



## Decision Tree - ID3 Algorithm

Day	Outlook	Temp	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

Four attributes : Outlook, Temp, Humidity, Wind  
 Target Output : Play Tennis

To draw Decision Tree,  
 We need to find the attribute giving maximum information.

Attribute that having max information gain will be the root node in decision tree.



Attribute : Outlook

Values (Outlook) = Sunny, Overcast, Rain

$S$  We need to calculate the entropy of whole dataset and entropy of the individual attribute.

Let  $S$  be whole dataset

$$S = [9+, 5-] \Rightarrow 9 \text{ Yes } 5 \text{ No}$$

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

$$S_{\text{Sunny}} = [2+, 3-]$$

$$\text{Entropy}(S_{\text{Sunny}}) = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.971$$

~~$S_{\text{Sunny}} S_{\text{Overcast}} = [4+, 0-]$~~

$$\text{Entropy}(S_{\text{Overcast}}) = -\frac{4}{4} \log_2 \frac{4}{4} - \frac{0}{4} \log_2 \frac{0}{4} = 0$$

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

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

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

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

=

$$= \text{Entropy}(S) - \frac{5}{14} \text{Entropy}(\text{Sunny}) - \frac{4}{14} \text{Entropy}(\text{Overcast}) \\ - \frac{5}{14} \text{Entropy}(\text{Rainy})$$

$$= 0.94 - \frac{5}{14} \times 0.971 - \frac{4}{14} \times 0.000 - \frac{5}{14} \times 0.971$$

$$= \underline{\underline{0.2464}}$$

Attribute : Temp

Values(Temp) = Hot, Mild, Cool.

$$S = [9+, 5-] \quad 9 \text{ Yes} \quad 5 \text{ No}$$

$$\text{Entropy}(S) = 0.94$$

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

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

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

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

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

$$\text{Entropy}(S_{\text{Cool}}) = -\frac{3}{8} \log_2 \frac{3}{2^8} - \frac{1}{8} \log_2 \frac{1}{2^8} = 0.8113$$

Grain (S, Temp)

$$0.94 - \frac{4}{14} \text{ Entropy (S Hot)} = \frac{6}{14} \text{ Entropy (S Mild)}$$
$$= \frac{4}{14} \text{ Entropy (S cool)}$$
$$= 0.94 - \frac{4}{14} * 1 = \frac{6}{14} \times 0.9183 = \frac{4 \times 0.8113}{14}$$
$$= \underline{\underline{0.289}}$$

Attribute : Humidity

Values (Humidity) - High, Normal.

S [9+, 5-]

$$\text{Entropy (S)} = 0.94$$

S<sub>High</sub> = [3+, 4-]

$$\text{Entropy (S High)} = -\frac{3}{7} \log_2 \frac{3}{27} + \frac{4}{7} \log_2 \frac{4}{27}$$
$$= 0.9852$$

S<sub>Normal</sub> = [5+, 2-]

$$\text{Entropy (S Normal)} = -\frac{5}{7} \log_2 \frac{5}{27} - \frac{2}{7} \log_2 \frac{2}{27}$$

$$= 0.5916$$

$$\text{Grain (S, Humidity)} = 0.94 - \frac{7}{14} \text{ Entropy (S High)}$$

$$= \frac{7}{14} \text{ Entropy (S normal)}$$



$$= 0.94 - \frac{7}{14} \times 0.9852 - \frac{7}{14} \times 0.5916 = \underline{\underline{0.4516}}$$

Attribute : Wind

Values(Wind) = Weak, Strong

S[9+, 5-]

Entropy (S) = 0.94

Sweak = [6+, 2-]

$$\begin{aligned} \text{Entropy (Sweak)} &= -\frac{6}{8} \log \frac{6}{28} - \frac{2}{8} \log \frac{2}{28} \\ &= 0.8113 \end{aligned}$$

Sstrong = [3+3-]

$$\begin{aligned} \text{Entropy (Strong)} &= -\frac{3}{6} \log \frac{3}{26} - \frac{3}{6} \log \frac{3}{26} \\ &= 1.0 \end{aligned}$$

$$\begin{aligned} \text{Gain}(S, \text{Wind}) &= 0.94 - \frac{8}{14} \times 0.8113 \\ &\quad - \frac{6}{14} \times 1.0 \\ &= \underline{\underline{0.478}} \end{aligned}$$

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

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

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

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

Max gain is for outlook = 0.2464

Consider outlook as the root node

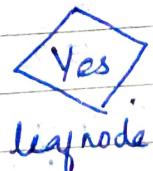
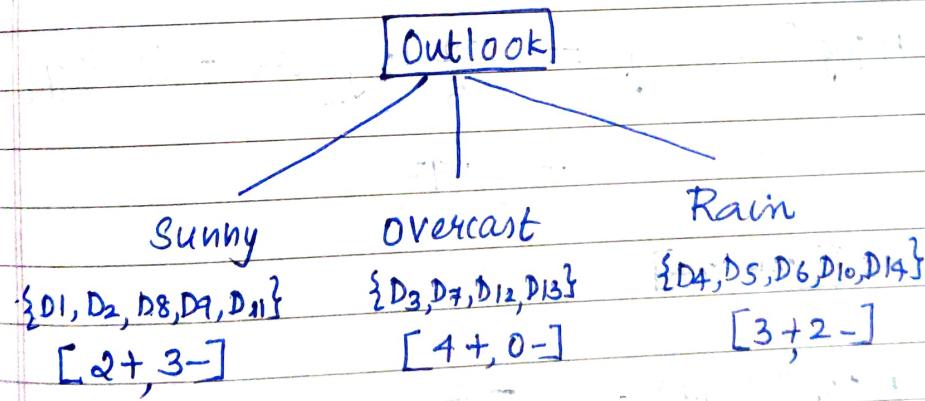
3 possibilities are there for outlook

- 1. sunny
- 2. overcast
- 3. rain

$\{D_1, D_2, \dots, D_4\}$

Decision Tree level 1

9+ 5-



Outlook = Sunny

	Day	Temp	Humidity	Wind	Play Tennis
D1	Hot	High	Weak	No	
D2	Hot	High	Strong	No	
D3	Mild	High	Weak	No	
D4	Cool	Normal	Weak	Yes	
D5	Mild	Normal	Strong	Yes	

Attribute - Temp

Values (Temp) = Hot, Mild, Cool

Sunny = [2+, 3-]

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

Hot = [0+, 2-]

$$\text{Entropy (Hot)} = \frac{-2}{2} \log_2 \frac{2}{2}$$
$$= -\frac{0}{2} \log_2 \frac{0}{2} - \frac{2}{2} \log_2 \frac{2}{2}$$
$$= 0.0$$

Mild = [1+, 1-]

$$\text{Entropy (Mild)} = \frac{-1}{2} \log_2 \frac{1}{2} + -\frac{1}{2} \log_2 \frac{1}{2}$$
$$= 1.0$$



$$S_{ool} = [1+, 0-]$$

$$= -\frac{1}{2} \log_2 \frac{1}{2} = 0$$

Brain (Sunny, Temp)

$$= \text{Entropy}(s) - \sum_{v \in \{hot, \text{Mild}\}} \frac{|sv|}{|S|} \text{Entropy}(S_v)$$

$$= \text{Entropy}(s) - \frac{2}{5} \times \text{Entropy}(S_{hot}) -$$

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

$$\frac{1}{5} \times \text{Entropy}(S_{ool})$$

$$= 0.97 - \frac{2}{5} \times 0.97 - \frac{2}{5} \times 1 - \frac{1}{5} \times 0$$

$$= \underline{\underline{0.570}}$$

Attribute : Humidity

Values [Humidity] = High, Normal

$S_{humidity} [2+, 3-]$

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

$$= \underline{\underline{0.97}}$$

$S_{normal} = [2+, 0-]$

$$\begin{aligned} \text{Entropy}(S_{normal}) &= -\frac{2}{2} \cdot \log_2 \frac{2}{2} - 0 \\ &= 0 \end{aligned}$$

$S_{high} = [0+, 3-]$

$$\begin{aligned} \text{Entropy}(S_{high}) &= -\frac{0}{3} \log_2 \frac{0}{3} - \frac{3}{3} \log_2 \frac{3}{3} \\ &= 0 \end{aligned}$$

Gain(Sunny, Humidity)

$$\begin{aligned} &= \text{Entropy}(s) - \frac{3}{5} \frac{\text{Entropy}(S_{high}) - \frac{2}{5}}{\text{Entropy}(S_{normal})} \\ &= \underline{0.97} - \underline{0} = \underline{0.97} \end{aligned}$$

Attribute : Wind

Values(Wind) - Strong, Weak

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

$$\text{Entropy}(s) = 0.97$$

$S_{strong} = [1+, 1-]$

$$\text{Entropy}(s_{strong}) = 1$$

Sweak = [1+, 2-]

$$\text{Entropy (Sweak)} = \frac{-1}{3} \log_2 \frac{1}{2} \cdot \frac{2}{3} - \frac{2}{3} \log_2 \frac{2}{3}$$
$$= \underline{\underline{0.9183}}$$

$$\text{Gain (Ssunny, Wind)} = \text{Entropy (S)} - \frac{2}{5} \times \text{Entropy (BS strong)}$$
$$- \frac{3}{5} \times \text{Entropy (Sweak)}$$

$$= 0.97 - \frac{2}{5} \times 1 - \frac{3}{5} \times 0.918 = \underline{\underline{0.0192}}$$

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

$$\text{Gain (Ssunny, Humidity)} = 0.97$$

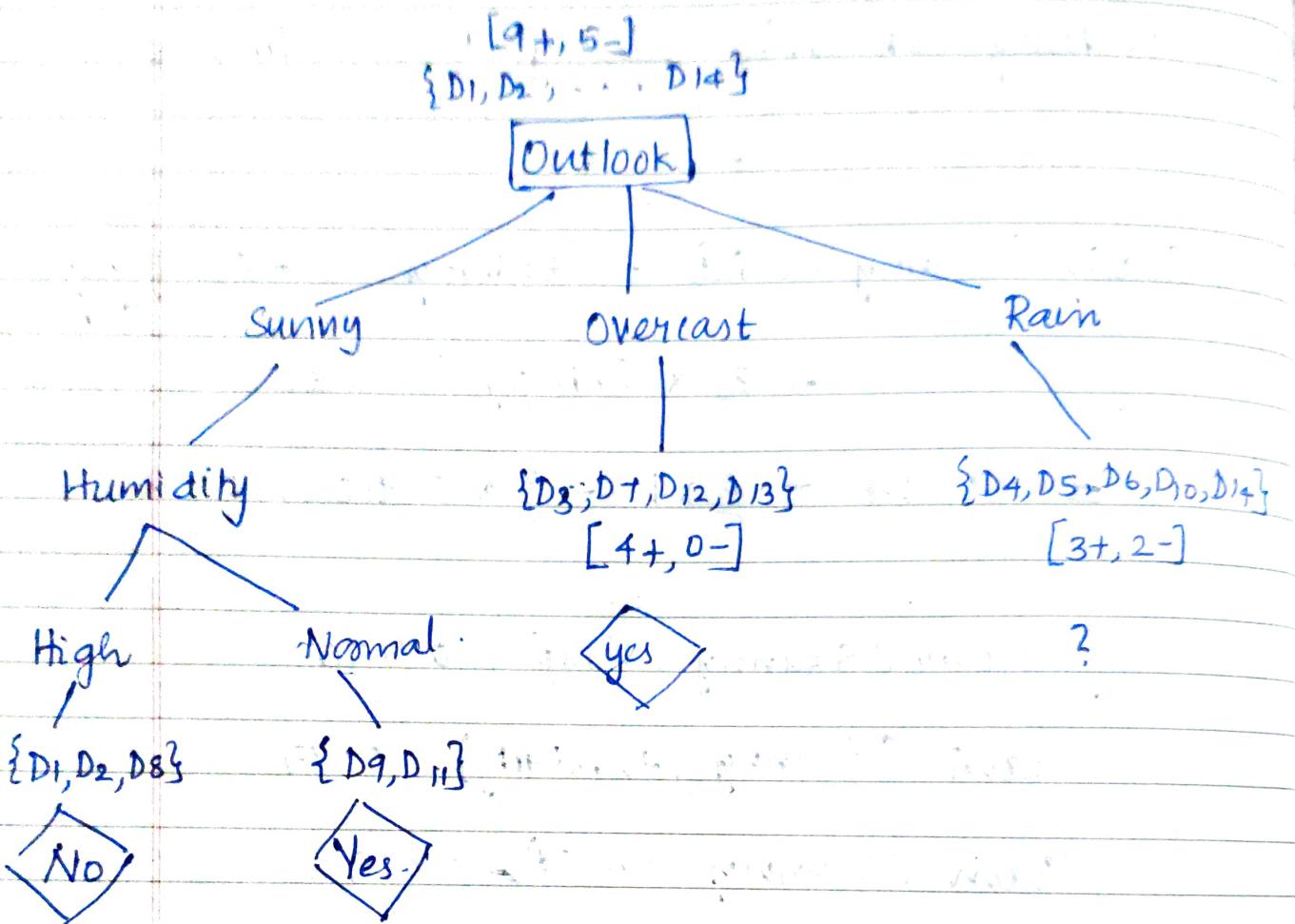
$$\text{Gain (Ssunny, Wind)} = 0.0192$$

Humidity is having high information gain  
so consider humidity as this level.

when we consider humidity the values are  
normal & high  
 $3 \text{ high} \Rightarrow \text{No}$  & Normal  $\Rightarrow \text{Yes}$ .

(contd)

## Decision tree level 2



Rain      Outlook = Rain

Attribute :

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

$$S_{\text{rain}} = [3+, 2-]$$

$$\text{Entropy}(S_{\text{rain}}) = \frac{-3}{5} \log \frac{3}{25} + \frac{2}{5} \log \frac{2}{5}$$

$$= 0.97$$

$$S_{\text{mild}} = [2+, 1-]$$

$$\text{Entropy}(S_{\text{mild}}) = \frac{-2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{2}$$

$$= 0.9183$$

$$S_{\text{cool}} = [1+, 1-]$$

$$\text{Entropy}(S_{\text{cool}}) = \frac{-1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}$$

$$= 1.$$

$$S_{\text{rain, Temp}} = \text{Entropy}(S) - \frac{0}{5} \times \text{Entropy}(S_{\text{hot}})$$

$$- \frac{3}{5} \text{Entropy}(S_{\text{mild}}) - \frac{2}{5} \text{Entropy}(S_{\text{cool}})$$

$$= 0.97 - \frac{0}{5} \cdot 0.0 - \frac{3}{5} \times 0.918 - \frac{2}{5} \times 1.0 = \underline{\underline{0.0192}}$$

Date: \_\_\_\_\_



Attribute: Temp

Values(Temp) = Hot, Mild, Cool

Strain = [3+, 2-]

$$\text{Entropy(Strain)} = \frac{-3}{5} \log_2 \frac{3}{2^5} + \frac{2}{5} \log_2 \frac{2}{2^5}$$
$$= 0.97$$

So Strain = [2+, 1-]

$$\text{Entropy(Smild)} = \frac{-2}{3} \log_2 \frac{2}{2^3} + \frac{1}{3} \log_2 \frac{1}{2^3}$$
$$= 0.9183$$

S Cool = [1+, 1-]

$$\text{Entropy(SCool)} = \frac{-1}{2} \log_2 \frac{1}{2^2} + \frac{1}{2} \log_2 \frac{1}{2^2}$$
$$= 1.$$

$$\text{Grain(Strain, Temp)} = \text{Entropy}(S) - \frac{0}{5} \times \text{Entropy}(Hot)$$

$$- \frac{3}{5} \text{ Entropy(Smild)} - \frac{2}{5} \text{ Entropy(SCool)}$$

$$= 0.97 - \frac{0}{5} \cdot 0.0 - \frac{3}{5} \times 0.918 - \frac{2}{5} \times 1.0 = \underline{\underline{0.0192}}$$

Attribut : Humidity

Values (humidity) = high, normal

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

$$\text{Entropy} = 0.97$$

$$S_{\text{Not Rain}} = [0+, 0-]$$

$$S_{\text{high}} = [1+, 1-]$$

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

$$S_{\text{normal}} = [2+, 1-]$$

$$\begin{aligned}\text{Entropy}(S_{\text{normal}}) &= -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} \\ &= 0.9183\end{aligned}$$

Gram ( $S_{\text{Rain}}, S_{\text{humidity}}$ ) =

$$\text{Entropy}(S) = \frac{2}{5} \text{Entropy}(S_{\text{high}}) + \frac{3}{5} \text{Entropy}(S_{\text{normal}})$$

$$= 0.97 - \frac{2}{5} \times 1.0 - \frac{3}{5} \times 0.918$$

$$= 0.0192$$



Date \_\_\_\_\_

Attribute : Wind

Values(Wind) = strong, weak.

Strain = [3+, 2-] Entropy(Strain) = 0.97

Sstrong = [0+, 2-] Entropy(Sstrong) = 0.

Sweak = [3+, 0-] Entropy(Sweak) = 0

$$\text{Grain}(Strain, Wind) = \text{Entropy}(S) - \frac{2}{5} \times 0 + \frac{3}{5} \times 0$$

$$= \underline{\underline{0.97}}$$

$$\text{Grain}(Strain, Stemp) = 0.0192$$

$$\text{Grain}(Strain, Humidity) = 0.0192$$

$$\text{Grain}(Strain, Wind) = 0.97$$

Consider Wind as node

Wind = Weak  $\Rightarrow$  yes

Wind = strong  $\Rightarrow$  No.

$\{D_1, D_2, \dots, D_{14}\}$   
 $[9+, 5-]$

Outlook

Sunny

Overcast

Rain

Humidity

High

Normal

$\{D_3, D_7, D_{12}, D_{13}\}$

$[4+, 0-]$

Wind

Strong

Weak

$\{D_1, D_2, D_8\}$

No

$\{D_9, D_{11}\}$

Yes.

$\{D_6, D_{14}\}$

No

$\{D_4, D_5, D_{10}\}$

yes.