

Decision Tree

ID3 - Example (Numerical)

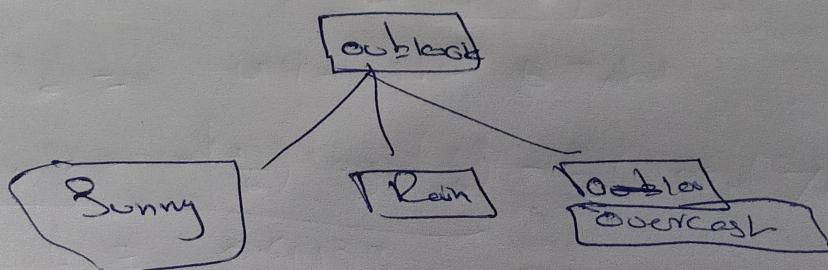
1st stage

$$\text{Entropy } (S) = -(P(Y)(\log_2(P_Y))) - (P(N)(\log_2(P_N)))$$

$$\text{Information gain} = \text{Whole Entropy } (S) - \sum_{(R_i)} \frac{|S_i|}{|S|} \text{Entropy}_{(R_i)}$$

1st Attribute

value (outlook) = sunny, overcast, Rain,



1st Outlook Entropy

①

$$= \text{outlook} = 9Y / 5N = 1.4$$

$$S = -\left(\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14} \log_2\left(\frac{5}{14}\right)\right)\right)$$

$$S = -(0.6324 \cdot -0.6324) - (0.357 \cdot -1.4)$$

$$S = -(-0.4232) - (0.357 \cdot -1.4854)$$

$$S = +0.4232 + 0.521$$

$$S = 0.94$$

- entropy for outlook attribute
- outlook

2nd Entropy

Dry / Sunny / Rain / overcast

②

$$\text{Entropy}_{\text{sunny}} = 3N / 2Y = 5$$

$$S_{\text{sunny}} = -\frac{2}{5} \log\left(\frac{2}{5}\right) - \frac{3}{5} \log\left(\frac{3}{5}\right) = 0.971$$

En

$$S_{\text{sunny}} = 0.971$$

③

$$\text{Entropy}_{\text{Rain}} = 3Y / 2N = 5$$

$$S_{\text{Rain}} = -\frac{3}{5} \log\left(\frac{3}{5}\right) - \frac{2}{5} \log\left(\frac{2}{5}\right) = 0.971$$

$$S_{\text{Rain}} = 0.971$$

④

$$\text{Entropy}_{\text{overcast}} \Rightarrow 4Y / 0N = 0^4$$

$$S_{\text{overcast}} = -4/4 \log 1/4 = 0$$

$$S_{\text{overcast}} = 0$$

Information

gain for outlook.

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

$$= \text{Entropy}(S) - \frac{5}{14} \text{Entropy}(S_{\text{sunny}}) - \frac{4}{14} \text{Entropy}(S_{\text{overcast}}) - \frac{5}{14} \text{Entropy}(S_{\text{rainy}})$$

$$= 0.94 - \frac{5}{14} (0.97) - \frac{4}{14} (0) - \frac{5}{14} (0.971)$$

Information

$$\text{Gain}_{\text{outlook}} =$$

$$0.2467$$

prob

② Attribute (Temp)

$$S = \text{Overall entropy} = \frac{9Y + 5N}{14} = 14 \text{ values}$$

$$S = \left(\frac{9}{14} \log_2 \frac{9}{14} \right) + \frac{5}{14} \log_2 \left(\frac{5}{14} \right) = 0.94 = \text{Temp}$$

$$S_{\text{Hot}} = 2N + 2Y = 4 \text{ values}$$

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

$$S_{\text{Hot}} = 1.0$$

B (Mild)

$$S_{\text{mild}} = 4Y + 2N = 6 \text{ values}$$

$$S_{\text{mild}} = -\left(\frac{4}{6} \log_2 \left(\frac{4}{6} \right) \right) - \frac{2}{6} \left(\log_2 \left(\frac{2}{6} \right) \right)$$

$$S_{\text{mild}} = 0.9183$$

$$S_{cool} = 3Y + 1N = 4 \text{ bits}$$

$$S_{cool} = -\frac{3}{4} \log_2 \left(\frac{3}{4}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right) = 0.8113$$

$$S_{cool} = 0.8113$$

$$\text{Gain}_{\text{Temp}} = \text{Entropy}(S) - \sum_{v \in \text{temp}(H, N)} \left(\frac{P_{Sv}}{S} \text{ entropy}(S_v) \right)$$

$$\text{Gain}_{\text{Temp}} = 0.84 - \frac{4}{14} \times 1 - \frac{6}{14} \times 0.9182 - \frac{1}{14} \times 0.8113$$

$$G_{\text{Temp}} = 0.0289$$

③ Attribute (Humidity)

$$3Y + 5N =$$

$$\Sigma(S) = \cancel{\frac{9}{14} \log \frac{9}{14}} - \frac{9}{14} \log \frac{9}{14} - \frac{5}{14} \log \frac{5}{14}$$

$$\Sigma(S) = 0.94$$

$$\Sigma_{\text{High}} = 3Y + 2N = 7N$$

$$= \left(\frac{3}{7} \log \frac{3}{7} \right) - \left(\frac{4}{7} \log \frac{4}{7} \right) = 0.9852$$

$$S_{\text{High}} = 0.9852$$

$$S_{\text{Normal}} = 6Y + 1N = 7N$$

$$= -\frac{6}{7} \log \frac{6}{7} - \frac{1}{7} \log \frac{1}{7} = 0.5916$$

$$S_N = 0.5916$$

$$\text{Inf. Grav.}_{\text{Homody}} = S - \sum_{v \in (H, N)} \frac{|S_v|}{S} \text{ Entropy}_v$$

$$= 0.94 - \frac{7}{14} \times 0.98 - \frac{7}{14} \times 0.5918$$

If $\text{grav.}_{\text{Homody}} = 0.1516$ \Rightarrow Homodyne Inf.

$$0.1516 = \text{Homodyne Inf.}$$

Wind Attribute
 $3Y + 5N = 14 \text{ values}$

$$S = -\frac{9}{14} \log \frac{9}{14} - \frac{5}{14} \log \left(\frac{5}{14} \right) = 0.94$$

$$S_{\text{weak}} = 6Y + 2N = 8 \text{ values}$$

$$S_{\text{weak}} = -\frac{6}{8} \log \frac{6}{8} - \frac{2}{8} \log \frac{2}{8} = 0.8113 = S_{\text{weak}}$$

$$S_{\text{strong}} = 3Y + 3N = 6 \text{ values}$$

$$= -\frac{3}{6} \log \left(\frac{3}{6} \right) - \frac{3}{6} \log \left(\frac{3}{6} \right) = 0.5 + 0.5$$

$\text{Strong} = 0.1$

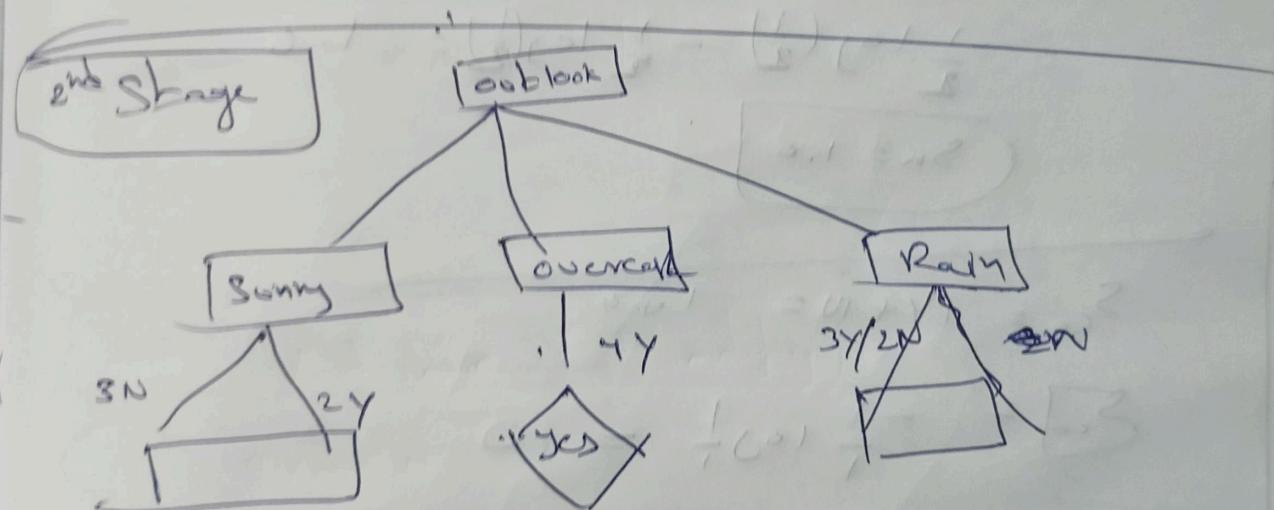
$$\text{Inf. Grav.} = 0.94 - \frac{6}{14} \times 1.0 - \frac{8}{14} \times 0.8113 = 0.078$$

$$If \text{ outlook} = 0.246 \quad (\text{Higher than others})$$

$$If \text{ temp} > 0.0289$$

$$If \text{ Humid} = 0.1516$$

$$If \text{ Wind} = 0.0478$$



2nd stage

Sunny Sunny table

	Temp	Humidity	wind	Y/N
1	H	H	W	N
2	M	M	S	N
3	C	M	S	Y
4	M	M	S	Y

3rd stage Rain. table

Rain

	Temp	Humidity	wind	Y/N

$$S_{\text{sunny}} = \frac{3Y+2N}{5} = \frac{3Y}{5} = \underline{\text{sunny}}$$

$$= -\frac{2}{5} \log\left(\frac{2}{5}\right) - \frac{3}{5} \log\left(\frac{3}{5}\right) = 0.97$$

$$S_{\text{temp}}(H, M, C)$$

$$= 0Y + 2N = 2N \text{ bits}$$

$$S_H = -\frac{2}{2} \log\left(\frac{2}{2}\right) = 0 = \boxed{S_H = 0}$$

$$S_M = 1N + 1Y = 2 \text{ bits}$$

$$= -\frac{1}{2} \log\left(\frac{1}{2}\right) - \frac{1}{2} \log\left(\frac{1}{2}\right) = 1.0$$

$$\boxed{S_N = 1.0}$$

$$S_{Co} = 1Y + 1N = 1 \text{ bit}$$

$$\boxed{S_{Co} = -\frac{1}{1} \log\frac{1}{1} = 0}$$

$$\text{If } I_f^{\text{temp}} = 0.97 - \frac{2}{5} \times 0 - \frac{2}{5} \times 1 = \frac{1}{5} = 0.20$$

$$\boxed{I_f^{\text{temp}} = 0.570}$$

Humidity \leftarrow sunny

$$\cancel{3+2} \quad S_{\text{vday}} = 3N + 2Y$$

$$S_{\text{A}} = 3Y + 2N = 0.97 \quad S_{\text{Total (H)}} = 0.97$$

$$S_H = 3 = 3N + 0Y = 3N$$

$$S_H = -\frac{3}{3} \log\left(\frac{3}{3}\right) = 0 = S_H$$

$$S_N = 82Y + 6N =$$

$$S_N = -\frac{2}{2} \log\left(\frac{2}{2}\right) = 0$$

$$S_N = 0$$

$$S_{\text{Humid}} = Tf = S_H - \frac{3}{3} \times 0 - \frac{2}{3} \times 0$$

$$Tf = 0.97$$

③

wild \leftarrow sunny

$$S_w = 3Y + 2N \approx 0.97$$

$$S_w = 2N + 1Y = 3$$

$$S_w = -\frac{3}{3} \log\left(\frac{1}{3}\right) - \left(\frac{2}{3}\right) \log\left(\frac{2}{3}\right) = 0.9183$$

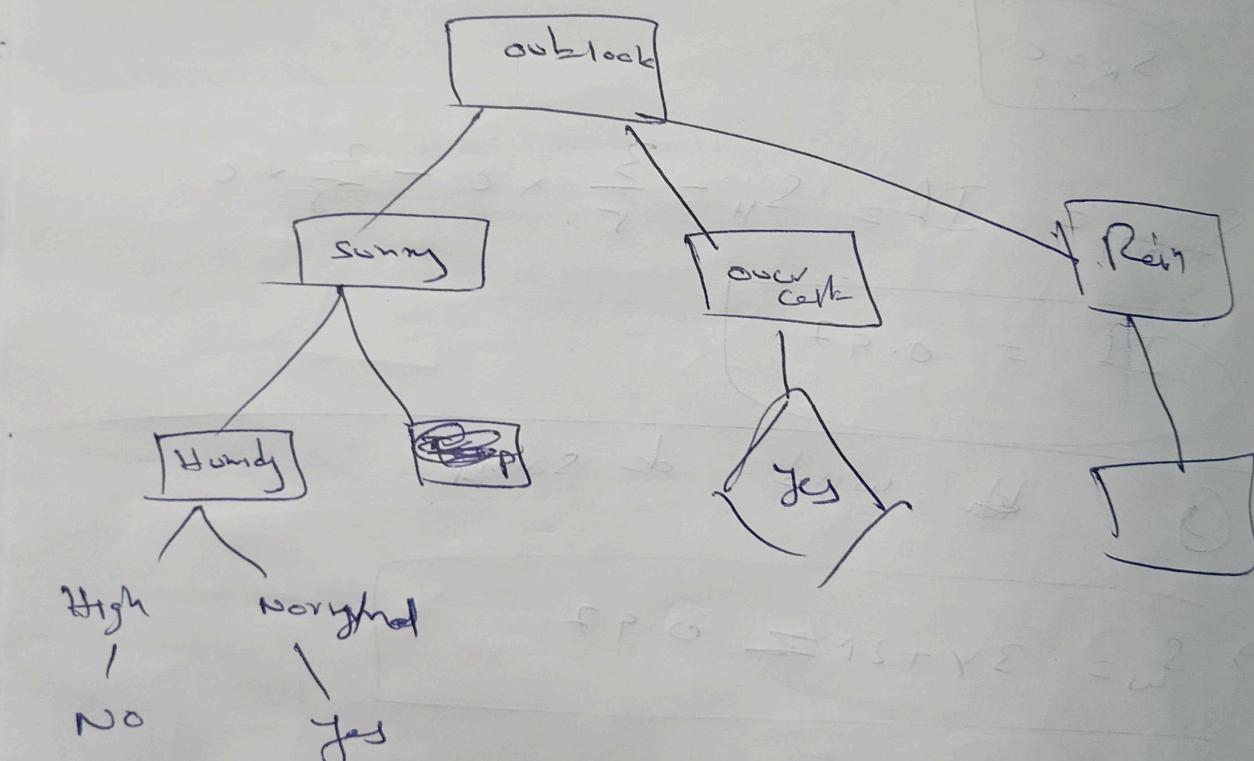
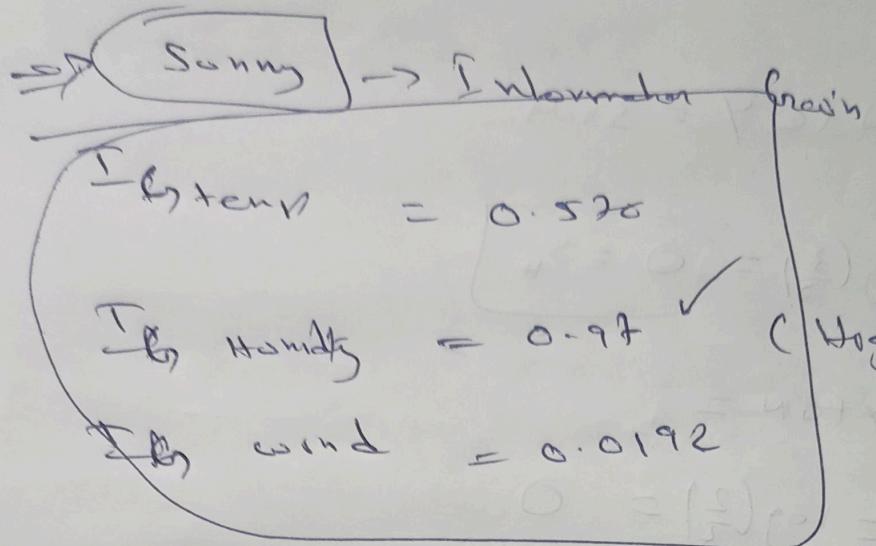
$$S_{\text{stray}} = -\frac{1}{2} \log\left(\frac{1}{2}\right) - \frac{1}{2} \log\left(\frac{1}{2}\right) = 1$$

$$S_s = 1$$

$$S_w = 0.9183$$

$$I_B \text{ wind} = 0.97 - \frac{3}{5} \times 0.98 = 2 \times 0.98 \times 0.1$$

$$= 0.0192$$



$$E(I^P, O) = \left(\frac{1}{2}\right) \text{col} \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) \text{col} \frac{1}{3} = \frac{1}{2}$$

$$1 - \left(\frac{1}{2}\right) \text{col} \frac{1}{2} - \left(\frac{1}{2}\right) \text{col} \frac{1}{3} = \frac{1}{2}$$

$E(I^P, O) = \frac{1}{2}$

$1 - \frac{1}{2}$

attribute Realy Information Gain

$$P_{\text{rain}} \text{ Entropy } (S) = P_{\text{rain}} = 3y + 2N = 5 \quad | \text{Temp}$$

$$\rightarrow \text{entropy } (S) = -\frac{3}{5} \log\left(\frac{3}{5}\right) - \frac{2}{5} \log\left(\frac{2}{5}\right) = 0.97$$

S_{Temp}

$$\cancel{\text{entropy}}_{\text{Temp}} =$$

$$S_{\text{Temp}} = 3y + 2N =$$

$$S_{\text{mild}} = -\frac{2}{3} \log\left(\frac{2}{3}\right) - \frac{1}{3} \log\left(\frac{1}{3}\right) = 0.9183$$

$$S_{\text{Hot}} = -0 - 0 = 0$$

$$S_{\text{cold}} = -\frac{1}{2} \log\left(\frac{1}{2}\right) - \frac{1}{2} \log\left(\frac{1}{2}\right) = 1$$

$$\text{Information} = S_{\text{rain}} - \sum_{v \in \{\text{mild, Hot, cold}\}} \frac{S_v}{S}$$

$$= 0.97 - \frac{2}{5} S_{(H)} - \frac{3}{5} (0.9183) - \frac{2}{5} (1)$$

$$I_{\text{Temp}} = 0.0192$$

Rain - Humidity

$$S_w = -\frac{3}{5} \log\left(\frac{3}{5}\right) - \frac{2}{5} \log\left(\frac{2}{5}\right) = 0.97$$

$$S_{Hyp} = -\frac{1}{2} \log\left(\frac{1}{2}\right) - \frac{1}{2} \log\left(\frac{1}{2}\right) = 1$$

$$S_{Norm} = -\frac{2}{3} \log\left(\frac{2}{3}\right) - \frac{1}{3} \log\left(\frac{1}{3}\right) = 0.918$$

$$I_G_{Humidity} = 0.97 - \frac{2}{5}(1) - \frac{3}{5}0.918 = 0.0192$$

$$I_G_{Hyp} = 0.0192$$

Rain - wind

$$S_{wind} = -\frac{3}{5} \log\left(\frac{3}{5}\right) - \frac{2}{5} \log\left(\frac{2}{5}\right) = 0.97$$

$$S_{cat} = -\frac{3}{5} \log\left(\frac{3}{5}\right) = 0$$

$$S_{shoy} = -\frac{2}{2} \log\left(\frac{2}{2}\right) = 0$$

$$I_B = 0.97 - 0 - 0 = 0.97$$

Rain - wind

I_G

$$\sum G_{\text{temp}} = 0.0192$$

$$I_G \text{ Humidity} = 0.0192$$

$$I_G \text{ wind} = 0.0192 \quad \checkmark$$

