



同济大学
TONGJI UNIVERSITY

地址: 中国上海市四平路1239号 邮编: 200092
1239 SIPING ROAD SHANGHAI CHINA 200092
电话(TEL): +86 21- 传真(FAX): +86 21-
网址(WEB): www.tongji.edu.cn

1. 计算 play 的信息熵

play = "no" : 5

play = "yes" : 9

信息熵计算:

$$H(\text{play}) = - \sum p(x_i) \log_2(p(x_i))$$

$$= - \frac{5}{14} \log_2 \frac{5}{14} - \frac{9}{14} \log_2 \frac{9}{14} \approx 0.9403$$

outlook 信息熵:

overcast	4	$\frac{4}{14}$	play = yes	play = no
rainy	5	$\frac{5}{14}$	4	0
Sunny	5	$\frac{5}{14}$	3	2
			2	3

$$H(\text{play} | \text{outlook}) = \frac{4}{14} \left(-\frac{4}{4} \log_2 \frac{4}{4} - \frac{4}{4} \log_2 \frac{4}{4} \right) + \frac{5}{14} \left(-\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} \right) + \frac{5}{14} \left(-\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} \right) = 0.6935$$

$$H(\text{play} | \text{temperature}) = \frac{4}{14} \times \left(-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} \right) + \frac{6}{14} \left(-\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} \right) + \frac{4}{14} \times \left(-\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} \right) = 0.9111$$

$$H(\text{play} | \text{humidity}) = \left(\frac{7}{14} \times 0.9852 + \frac{7}{14} \times 0.5917 \right) = 0.7885$$

$$H(\text{play} | \text{windy}) = \frac{8}{14} \times 0.8913 + \frac{6}{14} \times 1 = 0.8922$$

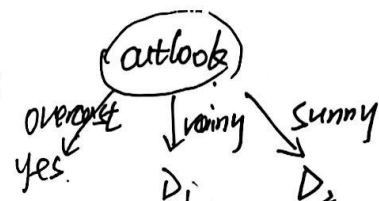
$$g(\text{play}, \text{outlook}) = 0.2467$$

$$g(\text{play}, \text{temperature}) = 0.0292$$

$$g(\text{play}, \text{humidity}) = 0.1518$$

$$g(\text{play}, \text{windy}) = 0.0481$$

构建决策树:



D1 计算 $H(play) = -(\frac{2}{5}) \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.9710$

$H(play|temperature) = \frac{2}{5} \times (-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2}) + \frac{3}{5} (-\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3}) = 0.9110$

$H(play|humidity) = \frac{2}{5} (-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2}) + \frac{3}{5} (-\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3}) = 0.9510$

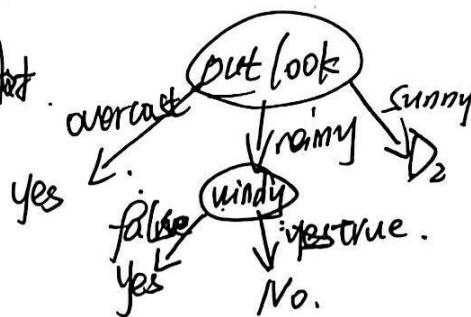
$H(play|windy) = \frac{2}{5} \times 0 + \frac{3}{5} \times 0 = 0$

$G(play, temperature) = 0.02$

$G(play, humidity) = 0.02$

$G(play, windy) = \underline{\underline{0.9710}}$

构建决策树



D2 计算 $H(play) = -(\frac{2}{5}) \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.9710$

$H(play|temperature) = \frac{2}{5} \times 0 + \frac{3}{5} \times (-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2}) + \frac{1}{5} \times 0 = 0.4$

$H(play|humidity) = \frac{2}{5} \times 0 + \frac{3}{5} \times 0 = 0$

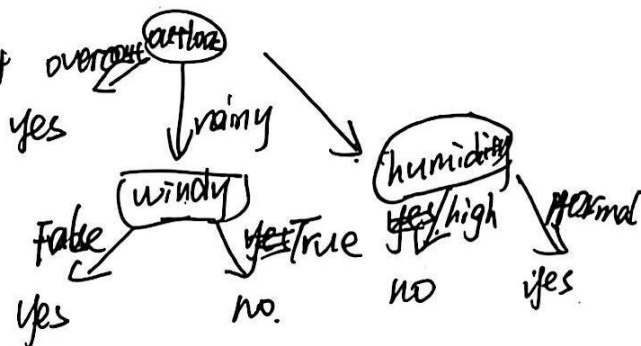
$H(play|windy) = \frac{2}{5} \times (-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2}) + \frac{3}{5} (-\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3}) = 0.9110$

$g(play|temperature) = 0.5710$

$g(play|humidity) = 0.9710$

$g(play|windy) = 0.2$

构建决策树



最终结果

