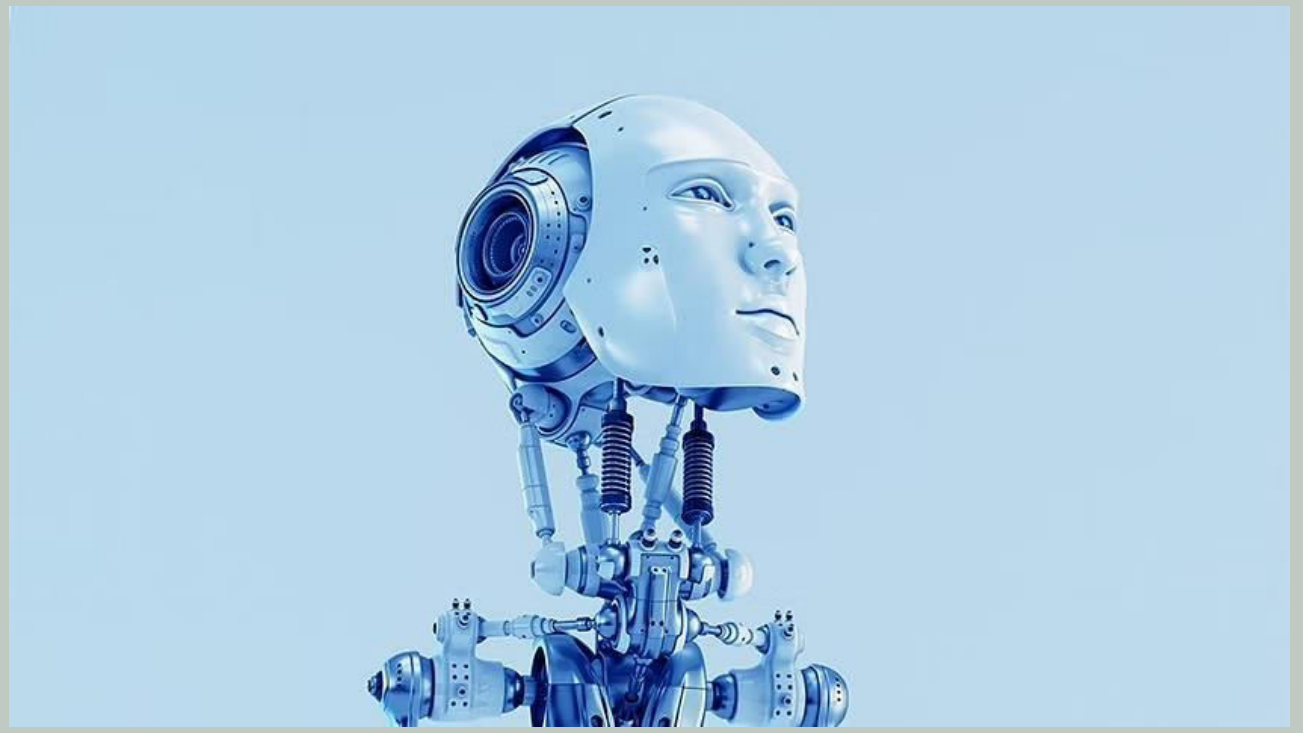
Supervised Learning



September 24, 2022

A023119820006

Topics

01 Introduction

02 Classification and regression

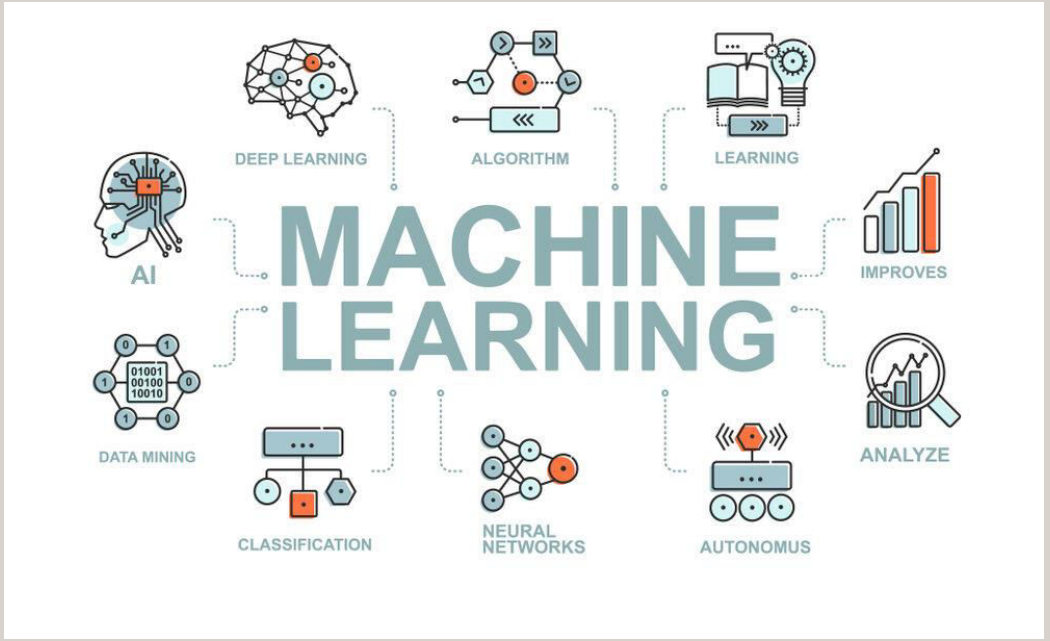
03 Regression

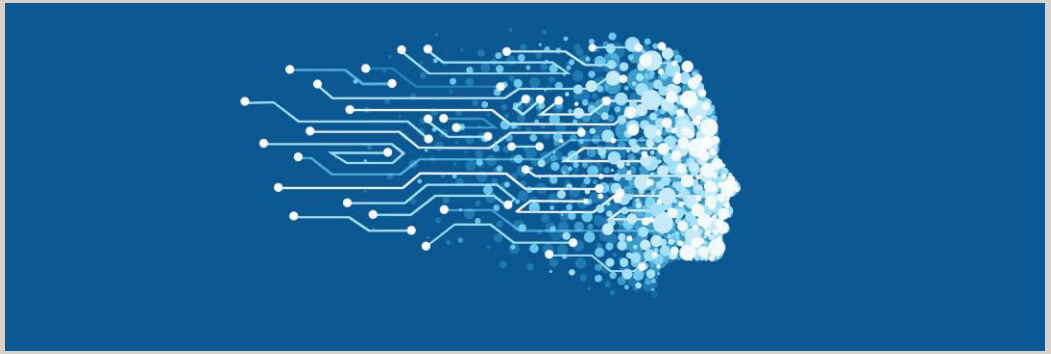
04 Classification

05 Gradient Decent

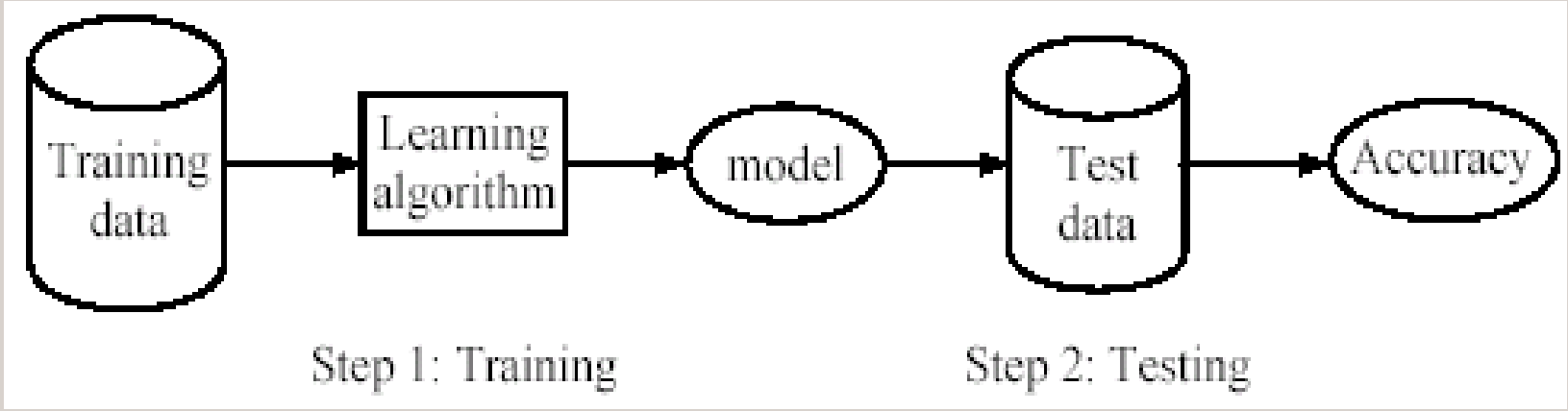
01 Introduction

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.





Process



Learning (training):

Learn a model using the training data

$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$

Testing:

Test the model using unseen test data to assess the model accuracy

02 Types of Supervised Learning

Types

Regression

Regression algorithms predict a continuous value based on the input variables.

The main goal of regression problems is to give an estimate value of any prediction in number form.

For example-

- Simple Regression
- Multiple Regression
- Polynomial Regression

Classification

A classification model attempts to draw some conclusion from observed values.

A classification algorithm can have both discrete and real-valued variables. The output is often yes or no , 1 or 2 etc.

For example-

- Logistical Regression
- Decision Tree
- Support Vector Machine

03

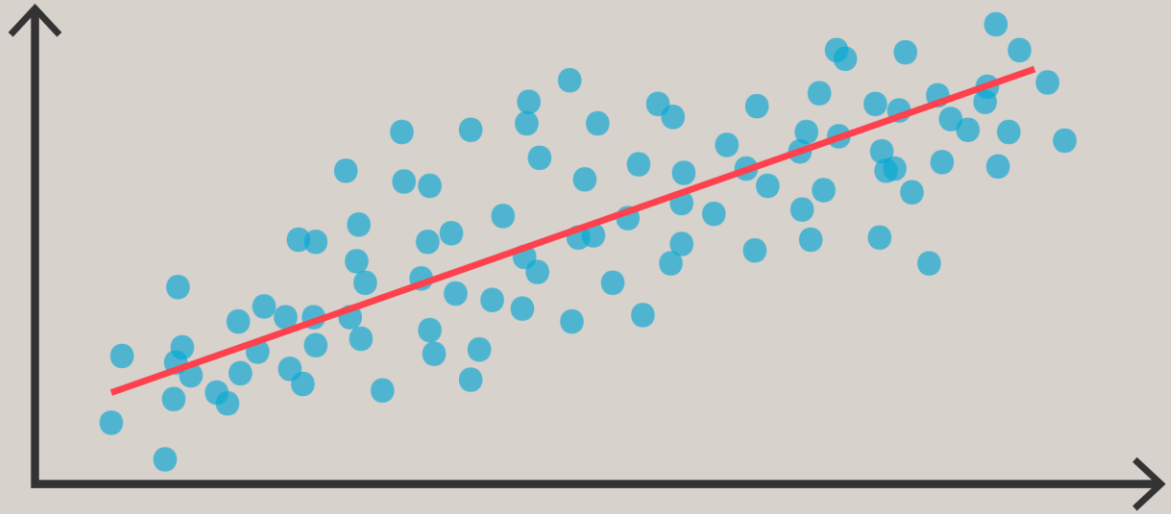
Regression Model

03 Regression

Definition:

Regression analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables.

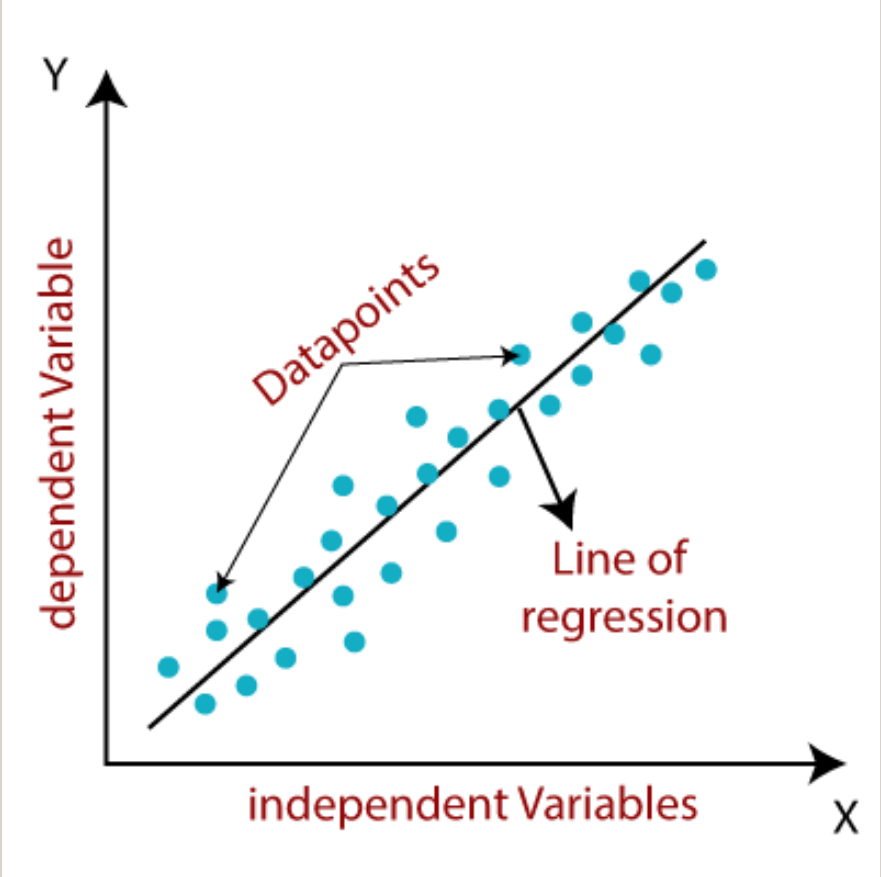
It predicts continuous/real values such as temperature, age, salary, price, etc.



03 Linear Regression:

It is one of the many algorithms which works on regression and shows the relationship between the continuous variables.

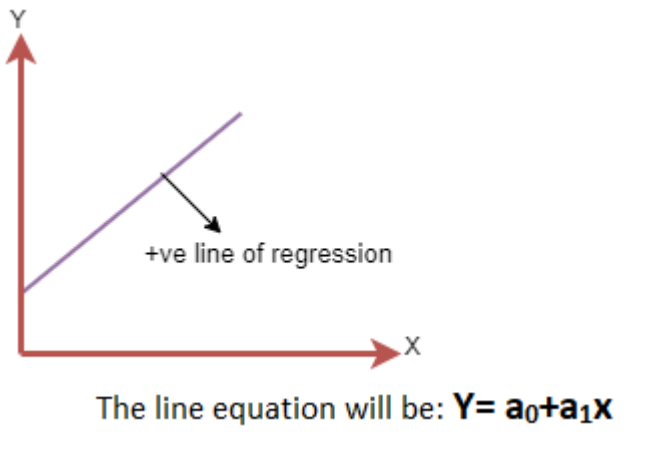
If there is only one input variable (x), then such linear regression is called simple linear regression. And if there is more than one input variable, then such linear regression is called multiple linear regression.

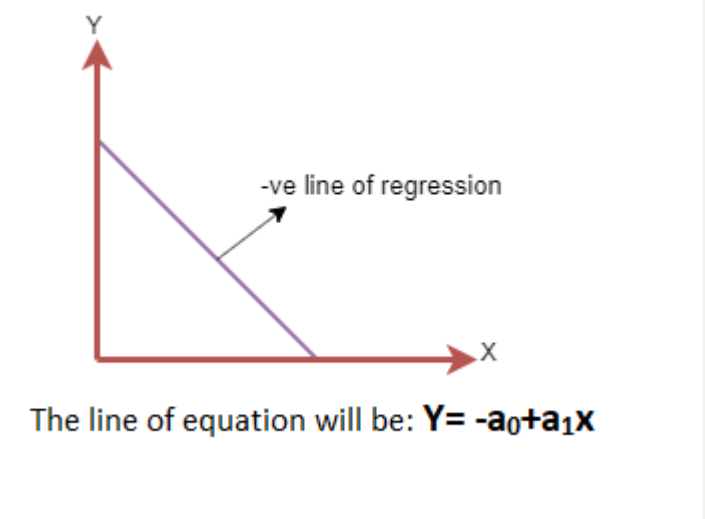


Linear Regression

- **Mathematically, we can represent a linear regression as:**

$$y= a0+a1x+ \varepsilon$$





Linear Regression

- **Cost Function-**
- The different values for weights or coefficient of lines (a0, a1) gives the different line of regression, and the cost function is used to estimate the values of the coefficient for the best fit line.
- Cost function optimizes the regression coefficients or weights. It measures how a linear regression model is performing.
- For Linear Regression, we use the Mean Squared Error (MSE) cost function, which is the average of squared error occurred between the predicted values and actual values.

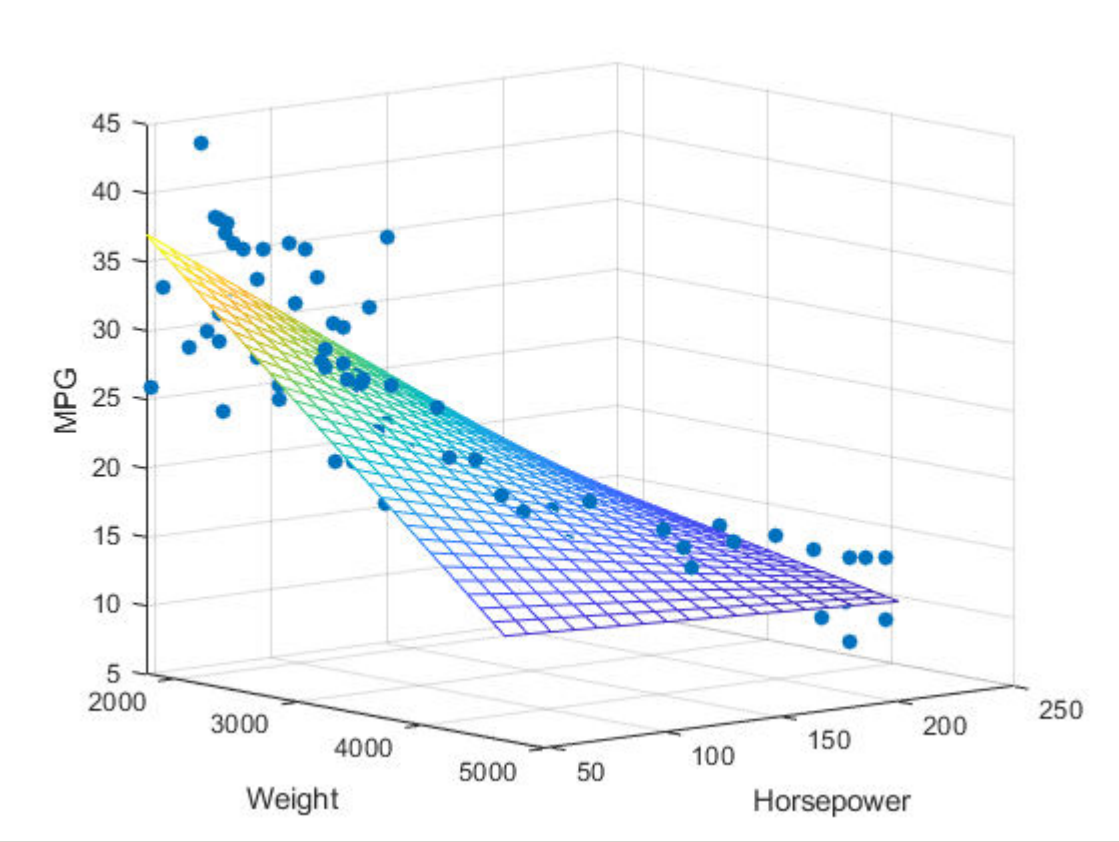
$$MSE = 1\frac{1}{N}\sum_{i=1}^{n}(y_i - (a_1 x_i + a_0))^2$$

03 Multiple Regression:

Multiple Linear Regression is one of the important regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon$$

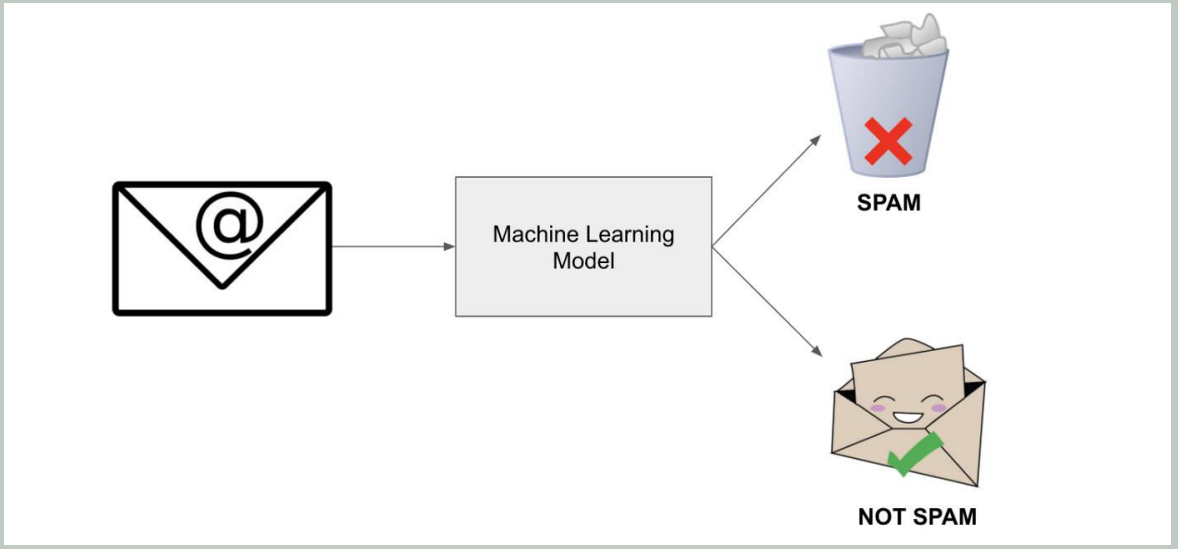
Prediction of CO2 emission based on engine size and number of cylinders in a car.



04 Classification Models

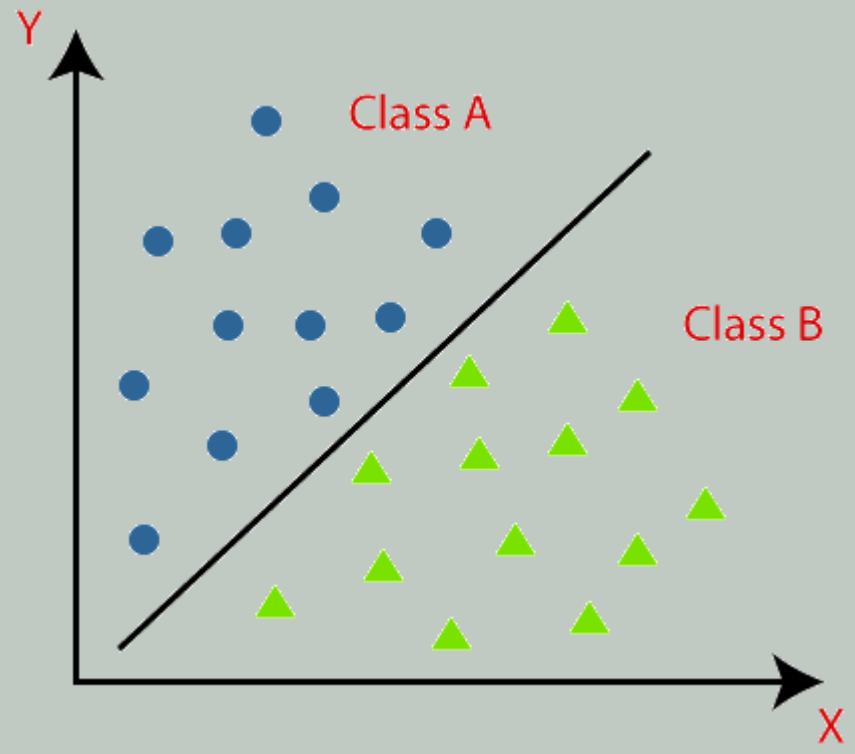
04 Classification model

- The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations based on training data.
- In Classification, a program learns from the given dataset or observations and then classifies new observation into several classes or groups.



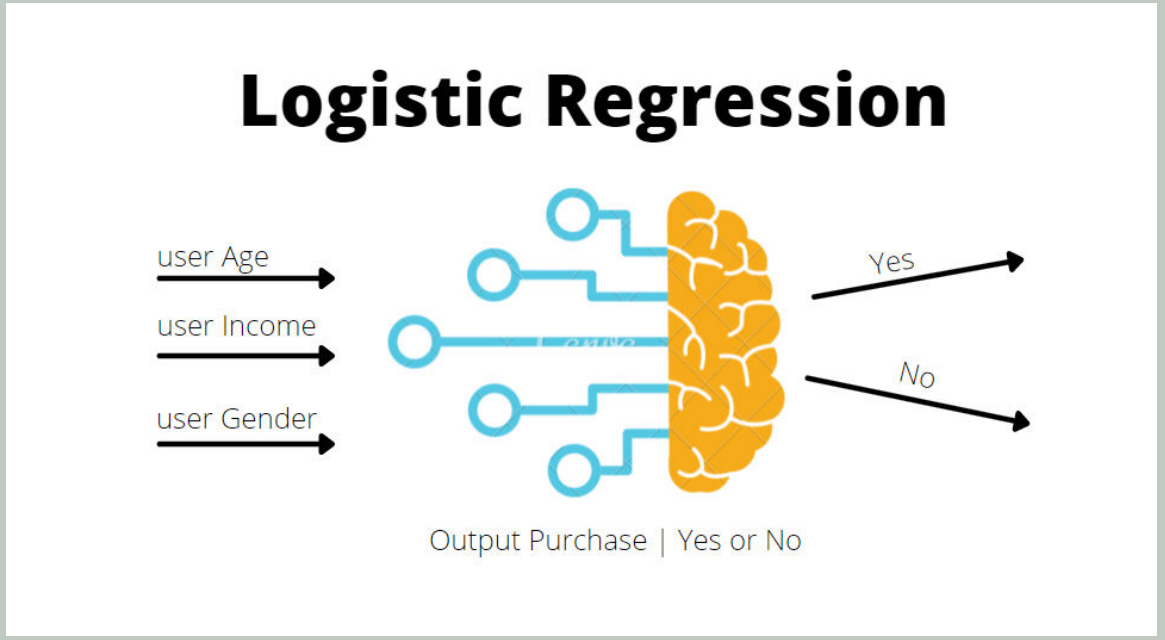
04 Classification model

- The main goal of the Classification algorithm is to identify the category of a given dataset, and these algorithms are mainly used to predict the output for the categorical data .
- Classification algorithms can be better understood using the diagram. In the diagram, there are two classes, class A and Class B. These classes have features that are similar to each other and dissimilar to other classes.



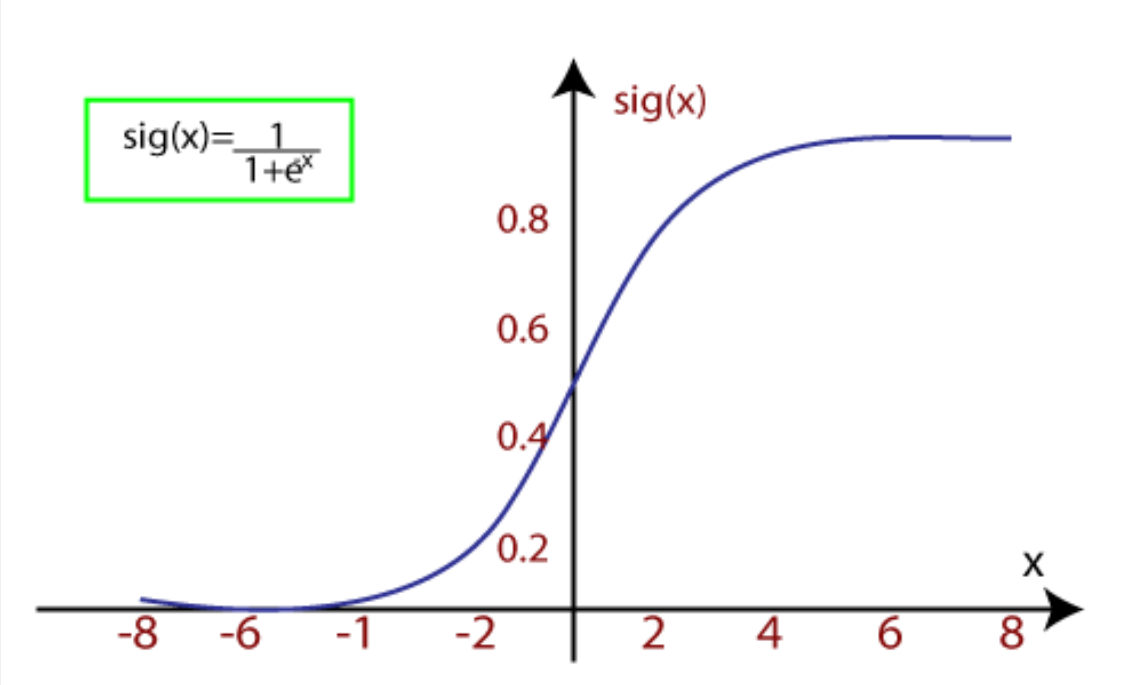
04 Logistical Regression

- 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.
- It is a predictive analysis algorithm which works on the concept of probability.



Logistic Regression

- Logistic regression uses sigmoid function or logistic function which is a complex cost function. This sigmoid function is used to model the data in logistic regression.



04 Naïve Bayes

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems . It is a predictive analysis algorithm which works on the concept of probability.
- It is a probabilistic classifier, which means it predicts based on the probability of an object.
- It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Hence each feature individually contributes to identify that it is an apple without depending on each other.
- It is called Bayes because it depends on the principle of Bayes' Theorem.

Naïve Bayes

- The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.
- P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.
- P(A) is Prior Probability: Probability of hypothesis before observing the evidence.
- P(B) is Marginal Probability: Probability of Evidence

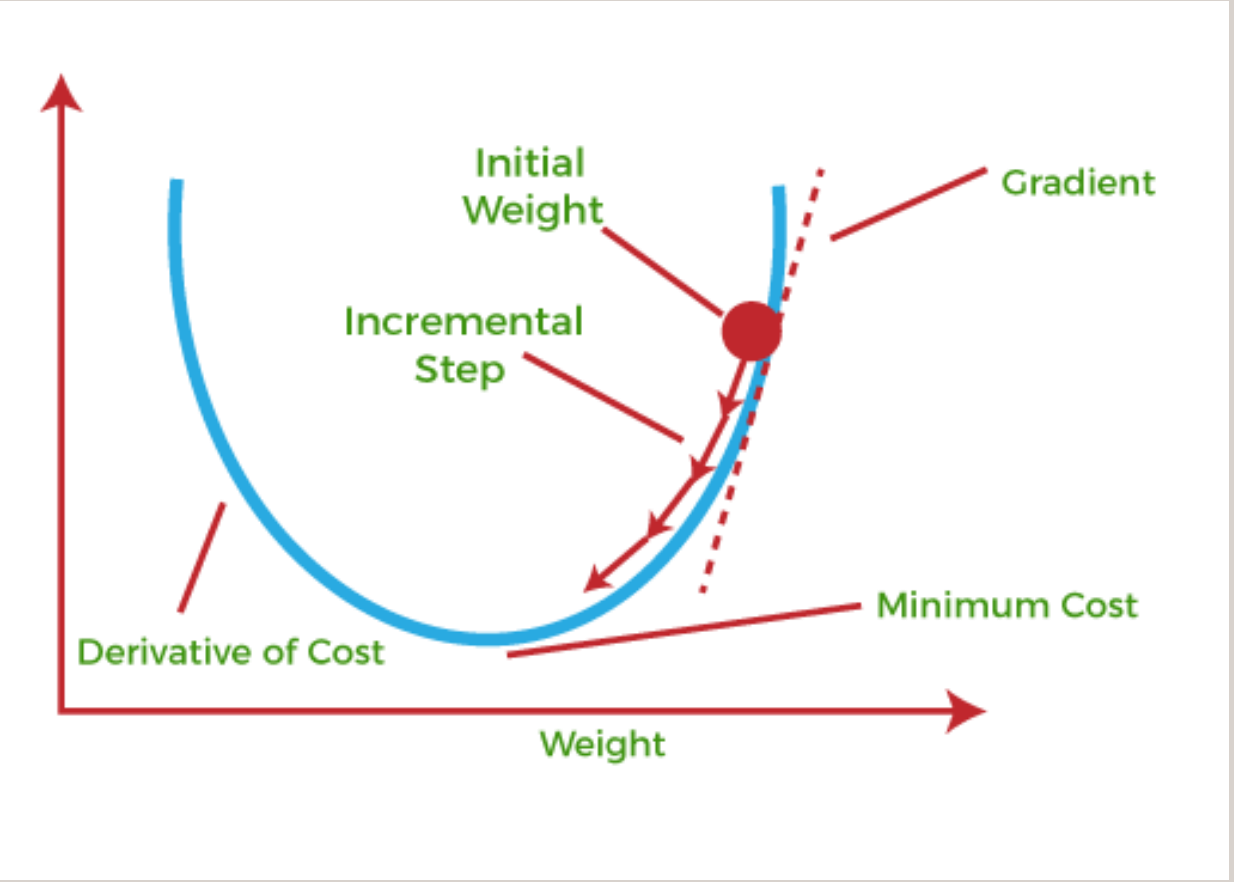
05

Gradient Decent

05 Gradient Decent

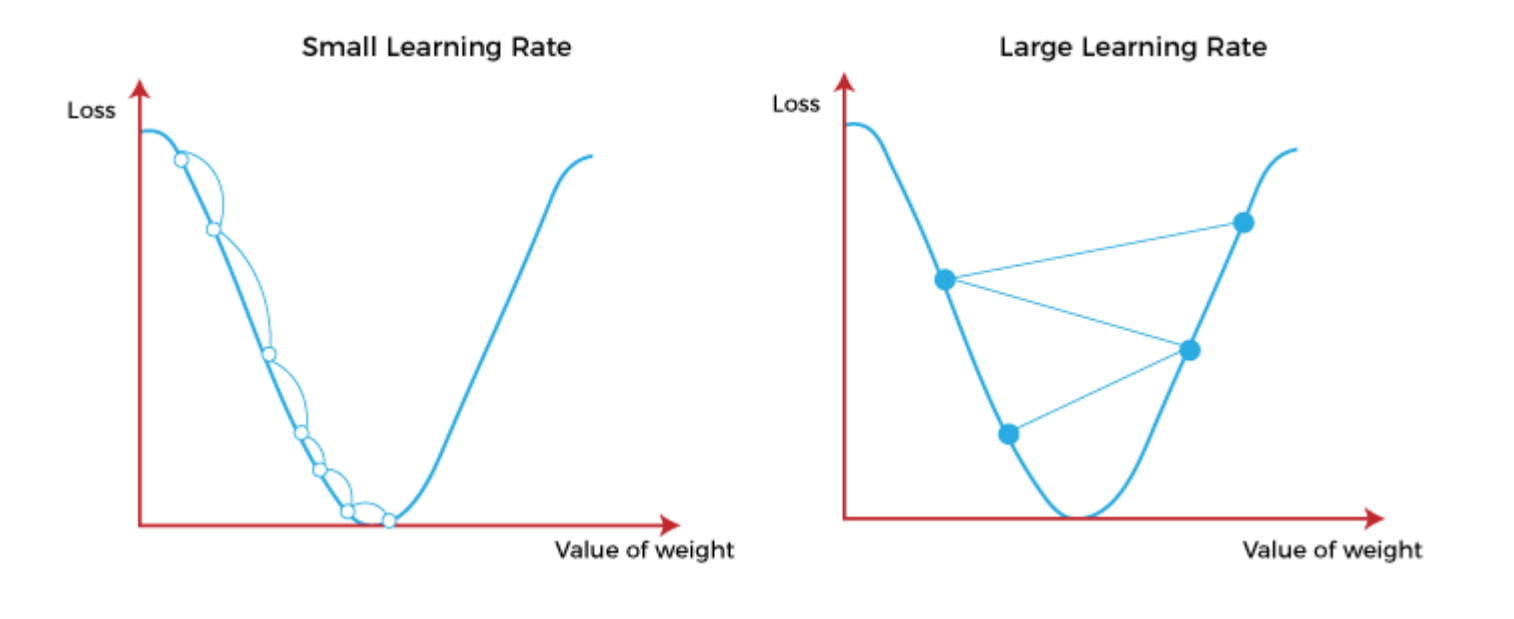
Gradient Descent is an optimization algorithms used to train machine learning models by means of minimizing errors between actual and expected results. Further.

It helps in finding the local minimum of a function.



05 Gradient Decent

Learning Rate: It is defined as the step size taken to reach the minimum or lowest point. If the learning rate is high, it results in larger steps but also leads to risks of overshooting the minimum. At the same time, a low learning rate shows the small step sizes, which compromises overall efficiency but gives the advantage of more precision.



05 Gradient Decent Types

Stochastic

It divides the training datasets into small batch sizes then performs the updates on those batches separately.

As it requires only one training example at a time, hence it is easier to store in allocated memory.

It is more efficient for large datasets.

Batch

It is used to find the error for each point in the training set and update the model after evaluating all training examples.

In simple words, it is a greedy approach where we must sum over all examples for each update.

It produces less noise in comparison to other gradient descent.

06 Pros and Cons of Supervised Learning

pros

Supervised learning in Machine Learning allows you to collect data or produce a data output from the previous experience

Helps you to optimize performance criteria using experience

Supervised machine learning helps you to solve various types of real-world computation problems.

Cons

Decision boundary might be overtrained if your training set which doesn’t have examples that you want to have in a class

Classifying big data can be a real challenge.

Training for supervised learning needs a lot of computation time.

Thank You

A023119820006

bhrigu.patel@s.amity.edu