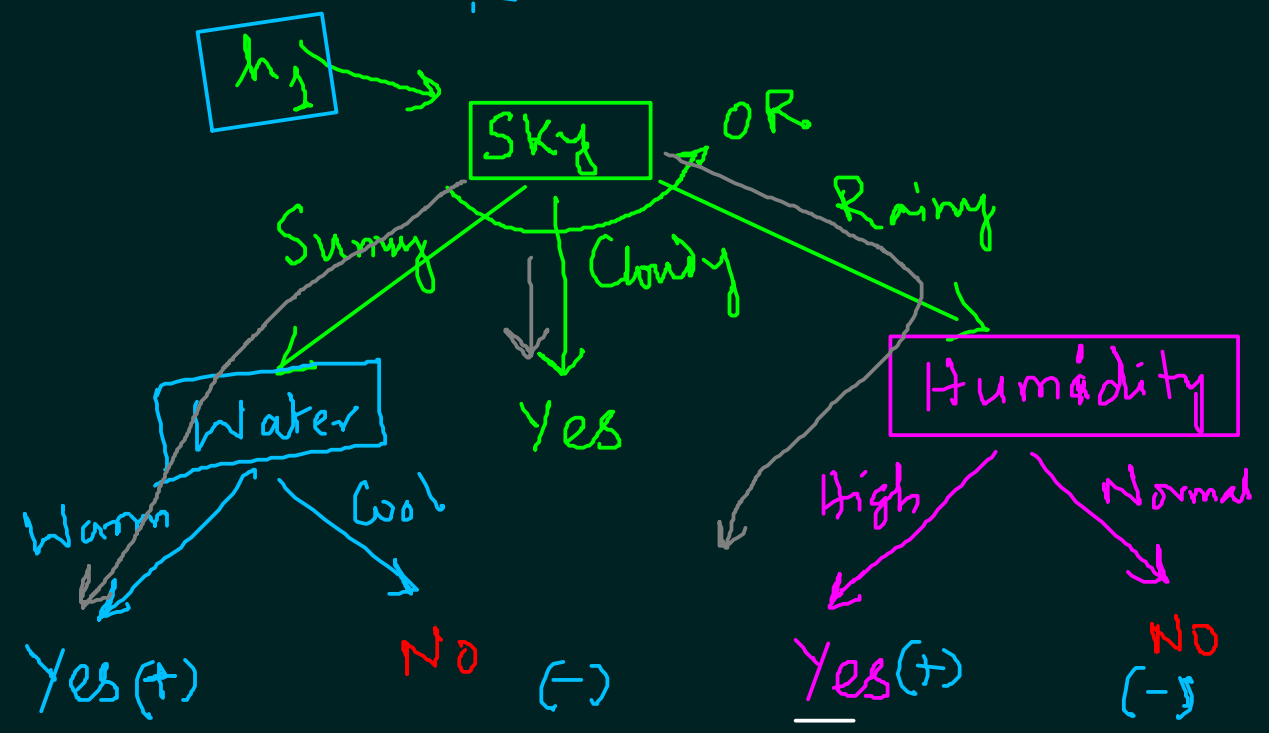
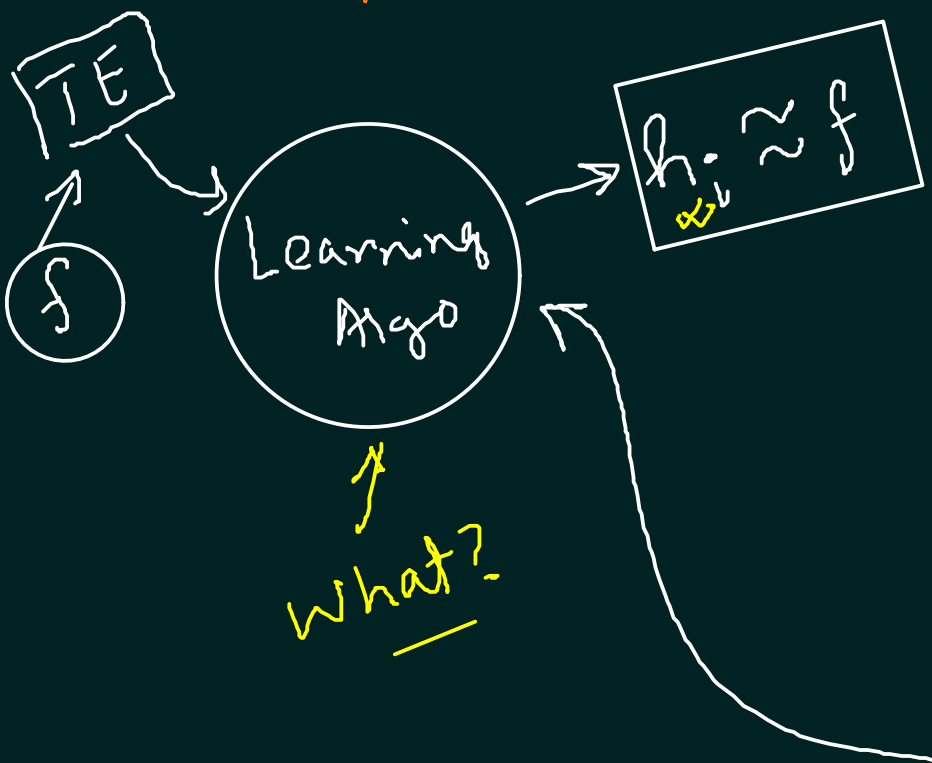


Boolean AND formulas $h_1 = (Sky = Sunny \wedge Water = Warm) \vee (Sky = Cloudy) \vee (Sky = Raining \wedge Humidity = High)$

$\mathcal{H} = \{ (A) \vee (B) \}$

— Can/How we learn?



Decision Tree Learning

$\mathcal{H} = \{ \text{Decision Tree} \}$

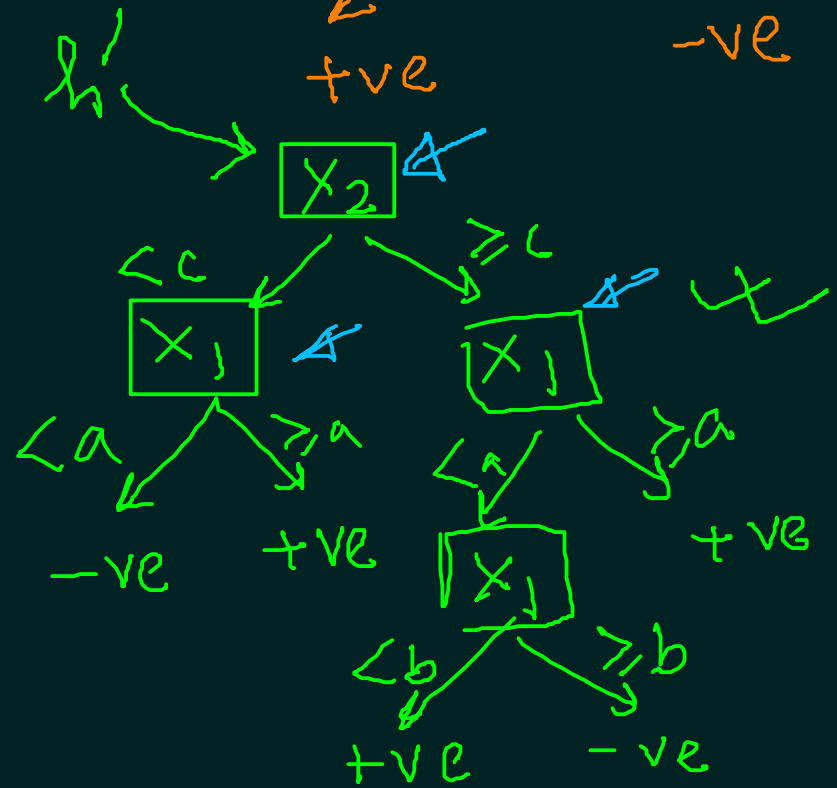
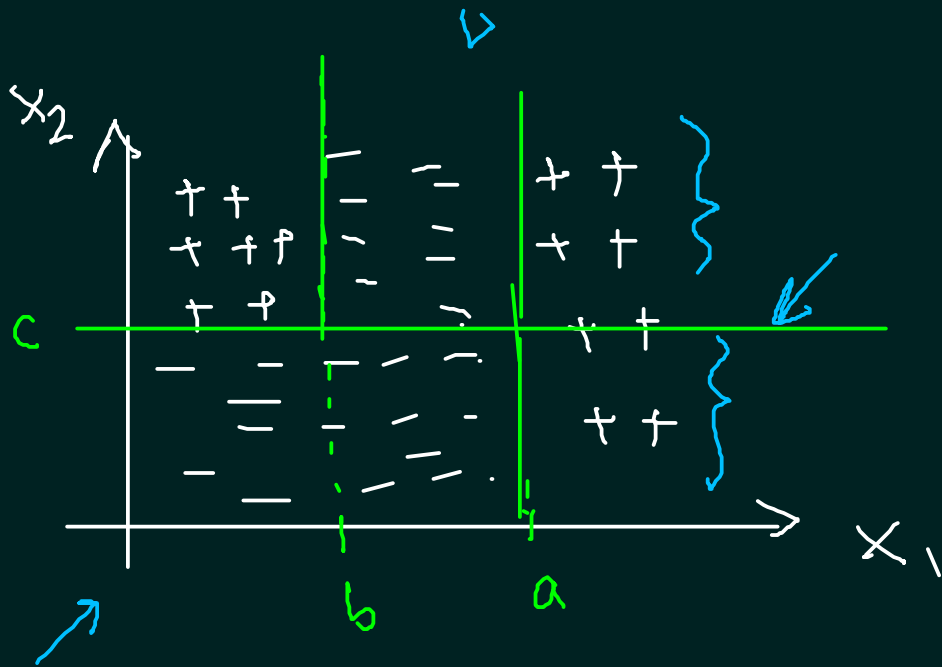
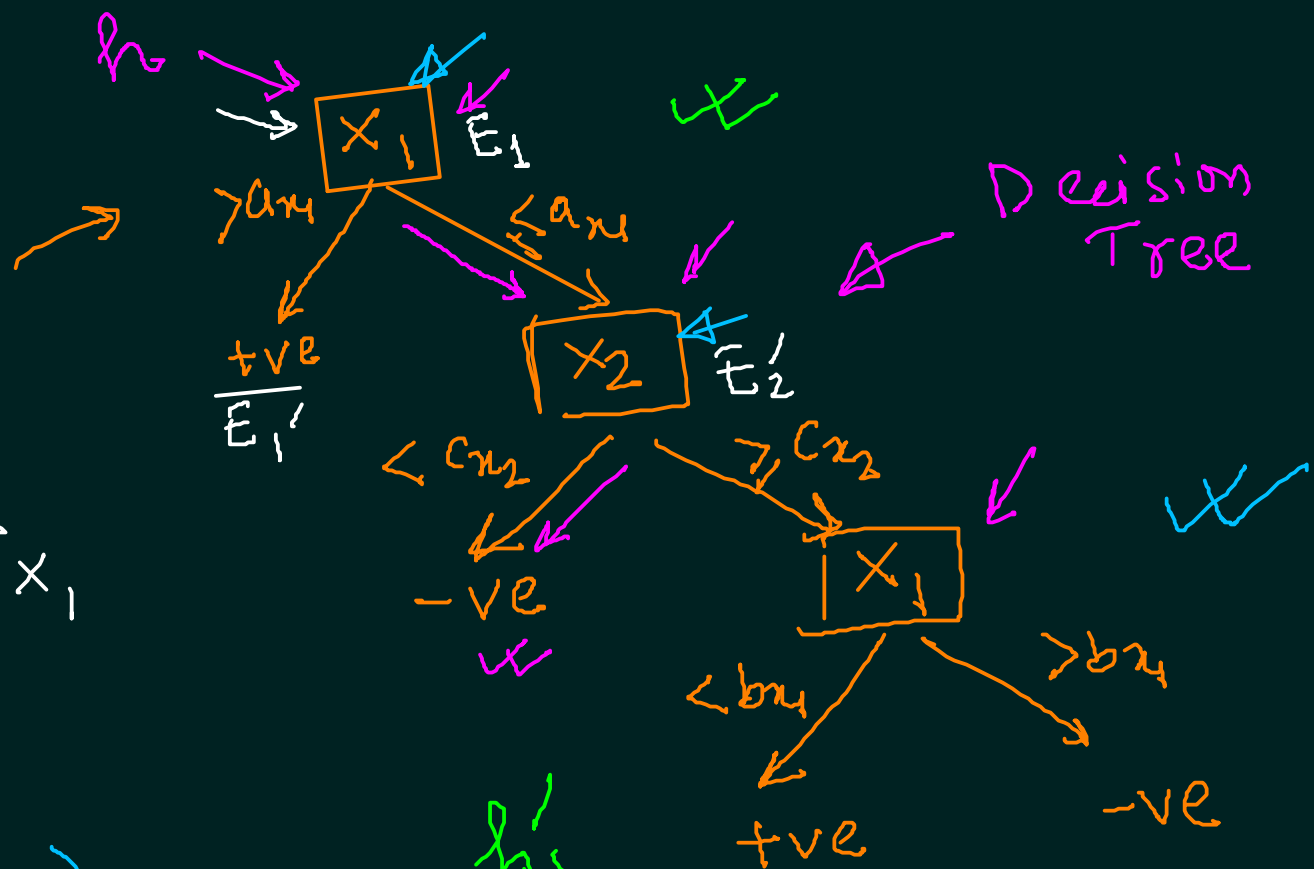
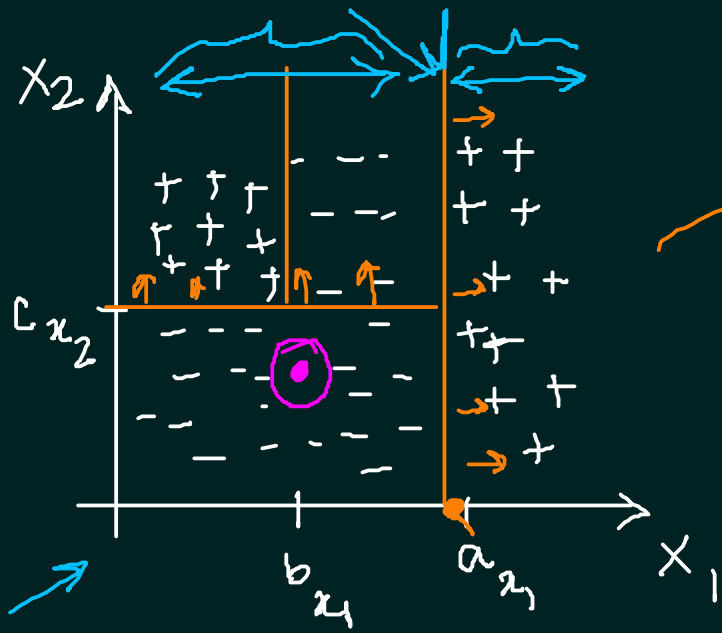
$(A \wedge B) \vee (C \wedge D) \vee (\neg A \wedge C \wedge E)$

$\{v_1, v_2, v_3\}$

$a \neq v_1$

TError

Attr. value miss



⇒ Which attribute should we choose first/next?

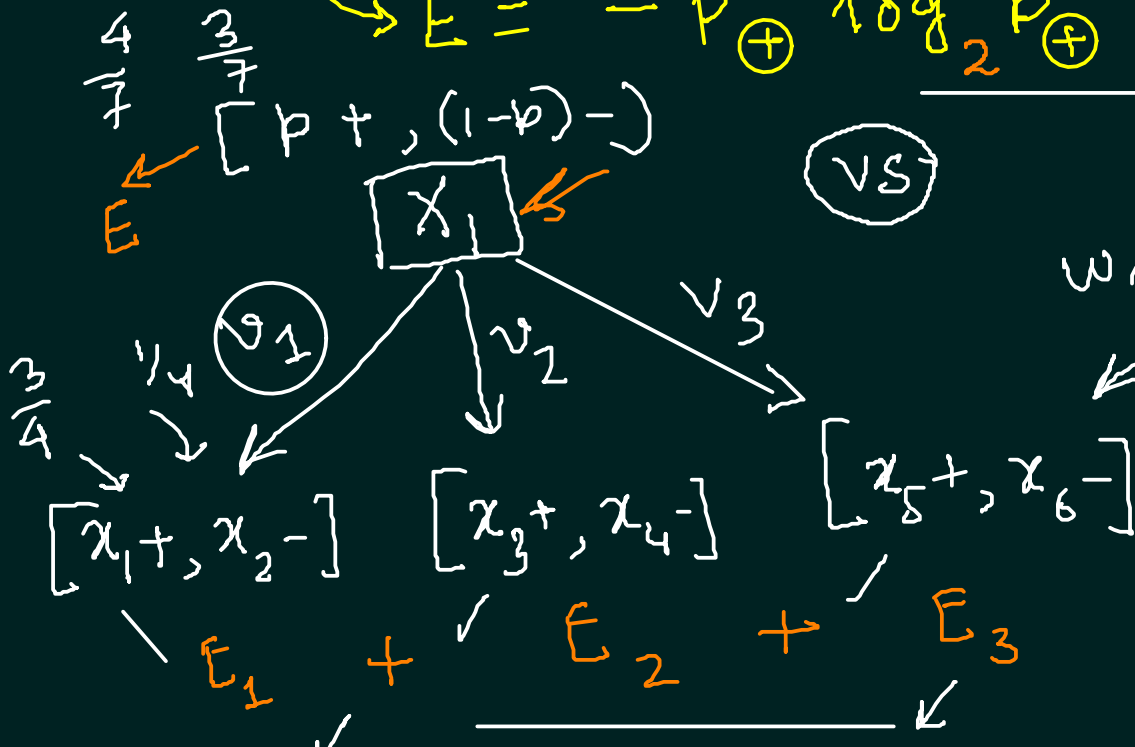
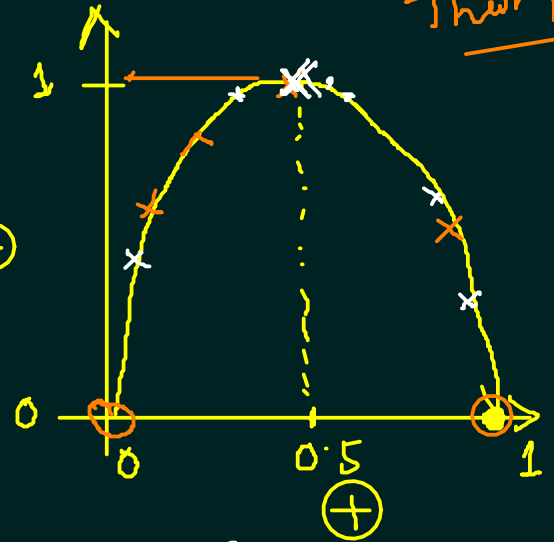
Inf. Theory

Entropy

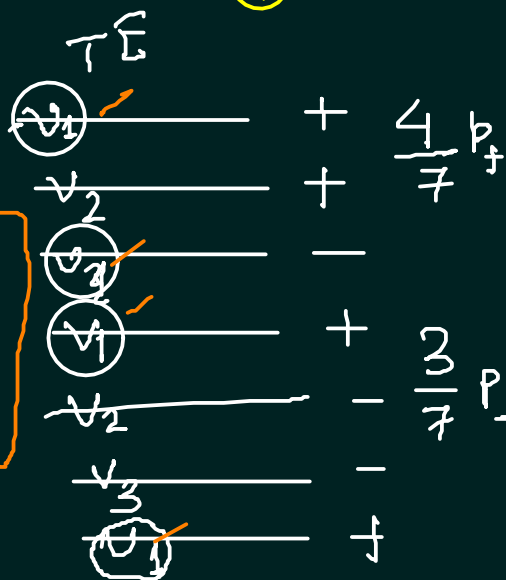
p_{\oplus}

p_{\ominus}

$$E = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$



$$E = -\sum_{c=1}^m p_c \log_2 p_c$$



$$IG = E - \left(\frac{4}{7} E_1 + \frac{2}{7} E_2 + \frac{1}{7} E_3 \right)$$

↑ low

$$IG(S, A) = E(S) - \sum_{v=1}^{|A|} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Wind = { Weak, Strong }

$$S_{W=Weak} = [6+, 2-]$$

$$S_{W=Strong} = [3+, 3-]$$

IG (S, Wind)

$$IG = E(S) - \left[\frac{8}{14} E(S_{Weak}) + \frac{6}{14} E(S_{Strong}) \right]$$

$$= 0.94 - \left[\frac{8}{14} \times 0.811 + \frac{6}{14} \times 1 \right]$$

$$= 0.048$$

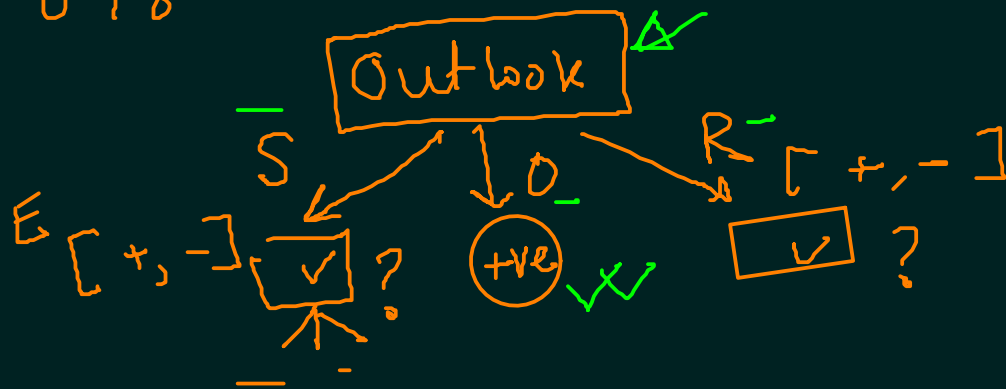
Day	Outlook	Temp	Humidity	Wind	PlayTennis?
D1	Sunny	Hot	High	Weak ✓	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak ✓	Yes
D4	Rain	Mild	High	Weak ✓	Yes
D5	Rain	Cool	Normal	Weak ✓	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak ✓	No
D9	Sunny	Cool	Normal	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

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

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

$$= 0.94$$

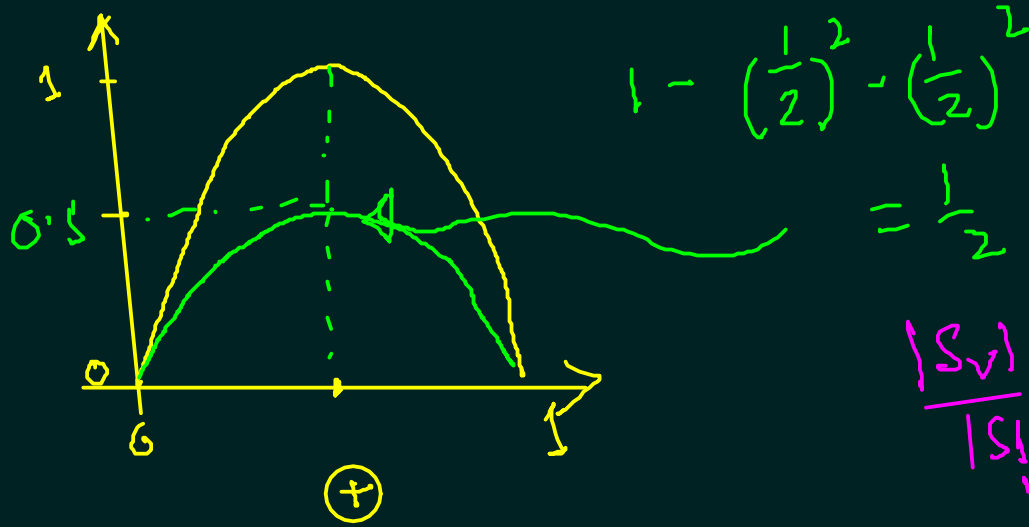
IG



Information Gain (Entropy)

Gini Impurity:

$$Imp = 1 - \sum_{i=1}^3 p_i^2$$



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

$$\frac{|S_v|}{|S|}$$

$$Inf. Gain = Imp(S) - \sum [f_i Imp(S_v)]$$

Gini Index

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

Computationally Easier!!

H/W: Decision Tree Model

Puzzle → CElim - X
 → ID3 - ✓