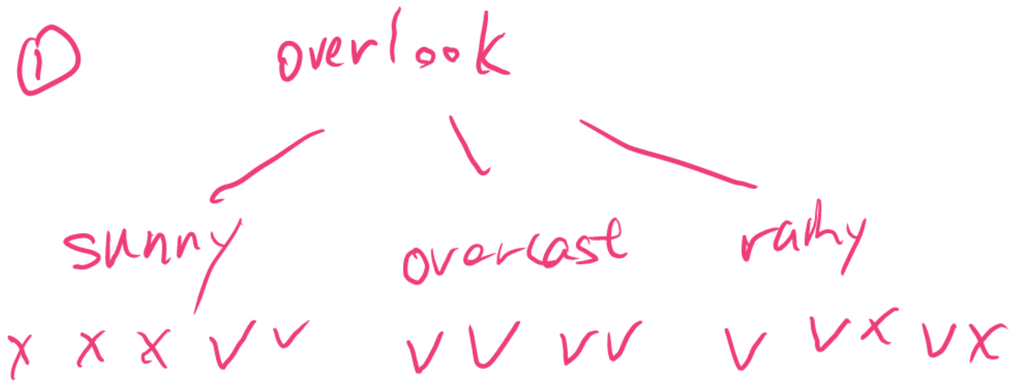
数据：

14天的气象数据(属性：outlook，temperature，humidity，windy)，并已知这些天气是否打球(play)。

问题：

根据决策树算法，构建一棵是否打球的决策树。并给出每个结点选择分裂属性时所作的计算。

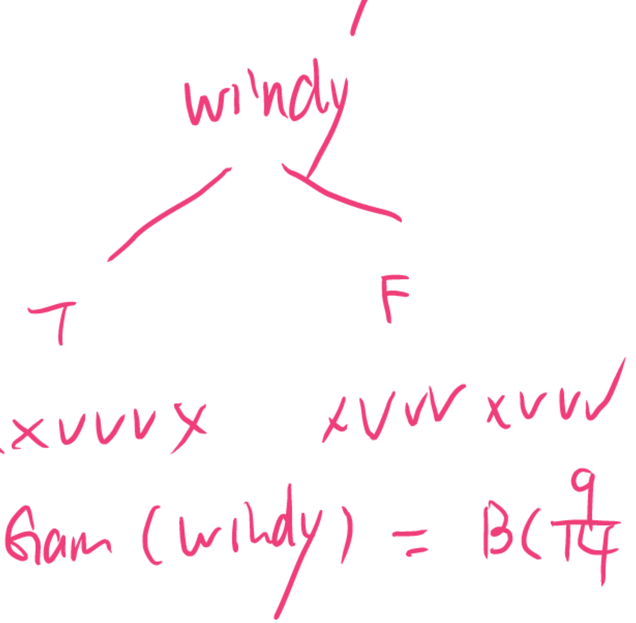
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no



$$Gain(outlook) = B\left(\frac{9}{14}\right) - \left(\frac{5}{14}B\left(\frac{2}{5}\right) + \frac{4}{14}B\left(\frac{4}{4}\right) + \frac{5}{14}B\left(\frac{3}{5}\right)\right) = 0.25$$

$$Gain(temperature) = B\left(\frac{9}{14}\right) - \left(\frac{4}{14}B\left(\frac{2}{4}\right) + \frac{6}{14}B\left(\frac{4}{6}\right) + \frac{4}{14}B\left(\frac{3}{4}\right)\right) = 0.03$$

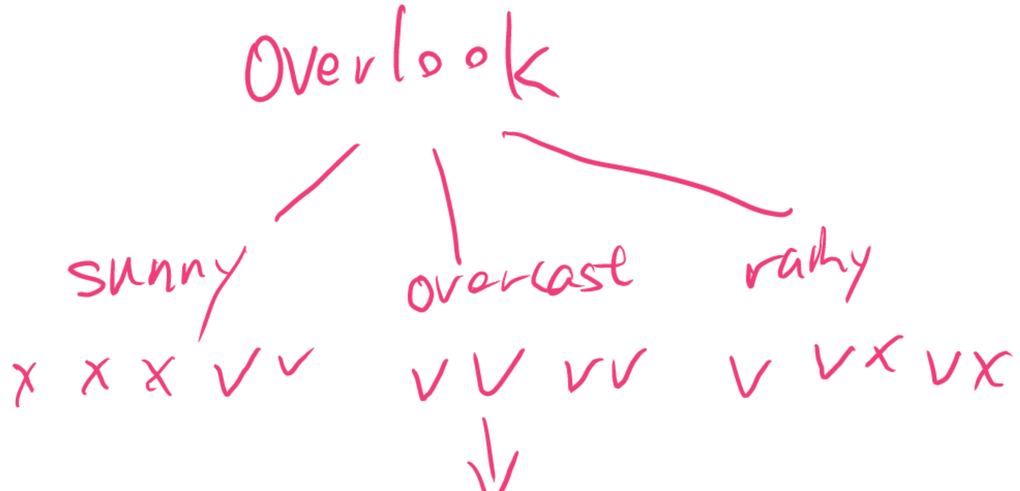
$$Gain(humidity) = B\left(\frac{9}{14}\right) - \left(\frac{7}{14}B\left(\frac{3}{7}\right) + \frac{7}{14}B\left(\frac{6}{7}\right)\right) = 0.15$$



$$Gain(windy) = B\left(\frac{9}{14}\right) - \sum \frac{6}{14} B\left(\frac{3}{6}\right) + \frac{8}{14} B\left(\frac{6}{8}\right) \approx 0.048$$

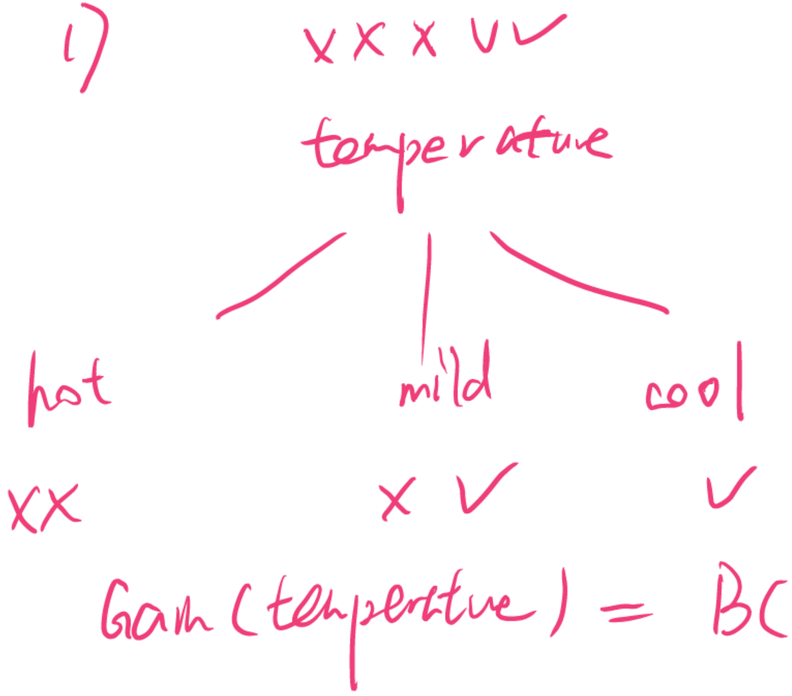
~~其~~ overlook

(2)



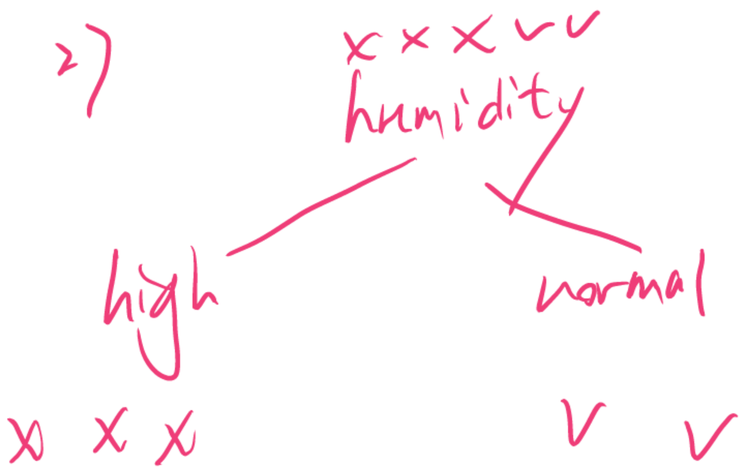
yes

1) x x x v v



$$Gain(temperature) = B\left(\frac{2}{5}\right) - \left[\frac{2}{5}B\left(\frac{0}{2}\right) + \frac{2}{5}B\left(\frac{1}{2}\right) + \frac{1}{5}B\left(\frac{1}{1}\right)\right] = 0.57$$

2) x x x v v



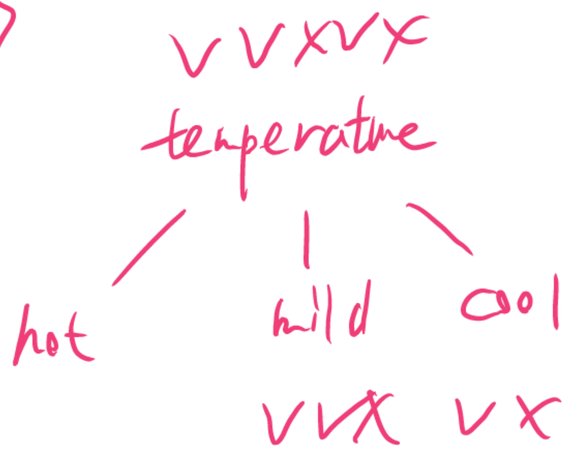
Gain = $B(\frac{2}{5}) - [\frac{3}{5}B(\frac{0}{3}) + \frac{2}{5}B(\frac{2}{2})] = 0.97$

由于 0.97 的 Reminder(humidity 是 0，不用管 windy)

又由于 rainy 分支最大为 $B(\frac{3}{5}) = 0.97$ 所以选 humidity

(3)

1)



$$Gain(temperature) = B\left(\frac{3}{5}\right) - \left[\frac{0}{5} + \frac{3}{5}B\left(\frac{2}{3}\right) + \frac{2}{5}B\left(\frac{1}{2}\right)\right] = 0.02$$

2)

Gain(outlook) = $B(\frac{5}{5}) - [\frac{5}{5} B(\frac{3}{5}) + \frac{5}{5} B(\frac{2}{2})] = 0.97$