

Candidate Elimination Technique

Training Data (Hypothesis Table)

	Citation	Size	In Library	Prize	Edition	Buy
1	Same	Small	No	Affordable	Many	No
2	Many	Big	No	Expensive	One	Yes
3	Same	Big	Always	Expensive	Few	No
4	Many	Medium	No	Expensive	Many	Yes
5	Many	Small	No	Affordable	Many	Yes

ANSWER:

Specific (S): Most specific hypothesis

General (G): Most general hypothesis

S0: { \emptyset , \emptyset , \emptyset , \emptyset , \emptyset }

S1: { \emptyset , \emptyset , \emptyset , \emptyset , \emptyset }

S2: {Many, Big, No, Expensive, one}

S3: {Many, ?, No, Expensive, ?}

S4: {Many, ?, No, Expensive, ?}

S5: {Many, ?, No, ?, ?}

G5: {<Many, ?, ?, ?, ?>}

G4: {<Many, ?, ?, ?, ?>}

G3: {<Many, ?, ?, ?, ?>, <?, ?, ?, ?, one>}

G2: {<Many, ?, ?, ?, ?>, <?, Big, ?, ?, ?>, <?, ?, ?, expensive, ?>, <?, ?, ?, ?, one>}

G1: { <Many, ?, ?, ?, ?>, <?, Medium, ?, ?, ?>, <?, Big, ?, ?, ?>, <?, ?, always ?, ?>, <?, ?, ?, expensive, ?>, <?, ?, ?, ?, one>, <?, ?, ?, ?, few>}

G0: { ?, ?, ?, ?, ? }

VERSION SPACE: {Many, ?, No, ?, ?}, {<Many, ?, ?, ?, ?>}

Decision Tree Technique

Play Tennis Dataset

Day	Outlook	Temperature	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
D15	Sunny	Hot	Normal	Strong	No

ANSWER

$$\begin{aligned}\text{Entropy} &= -9/15 \log (9/15) - 6/15 \log (6/15) \\ &= 0.970951\end{aligned}$$

Attribute= Outlook

Values(outlook) = {Sunny, Overcast, Rain}

$$S(\text{Sunny}) = \{2+, 4-\}$$

$$\text{Entropy}(S(\text{Sunny})) = -2/6 \log (2/6) - 4/6 \log (4/6) = 0.918296$$

$$S(\text{Overcast}) = \{4+, 0-\}$$

$$\text{Entropy}(S(\text{Overcast})) = -4/6 \log (4/6) - 0 \log 0 = 0.0$$

$$S(\text{Rain}) = \{3+, 2-\}$$

$$\text{Entropy}(S(\text{Rain})) = -3/5 \log (3/5) - 2/5 \log (2/5) = 0.971$$

$$\begin{aligned}\text{Information Gain}(S(\text{outlook})) &= 0.970951 - (6/15 * 0.918296 + 4/15 * 0 + 5/15 * 0.971) \\ &= 0.279966\end{aligned}$$

Attribute = Temperature

Values(temperature) = {Hot, Cool, Mild}

$$S(\text{Hot}) = \{2+, 3-\}$$

$$\text{Entropy}(S(\text{Hot})) = -2/5 \log (2/5) - 3/5 \log (3/5) = 0.971$$

$$S(\text{Mild}) = \{4+, 2-\}$$

$$\text{Entropy}(S(\text{Mild})) = -4/6 \log (4/6) - 2/6 \log (2/6) = 0.9183$$

$$S(\text{Cool}) = \{3+, 1-\}$$

$$\text{Entropy}(S(\text{Cool})) = -3/4 \log (3/4) - 1/4 \log (1/4) = 0.8113$$

$$\begin{aligned}\text{Information Gain}(S(\text{Temperature})) &= 0.970951 - \{5/15 * 0.971 + 6/15 * 0.9183 + 4/15 * 0.8113\} \\ &= 0.06362\end{aligned}$$

Attribute = Humidity

Values (humidity) = {High, Normal}

$S(\text{high}) = \{3+, 4-\}$

$\text{Entropy}(S(\text{high})) = -3/7 \log(3/7) - 4/7 \log(4/7) = 0.9852$

$S(\text{Normal}) = \{6+, 2-\}$

$\text{Entropy}(S(\text{Normal})) = -6/8 \log(6/8) - 2/8 \log(2/8) = 0.8112784$

$\text{Information Gain (Humidity)} = 0.970951 - \{7/15 * 0.9852 + 8/15 * 0.8112784\} = 0.0785$

Attribute = Wind

Values (Wind) = {Weak, Strong}

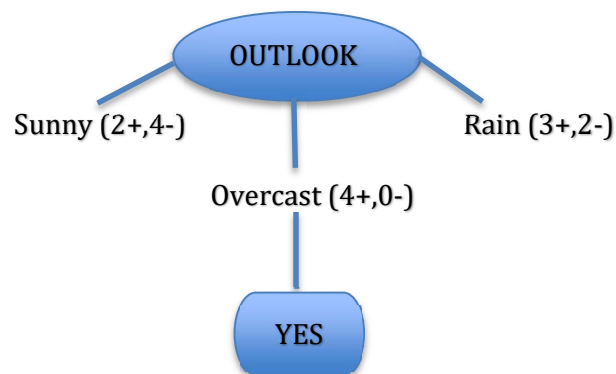
$S(\text{Strong}) = \{3+, 4-\}$

$E(S(\text{Strong})) = -3/7 \log(3/7) - 4/7 \log(4/7) = 0.9852$

$S(\text{Weak}) = \{6+, 2-\}$

$E(S(\text{Weak})) = -6/8 \log(6/8) - 2/8 \log(2/8) = 0.8113$

$\text{Information Gain (wind)} = 0.970951 - \{7/15 * 0.9852 + 8/15 * 0.8113\} = 0.0785$



Dataset (Sunny)

Day	Temperature	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
D15	Hot	Normal	Strong	No

$$\text{Entropy}(D(\text{Sunny})) = -2/6 \log(2/6) - 4/6 \log(4/6) = 0.9183$$

Attribute= Temperature

$$D(\text{Hot}) = \{0+, 3-\}$$

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

$$D(\text{Mild}) = \{1+, 1-\}$$

$$\text{Entropy}(S(\text{Mild})) = 1.0$$

$$D(\text{cool}) = \{1+, 0-\}$$

$$\text{Entropy}(S(\text{cool})) = 0.0$$

$$\text{Information Gain (Temperature)} = 0.9183 - (3/6 * 0 + 2/6 * 1 + 1/6 * 0) = 0.585$$

Attribute= Humidity

$$D(\text{High}) = \{0+, 3-\}$$

$$\text{Entropy}(D(\text{High})) = 0.0$$

$$D(\text{Normal}) = \{2+, 1-\}$$

$$\text{Entropy}(D(\text{Normal})) = -2/3 \log(2/3) - 1/3 \log(1/3) = 0.9183$$

$$\text{Information Gain (Humidity)} = 0.9183 - \{3/6 * 0 + 3/6 * 0.9183\} = 0.459$$

Attribute = Wind

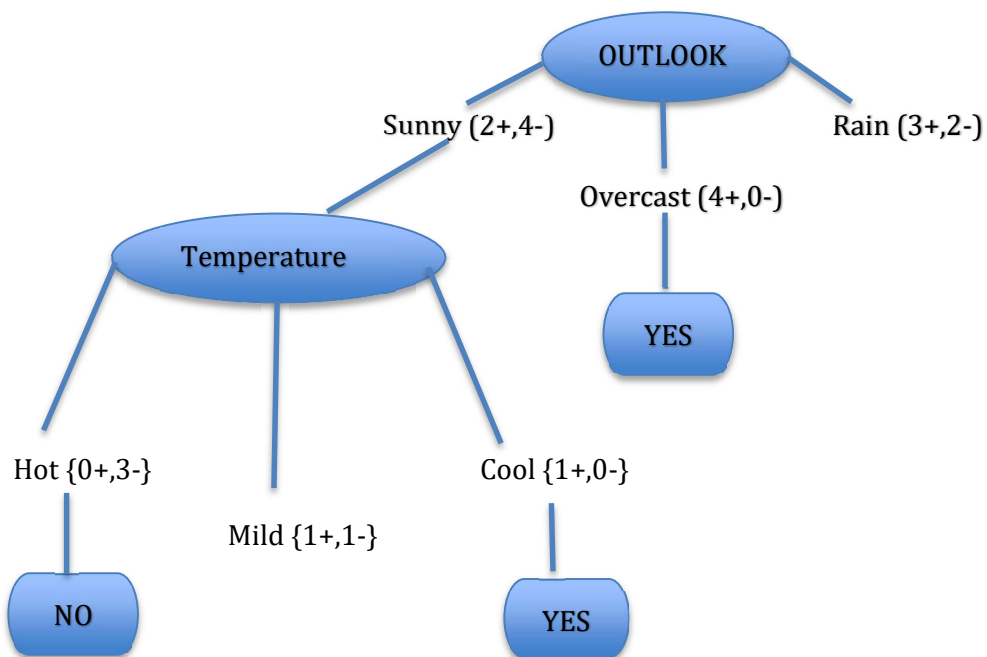
$D(\text{Strong}) = \{1+, 2-\}$

$\text{Entropy}(D(\text{Strong})) = -1/3 \log(1/3) - 2/3 \log(2/3) = 0.9183$

$D(\text{weak}) = \{1+, 2-\}$

$\text{Entropy}(D(\text{weak})) = -1/3 \log(1/3) - 2/3 \log(2/3) = 0.9183$

$\text{Information Gain}(\text{Wind}) = 0.9183 - (3/6 * 0.9183 + 3/6 * 0.9183) = 0.0$



Dataset 3 (Sunny-> Mild)

Day	Humidity	Wind	Play Tennis
D8	High	Weak	No
D11	Normal	Strong	Yes

$$\text{Entropy} = 1.0$$

Attribute- Humidity

$$D(\text{High}) = \{0+, 1-\}$$

$$\text{Entropy}(D(\text{High})) = 0.0$$

$$D(\text{Normal}) = \{1+, 0-\}$$

$$\text{Entropy}(D(\text{Normal})) = 0.0$$

$$\text{Information Gain (Humidity)} = 1 - 0.0 \cdot 1/2 - 0.0 \cdot 1/2 = 1$$

Attribute- Wind

$$D(\text{Weak}) = \{0+, 1-\}$$

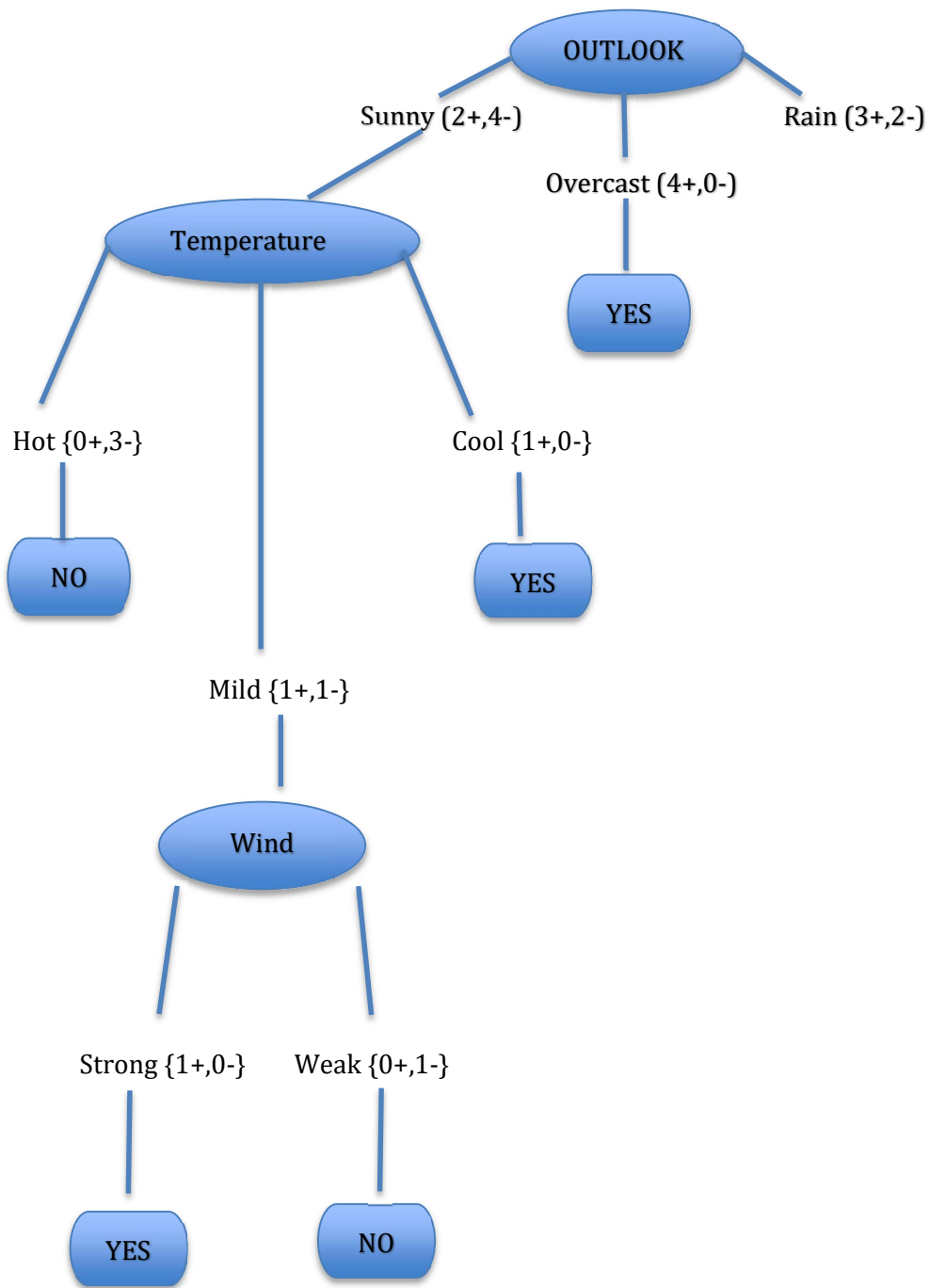
$$\text{Entropy}(D(\text{Weak})) = 0.0$$

$$D(\text{Strong}) = \{1+, 0-\}$$

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

$$\text{Information Gain (Wind)} = 1 - 0.0 \cdot 1/2 - 0.0 \cdot 1/2 = 1$$

Since Entropies are same choose any one. Let's choose Wind



Dataset 2

Day	Temperature	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

$$\text{Entropy} = -3/5 \log 3/5 - 2/5 \log 2/5 = 0.97$$

Attribute= Temperature

$$D(\text{Mild}) = \{2+, 1-\}$$

$$\text{Entropy}(D(\text{Mild})) = -2/3 \log 2/3 - 1/3 \log 1/3 = 0.9183$$

$$D(\text{Cool}) = \{1+, 1-\}$$

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

$$\text{Information Gain (Temperature)} = 0.97 - 3/5 * 0.9183 - 2/5 * 1 = 0.019$$

Attribute= Humidity

$$D(\text{High}) = \{1+, 1-\}$$

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

$$D(\text{Normal}) = \{2+, 1-\}$$

$$\text{Entropy}(D(\text{Normal})) = -2/3 \log 2/3 - 1/3 \log 1/3 = 0.9183$$

$$\text{Information Gain (Humidity)} = 0.97 - 1 * 2/5 - 0.9183 * 3/5 = 0.01902$$

Attribute= Wind

$$D(\text{Weak}) = \{3+, 0-\}$$

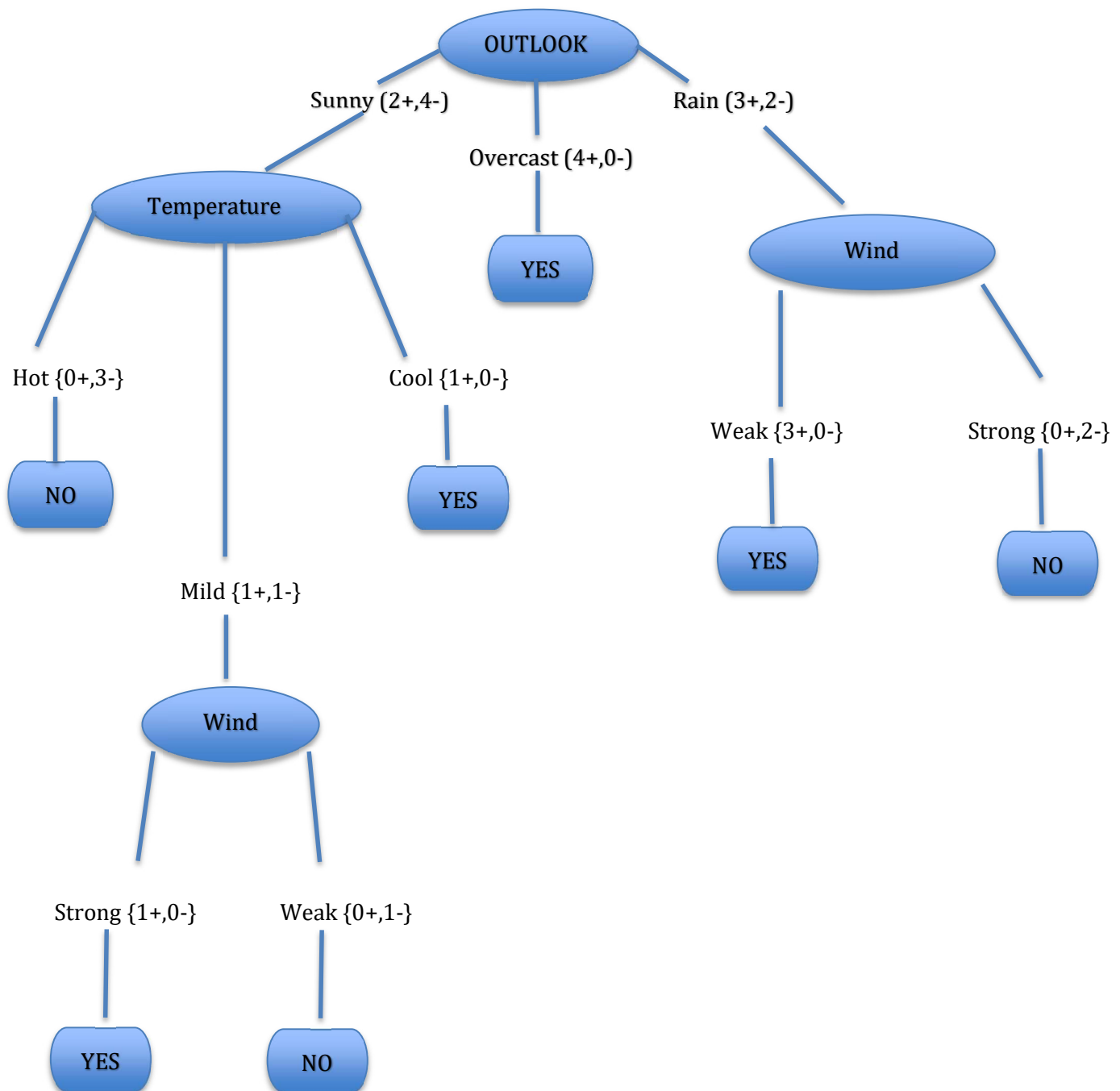
$$\text{Entropy}(D(\text{Weak})) = 0.0$$

$$D(\text{Strong}) = \{0+, 2-\}$$

Entropy(D(Strong)) = 0.0

Information Gain (Wind) = $0.97 - 0.0 * 3/5 - 0.0 * 2/5 = 0.97$

FINAL DECISION TREE



FINAL ANSWER: {<Sunny, Cool, ?, ?>, <Sunny, Mild, ?, Strong>, <Overcast, ?, ?, ?>, <Rain, ?, ?, Weak>}

Name: DARSHAN GOWDA M