**HOME LEARNING TASK ON MACHINE LEARNING**

Select one or more choices from the list of common Machine Learning Algorithms, do some investigations, and write me a short summary. I am looking for the following:

•Is it Supervised/Unsupervised/Reinforcement learning?

•What does the algorithm do?

•In which situations will it be most useful?

•(Optional) Can you find any examples of where this algorithm has been used?

ANSWER

Machine Learning is**the process of using mathematical models of data to help a computer learn without direct instruction.** It’s considered a subset of artificial intelligence (AI).

It focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Machine learning is an important component of the growing field of data science. Using statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision-making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

There are different types of Machine Learning Algorithms namely;

* Linear Regression, Logistic Regression, Decision Tree, SVM (Support Vector Machine), Naive Bayes, KNN (K-Nearest Neighbours), K-Means, Random Forest

This paper would deal on Linear Regression as a Machine Learning Algorithm tool.

Linear Regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc. A linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression.

It is a supervised learning algorithm method, and it Is used for classification and making predictions in machine learning. The linear regression model is best used to display the relationship between variables utilizing a sloped straight line.

Mathematically, we can represent a linear regression as:

y= a0+a1x+ ε

**Here,**

Y= Dependent Variable (Target Variable)  
X= Independent Variable (predictor Variable)  
a0= intercept of the line (Gives an additional degree of freedom)  
a1 = Linear regression coefficient (scale factor to each input value).  
ε = random error

The values for x and y variables are training datasets for Linear Regression model representation.

Types of Linear Regression

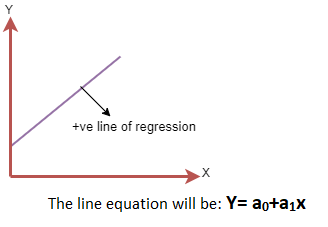
Linear regression can be further divided into two types of the algorithm:

* **Simple Linear Regression:**  
  If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.
* **Multiple Linear regression:**  
  If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

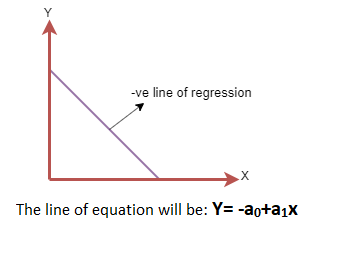
Linear Regression Line

A linear line showing the relationship between the dependent and independent variables is called a **regression line**. A regression line can show two types of relationships:

* **Positive Linear Relationship:**  
  If the dependent variable increases on the Y-axis and the independent variable increases on X-axis, then such a relationship is termed as a positive linear relationship.



* **Negative Linear Relationship:**  
  If the dependent variable decreases on the Y-axis and the independent variable increases on the X-axis, then such a relationship is called a negative linear relationship.



**Finding the best fit line:**

When working with linear regression, our main goal is to find the best fit line which means the error between predicted values and actual values should be minimized. The best fit line will have the least error.

The different values for weights or the coefficient of lines (a0, a1) give a different line of regression, so we need to calculate the best values for a0 and a1 to find the best fit line, so to calculate this we use the cost function.

For Linear Regression, we use the **Mean Squared Error (MSE)** cost function, which is the average of squared error that occurred between the predicted values and actual values.