CS 4104
APPLIED MACHINE LEARNING

Dr. Hashim Yasin

National University of Computer and Emerging Sciences,

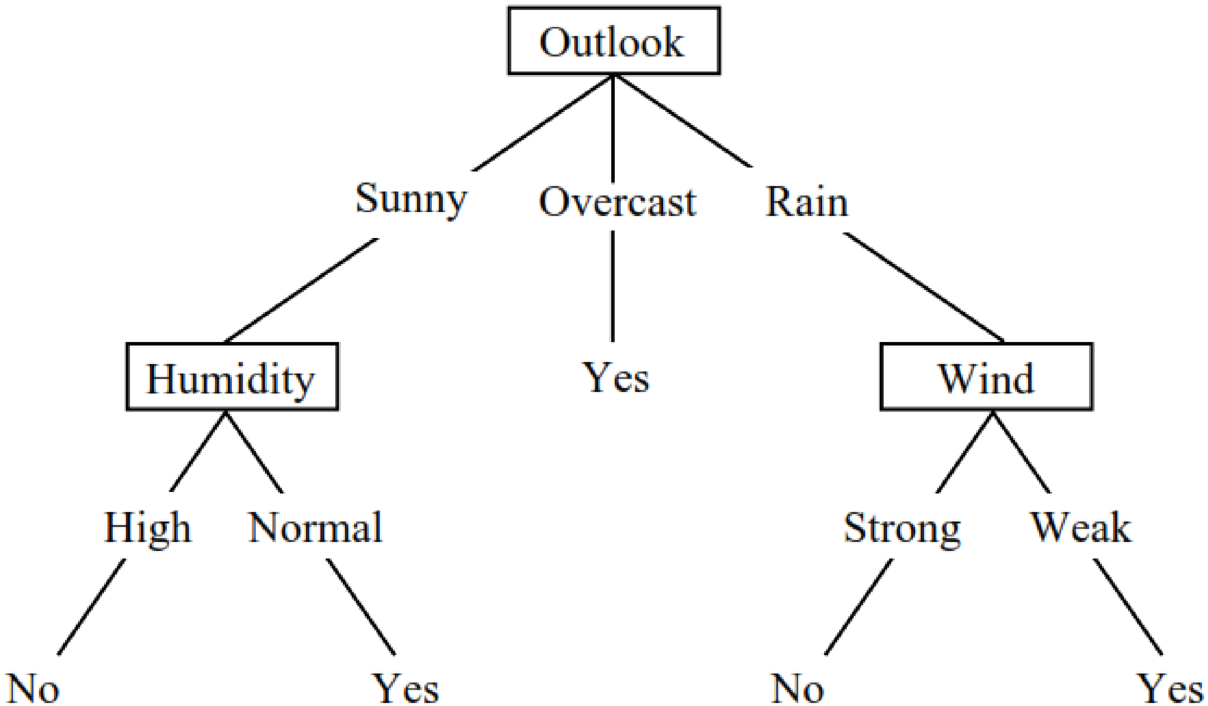
Faisalabad, Pakistan.

DECISION TREE

Decision Tree

- A Decision tree for

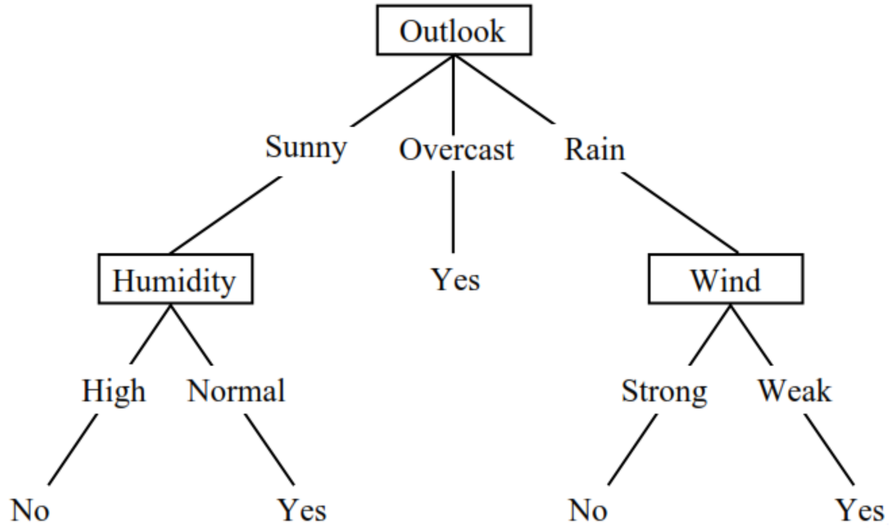
<Outlook, Temperature, Humidity, Wind> →PlayTennis?



Decision Tree

- A Decision tree for

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

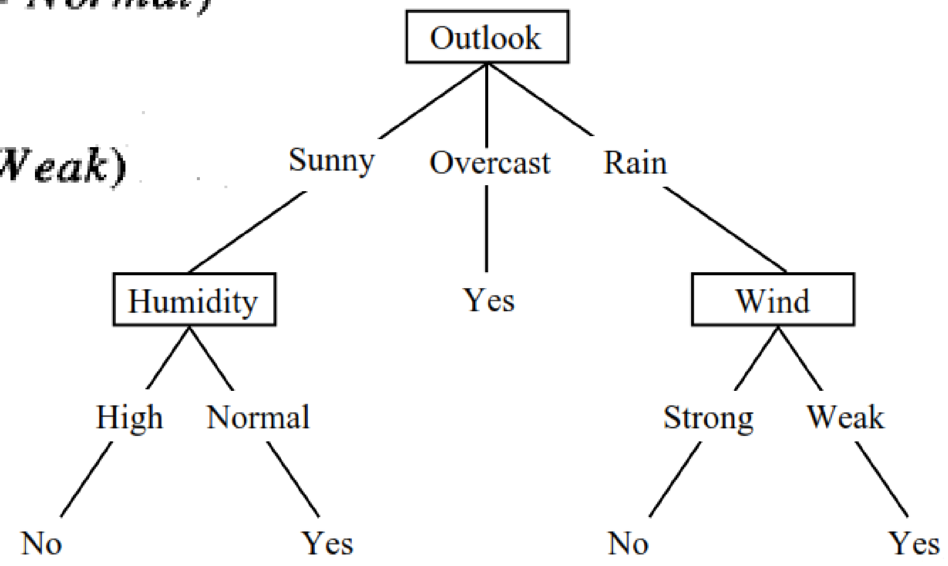


- **Each internal node:** test one attribute X_i
- **Each branch from a node:** selects one value for X_i
- **Each leaf node:** predict Y

Decision Tree

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

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



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

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

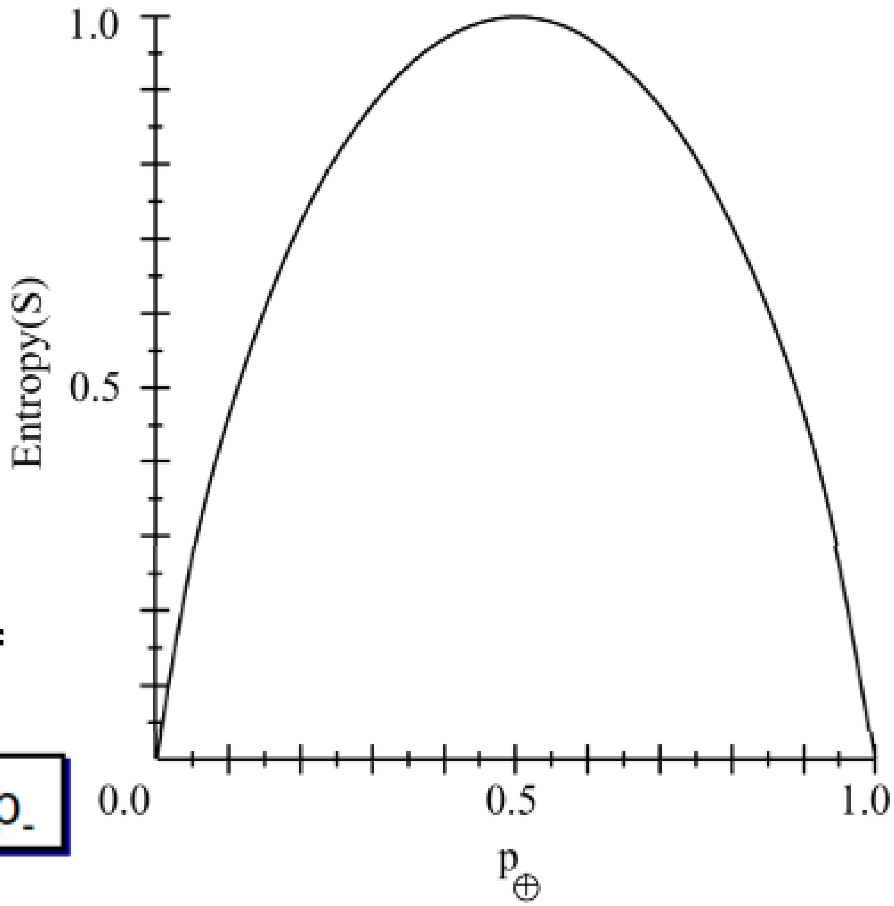
- $p_{\oplus}$ is the proportion of positive example in S
- $p_{\ominus}$ is the proportion of negative example in S

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

Information Gain

Which attribute is the best classifier?

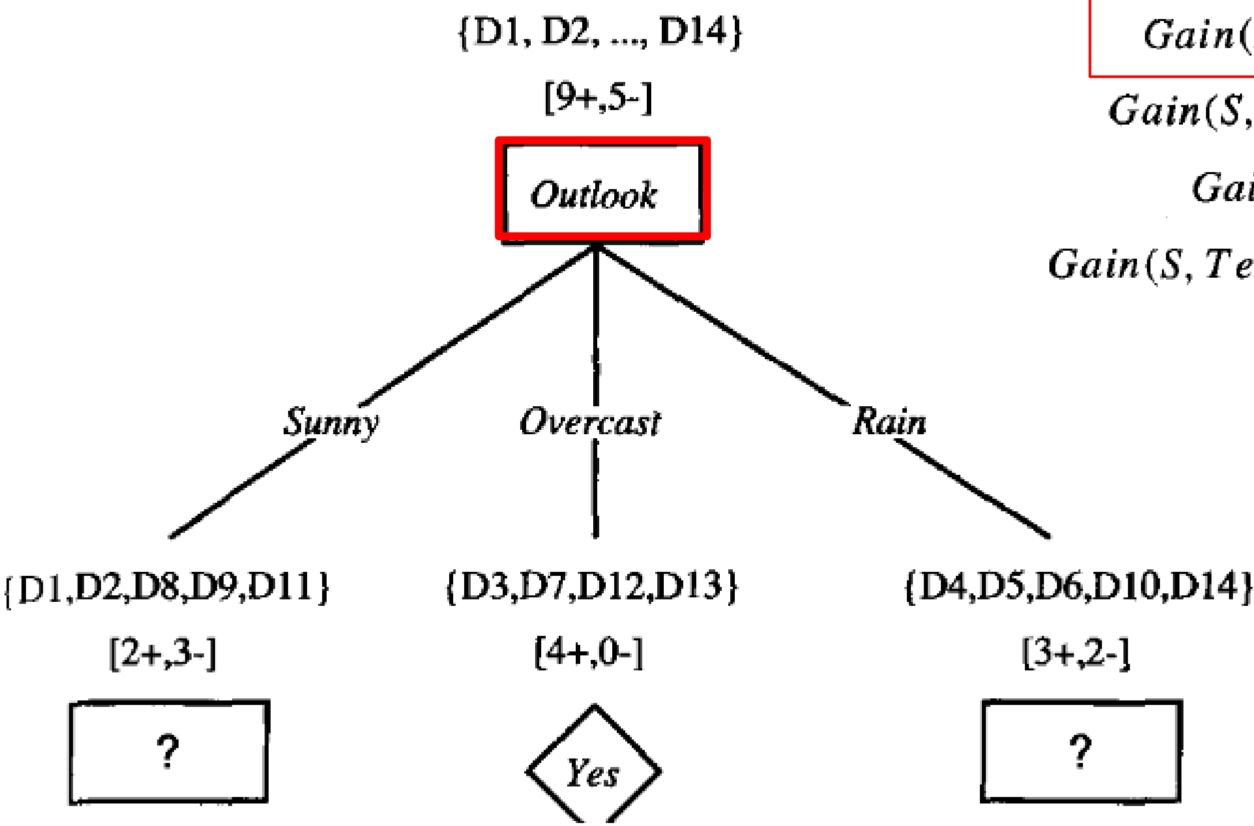
$$\boxed{Gain(S, Outlook) = 0.246}$$

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

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

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

Decision Tree



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

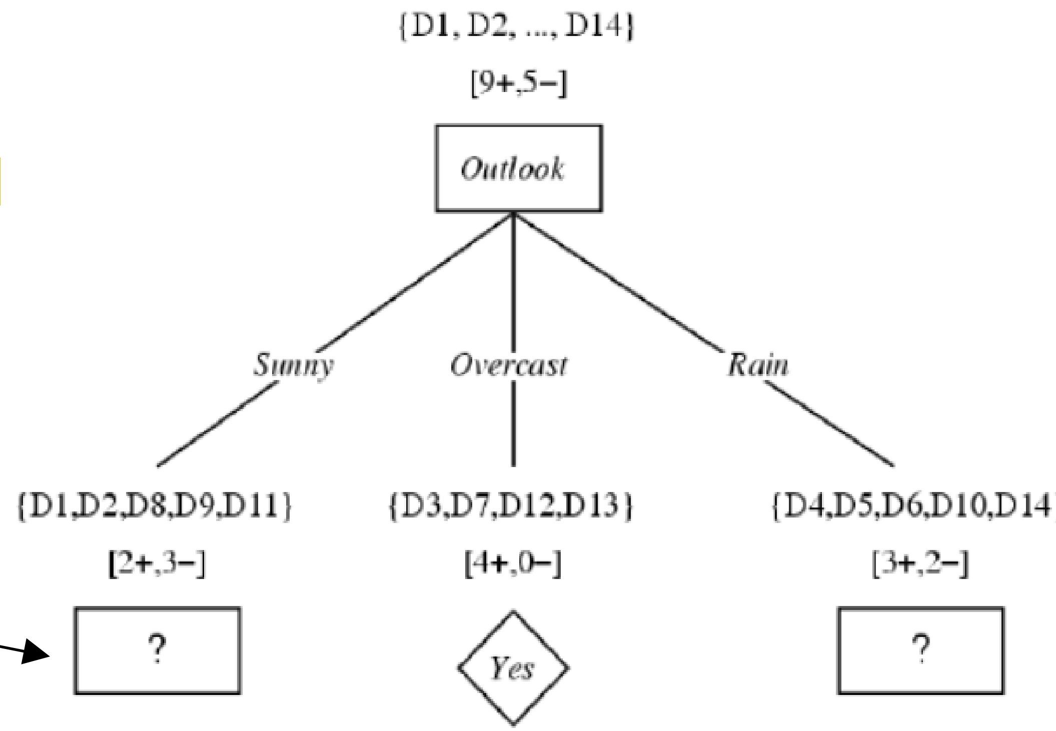
$$Gain(S, Humidity) = 0.151$$

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

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

Decision Tree

Which attribute should be tested here?



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

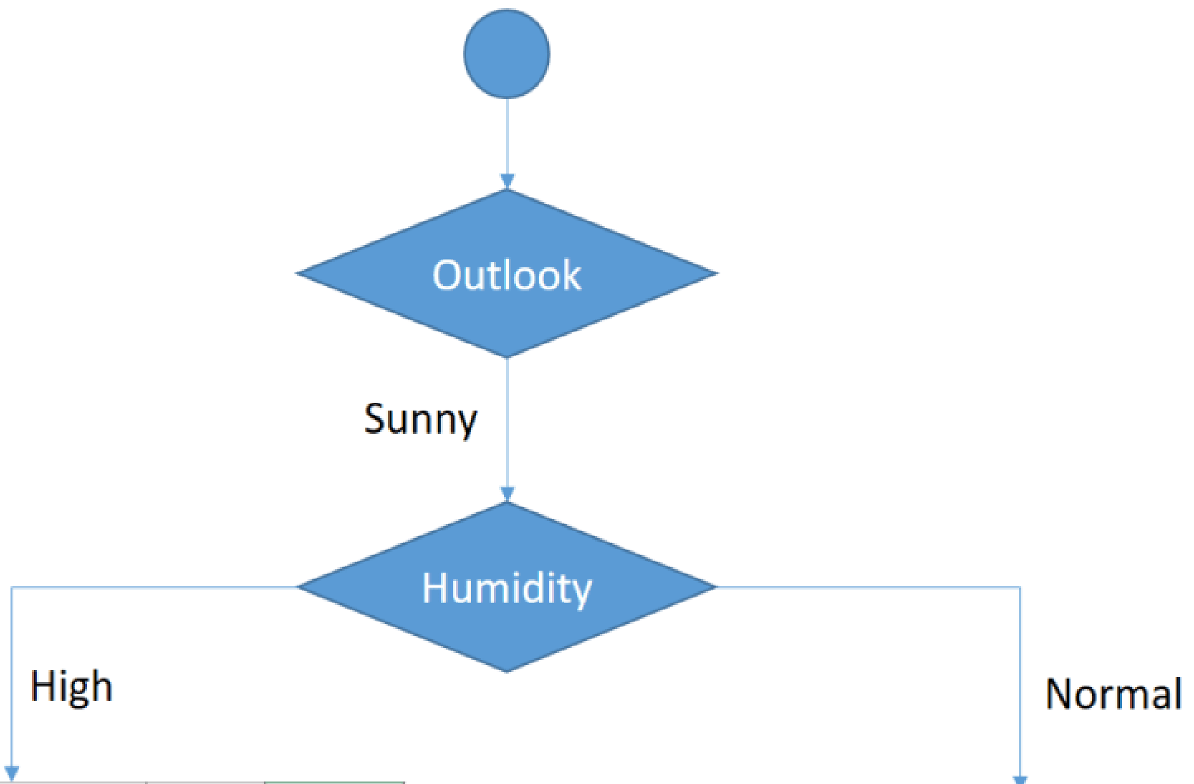
$Gain\ (S_{sunny}\ ,\ Humidity)$

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

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

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

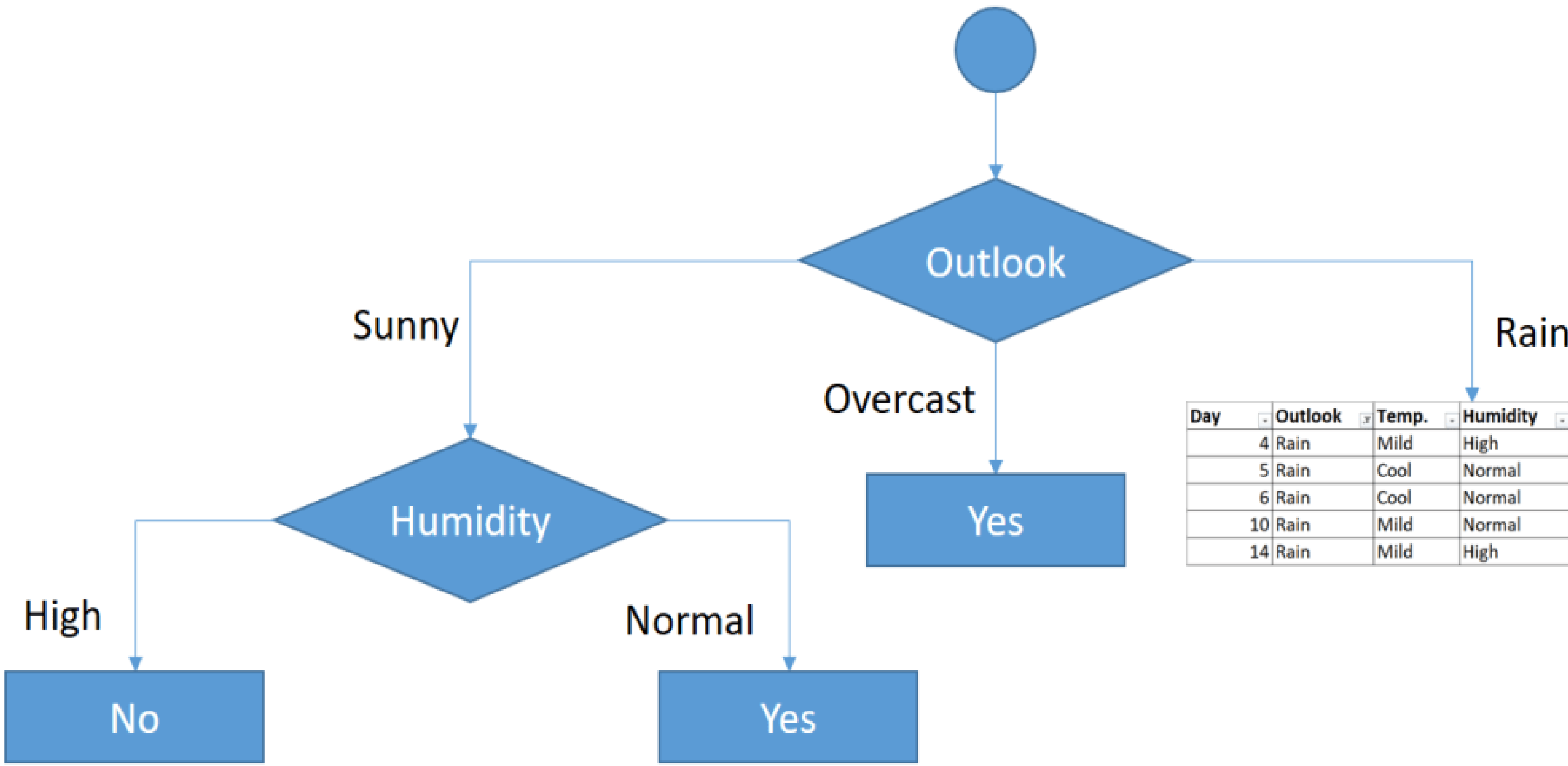
Decision Tree



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

CONTINUOUS-VALUED ATTRIBUTES

Continuous-Valued Attributes

- For an attribute A that is continuous-valued, the algorithm can dynamically create a new Boolean attribute A, that is true if $A < t$ and false otherwise.
- The only question is how to select the best value for the threshold t.

Temperature:	40	48	60	72	80	90
PlayTennis:	No	No	Yes	Yes	Yes	No

Continuous-Valued Attributes

- ***Sort the values.***
- There are ***two candidate thresholds***, corresponding to the values of Temperature at which the value of PlayTennis changes:

$$(48 + 60)/2 = 54$$

$$(80 + 90)/2 = 85$$

Temperature:	40	48	60	72	80	90
PlayTennis:	No	No	Yes	Yes	Yes	No

Continuous-Valued Attributes

- The information gain can then be computed for each of the candidate attributes, $\boldsymbol{Temperature_{>54}}$ and $\boldsymbol{Temperature_{>85}}$
- And the best can be selected ($\boldsymbol{Temperature_{>54}}$)

Temperature:	40	48	60	72	80	90
PlayTennis:	No	No	Yes	Yes	Yes	No

ID3 … Capabilities & Limitations

- ID3 hypothesis space of all decision trees is a ***complete* space** of finite discrete-valued functions, relative to the available attributes.
- ID3 maintain only **single current hypothesis**.
- ID3 **uses all training examples** at each step in the search.
- ID3 performs **no backtracking** in its search … as a result, it has risk in **converging to locally optimal solutions** that are not globally optimal.

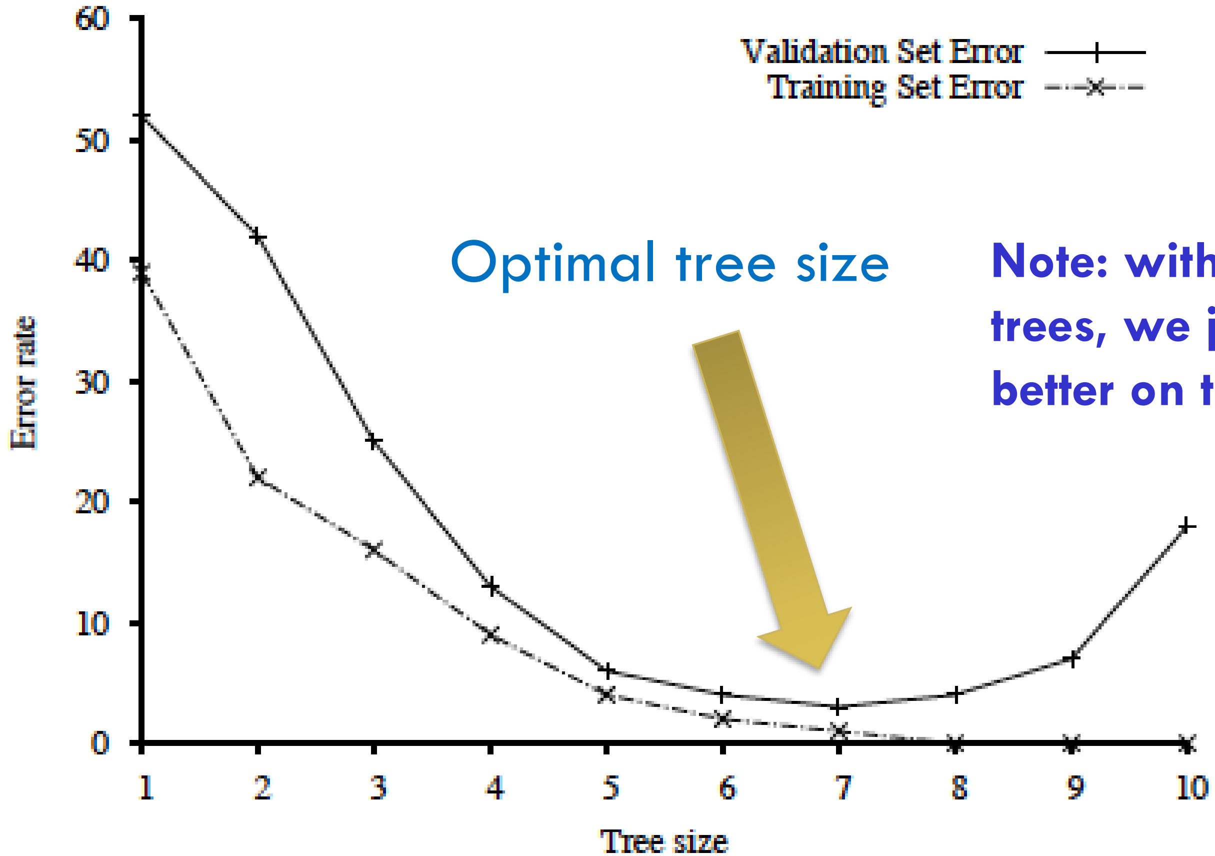
ID3 … Capabilities & Limitations

- The "information gain" favors **attributes with many values** over those with few values.
- *It is problematic: attributes with a large number of values* (extreme case: ID code)
- Information gain is biased towards choosing attributes with a large number of values
  - This may result in *overfitting*(selection of an attribute that is non-optimal for prediction)

ID3 … Capabilities & Limitations

ID code	Outlook	Temp.	Humidity	Windy	Play
A	Sunny	Hot	High	False	No
B	Sunny	Hot	High	True	No
C	Overcast	Hot	High	False	Yes
D	Rainy	Mild	High	False	Yes
E	Rainy	Cool	Normal	False	Yes
F	Rainy	Cool	Normal	True	No
G	Overcast	Cool	Normal	True	Yes
H	Sunny	Mild	High	False	No
I	Sunny	Cool	Normal	False	Yes
J	Rainy	Mild	Normal	False	Yes
K	Sunny	Mild	Normal	True	Yes
L	Overcast	Mild	High	True	Yes
M	Overcast	Hot	Normal	False	Yes
N	Rainy	Mild	High	True	No

Overfitting



Note: with larger and larger trees, we just do better and better on the training set!

Acknowledgement

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