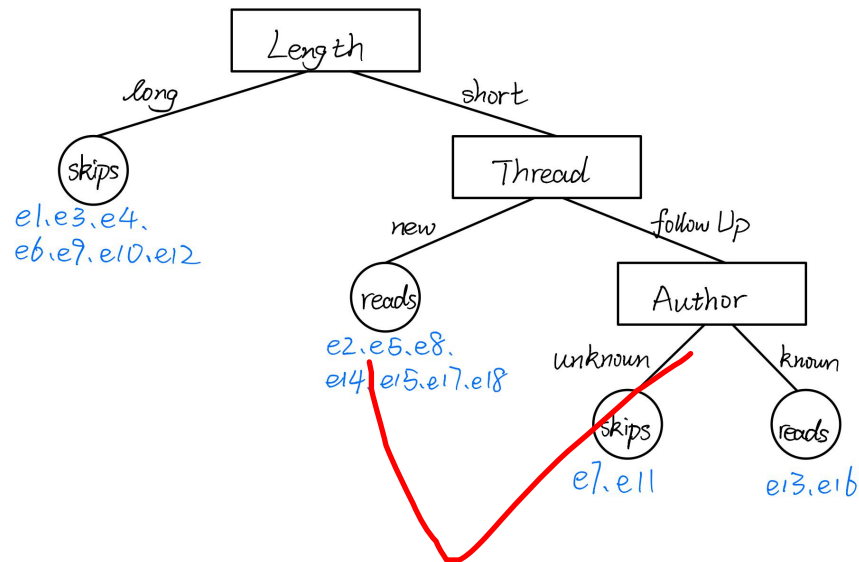
③ 选择 Author 继续划分:

Author = unknown 的有 e7, e11, 标签全为 skips, 叶节点,

Author = known 的有 e13, e16, 标签全为 reads, 叶节点,

无需进一步划分

至此, 决策树构造完成:



2 Q2

$$d = \langle d_1, d_2, d_3, d_4, d_5 \rangle = \langle \text{lime}, \text{cherry}, \text{cherry}, \text{lime}, \text{lime} \rangle$$

① Bayesian 学习

$$\text{似然: } P(d|h_1) = 0$$

$$P(d|h_2) = 0.75^2 \times 0.25^3 = \frac{9}{1024}$$

$$P(d|h_3) = 0.5^2 \times 0.5^3 = \frac{1}{32}$$

$$P(d|h_4) = 0.25^2 \times 0.75^3 = \frac{27}{1024}$$

$$P(d|h_5) = 0$$

$$P(d) = \sum_i P(d|h_i) P(h_i) = 0.2 \times \frac{9}{1024} + 0.4 \times \frac{1}{32} + 0.2 \times \frac{27}{1024} = \frac{5}{256}$$

$$\text{后验概率: } P(h_1|d) = \frac{P(d|h_1) P(h_1)}{P(d)} = 0$$

$$P(h_2|d) = \frac{P(d|h_2) P(h_2)}{P(d)} = 0.09$$

$$P(h_3|d) = \frac{P(d|h_3) P(h_3)}{P(d)} = 0.64$$

$$P(h_4|d) = \frac{P(d|h_4) P(h_4)}{P(d)} = 0.27$$

$$P(h_5|d) = \frac{P(d|h_5) P(h_5)}{P(d)} = 0$$

$$\text{预测: } P(\text{lime}|d) = \sum_i P(\text{lime}|h_i) P(h_i|d)$$

$$= 0 + 0.25 \times 0.09 + 0.5 \times 0.64 + 0.75 \times 0.27 + 0$$

$$= 0.545$$

$$P(\text{cherry}|d) = \sum_i P(\text{cherry}|h_i) P(h_i|d)$$

$$= 0 + 0.75 \times 0.09 + 0.5 \times 0.64 + 0.25 \times 0.27 + 0$$

$$= 0.455$$

$\therefore P(\text{lime}|d) > P(\text{cherry}|d)$, 故预测第6个是 lime.

② MAP 学习

$$h_{\text{MAP}} = \arg \max_{h_i} P(h_i|d) = h_3 \quad (\text{用①中的计算结果})$$

$$P(\text{lime}|h_{\text{MAP}}) = P(\text{cherry}|h_{\text{MAP}}) = 0.5$$

预测为 lime 或 cherry 的概率相同.

③ ML 学习

$$h_{ML} = \arg \max_{h_i} P(d|h_i) = h_3 \quad (\text{用①中的计算结果})$$

$$P(\text{lime} | h_{MAP}) = P(\text{cherry} | h_{MAP}) = 0.5$$

预测为 lime 或 cherry 的概率相同.



3 Q3

首先列出真值表，按要求删去最后一行(A=B=C=D=1)

A	B	C	D	E
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	0
0	1	0	0	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0
1	0	0	0	0
1	0	0	1	1
1	0	1	0	1
1	0	1	1	0
1	1	0	0	0
1	1	0	1	0
1	1	1	0	0

由朴素贝叶斯的假设：

$$\begin{aligned}
 & P(E|A,B,C,D) \\
 &= \frac{P(A,B,C,D|E)P(E)}{P(A,B,C,D)} \\
 &= \frac{1}{P(A,B,C,D)} \times P(E) \times P(A|E)P(B|E)P(C|E)P(D|E)
 \end{aligned}$$

记 $\alpha = \frac{1}{P(A,B,C,D)} > 0$ ，可理解为归一化因子，计算时可忽略。

预测E=1的概率： $P(e|a,b,c,d)$

$$= \alpha P(e)P(a|e)P(b|e)P(c|e)P(d|e)$$

$$= \alpha \frac{4}{15} \times \frac{2}{4} \times \frac{2}{4} \times \frac{2}{4} \times \frac{2}{4} = \frac{1}{60} \alpha \approx 0.017 \alpha$$

预测E=0的概率： $P(\neg e|a,b,c,d)$

$$= \alpha P(\neg e)P(a|\neg e)P(b|\neg e)P(c|\neg e)P(d|\neg e)$$

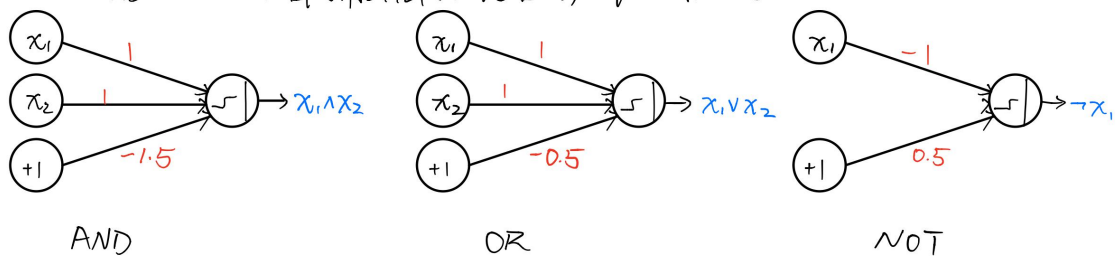
$$= \alpha \frac{11}{15} \times \frac{5}{11} \times \frac{5}{11} \times \frac{5}{11} \times \frac{5}{11} = \frac{125}{3993} \alpha \approx 0.031 \alpha$$

由于 $P(\neg e|a,b,c,d) > P(e|a,b,c,d)$

所以预测A=B=C=D=1时，E=0

4 Q4

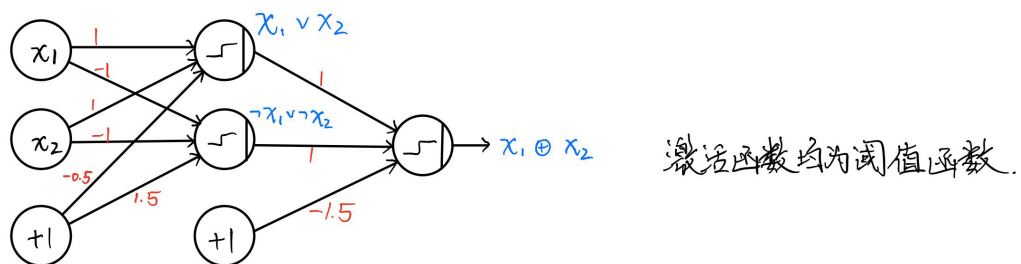
已知 AND、OR、NOT 都可用单层(无隐含层)神经网络表示:



其中输出层激活函数为阈值函数 $g(z) = \begin{cases} 1, & z > 0 \\ 0, & z \leq 0 \end{cases}$ 边上的红色数字为权重.

$$\therefore x_1 \oplus x_2 = (x_1 \vee x_2) \wedge (\neg x_1 \vee \neg x_2)$$

\therefore XOR 可以用 AND、OR、NOT 的组合表示, 即包含一个隐含层的神经网络:



验证:

$$(x_1, x_2) = (0, 0): g(g(0 \times 1 + 0 \times 1 - 0.5) + g(0 \times 1 + 0 \times 1 + 1.5) - 1.5) \\ = g(-1 + 1 - 1.5) = 0 = 0 \oplus 0$$

$$(x_1, x_2) = (0, 1): g(g(0 \times 1 + 1 \times 1 - 0.5) + g(0 \times 1 + 1 \times 1 + 1.5) - 1.5) \\ = g(1 + 1 - 1.5) = 1 = 0 \oplus 1$$

$$(x_1, x_2) = (1, 0): g(g(1 \times 1 + 0 \times 1 - 0.5) + g(1 \times 1 + 0 \times 1 + 1.5) - 1.5) \\ = g(1 + 1 - 1.5) = 1 = 1 \oplus 0$$

$$(x_1, x_2) = (1, 1): g(g(1 \times 1 + 1 \times 1 - 0.5) + g(1 \times 1 + 1 \times 1 + 1.5) - 1.5) \\ = g(1 + 0 - 1.5) = 0 = 1 \oplus 1$$

符合 XOR 的真值表.

5 Q5

(a) 激活函数: $g(x) = \frac{1}{1+e^{-x}}$, $g'(x) = g(x)[1-g(x)]$

$$in_{h_1} = w_1 i_1 + w_2 i_2 + b_1 = 0.15 \times 0.05 + 0.20 \times 0.10 + 0.35 = 0.3775$$

$$out_{h_1} = g(in_{h_1}) \approx 0.59326999$$

$$in_{h_2} = w_3 i_1 + w_4 i_2 + b_1 = 0.25 \times 0.05 + 0.30 \times 0.10 + 0.35 = 0.3725$$

$$out_{h_2} = g(in_{h_2}) \approx 0.59688439$$

$$in_{o_1} = w_5 out_{h_1} + w_6 out_{h_2} + b_2 \approx 1.10590597$$

$$out_{o_1} = g(in_{o_1}) \approx 0.75136507$$

下面直接用链式法则推导, 而不用反向传播算法

$$\begin{aligned} \frac{\partial Loss_{o_1}}{\partial w_1} &= \frac{\partial Loss_{o_1}}{\partial out_{o_1}} \times \frac{\partial out_{o_1}}{\partial in_{o_1}} \times \frac{\partial in_{o_1}}{\partial out_{h_1}} \times \frac{\partial out_{h_1}}{\partial in_{h_1}} \times \frac{\partial in_{h_1}}{\partial w_1} \\ &= \left[\frac{\partial}{\partial out_{o_1}} (y_{o_1} - out_{o_1})^2 \right] \times g'(in_{o_1}) \times \left[\frac{\partial}{\partial out_{h_1}} (w_5 out_{h_1} + w_6 out_{h_2} + b_2) \right] \times g'(in_{h_1}) \times \left[\frac{\partial}{\partial w_1} (w_1 i_1 + w_2 i_2 + b_1) \right] \\ &= -2(y_{o_1} - out_{o_1}) \times g'(in_{o_1}) \times w_5 \times g'(in_{h_1}) \times i_1 \\ &= -2(0.01 - 0.75136507) \times 0.18681560 \times 0.40 \times 0.24130071 \times 0.05 \\ &\approx 0.00133679 \end{aligned}$$

若使用反向传播:

$$\Delta_{o_1} = g'(in_{o_1})(y_{o_1} - out_{o_1}) \approx -0.13849856$$

$$\Delta_{h_1}^{(o_1 \text{ 分量})} = g'(in_{h_1}) w_5 \Delta_{o_1} \approx -0.01336792$$

$$\therefore \frac{\partial Loss_{o_1}}{\partial w_5} \propto -i_1 \Delta_{h_1}^{(o_1 \text{ 分量})} = 0.00066839,$$

该结果为上面结果的 1/2 倍, 因为 BP 公式中在处理 Loss 时略去了系数 2.

(b) ~~激活函数~~ $\tanh(x) = 2g(2x) - 1$, $\tanh'(x) = 1 - \tanh^2(x)$

$$in_{n_1} = 0.3775$$

$$out_{n_1} = \tanh(in_{n_1}) \approx 0.36053439$$

$$in_{n_2} = 0.3925$$

$$out_{n_2} = \tanh(in_{n_2}) \approx 0.37351345$$

$$in_{o_2} = w_7 out_{n_1} + w_8 out_{n_2} + b_2 \approx 0.98569959$$

$$out_{o_2} = \tanh(in_{o_2}) \approx 0.75552264$$

$$\begin{aligned} \frac{\partial Loss_{o_2}}{\partial w_4} &= \frac{\partial Loss_{o_2}}{\partial out_{o_2}} \times \frac{\partial out_{o_2}}{\partial in_{o_2}} \times \frac{\partial in_{o_2}}{\partial out_{n_2}} \times \frac{\partial out_{n_2}}{\partial in_{n_2}} \times \frac{\partial in_{n_2}}{\partial w_4} \\ &= -2(y_{o_2} - out_{o_2}) \times \tanh'(in_{o_2}) \times w_8 \times \tanh'(in_{n_2}) \times iz \\ &= -0.00952469 \end{aligned}$$