

Decision Tree

- ① The first thing which attribute that we need to find is giving more information.
- ② To find this, we calculate information gain of all the attributes and the maximum of them becomes the root node of decision tree.

$$\text{Information Gain} = \text{Entropy}(S) - \sum_{v \in \{v_1, v_2, v_3\}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Attribute: Outlook

Values (Outlook) = Sunny, Overcast, Rain

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

$$S_{\text{sunny}} [2+, 3-] \quad \text{Entropy}(S) \\ = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} \\ = 0.971$$

S_{overcast} (+4, 0)

$$= -\frac{4}{4} \log_2 \frac{4}{4} - \frac{0}{4} \log_2 \frac{0}{4} = 0$$

S_{Rain} (+3, -2)

Entropy S_{Rain} = $-\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5}$

~~Entropy~~

$$= 0.971$$

IG (S_{outlook})

$$= \text{Entropy}(S) - \frac{5}{14} \text{Entropy}(S_{\text{sunny}})$$

$$- \frac{4}{14} \text{Entropy}(S_{\text{overcast}})$$

$$- \frac{5}{14} \text{Entropy}(S_{\text{Rain}})$$

$$= 0.94 - \frac{5}{14} (0.971) - \frac{4}{14} (0) - \frac{5}{14} (0.971)$$

$$= 0.2464$$

Similarly find for Temp,
humidity and wind

$$\checkmark \text{Gain}(S, \text{Outlook}) = 0.2464$$

$$\text{Gain}(S, \text{Temp}) = 0.0289$$

$$\text{Gain}(S, \text{Humidity}) = 0.1516$$

$$\text{Gain}(S, \text{Wind}) = 0.0478$$

Similar steps for Sunny as
root

$$\text{Gain}(S_{\text{Sunny}}, \text{Temp}) = 0.570$$

$$\checkmark \text{Gain}(S_{\text{Sunny}}, \text{Humidity}) = 0.97$$

$$\text{Gain}(S_{\text{Sunny}}, \text{Wind}) = 0.0192$$

Information Gain with Rain as
root.

$$\text{Gain}(S_{\text{Rain}}, \text{Temp}) = 0.0192$$

$$\text{Gain}(S_{\text{Rain}}, \text{Humidity}) = 0.0192$$

$$\checkmark \text{Gain}(S_{\text{Rain}}, \text{Wind}) = 0.97$$

