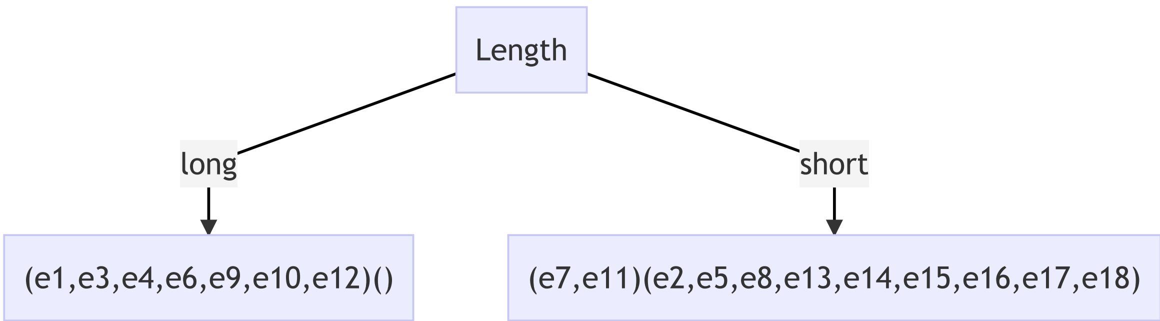
人工智能作业四

张天祎 18340215 计科8班

1

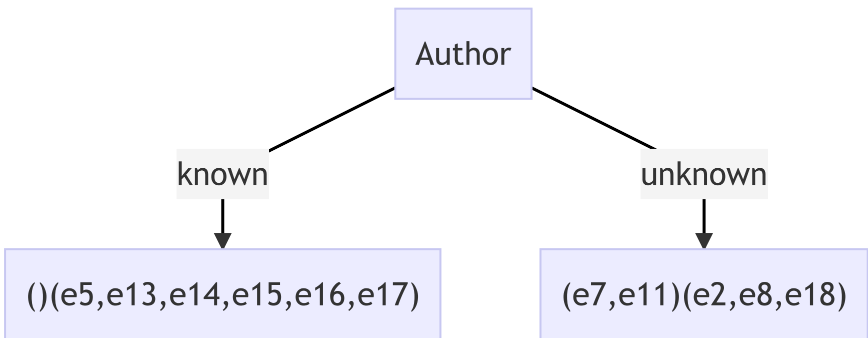
所有数据都是离散数据，使用ID3算法。以下叶结点的前一个集合表示skips标签，后一个集合表示reads标签。

首先依题按照Length决策，如下。

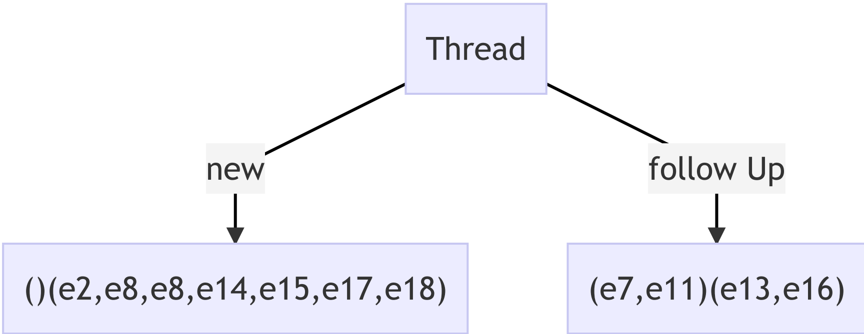


信息熵为$0-\frac{9}{11}\log\frac{9}{11}-\frac{2}{11}\log\frac{2}{11}\approx 0.684$。

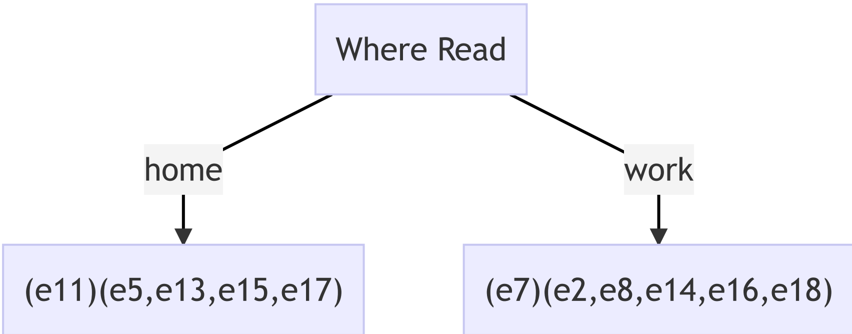
接着对右子树根据三个属性分别计算信息增益。



信息增益为$0.648-(0-\frac{5}{11}\times(\frac{2}{5}\log\frac{2}{5}+\frac{3}{5}\log\frac{3}{5}))\approx 0.243$

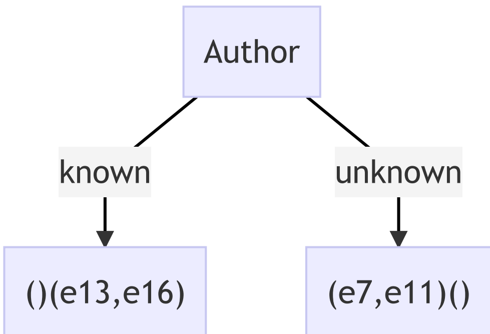


信息增益为$0.648-(0-\frac{4}{11}\times(\frac{2}{4}\log\frac{2}{4}+\frac{2}{4}\log\frac{2}{4}))\approx 0.320$

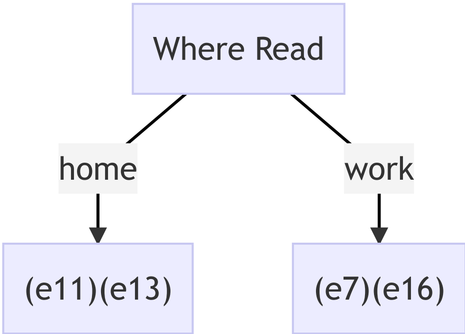


信息增益为$0.648-(-\frac{5}{11}\times(\frac{1}{5}\log\frac{1}{5}+\frac{4}{5}\log\frac{4}{5})-\frac{6}{11}\times(\frac{1}{6}\log\frac{1}{6}+\frac{5}{6}\log\frac{5}{6}))\approx 0.001$

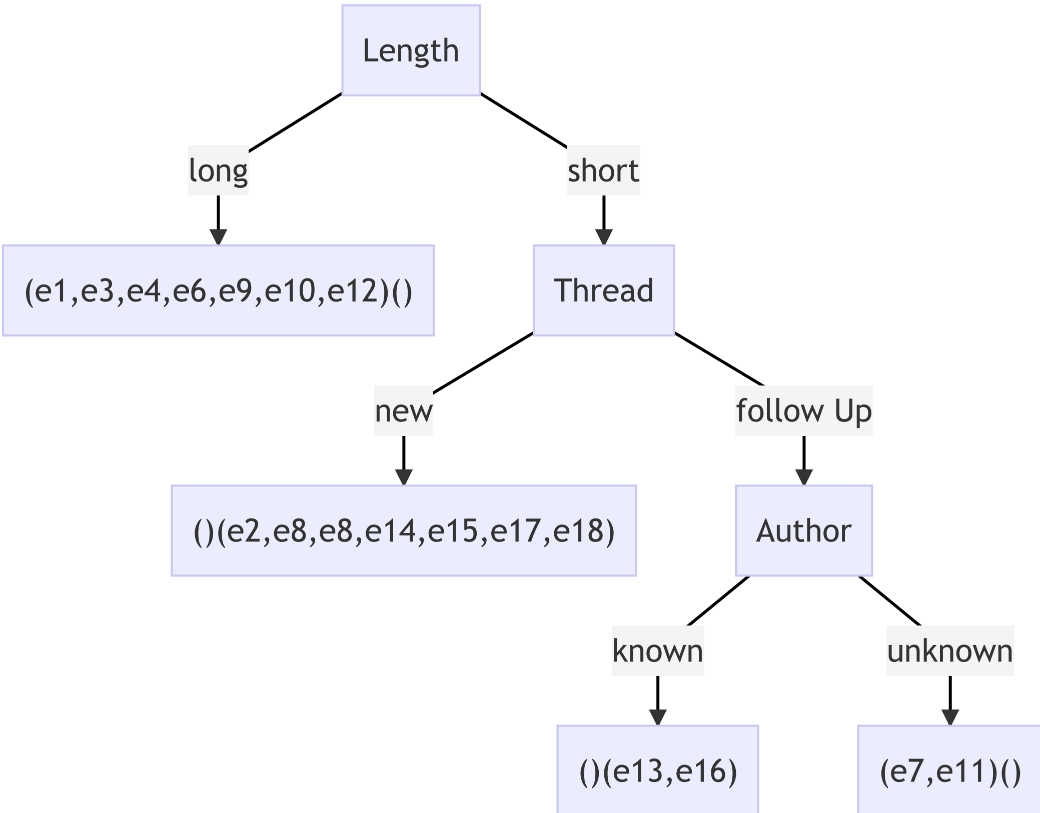
所以选Thread，信息熵为$-\frac{1}{2}\log\frac{1}{2}-\frac{1}{2}\log\frac{1}{2}=1$继续划分。



信息增益为$1-0=1$。



信息增益为$1-(-\frac{1}{2}\times(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2})-\frac{1}{2}\times(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}))=0$。
所以选Author。
整棵决策树如下：



2

(a)

$$
\begin{aligned}
&d=\{lime, cherry, cherry, lime, lime\}\\
&p(d|h_1)=0\\
&p(d|h_2)=0.75^2\times 0.25^3=\frac{9}{1024}\\
&p(d|h_3)=0.5^2\times 0.5^3=\frac{32}{1024}\\
&p(d|h_4)=0.25^2\times 0.75^3=\frac{27}{1024}\\
&p(d|h_5)=0\\
&p(d)=0.2\times\frac{9}{1024}+0.4\times\frac{32}{1024}+0.2\times\frac{27}{1024}=\frac{20}{1024}\\
&\text{由贝叶斯公式} p(h_i|d)=\frac{p(d|h_i)\times p(h_i)}{p(d)}\\
&p(h_1|d)=0\\
&p(h_2|d)=\frac{0.2\times\frac{9}{1024}}{\frac{20}{1024}}=0.09\\
&p(h_3|d)=\frac{0.4\times\frac{32}{1024}}{\frac{20}{1024}}=0.64\\
&p(h_4|d)=\frac{0.2\times\frac{27}{1024}}{\frac{20}{1024}}=0.27\\
&p(h_5|d)=0\\
&p(lime|d)=\sum_{i=1}^{5}p(lime|h_i)p(h_i|d)\\
&=0+0.25\times 0.09+0.5\times 0.64+0.75\times 0.27=0.545>0.5
\end{aligned}
$$

所以选lime作为预测。

(b)

根据（a）中的计算，$p(h_3)p(d|h_3)$为最大，故选择h_3作为假设。故lime和cherry的预测概率时一样的。

(c)

根据（a）中的计算，$p(d|h_3)$最大，仍选择h_3作为假设。故lime和cherry的预测概率时一样的。

3

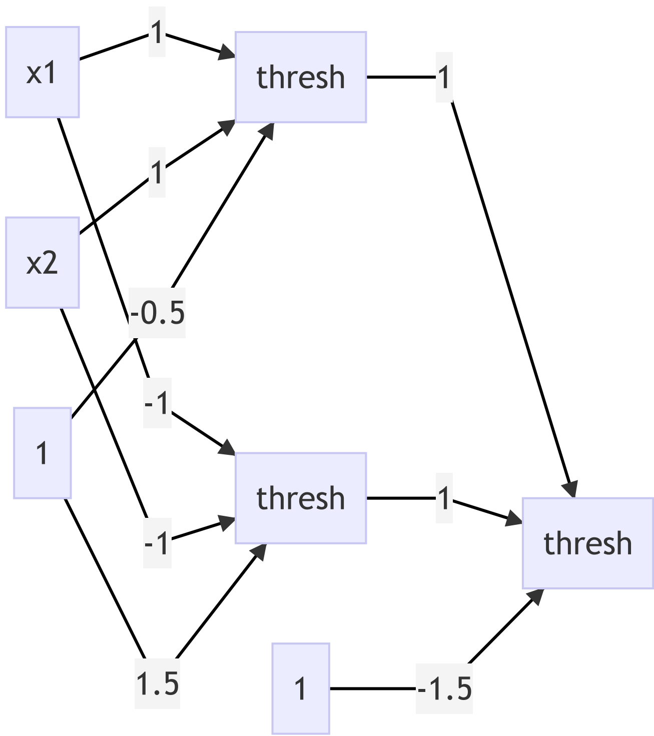
	A	B	C	D	E
1	0	0	0	0	0
2	0	0	0	1	0
3	0	0	1	0	0
4	0	0	1	1	0
5	0	1	0	0	0
6	0	1	0	1	1
7	0	1	1	0	1
8	0	1	1	1	0
9	1	0	0	0	0
10	1	0	0	1	1
11	1	0	1	0	1
12	1	0	1	1	0
13	1	1	0	0	0
14	1	1	0	1	0
15	1	1	1	0	0

$$
\begin{aligned}
&\text{由贝叶斯定理有} p(E|A,B,C,D) \\
&= \frac{p(A,B,C,D|E)p(E)}{p(A,B,C,D)} \\
&= Cp(A,B,C,D|E)p(E) \\
&= Cp(A|E)p(B|E)p(C|E)p(D|E)p(E) \\
&p(e|a,b,c,d) \\
&= Cp(a|e)p(b|e)p(c|e)p(d|e)p(e) \\
&= \frac{2}{4} \times \frac{2}{4} \times \frac{2}{4} \times \frac{2}{4} \times \frac{4}{15}C = \frac{1}{60}C \\
&p(\neg e|a,b,c,d) \\
&= Cp(a|\neg e)p(b|\neg e)p(c|\neg e)p(d|\neg e)p(\neg e) \\
&= \frac{5}{11} \times \frac{5}{11} \times \frac{5}{11} \times \frac{5}{11} \times \frac{11}{15}C = \frac{125}{3993}C > \frac{1}{60}C
\end{aligned}
$$

所以预测E=0。

4

由于$x_1 \oplus x_2 = (x_1 \vee x_2) \wedge (\neg(x_1 \wedge x_2))$，设计神经网络如下。



$$\text{激活函数为} g(x) = thresh(x) = \begin{cases} 1 & x > 0 \\ 0 & x \le 0 \end{cases}$$
$$h_1 = g(x_1 + x_2 - 0.5)$$
$$h_2 = g(-x_1 - x_2 + 1.5)$$
$$out = g(h_1 + h_2 - 1.5)$$

	(x1,x2)	**(h1,h2)**	**out**
1	(0,0)	(0,1)	0
2	(0,1)	(1,1)	1
3	(1,0)	(1,1)	1
4	(1,1)	(1,0)	0

计算过程如上表，符合异或的真值表。

5

(a)

$$
\begin{aligned}
&\frac{\partial Loss_{o1}}{\partial w_1} \\
&= \frac{\partial(sigmod(w_5 h_{1out} + w_6 h_{2out} + b_2) - y_1)^2}{\partial w_1} \\
&= \frac{\partial(out_{o1} - y_1)^2}{\partial out_{o1}} \times \frac{\partial sigmod(w_5 out_{h1} + w_6 out_{h2} + b_2)}{\partial(w_5 out_{h1} + w_6 out_{h2} + b_2)} \\
&\times \frac{\partial(w_5 out_{h1} + w_6 out_{h2} + b_2)}{\partial out_{h1}} \times \frac{\partial sigmod(w_1 i_1 + w_2 i_2 + b_1)}{\partial w_1} \\
&= 2(out_{o1} - y_1) \times (1 - out_{o1}) out_{o1} \times w_5 \times (1 - out_{h1}) out_{h1} i_1 \\
&= 2 \times (0.751 - 0.01)(1 - 0.751) \times 0.751 \times 0.4 \times (1 - 0.593) \times 0.593 \times 0.05 \\
&\approx 0.001337
\end{aligned}
$$

(b)

$$
\begin{aligned}
&\frac{\partial Loss_{o2}}{\partial w_4} \\
&= \frac{\partial(tanh(w_7 h_{1out} + w_8 h_{2out} + b_2) - y_2)^2}{\partial w_4} \\
&= \frac{\partial(out_{o2} - y_2)^2}{\partial out_{o2}} \times \frac{\partial tanh(w_7 out_{h1} + w_8 out_{h2} + b_2)}{\partial(w_7 out_{h1} + w_8 out_{h2} + b_2)} \\
&\times \frac{\partial(w_7 out_{h1} + w_8 out_{h2} + b_2)}{\partial out_{h2}} \times \frac{\partial tanh(w_3 i_1 + w_4 i_2 + b_1)}{\partial w_4} \\
&= 2(out_{o2} - y_2) \times (1 - out_{o2}^2) \times w_8 \times (1 - out_{h2}^2) i_2 \\
&= 2 \times (0.755 - 0.99) \times (1 - 0.755^2) \times 0.55 \times (1 - 0.373^2) \times 0.1 \\
&\approx -0.000952
\end{aligned}
$$