CS 4641

Homework 1

Hongrui Zheng

September 12, 2016

1 Decision Tree

(a)

$$S = [9+, 5-]$$
$$Entropy(S) = -\frac{9}{14} log_2 \frac{9}{14} - \frac{5}{14} log_2 \frac{5}{14} = 0.94.$$

For Outlook:

$$Values(Outlook) = sunny, overcast, rain$$
$$S_{sunny} = [2+, 3-]$$
$$S_{overcast} = [4+, 0-]$$
$$S_{rain} = [3+, 2-]$$
$$Entropy(S_{sunny}) = -\frac{2}{5} log_2 \frac{2}{5} - \frac{3}{5} log_2 \frac{3}{5} = 0.97$$
$$Entropy(S_{overcast}) = 0$$
$$Entropy(S_{rain}) = -\frac{3}{5} log_2 \frac{3}{5} - \frac{2}{5} log_2 \frac{2}{5} = 0.97$$

The information gain associated with the Outlook is:

$$Gain(S, Outlook) = 0.94 - \frac{5}{14} Entropy(S_{sunny}) - \frac{4}{14} Entropy(S_{overcast}) - \frac{5}{14} Entropy(S_{rain}) = 0.25$$

For Humidity:

$$Values(Humidity) = \leq 75, > 75$$
$$S_{\leq 75} = [4+, 1-]$$
$$S_{>75} = [5+, 4-]$$
$$Entropy(S_{\leq 75}) = -\frac{4}{5} log_2 \frac{4}{5} - \frac{1}{5} log_2 \frac{1}{5} = 0.72$$
$$Entropy(S_{>75}) = -\frac{5}{9} log_2 \frac{5}{9} - \frac{4}{9} log_2 \frac{4}{9} = 0.99$$

$$Gain(S, Humidity) = 0.94 - \frac{5}{14} Entropy(S_{\leq 75}) - \frac{9}{14} Entropy(S_{>75}) = 0.046$$

(b)

$$SplitInformation(S, Outlook) = -\frac{5}{14} log_2 \frac{5}{14} - \frac{4}{14} log_2 \frac{4}{14} - \frac{5}{14} log_2 414 - \frac{5}{14} log_2 \frac{5}{14} = 1.58$$
$$GainRatio(S, Outlook) = \frac{Gain(S, Outlook)}{SplitInformation(S, Outlook)} = \frac{0.25}{1.58} = 0.158$$
$$SplitInformatiion(S, Humidity) = -\frac{5}{14} log_2 \frac{5}{14} - \frac{9}{14} log_2 \frac{9}{14} = 0.94$$
$$GainRatio(S, Humidity) = \frac{Gain(S, Humidity)}{SplitInformation(S, Humidity)} = \frac{0.046}{0.94} = 0.049$$

(c)