

习题1

属性：性别

$$Ent(D) = -(\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}) = 0.971$$

$$Ent(D^1) = -(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}) = 0.918$$

$$Ent(D^2) = -(1\log_2 1 + 0) = 0$$

$$\begin{aligned} Gain(D, \text{性别}) &= Ent(D) - \sum_{v=1}^2 \frac{|D^v|}{|D|} Ent(D^v) \\ &= 0.971 - (\frac{3}{5} \times 0.918 + \frac{2}{5} \times 0) \\ &= 0.382 \end{aligned}$$

属性：喜欢ML作业

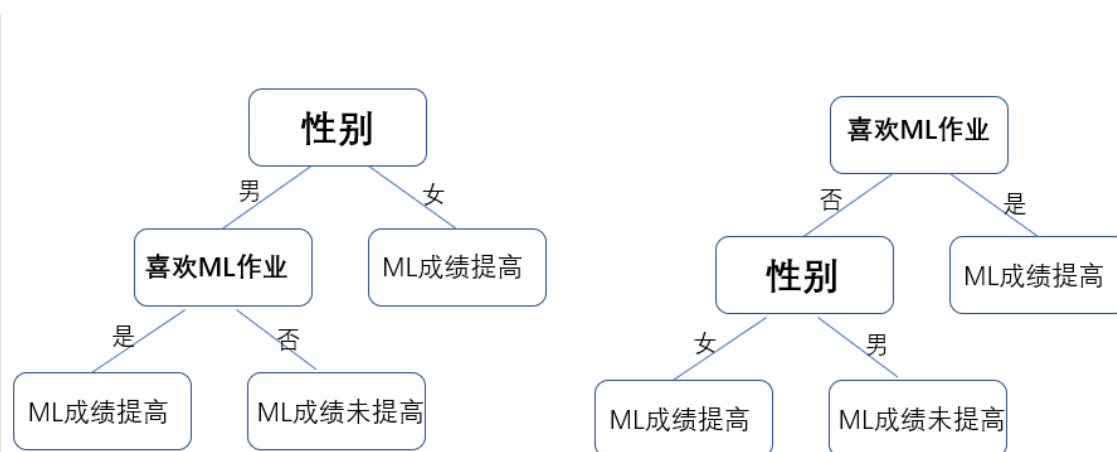
$$Ent(D^1) = -(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}) = 0.918$$

$$Ent(D^2) = -(1\log_2 1 + 0) = 0$$

$$\begin{aligned} Gain(D, \text{喜欢ML作业}) &= Ent(D) - \sum_{v=1}^2 \frac{|D^v|}{|D|} Ent(D^v) \\ &= 0.971 - (\frac{3}{5} \times 0.918 + \frac{2}{5} \times 0) \\ &= 0.382 \end{aligned}$$

$$Gain(D, \text{喜欢ML作业}) = Gain(D, \text{性别})$$

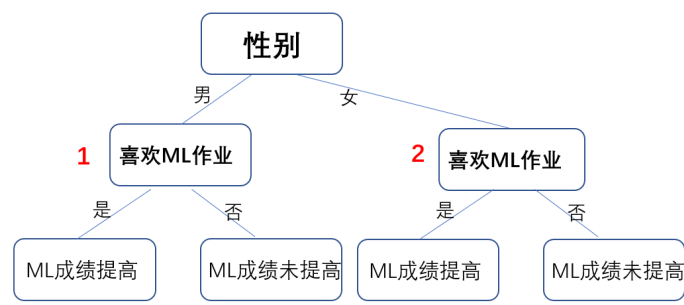
所以二者皆可



预剪枝

划分前	划分后	决策	训练集准确率	验证集准确率
1\4	1\4	禁止划分	3\5	1\4

后剪枝



	划分前	划分后	决策
1 处	1\2	1\4	不剪枝
2 处	1\2	1\2	剪枝
性别	1\2	1\4	不剪枝

训练集准确率： 1

验证集准确率： 1\2

后剪枝拟合能力更强