

ID3 (Decision tree)

1400/08/30

Concept learning

Entropy :

$$\sum_{i=1} -p_i * \log_2(p_i)$$

Entropy(S) :

$$(-p_+ \log_2 p_+) - (-p_- \log_2 p_-)$$

Gain(S, A):

$$Entropy(S) - \sum_{v=Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

ID3

Day	Outlook	Temperature	Humidity	Windy	Play
1	Overcast	Hot	High	False	Yes
2	Rainy	Mild	High	False	Yes
3	Rainy	Cool	Normal	False	NO
4	Sunny	Mild	High	False	No
5	Overcast	Mild	High	False	Yes
6	Sunny	Cool	Normal	True	No
7	Sunny	Hot	Normal	True	Yes
8	Rainy	Cool	High	False	Yes
9	Sunny	Cool	High	False	Yes
10	Overcast	Cool	Normal	True	Yes
11	Sunny	Hot	High	True	Yes
12	Rainy	Hot	High	True	Yes

ID3

Entropy

$$\sum_{i=1} -p_i * \log_2(p_i) = -\frac{3}{12} \log_2 \frac{3}{12} - \frac{9}{12} \log_2 \frac{9}{12} = 0.811$$

ID3

Gain outlook:

$$\text{Gain}(S, \text{outlook}) = 0.811 - (0 + 0.27 + 0.40) = 0.141$$

$$\left\{ \begin{array}{l} \text{entropy}(\text{overcast}) \Rightarrow -\frac{3}{12} \left(\frac{3}{3} \log \frac{3}{3} + \frac{0}{3} \log \frac{0}{3} \right) \\ \text{entropy}(\text{rainy}) \Rightarrow -\frac{4}{12} \left(\frac{3}{4} \log \frac{3}{4} + \frac{1}{4} \log \frac{1}{4} \right) \\ \text{entropy}(\text{sunny}) \Rightarrow -\frac{5}{12} \left(\frac{3}{5} \log \frac{3}{5} + \frac{2}{5} \log \frac{2}{5} \right) \end{array} \right.$$

$$\text{Gain}(S, \text{outlook}) = 0.811 - (0 + 0.27 + 0.4) = 0.141$$

ID3

Gain Temperature:

$$\text{Gain}(S, \text{Temperature}) = 0.811 - (0 + 0.4 + 0.29) = 0.121$$

$$\left\{ \begin{array}{l} \text{entropy}(\text{hot}) \Rightarrow -\frac{4}{12} \left(\frac{4}{4} \log \frac{4}{4} + \frac{0}{4} \log \frac{0}{4} \right) \\ \text{entropy}(\text{mild}) \Rightarrow -\frac{3}{12} \left(\frac{2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{3} \right) \\ \text{entropy}(\text{cool}) \Rightarrow -\frac{5}{12} \left(\frac{3}{5} \log \frac{3}{5} + \frac{2}{5} \log \frac{2}{5} \right) \end{array} \right.$$

$$\text{Gain}(S, \text{Temperature}) = 0.811 - (0 + 0.4 + 0.29) = 0.121$$

ID3

Gain Humidity:

$$\text{Gain}(S, \text{Humidity}) = 0.811 - (0.33 + 0.36) = 0.121$$

$$\begin{cases} \text{entropy}(\text{high}) \Rightarrow -\frac{8}{12} \left(\frac{7}{8} \log \frac{7}{8} + \frac{1}{8} \log \frac{1}{8} \right) \\ \text{entropy}(\text{normal}) \Rightarrow -\frac{4}{12} \left(\frac{2}{3} \log \frac{2}{3} + \frac{2}{3} \log \frac{2}{3} \right) \end{cases}$$

$$\text{Gain}(S, \text{Humidity}) = 0.811 - (0.33 + 0.36) = 0.121$$

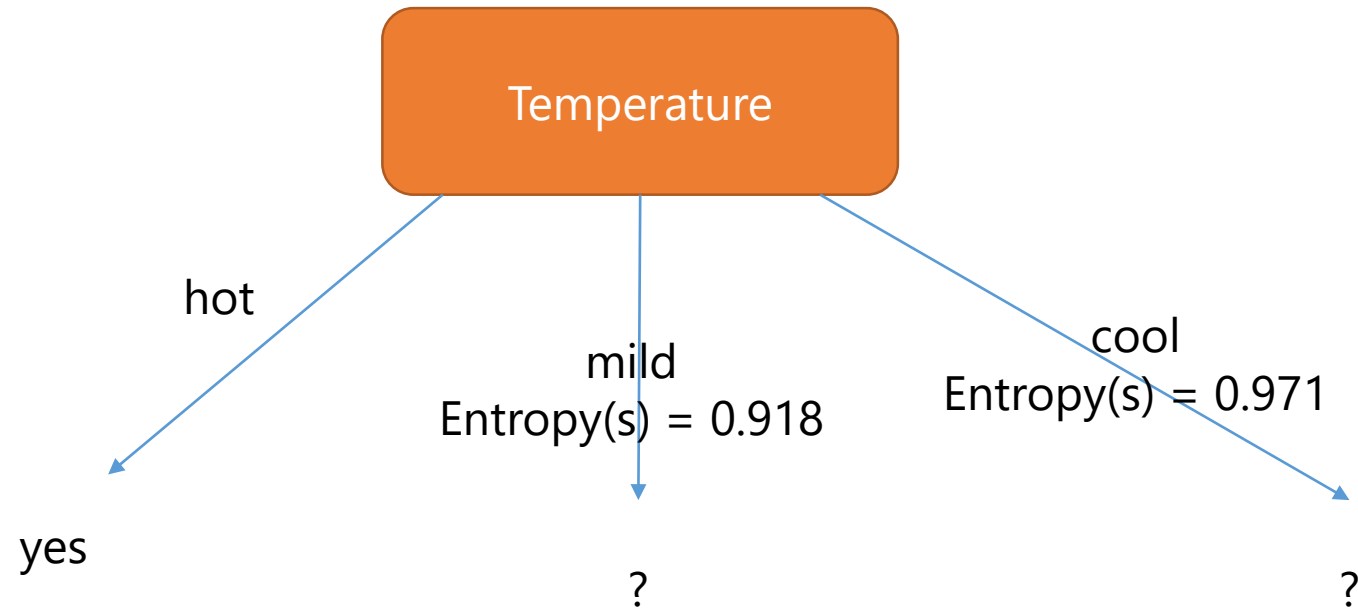
ID3

Gain Windy:

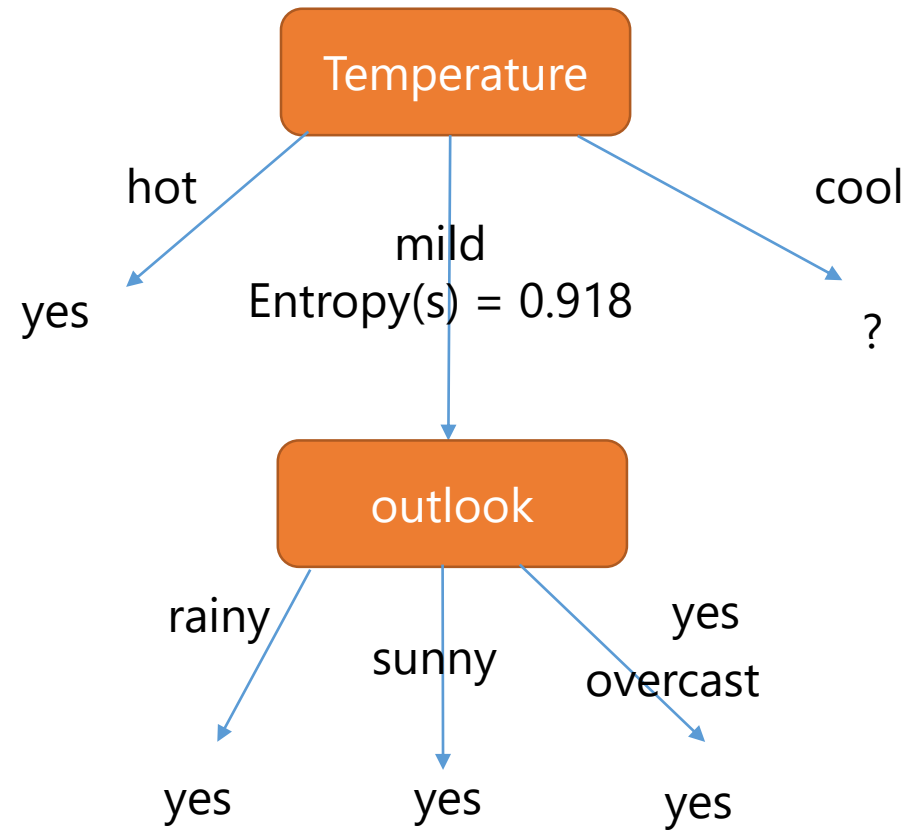
$$\text{Gain}(S, \text{Windy}) = 0.811 - (0.3 + 0.3) = 0.011$$

$$\begin{cases} \text{entropy}(\text{false}) \Rightarrow -\frac{7}{12} \left(\frac{5}{7} \log \frac{5}{7} + \frac{2}{7} \log \frac{2}{7} \right) \\ \text{entropy}(\text{true}) \Rightarrow -\frac{5}{12} \left(\frac{4}{5} \log \frac{4}{5} + \frac{1}{5} \log \frac{1}{5} \right) \end{cases}$$
$$\text{Gain}(S, \text{Windy}) = 0.811 - (0.5 + 0.3) = 0.011$$

ID3



ID3



ID3

Gain outlook: (temperature = cool)

$$\text{Gain}(\text{temperature}, \text{outlook}) = 0.918 - (0 + 0.4 + 0.4) = 0.118$$

$$\left\{ \begin{array}{l} \text{entropy}(\text{overcast}) \Rightarrow -\frac{1}{5} \left(\frac{1}{1} \log \frac{1}{1} + \frac{0}{1} \log \frac{0}{1} \right) \\ \text{entropy}(\text{rainy}) \Rightarrow -\frac{2}{5} \left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) \\ \text{entropy}(\text{sunny}) \Rightarrow -\frac{2}{5} \left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) \end{array} \right.$$

$$\text{Gain}(\text{temperature}, \text{outlook}) = 0.918 - (0 + 0.4 + 0.4) = 0.118$$

ID3

Gain Humidity: (temperature = cool)

$$\text{Gain}(\text{temperature}, \text{Humidity}) = 0.918 - (0 + 0.36) = 0.121$$

$$\begin{cases} \text{entropy}(\text{high}) \Rightarrow -\frac{2}{5} \left(\frac{2}{2} \log \frac{2}{2} + \frac{0}{2} \log \frac{0}{2} \right) \\ \text{entropy}(\text{normal}) \Rightarrow -\frac{3}{5} \left(\frac{1}{3} \log \frac{1}{3} + \frac{2}{3} \log \frac{2}{3} \right) \end{cases}$$

$$\text{Gain}(\text{temperature}, \text{Humidity}) = 0.918 - (0 + 0.55) = 0.368$$

ID3

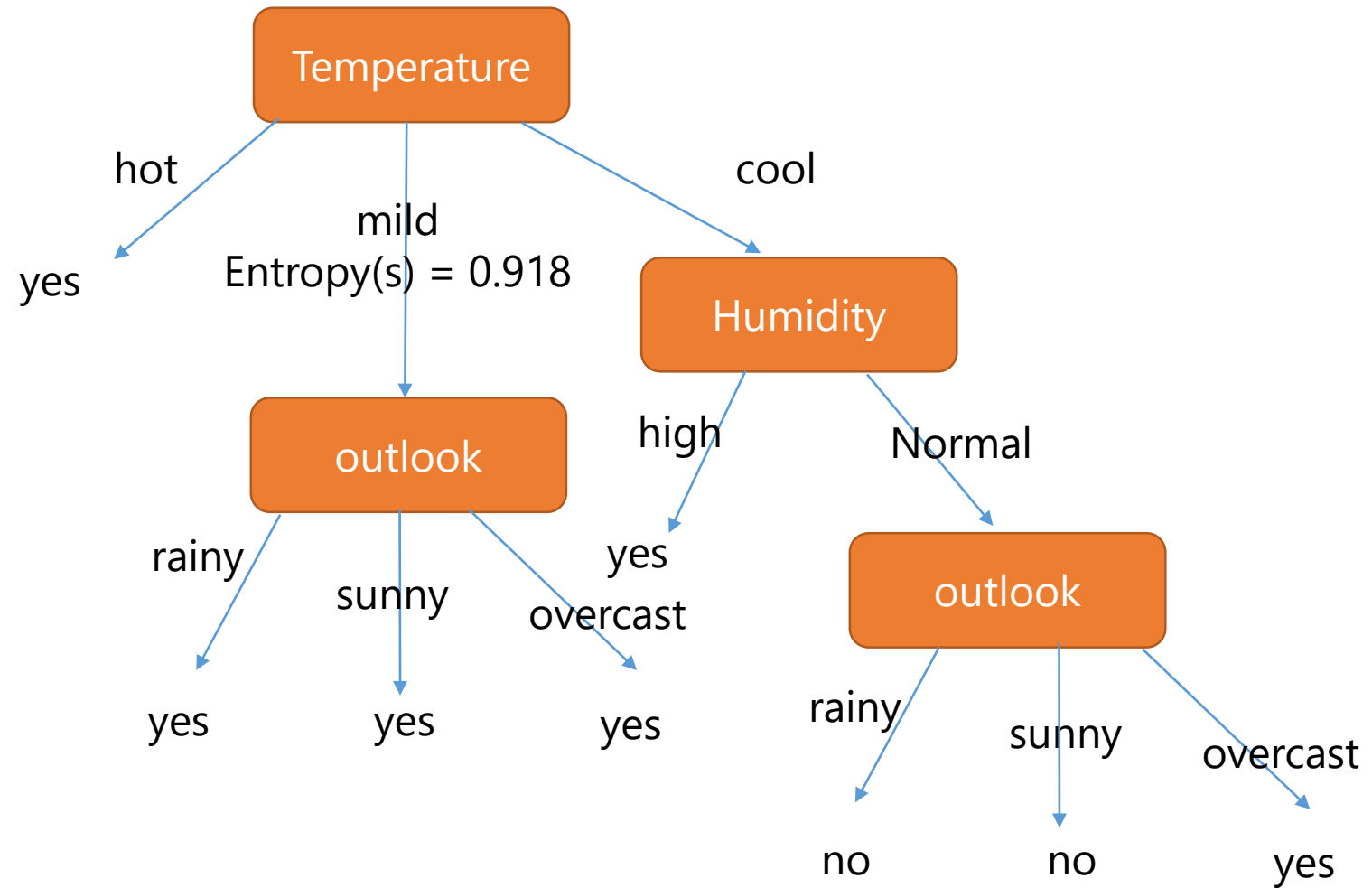
Gain Windy: (temperature = cool)

$$\text{Gain}(\text{temperature}, \text{Windy}) = 0.918 - (0.4 + 0.5) = 0.018$$

$$\begin{cases} \text{entropy}(\text{true}) \Rightarrow -\frac{2}{5} \left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) \\ \text{entropy}(\text{false}) \Rightarrow -\frac{3}{5} \left(\frac{2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{3} \right) \end{cases}$$

$$\text{Gain}(\text{temperature}, \text{Windy}) = 0.918 - (0.4 + 0.5) = 0.018$$

ID3



References

Machine learning book - Thomas Mitchell