161910126 赵安

习题1

属性: 性别

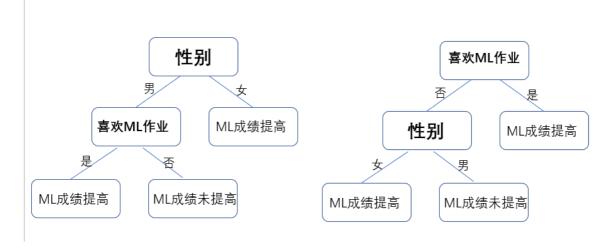
$$Ent(D) = -(rac{3}{5}log_2rac{3}{5} + rac{2}{5}log_2rac{2}{5}) = 0.971$$
 $Ent(D^1) = -(rac{1}{3}log_2rac{1}{3} + rac{2}{3}log_2rac{2}{3}) = 0.918$
 $Ent(D^2) = -(1log_21 + 0) = 0$
 $Gain(D, 性别) = Ent(D) - \sum_{v=1}^{2} rac{|D^v|}{|D|} Ent(D^v)$
 $= 0.971 - (rac{3}{5} \times 0.981 + rac{2}{5} \times 0)$
 $= 0.382$

属性: 喜欢ML作业

$$egin{aligned} Ent(D^1) &= -(rac{1}{3}log_2rac{1}{3} + rac{2}{3}log_2rac{2}{3}) = 0.918 \ Ent(D^2) &= -(1log_21 + 0) = 0 \ \\ Gain(D, 喜欢 ML作业) &= Ent(D) - \sum\limits_{v=1}^2rac{|D^v|}{|D|}Ent(D^v) \ \\ &= 0.971 - (rac{3}{5} imes 0.981 + rac{2}{5} imes 0) \ \\ &= 0.382 \end{aligned}$$

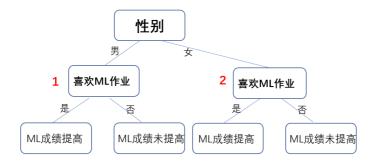
Gain(D,喜欢ML作业) = Gain(D,性别)

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



	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