UNIT 3

Supervised Learning

By Dr. Purvi Tandel

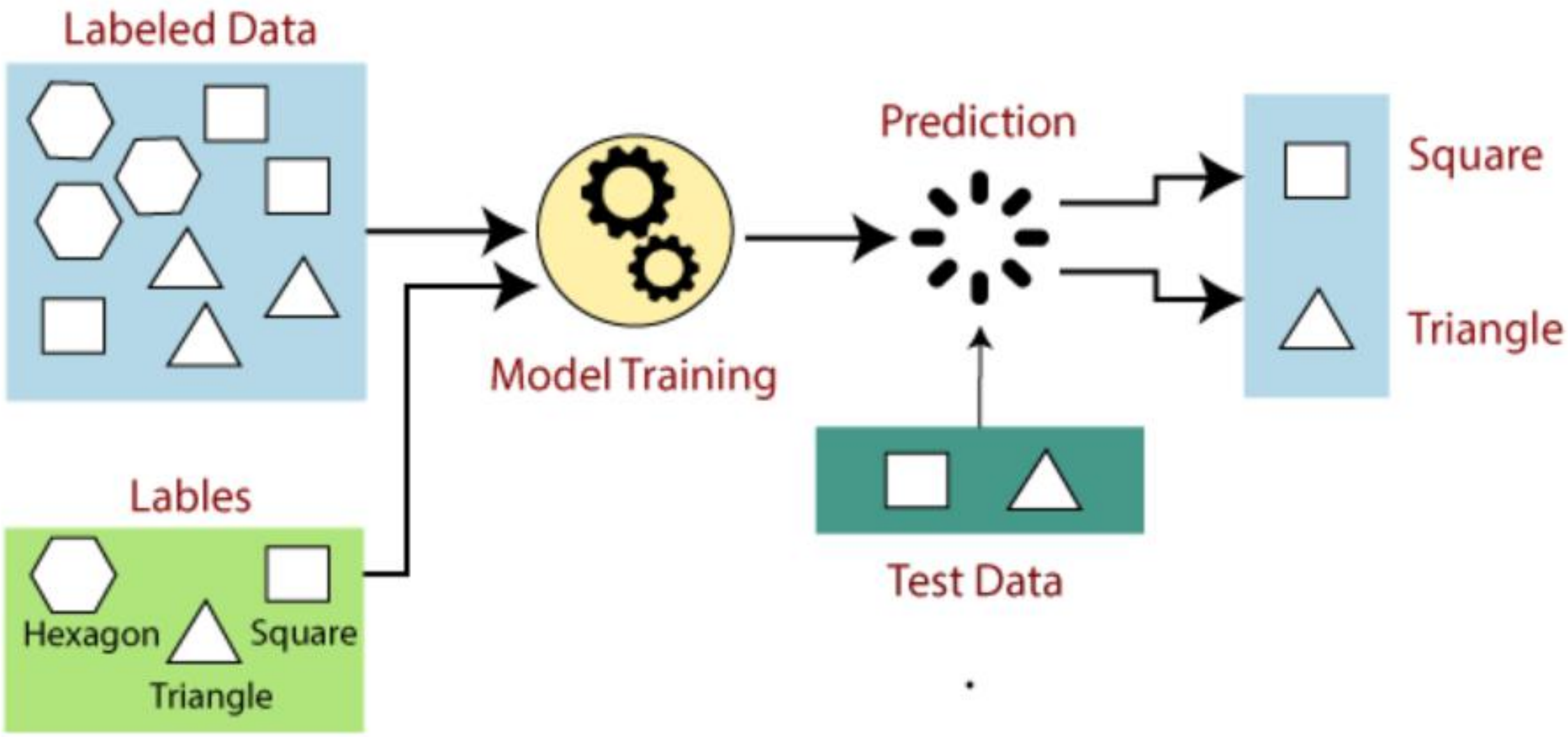
Topics:

- Linear Regression
- Logistic Regression
- K Nearest Neighbours
- Overfitting and Regularization
- Support Vector Machine
- Decision Trees

Supervised Learning

Supervised learning is the type of machine learning in which machines are trained using well "**labelled**" training data, and based on that data, machines predict the output.

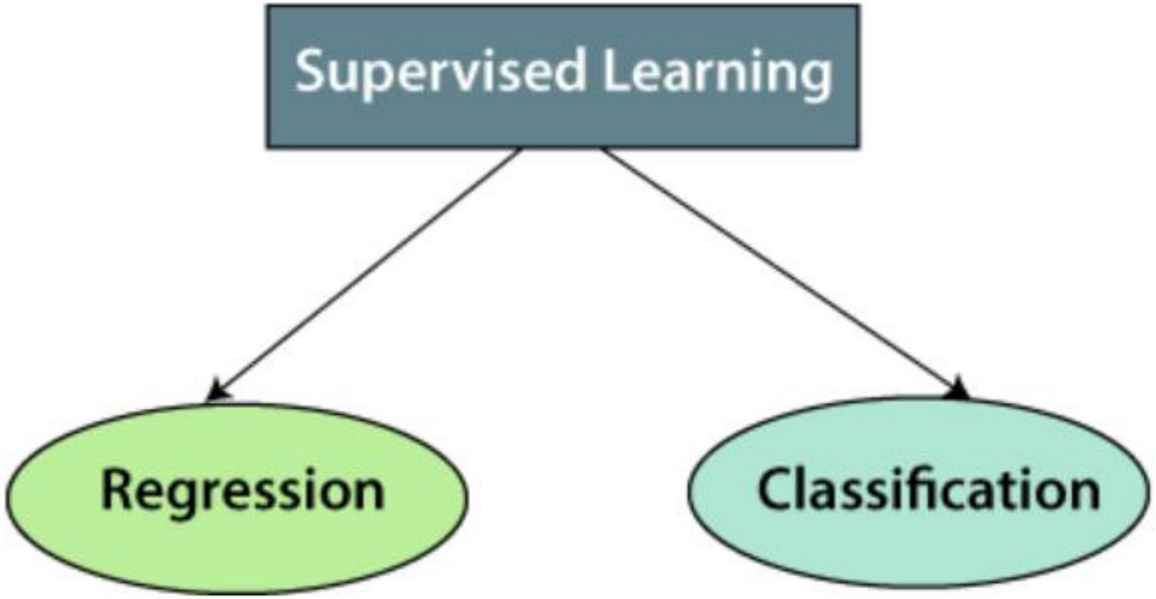
The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).



Types of supervised Machine learning Algorithms:

Regression is a statistical technique that finds a relationship between dependent and independent variables.

Classification algorithms are used when the output variable is two categorical, which means there are classes such that Yes-No, Male-Female, True-False, etc.



Types of supervised Machine learning Algorithms:

Regression:

Regression is a statistical technique that finds a relationship between dependent and independent variables.

Classification:

Classification algorithms are used when the output variable is two categorical, which means there are classes such that Yes-No, Male-Female, True-False, etc.

Linear Regression:

Linear regression is the most commonly used regression model in machine learning.

It may be defined as the statistical model that analyzes the linear relationship between a dependent variable with a given set of independent variables.

Linear regression is further divided into two subcategories:

- Simple linear regression
- Multiple linear regression

In simple linear regression, a single independent variable (or predictor) is used to predict the dependent variable.

Linear Regression:

Simple linear regression in machine learning is a type of linear regression. When the linear regression algorithm deals with a **single independent variable**, it is known as simple linear regression.

When there is **more than one independent variable** (feature variables), it is known as multiple linear regression.

Mathematically, the **simple linear regression** can be represented as follows –

$$Y=mX+b$$

Where,

- Y is the dependent variable we are trying to predict.
- X is the dependent variable we are using to make predictions.
- m is the slope of the regression line, which represents the effect X has on Y.
- b is a constant known as the Y-intercept. If $X = 0$, Y would be equal to b.

Linear Regression:

Independent Variable

- The feature inputs in the dataset are termed as the independent variables. There is only a single independent variable in simple linear regression. An independent variable is also known as a **predictor variable** as it is used to predict the target value. It is plotted on a **horizontal axis**.

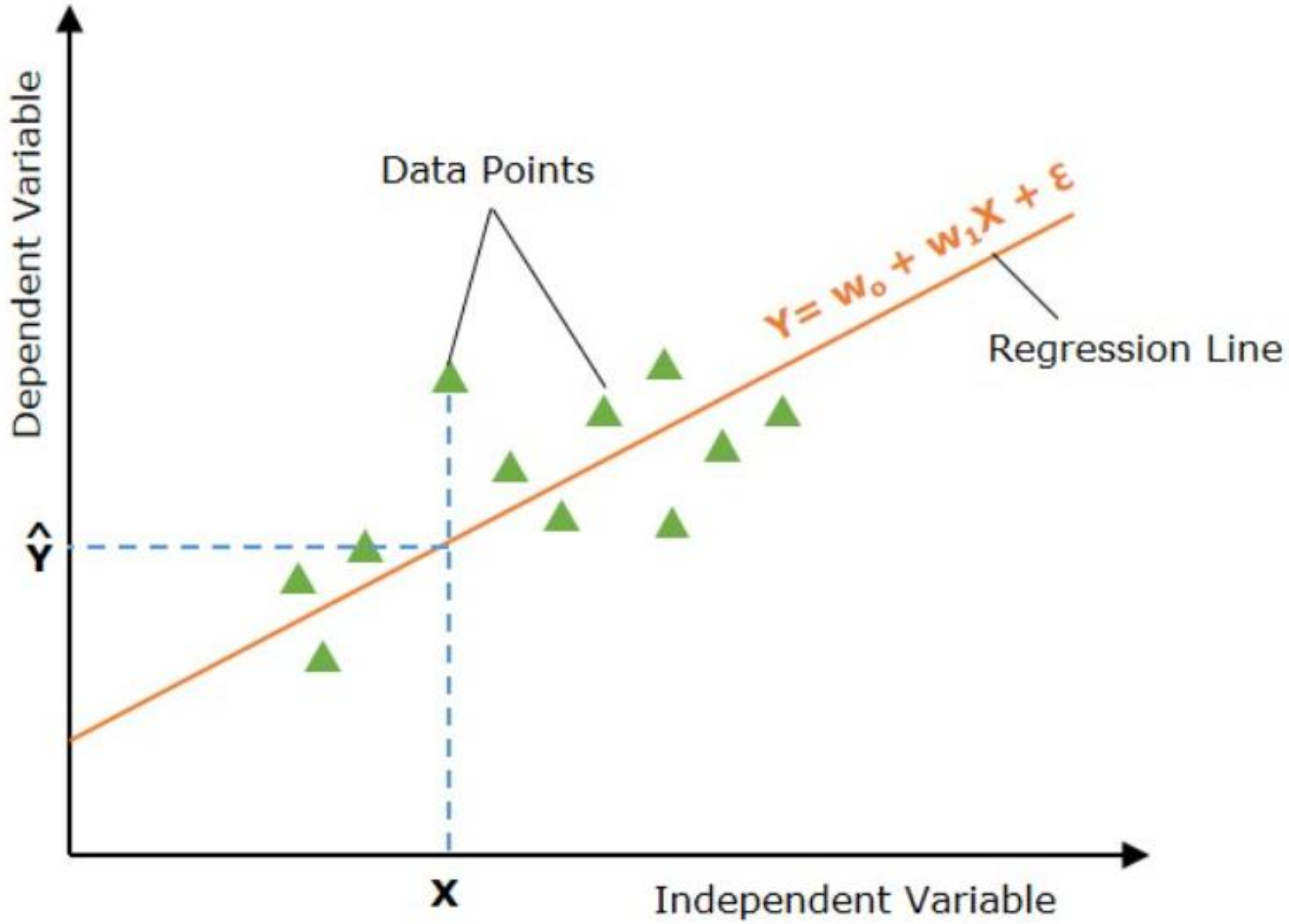
Dependent Variable

- The target value in the dataset is termed as the dependent variable. It is also known as a response variable or **predicted variable**. It is plotted on a **vertical axis**.

Line of Regression

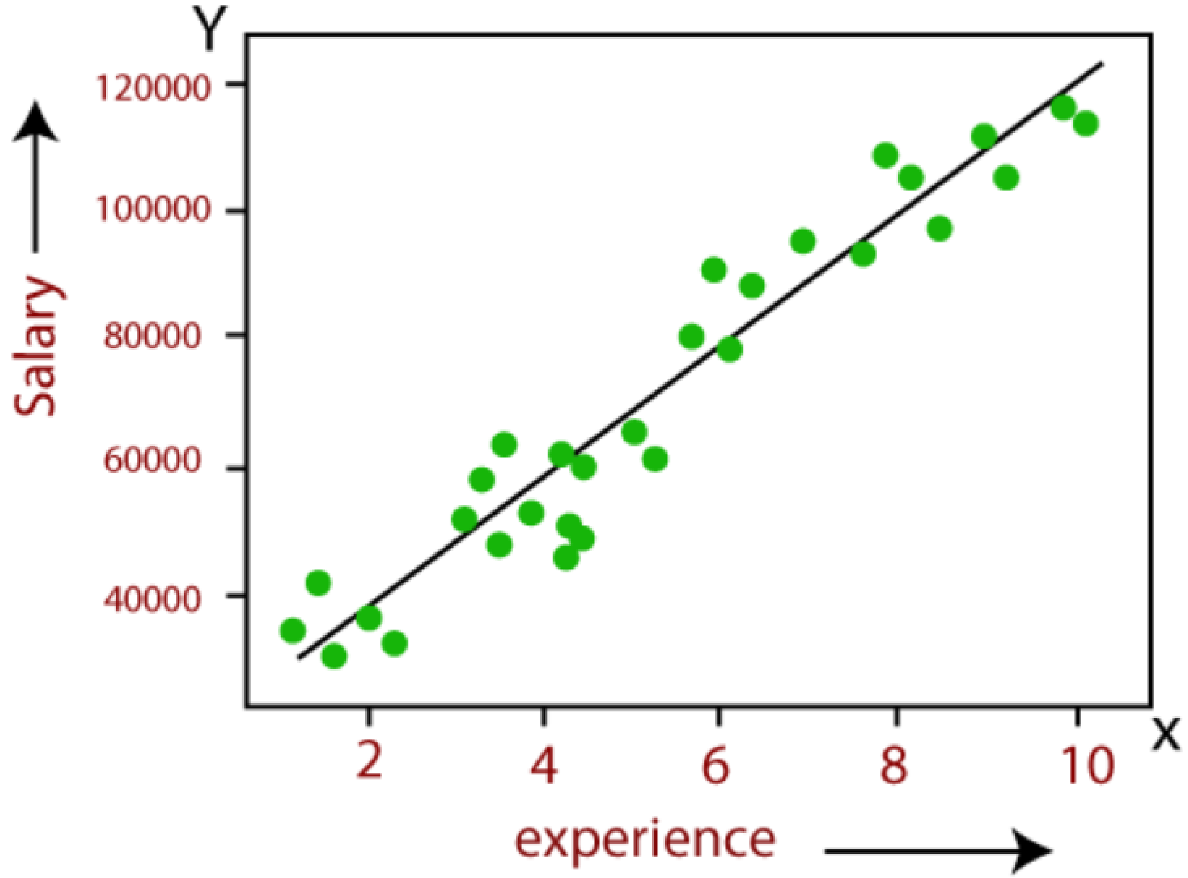
- In simple linear regression, a line of regression is a straight line that best fits the data points and is used to show the relationship between a dependent variable and an independent variable.

Linear Regression:



Linear Regression:

Here we are predicting the salary of an employee on the basis of **the year of experience**.



Linear Regression:

Example:

A company collects the following data to understand the relationship between the number of hours their employees spend in training (x) and their productivity score (y).

Perform linear regression to find the equation of the line $y=mx+c$.

Employee	Training Hours (x)	Productivity Score (y)
1	2	3
2	4	5
3	6	7
4	8	8
5	10	11

Linear Regression:

Employee	Training Hours (x)	Productivity Score (y)
1	2	3
2	4	5
3	6	7
4	8	8
5	10	11

Step 1: Calculate means of x and y

$$\text{Mean of } x, X = \frac{\text{Sum of } x}{\text{Number of data points}} = \frac{2+4+6+8+10}{5} = 6$$

$$\text{Mean of } y, Y = \frac{\text{Sum of } y}{\text{Number of data points}} = \frac{3+5+7+8+11}{5} = 6.8$$

Linear Regression:

Step 2: Calculate (x_i-X), (y_i-Y), $(x_i-X)^2$, and $(x_i-X)(y_i-Y)$

Create a table:

x_i	y_i	x_i-X	y_i-Y	$(x_i-X)^2$	$(x_i-X)(y_i-Y)$
2	3	-4	-3.8	16	15.2
4	5	-2	-1.8	4	3.6
6	7	0	0.2	0	0
8	8	2	1.2	4	2.4
10	11	4	4.2	16	16.8

Linear Regression:

Step 3: Compute sums

$$\sum(x_i - X)^2 = 16 + 4 + 0 + 4 + 16 = 40$$

$$\sum(x_i - X)(y_i - Y) = 15.2 + 3.6 + 0 + 2.4 + 16.8 = 38$$

Step 4: Calculate slope m

The formula for the slope is:

$$m = \frac{\sum(x_i - X)(y_i - Y)}{\sum(x_i - X)^2}$$

Substitute the values:

$$m = \frac{38}{40} = 0.95$$

Linear Regression:

Step 5: Calculate intercept c

The formula for the intercept is:

$$c = Y - mX$$

Substitute $Y = 6.8$, $m = 0.95$, and $X = 6$:

$$c = 6.8 - (0.95 \times 6) = 6.8 - 5.7 = 1.1$$

Step 6: Write the regression equation

The linear regression equation is:

$$y = 0.95x + 1.1$$

Linear Regression:

Step 6: Write the regression equation

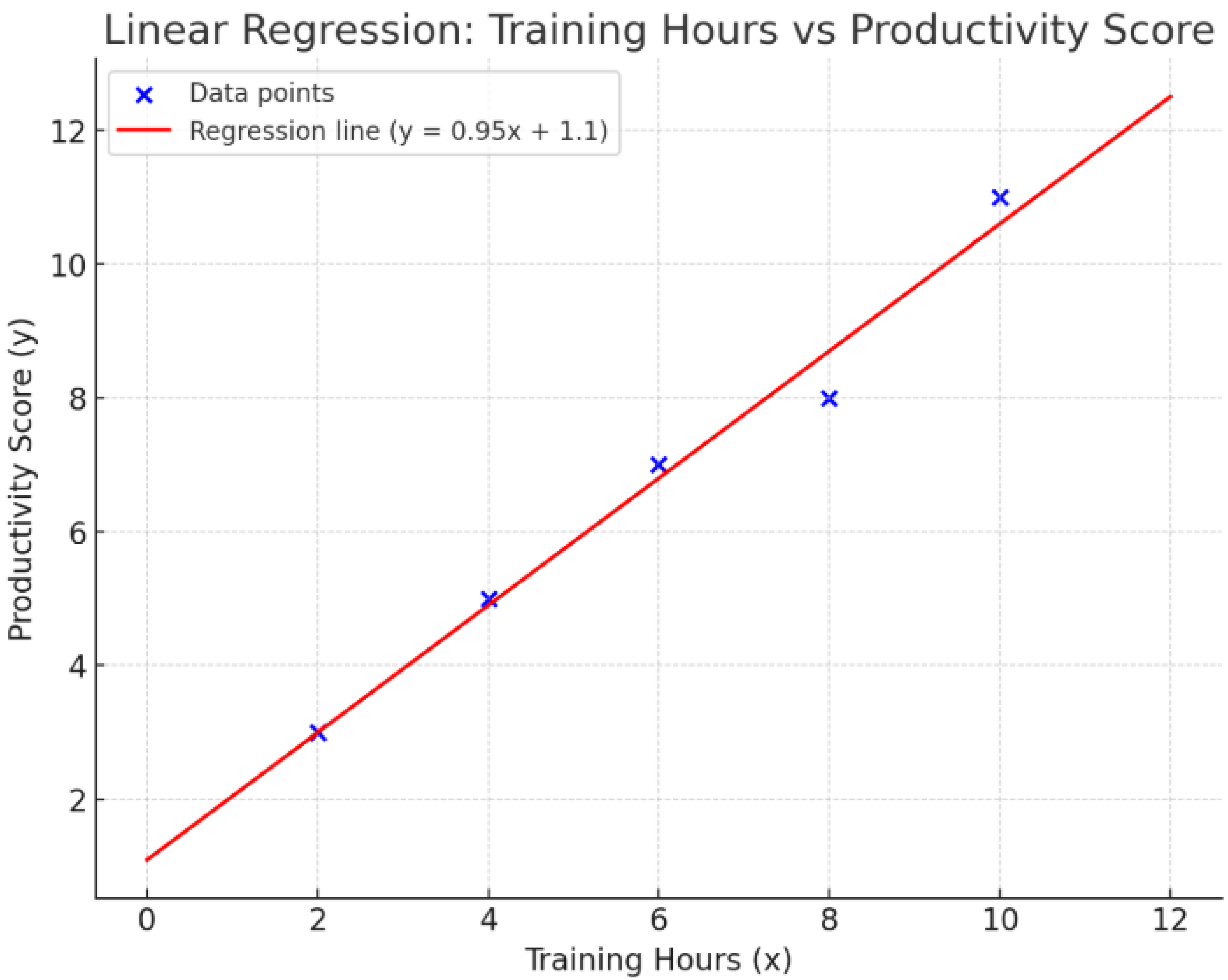
The linear regression equation is:

$$y = 0.95x + 1.1$$

The slope m = 0.95 means that for every additional hour of training, the productivity score increases by 0.95 on average.

The intercept c = 1.1 means that the predicted productivity score is 1.1 when the training hours are 0.

Linear Regression:



Linear Regression:

Applications of simple linear regression:

Sales Prediction
Predicting sales revenue (Y) based on advertising spending (X).

Temperature Forecasting
Estimating the temperature (Y) based on the time of day (X).

Crop Yield Estimation
Predicting crop yield (Y) based on rainfall (X).

Health Impact Assessment
Estimating blood pressure (Y) based on daily salt intake (X).

Linear Regression:

Multiple Linear Regression:

In machine learning, multiple linear regression (MLR) is a statistical technique that is used to predict the outcome of a dependent variable based on the values of multiple independent variables.

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_nX_n$$

Where:

- Y: Dependent variable (what we are trying to predict).
- $X_1, X_2, ..., X_n$: Independent variables (factors influencing Y).
- b_0: Intercept of the regression line.
- $b_1, b_2, ..., b_n$: Coefficients for each independent variable.

Linear Regression:

Example: Predicting House Prices

Suppose we want to predict the price of a house (Y) based on:

Size of the house (X1) in square feet.

Number of bedrooms (X2)

Distance to the city center (X3) in miles

House ID	Size (X1) (sq ft)	Bedrooms (X2)	Distance (X3) (miles)	Price (Y) ($)
1	1500	3	2	300,000
2	1800	4	5	320,000
3	2000	4	3	340,000
4	2100	5	4	360,000
5	2500	5	6	400,000

Parameter	Linear (Simple) Regression	Multiple Regression
Definition	Models the relationship between one dependent and one independent variable.	Models the relationship between one dependent and two or more independent variables.
Equation	$Y = C_0 + C_1X + e$	$Y = C_0 + C_1X_1 + C_2X_2 + C_3X_3 + ... + C_nX_n + e$
Complexity	It is simpler to deal with one relationship.	More complex due to multiple relationships.
Use Cases	Suitable when there is one clear predictor.	Suitable when multiple factors affect the outcome.
Assumptions	Linearity, Independence, Homoscedasticity, Normality	Same as linear regression, with the added concern of multicollinearity.
Visualization	Typically visualized with a 2D scatter plot and a line of best fit.	Requires 3D or multi-dimensional space, often represented using partial regression plots.
Risk of Overfitting	Lower, as it deals with only one predictor.	Higher, especially if too many predictors are used without adequate data.
Multicollinearity Concern	Not applicable, as there's only one predictor.	A primary concern; having correlated predictors can affect the model's accuracy and interpretation.
Applications	Basic research, simple predictions, understanding a singular relationship.	Complex research, multifactorial predictions, studying interrelated systems.

Linear Regression:

Applications of multiple linear regression:

Real Estate Pricing
Size (X_1), number of bedrooms (X_2), and distance to the city center (X_3).
Example: A realtor uses these factors to appraise properties.
Employee Performance and Salary Prediction
Years of experience (X_1), education level (X_2), and performance rating (X_3).
Example: HR teams determine pay scales based on these criteria.
Customer Behavior Analysis
Age (X_1), income level (X_2), and credit score (X_3).
Example: E-commerce platforms assess spending habits.
Energy Consumption Forecasting
Temperature (X_1), household size (X_2), and time of day (X_3).
Example: Power companies predict peak load periods to optimize energy distribution.

Logistic Regression:

- Logistic regression is another supervised learning algorithm which is used to solve the **classification** problems.
- In classification problems, we have dependent variables in a binary or discrete format such as 0 or 1.
- Logistic regression algorithm works with the categorical variable such as 0 or 1, Yes or No, True or False, Spam or not spam, etc.
- It is a predictive analysis algorithm which works on the concept of probability.
- Logistic regression is used for **binary classification** where we use **sigmoid** function, that takes input as independent variables and produces a probability value between 0 and 1.

Logistic Regression:

- For example, we have two classes Class 0 and Class 1 if the value of the logistic function for an input is greater than 0.5 (threshold value) then it belongs to Class 1 otherwise it belongs to Class 0.

- It's referred to as regression because it is the extension of linear regression but is mainly used for classification problems.

- The function can be represented as:

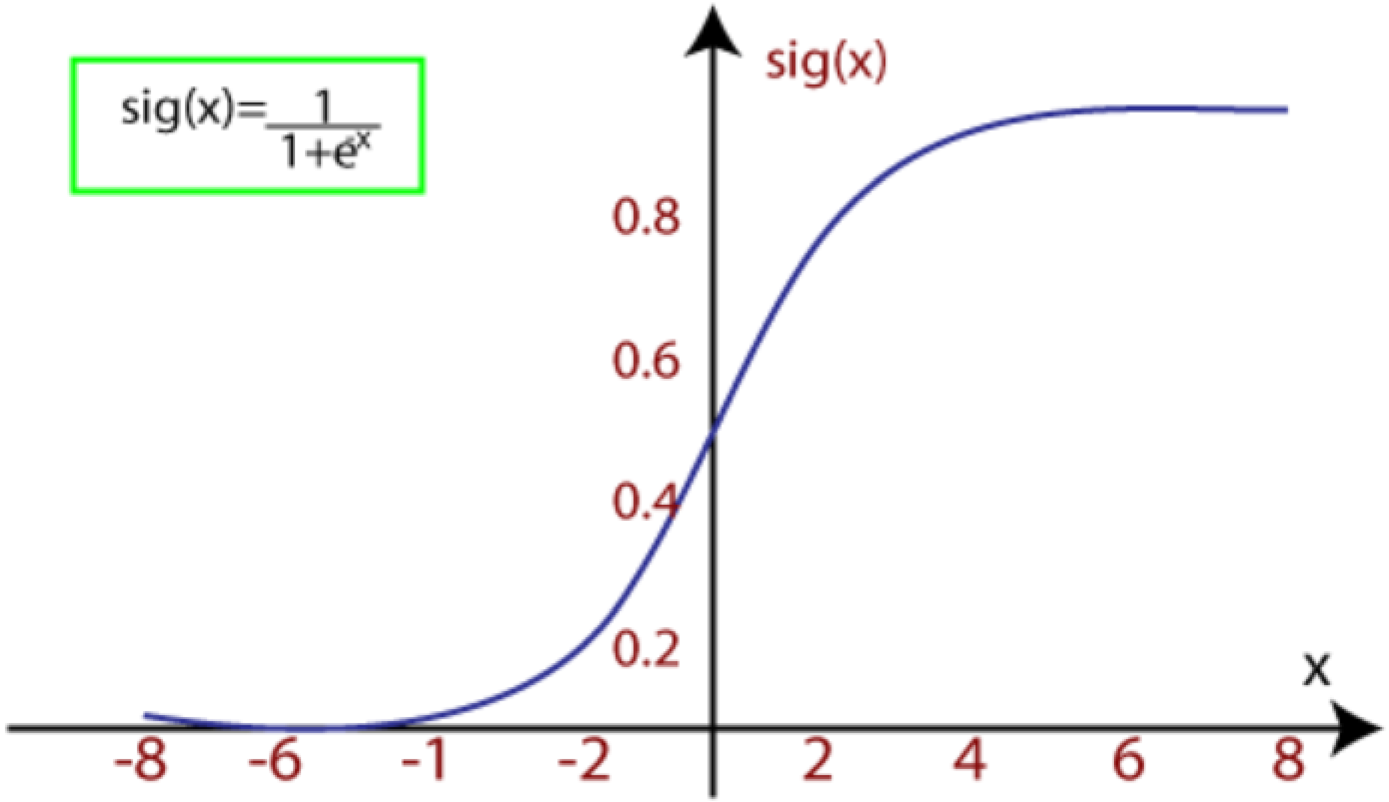
$$f(x)= \frac{1}{1+e^{-x}}$$

f(x)= Output between the 0 and 1 value.
x= input to the function
e= base of natural logarithm.

By Dr. Purvi Tandel

Logistic Regression:

- When we provide the input values (data) to the function, it gives the S-curve as follows:



Logistic Regression:

The logistic function is given by:

$$P(Y = 1|X) = \hat{y} = \frac{1}{1 + e^{-(b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n)}}$$

Where:

- $P(Y = 1|X)$: Probability that the outcome Y belongs to class 1.
- $b_0, b_1, \ldots, b_n$: Coefficients (weights) of the model.
- $X_1, X_2, \ldots, X_n$: Predictor variables.

There are three types of logistic regression:

- Binary(0/1, pass/fail)
- Multi(cats, dogs, lions)
- Ordinal(low, medium, high)

Logistic Regression:

Applications of logistic regression:

Medical Diagnosis

Predicting whether a patient has a disease (Y) based on features like age, blood pressure, and cholesterol levels.

Example: Classifying patients as diabetic or non-diabetic based on test results.

Email Spam Detection

Classifying emails as "spam" or "not spam" (Y) based on keywords, sender information, and other email characteristics (X_1 ,X_2 ,...).

Example: Email services filter messages into the spam folder using logistic regression.

Fraud Detection

Predicting whether a transaction is fraudulent (Y) based on features such as transaction amount, location, and time.

Example: Banks use logistic regression to identify suspicious credit card transactions.

Polynomial Regression:

- Polynomial Regression is a type of regression which models the non-linear dataset using a linear model.

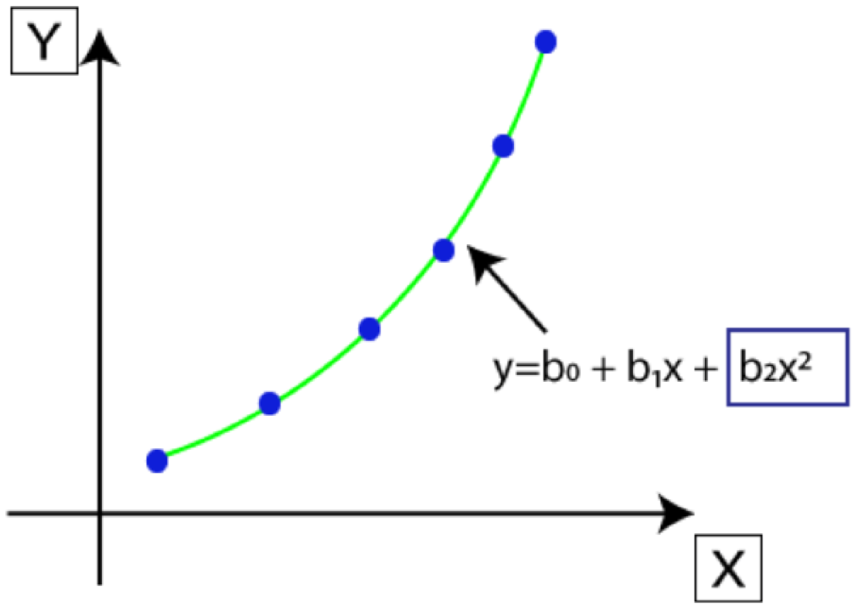
- It is similar to multiple linear regression, but it fits a non-linear curve between the value of x and corresponding conditional values of y.

- Suppose there is a dataset which consists of datapoints which are present in a non-linear fashion, so for such case, linear regression will not best fit to those datapoints. To cover such datapoints, we need Polynomial regression.

- In Polynomial regression, the original features are transformed into polynomial features of given degree and then modeled using a linear model. Which means the datapoints are best fitted using a polynomial line.

Polynomial Regression:

- The equation for polynomial regression also derived from linear regression equation that means Linear regression equation Y= b0+ b1x, is transformed into Polynomial regression equation $Y= b_0+b_1x+ b_2x^2+ b_3x^3+.....+ b_nx^n$.
- Here Y is the predicted/target output, b_0, b_1,... b_n are the regression coefficients. x is our independent/input variable.
- The model is still linear as the coefficients are still linear with quadratic



By Dr. Purvi Tandel

Polynomial Regression:

Note: This is different from Multiple Linear regression in such a way that in Polynomial regression, a single element has different degrees instead of multiple variables with the same degree.

K Nearest Neighbours:

K-Nearest Neighbors (KNN) is a supervised learning algorithm that can be used for **both classification and regression** problems.

The main idea behind KNN is **to find the k-nearest data points to a given test data point and use these nearest neighbors to make a prediction**.

Choose the Number of Neighbors (k):
k is the number of nearest neighbors to consider.
For classification: The majority class among the k neighbors determines the class.
For regression: The mean or weighted mean of the k neighbors gives the predicted value.

K Nearest Neighbours:

Calculate Distance:

•Compute the distance between the query point and all points in the dataset.

•Common distance metrics:

Euclidean distance: $\sqrt{\sum(xi\ - yi)^2}$

Find the k-Nearest Neighbors:

•Select the k data points with the smallest distances to the query point.

Predict the Output:

•For classification: Assign the majority class among the k neighbors.

•For regression: Compute the average or weighted average of the target values of the k neighbors.

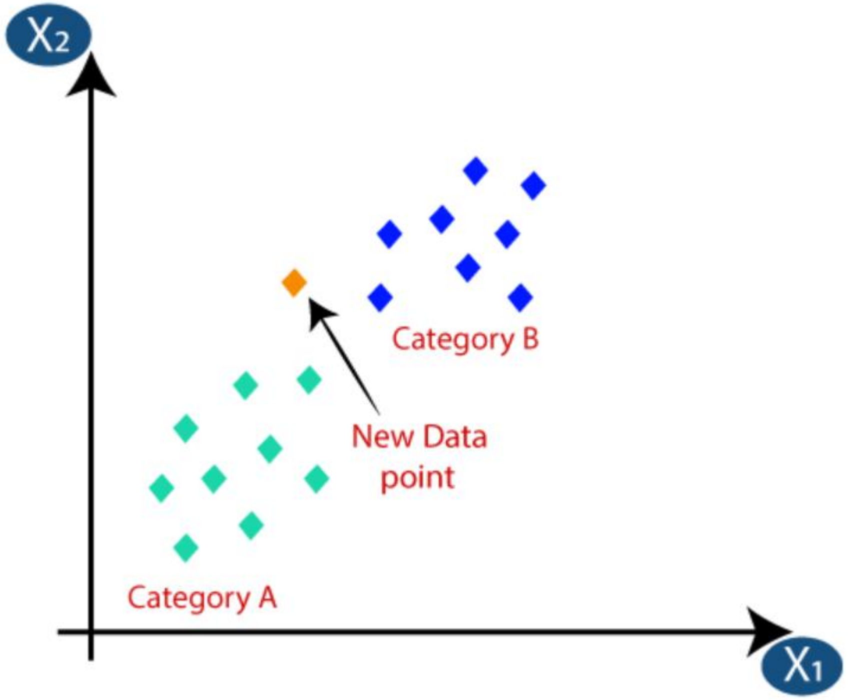
K Nearest Neighbours:

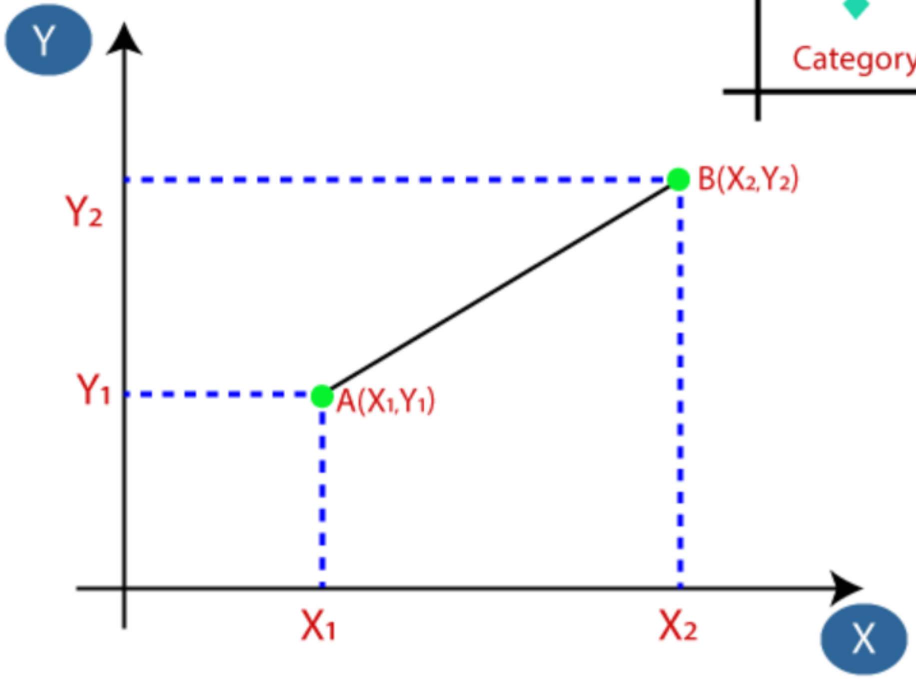
Working of KNN:

The K-NN working can be explained on the basis of the below algorithm:

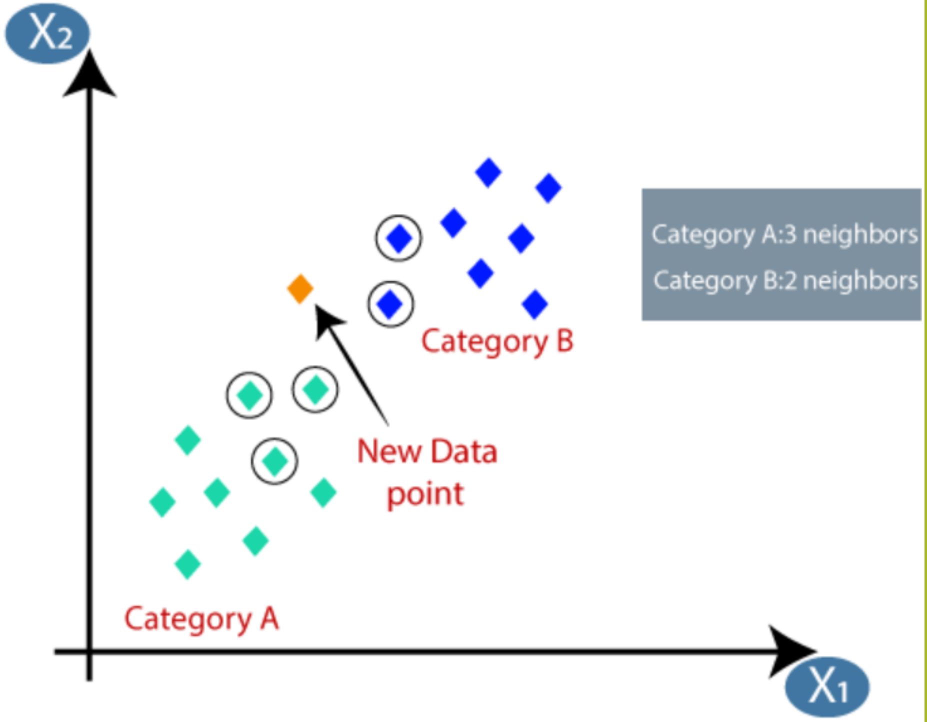
Step-1: Select the number K of the neighbors
Step-2: Calculate the Euclidean distance of K number of neighbors
Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
Step-4: Among these k neighbors, count the number of the data points in each category.
Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
Step-6: Our model is ready.

K Nearest Neighbours:



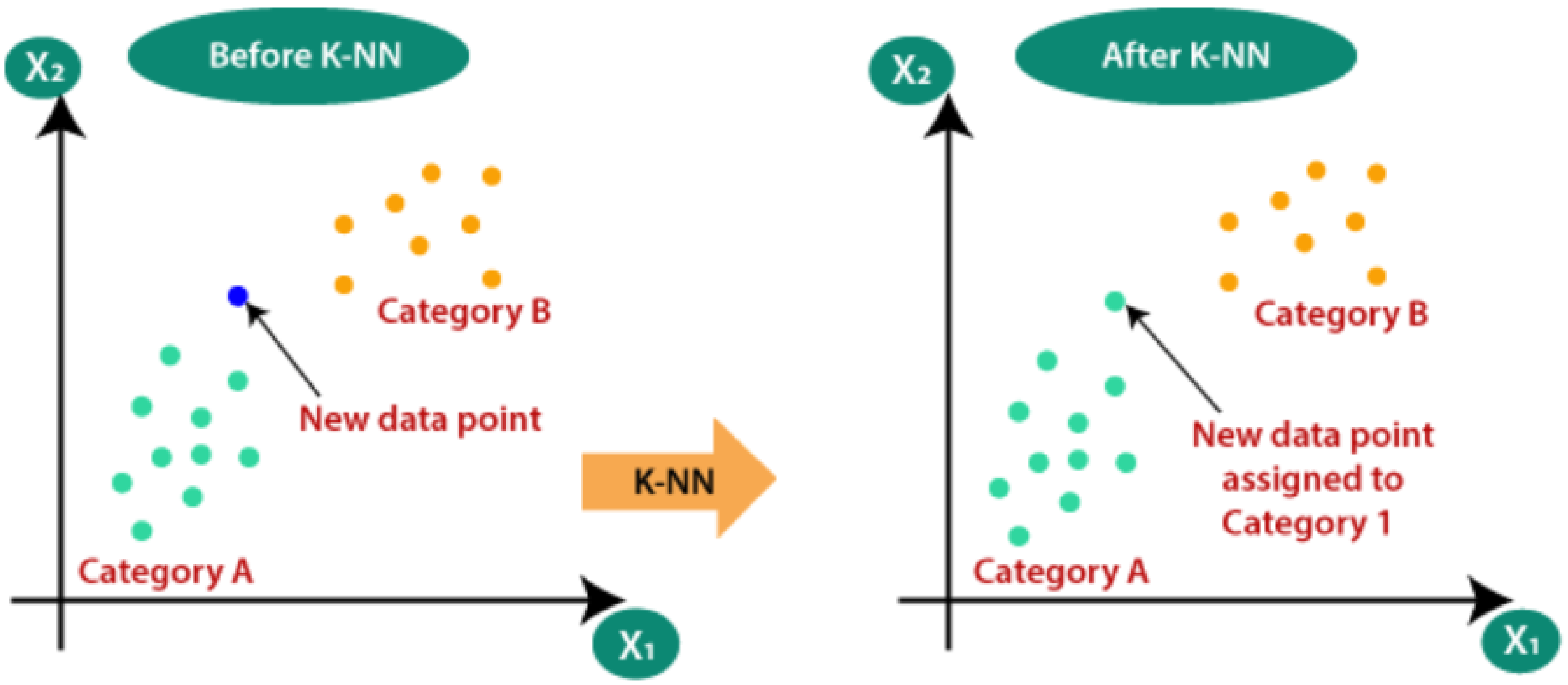


Euclidean Distance between A_1 and $B_2 = \sqrt{(X_2-X_1)^2+(Y_2-Y_1)^2}$



By Dr. Purvi Tandel

K Nearest Neighbours:



By Dr. Purvi Tandel

K Nearest Neighbors:

How to select the value of K in the K-NN Algorithm?

Below are some points to remember while selecting the value of K in the K-NN algorithm:

- There is no particular way to determine the best value for "K", so we need to try some values to find the best out of them.

- The most preferred value for K is 5.
- A very low value for K such as K=1 or K=2, can be noisy and lead to the effects of outliers in the model.
- Large values for K are good, but it may find some difficulties.

K Nearest Neighbors:

Advantages of KNN Algorithm:

It is simple to implement.
It is robust to the noisy training data
It can be more effective if the training data is large.

Disadvantages of KNN Algorithm:

Always needs to determine the value of K which may be complex some time.
The computation cost is high because of calculating the distance between the data points for all the training samples.

K Nearest Neighbors:

Problem 1:
Classify a new data point into one of two categories: "Pass" or "Fail" based on study hours and attendance. Find the result for study hours 5 and attendance 82.

Dataset:

Study Hours	Attendance (%)	Outcome
2	70	Fail
4	80	Fail
6	85	Pass
8	90	Pass
10	95	Pass

Study Hours	Attendance (%)	Outcome
2	70	Fail
4	80	Fail
6	85	Pass
8	90	Pass
10	95	Pass

1. **Compute Distances:** Using Euclidean distance:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Distances to the query point:

- To (2, 70): $\sqrt{(5-2)^2 + (82-70)^2} = \sqrt{3^2 + 12^2} = \sqrt{9+144} = 12.37$
- To (4, 80): $\sqrt{(5-4)^2 + (82-80)^2} = \sqrt{1^2 + 2^2} = \sqrt{1+4} = 2.24$
- To (6, 85): $\sqrt{(5-6)^2 + (82-85)^2} = \sqrt{(-1)^2 + (-3)^2} = \sqrt{1+9} = 3.16$
- To (8, 90): $\sqrt{(5-8)^2 + (82-90)^2} = \sqrt{(-3)^2 + (-8)^2} = \sqrt{9+64} = 8.54$
- To (10, 95): $\sqrt{(5-10)^2 + (82-95)^2} = \sqrt{(-5)^2 + (-13)^2} = \sqrt{25+169} = 13.93$

Study Hours	Attendance (%)	Outcome
2	70	Fail
4	80	Fail
6	85	Pass
8	90	Pass
10	95	Pass

2. **Find the 3 Nearest Neighbors ($k = 3$):**
   - Neighbors: (4, 80), (6, 85), (8, 90).
3. **Predict Outcome:**
   - Outcomes of neighbors: "Fail", "Pass", "Pass".
   - Majority class: "Pass".

Prediction: The query point belongs to class "Pass".

K Nearest Neighbors:

Problem 2:
Predict the price of a house based on its size and distance to the city center. Predict the house prize for size 1900 and Distance 4.

Dataset:

Size (sq ft)	Distance (miles)	Price ($)
1500	2	300,000
1800	5	320,000
2000	3	340,000
2100	4	360,000
2500	6	400,000

By Dr. Purvi Tandel

Size (sq ft)	Distance (miles)	Price ($)
1500	2	300,000
1800	5	320,000
2000	3	340,000
2100	4	360,000
2500	6	400,000

1. **House (1500, 2, $300,000$):**

$$d = \sqrt{(1900 - 1500)^2 + (4 - 2)^2} = \sqrt{400^2 + 2^2} = \sqrt{160000 + 4} = 400.005$$

2. **House (1800, 5, $320,000$):**

$$d = \sqrt{(1900 - 1800)^2 + (4 - 5)^2} = \sqrt{100^2 + (-1)^2} = \sqrt{10000 + 1} = 100.005$$

3. **House (2000, 3, $340,000$):**

$$d = \sqrt{(1900 - 2000)^2 + (4 - 3)^2} = \sqrt{(-100)^2 + 1^2} = \sqrt{10000 + 1} = 100.005$$

4. **House (2100, 4, $360,000$):**

$$d = \sqrt{(1900 - 2100)^2 + (4 - 4)^2} = \sqrt{(-200)^2 + 0^2} = \sqrt{40000} = 200.0$$

5. **House (2500, 6, $400,000$):**

$$d = \sqrt{(1900 - 2500)^2 + (4 - 6)^2} = \sqrt{(-600)^2 + (-2)^2} = \sqrt{360000 + 4} = 600.003$$

Size (sq ft)	Distance (miles)	Price ($)
1500	2	300,000
1800	5	320,000
2000	3	340,000
2100	4	360,000
2500	6	400,000

Step 2: Sort by Distance

House (Size, Distance, Price)	Distance
(1800, 5, $320,000$)	100.005
(2000, 3, $340,000$)	100.005
(2100, 4, $360,000$)	200.0
(1500, 2, $300,000$)	400.005
(2500, 6, $400,000$)	600.003

Size (sq ft)	Distance (miles)	Price ($)
1500	2	300,000
1800	5	320,000
2000	3	340,000
2100	4	360,000
2500	6	400,000

Step 2: Sort by Distance

House (Size, Distance, Price)	Distance
(1800, 5, $320,000$)	100.005
(2000, 3, $340,000$)	100.005
(2100, 4, $360,000$)	200.0
(1500, 2, $300,000$)	400.005
(2500, 6, $400,000$)	600.003

House (Size, Distance, Price)	Distance
(1800, 5, $320,000$)	100.005
(2000, 3, $340,000$)	100.005
(2100, 4, $360,000$)	200.0
(1500, 2, $300,000$)	400.005
(2500, 6, $400,000$)	600.003

Step 3: Select $k = 3$ Nearest Neighbors

The nearest neighbors are:

1. (1800, 5, $320,000$)
2. (2000, 3, $340,000$)
3. (2100, 4, $360,000$)

Step 4: Predict Price

The predicted price is the mean of the prices of the $k = 3$ nearest neighbors:

$$\text{Predicted Price} = \frac{320000 + 340000 + 360000}{3} = \frac{1020000}{3} = 340000$$

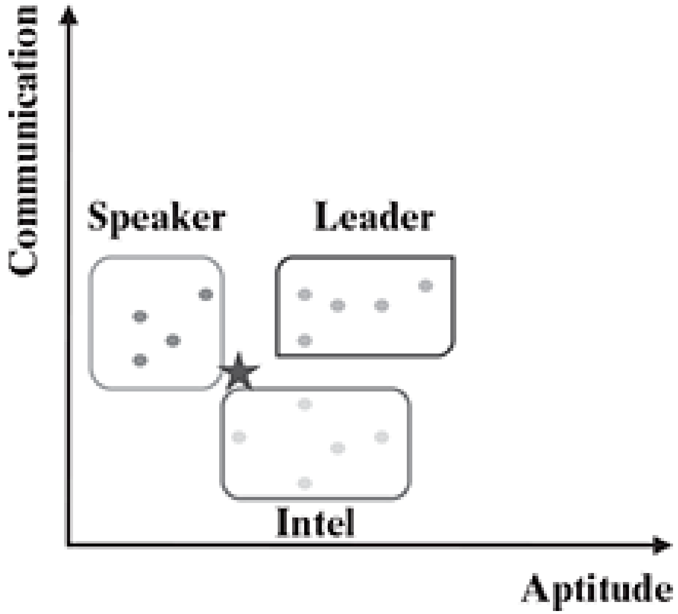
Problem 3: Predict or test the class of Josh.

Name	Aptitude	Communication	Class
Karuna	2	5	Speaker
Bhuvna	2	6	Speaker
Gaurav	7	6	Leader
Parul	7	2.5	Intel
Dinesh	8	6	Leader
Jani	4	7	Speaker
Bobby	5	3	Intel
Parimal	3	5.5	Speaker
Govind	8	3	Intel
Susant	6	5.5	Leader
Gouri	6	4	Intel
Bharat	6	7	Leader
Ravi	6	2	Intel
Pradeep	9	7	Leader
Josh	5	4.5	Intel

Problem 3:

	Name	Aptitude	Communication	Class
Training Data	Karuna	2	5	Speaker
	Bhuvna	2	6	Speaker
	Gaurav	7	6	Leader
	Parul	7	2.5	Intel
	Dinesh	8	6	Leader
	Jani	4	7	Speaker
	Bobby	5	3	Intel
	Parimal	3	5.5	Speaker
	Govind	8	3	Intel
	Susant	6	5.5	Leader
	Gouri	6	4	Intel
	Bharat	6	7	Leader
	Ravi	6	2	Intel
	Pradeep	9	7	Leader
Test Data	Josh	5	4.5	Intel

Problem 3:



Name	Aptitude	Communication	Class
Karuna	2	5	Speaker
Bhuvna	2	6	Speaker
Gaurav	7	6	Leader
Parul	7	2.5	Intel
Dinesh	8	6	Leader
Jani	4	7	Speaker
Bobby	5	3	Intel
Parimal	3	5.5	Speaker
Govind	8	3	Intel
Susant	6	5.5	Leader
Gouri	6	4	Intel
Bharat	6	7	Leader
Ravi	6	2	Intel
Pradeep	9	7	Leader
Josh	5	4.5	???

Name	Aptitude (X_1)	Communication (X_2)	Class	Distance to Josh
Karuna	2	5	Speaker	$\sqrt{(5-2)^2+(4.5-5)^2}=3.04$
Bhuvna	2	6	Speaker	$\sqrt{(5-2)^2+(4.5-6)^2}=3.54$
Gaurav	7	6	Leader	$\sqrt{(5-7)^2+(4.5-6)^2}=2.5$
Parul	7	2.5	Intel	$\sqrt{(5-7)^2+(4.5-2.5)^2}=3.61$
Dinesh	8	6	Leader	$\sqrt{(5-8)^2+(4.5-6)^2}=3.35$
Jani	4	7	Speaker	$\sqrt{(5-4)^2+(4.5-7)^2}=2.69$
Bobby	5	3	Intel	$\sqrt{(5-5)^2+(4.5-3)^2}=1.5$
Parimal	3	5.5	Speaker	$\sqrt{(5-3)^2+(4.5-5.5)^2}=2.24$
Govind	8	3	Intel	$\sqrt{(5-8)^2+(4.5-3)^2}=3.35$
Susant	6	5.5	Leader	$\sqrt{(5-6)^2+(4.5-5.5)^2}=1.41$
Gouri	6	4	Intel	$\sqrt{(5-6)^2+(4.5-4)^2}=1.12$
Bharat	6	7	Leader	$\sqrt{(5-6)^2+(4.5-7)^2}=2.69$
Ravi	6	2	Intel	$\sqrt{(5-6)^2+(4.5-2)^2}=2.69$
Pradeep	9	7	Leader	$\sqrt{(5-9)^2+(4.5-7)^2}=5.31$

By Dr. Purvi Tandel

Step 3: Sort Distances

Name	Distance	Class
Gouri	1.12	Intel
Susant	1.41	Leader
Bobby	1.5	Intel
Parimal	2.24	Speaker
Gaurav	2.5	Leader
Ravi	2.69	Intel
Bharat	2.69	Leader
Jani	2.69	Speaker
Karuna	3.04	Speaker
Bhuvna	3.54	Speaker
Dinesh	3.35	Leader
Govind	3.35	Intel
Parul	3.61	Intel
Pradeep	5.31	Leader

Step 4: Choose $k = 3$

The 3 nearest neighbors are:

1. Gouri (Intel)
2. Susant (Leader)
3. Bobby (Intel)

Step 5: Predict the Class

- Intel: 2 votes
- Leader: 1 vote

The predicted class for **Josh** is **Intel**.

Step 4: Choose k =5

- **Intel**: 2 votes (Gouri, Bobby)
- **Leader**: 2 votes (Susant, Gaurav)
- **Speaker**: 1 vote (Parimal)

Step 5: Predict the Class

Since there is a tie between **Intel and Leader**, we apply a tie-breaking rule.

The tie can be resolved by choosing the class of the closest point, which is **Gouri (Intel).**

Thus, for k=5, the predicted class for Josh is **Intel**.

Decision Tree:

- Decision Tree is a Supervised learning technique that can be used for both **classification and Regression** problems, but mostly it is preferred for solving Classification problems.

- It is a tree-structured classifier, where **internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome**.

- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

Decision Tree:

- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

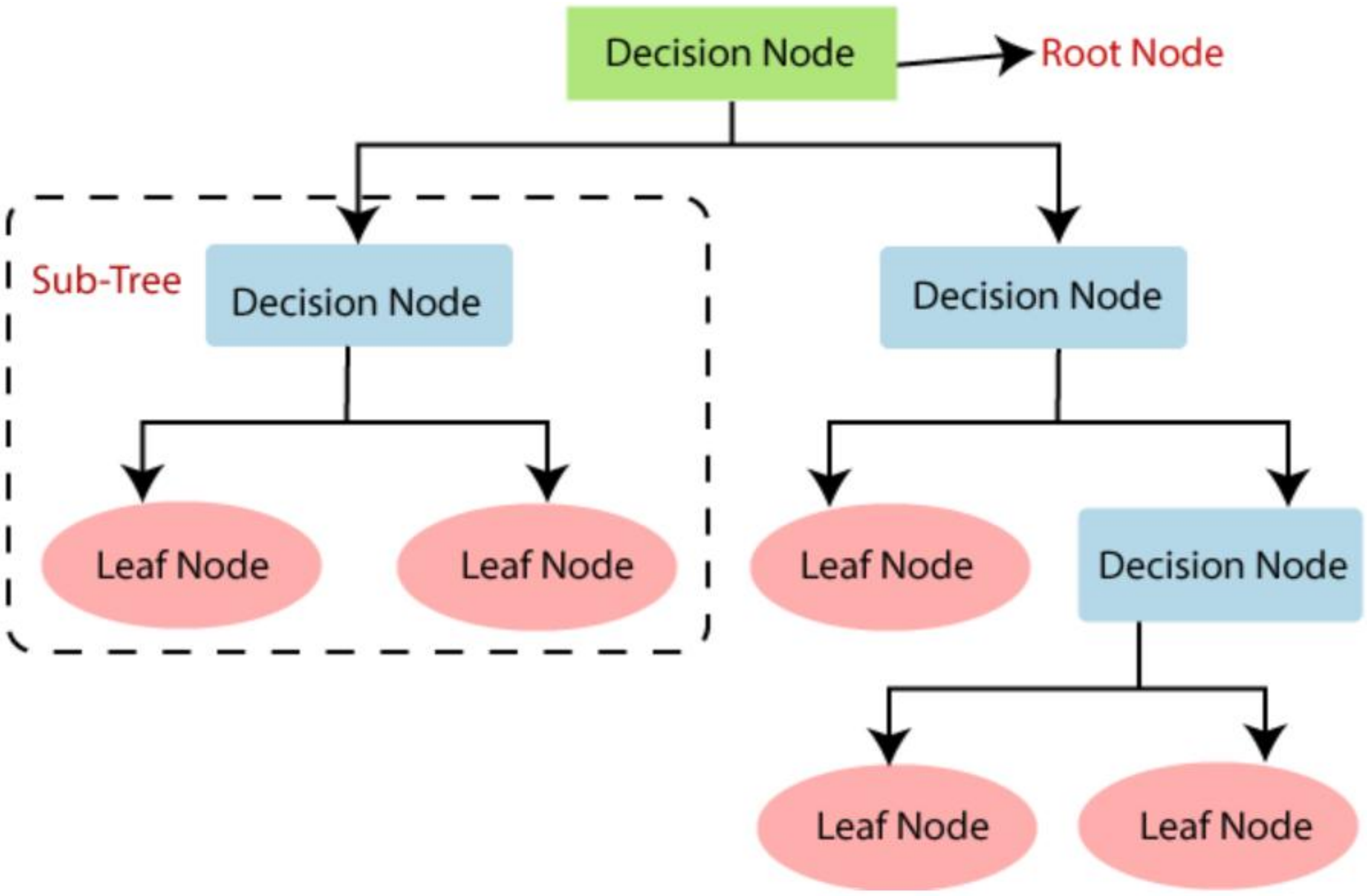
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

- In order to build a tree, we use the CART algorithm, which stands for **Classification and Regression Tree algorithm**.

- A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

- Note: A decision tree can contain categorical data (YES/NO) as well as numeric data.

Decision Tree:



Decision Tree:

Below are the two reasons for using the Decision tree:

- Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
- The logic behind the decision tree can be easily understood because it shows a tree-like structure.

Decision Tree:

Decision Tree Terminologies:
Root Node: Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.

Leaf Node: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.

Splitting: Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.

Branch/Sub Tree: A tree formed by splitting the tree.

Pruning: Pruning is the process of removing the unwanted branches from the tree.

Parent/Child node: The root node of the tree is called the parent node, and other nodes are called the child nodes.

Decision Tree:

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).

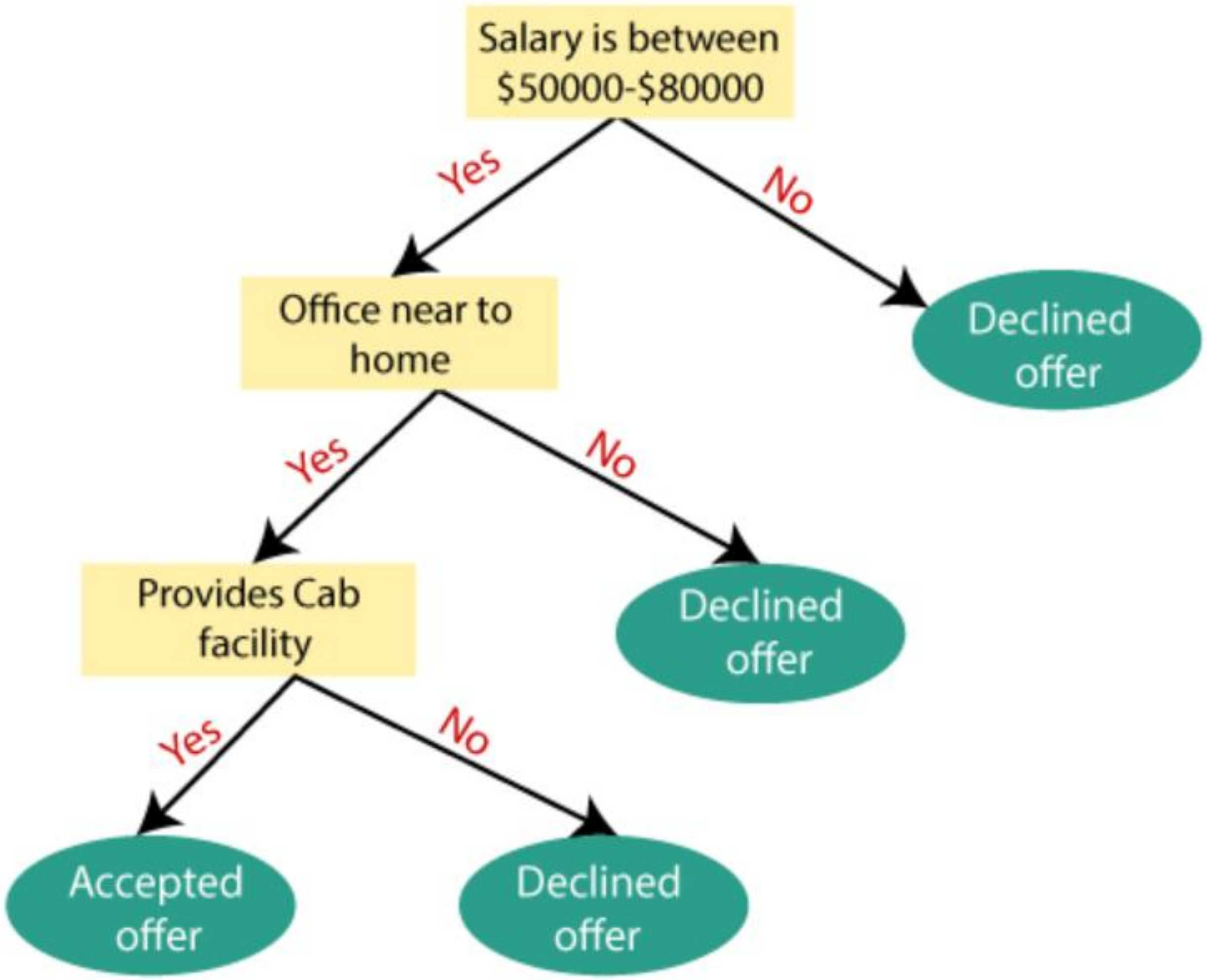
Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Decision Tree:

Example: Suppose there is a candidate who has a job offer and wants to decide whether he should accept the offer or Not.



Decision Tree:

Attribute Selection Measures:

While implementing a Decision tree, the main issue arises that how to select the best attribute for the root node and for sub-nodes.

So, to solve such problems there is a technique which is called as **Attribute selection measure or ASM**. By this measurement, we can easily select the best attribute for the nodes of the tree.

There are two popular techniques for ASM, which are:

- Information Gain
- Gini Index

Decision Tree:

Example: Decision Tree for Loan Approval

Dataset:

Applicant Age	Income	Loan Amount	Credit History	Loan Approved
25	High	Low	Good	Yes
45	Medium	High	Poor	No
35	Low	Medium	Good	Yes
30	Medium	Medium	Poor	No
50	High	Low	Good	Yes

Decision Tree:

Decision Tree Construction:

1. **Root Node:**
   - The tree starts with the feature `Credit History`, as it provides the best split:
     - If `Credit History = Good`, the likelihood of loan approval is high.
     - If `Credit History = Poor`, further checks are needed.
2. **Second Split (for Poor Credit History):**
   - Use `Income` as the next feature:
     - If `Income = High`, the loan is likely to be approved.
     - If `Income = Low/Medium`, the loan is unlikely to be approved.

Decision Tree:

Testing a New Applicant:

Applicant Age	Income	Loan Amount	Credit History
40	Medium	High	Poor

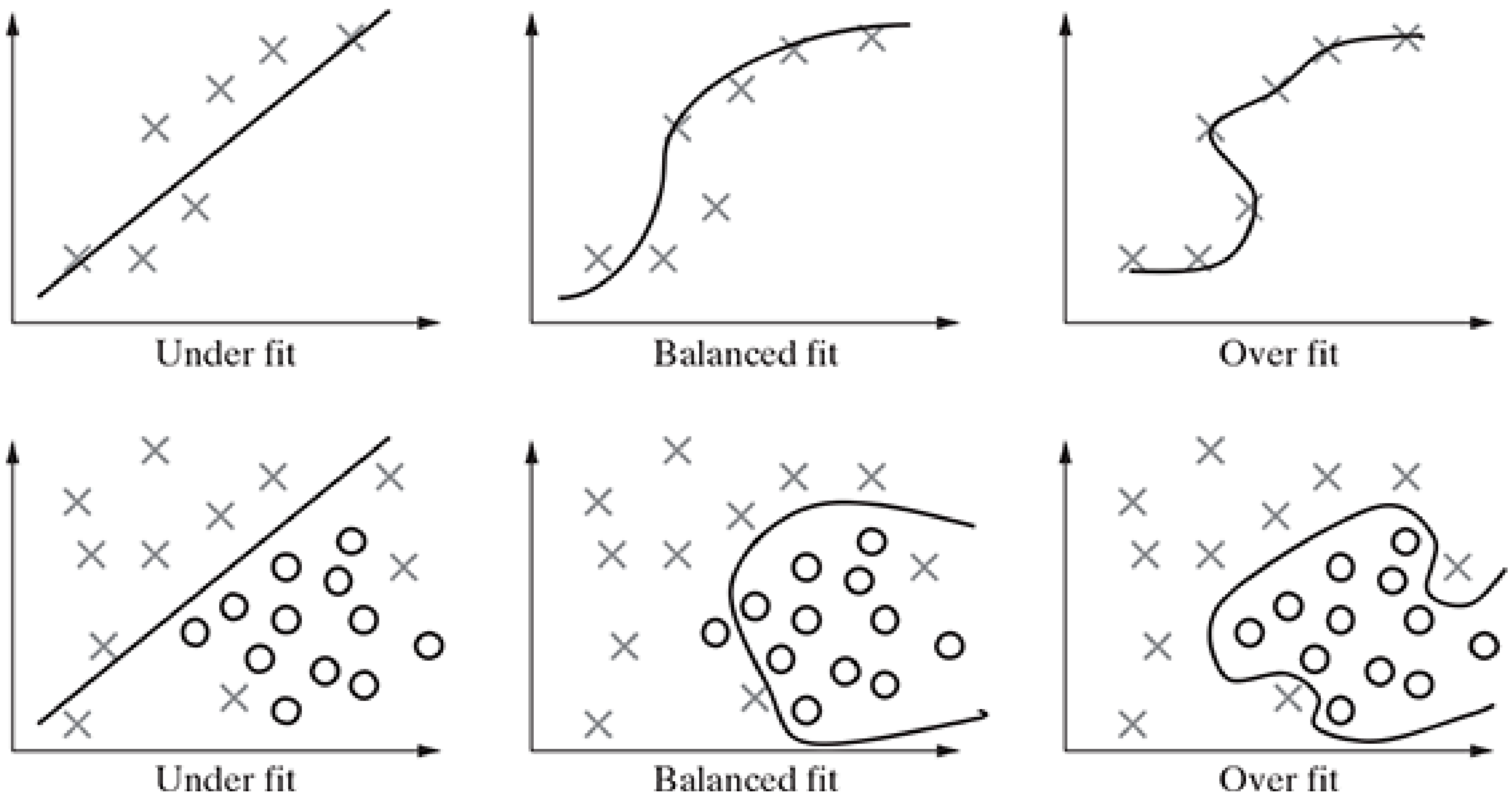
```
          [Credit History]
           /          \
       Good            Poor
       /                 \
    Yes               [Income]
                    /    |    \
                 Low  Medium  High
                  No     No    Yes
```

- **Decision Path:**
    - Credit History = Poor → Check Income.
    - Income = Medium → Loan Approved = No.

Predicted Outcome:

Loan Approved = **No**.

Overfitting and Regularization:



By Dr. Purvi Tandel