UNIT II

Learning from observations

- Learning is a process that improves the knowledge of an AI program by making observations about its environment.
- Machine Learning is a subset of artificial intelligence which focuses mainly on machine learning from their experience and making predictions based on its experience.

Forms of Learning

- To understand the different types of AI learning models, we can use two of the main elements of human learning processes: **knowledge and feedback.**
- From the **knowledge perspective**, learning models can be classified based on the representation of input and output data points
- In terms of the **feedback**, AI learning models can be classified based on the interactions with the outside environment, users and other external factors.

AI Learning Models: Knowledge-Based Classification

AI learning models can be classified in two main types:

- Inductive Learning
- Deductive Learning

Inductive Learning:

- This type of AI learning model is based on inferring a general rule from datasets of input-output pairs.

Deductive Learning:

- This type of AI learning technique starts with the series of rules and infers new rules that are more efficient in the context of a specific AI algorithm.

AI Learning Models: Feedback-Based Classification

Based on the feedback characteristics, AI learning models can be classified as

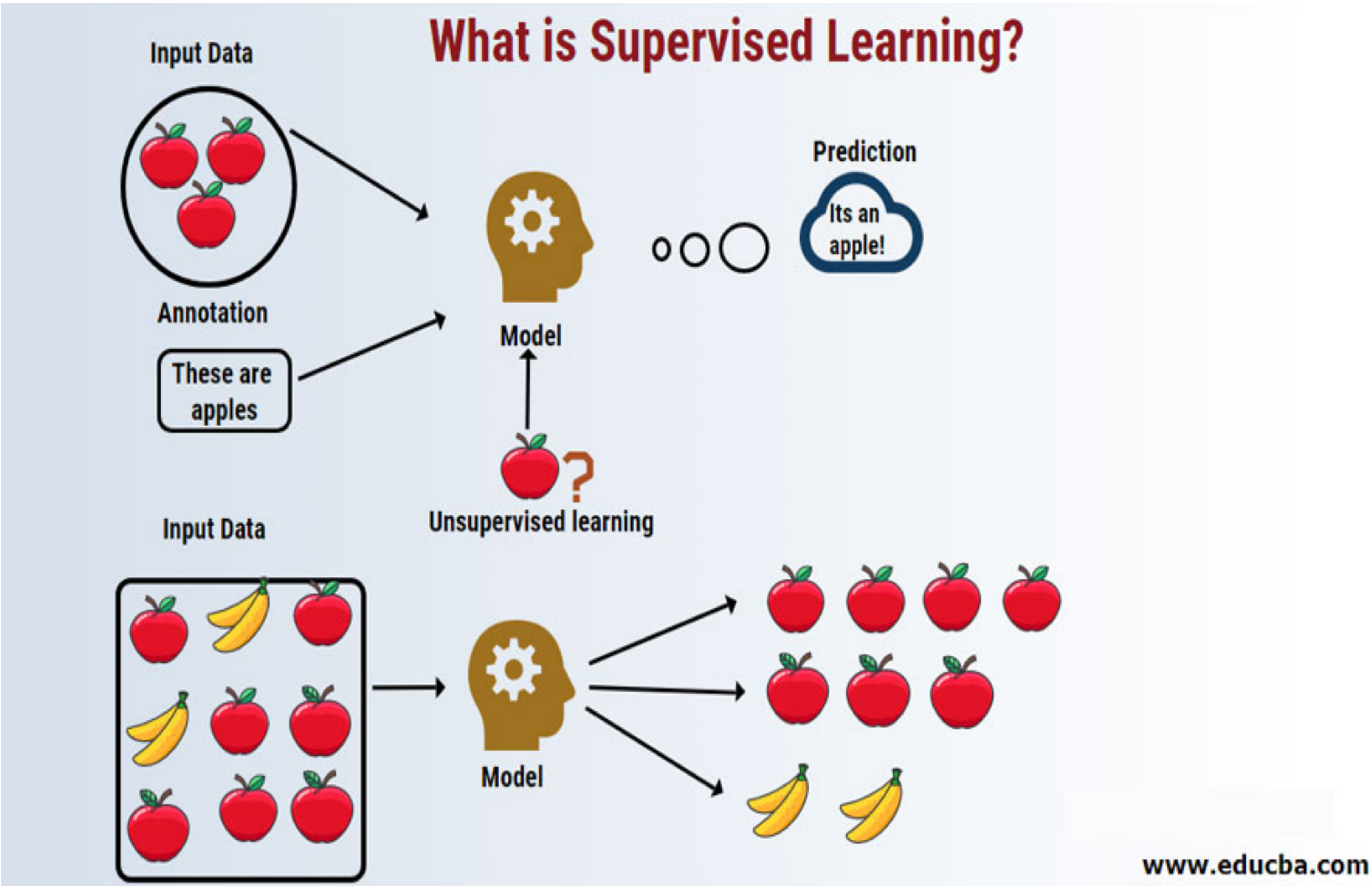
- **Supervised**
- **Unsupervised**
- **Semi-supervised or reinforced**.

Supervised Learning

- Supervised Learning is the one, where you can consider the learning is guided by a teacher. We have a dataset which acts as a teacher and its role is to train the model or the machine.
- Once the **model gets trained** it can start making a **prediction or decision** when new data is given to it.

- Multiple images of a cat, dog, orange, apple etc here the images are labelled. It is fed into the machine for training and the machine must identify the same.
- Just like a human child is shown a cat and told so, when it sees a completely different cat among others still identifies it as a cat, the same method is employed here.
- ***Popular Algorithms for Supervised Learning are:*** Linear Regression, Support Vector Machines (SVM), Neural Networks, Decision Trees, Naive Bayes, Nearest Neighbor.

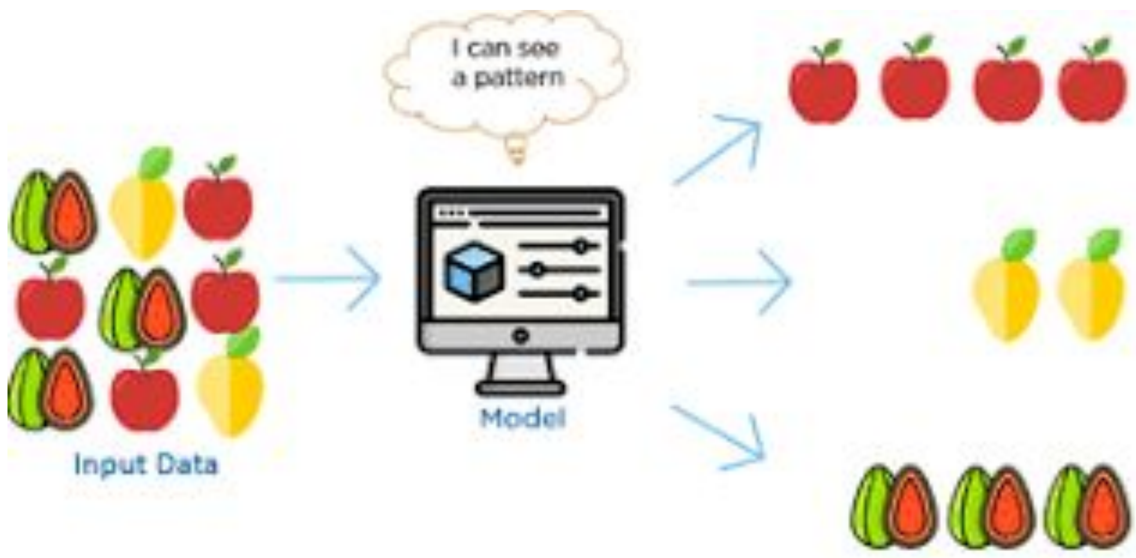
supervised


What is Supervised Learning?
Input Data
Annotation
These are apples
Model
Prediction
Its an apple!
Unsupervised learning
Input Data
Model
www.educba.com

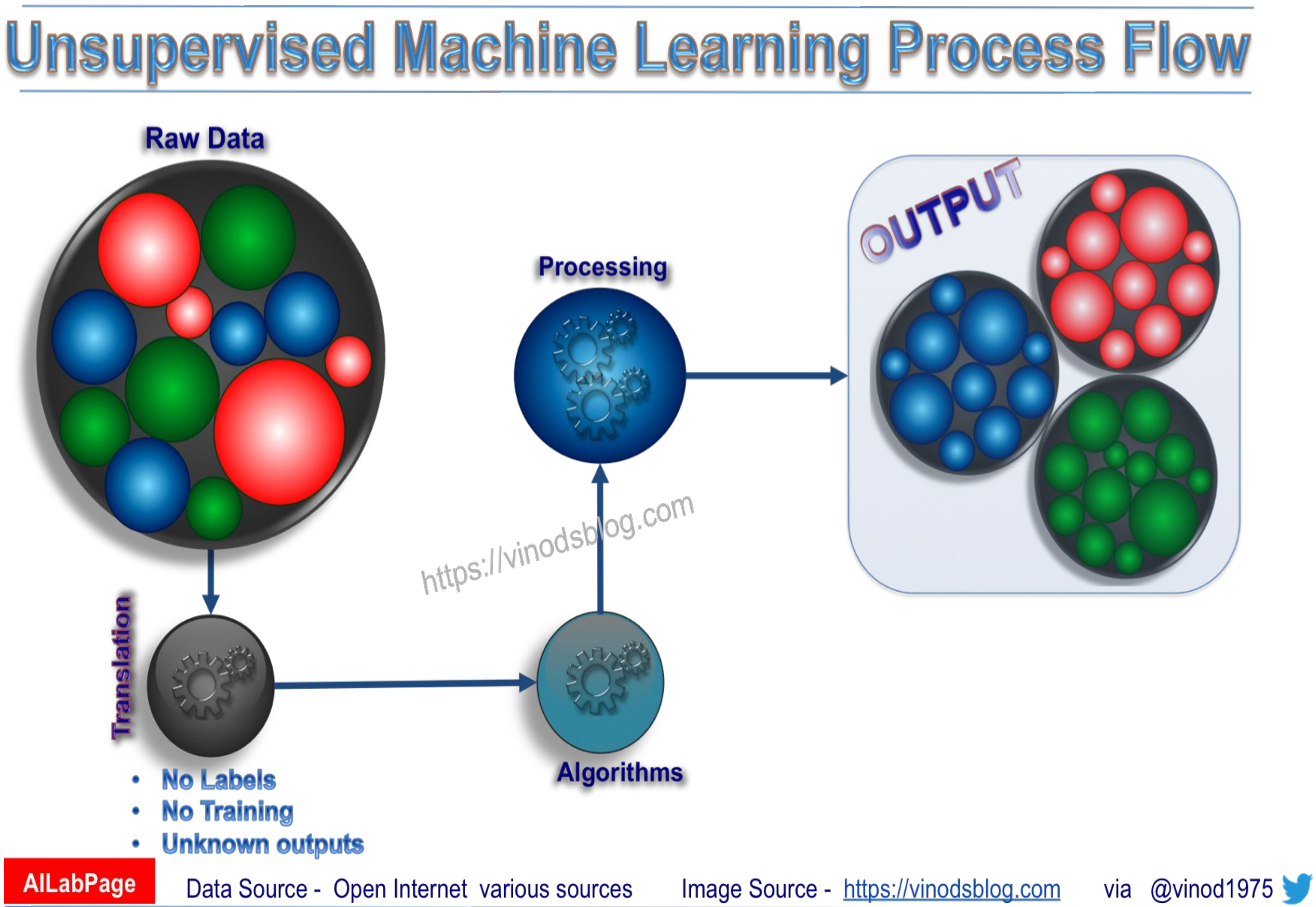
Unsupervised Learning

- The model learns through observation and finds structures in the data. Once the model is given a dataset, it automatically **finds patterns** and **relationships in the dataset** by creating clusters in it. What it cannot do is add labels to the cluster, like it cannot say this a group of apples or mangoes, but it will separate all the apples from mangoes.

- Suppose we presented images of apples, bananas and mangoes to the model, so what it does, based on some patterns and relationships it creates clusters and divides the dataset into those clusters. Now if a new data is fed to the model, it adds it to one of the created clusters.
- The main algorithms include ***Clustering algorithms and*** learning algorithms.



Unsuperwised



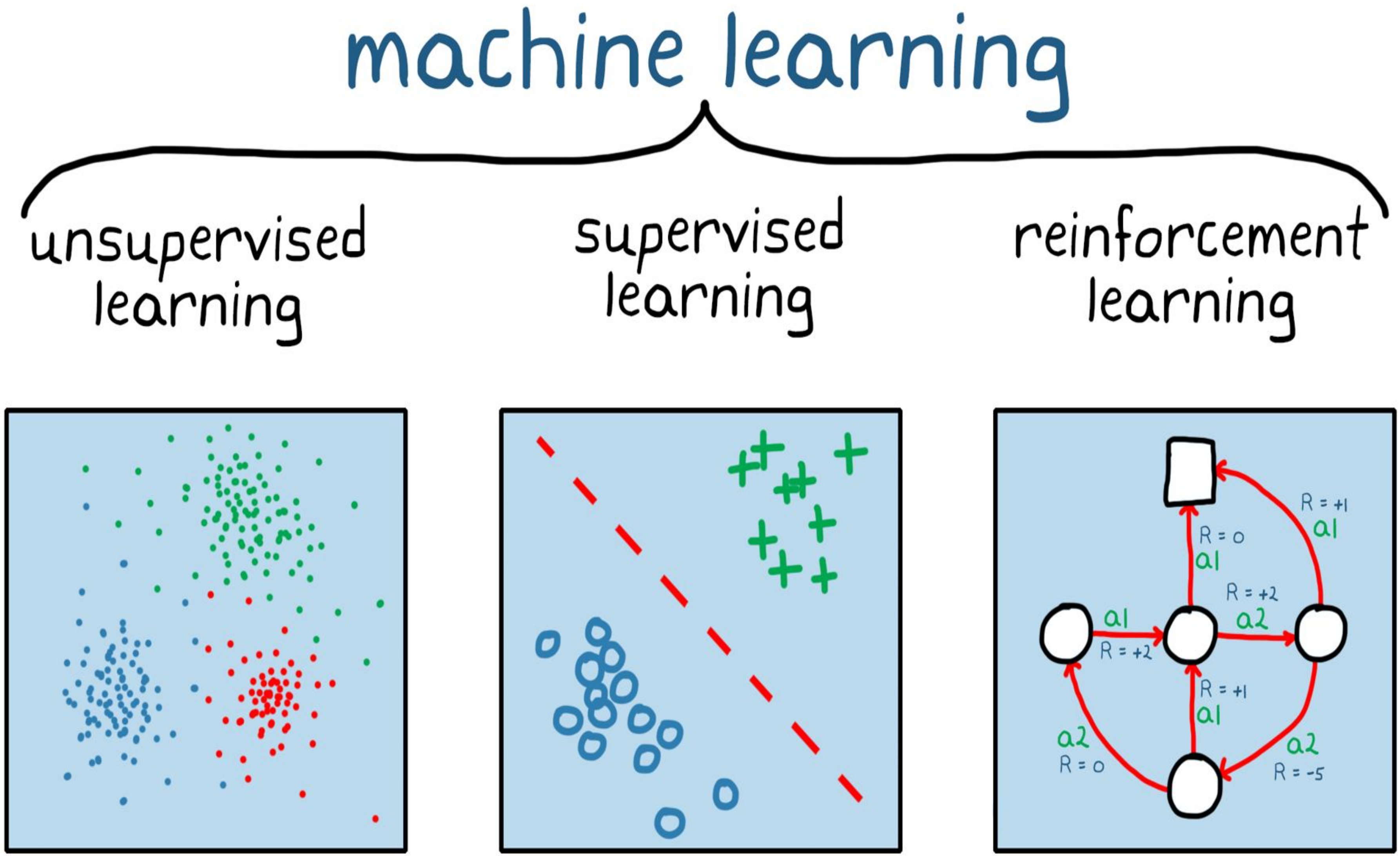
Semi-Supervised Learning

- **Semi-supervised Learning:** It is in-between that of ***Supervised and Unsupervised Learning***. Where the combination is used to produce the desired results and it is the most important in real-world scenarios where all the data available are a combination of ***labelled and unlabeled data***.

Reinforced Learning

- The machine is exposed to an ***environment where it gets trained by trial and error method***, here it is trained to make a much specific decision.
- The machine learns from **past experience** and tries to capture the best possible knowledge to make ***accurate decisions*** based on the feedback received.

Comparisons

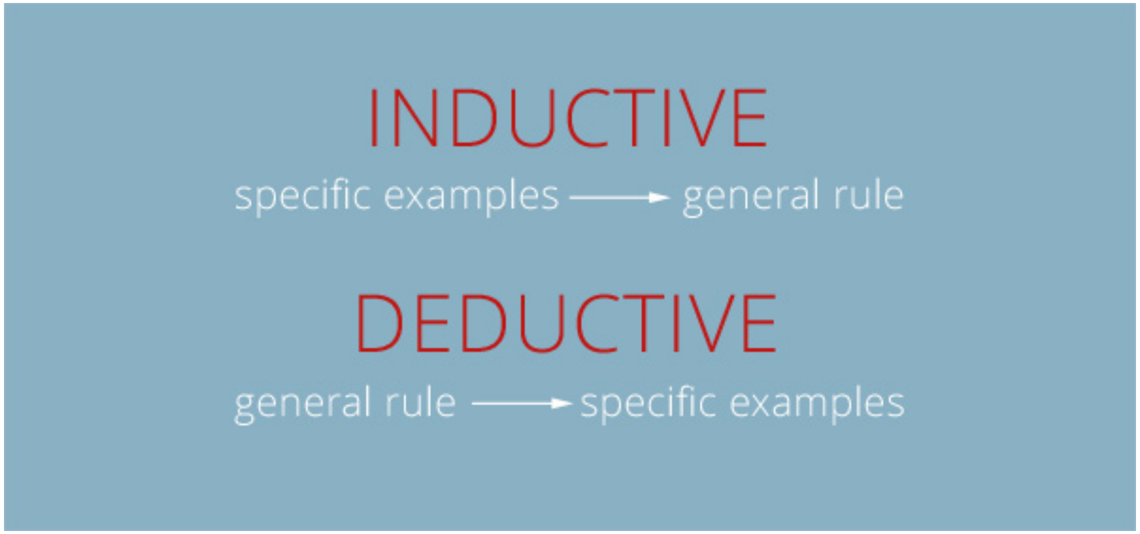


Comparisons

Criteria	Supervised ML	Unsupervised ML	Reinforcement ML
Definition	Learns by using labelled data	Trained using unlabelled data without any guidance.	Works on interacting with the environment
Type of data	Labelled data	Unlabelled data	No – predefined data
Type of problems	Regression and classification	Association and Clustering	Exploitation or Exploration
Supervision	Extra supervision	No supervision	No supervision
Algorithms	Linear Regression, Logistic Regression, SVM, KNN etc.	K – Means, C – Means, Apriori	Q – Learning, SARSA
Aim	Calculate outcomes	Discover underlying patterns	Learn a series of action
Application	Risk Evaluation, Forecast Sales	Recommendation System, Anomaly Detection	Self Driving Cars, Gaming, Healthcare

Inductive learning

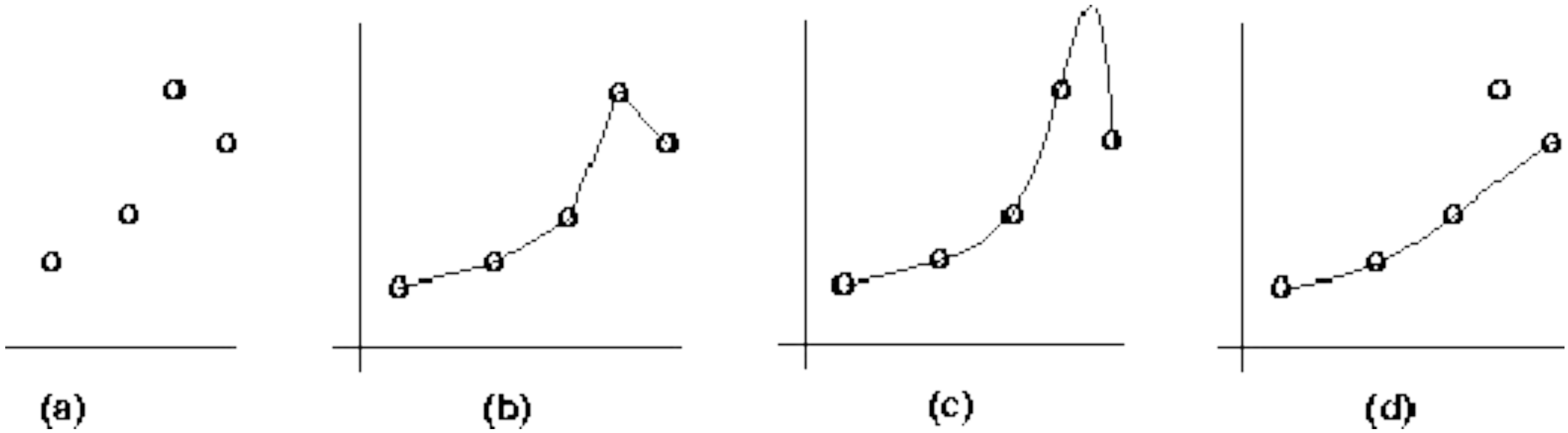
- Inductive learning, also known as discovery learning, is a process where the learner discovers rules by observing examples.



- From the perspective of inductive learning, we are given input samples (x) and output samples (f(x)) and the problem is to estimate the function (f).

- Specifically, the problem is to generalize from the samples and the mapping to be useful to estimate the output for new samples in the future.

Inductive Learning and Bias



- Suppose that we want to learn a function f(x) = y and we are given some sample (x,y) pairs, as in figure (a).
- There are several hypotheses we could make about this function, e.g.: (b), (c) and (d).
- A preference for one over the others reveals the **bias** of our learning technique, e.g.:
  - prefer piece-wise functions
  - prefer a smooth function
  - prefer a simple function and treat outliers as noise

Examples

- **Credit risk assessment**.
  - The x is the properties of the customer.
  - The f(x) is credit approved or not.
- **Disease diagnosis**.
  - The x are the properties of the patient.
  - The f(x) is the disease they suffer from.
- **Face recognition**.
  - The x are bitmaps of peoples faces.
  - The f(x) is to assign a name to the face.
- **Automatic steering**.
  - The x are bitmap images from a camera in front of the car.
  - The f(x) is the degree the steering wheel should be turned.

Decision Tree

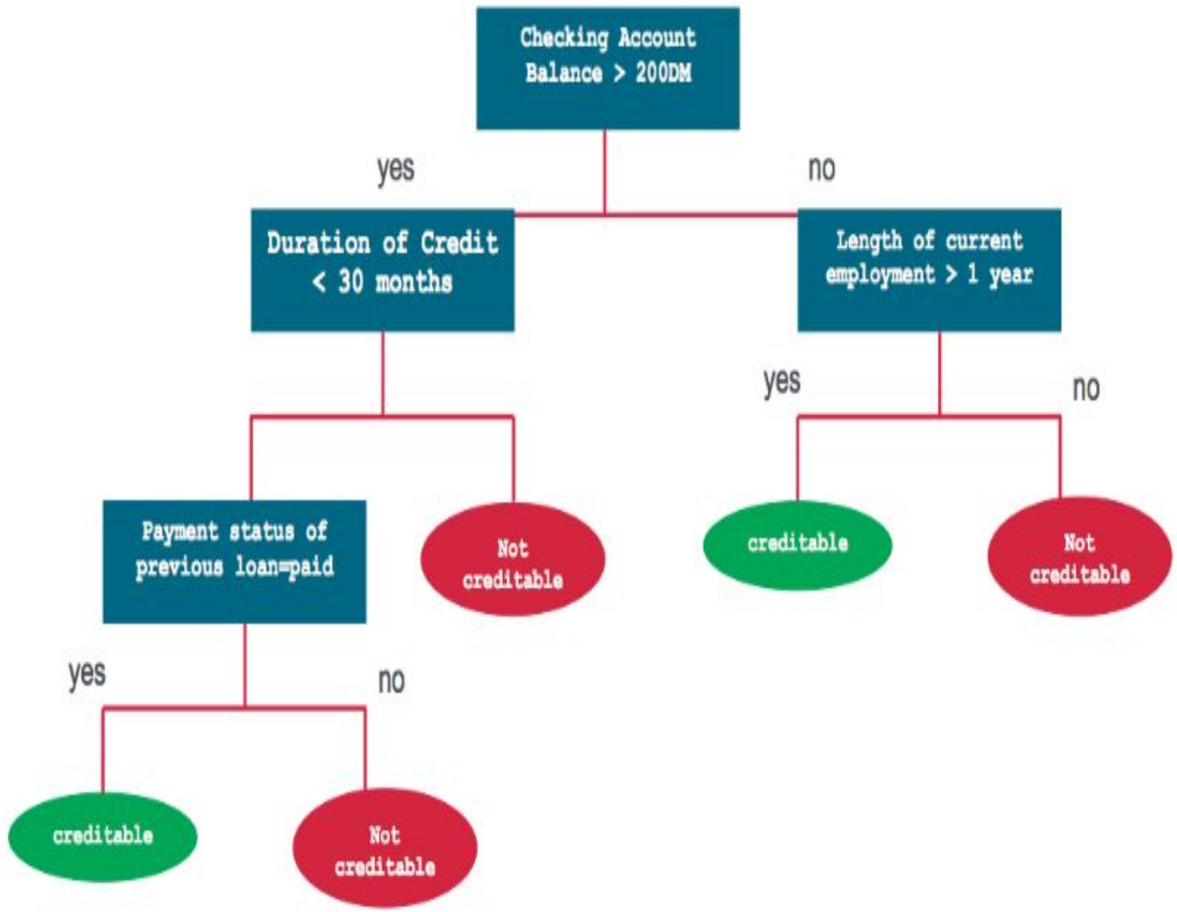
- A decision tree takes as input an object or situation described by a set of attributes and returns a decision-the predicted output value for the input.

The input attributes can be

- discrete - classification learning
- Continuous - Regression

Ex: Credit Risk Detection

- To minimize loss, the bank needs a decision rule to predict whom to give approval of the loan.
- An applicant's demographic (income, debts, credit history) and socio-economic profiles are considered.
- Data science can help banks recognize behavior patterns and provide a complete view of individual customers.



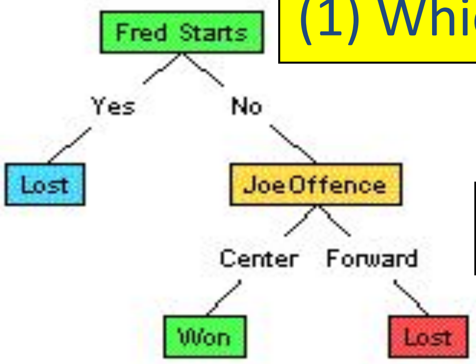
- A decision tree reaches its decision by performing a sequence of tests.
- Each internal node in the tree corresponds to a test of the value of one of the properties, and the branches from the node are labelled with the possible values of the test.
- Each leaf node in the tree specifies the value to be returned if that leaf is reached.

Definition

- Decision tree is a classifier in the form of a tree structure
  - Decision node: specifies a test on a single attribute
  - Leaf node: indicates the value of the target attribute
  - Arc/edge: split of one attribute
  - Path: a disjunction of test to make the final decision

- Decision trees classify instances or examples by starting at the root of the tree and moving through it until a leaf node.

Illustration



(1) Which to start? (root)

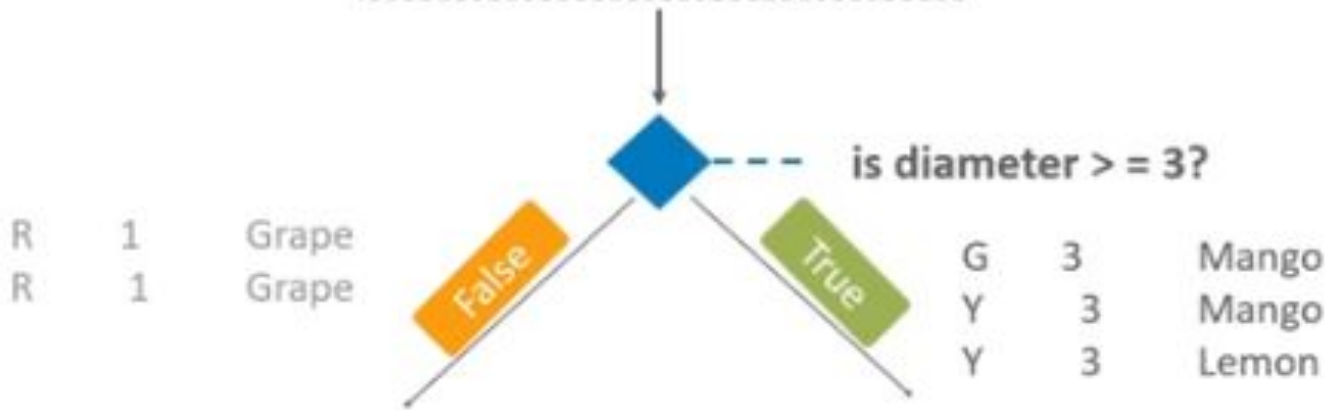
(2) Which node to proceed?

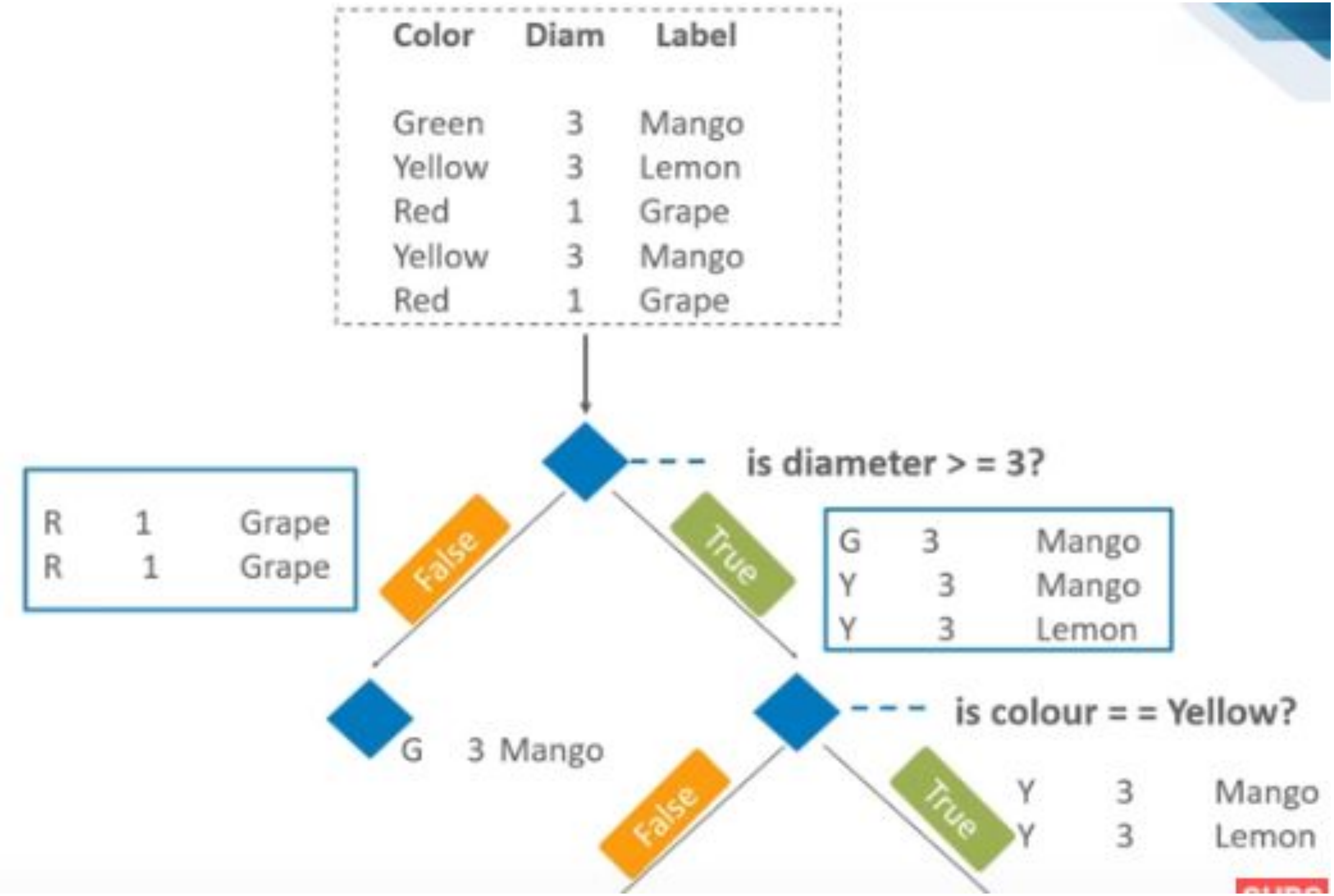
(3) When to stop/ come to conclusion?

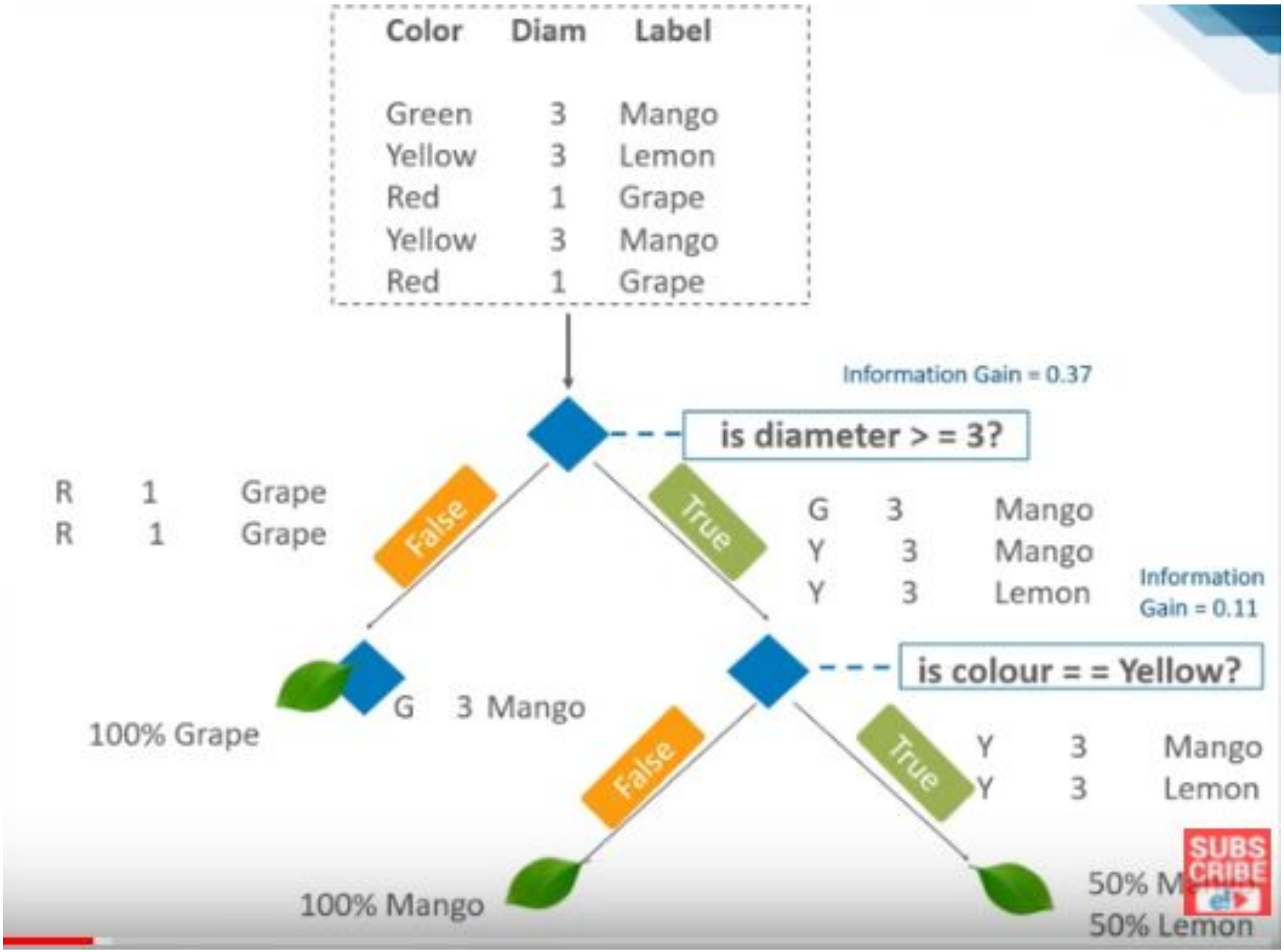
Decision Tree: Let us consider a data set as an example

Colour	Diameter	Label
Green	3	Mango
Yellow	3	Mango
Red	1	Grape
Red	1	Grape
Yellow	3	Lemon

Color	Diam	Label
Green	3	Mango
Yellow	3	Lemon
Red	1	Grape
Yellow	3	Mango
Red	1	Grape




Color Diam Label
Green 3 Mango
Yellow 3 Lemon
Red 1 Grape
Yellow 3 Mango
Red 1 Grape
is diameter > = 3?
False
True
R 1 Grape
R 1 Grape
G 3 Mango
Y 3 Mango
Y 3 Lemon
G 3 Mango
is colour = = Yellow?
False
True
Y 3 Mango
Y 3 Lemon


Color Diam Label
Green 3 Mango
Yellow 3 Lemon
Red 1 Grape
Yellow 3 Mango
Red 1 Grape
Information Gain = 0.37
is diameter > = 3?
False
True
R 1 Grape
R 1 Grape
G 3 Mango
Y 3 Mango
Y 3 Lemon
Information Gain = 0.11
is colour = = Yellow?
100% Grape
G 3 Mango
False
True
Y 3 Mango
Y 3 Lemon
100% Mango
50% Lemon

Green	3	Mango
Yellow	3	Lemon
Yellow	3	Mango

Is the colour green?

Is the diameter >=3

Is the colour yellow

TRUE

False

Is the colour green?

Is the diameter >=3

Is the colour yellow

TRUE			False		
Green	3	Mango	Yellow	3	Lemon
			Yellow	3	Mango

Let's take an example,

We have taken dataset consisting of:

- Weather information of last 14 days
- Whether match was played or not on that particular day

Now using the decision tree we need to predict whether the game will happen if the weather condition is

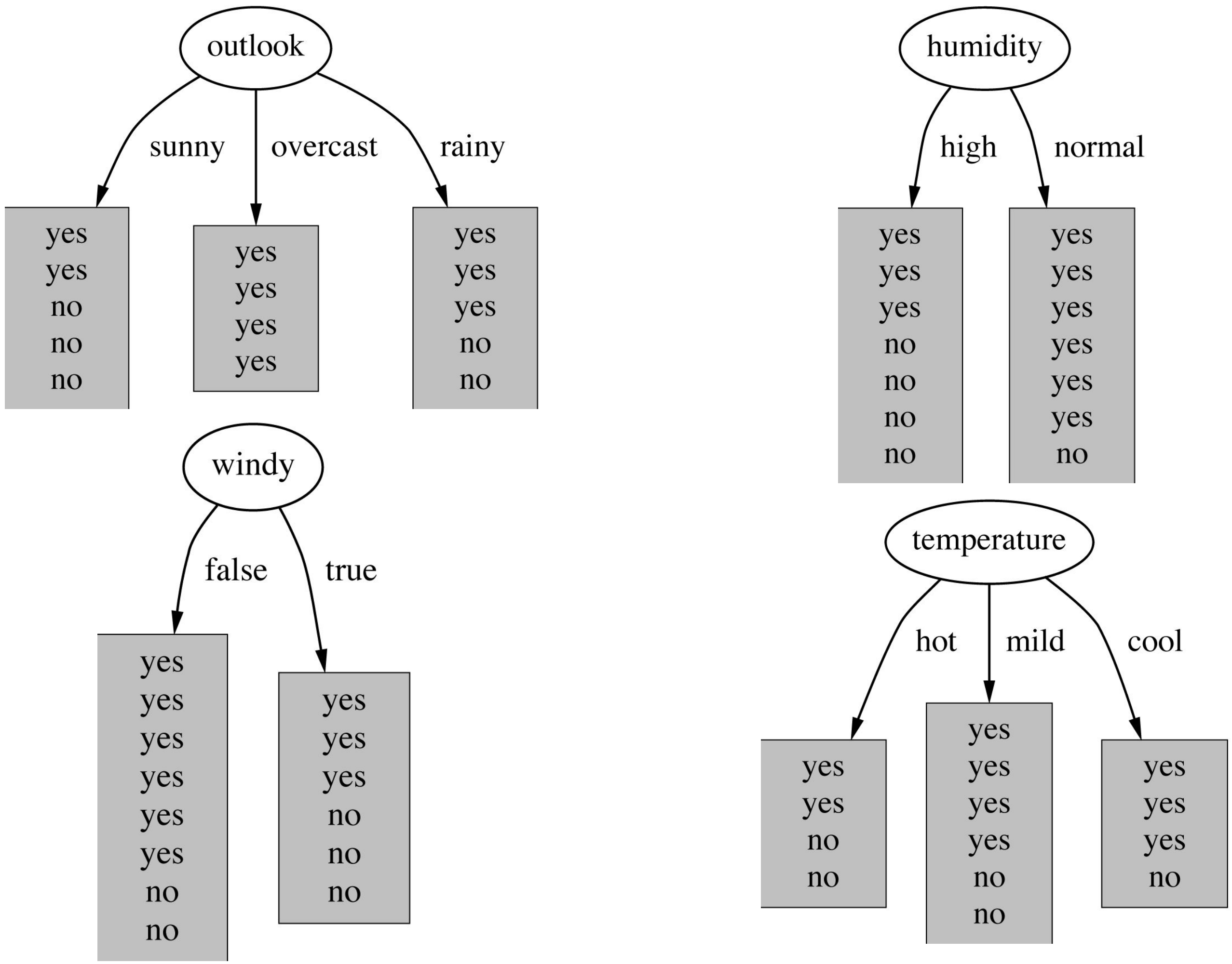
Outlook	=	Rain
Humidity	=	High
Wind	=	Weak
Play	**=**	**?**

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

Outlook	Temperature	Humidity	Wind	Played football(yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Here There are for independent variable to determine the dependent variable. The independent variables are Outlook, Temperature, Humidity, and Wind. The dependent variable is whether to play football or not.

Which attribute to select?



witten&eib

But How do we choose the best attribute?

Or

How does a tree decide where to split?

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

How does a Decision Tree decide where to split?

- Decision trees use **multiple algorithms** to decide to split a node in two or more sub-nodes. Decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes.

Multiple Algorithms are

- Gini Index
- Chi-Square
- Information Gain
- Reduction in Variance

To draw a decision tree from a dataset of some attributes:

- Each node corresponds to a **splitting attribute**.
- Each arc is a possible value of that attribute.
- **Splitting attribute** is selected to be the most informative among the attributes.
- **Entropy** is a factor used to measure how informative is a node.
- The algorithm uses the criterion of **information gain** to determine the goodness of a split.
- The attribute with the greatest information gain is taken as the splitting attribute, and the data set is split for all distinct values of the attribute values of the attribute.

Entropy

- A measure that characterize the purity of a dataset.
- Given a training set D with two classes, P and S of some target.

Let $p = \frac{number\ of\ P}{number\ of\ records}$,and $s = \frac{number\ of\ S}{number\ of\ records}$

$$\boldsymbol{E(D) = -p\log_2 p - s\, log_2\, s}$$

- For example if we have a collection D of 14 records including 9 records of **P** and 5 of **S**.

$p = \frac{9}{14} = 0.643$,and $s = \frac{5}{14} = 0.357$

$$\boldsymbol{E(D) = -(0.643)\, log_2(0.643) - (0.357)\, log_2(0.357) = 0.94}$$

#	Decision
1	P
2	S
3	P
4	S
5	P
6	P
7	P
8	S
9	S
10	P
11	P
12	S
13	P
14	P

Entropy

- If all records belongs to the same class.
  - Entropy = 0
- If records are equally distributed over collection class.
  - Entropy = 1
- General Formula of Entropy for a dataset of (k) classes:

$$E(D) = \sum_{i=1}^{k} -p_i \log_2 p_i$$

For example if k=4

$$E(D) = -p_1 \, log_2 \, p_1 - p_2 \, log_2 \, p_2 - p_3 \, log_2 \, p_3 - p_4 \, log_2 \, p_4$$

Information Gain

Given Entropy of a collection we can define a measure of the effectiveness of an attribute in classification.

Information Gain of an attribute is the expected reduction in entropy caused by partitioning the collection according this attribute.

$$\boldsymbol{Gain(D,V) = E(D) - \sum \frac{D_v}{D} \times E(D_v)}$$

$$Gain(D,V) = E(D) - \left(\frac{8}{14}E(X) + \frac{6}{14}E(Y)\right) = 0.94 - (0.892) = 0.048$$

#	Decision	Att. (V)
1	P	X
2	S	X
3	P	Y
4	S	Y
5	P	X
6	P	X
7	P	X
8	S	Y
9	S	X
10	P	Y
11	P	X
12	S	Y
13	P	Y
14	P	X

Step 1: Compute the entropy for the Data set

Out of 14 instances we have 9 YES and 5 NO

So we have the formula,

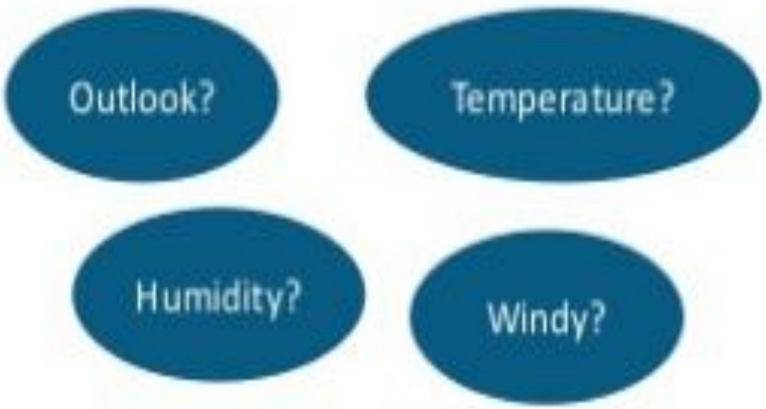
$E(S) = -P(Yes)\log_2 P(Yes) - P(No)\log_2 P(No)$

$E(S) = -(9/14)^* \log_2 9/14 - (5/14)^* \log_2 5/14$

$E(S) = 0.41 + 0.53 = 0.94$

	outlook	temp.	humidity	windy	play
D1	sunny	hot	high	false	no
D2	sunny	hot	high	true	no
D3	overcast	hot	high	false	yes
D4	rainy	mild	high	false	yes
D5	rainy	cool	normal	false	yes
D6	rainy	cool	normal	true	no
D7	overcast	cool	normal	true	yes
D8	sunny	mild	high	false	no
D9	sunny	cool	normal	false	yes
D10	rainy	mild	normal	false	yes
D11	sunny	mild	normal	true	yes
D12	overcast	mild	high	true	yes
D13	overcast	hot	normal	false	yes
D14	rainy	mild	high	true	no

Which Node To Select As Root Node?



outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Which Node To Select As Root Node: Outlook

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Which Node To Select As Root Node: Outlook

E(Outlook = Sunny) = $-2/5 \log_2 2/5 - 3/5 \log_2 3/5 = 0.971$

E(Outlook = Overcast) = $-1 \log_2 1 - 0 \log_2 0 = 0$

E(Outlook = Sunny) = $-3/5 \log_2 3/5 - 2/5 \log_2 2/5 = 0.971$

Information from outlook,

I(Outlook) = 5/14 x 0.971 + 4/14 x 0 + 5/14 x 0.971 = 0.693

Information gained from outlook,

Gain(Outlook) = E(S) – I(Outlook)

0.94 – 0.693 = 0.247

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Which Node To Select As Root Node: Outlook

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Which Node To Select As Root Node: Windy

E(Windy = True) = 1

E(Windy = False) = 0.811

Information from windy,

I(Windy) = 8/14 x 0.811+ 6/14 x 1 = 0.892

Information gained from outlook,

Gain(Windy) = E(S) – I(Windy)

0.94 – 0.892 = 0.048

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Similarly we need to calculate for the rest two

Outlook:

Info 0.693

Gain: 0.940-0.693 $\boxed{0.247}$

Temperature:

Info 0.911

Gain: 0.940-0.911 0.029

Humidity:

Info 0.788

Gain: 0.940-0.788 0.152

Windy:

Info 0.892

Gain: 0.940-0.982 0.048

Since Max gain = 0.247,

Outlook is our ROOT Node

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

This Is How Your Complete Tree Will Look Like

Decision Tree

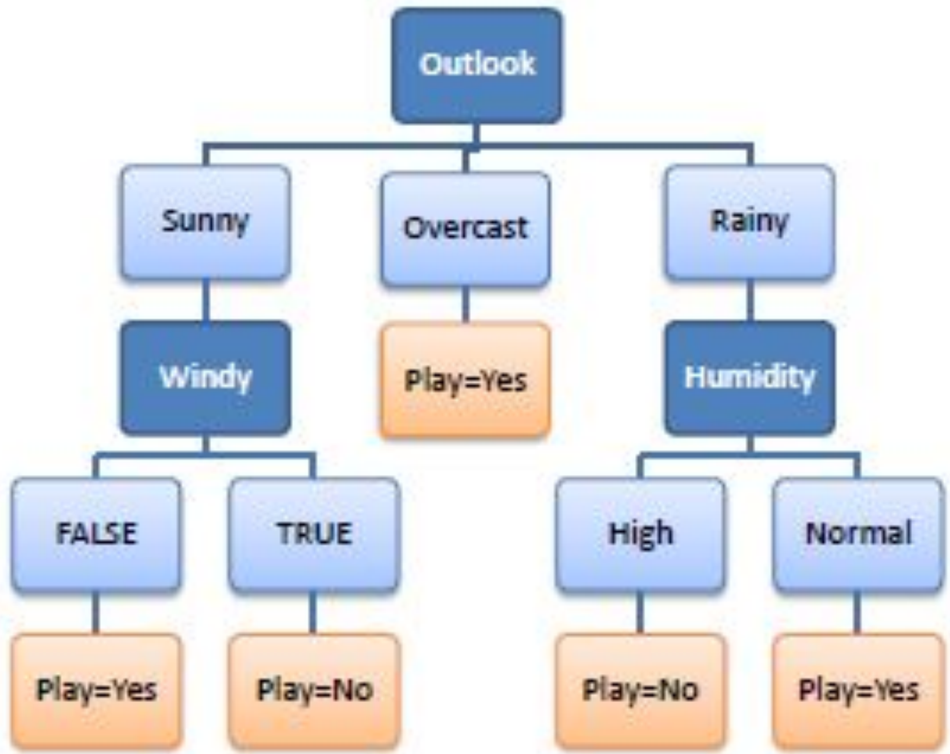
R_1: IF (Outlook=Sunny) AND (Windy=FALSE) THEN Play=Yes

R_2: IF (Outlook=Sunny) AND (Windy=TRUE) THEN Play=No

R_3: IF (Outlook=Overcast) THEN Play=Yes

R_4: IF (Outlook=Rainy) AND (Humidity=High) THEN Play=No

R_5: IF (Outlook=Rain) AND (Humidity=Normal) THEN Play=Yes



Why Learning Works

- Problem: How can you know if a theory will accurately predict the future?

OR

- How can you know that a hypothesis is close to the target function if you don't know what the target function is?
- Answers provided by Computational Learning Theory

Goal of learning theory

- To understand
- What kinds of tasks are learnable
- What kind of data is required for learnability
- What are the (time, space) requirements for learnability
- To develop and analyse models
- Develop algorithms that probably meet desired models.
- Prove guarantee for success algorithms