Example

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Decision Tree Model Building

$$\text{Entropy}(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

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

$$= -0.6428 \frac{\log(0.6428)}{\log 2} - 0.3571 \frac{\log(0.3571)}{\log 2}$$

$$= \frac{-0.6428 \times (-0.1919) - 0.3571 \times (-0.4472)}{0.3010}$$

$$= \frac{0.1233 + 0.1597}{0.3010} = \frac{0.283}{0.3010} = 0.9401$$

$$\text{Gain}(S, \text{Outlook}) = \text{Entropy}(S) - \sum_{v \in \{\text{Sunny, Overcast, Rain}\}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

[9+, 5-]

Outlook

- Sunny → {D1, D2, D8, D9, D11} [2+, 3-]
- Overcast → {D3, D7, D12, D13} [4+, 0-] → Yes → 0
- Rain → {D4, D5, D6, D10, D14} [3+, 2-]

$$E(\text{Outlook} = \text{Sunny}) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)$$

$$= -0.4 \log_2(0.4) - 0.6 \log_2(0.6)$$

$$= \frac{-0.4(-0.3979) - 0.6(-0.2218)}{0.3010}$$

$$= 0.9707$$

$$E(\text{Outlook} = \text{Rain}) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)$$

$$= 0.9709$$

Gain (S, Outlook) =

$$\text{Entropy}(S) - \frac{|S_{sunny}|}{|S|} E(\text{outlook} = \text{Sunny}) - \frac{|S_{overcast}|}{|S|} E(\text{Outlook} = \text{Overcast}) - \frac{|S_{Rain}|}{|S|} E(\text{Outlook} = \text{Rain})$$

$$= 0.94 - \frac{5}{14}(0.9707) - \frac{4}{14}(0) - \frac{5}{14}(0.9707)$$

$$= 0.94 - 0.6933$$

$$= 0.2468 \Rightarrow \text{Gain}(S, \text{Outlook}) = \underline{\underline{0.2468}}$$

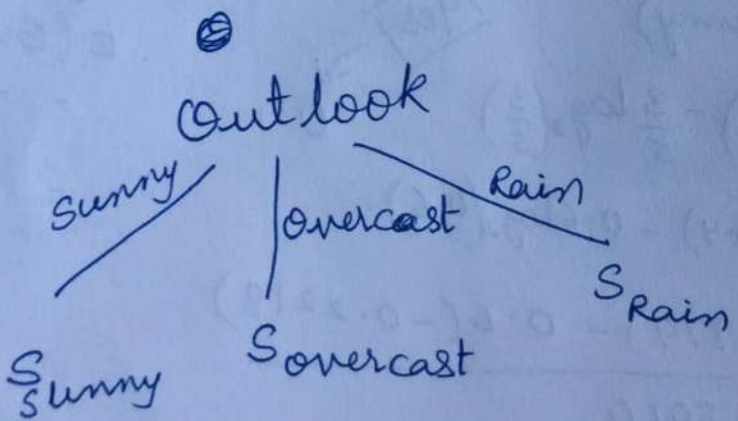
Similarly,

Gain (S, Humidity) = 0.151

Gain (S, Wind) = 0.084

Gain (S, Temperature) = 0.029

As "outlook" provides the best prediction, we pick "outlook" attribute to split the root node into three branches viz., "sunny, overcast, rain



Sunny

Temperature	Humidity	Wind	Play Tennis
Hot	High	Weak	No
Hot	High	Strong	No
Mild	High	Weak	No
Cool	Normal	Weak	Yes
Mild	Normal	Strong	Yes

Overcast

Temperature	Humidity	Wind	Play Tennis
Hot	High	Weak	Yes
Cool	Normal	Strong	Yes
Mild	High	Strong	Yes
Hot	Normal	Weak	Yes

Rain

Temperature	Humidity	Wind	Play Tennis
Mild	High	Weak	Yes
Cool	Normal	Weak	Yes
Cool	Normal	Strong	No
Mild	Normal	Weak	Yes
Mild	High	Strong	No

$E(Sunny) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$

Gain (S_{sunny}, Temperature)

$= E(S_{sunny}) - \left[\frac{2}{5}E(0,2) + \frac{2}{5}E(1,1) + \frac{1}{5}E(1,0)\right]$

$= 0.971 - 0.4 = 0.571$

$\therefore$ Gain (S_{sunny}, Temperature) $= 0.571$

IIIly Gain (S_{sunny}, Humidity) $= 0.971$

Gain (S_{sunny}, Windy) $= 0.020$