Pantech e Learning
DIGITAL LEARNING SIMPLIFIED

Amazon Web Services

MLOps with AWS

Masterclass

aws

Machine Learning Operations with AWS

Day -14

Amazon Web Services

MLOps with AWS

Masterclass

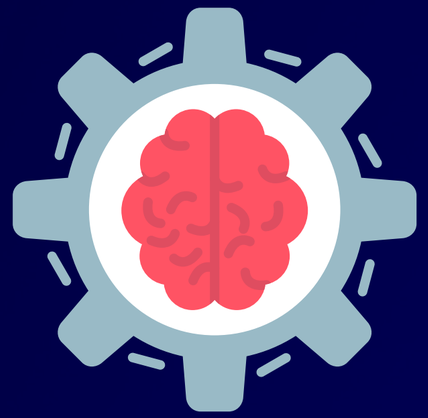
Pantech e Learning

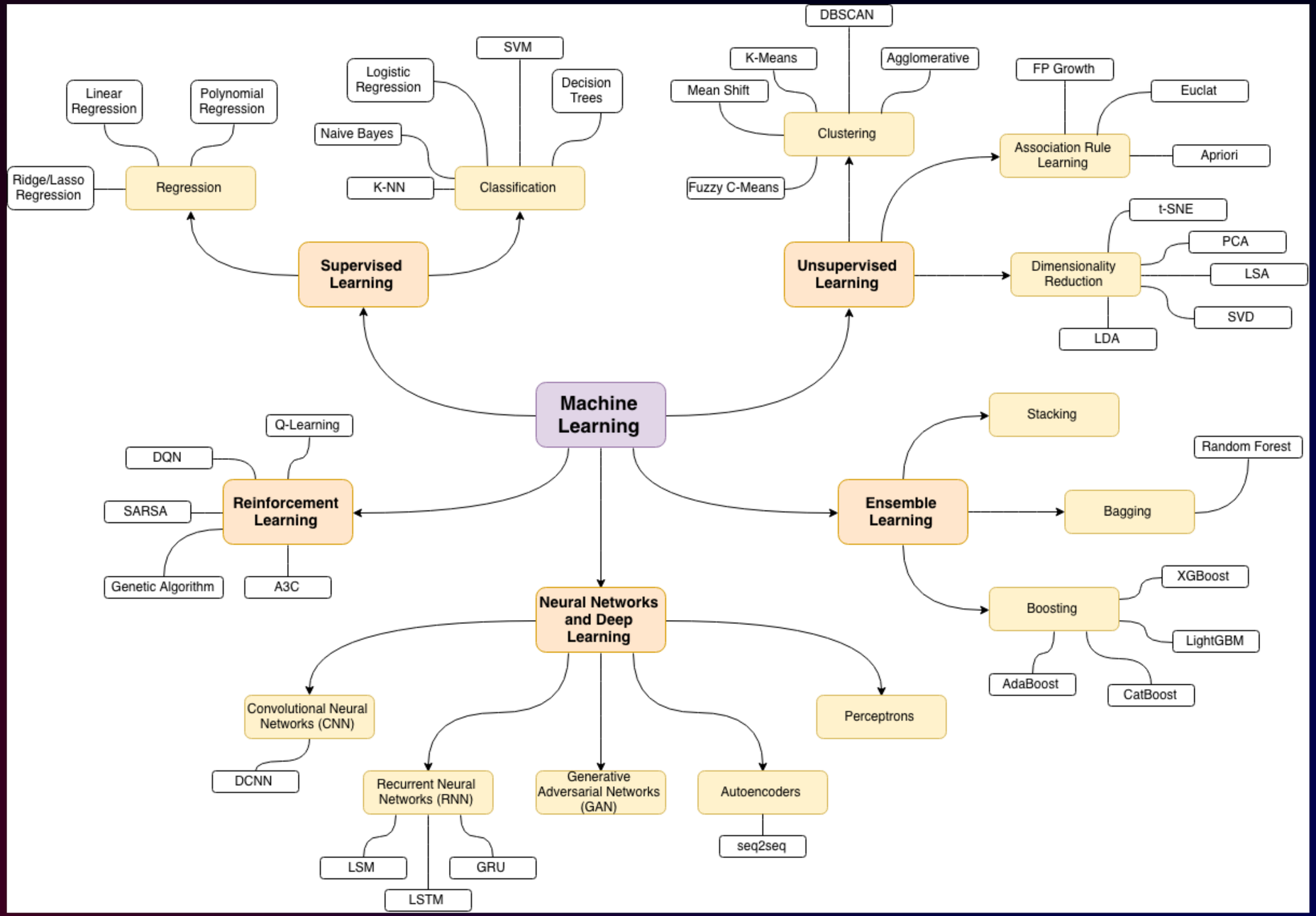
DIGITAL LEARNING SIMPLIFIED

HOW TO CHOOSE

THE RIGHT

MACHINE LEARNING ALGORITHM

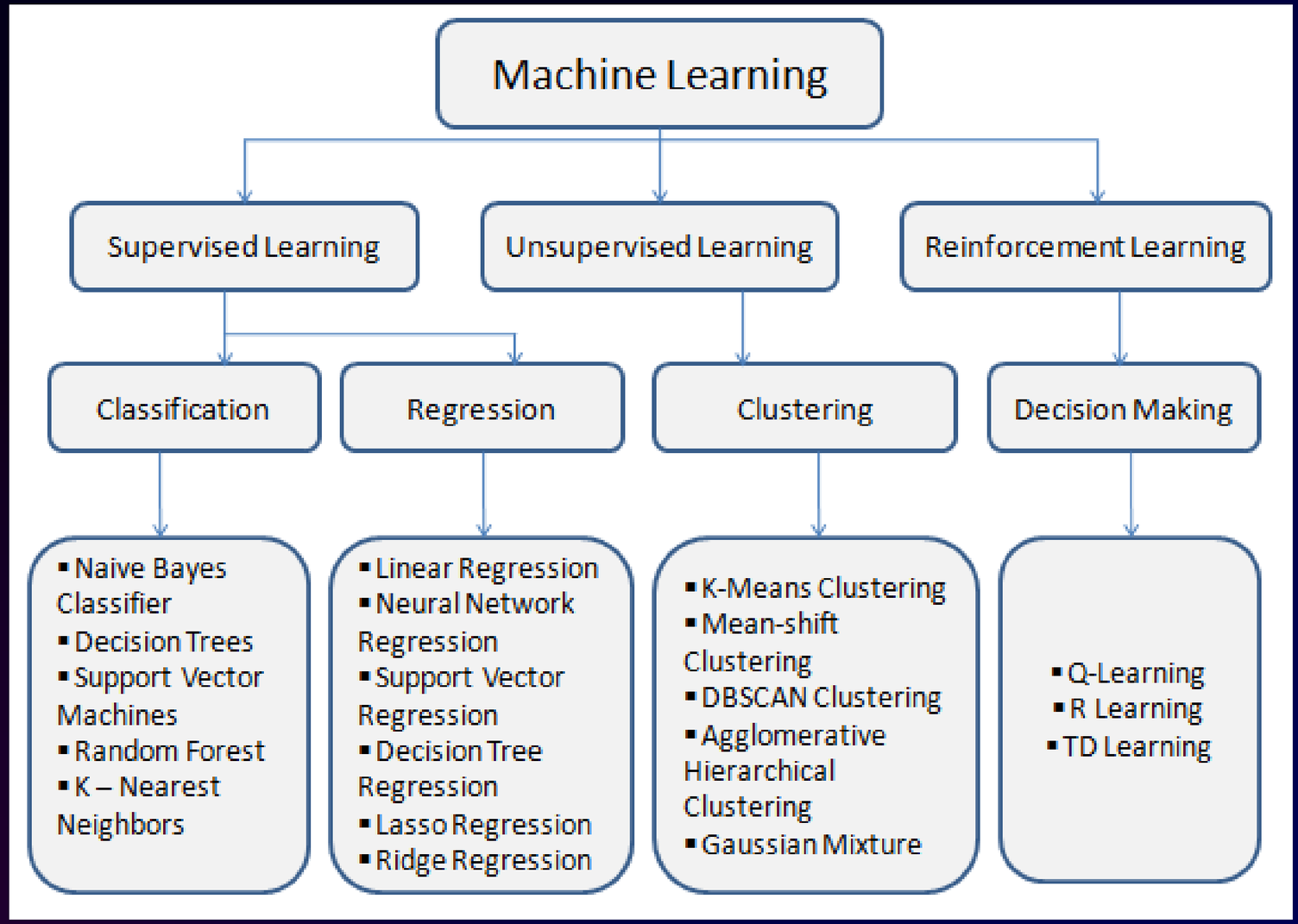



Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Ensemble Learning
Neural Networks and Deep Learning
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
Classification
Logistic Regression
SVM
Decision Trees
Naive Bayes
K-NN
Clustering
DBSCAN
K-Means
Agglomerative
Mean Shift
Fuzzy C-Means
Association Rule Learning
FP Growth
Euclat
Apriori
Dimensionality Reduction
t-SNE
PCA
LSA
SVD
LDA
Q-Learning
DQN
SARSA
Genetic Algorithm
A3C
Stacking
Bagging
Random Forest
Boosting
XGBoost
LightGBM
AdaBoost
CatBoost
Convolutional Neural Networks (CNN)
DCNN
Recurrent Neural Networks (RNN)
LSM
GRU
LSTM
Generative Adversarial Networks (GAN)
Autoencoders
seq2seq
Perceptrons

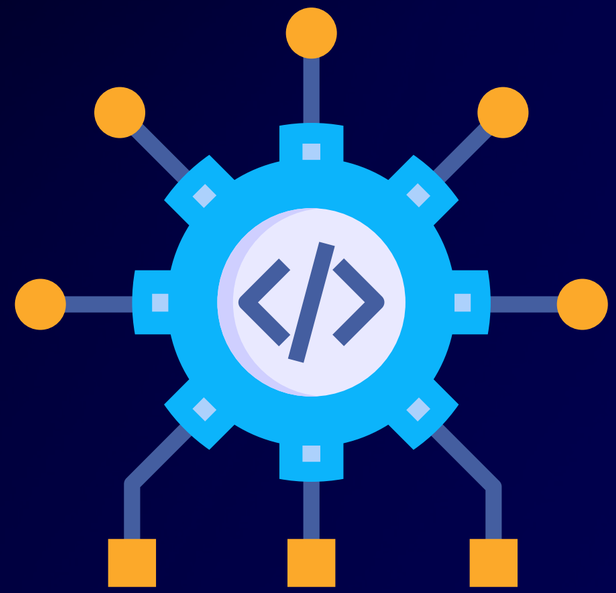
FACTORS TO BE CONSIDERED

- *Nature of the problem*
- *Nature of algorithms*
- *Performance comparison*

NATURE OF PROBLEM



NATURE OF ALGORITHM



LINEAR REGRESSION

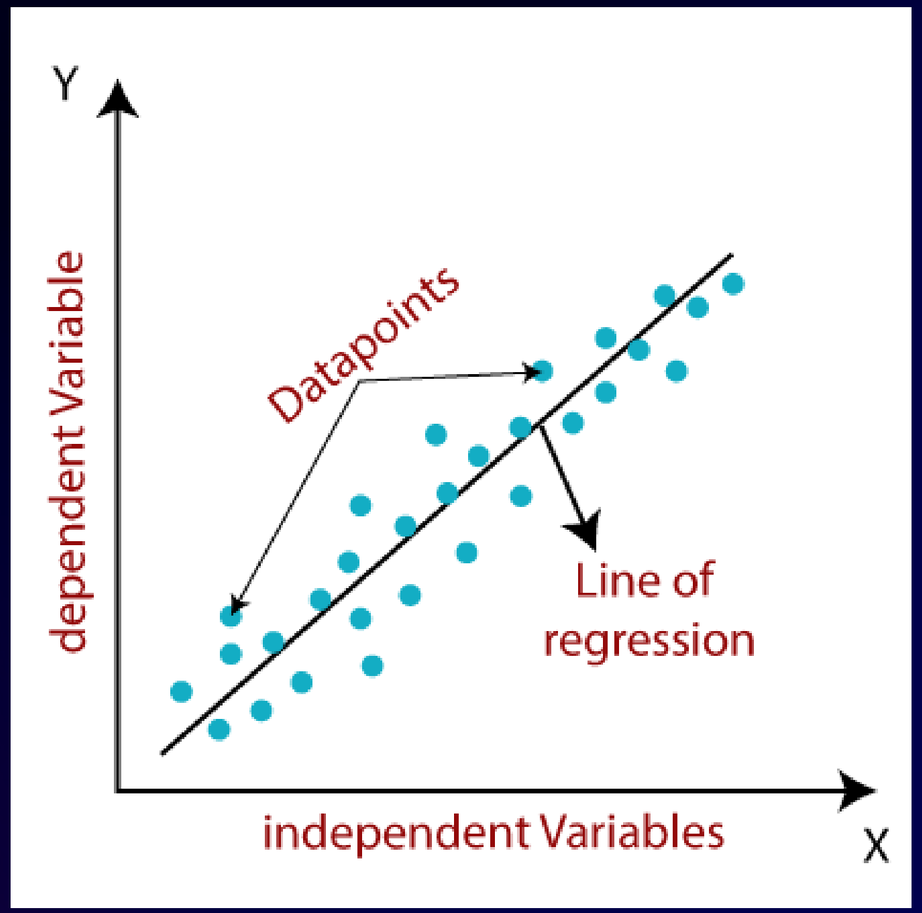
LINEAR REGRESSION

- Linear regression is a popular and widely-used supervised learning algorithm in machine learning.
- It is a statistical method that models the linear relationship between a dependent variable and one or more independent variables .
- Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

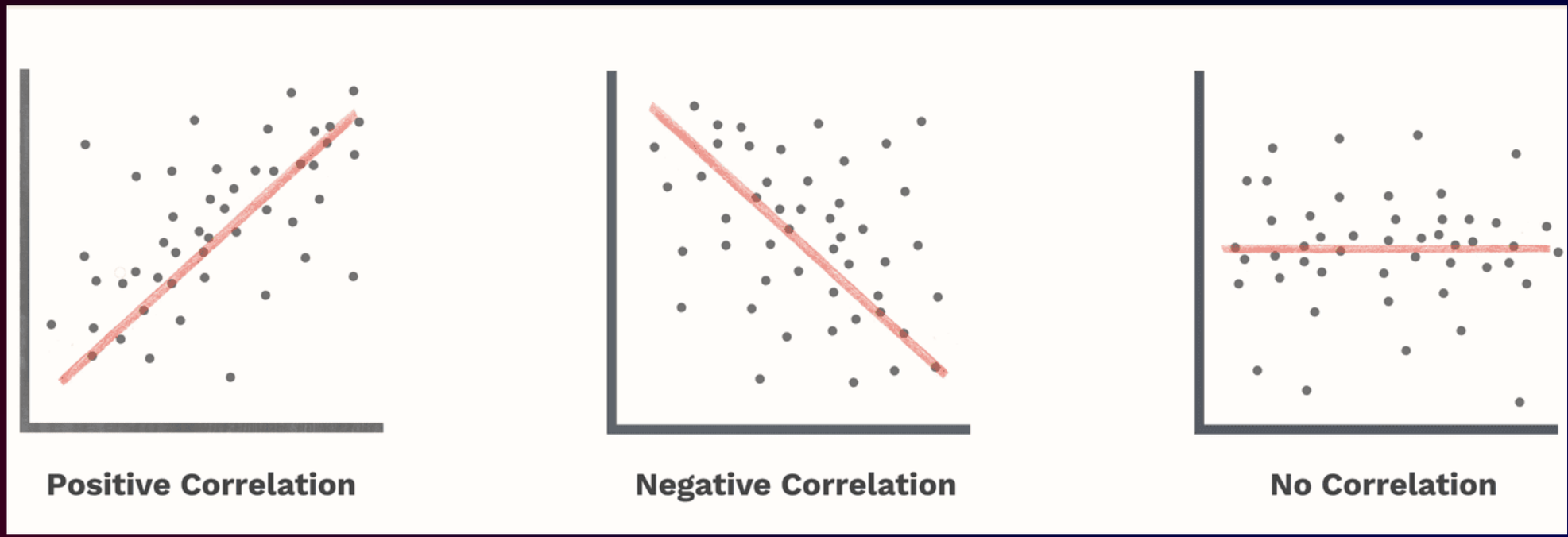
LINEAR REGRESSION

Price	Age	KM	FuelType	HP	MetColor	Automatic	CC	Doors	Weight
13500	23.0	46986	Diesel	90	1	0	2000.0	3	1165.0
13750	23.0	72937	Diesel	90	1	0	2000.0	3	1165.0
13950	24.0	41711	Diesel	90	1	0	2000.0	3	1165.0
14950	26.0	48000	Diesel	90	0	0	2000.0	3	1165.0
13750	30.0	38500	Diesel	90	0	0	2000.0	3	1170.0
12950	32.0	61000	Diesel	90	0	0	2000.0	3	1170.0
16900	27.0	94612	Diesel	90	1	0	2000.0	3	1245.0
18600	30.0	75889	Diesel	90	1	0	2000.0	3	1245.0
21500	27.0	19700	Petrol	192	0	0	1800.0	3	1185.0
12950	23.0	71138	Diesel	69	0	0	1900.0	3	1105.0

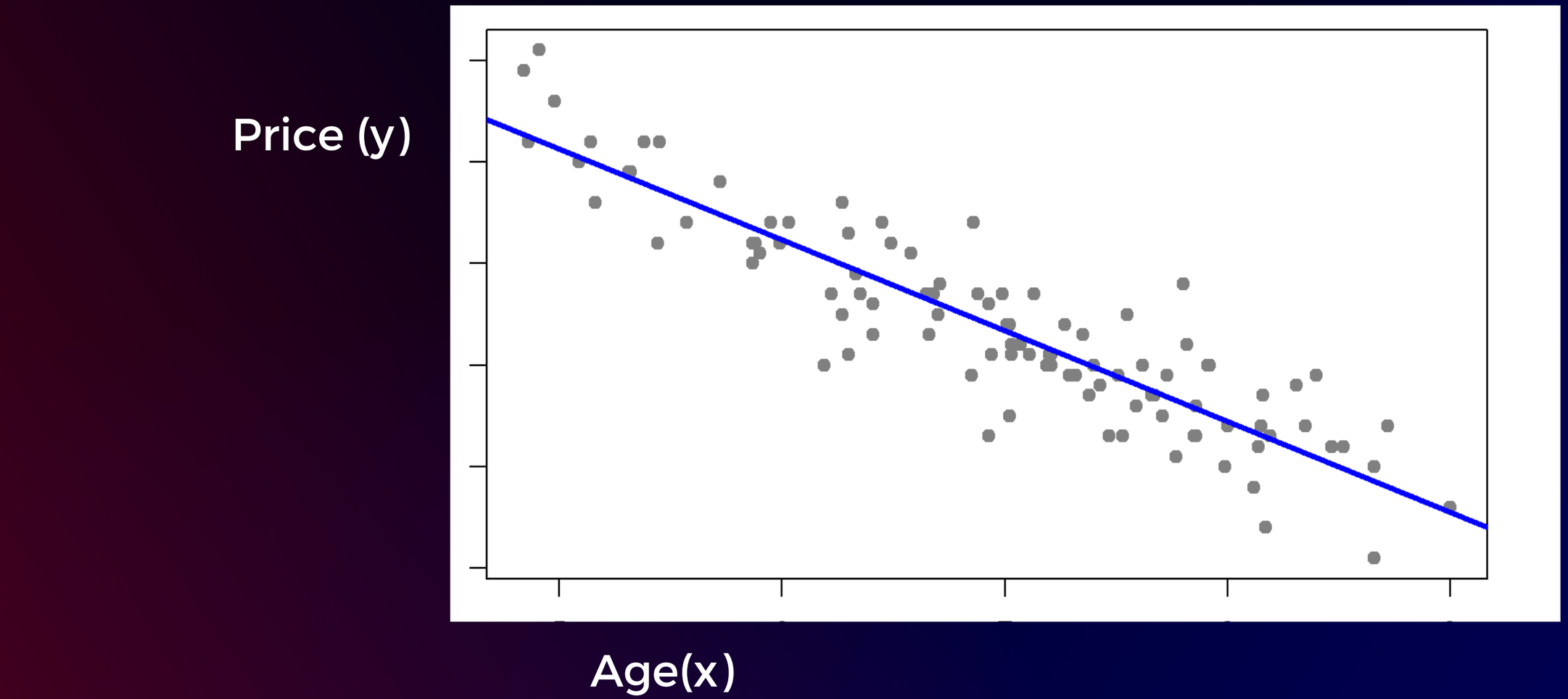
LINEAR REGRESSION



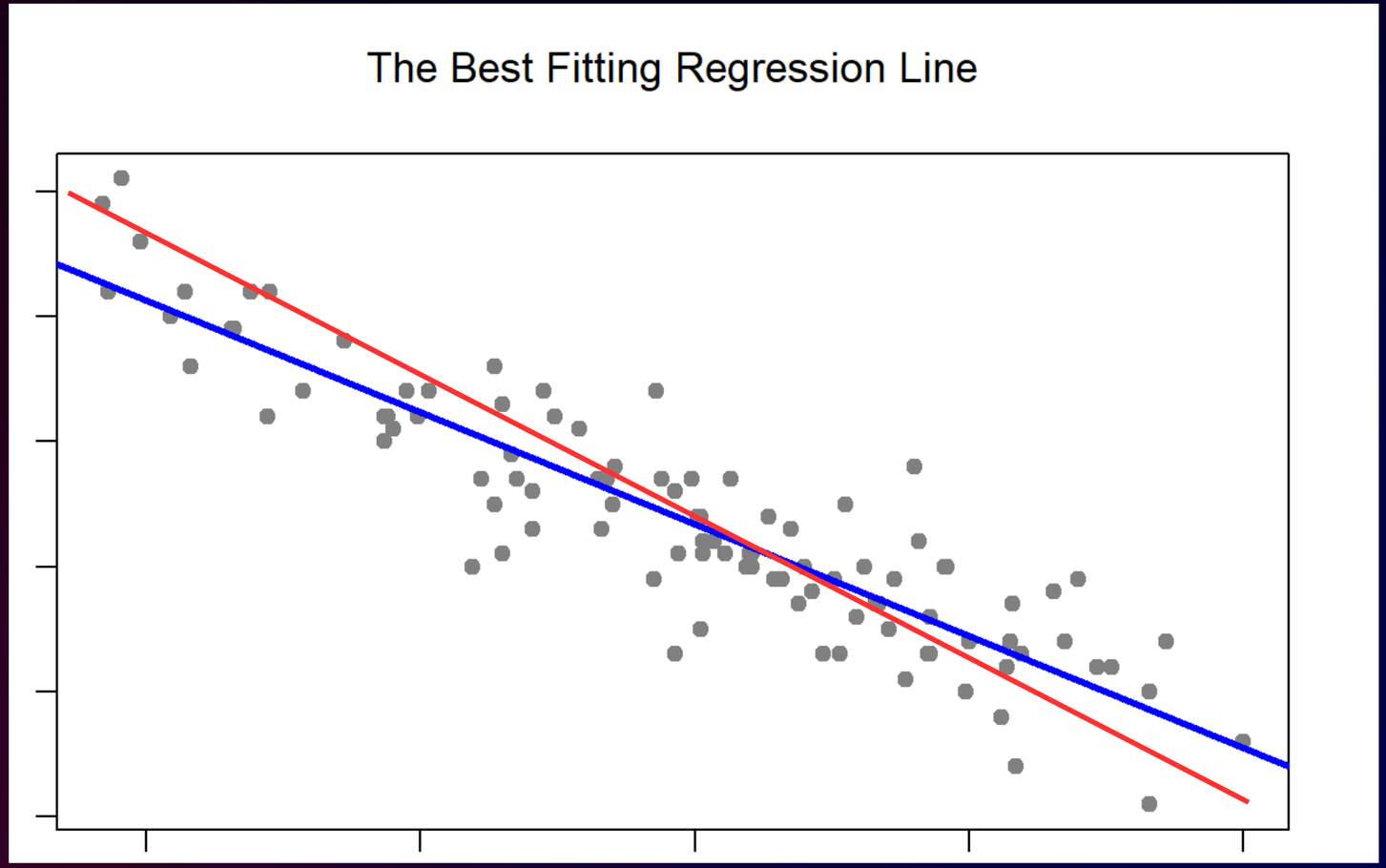
LINEAR REGRESSION



LINEAR REGRESSION

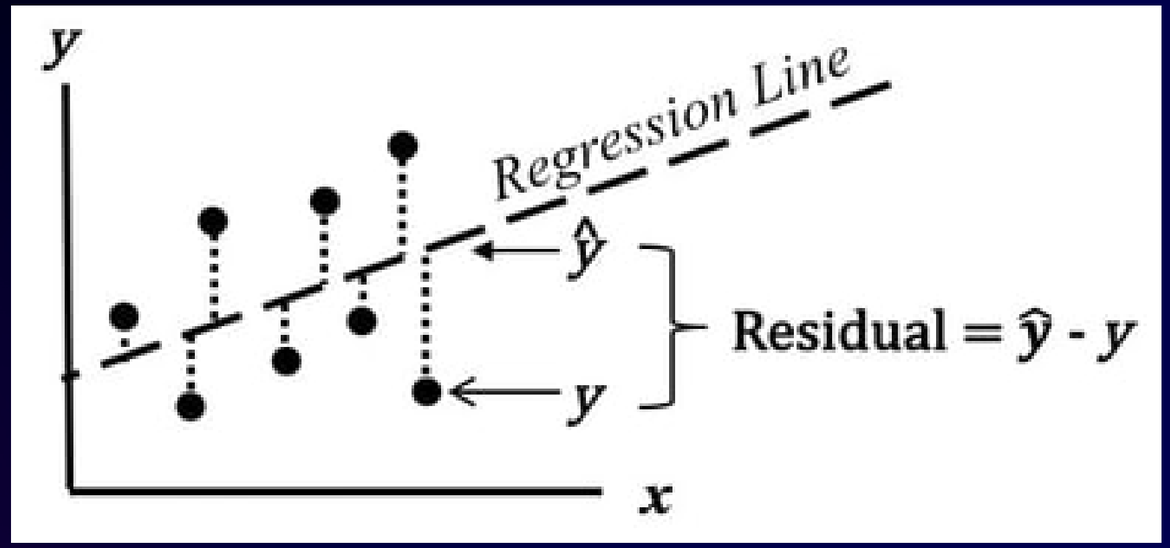


LINEAR REGRESSION

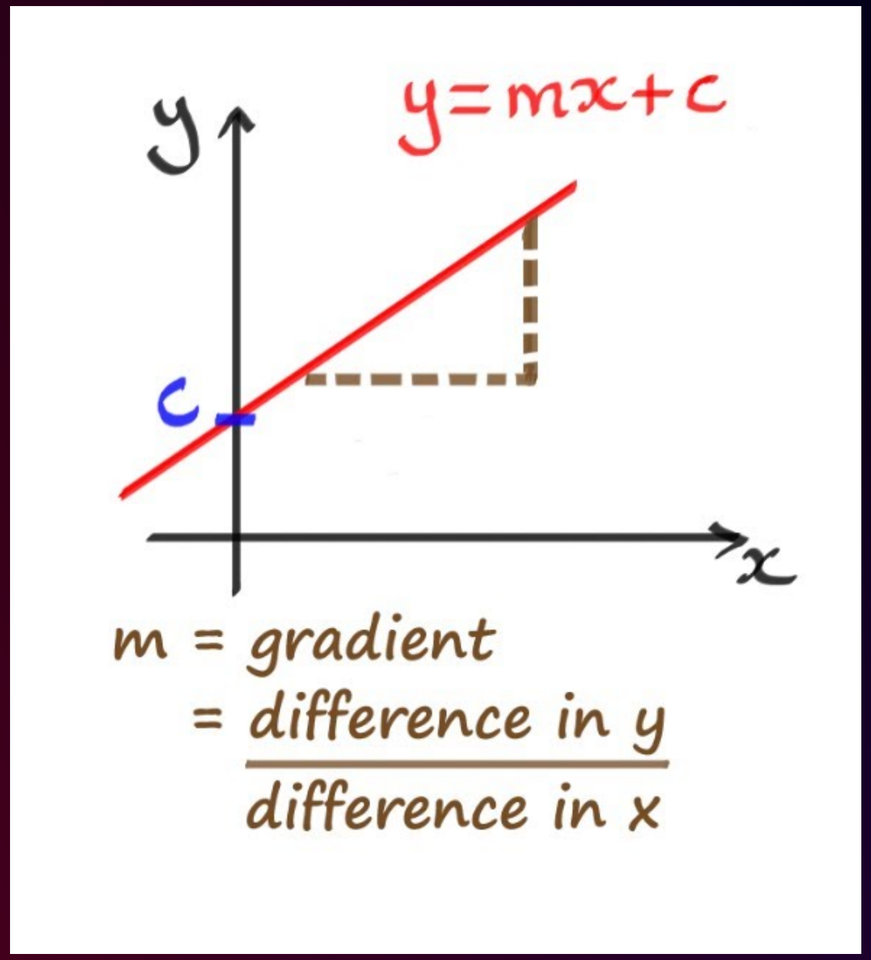


COST FUNCTION - SSE

- The goal of linear regression is to find the line of best fit that can accurately predict the value of the dependent variable based on the values of the independent variables.
- The line of best fit is determined by minimizing the sum of squared errors (SSE) between the actual and predicted values of the dependent variable.



EQUATION OF STRAIGHT LINE



- The Linear regression algorithm will try various combinations of m and c value to get low SSE, that will be the best fitting regression line

MULTIPLE FEATURES VS TARGET

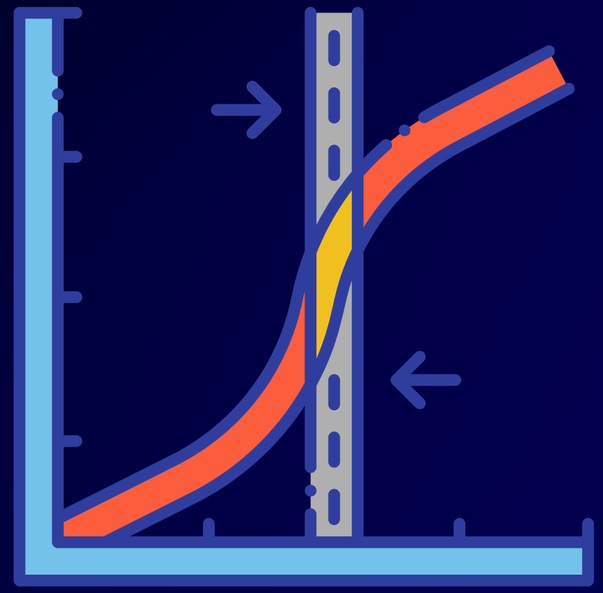
Simple Linear Regression

$$y = b_0 + b_1 * x_1$$

Multiple Linear Regression

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \ldots + b_n * x_n$$

LOGISTIC REGRESSION



LOGISTIC REGRESSION

- Logistic Regression is much similar to the Linear Regression except that how they are used.
- Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.
- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).

TYPES OF LOGISTIC REGRESSION

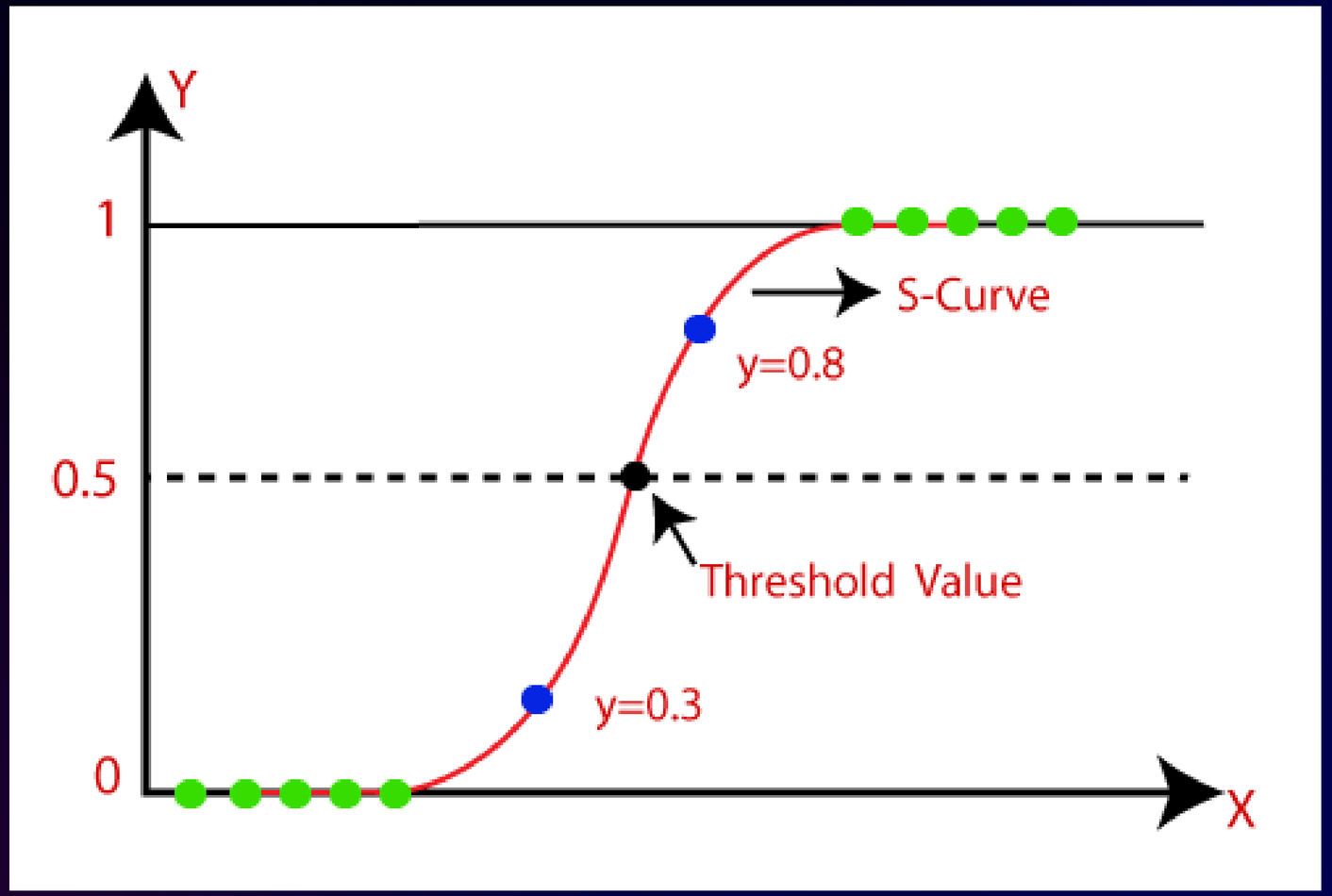
- Binomial: In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.

- Multinomial: In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"

LOGISTIC REGRESSION

	Pclass	Sex	SibSp	Parch	Fare	Embarked_C	Embarked_Q	Embarked_S	Survived
0	3	1	1	0	7.2500	0	0	1	0
1	1	0	1	0	71.2833	1	0	0	1
2	3	0	0	0	7.9250	0	0	1	1
3	1	0	1	0	53.1000	0	0	1	1
4	3	1	0	0	8.0500	0	0	1	0
5	3	1	0	0	8.4583	0	1	0	0
6	1	1	0	0	51.8625	0	0	1	0
7	3	1	3	1	21.0750	0	0	1	0
8	3	0	0	2	11.1333	0	0	1	1
9	2	0	1	0	30.0708	1	0	0	1

LOGISTIC REGRESSION

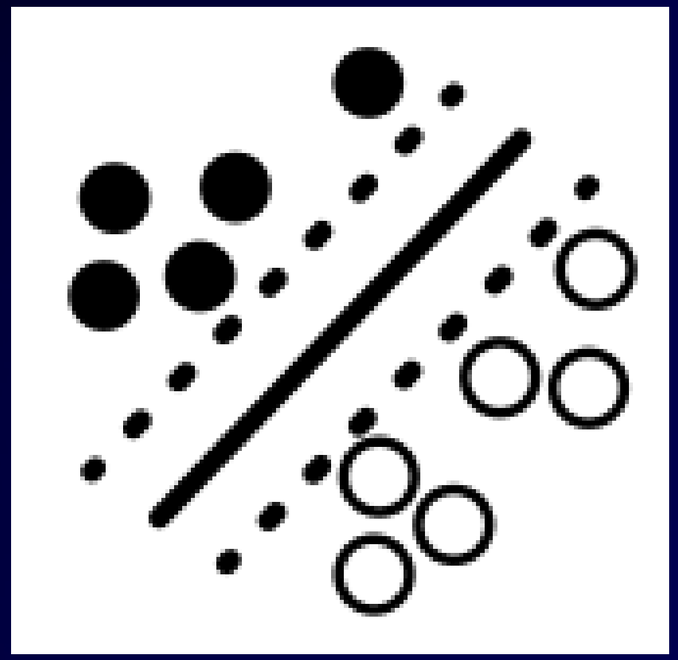


LOGISTIC REGRESSION

- The sigmoid function is a mathematical function used to map real value into another value within a range of 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form.
- The S-form curve is called the Sigmoid function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

$$y = \frac{e^{(b_0 + b_1 X)}}{1 + e^{(b_0 + b_1 X)}}$$

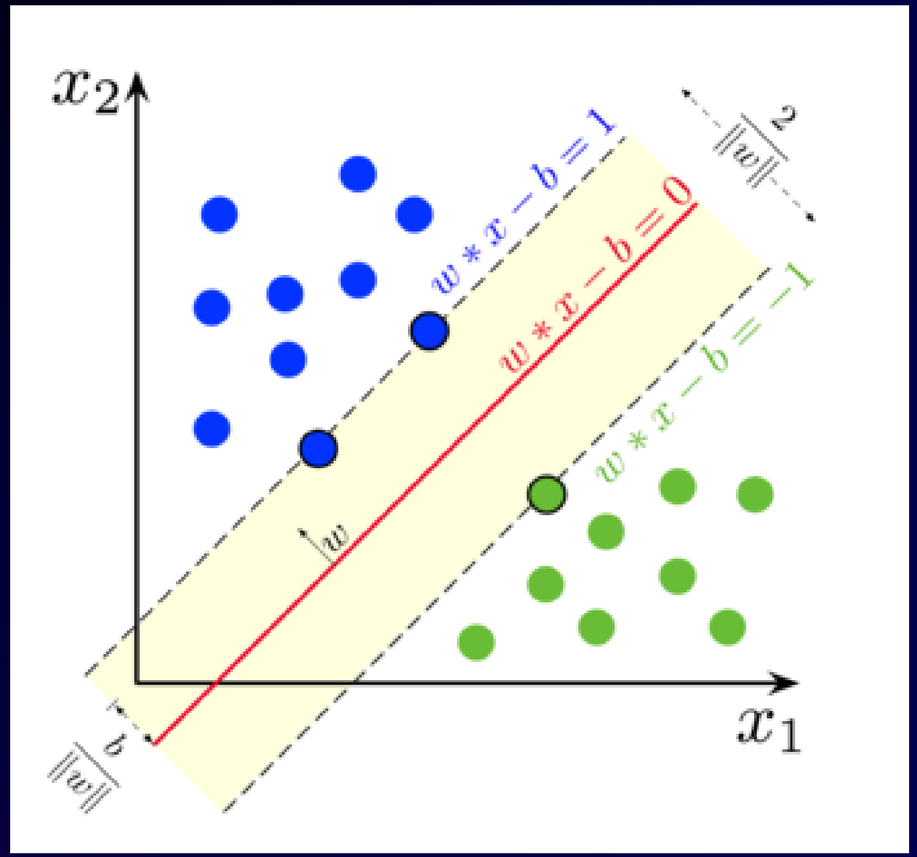
SUPPORT VECTOR MACHINE



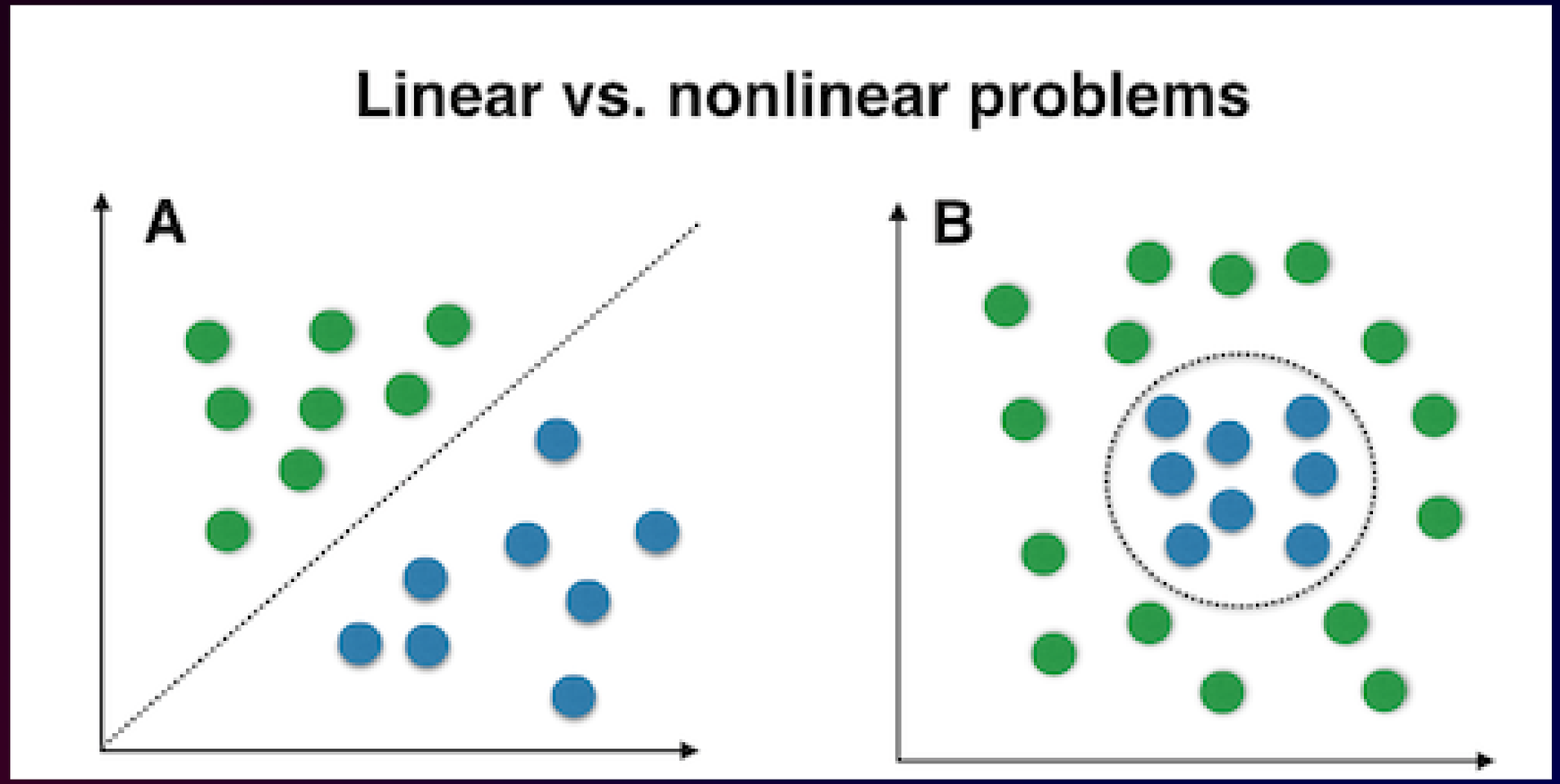
SUPPORT VECTOR MACHINES

- Support Vector Machines (SVMs) are a popular and powerful supervised learning algorithm used for classification and regression analysis.
- SVMs are based on the idea of finding the hyperplane that best separates the data into different classes, by maximizing the margin between the hyperplane and the closest data points.
- SVM is a very versatile algorithm and can handle both linear and nonlinear datasets.

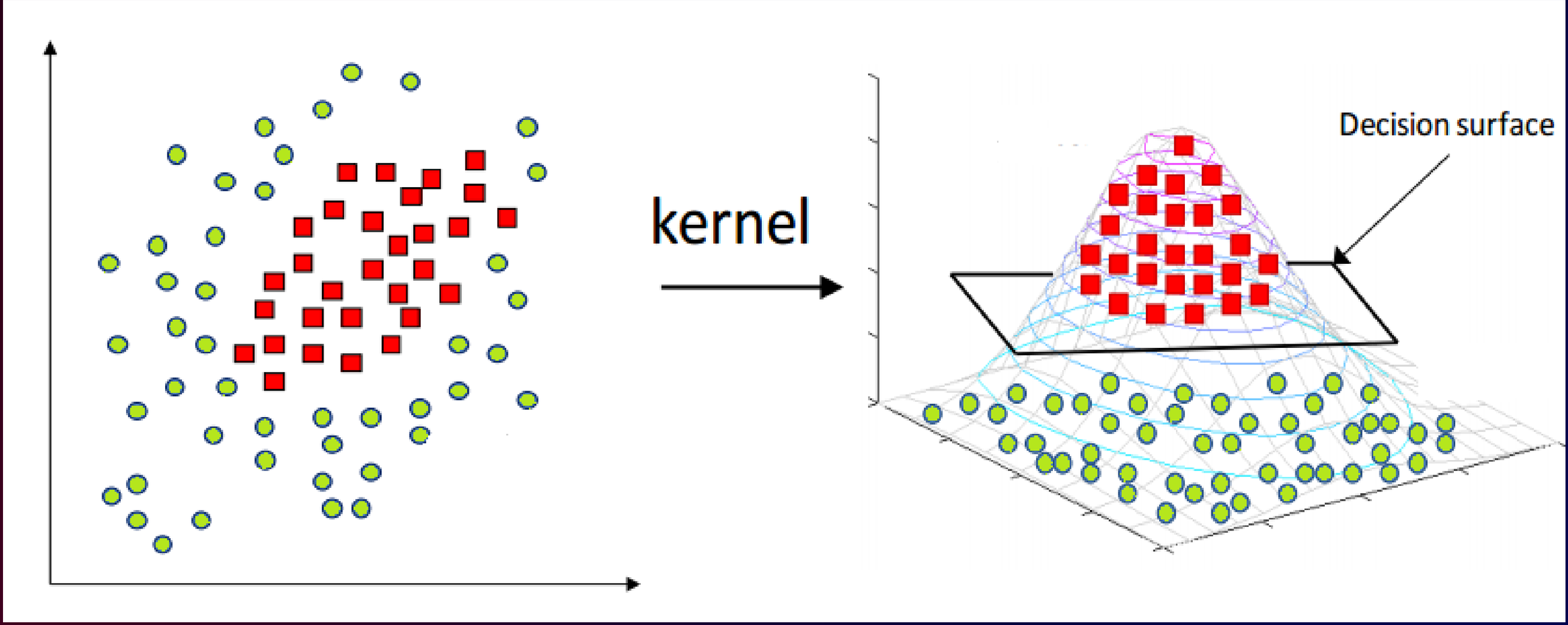
SUPPORT VECTOR MACHINE - CLASSIFIER



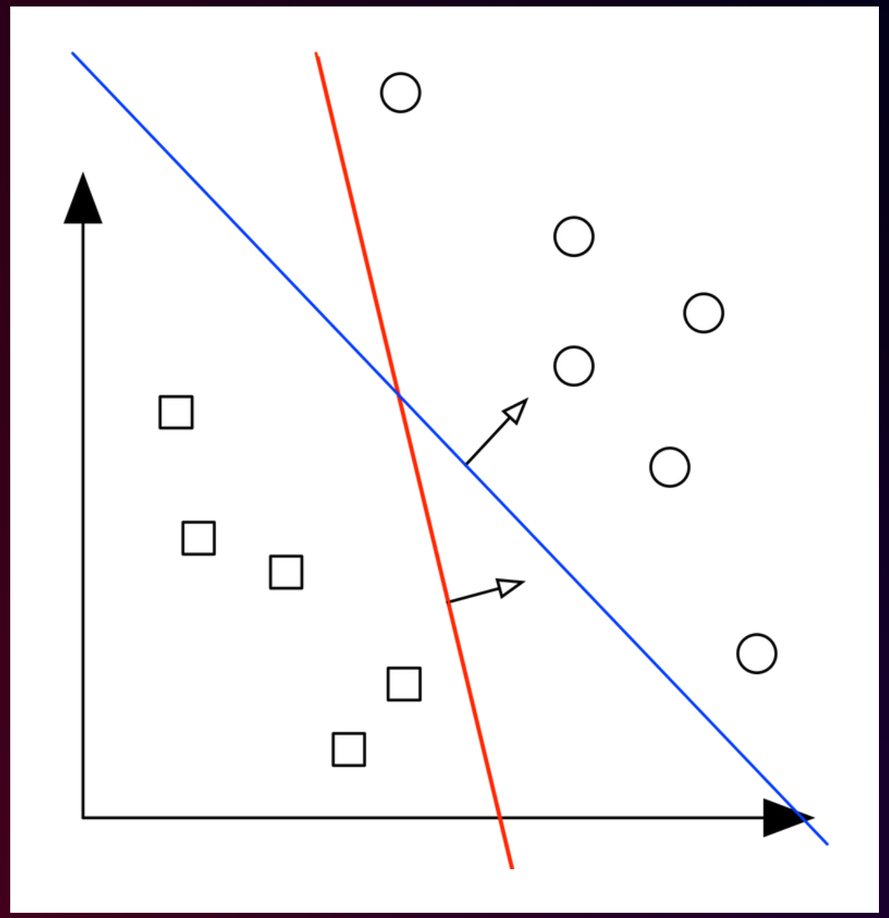
LINEAR VS NON LINEAR PROBLEM

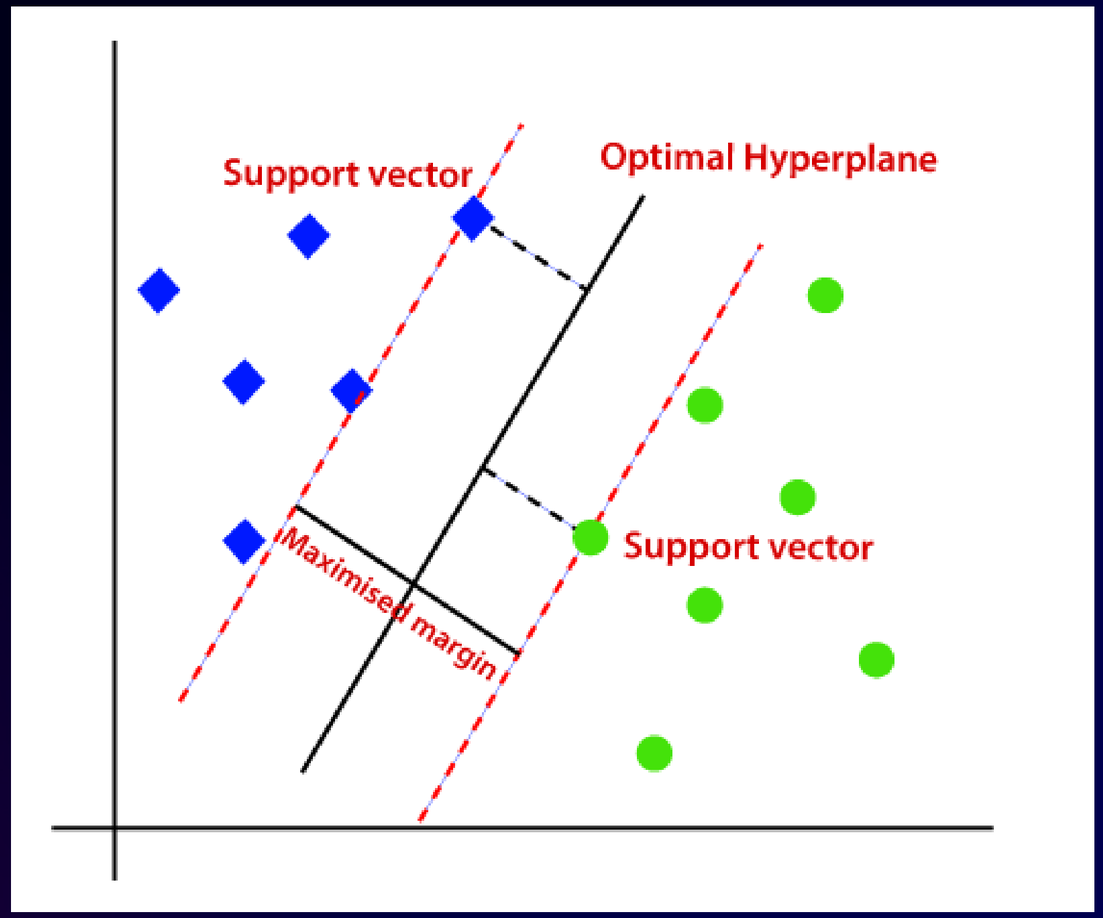


LINEAR VS NON LINEAR PROBLEM

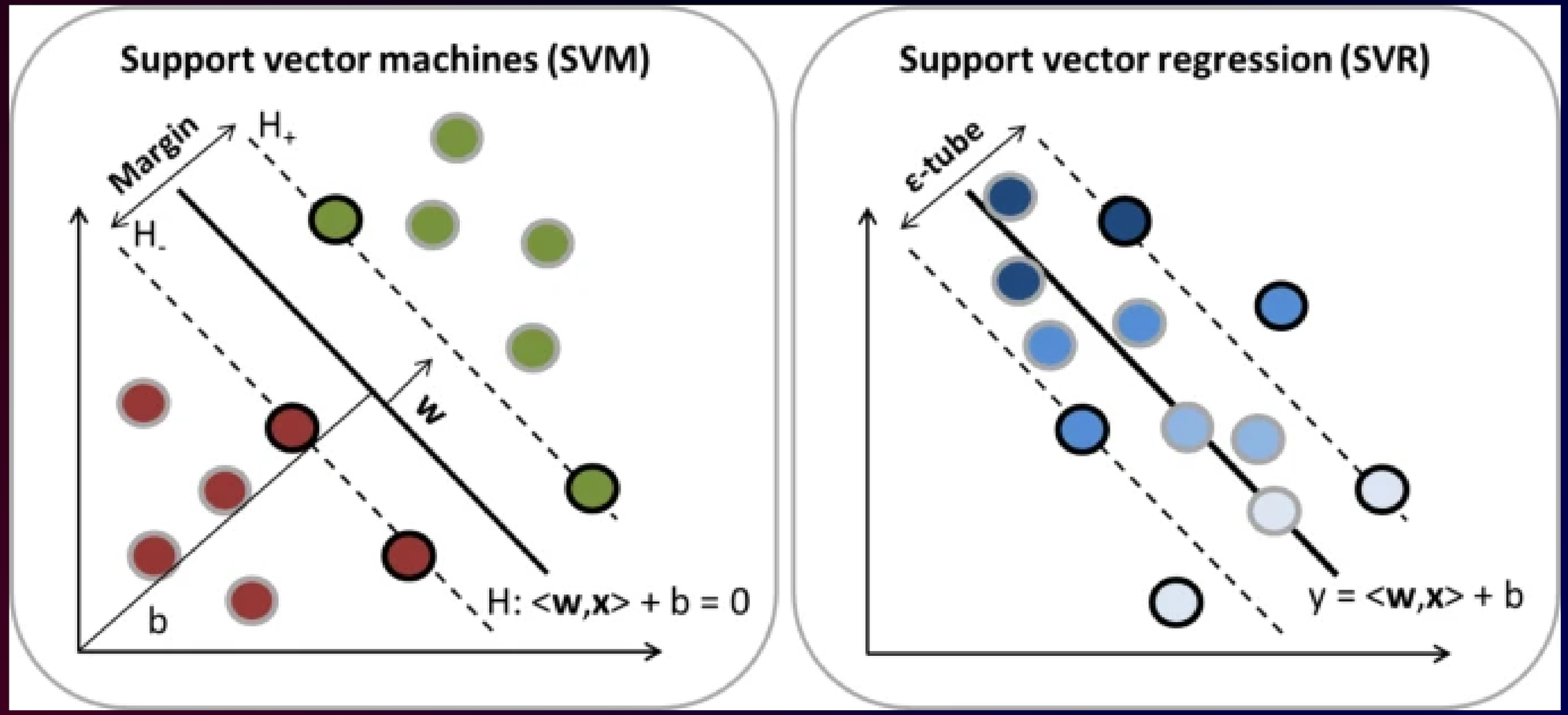


LINEAR VS NON LINEAR PROBLEM

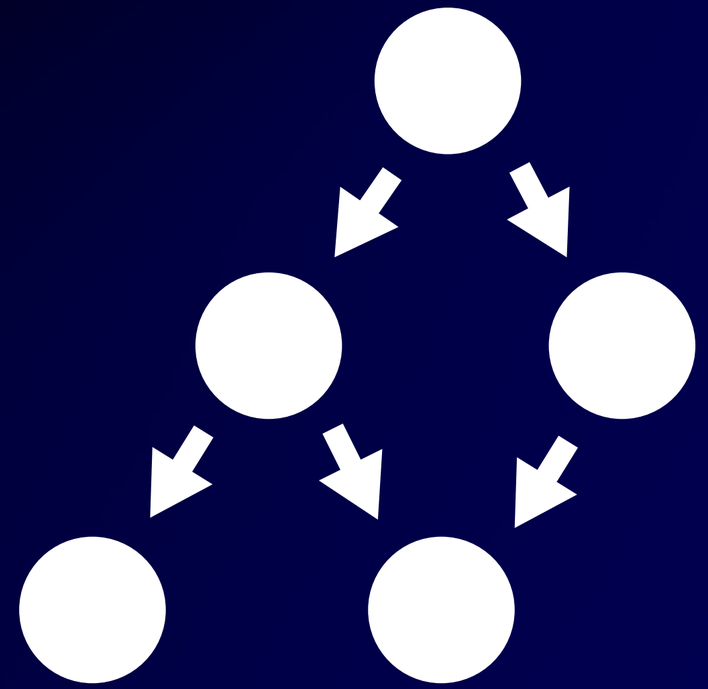




SUPPORT VECTOR REGRESSION



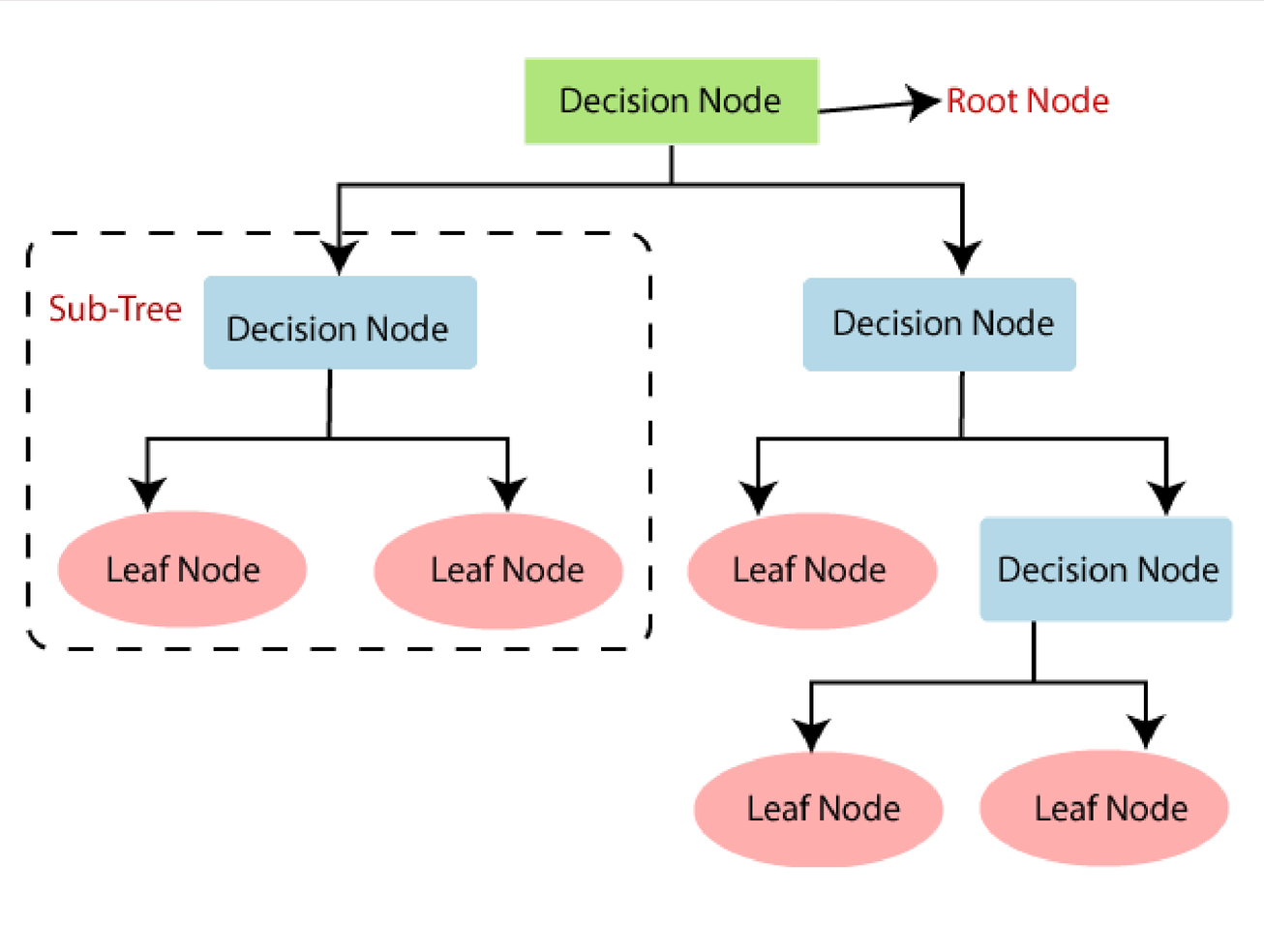
DECISION TREE

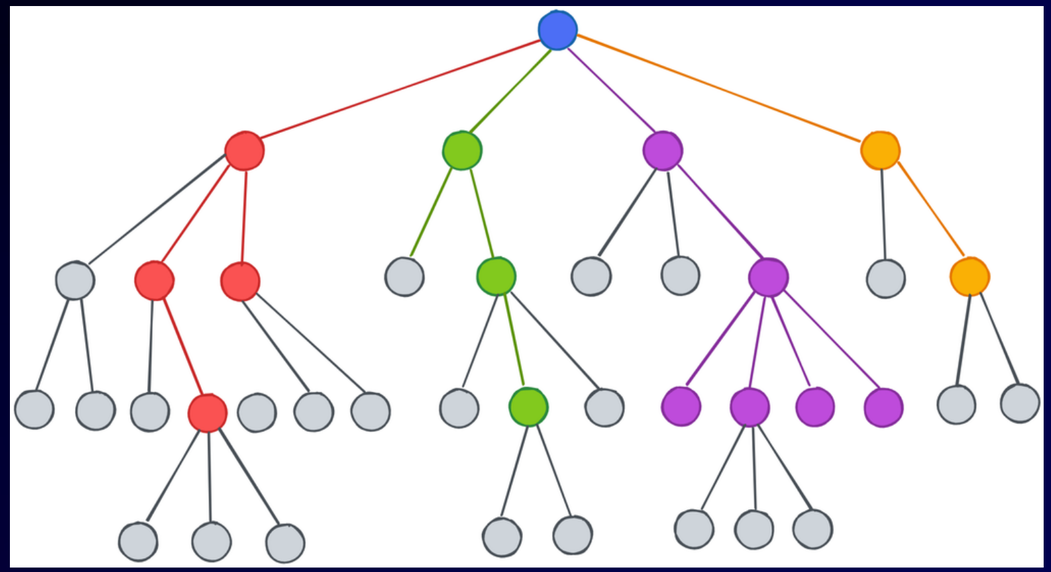


DECISION TREE

- 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.
- 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

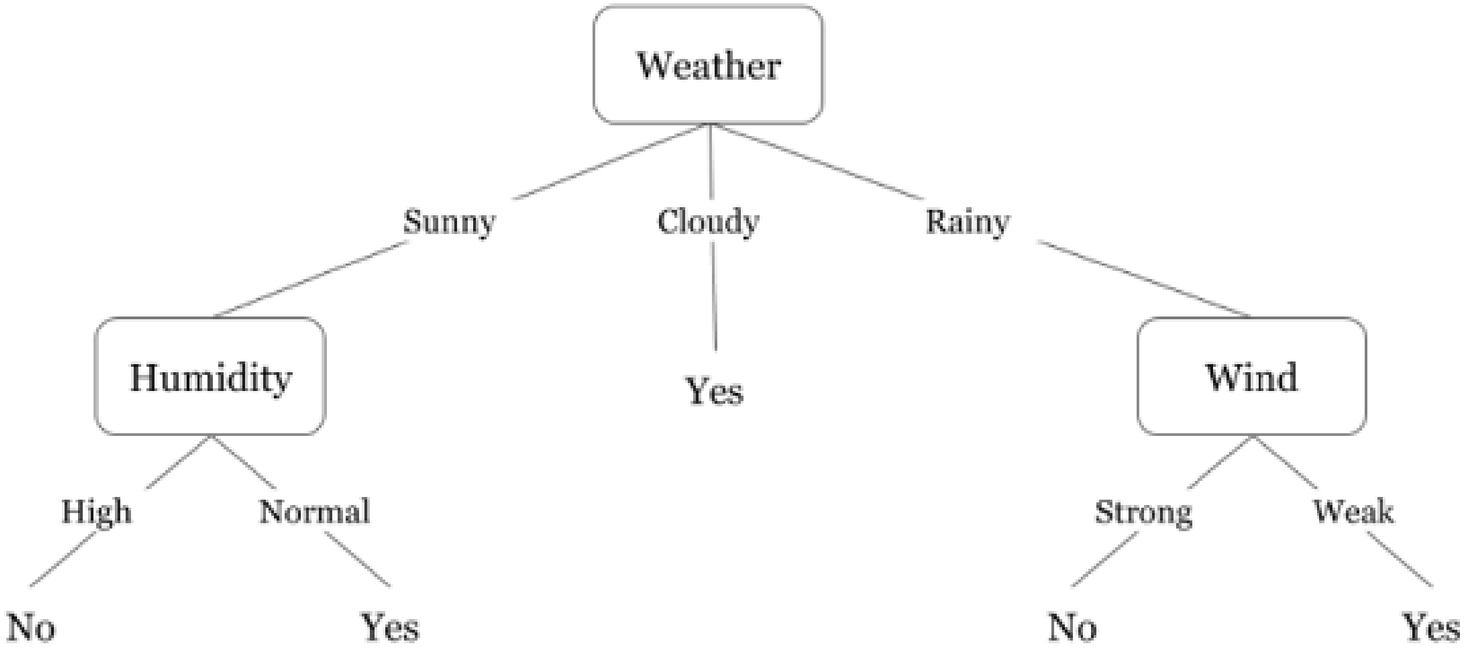




DECISION TREE

Day	Weather	Temperature	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Cloudy	Hot	High	Weak	Yes
3	Sunny	Mild	Normal	Strong	Yes
4	Cloudy	Mild	High	Strong	Yes
5	Rainy	Mild	High	Strong	No
6	Rainy	Cool	Normal	Strong	No
7	Rainy	Mild	High	Weak	Yes
8	Sunny	Hot	High	Strong	No
9	Cloudy	Hot	Normal	Weak	Yes
10	Rainy	Mild	High	Strong	No

Day	Weather	Temperature	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Cloudy	Hot	High	Weak	Yes
3	Sunny	Mild	Normal	Strong	Yes
4	Cloudy	Mild	High	Strong	Yes
5	Rainy	Mild	High	Strong	No
6	Rainy	Cool	Normal	Strong	No
7	Rainy	Mild	High	Weak	Yes
8	Sunny	Hot	High	Strong	No
9	Cloudy	Hot	Normal	Weak	Yes
10	Rainy	Mild	High	Strong	No



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

Play Golf	
Yes	No
9	5

Entropy(PlayGolf) = Entropy (5,9)

= Entropy (0.36, 0.64)

= - (0.36 $\log_2$ 0.36) - (0.64 $\log_2$ 0.64)

= 0.94

$$Gain(T, X) = Entropy(T) - Entropy(T, X)$$

parent entropy $-\left(\frac{14}{30}\cdot\log_2\frac{14}{30}\right)-\left(\frac{16}{30}\cdot\log_2\frac{16}{30}\right)=$ **0.996**

child 1 entropy $-\left(\frac{13}{17}\cdot\log_2\frac{13}{17}\right)-\left(\frac{4}{17}\cdot\log_2\frac{4}{17}\right)=$ **0.787**

child 2 entropy $-\left(\frac{1}{13}\cdot\log_2\frac{1}{13}\right)-\left(\frac{12}{13}\cdot\log_2\frac{12}{13}\right)=$ **0.391**

(Weighted) Average Entropy of children = $\left(\frac{17}{30}\cdot 0.787\right)+\left(\frac{13}{30}\cdot 0.391\right)=0.615$

Information Gain - 0.996 - 0.615 = 0.38 for this split

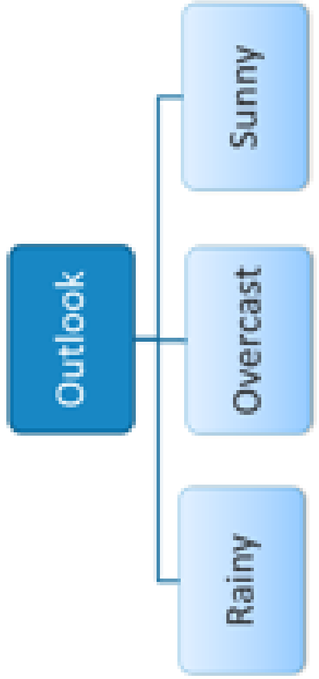
		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1
Gain = 0.029			

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1
Gain = 0.152			

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3
Gain = 0.048			

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			



Outlook	Temp	Humidity	Windy	Play Golf
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Sunny	Mild	Normal	FALSE	Yes
Sunny	Mild	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
Overcast	Cool	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes

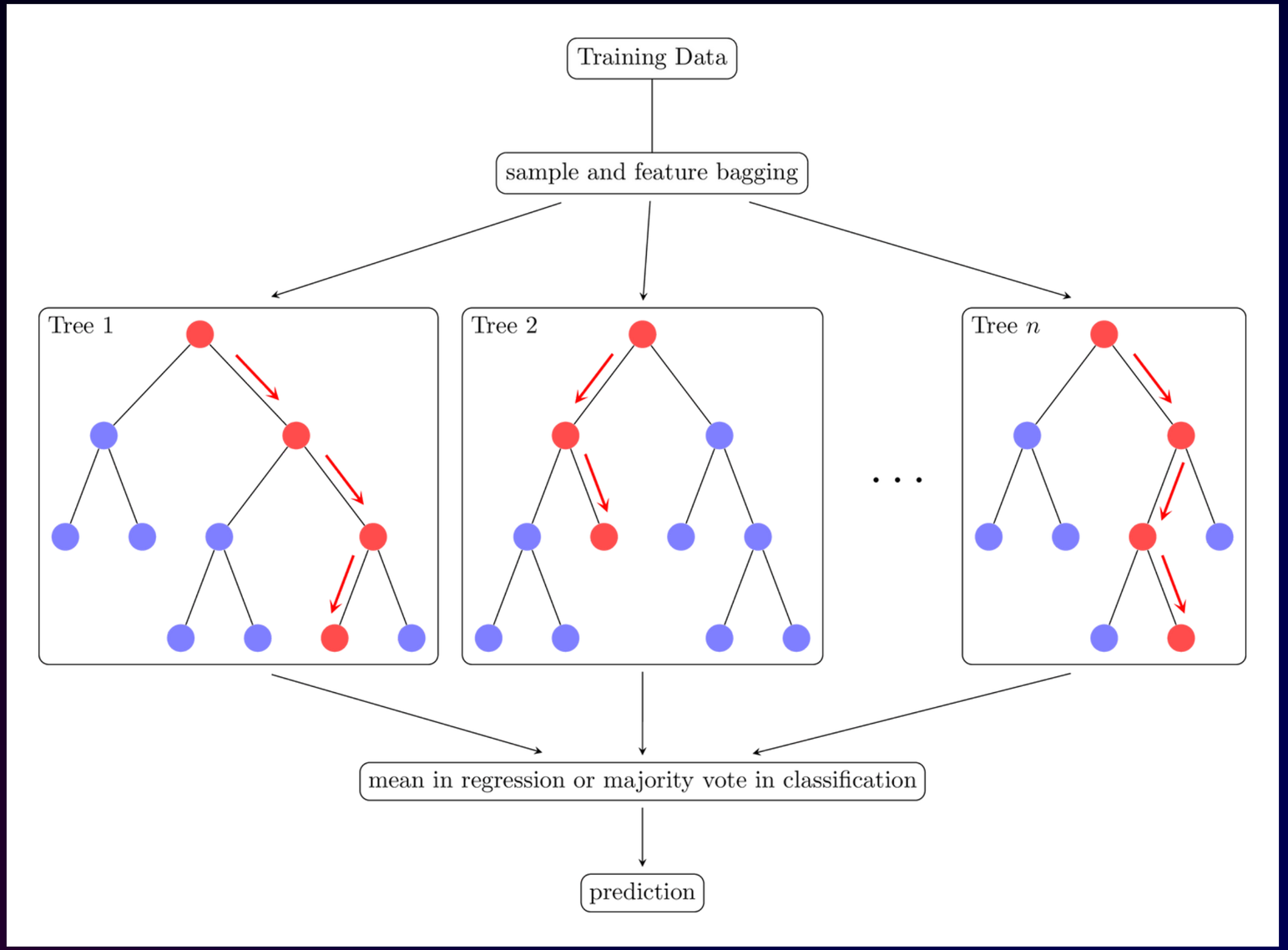
ENSEMBLE ALGORITHM

1. Bagging - Random Forest
2. Boosting - AdaBoost, XGBoost

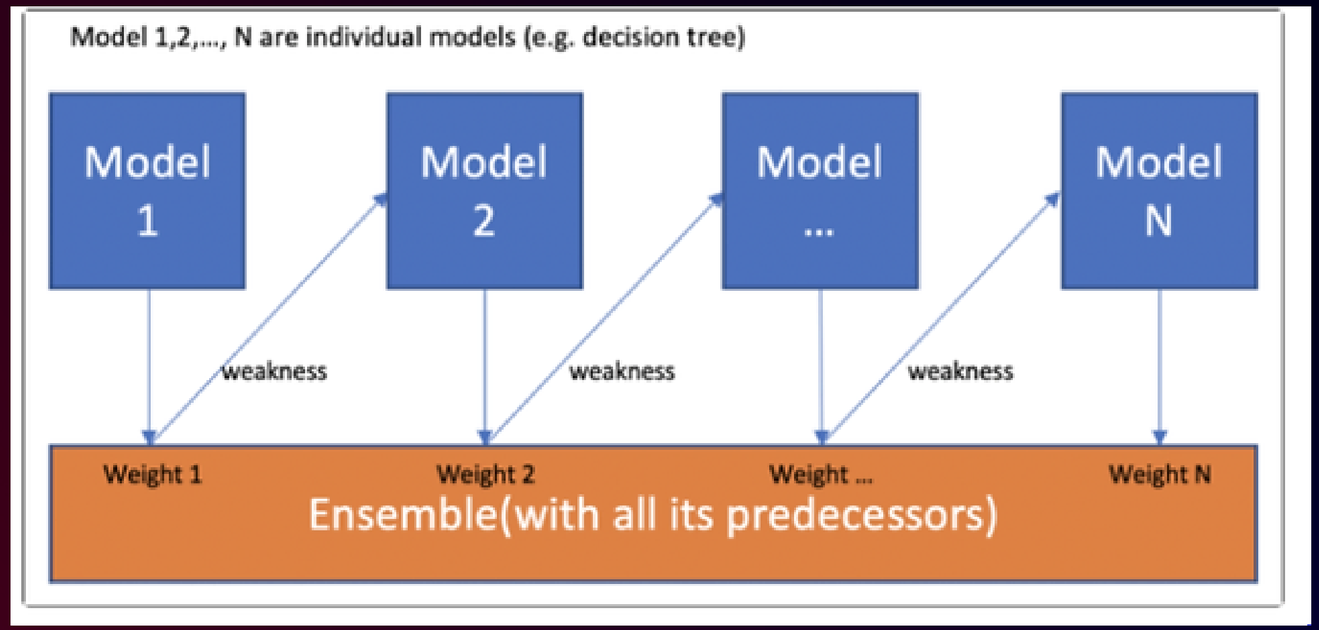
RANDOM FOREST

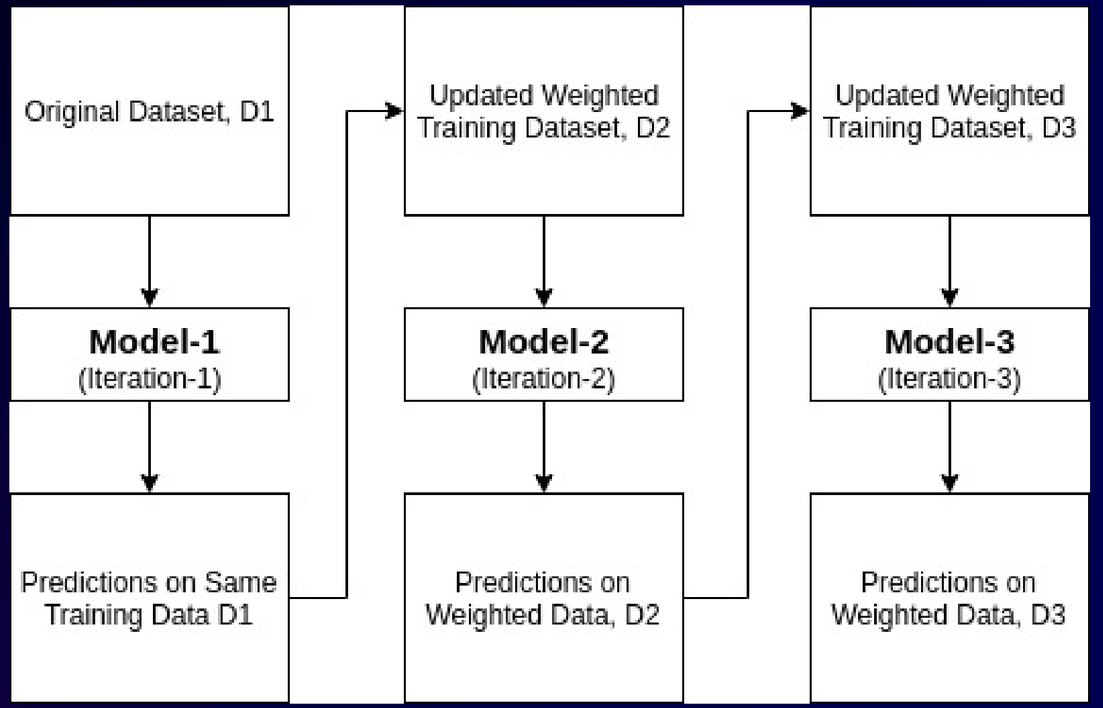
- Random forest is a popular machine learning algorithm that is used for both classification and regression tasks.
- It is an ensemble learning method that combines multiple decision trees to create a more robust and accurate model.
- The random forest algorithm works by building a large number of decision trees, each using a randomly selected subset of the available input features and a randomly selected subset of the training data.
- These decision trees are then combined to make predictions on new data by taking a vote or averaging the outputs of each individual tree.

RANDOM FOREST



ADABOOST

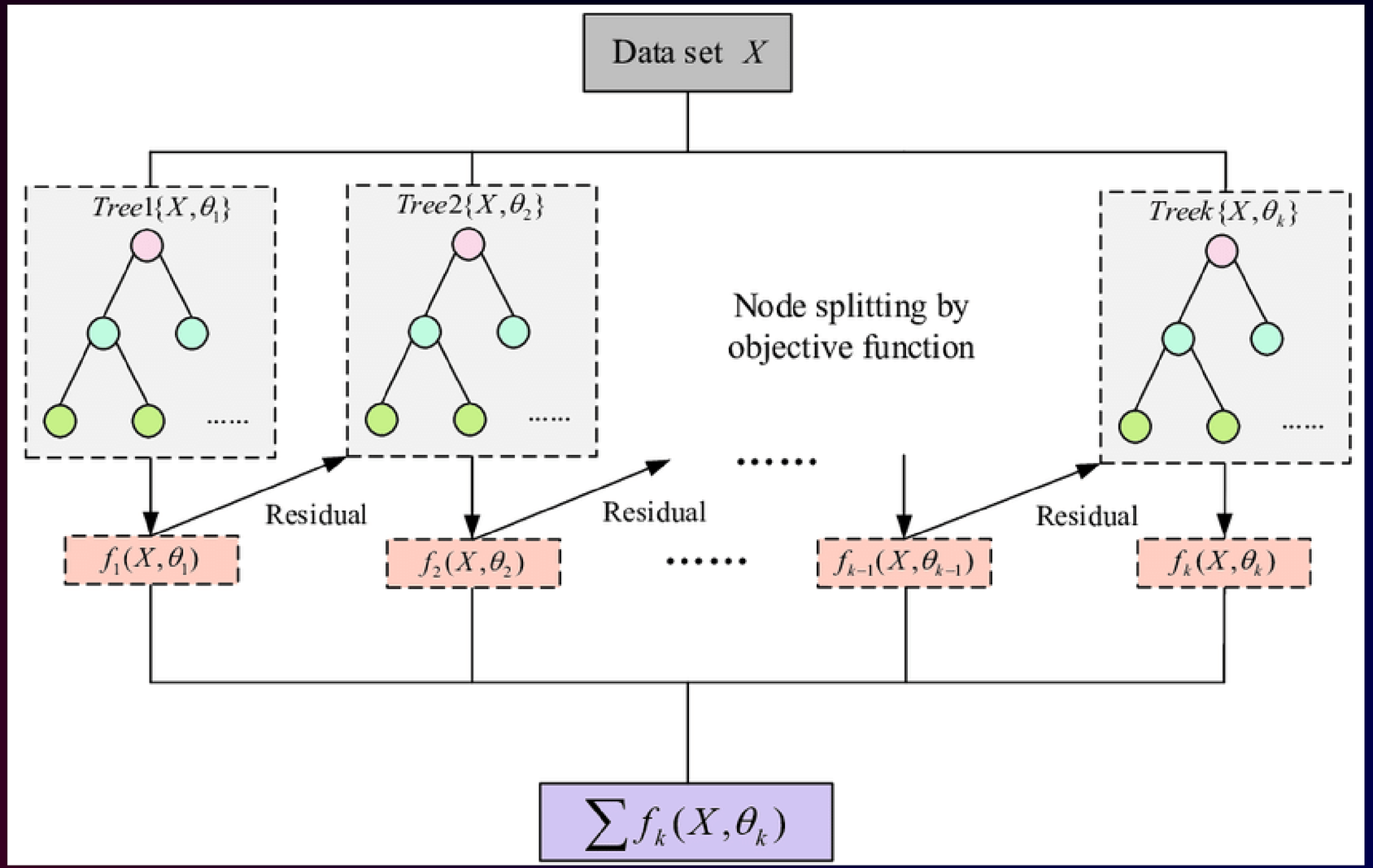

Model 1,2,..., N are individual models (e.g. decision tree)
Model 1
Model 2
Model ...
Model N
weakness
weakness
weakness
Weight 1
Weight 2
Weight ...
Weight N
Ensemble(with all its predecessors)


Original Dataset, D1
Updated Weighted Training Dataset, D2
Updated Weighted Training Dataset, D3
Model-1 (Iteration-1)
Model-2 (Iteration-2)
Model-3 (Iteration-3)
Predictions on Same Training Data D1
Predictions on Weighted Data, D2
Predictions on Weighted Data, D3

XG BOOST

- XGBoost (EXTREME GRADIENT Boost) works by building a series of decision trees in a gradient boosting framework.
- In this framework, each tree is built to correct the errors of the previous tree, with the final prediction being the sum of the predictions of all the individual trees.
- Another important feature of XGBoost is its ability to handle missing data.
- XGBoost can automatically learn how to handle missing data by assigning different directions in the tree to the missing values and using the best split based on the available data.

XG BOOST


Data set X
$Tree1\{X,\theta_1\}$
$Tree2\{X,\theta_2\}$
$Treek\{X,\theta_k\}$
.....
Node splitting by objective function
Residual
Residual
Residual
$f_1(X,\theta_1)$
$f_2(X,\theta_2)$
$f_{k-1}(X,\theta_{k-1})$
$f_k(X,\theta_k)$
$\sum f_k(X,\theta_k)$

Thank you