Applied Machine Learning

Dr. Harikrishnan N B
Computer Science and Information Systems

BITS Pilani
Pilani Campus

SE ZG568 / SS ZG568,
Applied Machine Learning
Lecture No. 8 [09- March-2025]

Decision Tree

Learning - Function Approximation

Problem Setting

- Set of possible instances *X*.
- Unknown target function *f: X—> Y*
- Set of function hypotheses $H = \{h \mid h: X \longrightarrow Y\}$

Input

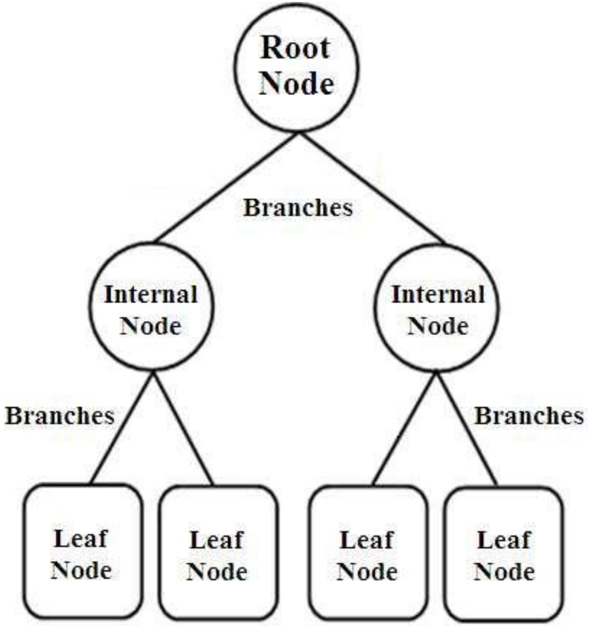
Training Examples of unknown target function *f.*

Output

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

Decision Tree

Decision Tree Learning is a method for approximating the target function **(*Y*)**, in which the **learned functions (h)** is represented by a **decision tree**.



$$h: X \longrightarrow Y$$

Decision Tree

- **ID3** (Iterative Dichotomiser 3)- **1986 Ross Quinlan**
- **C4.5** - is the successor to ID3 and removed the restriction that features must be categorical by dynamically defining a discrete attribute (based on numerical variables) that partitions the continuous attribute value into a discrete set of intervals.
- **C5.0** - is Quinlan's latest version release under a proprietary license. It uses less memory and builds smaller rulesets than C4.5 while being more accurate.
- **CART** - Classification and Regression Trees - s very similar to C4.5, but it differs in that it supports numerical target variables (regression) and does not compute rule sets. CART constructs binary trees using the feature and threshold that yield the largest information gain at each node.

scikit-learn uses an optimized version of the CART algorithm; however, the scikit-learn implementation does not support categorical variables for now.

https://scikit-learn.org/stable/modules/tree.html#classification

Play Tennis or Not?

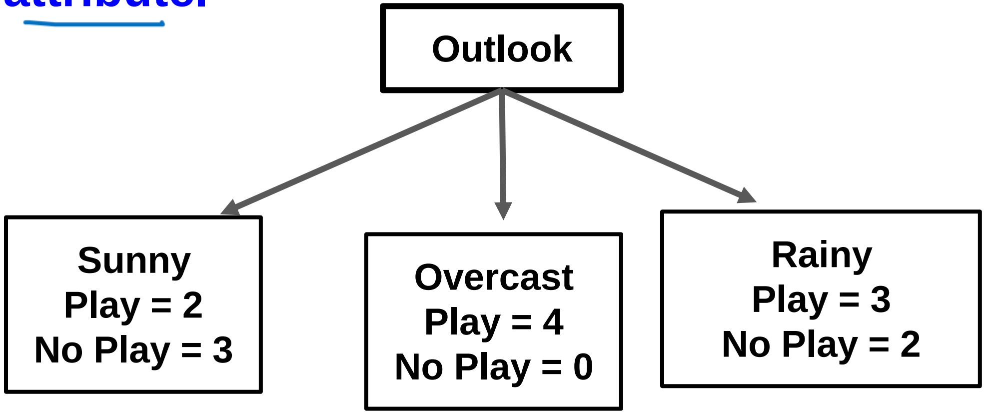
Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

ID3- Decision Tree algorithm for classification

ID3 algorithm

Step 1: Select an attribute to place on the root node and make one branch for each possible value of the attribute.

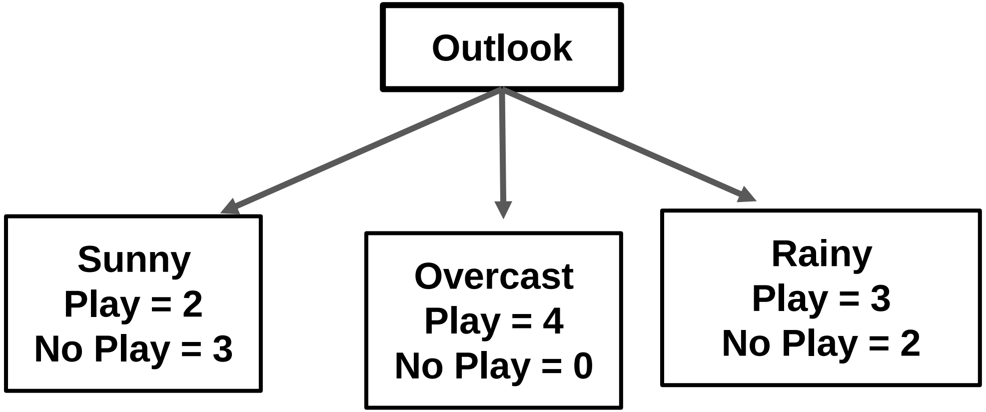
Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No



ID3 algorithm

Step 2: Make an assessment of the quality of the split

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

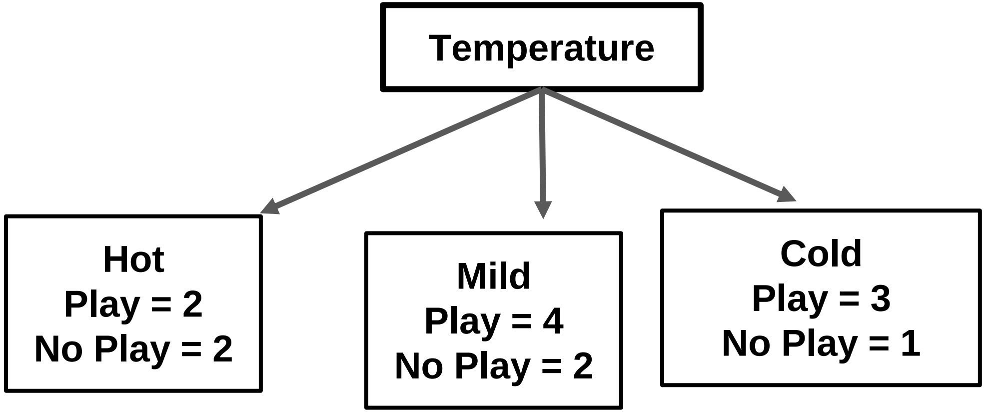


HOW DO YOU MEASURE THE QUALITY?

ID3 algorithm

Step 3: Repeat Step 1 and Step 2 for all other attributes

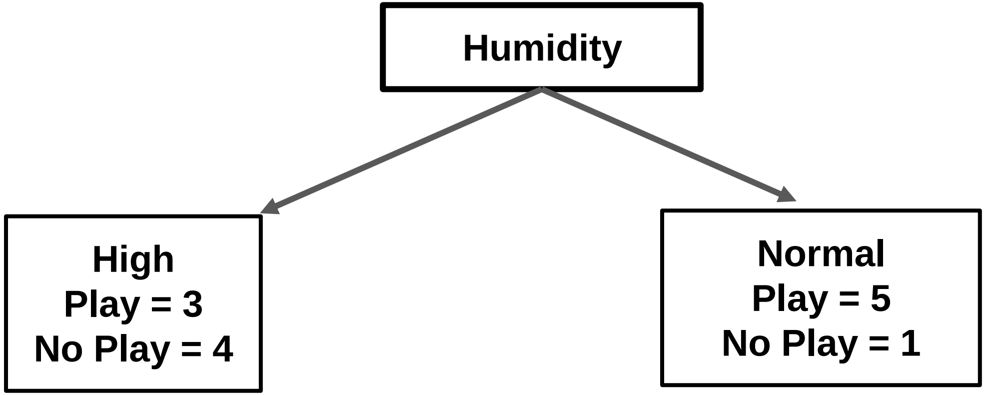
Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No



ID3 algorithm

Step 3: Repeat Step 1 and Step 2 for all other attributes

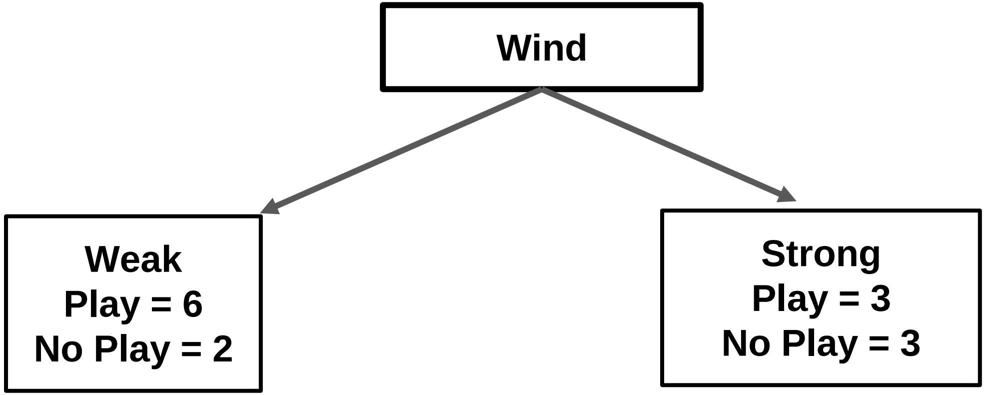
Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No



ID3 algorithm

Step 3: Repeat Step 1 and Step 2 for all other attributes

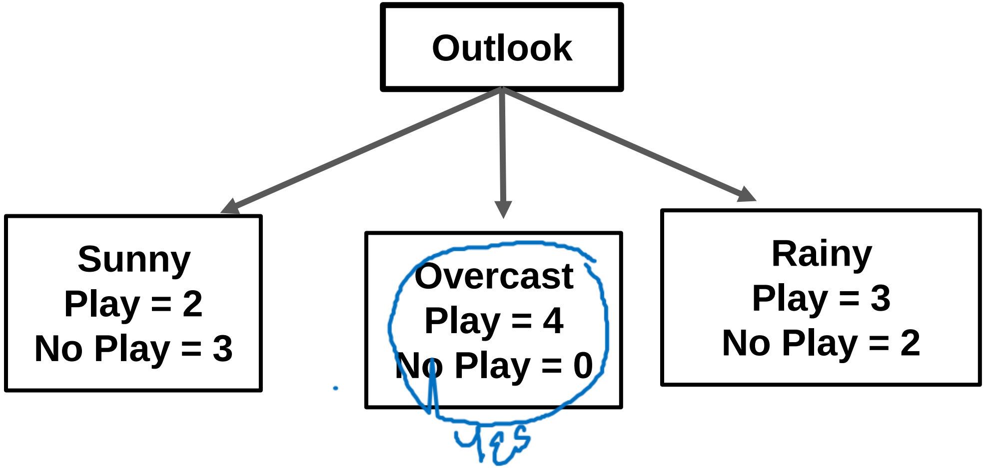
Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No



ID3 algorithm

Step 4: Depending on the QUALITY of the Partial Tree, we select one partial tree

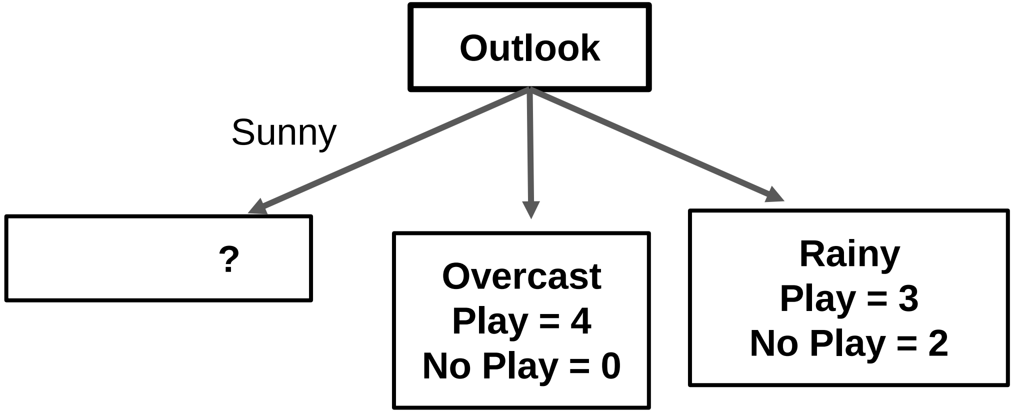
Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No



ID3 algorithm

Step 5: Repeat steps 1 to 4 for each daughter nodes of the selected partial tree

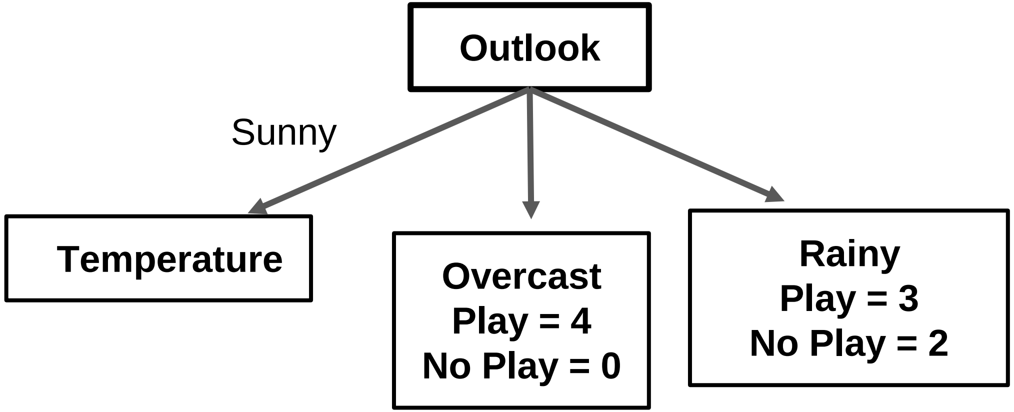
Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No



ID3 algorithm

Step 5: Repeat steps 1 to 4 for each daughter nodes of the selected partial tree

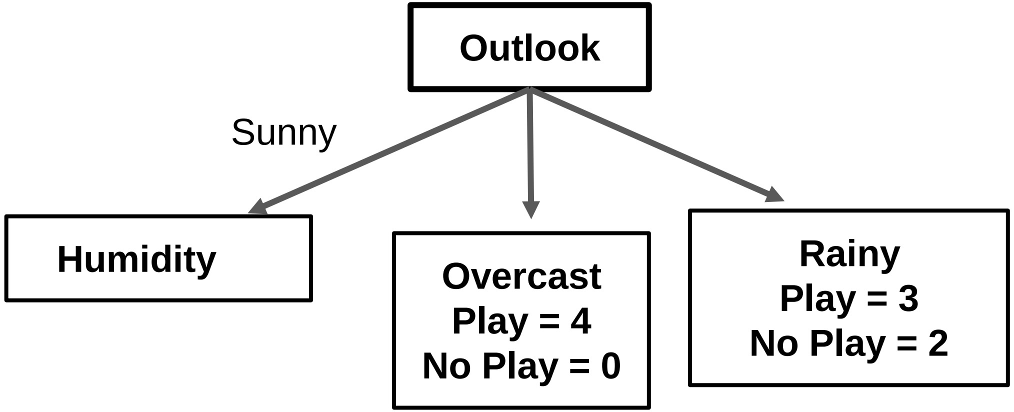
Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No



ID3 algorithm

Step 5: Repeat steps 1 to 4 for each daughter nodes of the selected partial tree

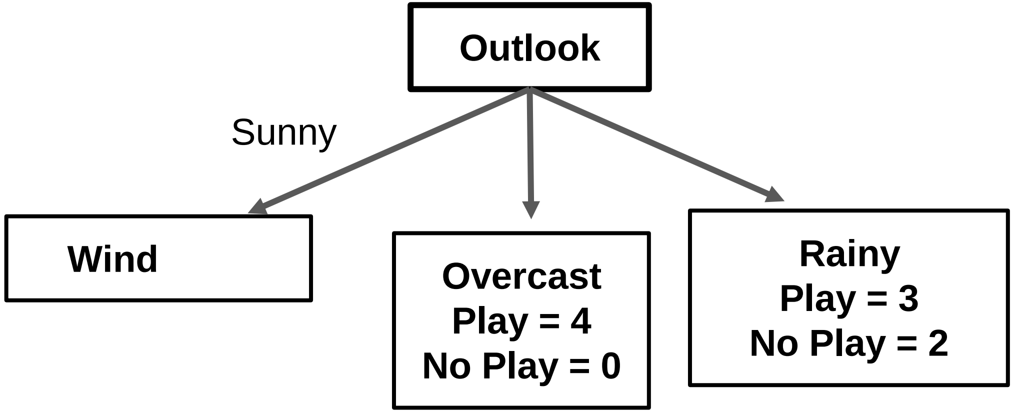
Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No



ID3 algorithm

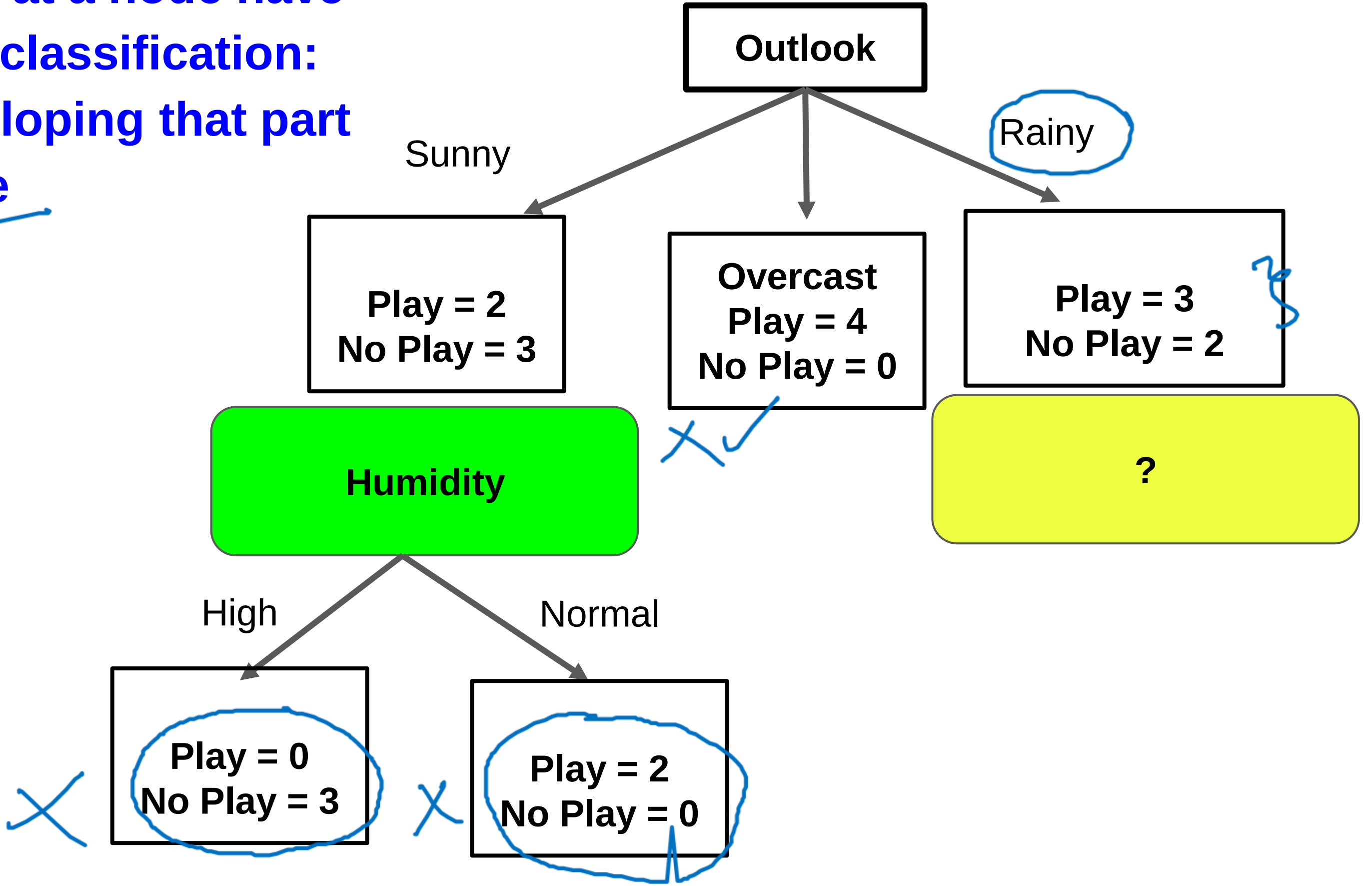
Step 5: Repeat steps 1 to 4 for each daughter nodes of the selected partial tree

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

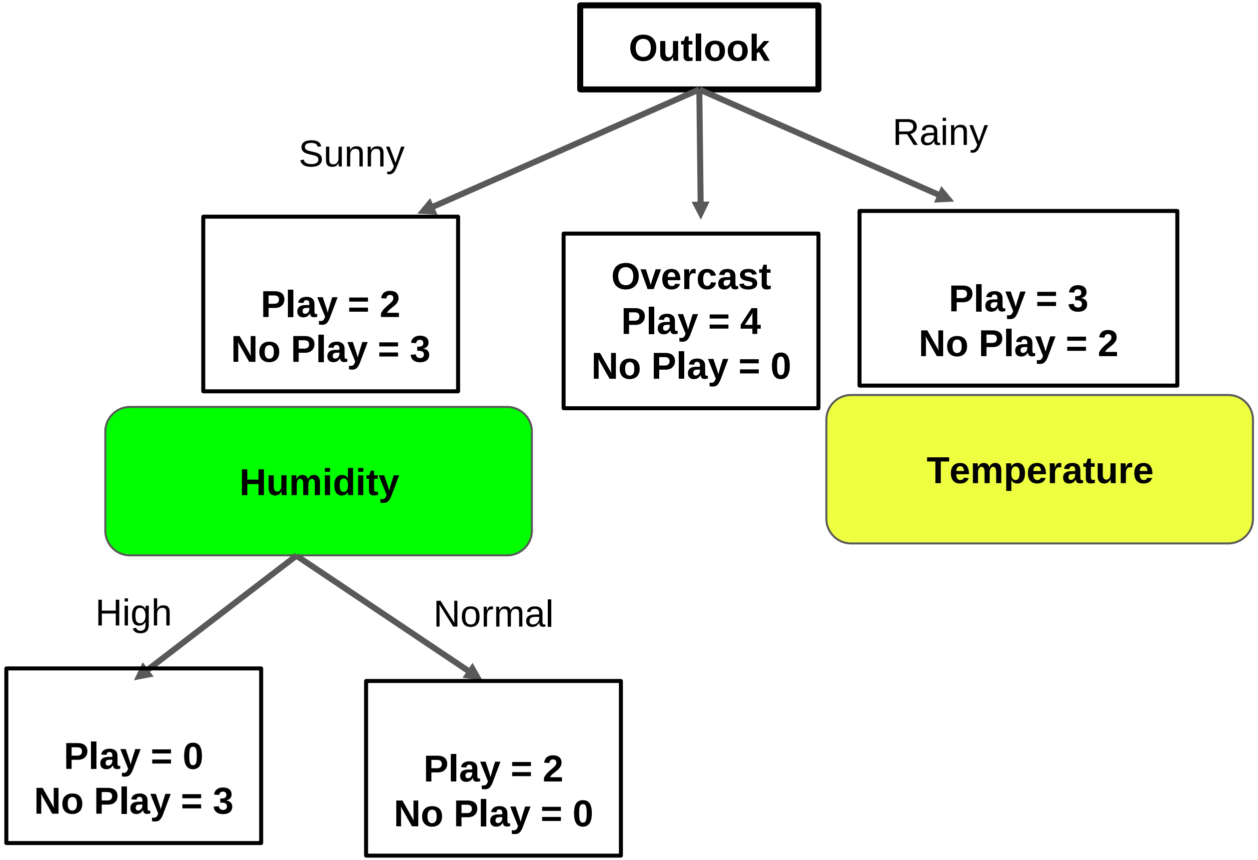


Decision Tree

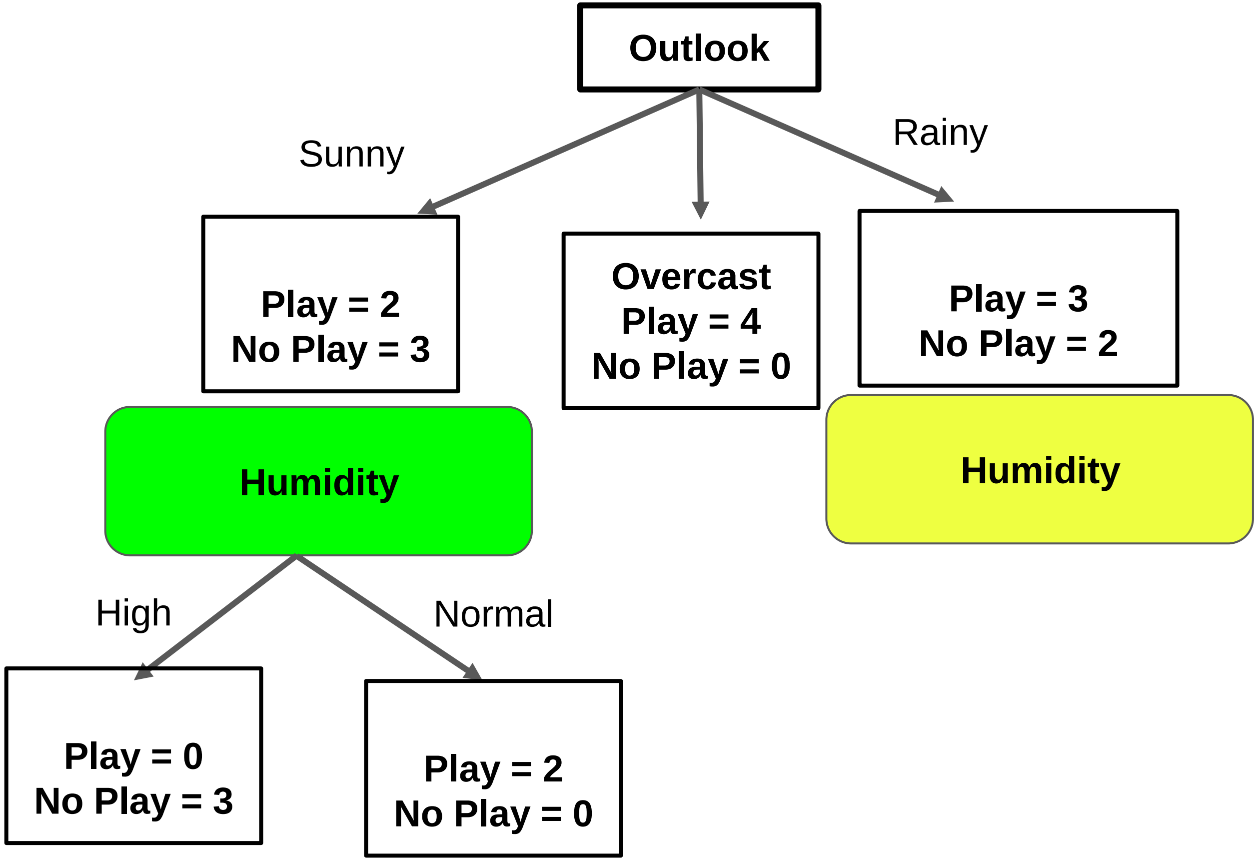
Step 6: If at any time, all instances at a node have the same classification: stop developing that part of the tree



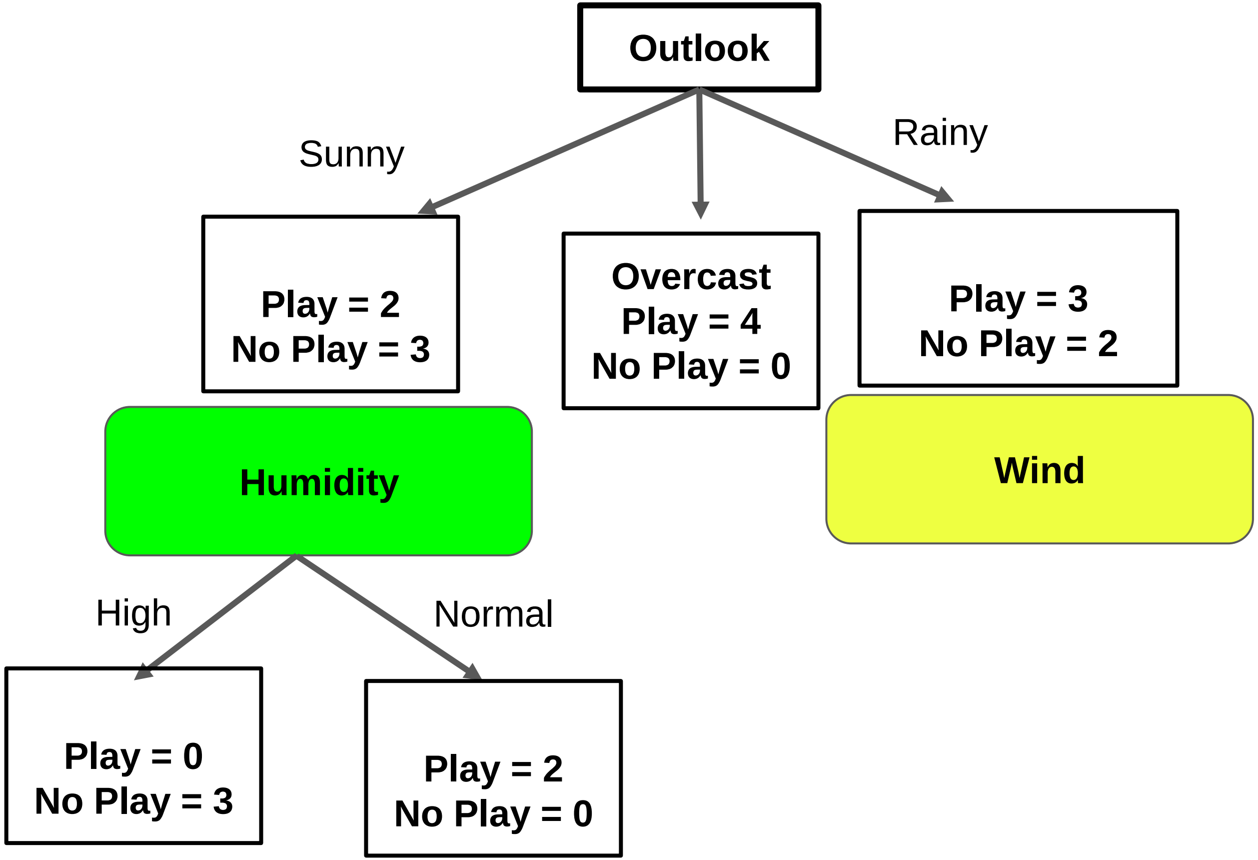
Decision Tree



Decision Tree



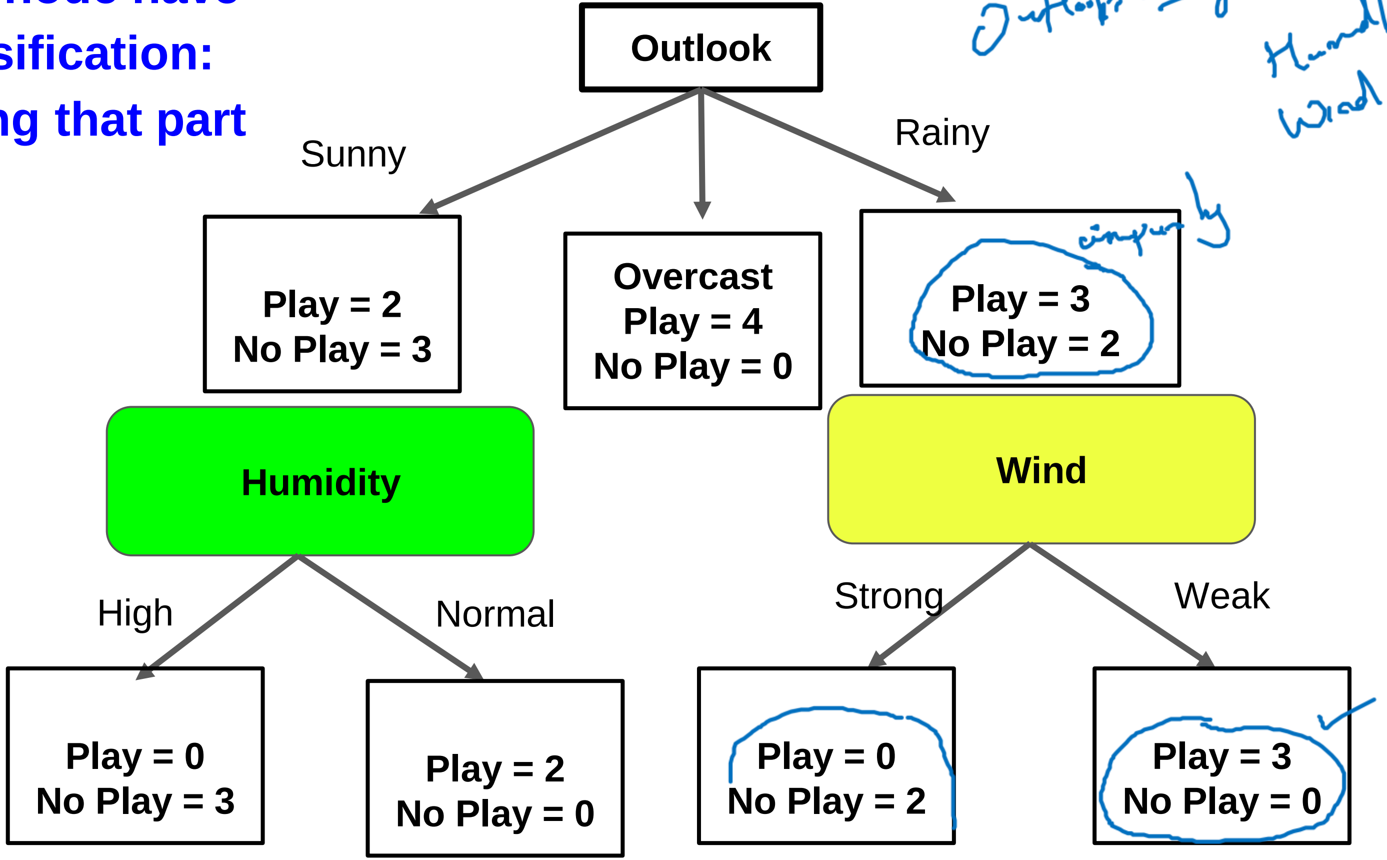
Decision Tree



Decision Tree

Step 6: If at any time, all instances at a node have the same classification: stop developing that part of the tree

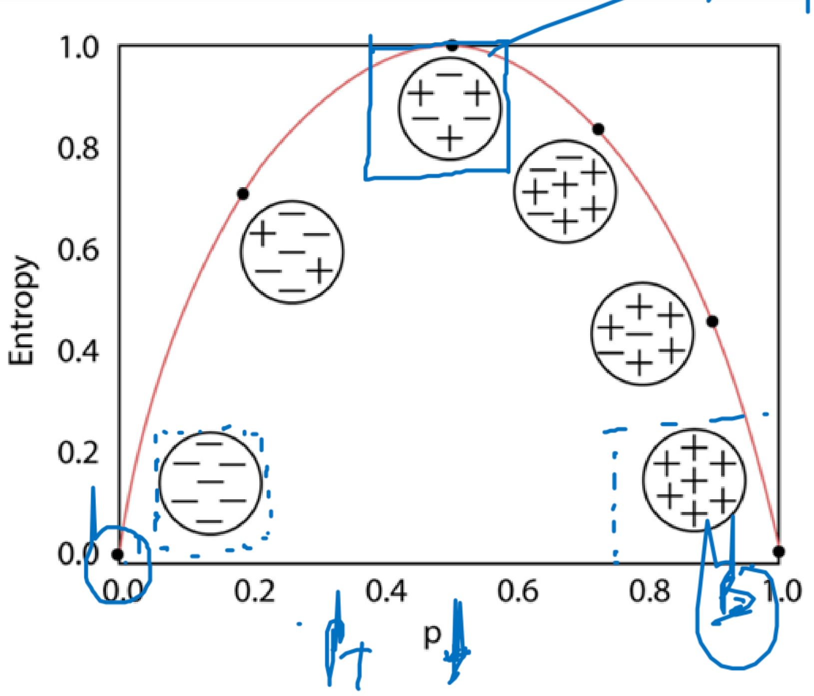
Outlook = Rainy, Temp = Hot, Humid = Normal, Wind Weak



Entropy and Information Gain

P(+ve sample) = 2/4 = 1/2

P(-ve sample) = 3/6 = 1/2



Suppose S is a set of class labels containing both positive (+) and negative (−) instances. For example,

$$S = \{+, +, -, -, +, -, -, +, -\}.$$

The entropy of S, denoted as $H(S)$, is a measure of uncertainty in the class distribution and is defined as:

$$H(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-,$$

where p_+ and p_- represent the proportions of positive and negative labels in S, respectively. The entropy is measured in bits.

p_+

$p = 0$

$p_+ = 1$

Image Source: https://blog.quantinsti.com/gini-index/

Find the Entropy of the Decision Variable?

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

Information Gain

$-P_N \log_2(P_N) - P_Y \log_2(P_Y)$

Information Gain measures **how well** a **given attribute** **separates the training examples according to their target classification**.

S

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

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

$S = \{N, N, Y, Y, Y, N, Y, N, Y, Y, Y, Y, Y, N\}$

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

$= \cancel{0.651}\ 0.94$

Information Gain for **A = Outlook**

Information Gain measures **how well** a **given attribute** **separates the training examples according to their target classification**.

H(S)

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

$$Gain(S, A = Outlook) = H(S) - \sum_{v \in Values(Outlook)} \frac{|S_v|}{|S|} H(S_v)$$

$$H(S) - \left[\frac{|S_{sunny}|}{|S|} \times H(S_{sunny}) + \frac{|S_{overcast}|}{|S|} H(S_{overcast}) + \frac{|S_{Rainfall}|}{|S|} H(S_{Rainfall}) \right]$$

H(S)

A = Outlook , How reduction in H(S) is possible)

A = Temp, how much reduction in H(S) is possible?

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

$H(S_v)$

$H(S)$

$S = \{1, 2, 3, 4, 5\}$

$|S| = 5$

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

Outlook

Sunny | Rainfall | Overcast

Total => 5

3 No 2Y

Prob $\frac{3}{5}$: No → 1

$\frac{2}{5}$: yes → 2

$$H(S_{sunny}) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)$$

$|Decision| = 14$

Information Gain for **A = Outlook**

$$\frac{5}{14} \times H(S_{sunny})$$

$$Gain(S, A = Outlook) = H(S) - \sum_{v \in Values(Outlook)} \frac{|S_v|}{|S|} H(S_v)$$

A = Outlook

cardinality of the set S_v

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

$P_{No} = \frac{3}{5}$

$P_{yes} = \frac{2}{5}$

Outlook

Sunny: $\frac{2}{5}$ yes, $\frac{3}{5}$ No

Overcast: 4 — 1 yes, 0 No

Rainfall: 5

$H(S_{sunny})$ $H(S_{overcast})$ $H(S_{Rainfall})$

S_v	$\lvert S_v \rvert$	P_{yes} count	P_{No} count	$H(S_v)$
S_{sunny}	5	~~3/5~~ 2	3	0.97
$S_{overcast}$	4	4	0	0
$S_{rainfall}$	5	3	2	0.97

$$H(S) - \left[\frac{5}{14} \times H(S_{sunny}) + \frac{4}{14} \times H(S_{overcast}) + \frac{5}{14} \times H(S_{Rainfall})\right]$$

Information Gain for **A = Outlook**

$$Gain(S, A = Outlook) = H(S) - \sum_{v \in Values(Outlook)} \frac{|S_v|}{|S|} H(S_v)$$

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

1. Entropy of "Outlook"

Outlook	Subset Count	Play=Yes	Play=No	Entropy
Sunny	5	2	3	$H(S_{\text{sunny}}) = -\left(\frac{2}{5}\log_2\frac{2}{5}\right) - \left(\frac{3}{5}\log_2\frac{3}{5}\right) \approx 0.971$
Overcast	4	4	0	$H(S_{\text{overcast}}) = 0$
Rainy	5	3	2	$H(S_{\text{rainy}}) = -\left(\frac{3}{5}\log_2\frac{3}{5}\right) - \left(\frac{2}{5}\log_2\frac{2}{5}\right) \approx 0.971$

$$H(S_{\text{Outlook}}) = \frac{5}{14}(0.971) + \frac{4}{14}(0) + \frac{5}{14}(0.971) \approx 0.693$$

$$IG(\text{Outlook}) = H(S) - H(S_{\text{Outlook}}) = 0.94 - 0.693 = 0.247$$

Information Gain for A = Temperature

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

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

S_v	$\|S_v\|$	P_{yes} count	P_{No}	$H(S_v)$
S_{HOT}	4	2	2	1
S_{MILD}	6	4	2	0.92
S_{COOL}	4	3	1	0.89

Information Gain for A = Temperature

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

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

2. Entropy of "Temperature"

Temperature	Subset Count	Play=Yes	Play=No	Entropy
Hot	4	2	2	$H(S_{\text{hot}}) = 1.0$
Mild	6	4	2	$H(S_{\text{mild}}) = 0.918$
Cool	4	3	1	$H(S_{\text{cool}}) = 0.811$

$$H(S_{\text{Temperature}}) = \boxed{\frac{4}{14}(1.0) + \frac{6}{14}(0.918) + \frac{4}{14}(0.811) \approx 0.911}$$

$$IG(\text{Temperature}) = H(S) - H(S_{\text{Temperature}}) = 0.94 - 0.911 = 0.029$$

$\Rightarrow 0.94 - 0.911 = \underline{0.029}$

Information Gain for **A = Humidity**

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

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

S_v	$\|S_v\|$	P_{yes}	P_{No}	$H(S_v)$
S_{HIGH}	7	3	4	0.985
S_{NORMAL}	7	6	1	0.591

Information Gain for **A = Humidity**

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

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

3. Entropy of "Humidity"

Humidity	Subset Count	Play=Yes	Play=No	Entropy
High	7	3	4	$H(S_{\text{high}}) = 0.985$
Normal	7	6	1	$H(S_{\text{normal}}) = 0.592$

$$H(S_{\text{Humidity}}) = \frac{7}{14}(0.985) + \frac{7}{14}(0.592) \approx 0.789$$

$$\boxed{IG(\text{Humidity}) = H(S) - H(S_{\text{Humidity}}) = 0.94 - 0.789 = 0.151}$$

Information Gain for **A = Wind**

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

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

S_v	$\|S_v\|$	P_{yes}	P_{No}	$H(S_v)$
S_{Weak}				
S_{Strong}				

IG = 0.048

Information Gain for **A = Wind**

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

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

4. Entropy of "Wind"

Wind	Subset Count	Play=Yes	Play=No	Entropy
Weak	8	6	2	$H(S_{\text{weak}}) = 0.811$
Strong	6	3	3	$H(S_{\text{strong}}) = 1.0$

$$H(S_{\text{Wind}}) = \frac{8}{14}(0.811) + \frac{6}{14}(1.0) \approx 0.892$$

$$IG(\text{Wind}) = H(S) - H(S_{\text{Wind}}) = 0.94 - 0.892 = 0.048$$

Summary of First Split

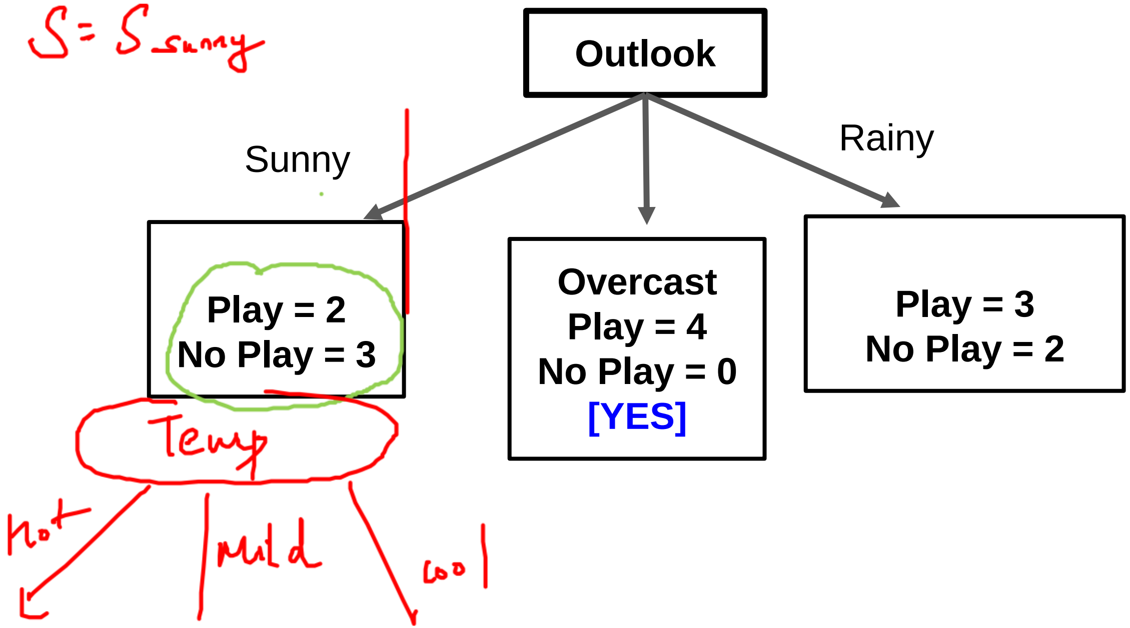
Step 3: Information Gain Summary Table

Attribute	Entropy	Information Gain
Outlook	0.693	**0.247**
Temperature	0.911	0.029
Humidity	0.789	0.151
Wind	0.892	0.048

Step 4: Choosing the Best Split

Since **Outlook** has the highest Information Gain (0.247), it is chosen as the root node for the decision tree.

Decision Tree



$S = S_{outlook}$, Find H(S)?

$S = S_{sunny}$. Find $H(S)$

$H(S)$

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

$S = S_{outlook}$, Find H(S)?

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

Subset for Outlook = Sunny

Temperature	Humidity	Wind	Play Tennis
Hot	High	Weak	No
Hot	High	Strong	No
Mild	High	Weak	No
Cool	Normal	Weak	Yes
Mild	Normal	Strong	Yes

The entropy for this subset is:

$$H(S_{\text{Sunny}}) = -\left(\frac{2}{5}\log_2\frac{2}{5}\right) - \left(\frac{3}{5}\log_2\frac{3}{5}\right)$$

$$= -(0.4 \times (-1.322)) - (0.6 \times (-0.737))$$

$$= 0.971$$

$S = S_{\cancel{\text{outlook}}\ \text{sunny}}$, A = Temperature

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

S_v	$\lvert S_v \rvert$	P_{yes}	P_{No}	$H(S_v)$
S_{HOT}	2	0	2	0
S_{MILD}	2	1	1	1.
S_{COOL}	1	1.	0	0

$S = S_{outlook}$, A = Temperature

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

Compute Information Gain for Each Attribute

1. Temperature

Temperature	Play Yes	Play No	Entropy
Hot	0	2	0
Mild	1	1	1.0
Cool	1	0	0

$$H(S_{\text{Temperature}}) = \frac{2}{5}(0) + \frac{2}{5}(1.0) + \frac{1}{5}(0)$$

$$= 0 + 0.4 + 0 = 0.4$$

$$IG(\text{Temperature}) = 0.971 - 0.4 = \underline{\underline{0.571}}$$

$S = S_{outlook}$, A = Humidity

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

S_v	$\|S_v\|$	P_{yes}	P_{No}	$H(S_v)$
S_{HIGH}	3	0	3	0
S_{NORMAL}	2	2	0	0

$S = S_{outlook}$, A = Humidity

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

2. Humidity

Humidity	Play Yes	Play No	Entropy
High	0	3	0
Normal	2	0	0

$$H(S_{\text{Humidity}}) = \frac{3}{5}(0) + \frac{2}{5}(0) = 0$$

$$IG(\text{Humidity}) = 0.971 - 0 = 0.971$$

IG max

$S = S_{outlook}$, A = Wind

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

S_v	$\|S_v\|$	P_{yes}	P_{No}	$H(S_v)$
S_{Weak}	3	1	2	
S_{Strong}	2	1	1	

$S = S_{outlook}$, $A = Wind$

Day	outlook	temperature	humidity	wind	Decision
1	sunny	hot	high	weak	No
2	sunny	hot	high	strong	No
3	overcast	hot	high	weak	Yes
4	rainfall	mild	high	weak	Yes
5	rainfall	cool	normal	weak	Yes
6	rainfall	cool	normal	strong	No
7	overcast	cool	normal	wtrong	Yes
8	sunny	mild	high	weak	No
9	sunny	cool	normal	weak	Yes
10	rainfall	mild	normal	weak	Yes
11	sunny	mild	normal	strong	Yes
12	overcast	mild	high	strong	Yes
13	overcast	hot	normal	weak	Yes
14	rainfall	mild	high	strong	No

3. Wind

Wind	Play Yes	Play No	Entropy
Weak	1	2	0.918
Strong	1	1	1.0

$$H(S_{\text{Wind}}) = \frac{3}{5}(0.918) + \frac{2}{5}(1.0)$$

$$= 0.5508 + 0.4 = 0.9508$$

$$IG(\text{Wind}) = 0.971 - 0.9508 = 0.020$$

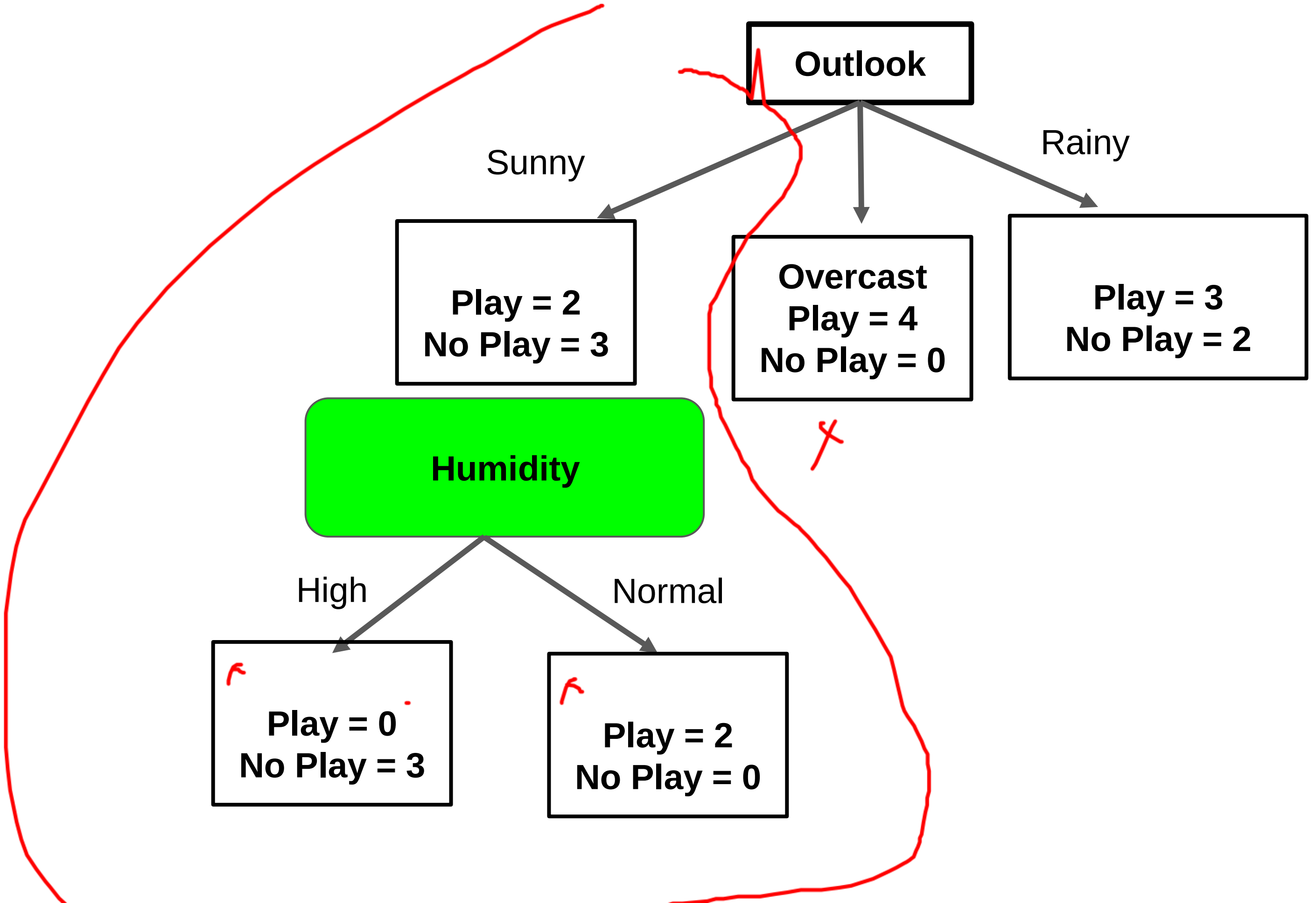
Summary of the Split

Information Gain Summary for Sunny Subset

Attribute	Entropy	Information Gain
Temperature	0.4	0.571
Humidity	0	**0.971**
Wind	0.9508	0.020

Since **Humidity has the highest Information Gain (0.971)**, it is the best attribute to split on.

Decision Tree



$S = S_{rainfall}$, A = Temperature

$S = S_{rainfall}$, A = Humidity

$S = S_{rainfall}$, A = Wind

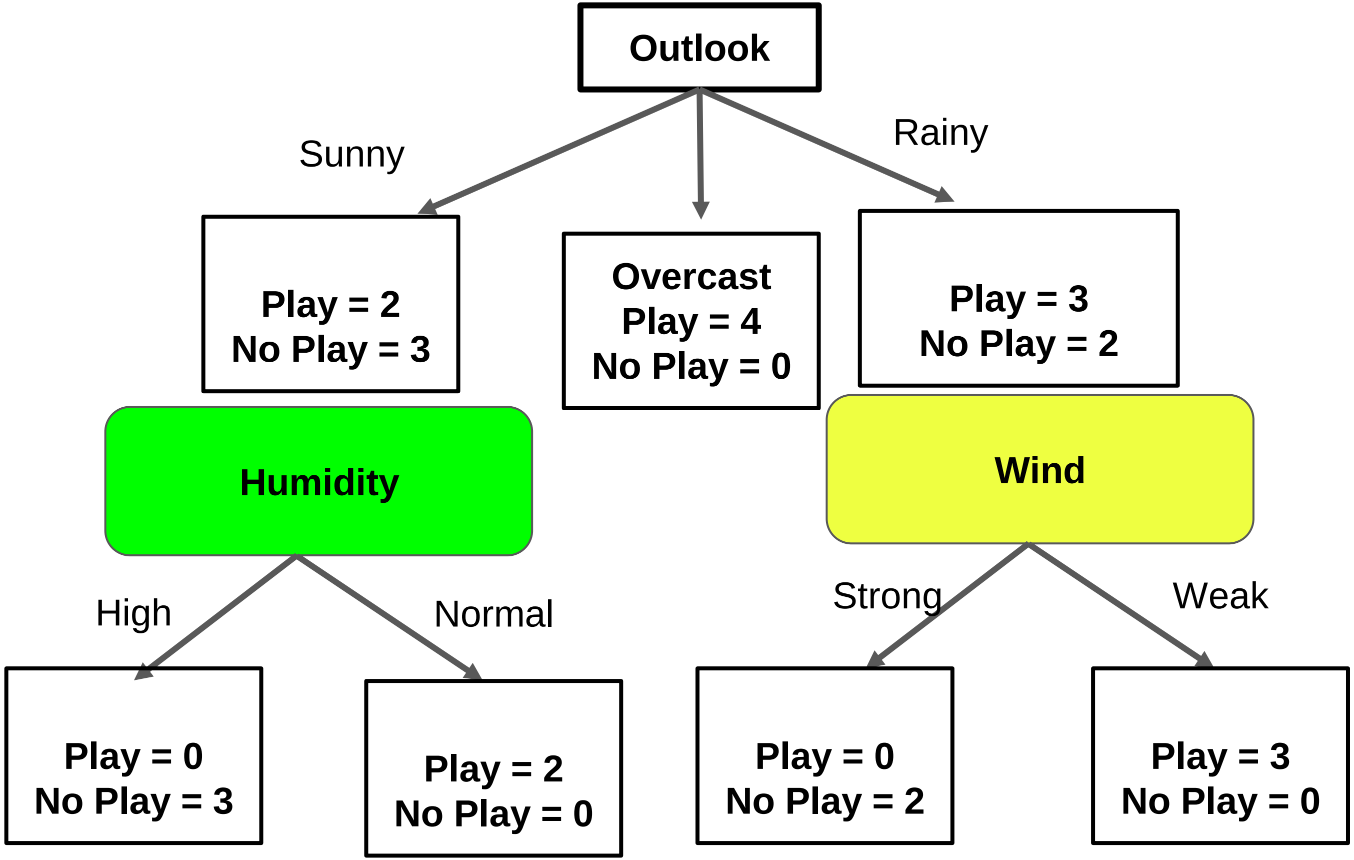
Summary of the Split

Information Gain Summary for Rainy Subset

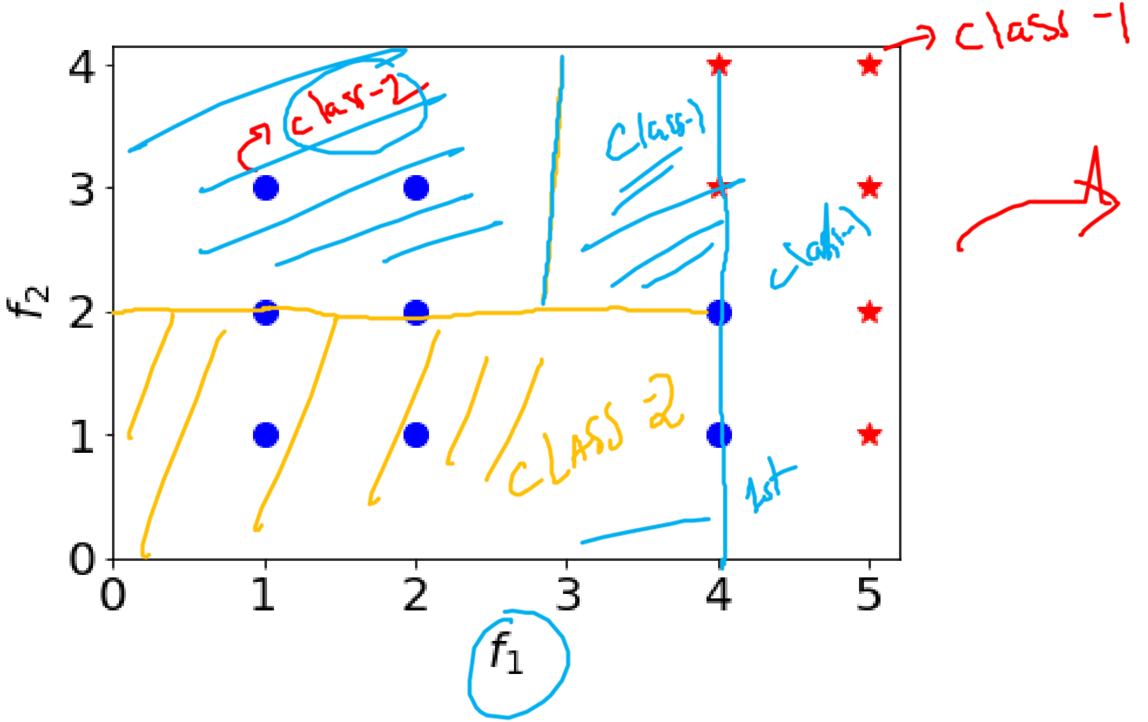
Attribute	Entropy	Information Gain
Temperature	0.9508	0.020
Humidity	0.9508	0.020
Wind	0	**0.971**

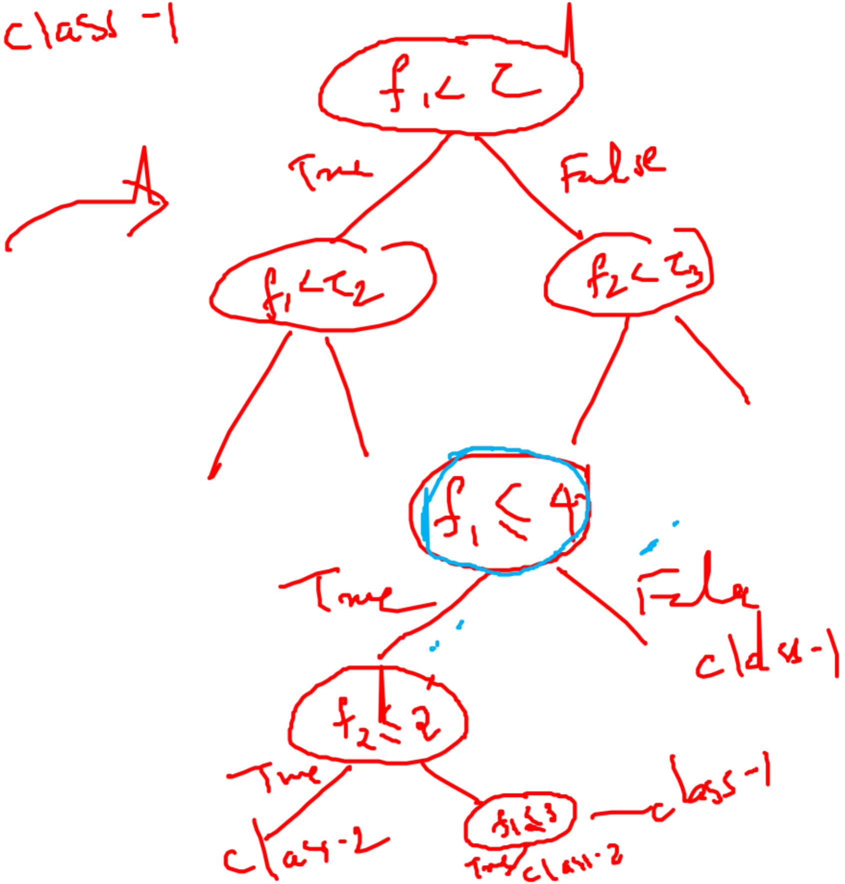
Since **Wind has the highest Information Gain (0.971)**, it is the best attribute to split on.

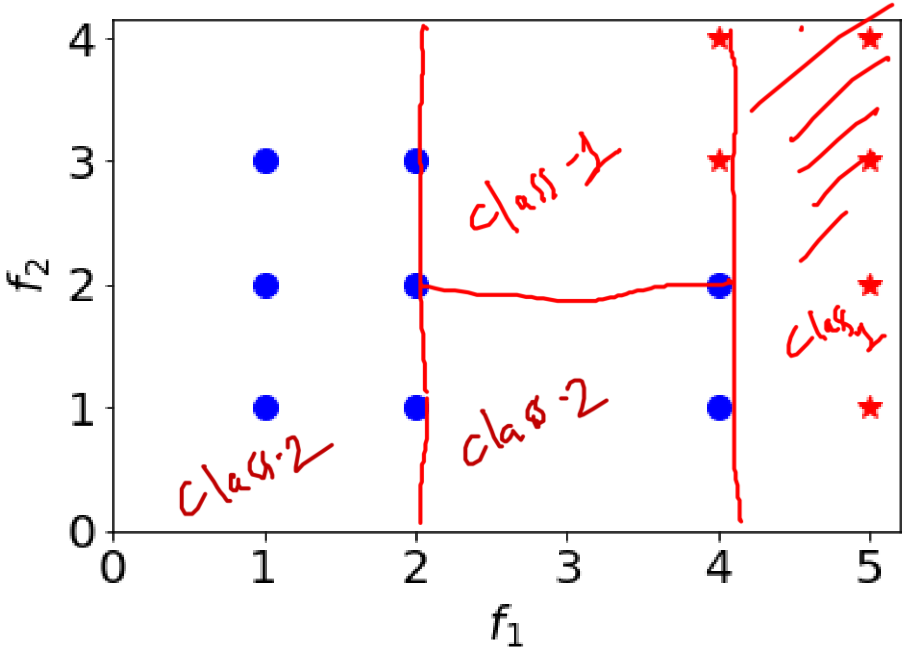
Decision Tree



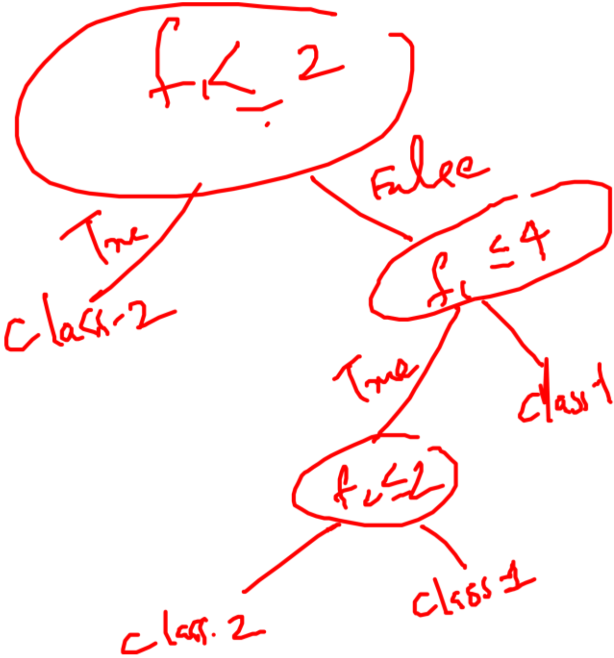
Decision Tree for Continuous Variables







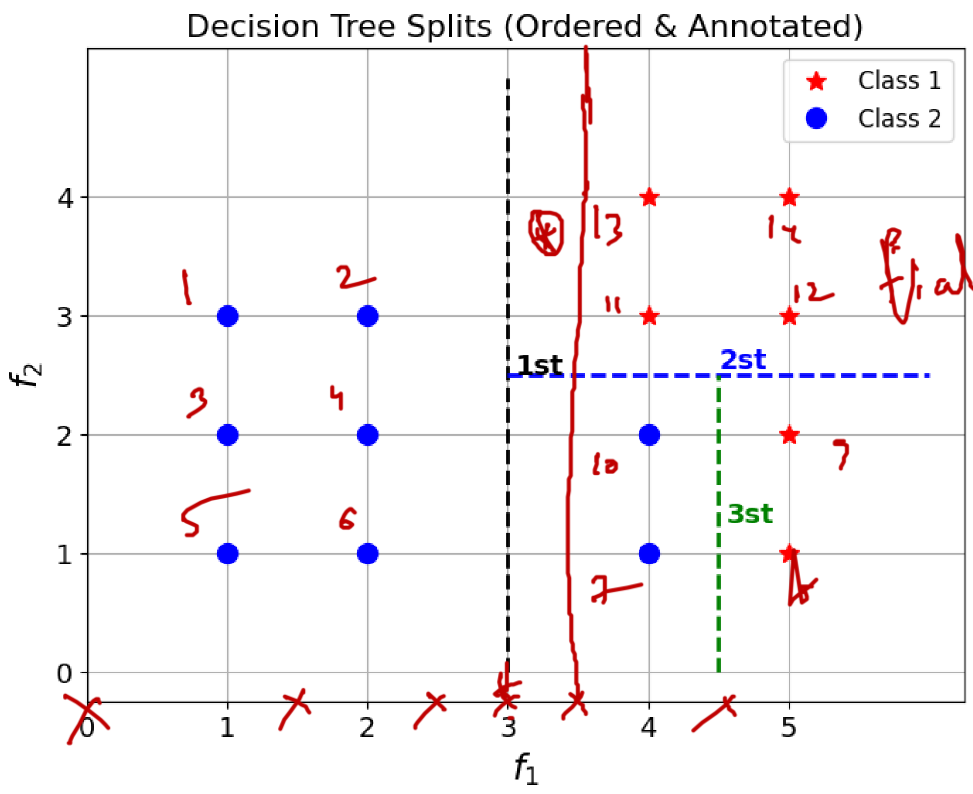
f_2
f_1
0 1 2 3 4 5
0 1 2 3 4
class-1
class-2
class-2
class-1

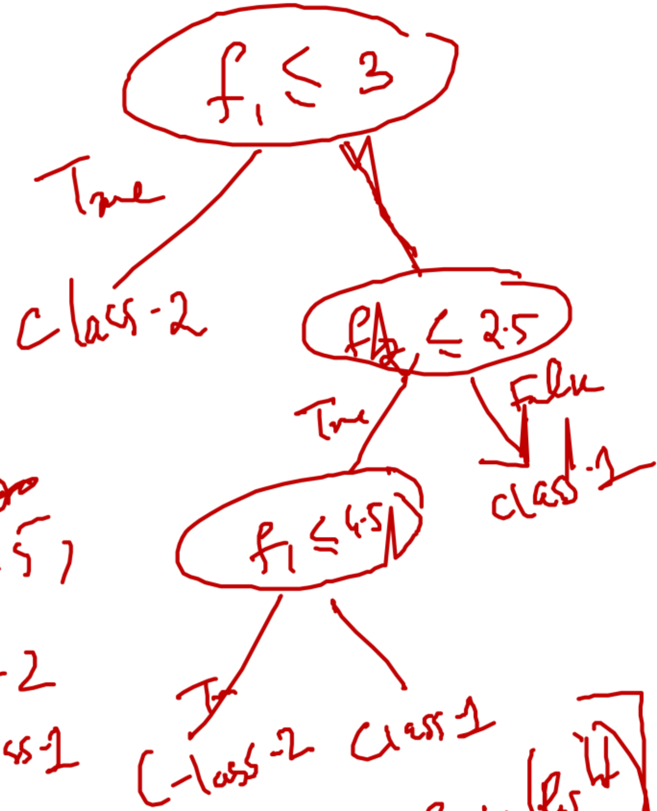


$f_1 \leq 2$
True
False
class-2
$f_1 \leq 4$
True
class-1
$f_2 \leq 2$
class-2
class-1

Decision Tree Splits

f_1: 0 to 5
f_2: 0 to 5



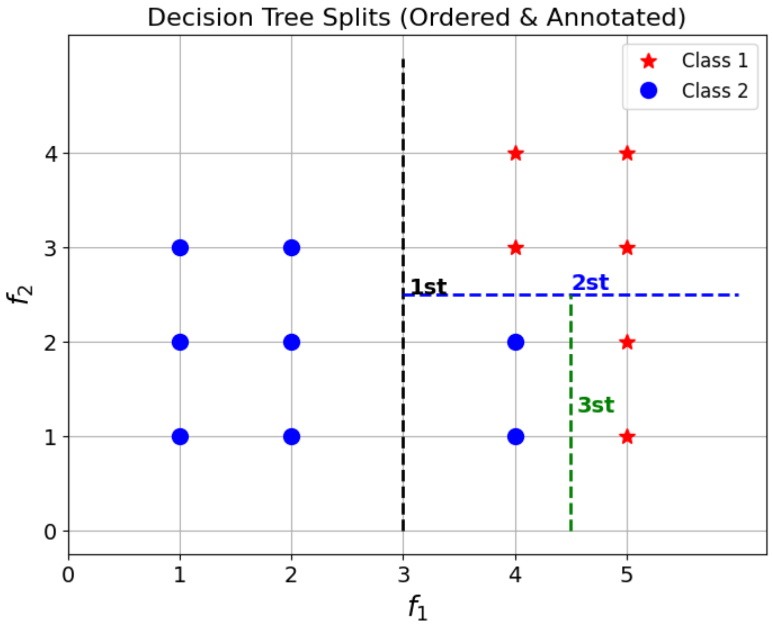


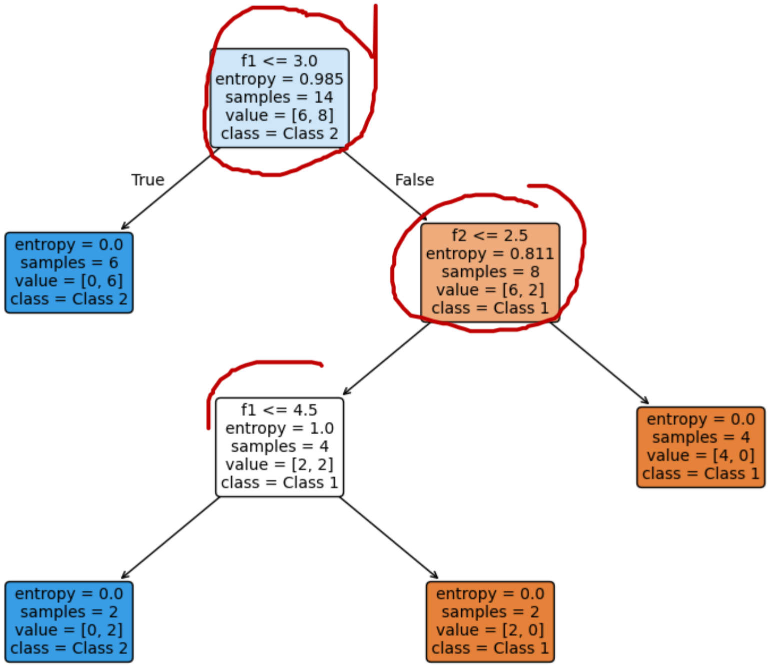
f_1 calculate

8 : class-2
6 : class-1

$$H(S) - \left[\frac{6}{14} H(\text{left}) + \frac{8}{14} H(\text{right})\right]$$

Decision Tree Splits





Hyperparameters in Decision Tree

criterion{“gini”, “entropy”, “log_loss”}, **default=”gini”**

max_depth **int, default=None**

for crits

for max-dep lb

for min-samples split

for min-samples leaf

min_samples_split : int or float, default=2

The minimum number of samples required to split an internal node:

- If int, then consider `min_samples_split` as the minimum number.
- If float, then `min_samples_split` is a fraction and `ceil(min_samples_split * n_samples)` are the minimum number of samples for each split.

min_samples_leafint or float, default=1

The minimum number of samples required to be at a leaf node. A split point at any depth will only be considered if it leaves at least `min_samples_leaf` training samples in each of the left and right branches. This may have the effect of smoothing the model, especially in regression.

- If int, then consider `min_samples_leaf` as the minimum number.
- If float, then `min_samples_leaf` is a fraction and `ceil(min_samples_leaf * n_samples)` are the minimum number of samples for each node.

max_features : **int, float or {“sqrt”, “log2”}, default=None**

1️⃣ `min_samples_split` (Controls When to Split a Node)

- Defines the **minimum number of samples required** to **split** an internal node.
- If a node has **fewer** than `min_samples_split` samples, it **won't split** further.
- Helps **prevent overfitting** by stopping unnecessary splits.

Examples

- `min_samples_split=5` → A node **must have at least 5 samples** to split.
- `min_samples_split=0.2` → A node **must have at least 20% of total samples** to split.

🔹 **Key Effect**: Controls the tree's depth by **stopping premature splits**.

2 `min_samples_leaf` (Controls Minimum Leaf Size)

- Defines the **minimum number of samples required** in a **leaf node** (final node after splitting).
- Prevents the tree from creating **tiny, unstable leaf nodes**.
- Useful for **smoothing predictions** in regression.

Examples

- `min_samples_leaf=2` → Each **leaf node must have at least 2 samples**.
- `min_samples_leaf=0.1` → Each **leaf must contain at least 10% of total samples**.

◆ **Key Effect**: Prevents **small leaves** and ensures meaningful splits.

HW

3 feature

Task:
Using the **Breast Cancer Wisconsin Dataset**, perform **Principal Component Analysis (PCA)** to reduce the feature dimensions to **2**. Then, conduct **hyperparameter tuning** using **3-fold cross-validation** to optimize the values of:

- `max_depth` (maximum depth of the tree)
- `min_samples_split` (minimum samples required to split a node)
- `min_samples_leaf` (minimum samples required in a leaf node)

Once the best hyperparameters are identified, evaluate the model's performance on the **test dataset**.

Report Accuracy, Precision, Recall, F1-score.