CS 4104
APPLIED MACHINE LEARNING

Dr. Hashim Yasin

National University of Computer and Emerging Sciences,

Faisalabad, Pakistan.

DECISION TREE

Decision Tree

Problem Setting:

- Set of possible instances X
  - each instance x in X is a feature vector
  - e.g., <*Humidity=low, Wind=weak, Outlook=rain, Temp=hot*>
- Unknown target function $f: X \rightarrow Y$
  - Y is discrete valued
- Set of function hypotheses $H = \{h \mid h: X \rightarrow Y\}$
  - each hypothesis h is a decision tree
  - trees sorts x to leaf, which assigns y

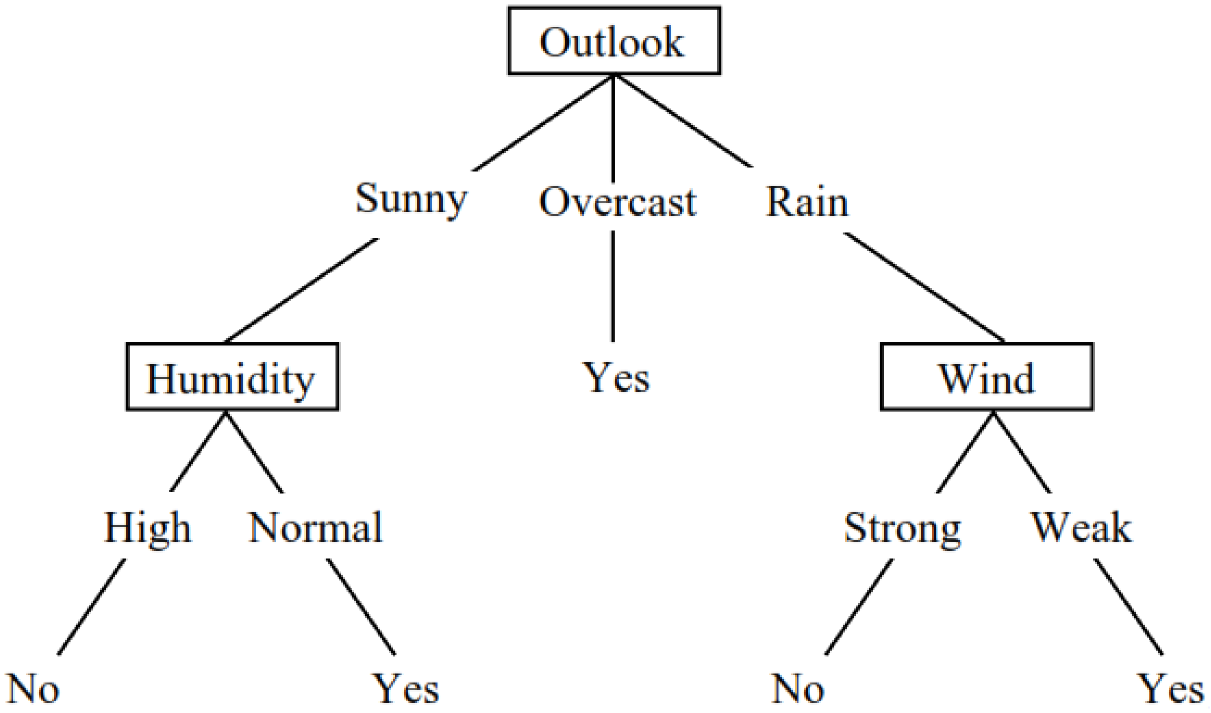
Decision Tree

	X				Y
Day	*Outlook*	*Temperature*	*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

Decision Tree

- A Decision tree for

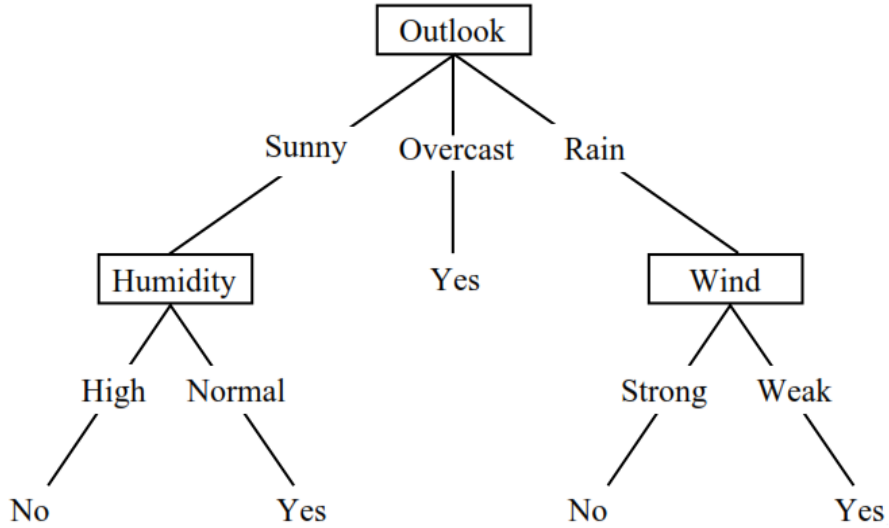
<Outlook, Temperature, Humidity, Wind> →PlayTennis?



Decision Tree

- A Decision tree for

  <Outlook, Temperature, Humidity, Wind> →PlayTennis?

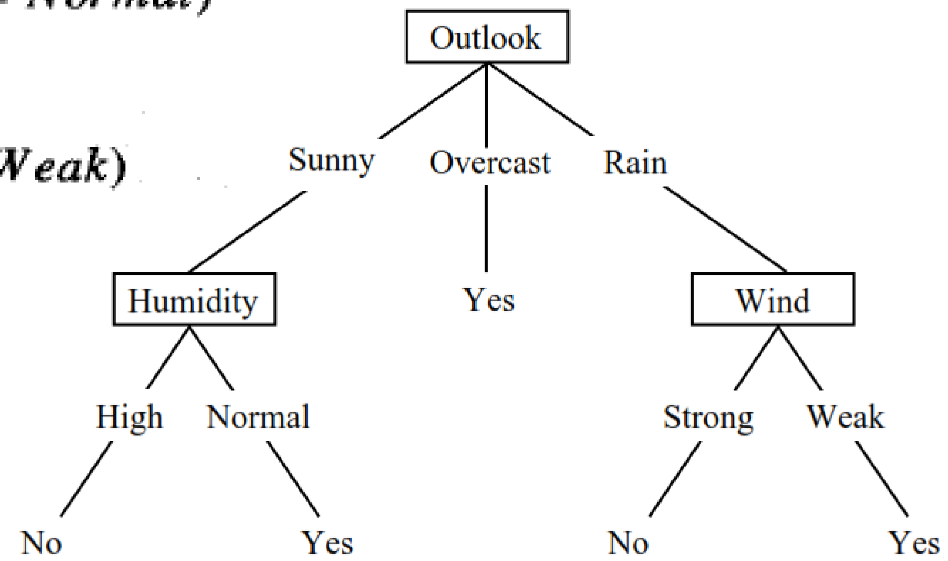


- **Each internal node:** test one attribute X_i
- **Each branch from a node:** selects one value for X_i
- **Each leaf node:** predict Y

Decision Tree

- In general, decision trees represent a disjunction of conjunctions of the attribute values,

$$
\begin{aligned}
& (Outlook = Sunny \ \wedge \ Humidity = Normal) \\
\vee \quad & (Outlook = Overcast) \\
\vee \quad & (Outlook = Rain \ \wedge \ Wind = Weak)
\end{aligned}
$$



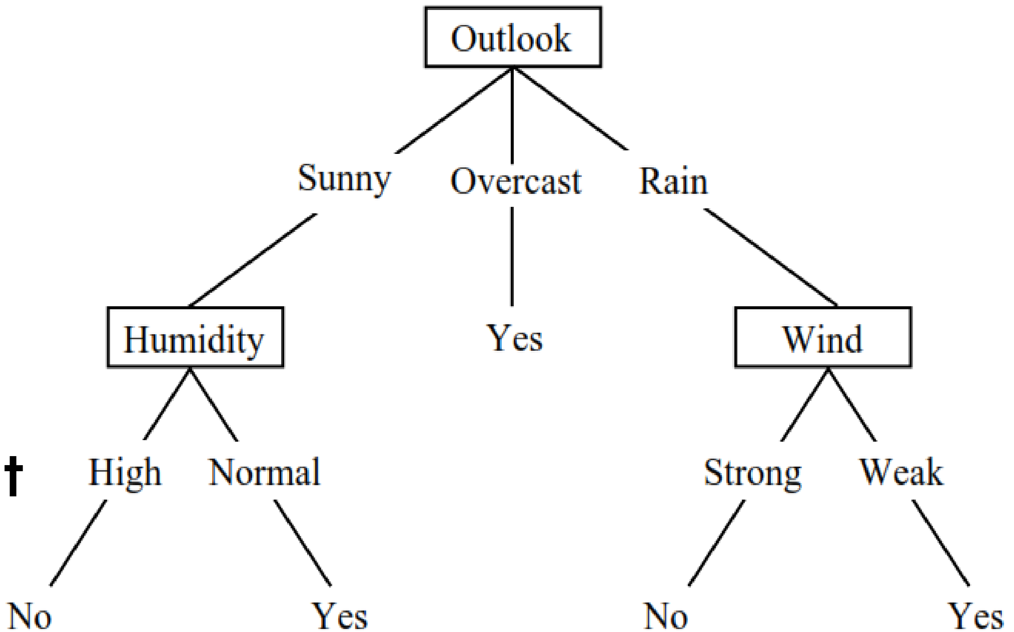
Decision Tree

Input:

- Training examples $\{x_i, y_i\}$ of unknown target function

Output:

- Hypothesis $h \in H$ that **best approximates** target function f

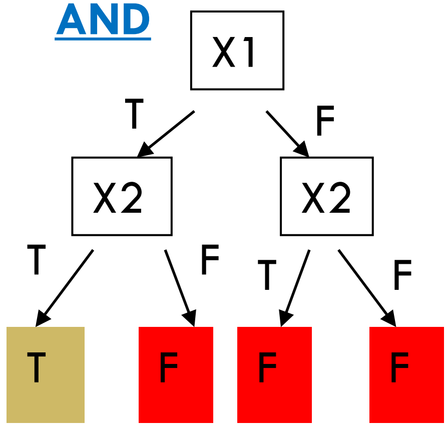


Decision Trees … Examples

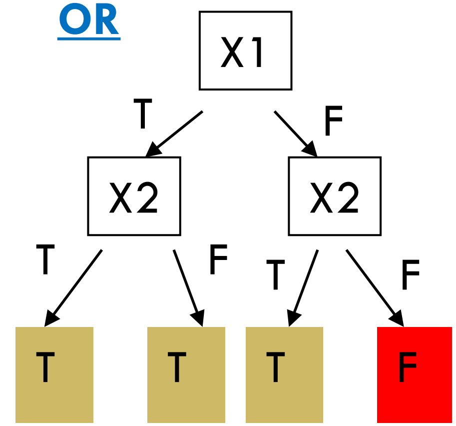
- Suppose $X = <X_1,\dots X_n>$, where X_i are Boolean variables
- How would you represent the followings:

$$Y = X_1 \wedge X_2 \qquad\qquad Y = X_1 \vee X_2$$

AND



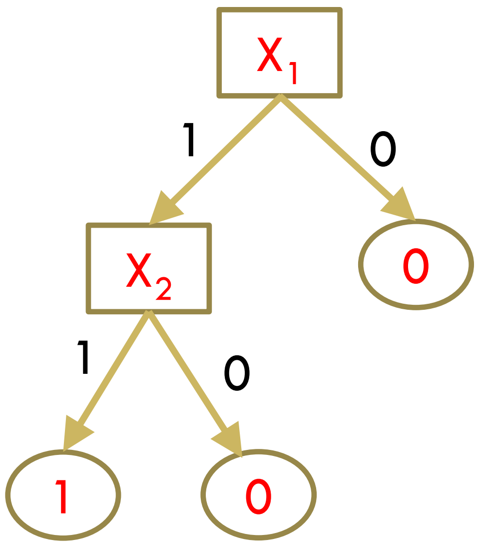
OR



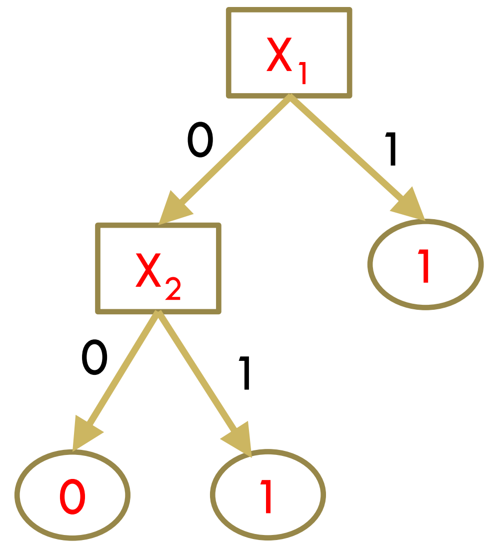
Decision Trees … Examples

- Suppose $X = \langle X_1, \ldots X_n \rangle$, where X_i are Boolean variables
- How would you represent the followings:

$$Y = X_1 \wedge X_2$$



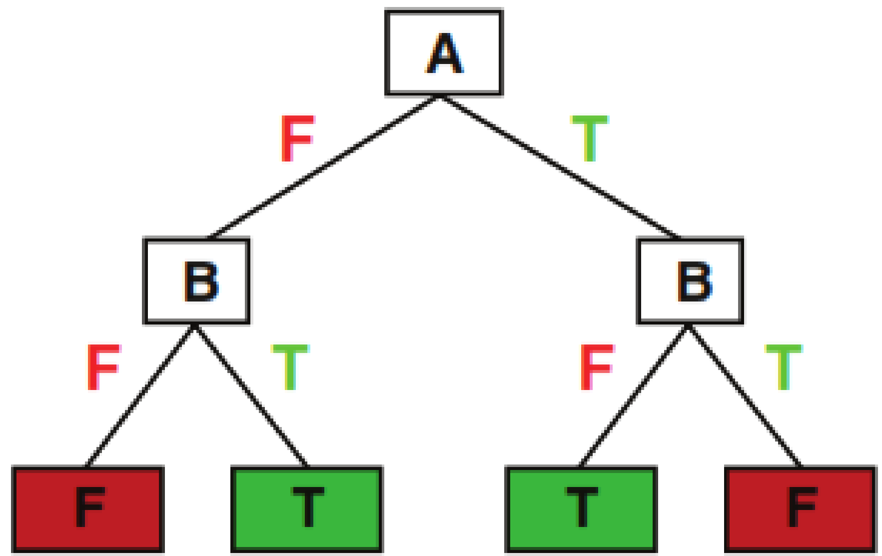
$$Y = X_1 \vee X_2$$



Decision Trees … Examples

- Suppose $X = \langle X_1, \ldots X_n \rangle$, where X_i are Boolean variables
- How would you represent the followings:

A	B	A xor B
F	F	F
F	T	T
T	F	T
T	T	F



Decision Tree Algorithm … ID3

Iterative Dichotomiser 3 (ID3)

ID3(Examples, Target_attribute, Attributes)

- Create a Root node for the tree
- **If all Examples are positive**, Return the single-node tree Root, with label = +
- **If all Examples are negative**, Return the single-node tree Root, with label = -
- **If Attributes is empty**, Return the single-node tree Root, with label = most common value of *Target_attribute* in Examples

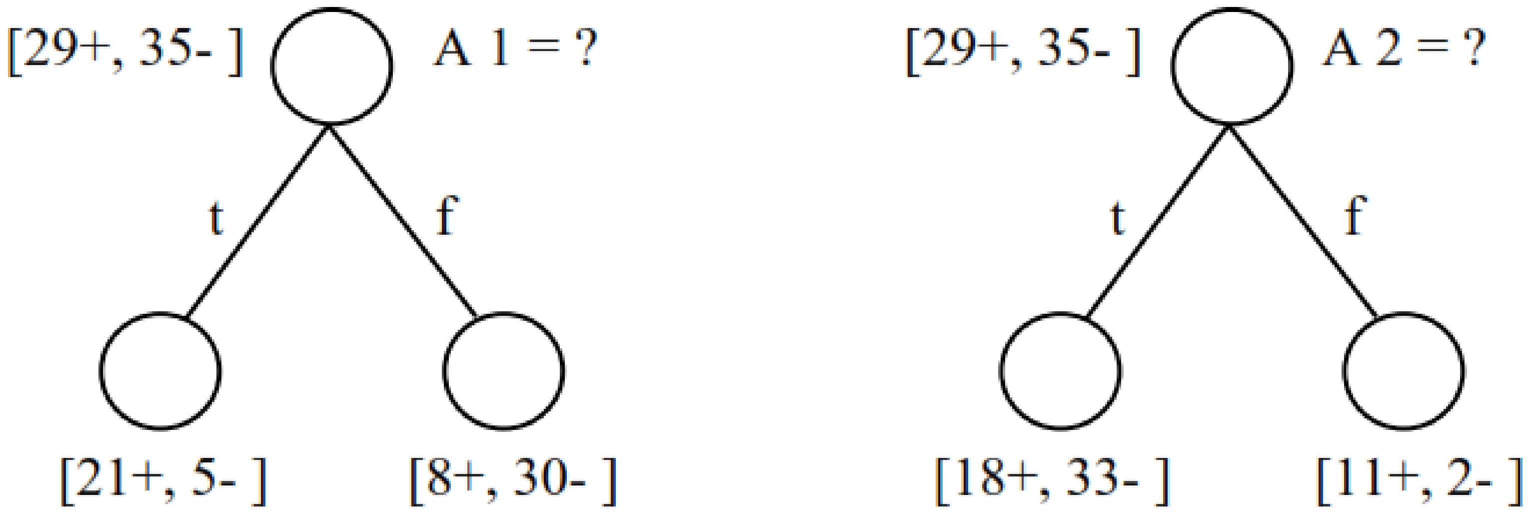
Decision Tree Algorithm … ID3

Otherwise Begin

- A ← the attribute from Attributes that **best*** classifies Examples
- Assign A as decision attribute for node
- For **each value** of A, create new decedent of node
- Sort training examples to leaf nodes
- If training examples are perfectly classified, then STOP otherwise iterate over new leaf nodes

Decision Tree

- Which attribute is the **best** attribute?



Information Gain measure the effectiveness of an attribute

Entropy

- **Entropy** characterizes the (im)purity of an arbitrary collection of examples S.

# of possible values of X

$$Entropy(S) = \sum_{i=1}^{n} -p_i \log_2 p_i$$

Entropy

Example

- Given a **collection S**, containing positive and negative examples of some target concept, the entropy of S relative to this **Boolean classification** is

$$\boxed{\text{Entropy}\ (S) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}}$$

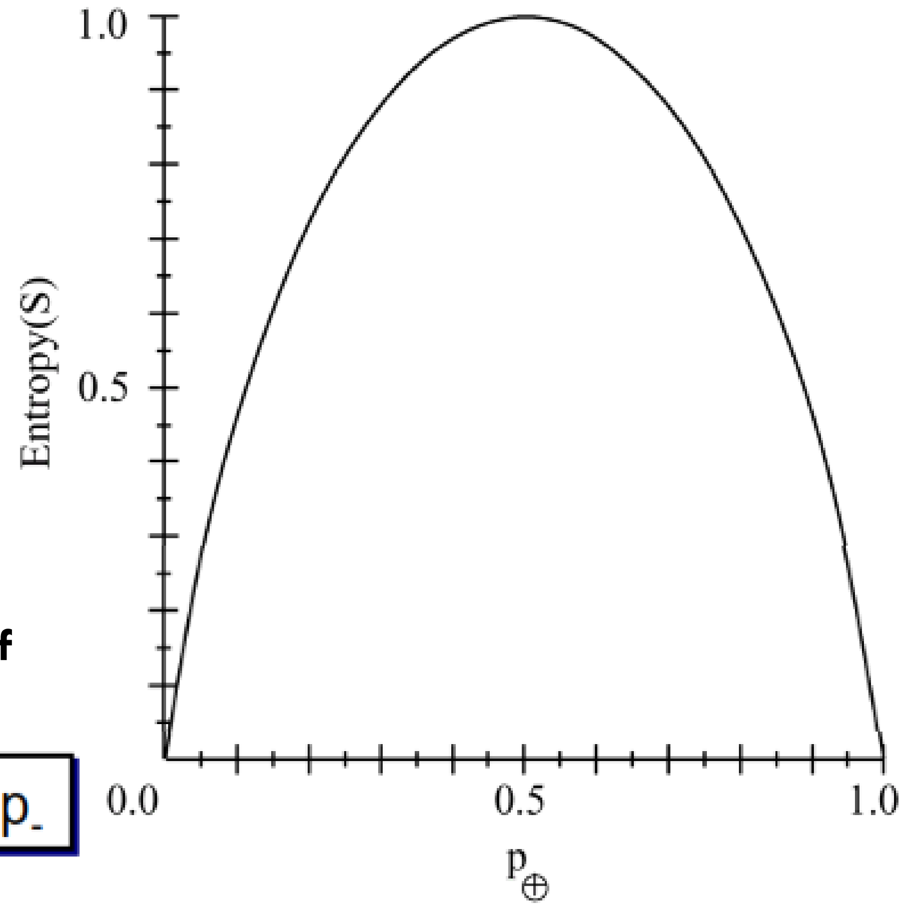
- $p_{\oplus}$ is the proportion of positive example in S
- $p_{\ominus}$ is the proportion of negative example in S

Entropy

- S is a sample of training examples
- p_+ is the proportion of positive examples in S
- p_- is the proportion of negative examples in S
- Entropy measures the impurity of S

Entropy is 0 if all members belong to same class
Entropy is 1 when there is equal no. of +ve and –ve examples

$$\text{Entropy}(S) \equiv -p_+ \log_2 p_+ \;-\; p_- \log_2 p_-$$



Information Gain

- Information Gain measure the effectiveness of an attribute
- It is simply the expected reduction in entropy

$$\text{Gain}(S, A) \equiv \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Where:

- ***Values(A)*** is the set of all possible values for attribute **A**
- S_v is the subset of **S** for which attribute **A** has value **v**.

EXAMPLE

Decision Tree

	X				Y
Day	*Outlook*	*Temperature*	*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

Entropy

Example

	X				Y
Day	*Outlook*	*Temperature*	*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

$$\boxed{\text{Entropy}\ (S) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}}$$

- In **play_tennis** example,

$$Entropy([9+, 5-]) = -(9/14) \log_2(9/14) - (5/14) \log_2(5/14)$$
$$= 0.940$$

Which Attribute?

- Which attribute should be selected for root node in play-tennis example?
  - Outlook
  - Temperature
  - Humidity
  - Wind

	X				Y
Day	*Outlook*	*Temperature*	*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

Information Gain (WIND)

- Suppose in **play-tennis** example, the attribute **WIND** which have values **Weak** and **Strong**, the ***information gain*** is:

$$\text{Gain}(S, A) \equiv \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Information Gain (WIND)

- Suppose in **play-tennis** example, the attribute **WIND** which have values **Weak** and **Strong**, the ***information gain*** is:

$$
\begin{aligned}
Values(Wind) &= Weak, Strong \\
S &= [9+, 5-] \\
S_{Weak} &\leftarrow [6+, 2-] \\
S_{Strong} &\leftarrow [3+, 3-] \\
Gain(S, Wind) &= Entropy(S) - \sum_{v \in \{Weak, Strong\}} \frac{|S_v|}{|S|} Entropy(S_v) \\
&= Entropy(S) - (8/14) Entropy(S_{Weak}) \\
&\quad - (6/14) Entropy(S_{Strong})
\end{aligned}
$$

$$Values(Wind) = Weak, Strong$$

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

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

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

Information Gain

- **Entropy S_{weak}**

$$\boxed{\text{Entropy}\ (S_{weak}) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}}$$

$$\text{Entropy}\ (S_{weak}) = -\left[\frac{6}{8}\log_2\frac{6}{8}\right] - \left[\frac{2}{8}\log_2\frac{2}{8}\right]$$

$$\text{Entropy}\ (S_{weak}) = -0.75(-0.415)\ - 0.25(-2)$$

$$\text{Entropy}\ (S_{weak}) = 0.311 + 0.5 = 0.811$$

Information Gain

$$Values(Wind) = Weak, Strong$$

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

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

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

- **Entropy S_{strong}**

$$\boxed{\text{Entropy}\left(S_{strong}\right) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}}$$

$$\text{Entropy}\left(S_{strong}\right) = -[\frac{3}{6}\log_2\frac{3}{6}] - [\frac{3}{6}\log_2\frac{3}{6}]$$

$$\text{Entropy}\left(S_{strong}\right) = -0.5(-1)\ -0.5(-1)$$

$$\text{Entropy}\left(S_{strong}\right) = 0.5 + 0.5 = 1$$

Information Gain (WIND)

$$Values(Wind) = Weak, Strong$$

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

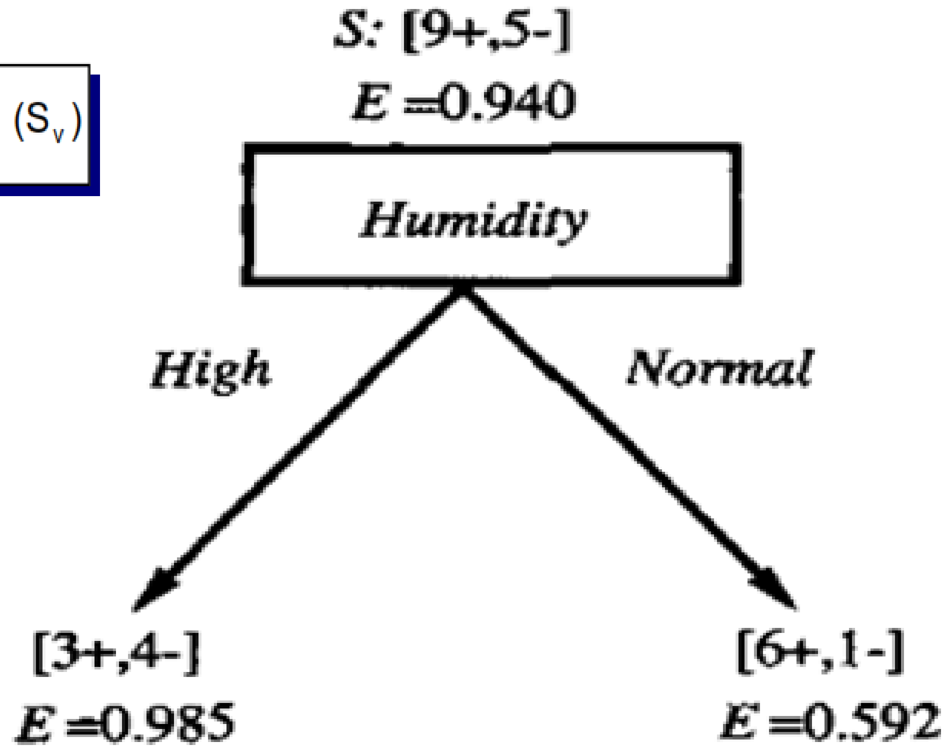
$$S_{Weak} \leftarrow [6+, 2-]$$

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

$$\begin{aligned} Gain(S, Wind) &= Entropy(S) - \sum_{v \in \{Weak, Strong\}} \frac{|S_v|}{|S|} Entropy(S_v) \\ &= Entropy(S) - (8/14) Entropy(S_{Weak}) \\ &\quad - (6/14) Entropy(S_{Strong}) \\ &= 0.940 - (8/14)0.811 - (6/14)1.00 \\ &= 0.048 \end{aligned}$$

Information Gain (HUMIDITY)

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

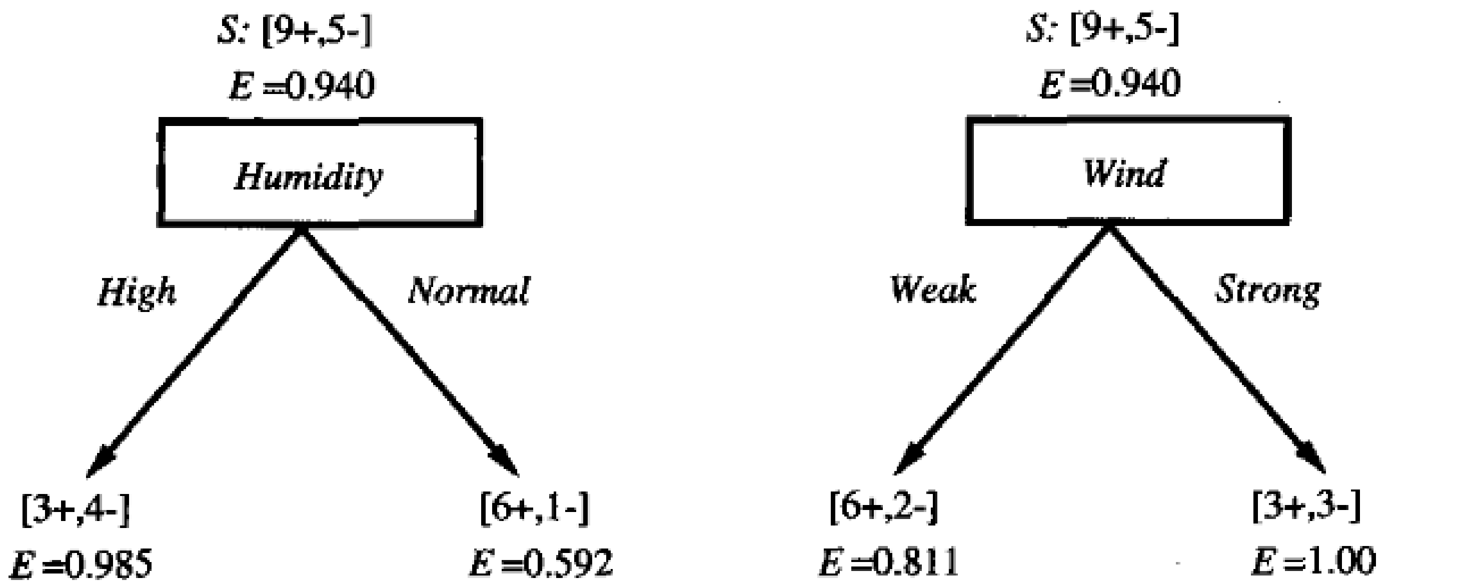


Gain (S, Humidity)

$= .940 - (7/14).985 - (7/14).592$
$= .151$

Information Gain

Which attribute is the best classifier?



Gain (S, Humidity)

= .940 - (7/14).985 - (7/14).592
= .151

Gain (S, Wind)

= .940 - (8/14).811 - (6/14)1.0
= .048

Humidity provide greater information gain than wind

Information Gain

Which attribute is the best classifier?

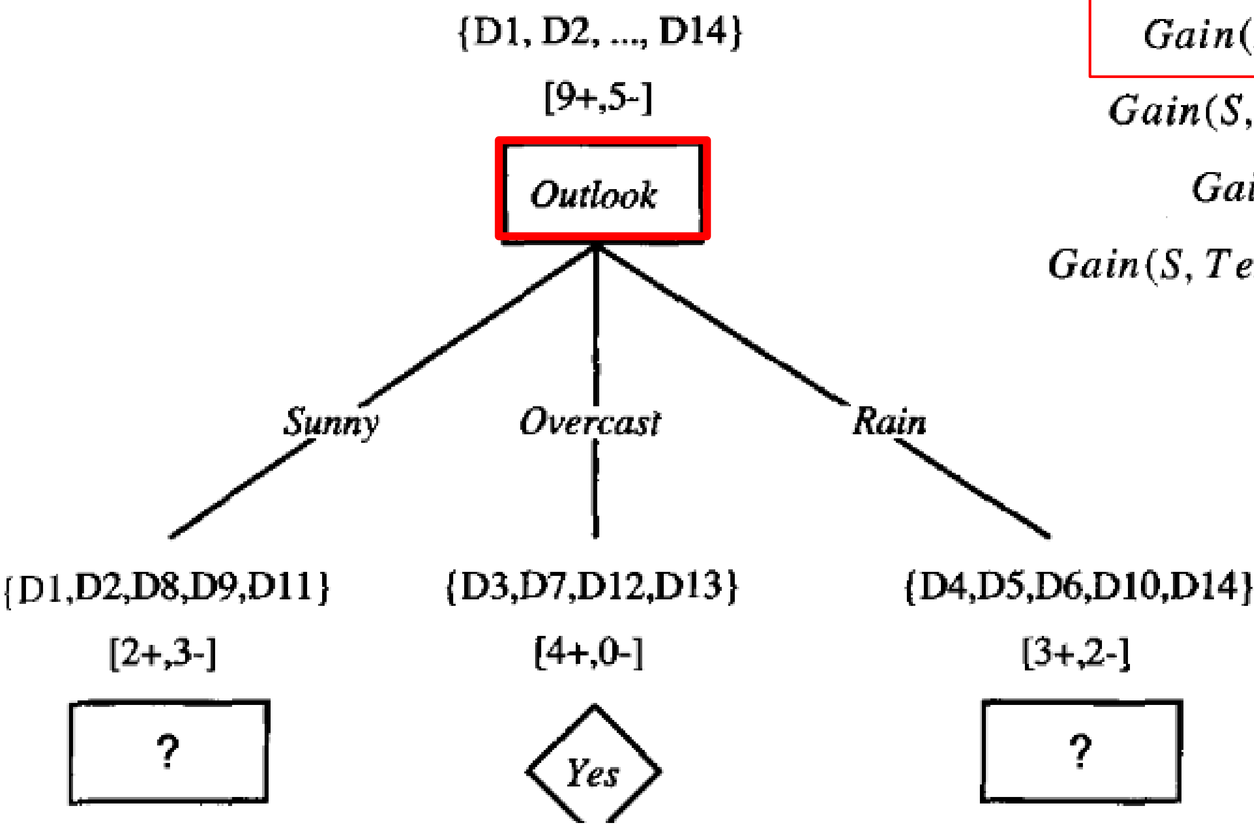
$$\boxed{Gain(S, Outlook) = 0.246}$$

$$Gain(S, Humidity) = 0.151$$

$$Gain(S, Wind) = 0.048$$

$$Gain(S, Temperature) = 0.029$$

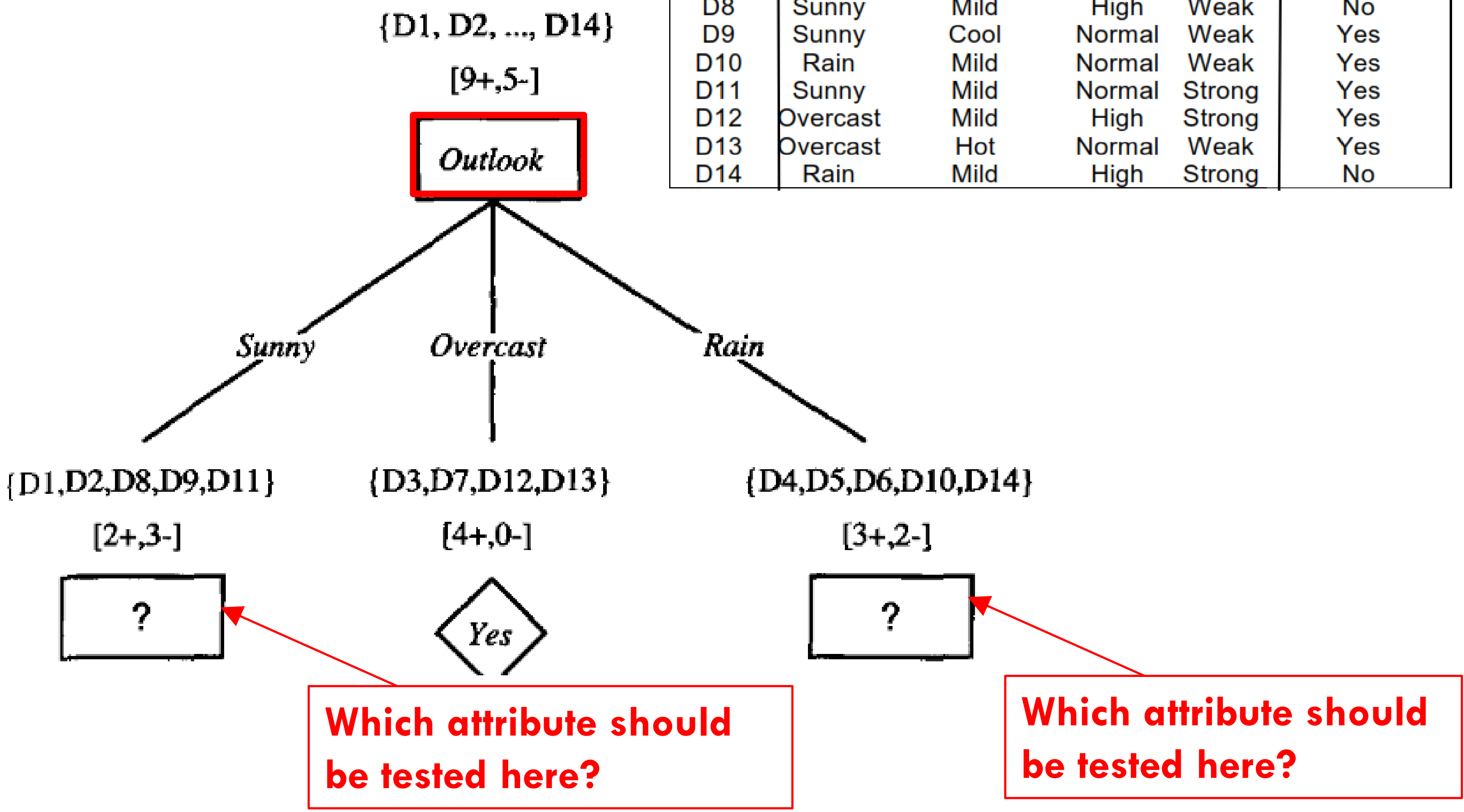
Decision Tree



$$\boxed{Gain(S, Outlook) = 0.246}$$
$$Gain(S, Humidity) = 0.151$$
$$Gain(S, Wind) = 0.048$$
$$Gain(S, Temperature) = 0.029$$

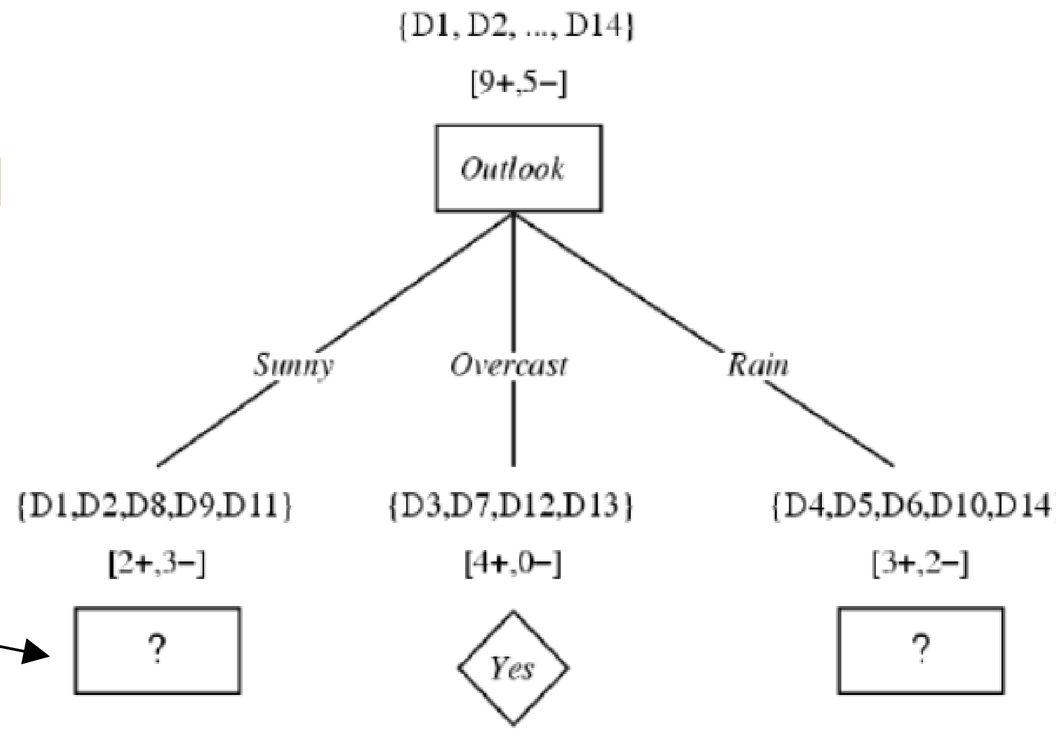
Decision Tree

Day	Outlook	Temperature	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



Decision Tree

Which attribute should be tested here?



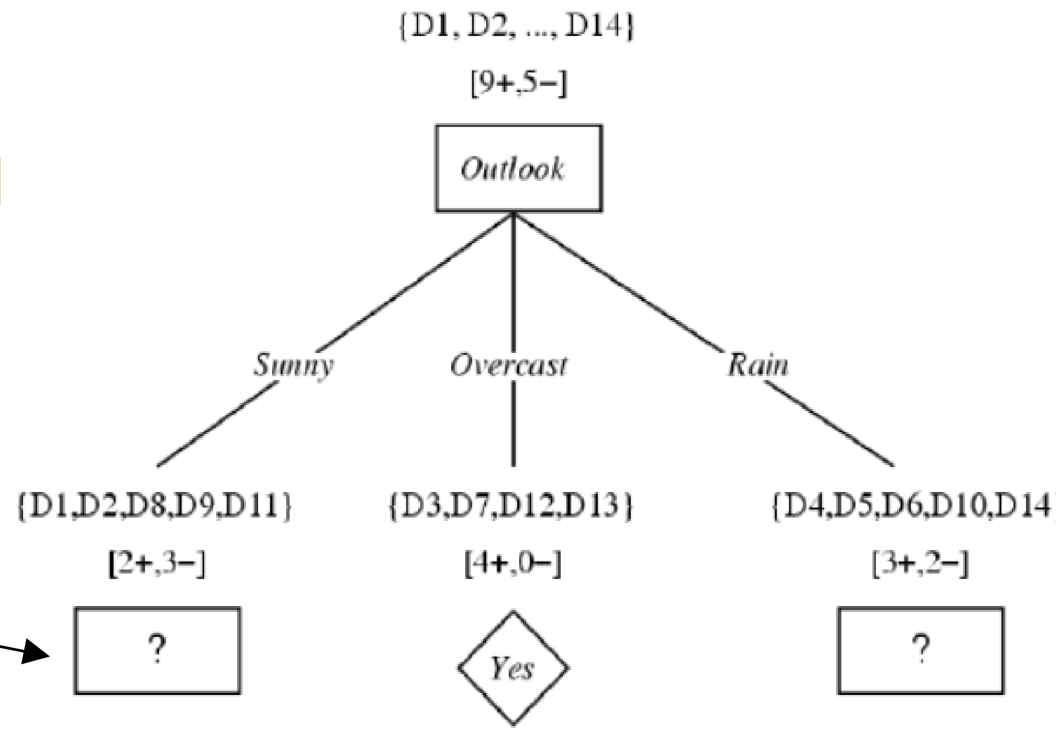
$$S_{sunny} = \{D1,D2,D8,D9,D11\}$$

$$Gain\,(S_{sunny}, Humidity)$$

$$Gain\,(S_{sunny}, Temperature)$$

$$Gain\,(S_{sunny}, Wind)$$

Decision Tree



Which attribute should be tested here?

$$S_{sunny} = \{D1,D2,D8,D9,D11\}$$

$$Gain\ (S_{sunny}\ ,\ Humidity)$$

$$Gain\ (S_{sunny}\ ,\ Temperature)$$

$$Gain\ (S_{sunny}\ ,\ Wind)$$

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

Information Gain

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

- $Gain(S_{sunny}, Humidity)$

$$\boxed{Gain\ (S, A) \equiv Entropy\ (S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy\ (S_v)}$$

$$Gain(S_{sunny}, Humidity) = Entropy(S_{sunny}) -$$

$$\left[\frac{Humidity_{high}}{S_{sunny}} [Entropy(Humidity_{high})] + \frac{Humidity_{normal}}{S_{sunny}} [Entropy(Humidity_{normal})] \right]$$

Information Gain

- $\boldsymbol{Gain(S_{sunny}, Humidity)}$

$$\text{Gain}(S, A) \equiv \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$Gain(S_{sunny}, Humidity) = Entropy(S_{sunny}) -$$

$$\begin{bmatrix} \frac{3}{5}[Entropy(Humidity_{high})] + \\ \frac{2}{5}[Entropy(Humidity_{normal})] \end{bmatrix}$$

Information Gain

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

1. $Entropy\,(S_{sunny})$

$$\boxed{Entropy\,(S_{sunny}) = -p_{\oplus}\log_2 p_{\oplus} - p_{\ominus}\log_2 p_{\ominus}}$$

$$Entropy\,(S_{sunny}) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}$$

$$Entropy\,(S_{sunny}) = 0.529 + 0.441 = 0.970$$

Information Gain

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

$$\text{Entropy } (S) = -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$

2. $\boldsymbol{Entropy(Humidity_{high})}$

$$\text{Entropy } (Humidity_{high}) = -\frac{0}{3}\log_2\frac{0}{3} - \frac{3}{3}\log_2\frac{3}{3}$$

$$\text{Entropy } (Humidity_{high}) = 0 + 0 = 0$$

3. $\boldsymbol{Entropy(Humidity_{normal})}$

$$\text{Entropy } (Humidity_{normal}) = -\frac{2}{2}\log_2\frac{2}{2} - \frac{0}{2}\log_2\frac{0}{2}$$

$$\text{Entropy } (Humidity_{normal}) = 0 + 0 = 0$$

Information Gain

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

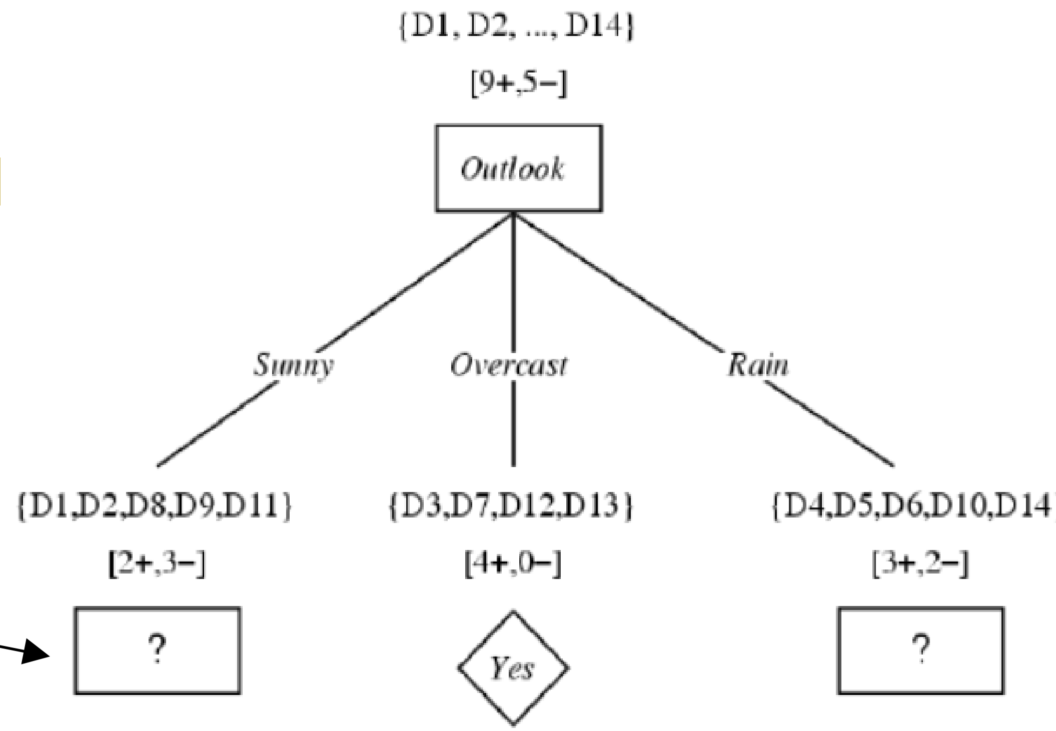
- $\boldsymbol{Gain(S_{sunny}, Humidity)}$

$$Gain(S_{sunny}, Humidity) = Entropy(S_{sunny}) - \left[\frac{3}{5}\left[Entropy(Humidity_{high})\right] + \frac{2}{5}\left[Entropy(Humidity_{normal})\right]\right]$$

$$Gain(S_{sunny}, Humidity) = 0.97 - \left[\frac{3}{5}(0) + \frac{2}{5}(0)\right] = 0.970$$

Decision Tree

Which attribute should be tested here?



$$S_{sunny} = \{D1, D2, D8, D9, D11\}$$

$$Gain\ (S_{sunny}, Humidity) = .970 - (3/5)\ 0.0 - (2/5)\ 0.0 \boxed{= .970}$$

$$Gain\ (S_{sunny}, Temperature) = .970 - (2/5)\ 0.0 - (2/5)\ 1.0 - (1/5)\ 0.0 = .570$$

$$Gain\ (S_{sunny}, Wind) = .970 - (2/5)\ 1.0 - (3/5)\ .918 = .019$$

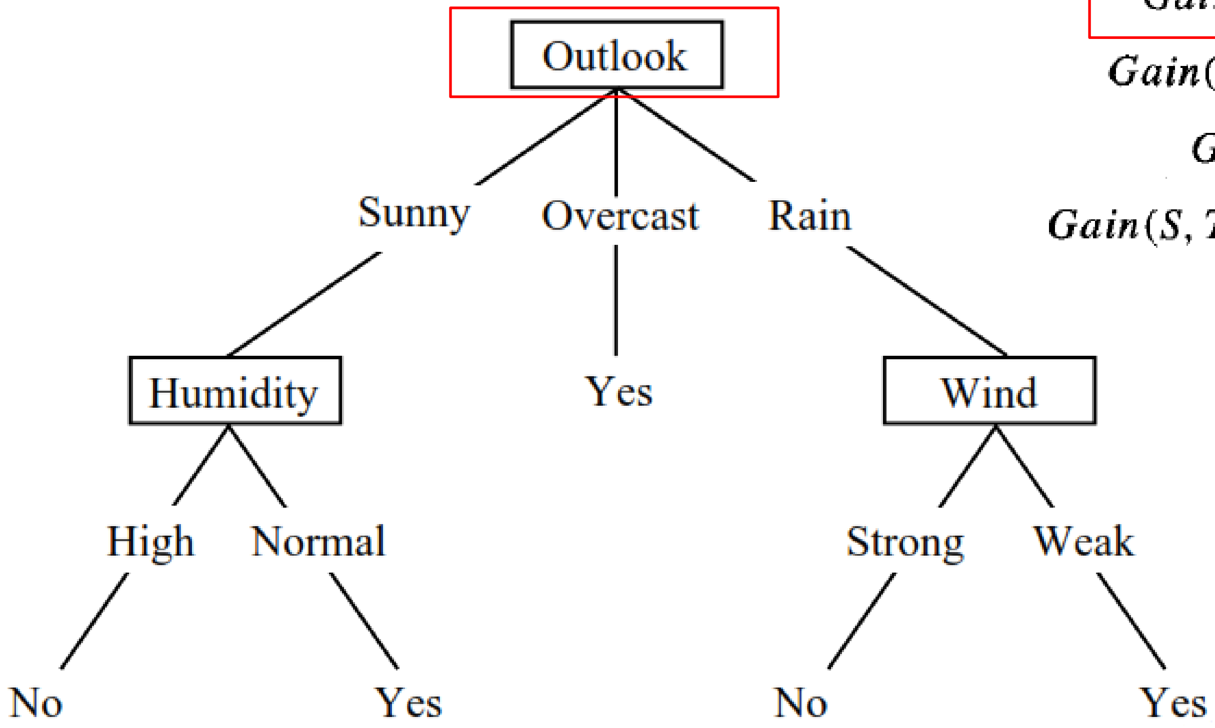
Decision Tree

$$Gain(S, Outlook) = 0.246$$

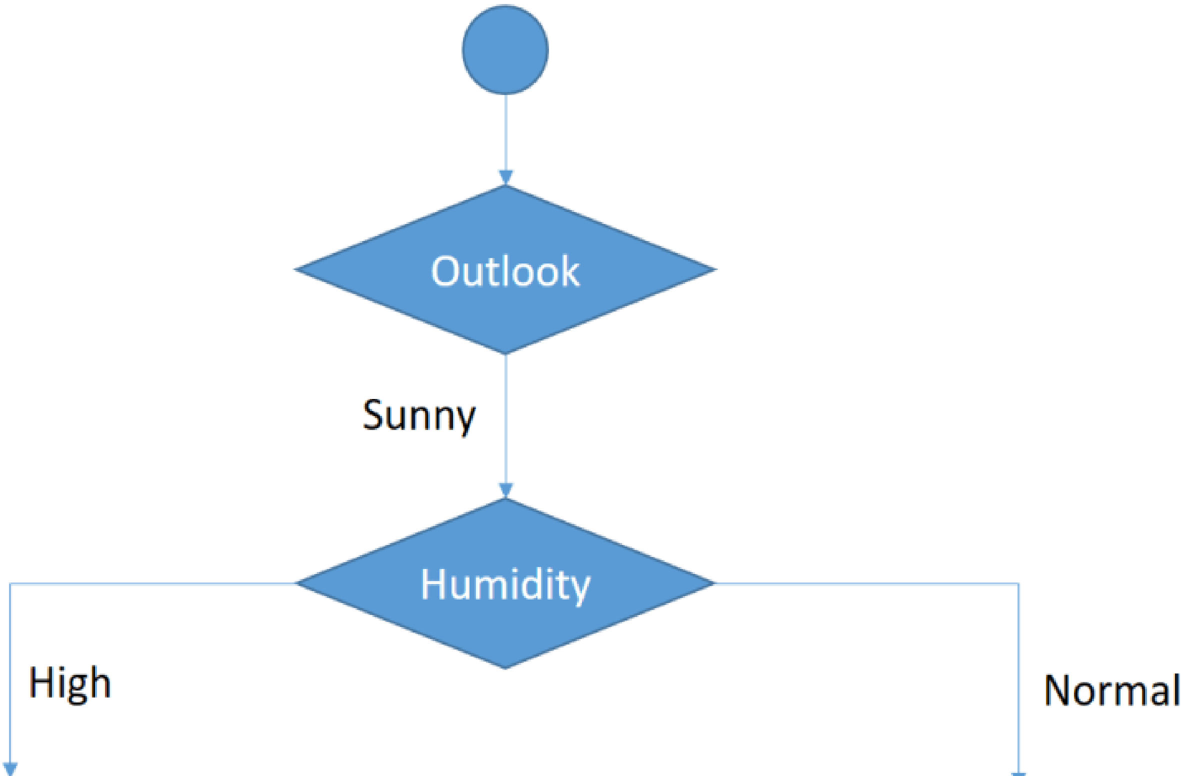
$$Gain(S, Humidity) = 0.151$$

$$Gain(S, Wind) = 0.048$$

$$Gain(S, Temperature) = 0.029$$



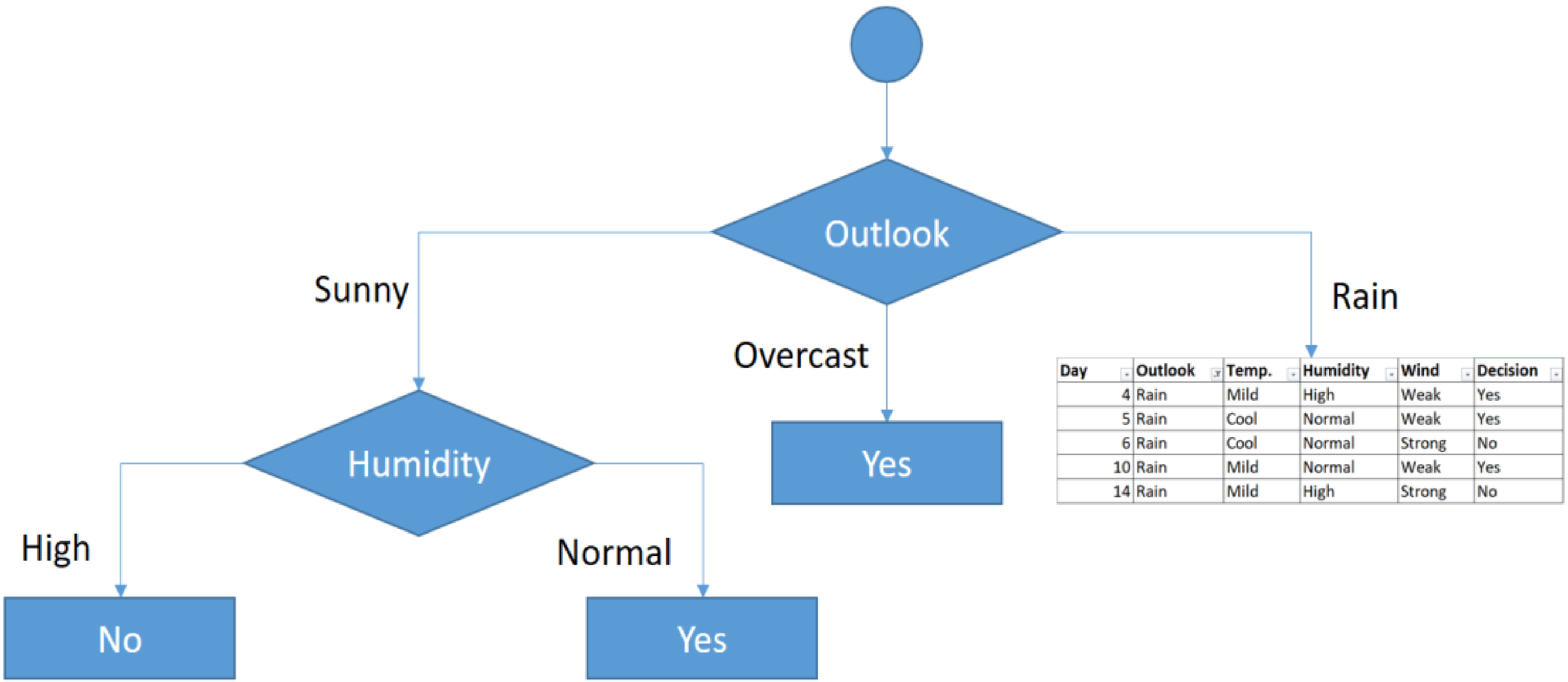
Decision Tree



Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No

Day	Outlook	Temp.	Humidity	Wind	Decision
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

Decision Tree



Day	Outlook	Temp.	Humidity	Wind	Decision
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
10	Rain	Mild	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Acknowledgement

Tom Mitchel, Russel & Norvig, Andrew Ng, Alpydin & Ch. Eick.