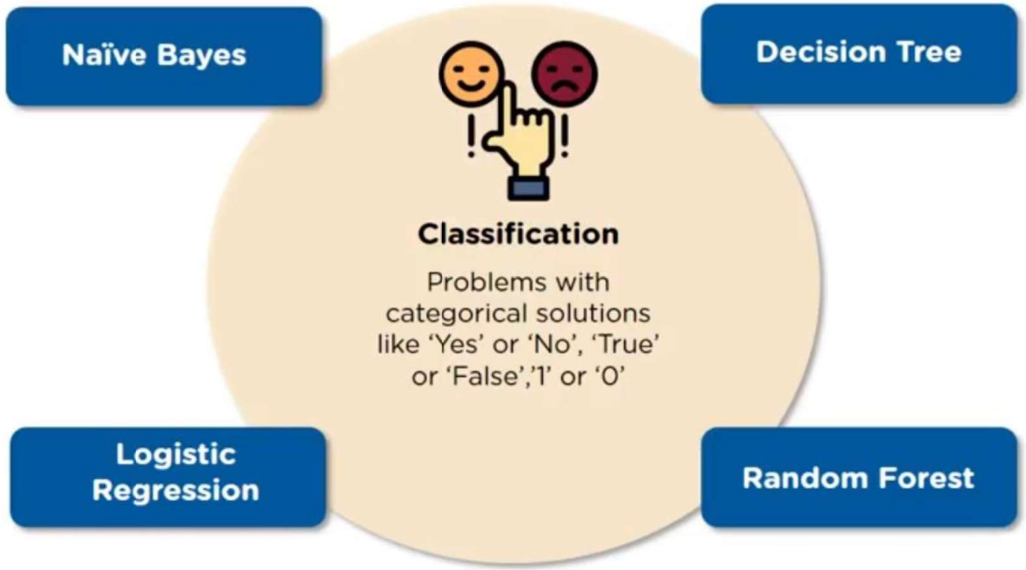
Problems in Machine Learning

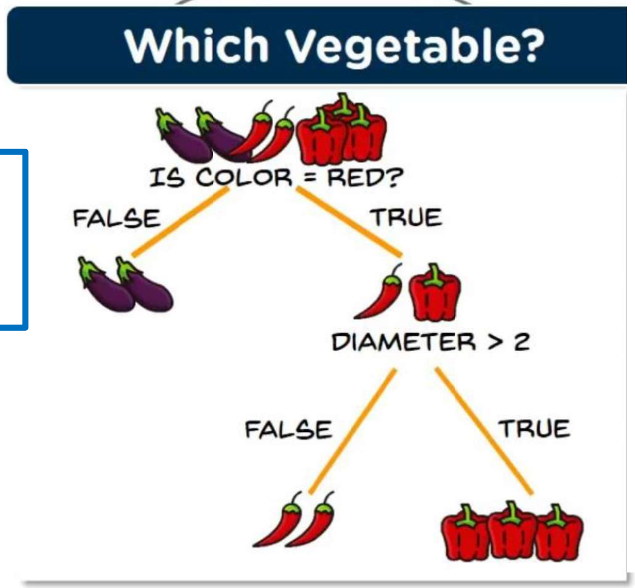


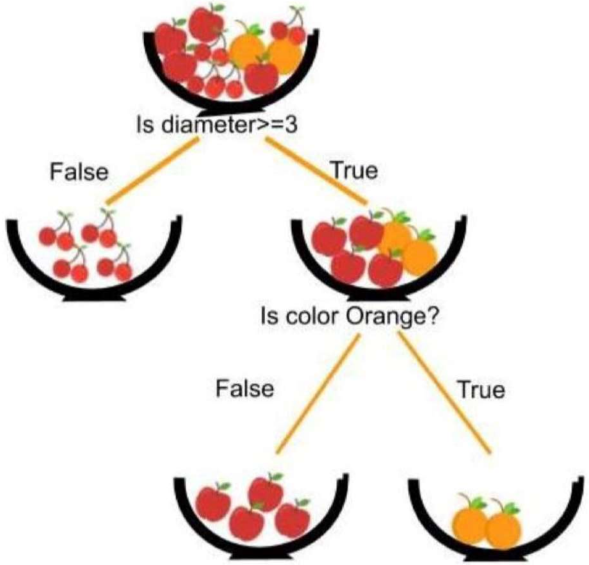
Random forest is simply a collection of decision trees whose results are aggregated into final result.

What is Decision Tree?

Decision Tree is a tree shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, occurrence or reaction

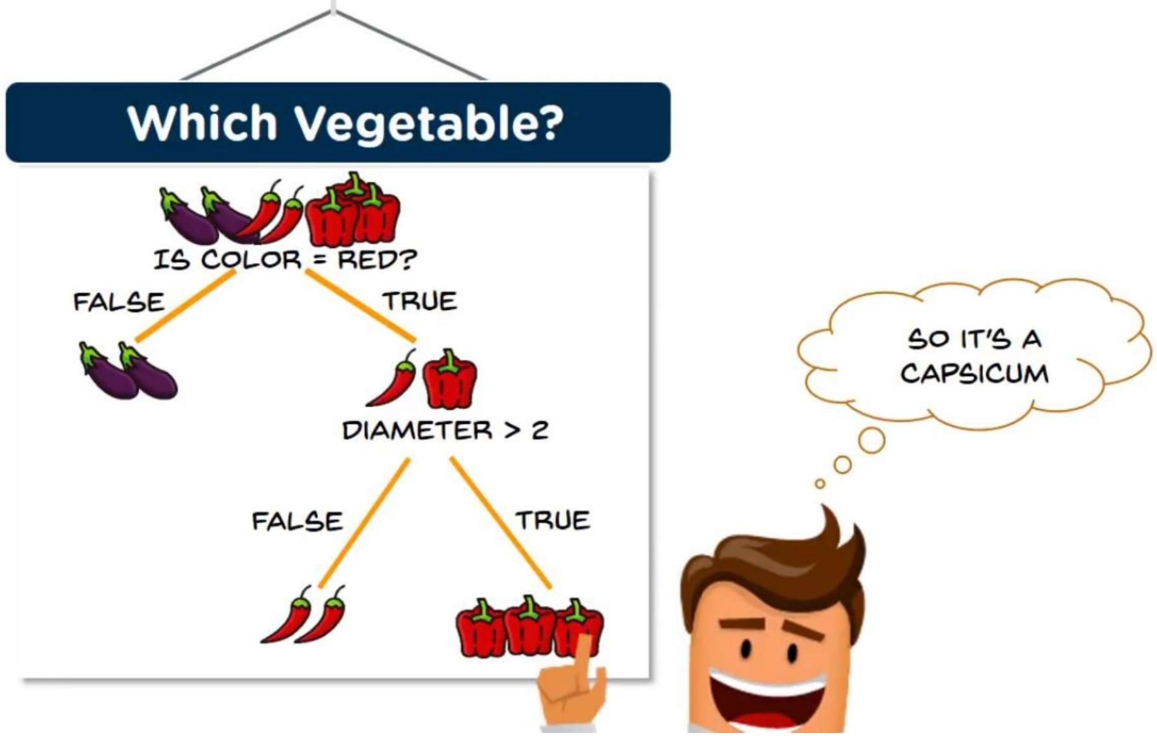
How do I identify a random vegetable from a shopping bag?

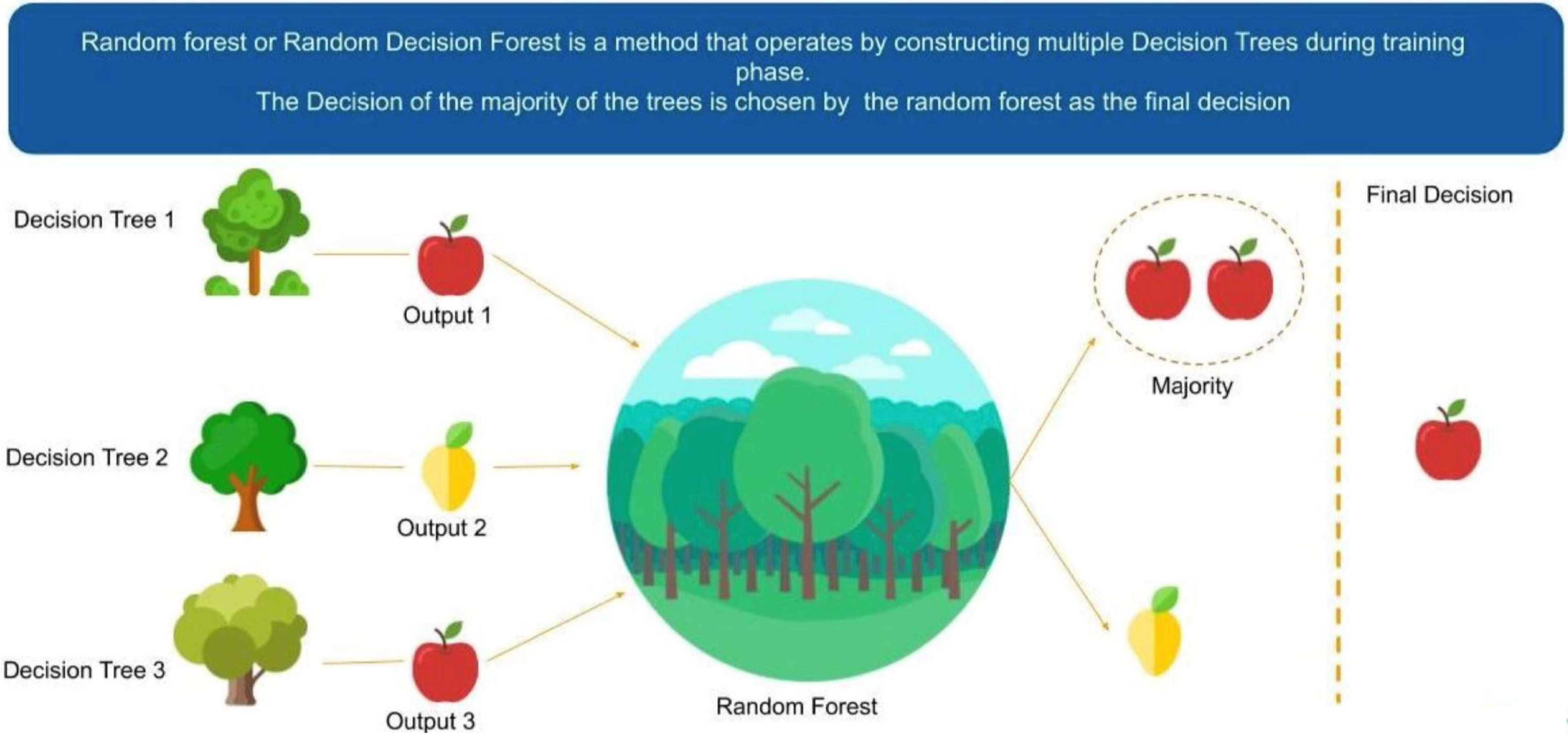




What is Decision Tree?

Decision Tree is a tree shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, occurrence or reaction

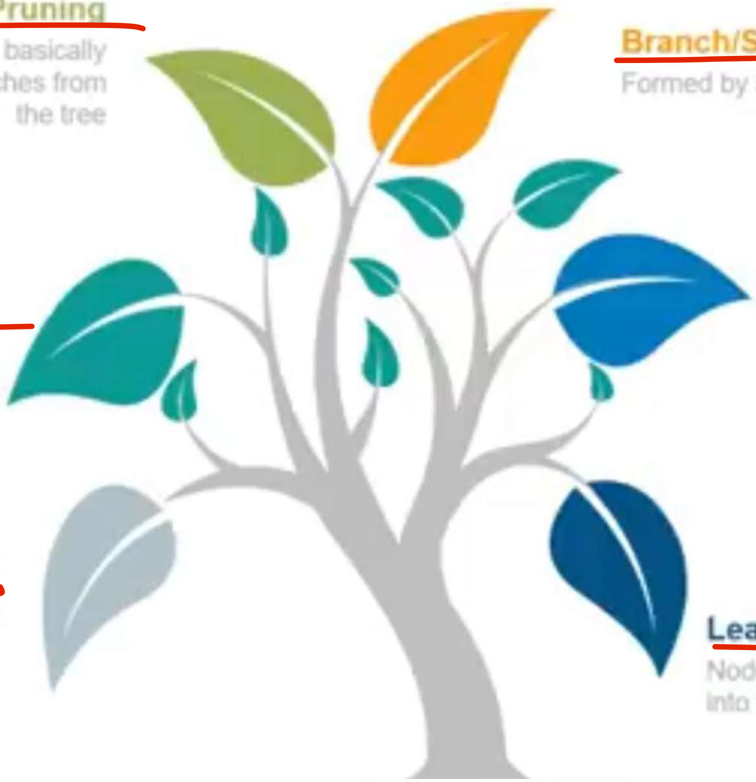



Random forest or Random Decision Forest is a method that operates by constructing multiple Decision Trees during training phase.
The Decision of the majority of the trees is chosen by the random forest as the final decision
Decision Tree 1
Output 1
Decision Tree 2
Output 2
Decision Tree 3
Output 3
Random Forest
Majority
Final Decision

Decision Tree Terminology

Pruning
Opposite of Splitting, basically removing unwanted branches from the tree

Branch/SubTree
Formed by splitting the tree/node



Parent/Child Node
Root node is the parent node and all the other nodes branched from it is known as child node

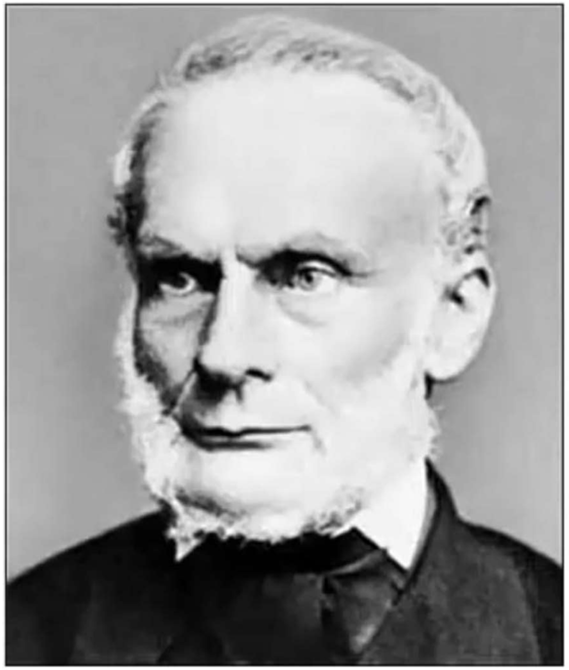
Splitting
Splitting is dividing the root node/sub node into different parts on the basis of some condition.

Root Node
It represents the entire population or sample and this further gets divided into two or more homogenous sets.

Leaf Node
Node cannot be further segregated into further nodes

Entropy and Information Gain

Inventor: **Rudolf Clausius**



We will use the concept of entropy, borrowed from physics, to build a classification method.

Entropy is a measure of disorder in a dataset.

Information gain a measure of the decrease in disorder achieved by partitioning the original data set.

$$Entropy(S) \equiv -p_+ \log_2 p_+ - p_- \log_2 p_-$$

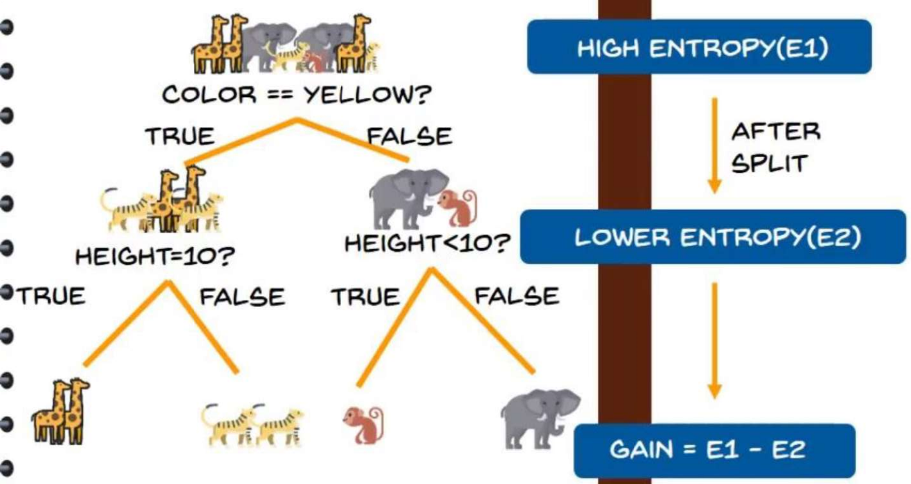
$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Decision Tree - Important Terms

INFORMATION GAIN

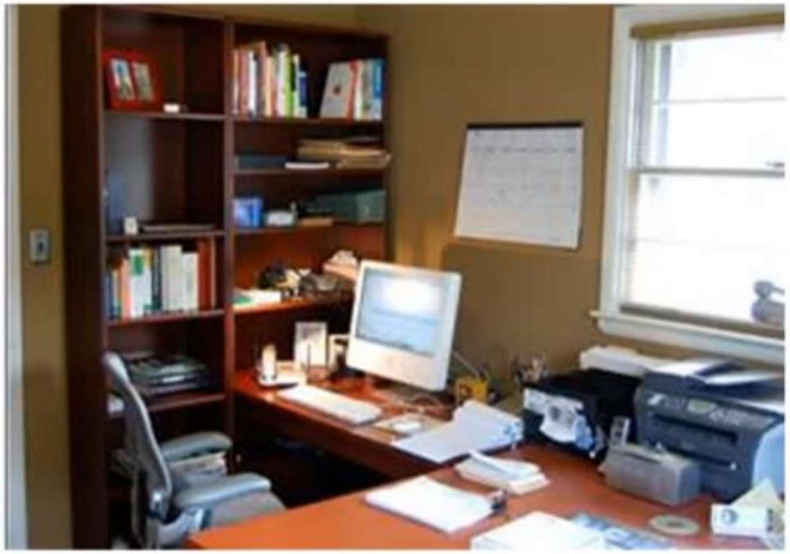
IT IS THE MEASURE OF DECREASE IN ENTROPY AFTER THE DATASET IS SPLIT

EXAMPLE

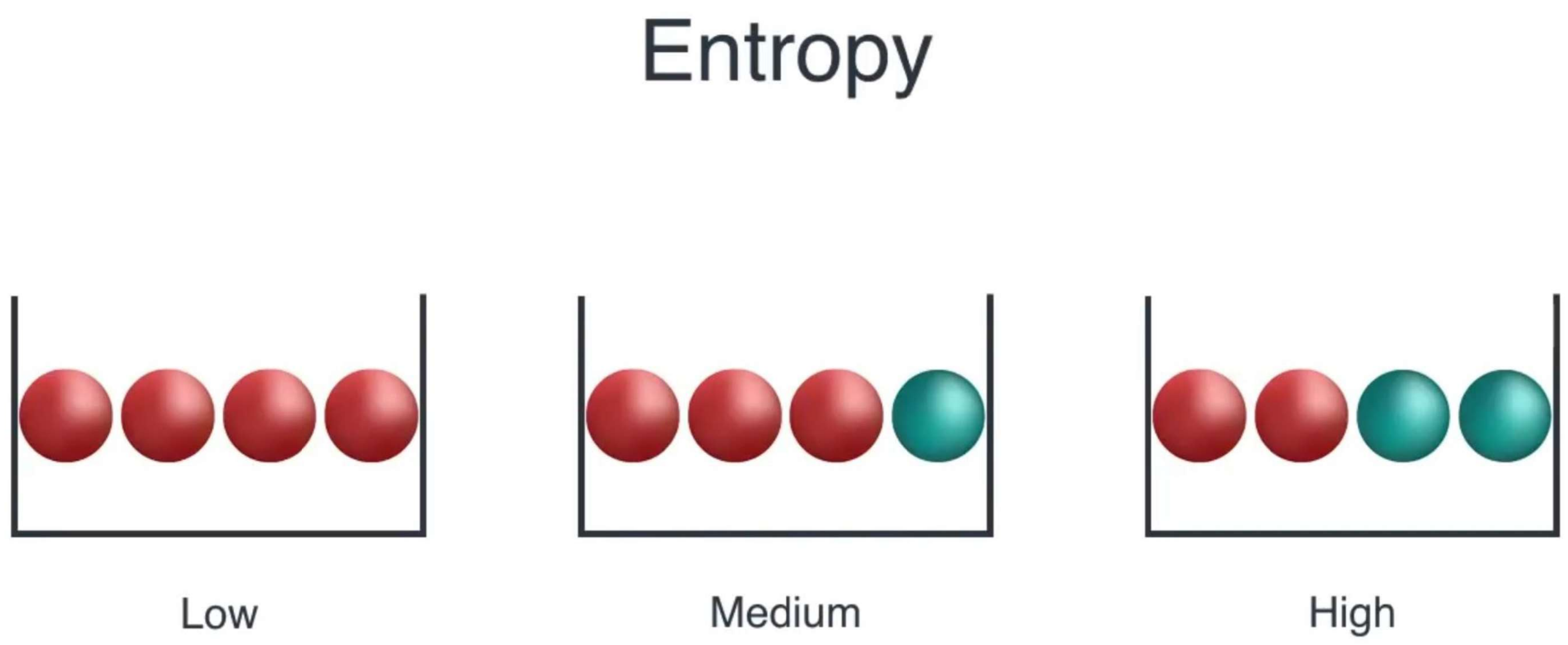


Entropy

High Entropy (messy)



Low Entropy (Clean)



Pure Set

Impure Set

$\frac{4}{4} = 1$

$\frac{0}{4} = 0$

Entropy

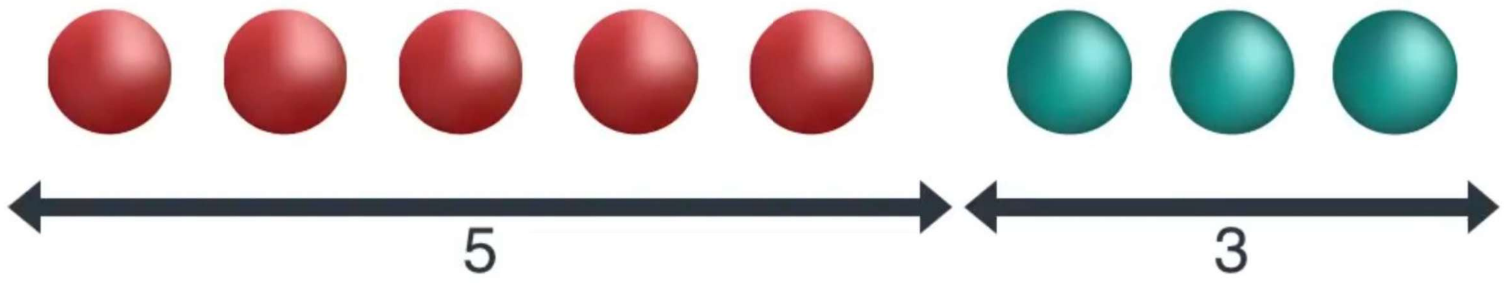
	P(red)	P(blue)	P(winning)	$-\log_2$(P(winning))	Entropy
	1	0	1 × 1 × 1 × 1 = 1	0+0+0+0	0
	0.75	0.25	0.75 × 0.75 × 0.75 × 0.25 = 0.105	0.415+0.415+0.415+2	0.81
	0.5	0.5	0.5 × 0.5 × 0.5 × 0.5 = 0.0625	1+1+1+1	1

$\frac{2}{4}$

$0.75 = \frac{3}{4}$

$\frac{1}{4} = 0.25$

Entropy



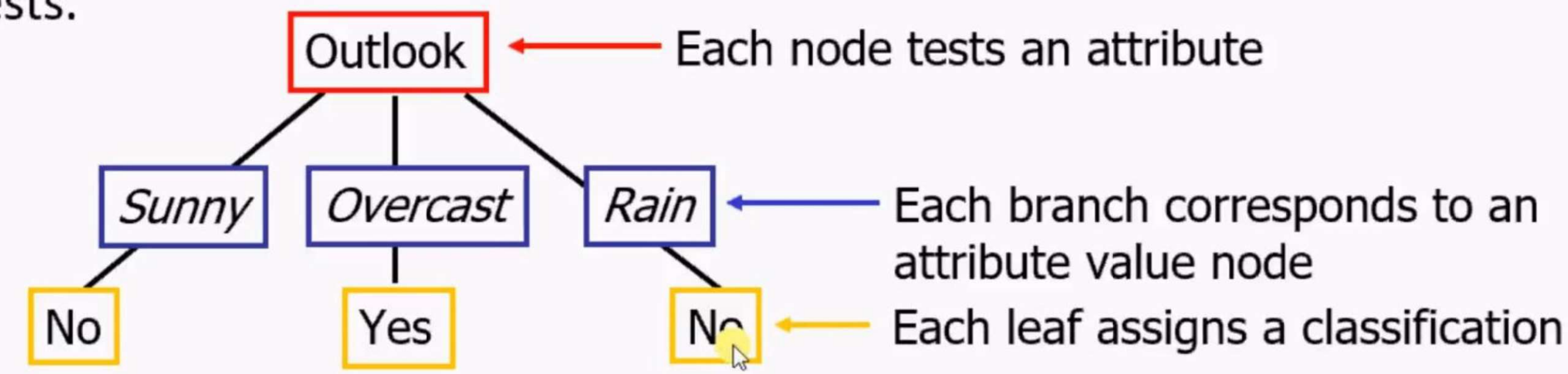
$$\text{Entropy} = -\frac{5}{8}\log_2\left(\frac{5}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) = 0.9544$$

Decision Tree

- A flow-chart-like tree structure
- Internal node denotes a test on an attribute
- Branch represents an outcome of the test
- Leaf nodes represent class labels or class distribution
- Constructed in a top down recursive divide and conquer approach

Decision Tree Learning

- A **Decision tree** represents a function that takes as input a vector of attribute values and returns a "decision"—a single output value.
- The input and output values can be discrete or continuous.
- A decision tree reaches its decision by performing a sequence of tests.



Decision Tree Learning Algorithm

- There are many specific decision-tree algorithms. Notable ones include:
- ID3 (Iterative Dichotomiser 3)
- C4.5 (successor of ID3)
- CART (Classification And Regression Tree)
- CHAID (Chi-squared Automatic Interaction Detector). Performs multi-level splits when computing classification trees.
- MARS: extends decision trees to handle numerical data better.

(Multivariate adaptive regression splines)

Definition: Entropy

MACHINE LEARNING

- It is a Measurement of Homogeneity of Examples
- Given a collection S, containing +ve and -ve examples of some target concept, the entropy of S is given by

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

- where $\mathbf{p}_+$ is the proportion of positive examples in S and $\mathbf{p}_-$ is the proportion of negative examples in S
- In all calculations involving entropy we define 0.log 0 = 0.
- In general for c class classification

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

Definition: Information Gain

MACHINE LEARNING

- It is the expected reduction in entropy caused by partitioning the examples according to some attribute A
- Split the node with attribute having highest Gain

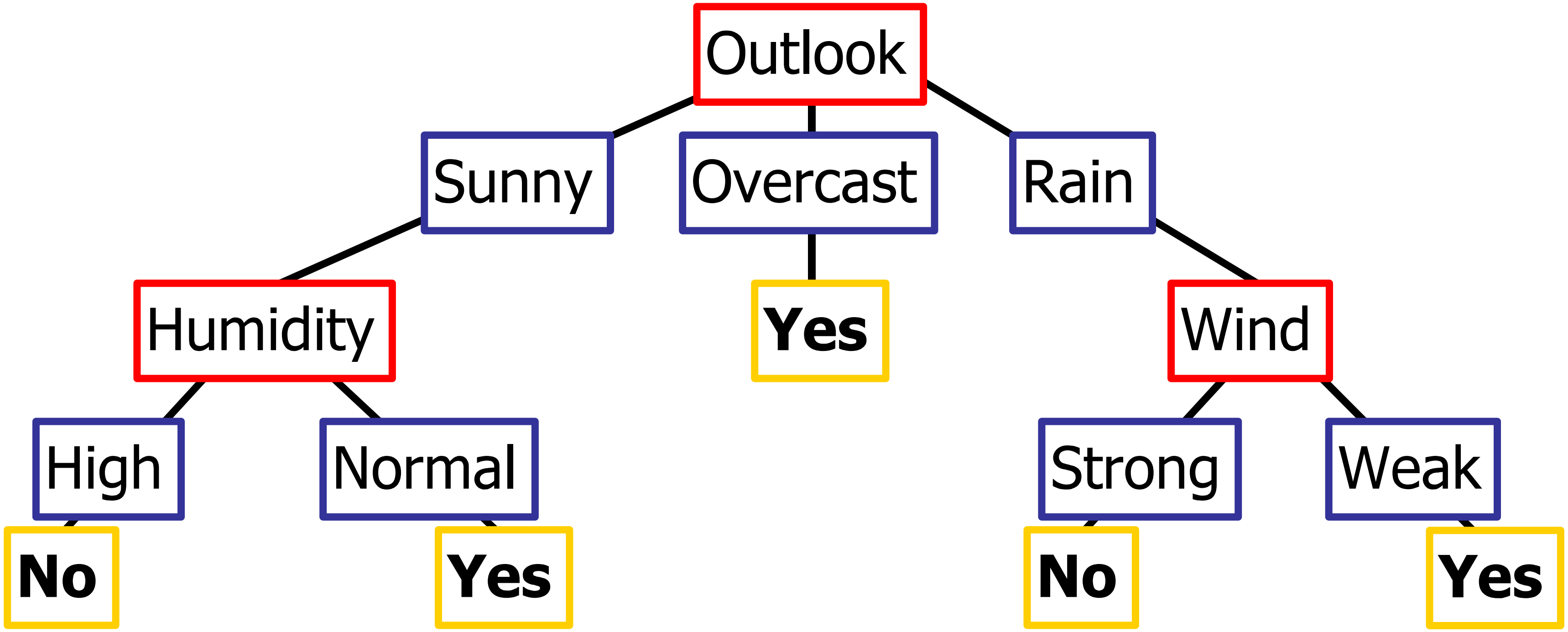
$$Gain(S,A) = \underbrace{Entropy(S)}_{\text{original entropy of S}} - \underbrace{\sum_{v \in values(A)} \frac{|S_v|}{|S|} \cdot Entropy(S_v)}_{\text{relative entropy of S}}$$

- S – a collection of examples
- A – an attribute
- *Values(A)* – possible values of attribute A;
- S_v – the subset of S for which attribute A has value v.

$$(\text{i.e., } S_v = \{s \in S | A(s) = v\}).$$

Decision Tree

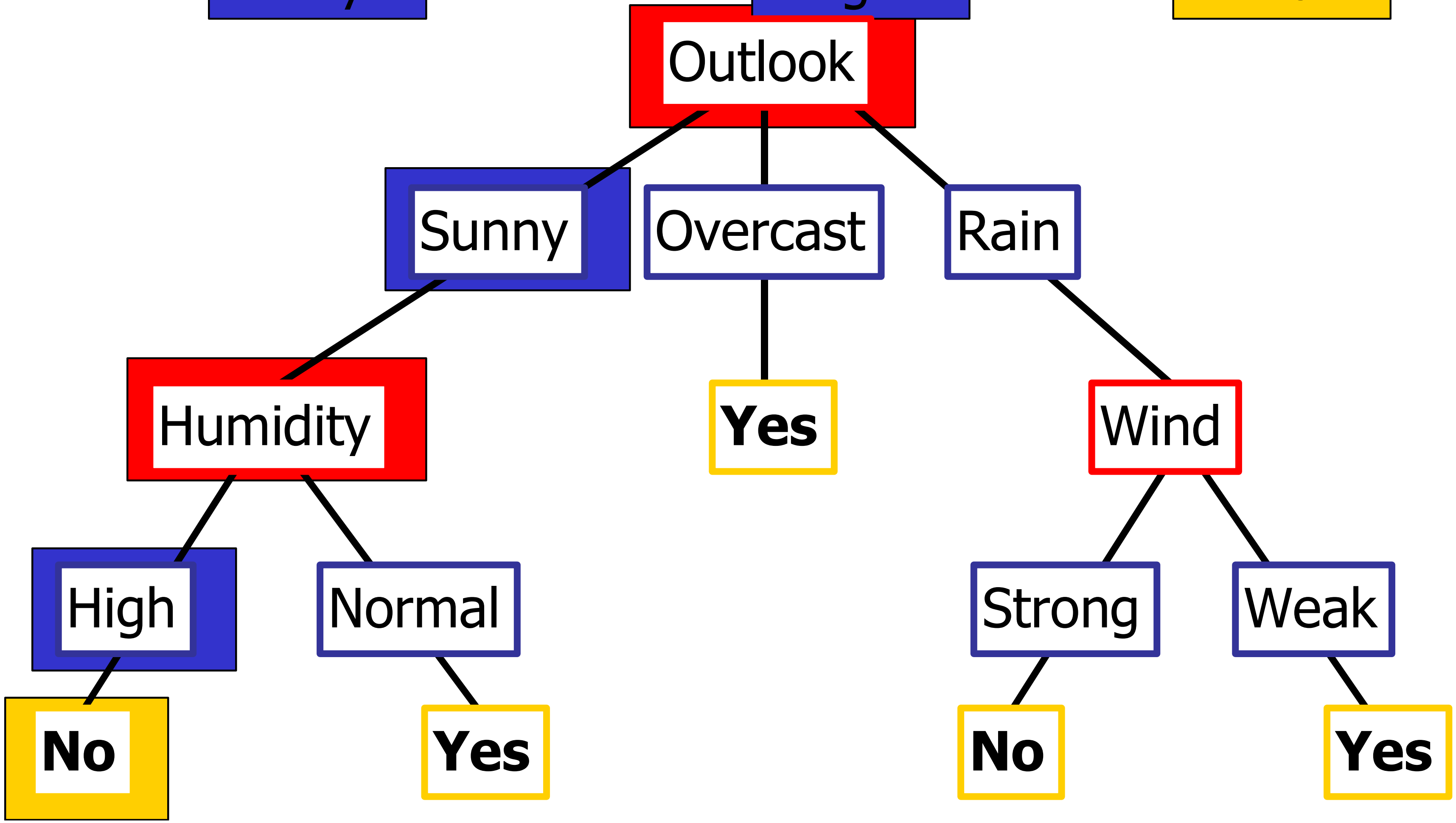
- decision trees represent disjunctions of conjunctions



(Outlook=Sunny $\wedge$ Humidity=Normal)
$\vee$ (Outlook=Overcast)
$\vee$ (Outlook=Rain $\wedge$ Wind=Weak)

Decision Tree for PlayTennis

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	?No



Decision Tree – ID3 Algorithm

- ID3 is one of the most common decision tree algorithm
- Dichotomisation means dividing into two completely opposite things.
- Algorithm iteratively divides attributes into two groups which are the most dominant attribute and others to construct a tree.
- Then, it calculates the Entropy and Information Gains of each attribute. In this way, the most dominant attribute can be founded.
- After then, the most dominant one is put on the tree as decision node.
- Entropy and Gain scores would be calculated again among the other attributes.
- Procedure continues until reaching a decision for that branch.

Decision Tree Steps

Step 1: Create a root node

- How to choose the root node?

The attribute that best classifies the training data, use this attribute at the root of the tree.

- How to choose the best attribute?

So from here, ***ID3 algorithm*** begins

Step 2: To choose best attribute

1. COMPUTE THE **ENTROPY** FOR DATA-SET **ENTROPY(S)**
2. FOR EVERY ATTRIBUTE/FEATURE:
    1. CALCULATE ENTROPY FOR ALL OTHER VALUES **ENTROPY(A)**
    2. TAKE **AVERAGE INFORMATION ENTROPY** FOR THE CURRENT ATTRIBUTE
    3. CALCULATE **GAIN** FOR THE CURRENT ATTRIBUTE
3. PICK THE **HIGHEST GAIN ATTRIBUTE**.
4. **REPEAT** UNTIL WE GET THE TREE WE DESIRED.

Sample – Decision Tree

Attribute				Classes
Gender	**Car Ownership**	**Travel Cost**	**Income Level**	**Transportation**
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Sample – Decision Tree

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

the problem

Name	Gender	Car ownership	Travel Cost	Income Level	Transportation
Alex	Male	1	Standard	High	? Train
Buddy	Male	0	Cheap	Medium	? Bus
Cherry	Female	1	Cheap	High	? Train

Sample – Decision Tree

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Probability :
-Bus → 4/10 = 0.4
-Train → 3/10 = 0.3
-Car → 3/10 = 0.3

10 data

Sample – Decision Tree

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Probability :
-Bus → 4/10 = 0.4
→ 3/10 = 0.3
-Car → 3/10 = 0.3

10 data

Measure the impurity :
- *Level of homogenous and heterogeneous*
- *There are three :*
  1. *Entropy*
  2. *Gini Index*
  3. *Classification Error*

Impurity using entropy :

$$E(S) = \sum - p(I) \log_2 p(I)$$

Entropy
– 0.4 log (0.4,2) – 0.3 log (0.3,2) – 0.3 log (0.3,2) = 1.571

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Sample – Decision Tree

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Female	Train
Female	Bus
Male	Bus
Male	Train
Female	Train
Female	Car
Male	Car
Female	Car

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Male	Bus
Male	Train
Male	Car

Probability :
Bus : 3/5 = 0.6
Train : 1/5= 0.2
Car : 1/5 = 0.2
Entropy→ 1.522

Attribute	Classes
Gender	Transportation
Female	Train
Female	Bus
Female	Train
Female	Car
Female	Car

Probability :
Bus : 1/5 = 0.2
Train : 2/5= 0.4
Car : 2/5 = 0.4
Entropy → 1.371

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

Information Gain →
*$\boxed{1.571}$ – $\boxed{(((5/10)}$ *1.522)+$\boxed{((5/10)}$ *1.371)) = 0.12*

Entropy all data

Pr...babili... of M...

Sample – Decision Tree

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Attribute	Classes
Car Ownership	Transportation
0	Bus
1	Bus
1	Train
0	Bus
1	Bus
0	Train
1	Train
1	Car
2	Car
2	Car

Attribute	Classes
Car Ownership	Transportation
0	Bus
0	Bus
0	Train

Probability:
Bus: 2/3=0.66
Train : 1/3= 0.33
Entropy : -2/3 log2(2/3)-1/3log2(1/3)=0.918

Attribute	Classes
Car Ownership	Transportation
1	Bus
1	Train
1	Bus
1	Train
1	Car

Probability:
Bus: 2/5=0.4
Train : 2/5= 0.4
Car : 1/5= 0.2
Entropy: -2/5log(2/5)-2/5log(w2/5) - 1/5log(1/5)=1.522

Attribute	Classes
Car Ownership	Transportation
2	Car
2	Car

Probability:
Car: 2/2=1
Entropy: 0

Information Gain →
*1.571 – (((3/10)*0.918)+((5/10)*1.522)+((2/10)*0)) = 0.534*

Sample – Decision Tree

Attributes				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Sample – Decision Tree

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Attribute	Classes
Travel Cost	Transportation
Cheap	Bus
Cheap	Bus
Cheap	Train
Cheap	Bus
Cheap	Bus
Standard	Train
Standard	Train
Expensive	Car
Expensive	Car
Expensive	Car

Attribute	Classes
Travel Cost	Transportation
Cheap	Bus
Cheap	Bus
Cheap	Train
Cheap	Bus
Cheap	Bus

Probability:
Bus: 4/5
Train : 1/5
Entropy : -4/5log(4/5)-1/5log(1/5)=0.722

Attribute	Classes
Travel Cost	Transportation
Standard	Train
Standard	Train

Probability:
Train : 2/2=1
Entropy: 0

Attribute	Classes
Travel Cost	Transportation
Expensive	Car
Expensive	Car
Expensive	Car

Probability:
Expencive: 3/3=1
Entropy: 0

Information Gain →
$1.571 - (((5/10)*0.722)+((2/10)*0)+((3/10)*0)) = 1.21$

Sample – Decision Tree

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Attribute	Classes
Income Level	Transportation
Low	Bus
Medium	Bus
Medium	Train
Low	Bus
Medium	Bus
Medium	Train
Medium	Train
High	Car
Medium	Car
High	Car

Attribute	Classes
Income Level	Transportation
Low	Bus
Low	Bus

Probability:
Bus : 2/2=1
Entropy: 0

Attribute	Classes
Income Level	Transportation
Medium	Bus
Medium	Train
Medium	Bus
Medium	Train
Medium	Train
Medium	Car

Probability:
Bus: 2/6
Train : 2/6
Car: 1/6
Entropy : 1.459

Attribute	Classes
Income Level	Transportation
High	Car
High	Car

Probability:
Car : 2/2=1
Entropy: 0

Information Gain →
*1.571 – (((2/10)*0)+((6/10)*1.459)+((2/10)*0)) = 0.695*

Sample – Decision Tree

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Attribute	*Information Gain*
Gender	0.125
Car	0.534
Travel cost	1.21
Income Level	0.695

Sample – Decision Tree

Attribute				Classes
Travel Cost	Gender	Car Ownership	Income Level	Transportation
Cheap	Male	0	Low	Bus
Cheap	Male	1	Medium	Bus
Cheap	Female	1	Medium	Train
Cheap	Female	0	Low	Bus
Cheap	Male	1	Medium	Bus
Standard	Male	0	Medium	Train
Standard	Female	1	Medium	Train
Expensive	Female	1	High	Car
Expensive	Male	2	Medium	Car
Expensive	Female	2	High	Car

Travel Cost

Attribute				Classes
Travel Cost	Gender	Car Ownership	Income Level	Transportation
Cheap	Male	0	Low	Bus
Cheap	Male	1	Medium	Bus
Cheap	Female	1	Medium	Train
Cheap	Female	0	Low	Bus
Cheap	Male	1	Medium	Bus

Attribute				Classes
Travel Cost	Gender	Car Ownership	Income Level	Transportation
Standard	Male	0	Medium	Train
Standard	Female	1	Medium	Train

Attribute				Classes
Travel Cost	Gender	Car Ownership	Income Level	Transportation
Expensive	Female	1	High	Car
Expensive	Male	2	Medium	Car
Expensive	Female	2	High	Car

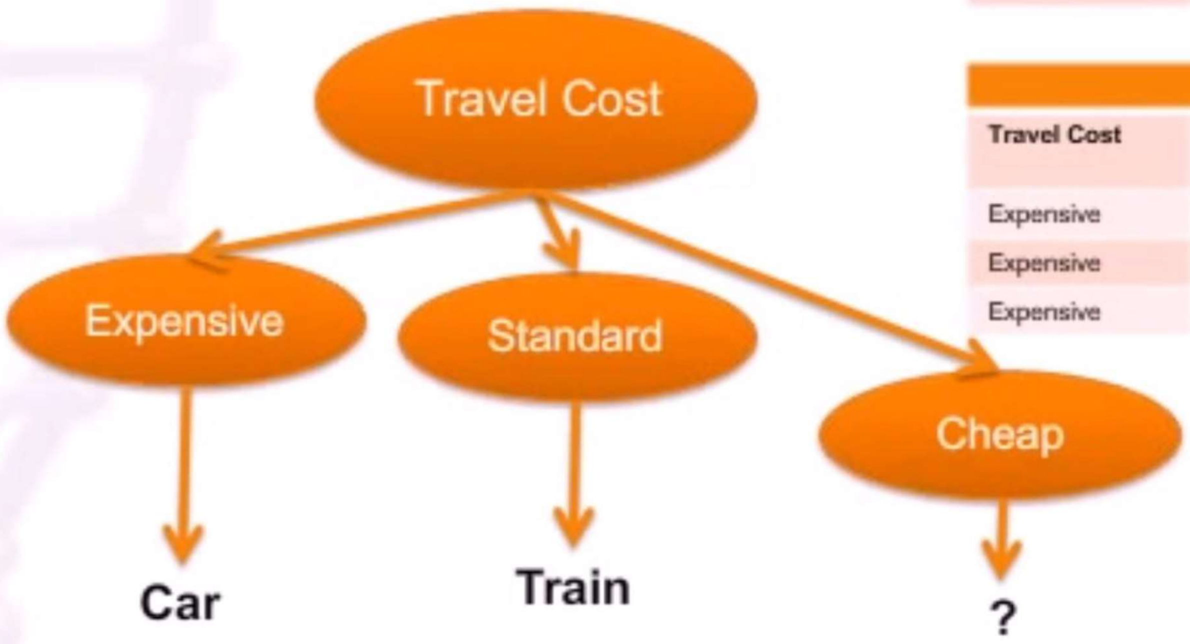
Sample – Decision Tree

Attribute				Classes
Travel Cost	Gender	Car Ownership	Income Level	Transportation
Cheap	Male	0	Low	Bus
Cheap	Male	1	Medium	Bus
Cheap	Female	1	Medium	Train
Cheap	Female	0	Low	Bus
Cheap	Male	1	Medium	Bus
Standard	Male	0	Medium	Train
Standard	Female	1	Medium	Train
Expensive	Female	1	High	Car
Expensive	Male	2	Medium	Car
Expensive	Female	2	High	Car

Attribute				Classes
Travel Cost	Gender	Car Ownership	Income Level	Transportation
Cheap	Male	0	Low	Bus
Cheap	Male	1	Medium	Bus
Cheap	Female	1	Medium	Train
Cheap	Female	0	Low	Bus
Cheap	Male	1	Medium	Bus

Attribute				Classes
Travel Cost	Gender	Car Ownership	Income Level	Transportation
Standard	Male	0	Medium	Train
Standard	Female	1	Medium	Train

Attribute				Classes
Travel Cost	Gender	Car Ownership	Income Level	Transportation
Expensive	Female	1	High	Car
Expensive	Male	2	Medium	Car
Expensive	Female	2	High	Car



Sample – Decision Tree

Attribute				Classes
Travel Cost	Gender	Car Ownership	Income Level	Transportation
Cheap	Male	0	Low	Bus
Cheap	Male	1	Medium	Bus
Cheap	Female	1	Medium	Train
Cheap	Female	0	Low	Bus
Cheap	Male	1	Medium	Bus
Standard	Male	0	Medium	Train
Standard	Female	1	Medium	Train
Expensive	Female	1	High	Car
Expensive	Male	2	Medium	Car
Expensive	Female	2	High	Car

Second iteration → Omit Travel Cost and only for cheap

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Probability :
Bus → 4/5 = 0.8
Train → 1/5 = 0.2
Entropy → 0.722

Sample – Decision Tree

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Female	Train
Female	Bus
Male	Bus

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Male	Bus

Probability
Bus : 3/3 = 1
Entropy → 0

Attribute	Classes
Gender	Transportation
Female	Train
Female	Bus

Sample – Decision Tree

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Female	Train
Female	Bus
Male	Bus

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Male	Bus

Probability :
Bus : 3/3 = 1
Entropy → 0

Attribute	Classes
Gender	Transportation
Female	Train
Female	Bus

Probability :
Bus : 1/2 = 0.5
Train : 1/2 = 0.5
Entropy → 1

Information Gain →
0.722 – (((3/5)*0)+((2/5)*1)) = 0.322

Sample – Decision Tree

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Attribute	Classes
Car Ownership	Transportation
0	Bus
1	Bus
1	Train
0	Bus
1	Bus

Attribute	Classes
Car Ownership	Transportation
0	Bus
0	Bus

Probability :
Bus : 2/2 = 1
Entropy → 0

Attribute	Classes
Car Ownership	Transportation
1	Bus
1	Train
1	Bus

Probability :
Bus : 2/3 = 0.67
Train : 1/3 = 0.33
Entropy → 0.918

Information Gain →
*0.722 – (((2/5)*0)+((3/5)*0.918)) = 0.171*

Sample – Decision Tree

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Attribute	Classes
Income Level	Transportation
Low	Bus
Medium	Bus
Medium	Train
Low	Bus
Medium	Bus

Attribute	Classes
Income Level	Transportation
Low	Bus
Low	Bus

Probability :
Bus : 2/2 = 1
Entropy → 0

Attribute	Classes
Income Level	Transportation
Medium	Bus
Medium	Train
Medium	Bus

Probability :
Bus : 2/3 = 0.67
Train : 1/3 = 0.33
Entropy → 0.918

Information Gain →
0.722 – (((2/5)*0)+((3/5)*0.918)) = 0.171

Sample – Decision Tree

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus
Male	0	Medium	Train
Female	1	Medium	Train
Female	1	High	Car
Male	2	Medium	Car
Female	2	High	Car

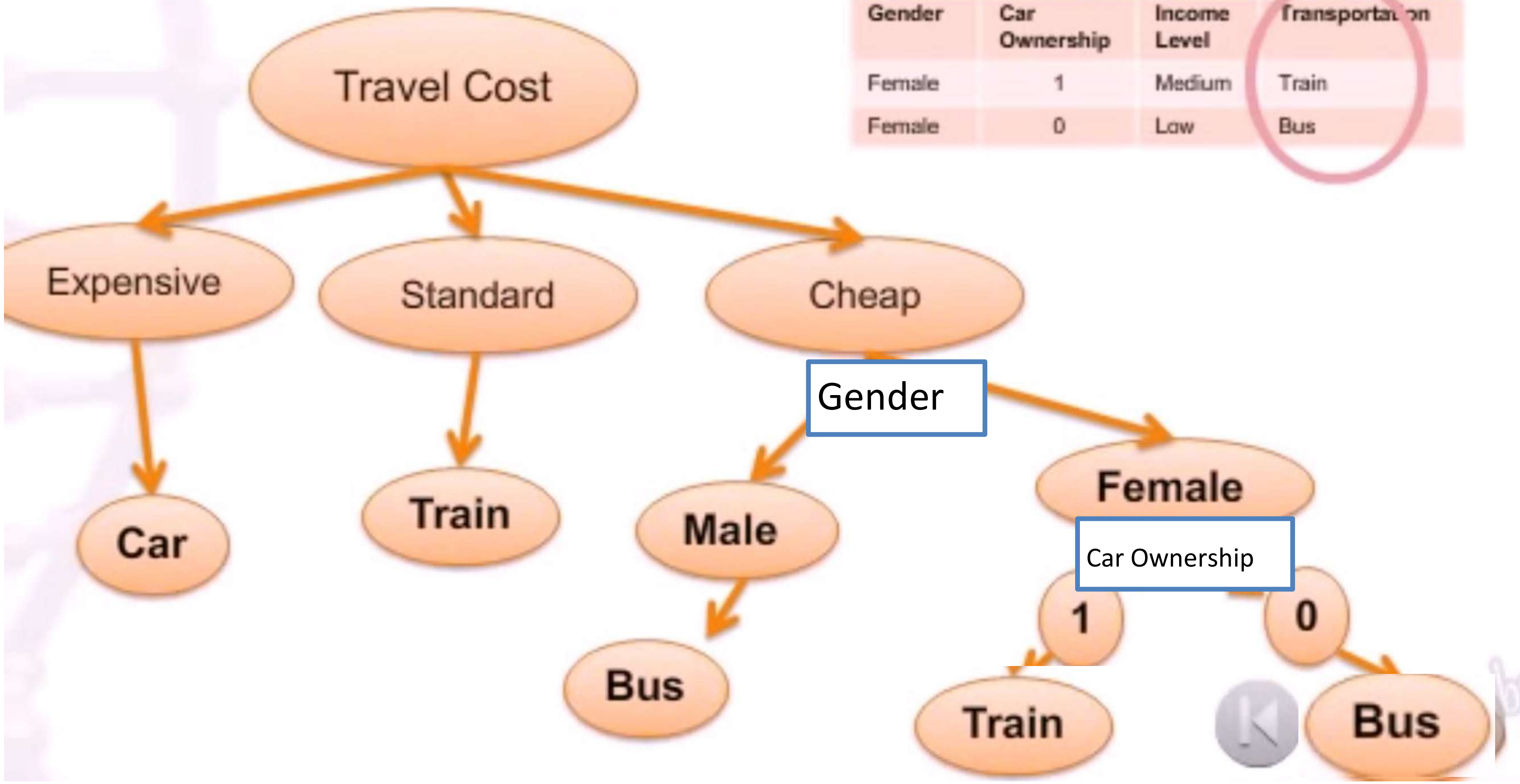
Attribute	*Information Gain*
Gender	0.322
Car	0.171
Income Level	0.171

Sample – Decision Tree

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

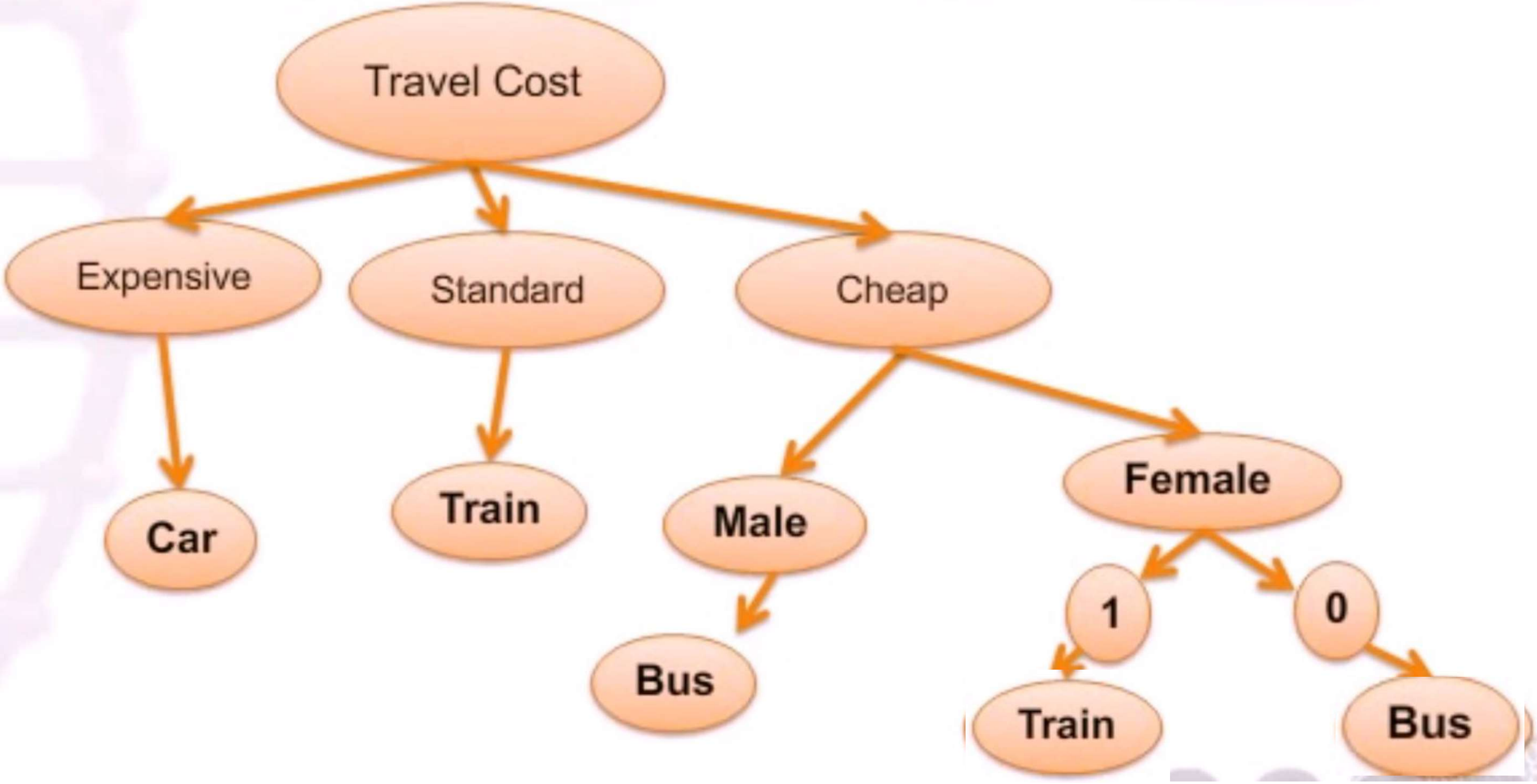
Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Male	1	Medium	Bus

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Female	1	Medium	Train
Female	0	Low	Bus



Sample – Decision Tree

Name	Gender	Car ownership	Travel Cost	Income Level	Transportation
Alex	Male	1	Standard	High	?
Buddy	Male	0	Cheap	Medium	?
Cherry	Female	1	Cheap	High	?



ID3 Algorithm

ID3(*Examples*, *Target_attribute*, *Attributes*)

Examples are the training examples. Target_attribute is the attribute whose value is to be *predicted by the tree. Attributes is a list of other attributes that may be tested by* the learned *decision tree. Returns a decision tree that correctly classifies the given Examples.*

- 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*
- Otherwise Begin
    - $A \leftarrow$ the attribute from *Attributes* that best* classifies *Examples*
    - The decision attribute for *Root* $\leftarrow A$
    - For each possible value, v_i, of A,
        - Add a new tree branch below *Root*, corresponding to the test $A = v_i$
        - Let $Examples_{v_i}$ be the subset of *Examples* that have value v_i for A
        - If $Examples_{v_i}$ is empty
            - Then below this new branch add a leaf node with label = most common value of *Target_attribute* in *Examples*
            - Else below this new branch add the subtree ID3($Examples_{v_i}$, *Target_attribute*, *Attributes* − {A}))
- End
- Return *Root*

* The best attribute is the one with highest *information gain*, as defined in Equation (3.4).

TABLE 3.1

Summary of the ID3 algorithm specialized to learning boolean-valued functions. ID3 is a greedy algorithm that grows the tree top-down, at each node selecting the attribute that best classifies the local training examples. This process continues until the tree perfectly classifies the training examples, or until all attributes have been used.

Example 2: database: playtennis

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	Normal	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	High	Strong	Yes
D8	Sunny	Mild	Normal	Weak	No
D9	Sunny	Hot	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Cool	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Eg. **Play Tennis:** **Probability**

positive examples: 09 —— $9/14 = 0.64$

negative examples: 05 —— $5/14 = 0.35$.

Entropy of S: $-(0.64)\log_2(0.64) - (0.35)\log(0.35)$

$= 0.940$

Info gain of the attribute outlook is calculated as: [i.e 1st col. & last col]

I

Outlook	PlayTennis
Sunny	NO
Sunny	NO
Sunny	NO
Sunny	yes
Sunny	yes

Probability:

NO: $3/5 = 0.6$

yes: $2/5 = 0.4$

Entropy $= -0.6\log_2 0.6 - 0.4\log_2 0.4$

$= -0.6 * (-0.7369) - 0.4 * (-1.3217)$

$= 0.52876 + 0.44214$

$= 0.970$

II

outlook	playTennis
overcast	yes
overcast	yes
overcast	yes
overcast	yes

Entropy = 0

III

outlook	playTennis
Rain	yes —— 1
Rain	yes —— 2
Rain	~~yes~~ NO —— (1)
Rain	yes —— 3
Rain	NO —— (2)

Entropy $= -3/5 \log_2 \frac{3}{5} - \frac{2}{5}\log_2 \frac{2}{5}$

$= (0.6 * (-0.7369) - (0.4 * (-1.3219))$

$= 0.44214 + 0.52876$

$= 0.970$

Information Gain: (S, outlook)

$= \text{Entropy}(S) - \left[\left(\frac{5}{14}\right) * 0.9709 + 0 + \left(\frac{5}{14}\right) * 0.9709 \right]$

$= 0.9409 - [0.3467 + 0.3467]$

$= 0.246$

|||ly

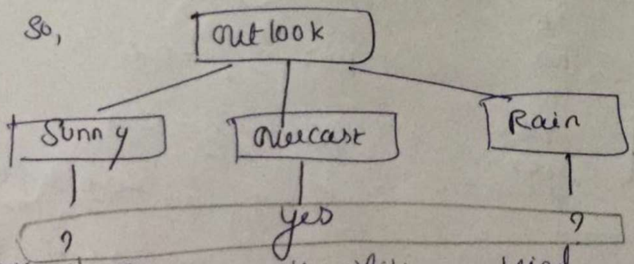
$\boxed{\text{Gain}(S, \text{outlook}) = 0.246}$ Root node

Gain (S, temperature) = 0.029 — [2nd col & last col]

Gain (S, Humidity) = 0.151 — [3rd col & last col]

Gain (S, wind) = 0.029 — [4th col & last col]

So,



outlook	Temp	Humidity	wind	playTennis.
Sunny	Hot	High	weak	NO
Sunny	Hot	High	Strong	NO
Sunny	Mild	High	weak	NO
Sunny	Cool	Normal	weak	Yes
Sunny	Mild	Normal	Strong	yes

Then, compare gain, First find Entropy 1st col with last col

|||ly 2nd " "
3rd, 4th

Probability : NO — $3/5$

yes — $2/5$

$\boxed{\text{Entropy : } 0.970}$

$\text{Entropy} = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}$
$= 0.722$

Now consider, [1st col — Temp & last col — play Tennis]

Temp	playTennis
Hot	NO
Hot	NO
mild	NO
cool	yes
mild	yes.

Entropy : 0

Temp	playTennis
Hot	NO
Hot	NO
~~mild~~	~~NO~~

Entropy : 0

Temp	playTennis
cool	yes
~~mild~~	~~yes~~

Entropy : 1

Temp	playTennis
mild	NO
mild	yes.

Information Gain :

i.e (Sunny, Temp)

$= 0.970 - \left[\left(\frac{2}{5}\right)*0 + \left(\frac{1}{5}\right)*0 + \left(\frac{2}{5}\right)*1\right]$

$= 0.970 - 0.4$

$= 0.570$.

IIIly for (sunny, Humidity) (sunny, wind)

Gain (Sunny, Humidity) $= 0.970 - \left[\left(\frac{3}{5}\right)*0.0 + \left(\frac{2}{5}\right)*0.0\right]$

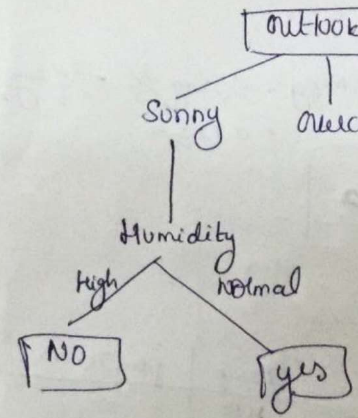
$\boxed{= 0.970}$ — Highest info gain (so next node)

Gain (Sunny, wind) $= 0.970 - \left[\left(\frac{3}{5}\right)*(0.918) + \left(\frac{2}{5}\right).1\right]$

$= 0.970 - 0.9508$

$= 0.0192$

So, attribute Humidity will be descendant node.



Now, find for Rain

Rain Table:

Outlook	Temp	Humidity	Wind	PlayTennis
Rain	Mild	High	Weak	yes
Rain	cool	Normal	Weak	yes
Rain	cool	Normal	Strong	No
Rain	Mild	Normal	Weak	yes
Rain	Mild	High	Strong	No

$$\text{Entropy}(S_{Rain}) = -\frac{3}{5}\log_2 3/5 - \frac{2}{5}\log_2 2/5$$
$$= 0.970$$

Now Consider:

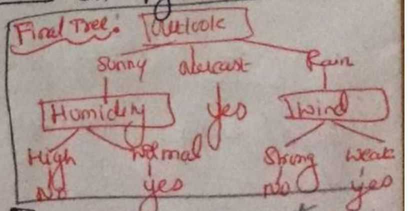
Temp	PlayTennis
Mild	yes
cool	yes
cool	No
Mild	yes
Mild	No

Temp	PlayTennis
mild	yes
mild	yes
mild	No

Yes : 2/3
No : 1/3
Entropy : 0.918

Entropy = 0

Temp	PlayTennis
cool	yes
cool	yes



$$\text{Gain}(Rain, Temp) = 0.970 - \left[\left(\frac{3}{5}\right)*0.918 + \left(\frac{2}{5}\right)*1\right] = 0.0198$$

$$\text{Gain}(Rain, Wind) = 0.970 - \left[\left(\frac{3}{5}\right)*0 + \left(\frac{2}{5}\right)*0\right] \boxed{= 0.970}$$

Next Highest info gain

$$\text{Gain}(Rain, Humidity) = 0.970 - \left[\left(2/5\right)*1 + \left(\frac{3}{5}\right)*0.917\right] = 0.0198$$

So, highest info gain is attribute 'Wind'.

Describe the concepts of entropy and information gain. Construct decision tree by taking the enjoy sports concepts and training instances given below.(using ID3 algorithm)

day	outlook	temperature	Humidity	Wind	Play ?
D1	sunny	hot	high	weak	Yes
D2	sunny	hot	high	strong	No
D3	overcast	hot	high	weak	Yes
D4	rain	mild	normal	weak	No
D5	rain	cool	normal	weak	Yes

day	outlook	temperature	Humidity	Wind	Play ?
D1	sunny	hot	high	weak	Yes
D2	sunny	hot	high	strong	No
D3	overcast	hot	high	weak	Yes
D4	rain	mild	normal	weak	No
D5	rain	cool	normal	weak	Yes

$$\text{entropy}(S) = -P_+ \log(P_+) - P_- \log(P_-)$$

$$= -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right)$$

$$= 0.9709.$$

All logs with base 2

For attribute Outlook:

Sunny	Yes	Entropy = 1
Sunny	No	
Overcast	Yes	Entropy = 0
Rain	No	Entropy = 1
Rain	Yes	

Information gain

$$G(S, A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0.9709 - \left[\frac{2}{5}(1) + \frac{1}{5}(0) + \frac{2}{5}(1)\right]$$

$$= 0.9709 - \left[\frac{4}{5}\right]$$

$$= 0.1709$$

For attribute temperature

Hot	Yes	$\text{Entropy} = -\frac{2}{3}\log\left(\frac{2}{3}\right) - \frac{1}{3}\log\left(\frac{1}{3}\right)$
Hot	No	$= 0.9182$
Hot	Yes	
Mild	No	Entropy = 0
Cool	Yes	Entropy = 0.

~~For~~ Info gain = $0.9709 - \left[\frac{3}{5}(0.9182) + 0 + 0\right]$

$= 0.60362$ — $0.4l998$

For attribute humidity:

High	Yes	Entropy = 0.9182
High	No	
High	Yes	
Normal	No	Entropy = 1
Normal	Yes	

Info gain = $0.9709 - \left[\frac{3}{5}(0.9182) + \frac{2}{5}\right]$

$= 0.0199$

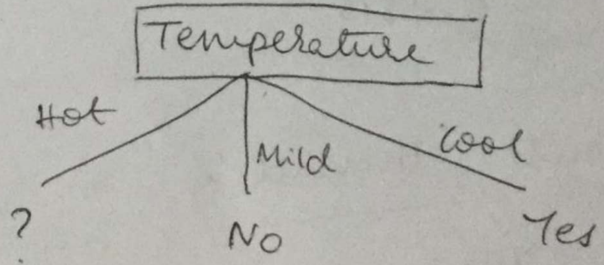
For attribute Wind

Weak	Yes	Entropy $= -\frac{3}{4}\log(\frac{3}{4}) - \frac{1}{4}\log(\frac{1}{4})$
Weak	Yes	
Weak	No	$= 0.8112$
Weak	Yes	
Strong	No	Entropy $= 0$.

$$\text{Info gain} = 0.9709 - \left[\frac{4}{5} \times 0.8112 + 0\right]$$
$$= 0.3219$$

Since information gain for Temperature is maximum, we take it as the root node.

day	outlook	temperature	Humidity	Wind	Play ?
D1	sunny	hot	high	weak	Yes
D2	sunny	hot	high	strong	No
D3	overcast	hot	high	weak	Yes
D4	rain	mild	normal	weak	No
D5	rain	cool	normal	weak	Yes



day	outlook	temperature	Humidity	Wind	Play ?
D1	sunny	hot	high	weak	Yes
D2	sunny	hot	high	strong	No
D3	overcast	hot	high	weak	Yes
D4	rain	mild	normal	weak	No
D5	rain	cool	normal	weak	Yes

Select the training eg having temperature value as "hot".

For attribute Outlook:

Sunny Yes
Sunny No   Entropy = 1

Overcast Yes   Entropy = 0.

~~Info~~ = Entropy (S) = $-\frac{2}{3}\log(2/3) - \frac{1}{3}\log(\frac{1}{3})$
$= 0.9182$

Info gain = $0.9182 - \left[\frac{2}{3}(1) + 0\right]$
$= 0.2515$

For attribute Humidity:

High Yes
High No   Entropy = 0.9182
High Yes

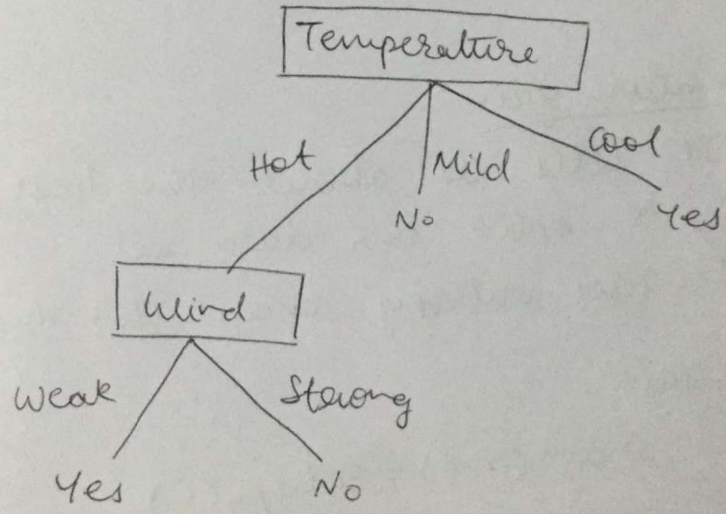
Info gain = $0.9182 - 0.9182$
$= 0.$

For attribute <u>Wind</u>:

~~Rof~~ Weak	Yes	Entropy = 0
Weak	Yes	
Strong	No	Entropy = 0.

Info gain $= 0.9182 - [0+0] = 0.9182$.

Info gain is max for wind attribute. Hence we take it for futher split.



Since all the branches terminate with leaf nodes, no further splitting is required. This is the final decision tree.

Question 1

Which of the following is true for a decision tree?

a. A decision tree is an example of a linear classifier.
b. The entropy of a node typically decreases as we go down a decision tree.
c. Entropy is a measure of purity.
d. An attribute with lower mutual information should be preferred to other attributes.

Correct Answer: b

Question 2

Consider the dataset, S given below:

Elevation	Road Type	Speed Limit	Speed
steep	Uneven	Yes	Slow
steep	Smooth	Yes	Slow
flat	Uneven	No	Fast
steep	Smooth	No	Fast

Elevation, Road Type and speed Limit are the features and Speed is the target label that we want to predict.

Find the entropy of the dataset, S as given above:

a. 0.5
b. 0
c. 1
d. 0.7

Correct Answer: c

Detailed solution:

For a dataset, S with C many classes, the entropy of the set, S given by H(S) is defined as:

$H(S) = -\sum_{c \in C} p_c \log p_c$ where p_c is the probability of an element of S belonging to a class, $c \in C$ where C is the set of all classes. In this case, P(slow) = 0.5; P(fast) = 0.5 and hence

H(S) = 1

Question 3

Find the information Gain if the dataset is split at the feature “Elevation”:

a. 1
b. 0
c. 0.675
d. 0.325

Correct Answer: c

Detailed Solution:

Information gain for a feature, F on a set, S is defined as:

Information Gain, $IG\ (S, F)\ =\ H(S)\ - \sum_{f\ \epsilon\ F}\ \frac{|S_f|}{|S|}\ H(S_f)$ where S_f is the subset of elements in S which have a value f for the feature F. $H(S_f)$ is the entropy of the set after split on feature, f ϵ F. The feature, Elevation has 2 values = {Steep, Flat}. For a split on the feature Elevation, one subtree would be of inputs having feature Steep and the other having feature, Flat.

For the feature, Steep, there are 3 examples, out of which P(slow) = $\frac{2}{3}$ and P(fast) = $\frac{1}{3}$. Thus, Entropy (Steep) = 0.9 [see question 4 for the formula] and Entropy(Flat) = 0.

Thus Information Gain = $1 - (\frac{3}{4} * 0.9 + \frac{1}{4} * 0)\ =\ 1 -\ 0.675\ =\ 0.325$

Question 4

Find the feature on which the parent node must be chosen to split the dataset, S based or information gain:

a. Speed Limit
b. Road Type
c. Elevation

Correct Answer: a

Detailed solution:

Using the Information Gain formula, we can find the information gain for each of the features. The values should be:

IG(S, Elevation) = 0.325

IG(S, Road Type) = 0

IG(S, Speed Limit) = 1

Since the decision tree is constructed on the feature having maximum information gain, the correct answer.

Elev	Road type	Speed limit	Speed
Steep	uneven	yes	Slow
Steep	Smooth	yes	Slow
flat	uneven	NO	Fast
Steep	Smooth	no	Fast

Probability

Slow: $\frac{2}{4} = 0.5$

Fast: $\frac{2}{4} = 0.5$

Entropy $= -p \log p$

$= -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1$

Probability:

Steep	Slow
Steep	Slow
flat	Fast
Steep	Fast

$\Rightarrow$

Steep	Slow
Steep	Slow
Steep	Fast

Flat	Fast

Entropy (flat): 0

Slow: $\frac{2}{3} = 0.66$

Fast: $\frac{1}{3} = 0.33$

Entropy (Steep) $: -p \log p$

$= -0.66 \log_2 0.66 - 0.33 \log_2 0.33$

$= 0.9$

Information gain $=$ Entropy $- \sum \frac{|S_v|}{S} E(S_v)$

$= 1 - \left(\left(\frac{3}{4} * 0.9\right) + \left(0 * \frac{1}{4}\right)\right)$

$= 1 - 0.675 = 0.325$

Road type	Speed
uneven	Slow
Smooth	Slow
uneven	Fast
Smooth	Fast

$\longrightarrow$

uneven	Slow
uneven	Fast

Entropy: 1

Smooth	Slow
Smooth	Fast

Entropy: 1

Information gain $= 1 - \sum\left(\left(\frac{\cancel{2}^1}{\cancel{4}_2} * 1\right) + \left(\frac{\cancel{2}^1}{\cancel{4}_2} * 1\right)\right)$

$= 1 - \left[\frac{1}{2} + \frac{1}{2}\right] \quad \frac{1+1}{2} = \frac{\cancel{2}}{\cancel{2}}$

$= 1 - 1 = 0$

Speed limit	Speed
yes	Slow
yes	Slow
No	Fast
NO	Fast

$\Rightarrow$

yes	Slow
yes	Slow

Entropy: 0

No	Fast
No	Fast

Entropy: 0

Info gain $= 1 - [0] = 1$

Give decision trees to represent the following Boolean functions.

1) A && ¬B

2) A XOR B

3) A ∨ [B && C]

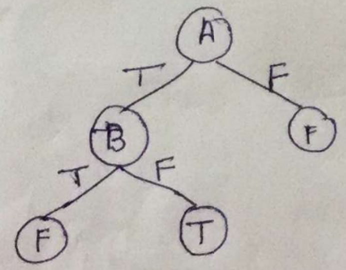
4) [A && B] ∨ [C && D]

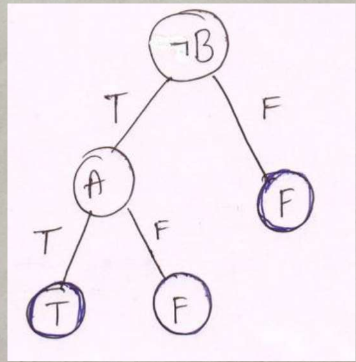
3) Give decision trees to represent the following Boolean functions:

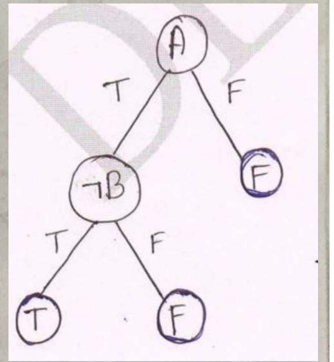
1) $\boxed{A \,\&\, \sim B}$ or $A \wedge \sim B$

Soln:

A	B	~B	A & ~B
T	T	F	F (-)
T	F	T	T (+)
F	T	F	F (-)
F	F	T	F (-)

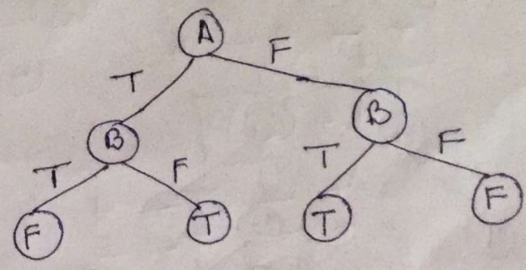






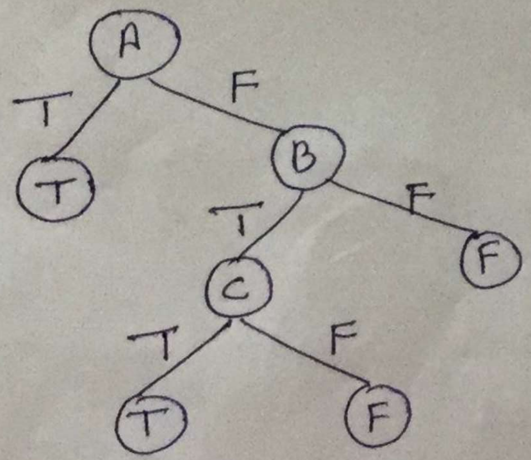
2) $\boxed{A \text{ XOR } B}$

A	B	A XOR B
T	T	F
T	F	T
F	T	T
F	F	F



3) $\boxed{A \vee [B \& C]}$ <u>&</u> $A \vee [B \wedge C]$

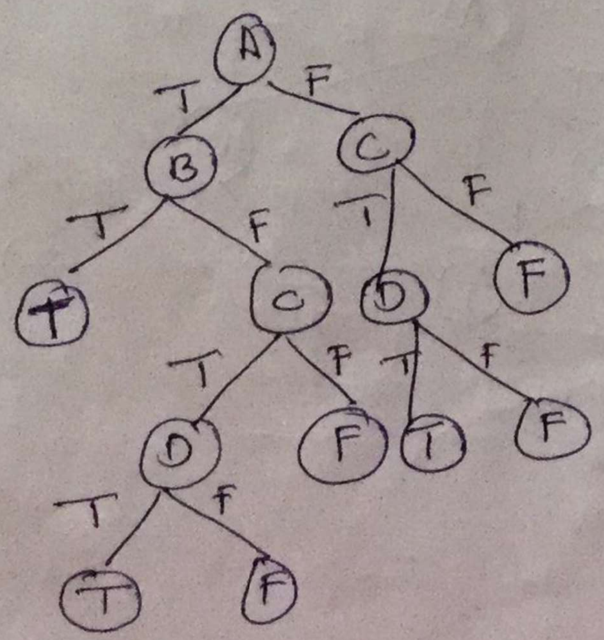
A	B	C	B & C	A v [B & C]
T	T	T	T	T
T	T	F	F	T
T	F	T	F	T
T	F	F	F	T
F	T	T	T	T
F	T	F	F	F
F	F	T	F	F
F	F	F	F	F

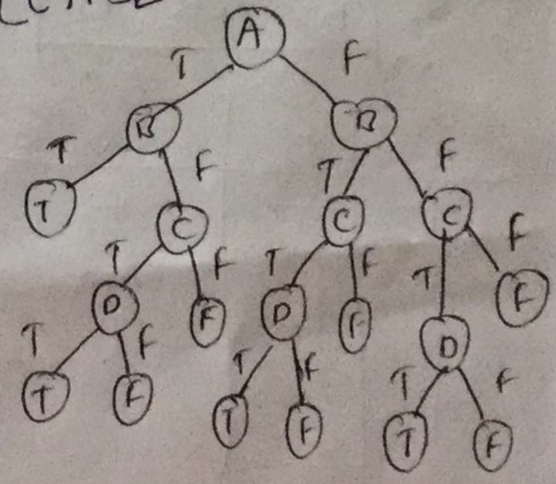


iv) [A && B] v [C && D]

A	B	C	D	A && B	C && D	(A && B) v (C && D)
T	T	T	T	T	T	T
T	T	T	F	T	F	T
T	T	F	T	T	F	T
T	T	F	F	T	F	T
T	F	T	T	F	T	T
T	F	T	F	F	F	F
T	F	F	T	F	F	F
T	F	F	F	F	F	F
F	T	T	T	F	T	T
F	T	T	F	F	F	F
F	T	F	T	F	F	F
F	T	F	F	F	F	F
F	F	T	T	F	T	T
F	F	T	F	F	F	F
F	F	F	T	F	F	F
F	F	F	F	F	F	F

4) $[A \,\&\, B] \vee [C \,\&\, D]$ or $[A \wedge B] \vee [C \wedge D]$



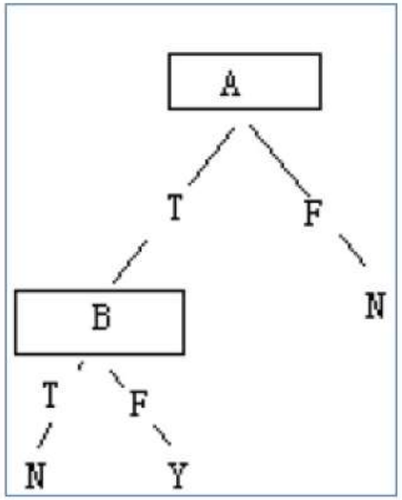


1) **Give decision trees to represent the following Boolean functions:**

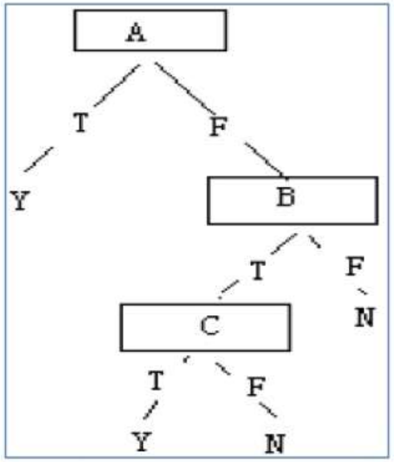
1) $A \wedge \neg B$
2) $A \vee [B \wedge C]$
3) A XOR B
4) $[A \wedge B] \vee [C \wedge D]$

Answer:

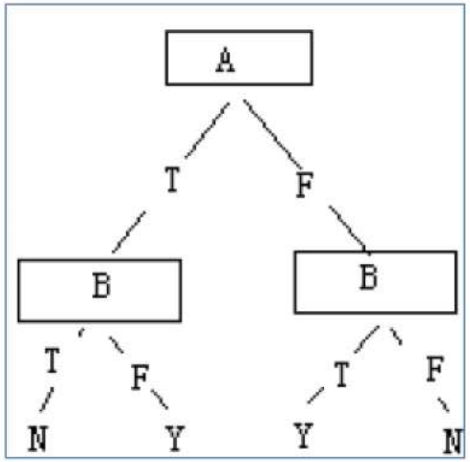
1) $A \wedge \neg B$



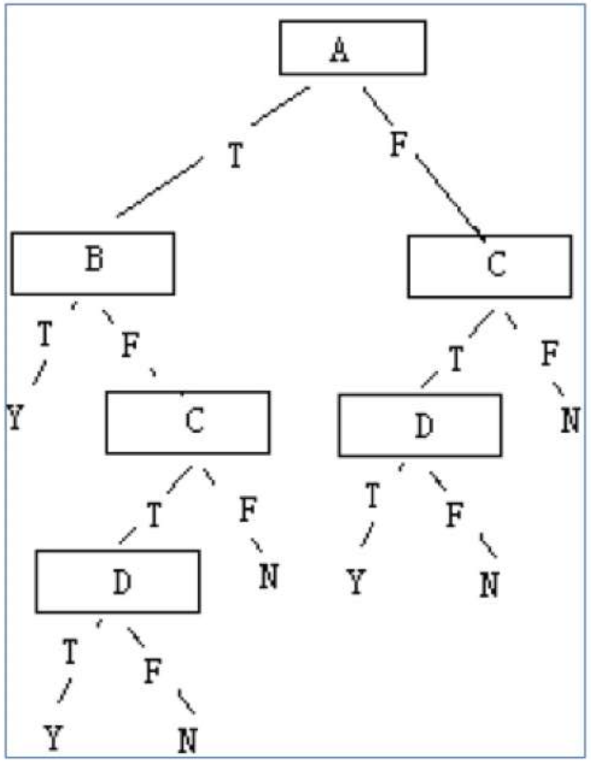
2) $A \vee [B \wedge C]$



3) A XOR B = $(A \wedge \neg B) \vee (\neg A \wedge B)$



4) $[A \wedge B] \vee [C \wedge D]$



HOME WORK

i) $[A \land B] \lor [C \land D]$

ii) $[A \land \sim B] \lor [\sim A \land B]$ or A XOR B

Home Work

Consider the following set of training examples.

a) What is the entropy of this collection of training example with respect to the target function classification?

b) What is the information gain of a_2 relative to these training examples?

Instance	Classification	a_1	a_2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

Decision Tree Algorithm – ID3 Solved Example

1. What is the entropy of this collection of training examples with respect to the target function classification?
2. What is the information gain of *a1* and *a2* relative to these training examples?
3. Draw decision tree for the given dataset.

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

Attribute: a1

$Values\,(a1) = T, F$

$S = [3+,\ 3-]$ $\qquad Entropy(S) = 1.0$

$S_T = [2+,\ 1-]$ $\qquad Entropy(S_T) = -\frac{2}{3}log_2\frac{2}{3} - \frac{1}{3}log_2\frac{1}{3} = 0.9183$

$S_F \leftarrow [1+,\ 2-]$ $\qquad Entropy(S_F) = -\frac{1}{3}log_2\frac{1}{3} - \frac{2}{3}log_2\frac{2}{3} = 0.9183$

$$Gain\,(S, a1) = Entropy(S) - \sum_{v \in \{T,F\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, a1) = Entropy(S) - \frac{3}{6}Entropy(S_T) - \frac{3}{6}Entropy(S_F)$$

$$Gain(S, a1) = 1.0 - \frac{3}{6} * 0.9183 - \frac{3}{6} * 0.9183 = 0.0817$$

Example - 2
Decision Tree Algorithm – ID3
Solved Example

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

Attribute: a2

$Values\,(a2) = T, F$

$S = [3+,\ 3-]$ $\quad Entropy(S) = 1.0$

$S_T = [2+,\ 2-]$ $\quad Entropy(S_T) = 1.0$

$S_F \leftarrow [1+,\ 1-]$ $\quad Entropy(S_F) = 1.0$

$$Gain\,(S, a2) = Entropy(S) - \sum_{v \in \{T,F\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, a2) = Entropy(S) - \frac{4}{6} Entropy(S_T) - \frac{2}{6} Entropy(S_F)$$

$$Gain(S, a2) = 1.0 - \frac{4}{6} * 1.0 - \frac{2}{6} * 1.0 = 0.0$$

Example - 2

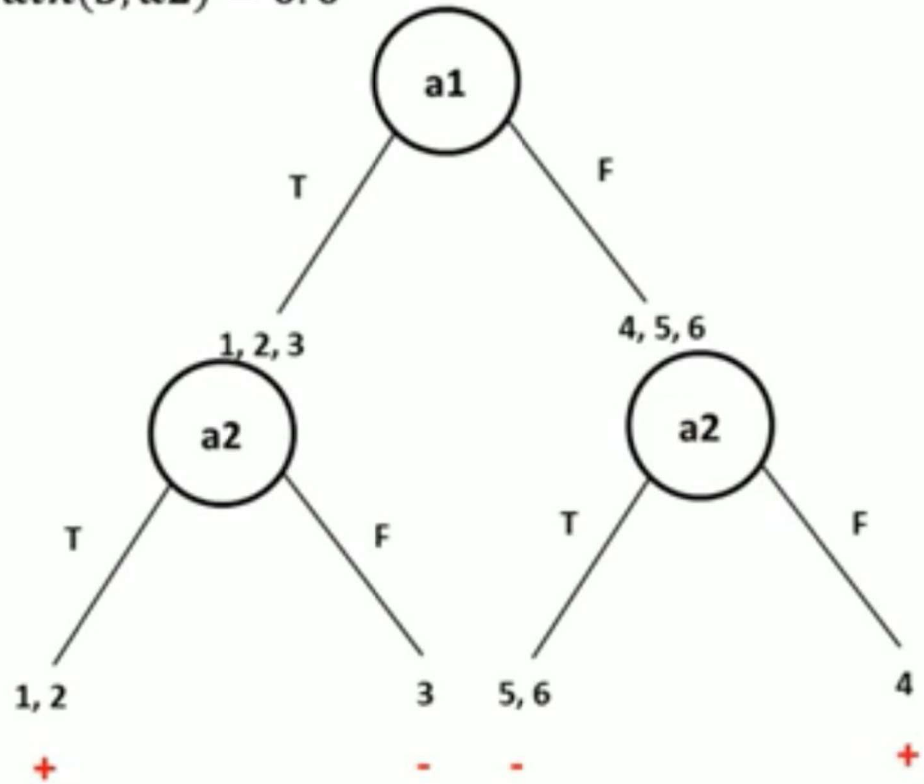
Decision Tree Algorithm – ID3

Solved Example

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

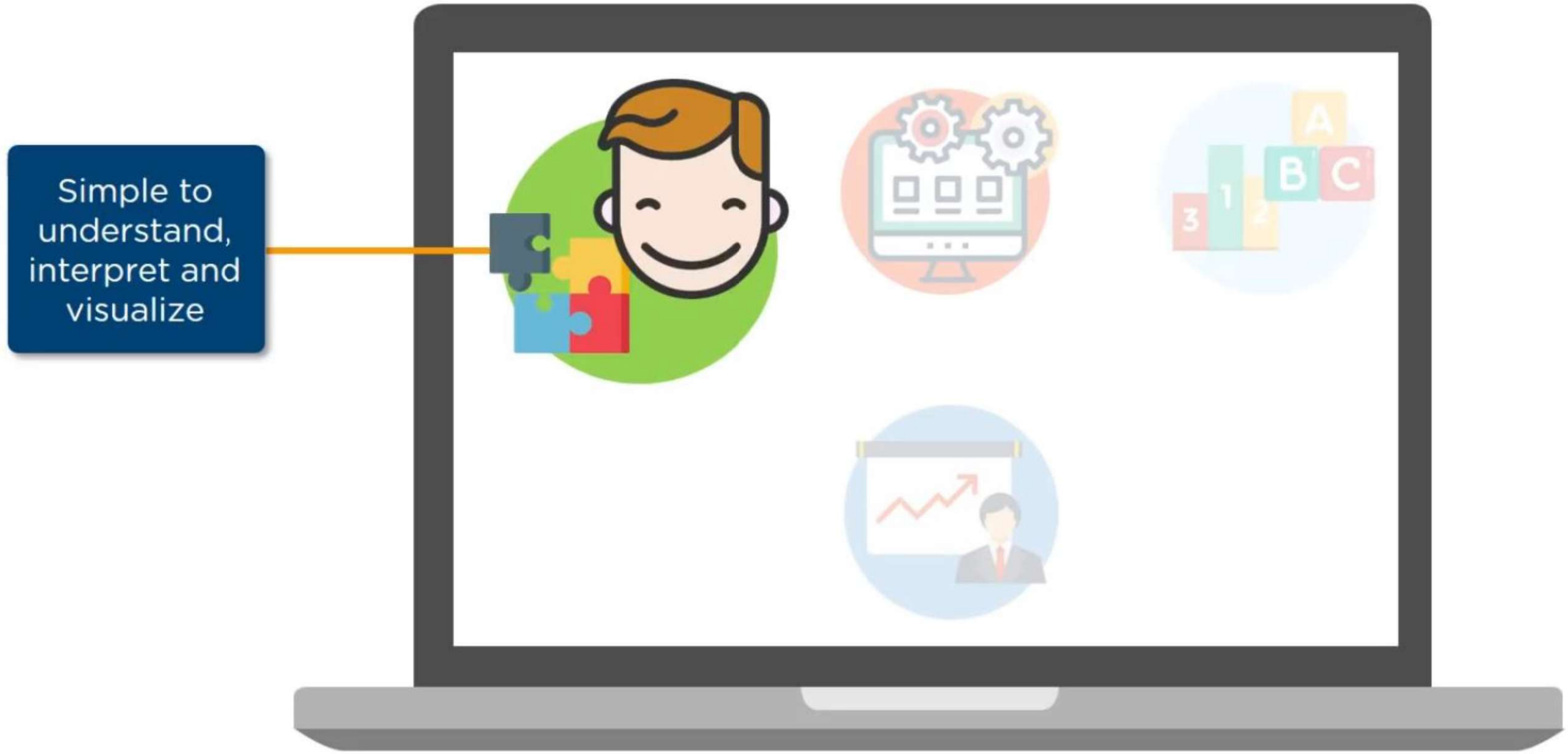
$Gain(S, a1) = 0.0817$ – *Maximum Gain*

$Gain(S, a2) = 0.0$



Example - 2
Decision Tree Algorithm – ID3
Solved Example

Advantages of Decision Tree



Advantages of Decision Tree

Little effort required for data preparation

Advantages of Decision Tree

Advantages of Decision Tree

Disadvantages of Decision Tree

Decision Trees are prone to overfitting, especially with noisy or complex datasets, leading to poor generalization on new data.

They can be unstable, as even small changes in the data may result in a completely different tree structure.

The model may also be biased towards features with more levels, skewing the importance of variables.

Decision Trees struggle to capture smooth, linear relationships between features, often resulting in step-function-like predictions.

Additionally, they tend to perform poorly on imbalanced datasets without proper tuning.

THANK YOU