

Day	Outlook	Temp	Humidity	Wind	Play Tennis	Attribute: Outlook
D1	Sunny	Hot	High	Weak	No	$\text{Values (Outlook)} = \text{Sunny, Overcast, Rain}$
D2	Sunny	Hot	High	Strong	No	$S = [9+, 5-]$
D3	Overcast	Hot	High	Weak	Yes	$S_{\text{Sunny}} \leftarrow [2+, 3-]$
D4	Rain	Mild	High	Weak	Yes	$S_{\text{Overcast}} \leftarrow [4+, 0-]$
D5	Rain	Cool	Normal	Weak	Yes	$S_{\text{Rain}} \leftarrow [3+, 2-]$
D6	Rain	Cool	Normal	Strong	No	$\text{Entropy}(S) = -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} = 0.94$
D7	Overcast	Cool	Normal	Strong	Yes	$\text{Entropy}(S_{\text{Sunny}}) = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.971$
D8	Sunny	Mild	High	Weak	No	$\text{Entropy}(S_{\text{Overcast}}) = -\frac{4}{4} \log_2 \frac{4}{4} - \frac{0}{4} \log_2 \frac{0}{4} = 0$
D9	Sunny	Cool	Normal	Weak	Yes	$\text{Entropy}(S_{\text{Rain}}) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.971$
D10	Rain	Mild	Normal	Weak	Yes	$\text{Gain}(S, \text{Outlook}) = \text{Entropy}(S) - \sum_{v \in \{\text{Sunny, Overcast, Rain}\}} \frac{ S_v }{ S } \text{Entropy}(S_v)$
D11	Sunny	Mild	Normal	Strong	Yes	$= \text{Entropy}(S) - \frac{5}{14} \text{Entropy}(S_{\text{Sunny}}) - \frac{4}{14} \text{Entropy}(S_{\text{Overcast}})$
D12	Overcast	Mild	High	Strong	Yes	$- \frac{5}{14} \text{Entropy}(S_{\text{Rain}})$
D13	Overcast	Hot	Normal	Weak	Yes	$\text{Gain}(S, \text{Outlook}) = 0.94 - \frac{5}{14} 0.971 - \frac{4}{14} 0 - \frac{5}{14} 0.971 = 0.2464$
D14	Rain	Mild	High	Strong	No	

$P_v \log_2 P_v + P_{\bar{v}} \log_2 P_{\bar{v}}$   
 $= -\sum P_i \log_2 P_i$

Day	Outlook	Temp	Humidity	Wind	Play Tennis
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

### Attribute: Temp

$\text{Values (Temp)} = \text{Hot, Mild, Cool}$

$$S = [9+, 5-]$$

$$\text{Entropy}(S) = -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} = 0.94$$

$$S_{\text{Hot}} \leftarrow [2+, 2-]$$

$$\text{Entropy}(S_{\text{Hot}}) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} = 1.0$$

$$S_{\text{Mild}} \leftarrow [4+, 2-]$$

$$\text{Entropy}(S_{\text{Mild}}) = -\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} = 0.9183$$

$$S_{\text{Cool}} \leftarrow [3+, 1-]$$

$$\text{Entropy}(S_{\text{Cool}}) = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} = 0.8113$$

$$\text{Gain}(S, \text{Temp}) = \text{Entropy}(S) - \sum_{v \in \{\text{Hot, Mild, Cool}\}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$\begin{aligned} \text{Gain}(S, \text{Temp}) &= \text{Entropy}(S) - \frac{4}{14} \text{Entropy}(S_{\text{Hot}}) - \frac{6}{14} \text{Entropy}(S_{\text{Mild}}) \\ &\quad - \frac{4}{14} \text{Entropy}(S_{\text{Cool}}) = 0.9183 \end{aligned}$$

### Attribute: Humidity

$\text{Values (Humidity)} = \text{High, Normal}$

$$S = [9+, 5-]$$

$$\text{Entropy}(S) = -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} = 0.94$$

$$S_{\text{High}} \leftarrow [3+, 4-]$$

$$\text{Entropy}(S_{\text{High}}) = -\frac{3}{7} \log_2 \frac{3}{7} - \frac{4}{7} \log_2 \frac{4}{7} = 0.9852$$

$$S_{\text{Normal}} \leftarrow [6+, 1-]$$

$$\text{Entropy}(S_{\text{Normal}}) = -\frac{6}{7} \log_2 \frac{6}{7} - \frac{1}{7} \log_2 \frac{1}{7} = 0.5916$$

$$\text{Gain}(S, \text{Humidity}) = \text{Entropy}(S) - \sum_{v \in \{\text{High, Normal}\}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

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

$$\begin{aligned} &= \text{Entropy}(S) - \frac{7}{14} \text{Entropy}(S_{\text{High}}) - \frac{7}{14} \text{Entropy}(S_{\text{Normal}}) \\ &= 0.1516 \end{aligned}$$

### Attribute: Wind

$\text{Values (Wind)} = \text{Strong, Weak}$

$$S = [9+, 5-]$$

$$\text{Entropy}(S) = -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} = 0.94$$

$$S_{\text{Strong}} \leftarrow [3+, 3-]$$

$$\text{Entropy}(S_{\text{Strong}}) = 1.0$$

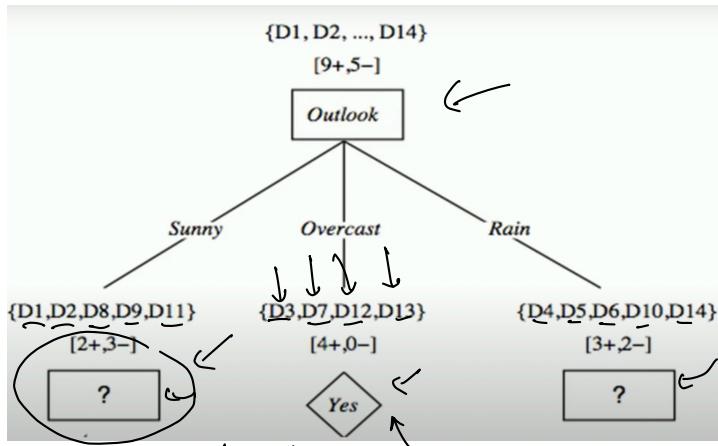
$$S_{\text{Weak}} \leftarrow [6+, 2-]$$

$$\text{Entropy}(S_{\text{Weak}}) = -\frac{6}{8} \log_2 \frac{6}{8} - \frac{2}{8} \log_2 \frac{2}{8} = 0.8113$$

$$\text{Gain}(S, \text{Wind}) = \text{Entropy}(S) - \sum_{v \in \{\text{Strong, Weak}\}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$\text{Gain}(S, \text{Wind}) = \text{Entropy}(S) - \frac{6}{14} \text{Entropy}(S_{\text{Strong}}) - \frac{8}{14} \text{Entropy}(S_{\text{Weak}})$$

$$0.0478$$



Day	Temp	Humidity	Wind	Play Tennis
D1	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
D11	Mild	Normal	Strong	Yes

### Attribute: Temp

Values (Temp) = Hot, Mild, Cool

$$S_{Sunny} = [2+, 3-] \quad Entropy(S_{Sunny}) = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.97$$

$$S_{Hot} \leftarrow [0+, 2-] \quad Entropy(S_{Hot}) = 0.0$$

$$S_{Mild} \leftarrow [1+, 1-] \quad Entropy(S_{Mild}) = 1.0$$

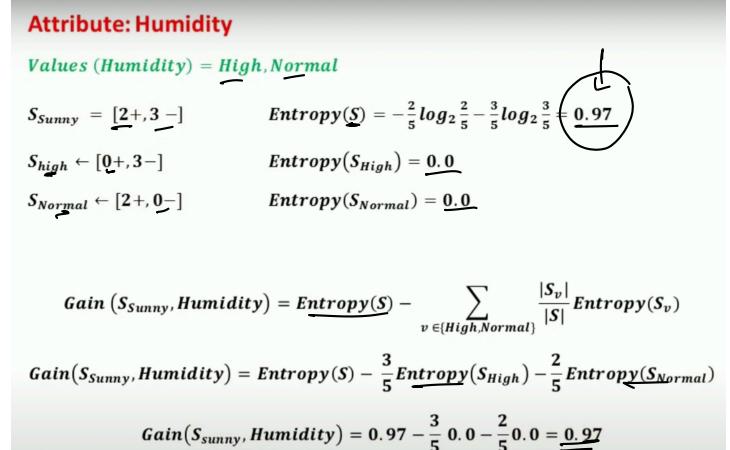
$$S_{Cool} \leftarrow [1+, 0-] \quad Entropy(S_{Cool}) = 0.0$$

$$Gain(S_{Sunny}, Temp) = Entropy(S) - \sum_{v \in \{Hot, Mild, Cool\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

Gain(S<sub>Sunny</sub>, Temp)

$$= Entropy(S) - \frac{2}{5} Entropy(S_{Hot}) - \frac{2}{5} Entropy(S_{Mild}) \\ - \frac{1}{5} Entropy(S_{Cool})$$

$$Gain(S_{Sunny}, Temp) = 0.97 - \frac{2}{5} 0.0 - \frac{2}{5} 1 - \frac{1}{5} 0.0 = 0.570$$



### Attribute: Wind

Values (Wind) = Strong, Weak

$$S_{Sunny} = [2+, 3-] \quad Entropy(S) = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.97$$

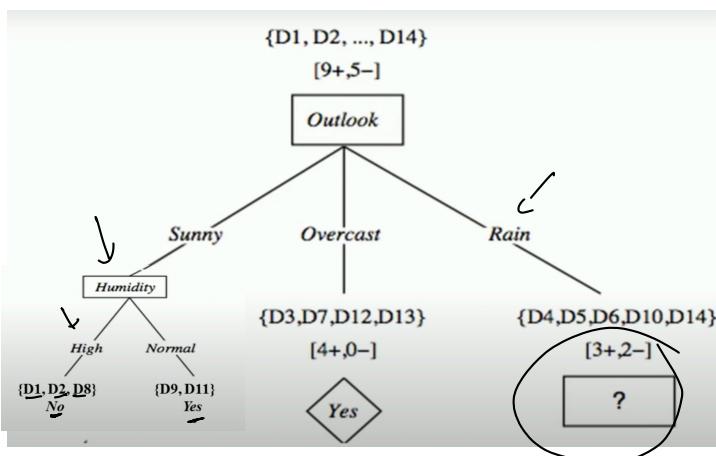
$$S_{Strong} \leftarrow [1+, 1-] \quad Entropy(S_{Strong}) = 1.0$$

$$S_{Weak} \leftarrow [1+, 2-] \quad Entropy(S_{Weak}) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.9183$$

$$Gain(S_{Sunny}, Wind) = Entropy(S) - \sum_{v \in \{Strong, Weak\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Sunny}, Wind) = Entropy(S) - \frac{2}{5} Entropy(S_{Strong}) - \frac{3}{5} Entropy(S_{Weak})$$

$$Gain(S_{Sunny}, Wind) = 0.97 - \frac{2}{5} 1.0 - \frac{3}{5} 0.918 = 0.0192$$



Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes ✓
D5	Cool	Normal	Weak	Yes ✓
D6	Cool	Normal	Strong	No ✓
D10	Mild	Normal	Weak	Yes ✓
D14	Mild	High	Strong	No ✓

### Attribute: Temp

Values (Temp) = Hot, Mild, Cool

$$S_{Rain} = [3+, 2-]$$

$$\text{Entropy}(S_{Sunny}) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.97$$

$$S_{Hot} \leftarrow [0+, 0-]$$

$$\text{Entropy}(S_{Hot}) = 0.0$$

$$S_{Mild} \leftarrow [2+, 1-]$$

$$\text{Entropy}(S_{Mild}) = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} = 0.9183$$

$$S_{Cool} \leftarrow [1+, 1-]$$

$$\text{Entropy}(S_{Cool}) = 1.0$$

$$\text{Gain}(S_{Rain}, \text{Temp}) = \text{Entropy}(S) - \sum_{v \in \{\text{Hot, Mild, Cool}\}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

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

$$= \text{Entropy}(S) - \frac{0}{5} \text{Entropy}(S_{Hot}) - \frac{3}{5} \text{Entropy}(S_{Mild})$$

$$- \frac{2}{5} \text{Entropy}(S_{Cool})$$

$$\text{Gain}(S_{Rain}, \text{Temp}) = 0.97 - \frac{0}{5} 0.0 - \frac{3}{5} 0.918 - \frac{2}{5} 1.0 = 0.0192$$

### Attribute: Humidity

Values (Humidity) = High, Normal

$$S_{Rain} = [3+, 2-]$$

$$\text{Entropy}(S_{Sunny}) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.97$$

$$S_{High} \leftarrow [1+, 1-]$$

$$\text{Entropy}(S_{High}) = 1.0$$

$$S_{Normal} \leftarrow [2+, 1-]$$

$$\text{Entropy}(S_{Normal}) = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} = 0.9183$$

$$\text{Gain}(S_{Rain}, \text{Humidity}) = \text{Entropy}(S) - \sum_{v \in \{\text{High, Normal}\}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$\text{Gain}(S_{Rain}, \text{Humidity}) = \text{Entropy}(S) - \frac{2}{5} \text{Entropy}(S_{High}) - \frac{3}{5} \text{Entropy}(S_{Normal})$$

$$\text{Gain}(S_{Rain}, \text{Humidity}) = 0.97 - \frac{2}{5} 1.0 - \frac{3}{5} 0.918 = 0.0192$$

### Attribute: Wind

Values (wind) = Strong, Weak

$$S_{Rain} = [3+, 2-]$$

$$\text{Entropy}(S_{Sunny}) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.97$$

$$S_{Strong} \leftarrow [0+, 2-]$$

$$\text{Entropy}(S_{Strong}) = 0.0$$

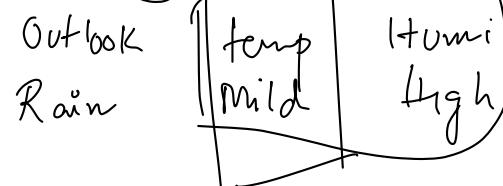
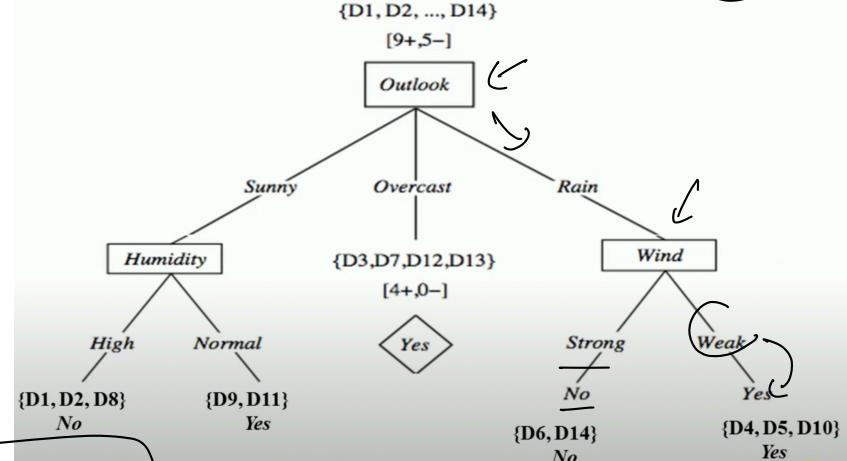
$$S_{Weak} \leftarrow [3+, 0-]$$

$$\text{Entropy}(S_{Weak}) = 0.0$$

$$\text{Gain}(S_{Rain}, \text{Wind}) = \text{Entropy}(S) - \sum_{v \in \{\text{Strong, Weak}\}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$\text{Gain}(S_{Rain}, \text{Wind}) = \text{Entropy}(S) - \frac{2}{5} \text{Entropy}(S_{Strong}) - \frac{3}{5} \text{Entropy}(S_{Weak})$$

$$\text{Gain}(S_{Rain}, \text{Wind}) = 0.97 - \frac{2}{5} 0.0 - \frac{3}{5} 0.0 = 0.97$$



wind  
weak

### ① Calculate Root Node

↳ Highest Information gain

↳ Lowest Entropy.

for example:

features = [outlook, weather (clouding)]

↳ Entropy formula =

↳  $-\log -$