Date:- 8th October, 2022

MACHINE LEARING (DAY-1)

Introduction to Machine Learning:-

Agenda:-

1) Machine Learning Introduction
2) AI vs ML vs DL vs DS
3) Simple Linear Regression ⟶ Mathematical Introduction

1) AI vs ML vs DL vs DS :-

a) What is Artificial Intelligence (A.I):-

⇒ A.I is Creating an Application where it Perform all its tasks without any Human Intervention.

✱) Example of Some A.I Applications are :-

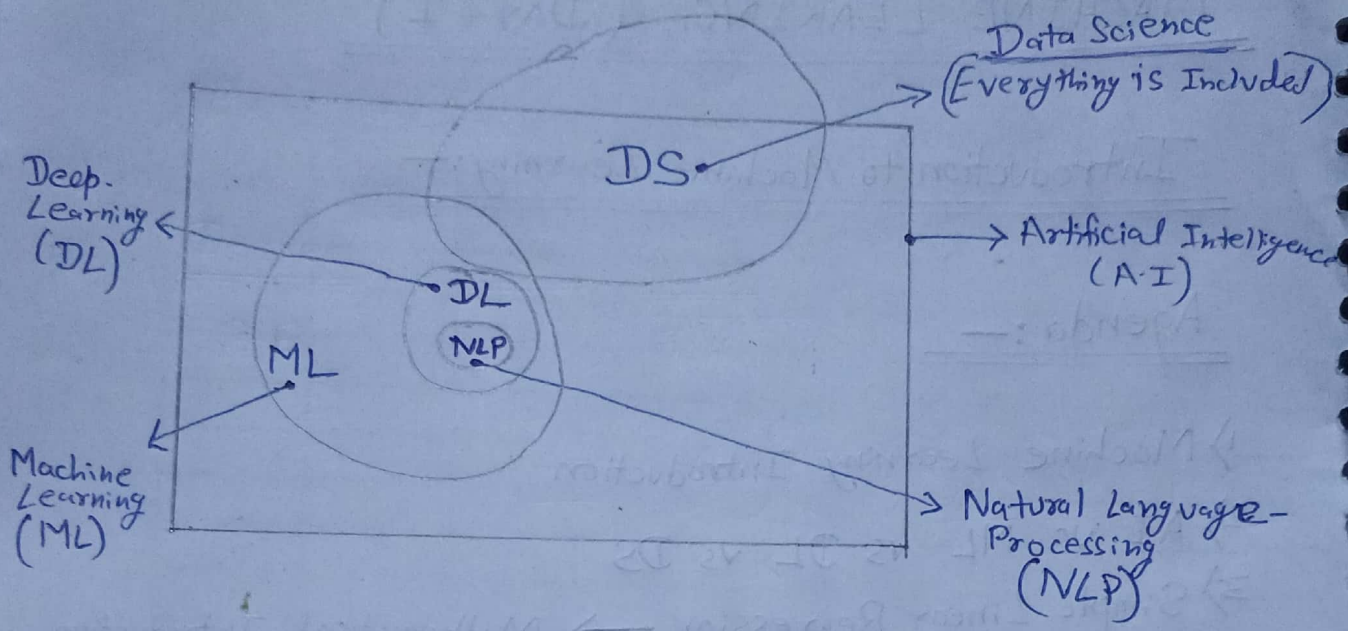
i) NetFlix App :- It is an Streaming Platform with an AI Recommentation System.

ii) Amazon.in :- Recommendation System.

iii) Chat Box :- AI Chat Box

iv) Alexa :- Speech Recognisation System.

a) Graphical Representation of A.I vs ML vs DL vs DS.



b) What is Machine Learning (ML) ?:-

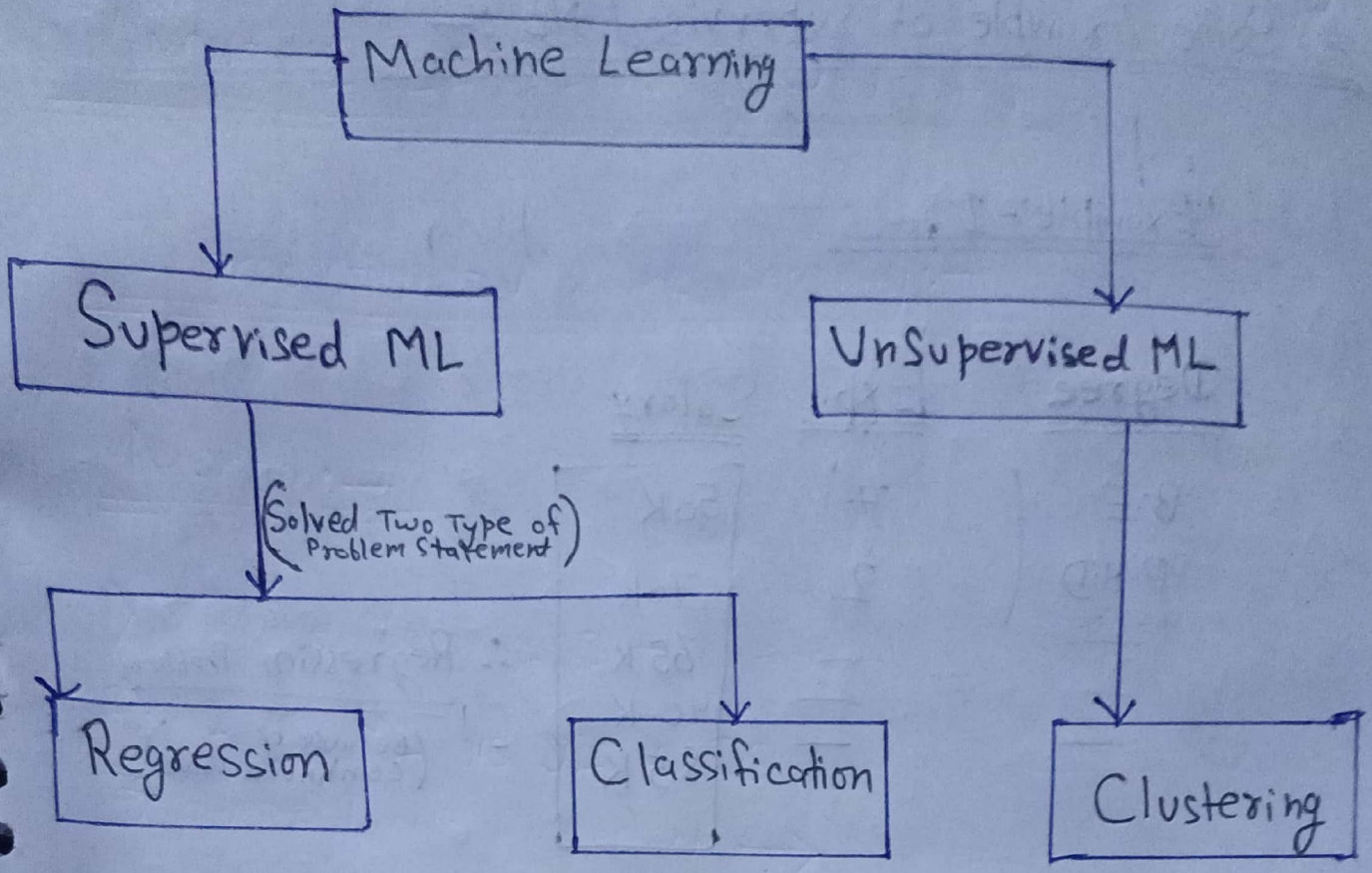
⇒ ML provide Stats tools to analyze, visualize, perform prediction and other task with the help of Data.

Note:- ML is a Subset of A.I.

c) What is Deep Learning (DL):-

⇒ Deep Learning (DL) is a Subset of Machine Learning Concerned with Algorithm inspired by the Structured and function of the brain called Artificial Neural Network.

Note:- DL is invented in 1950's and it is created to mimic Human Brain.



1) Linear Regression
2) Polynomial
3) SVR
4) Decision Tree
5) Random Forest
6) XG-Boost
7) Naive Bayes
8) KNN

① Logistic Regression
② SVM
③ Decision Tree
④ Random Forest
⑤ Naive Bayes
⑥ KNN

① DB Scan
② KMean
③ Hierarical

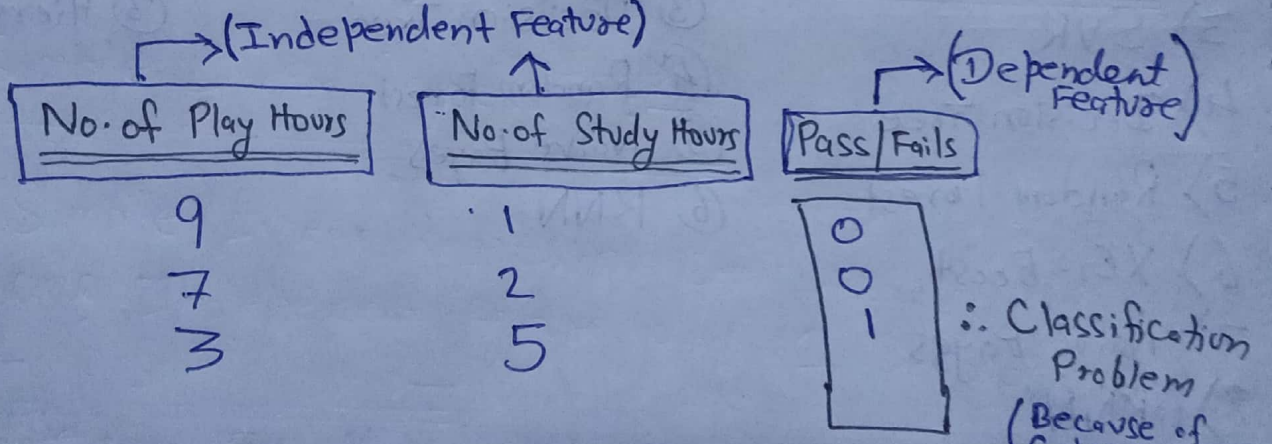
Some Example of Supervised ML :-

Example - 1 :-

Degree	Exp.	Salary
B.E	7	50K
P.H.D	2	70K
—	—	65K
—	—	90K
—	—	85K

∴ Regression Problem
(Because of Continuous Value)

Example - 2 :-



∴ Independent Feature means Input Feature

∴ Dependent Feature means ~~Output~~ Output Feature.

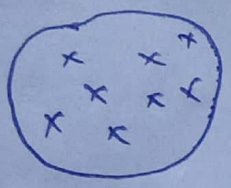
A) Some Example of Un-Supervised ML :—

Example

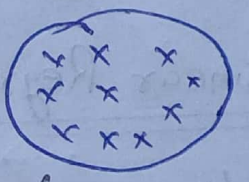
Age	Salary	Spending-Score (1-10)
24	70K	1
26	100K	9
—	—	—
21	20K	9
25	120K	2

Now,

Clustering the Data :—



(Earn More Spend More)



(Earn More Spend Less)

Here, we are Trying to create Clusters group so that we can Decide how much Percentage of Discount can be given to what Person based on their Salary & Spending.

Hence, After Doing this Clustering group this Sale will increase to nearly by 20%.

This Scenario is Called Customer Segmentation.

*) Identify the Prediction Methods for Every Problem Statement:

1) Flight Price Prediction → Regression
2) Algerian Fire Forest Prediction → Classification
3) Air Quality Index → Regression
4) Tommorrow will Rain / Not → Classification
5) Buy Day of the Person → Classification.

*) First Machine Learning Algorithm :-

1) Simple Linear Regression :-

⇒ Simple Linear Regression considers of only one Independent Feature and 1 Dependent Feature.

Note :- For Multi - Regression :- Its Considers of many Independent Feature and one Dependent Feature.

*) Some Example of Simple Linear Regression are :—

Example-1 :—

Dataset

Height	Weight

Aim :— We need to create a Model.

Input = Height , Predict = Weight

Example-2 :—

Dataset

No. of Rooms	Price
X-Feature	Y-Feature

Observation :—

If No. of Rooms ↑ , then Price of House also ↑

P.T.O

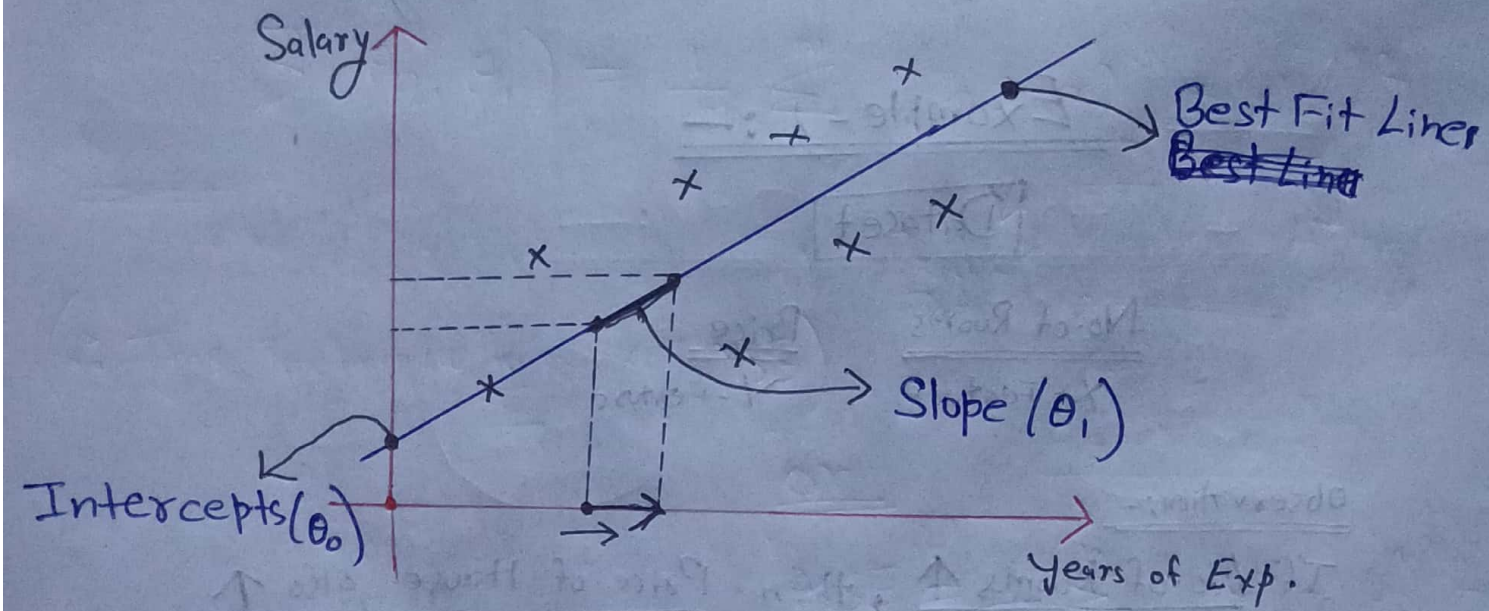
Example-3:-

Training Dataset :-

Year of Exp.	Salary

Aim:- Years of Experience and Salary.

Predicts:- Salary Based on Input years.



Note:- Best Fit Line:- Best Fit Line Which Will gave us Predicted Line. We need to find a way, so that these point has less Distance.

* **Best Fit Line** $\Rightarrow$ (Equation of a Stright Line)

$y = mx + c$ or $y = B_0 + B_1 X_1$

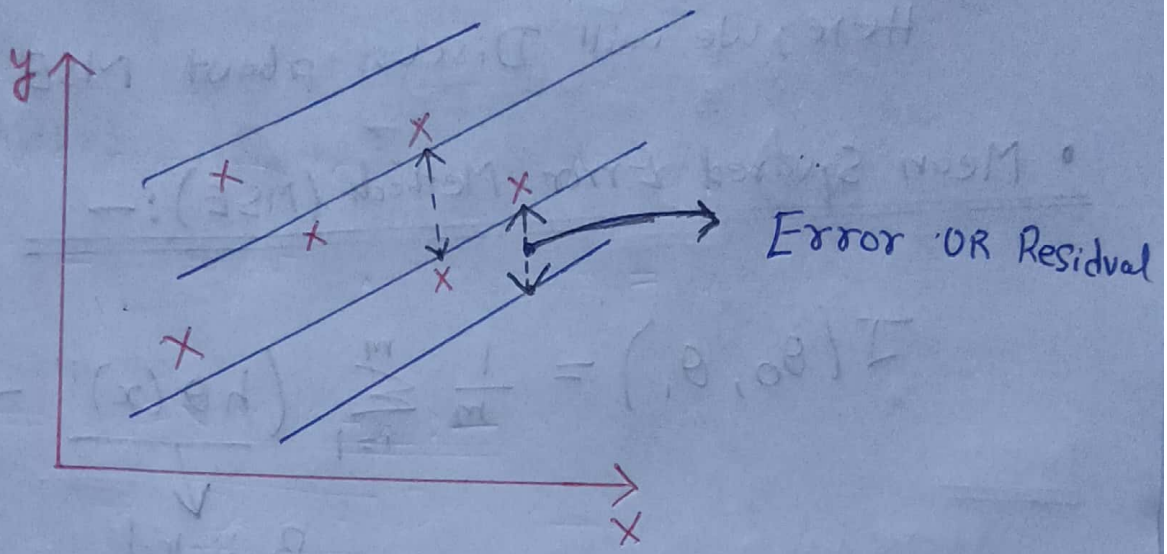
$$H_0(x) = \theta_0 + \theta_1 x \Rightarrow (\text{From Research Paper})$$

Intercepts (value of y with respect to x)

Slope $\Rightarrow$ Slope is the x-axis. What is the movement of in

✱ Training the Model :-

⇒ If we want to change the line then our intercept (θ_0) and Slope (θ_1) also change. To Change the line for the best Fit this Process is called Training the Model.



Note :- Error OR Residual ⇒ Error is a Different between Predicted Value and Actual Value.

For. Best Prediction, We have to Minimize the Error and it can be Done by COST Functions :-

✱ What is Cost Function :-

⇒ A Cost Function is an Important Parameter that Determines how well an Machine Learning Model Performs for a given Dataset.

∴ Its Calculated the Different Between the Expected Value and Predicted Value and Represents it as a Single real Number.

*) Cost Functions are Calculated by Various Methods.

1) MSE = Mean Squared Error
2) MAE = Mean Absolute Error
3) RMSE = Root Mean Squared Error.

Now,

Here, We will Discuss about MSE Methods :-

• Mean Squared Error Methods (MSE) :-

$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h\theta(x)^{(i)} - (y)^{(i)} \right)^2$$

$h\theta(x)^{(i)}$ → Predicted Data Point

$(y)^{(i)}$ → Actual Point

(∵ This Technique is called MSE)

or

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y_i} \right)^2$$

Where,

MSE = Mean Squared Error
n = Number of Data Point
y_i = Observed Value
$\hat{y_i}$ = Predicted Value

P.T.O

Final Aim :-

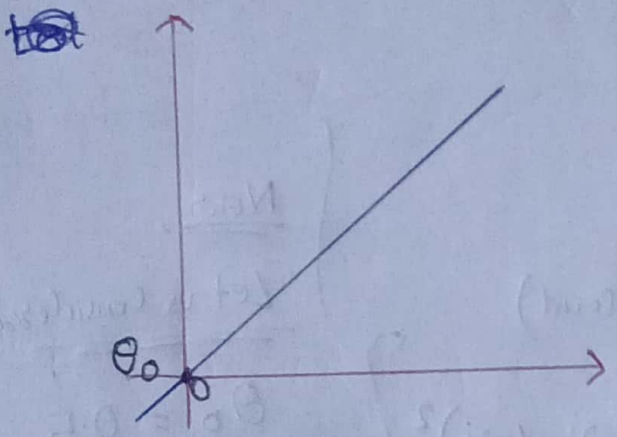
Minimize $J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - (y)^{(i)} \right)^2$

θ_0 ↓ Intercept, θ_1 ↘ Slope

$\left(h_\theta(x)^{(i)} - (y)^{(i)} \right)^2$ → MSE

Now,

$h\theta(x) = \theta_0 + \theta_1 x$



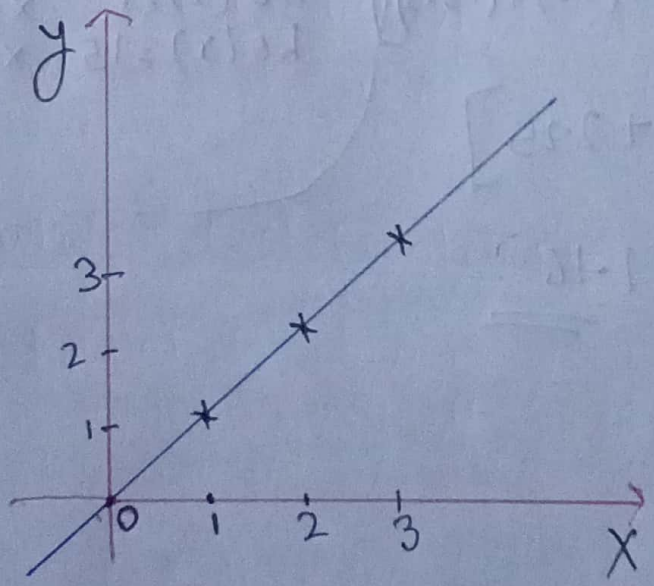
Now,

Let us Consider :-

$\theta_0 = 0$

So,

$h\theta(x) = \theta_1 x$



Now,

Training Dataset

X	Y
1	1
2	2
3	3

Here →

$\theta_0 = 1$

$h\theta(x) = \theta_1 x$

$h\theta(x) = 1 \quad x = 1$
$h\theta(x) = 2 \quad x = 2$
$h\theta(x) = 3 \quad x = 3$

Now,

Here, $m = 3$ (No. of Data Point)

$$\therefore J(\theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h\theta(x)^i - (y)^i\right)^2$$

$$\Rightarrow \frac{1}{3} \left[(1-1) + (2-2) + (3-3)\right]^2$$

$$\Rightarrow \frac{1}{3} \left[0+0+0\right]^2$$

$$\Rightarrow 0$$

Now,

Here, $m = 3$ (No. of Data Point)

$$\therefore J(\theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h\theta(x)^i - (y)^i\right)^2$$

$$= \frac{1}{3} \left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$$

$$= \frac{1}{3} \left[0.25 + 1 + 2.25\right]$$

$$= \frac{3.5}{3} \Rightarrow 1.16$$

Now,

Let us Considered:-

$\theta_0 = 0.5$

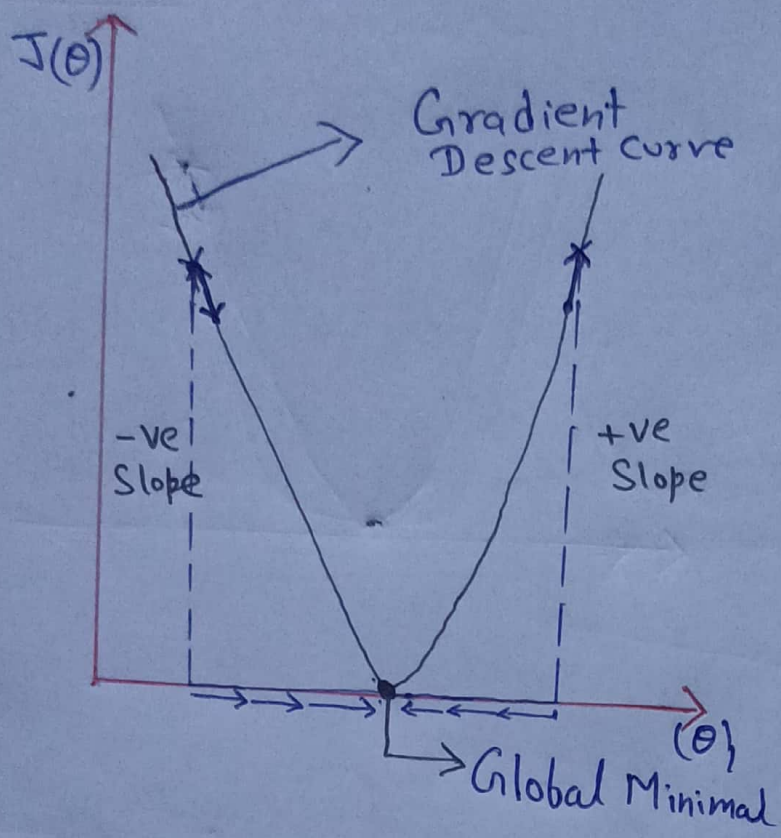
$h\theta(x) = \theta_1 x$

$h\theta(x) = 0.5 \quad x = 1$

$h\theta(x) = 1 \quad x = 2$

$h\theta(x) = 1.5 \quad x = 3$

#) Convergence Algorithm :— (Optimize the Change of θ_1 Value.)



Note :—

Gradient Descent is an Optimization Algorithm. Used for ~~Minimizing the~~ Minimizing the cost Function which is useful in ML Algorithm. It is basically used for Updating the Parameter of the Learning Model.

Repeat Until Convergence :—

Learning Rate

$$\theta_j = \theta_j - \boxed{\alpha}\left[\frac{\partial}{\partial \theta_j} J(\theta_j)\right]$$

Derivative Slope

- Main Aim is to Come over Global Minimal Because there Different is very less.
- ∴ α (Learning Rate) Decides the Speed of convergence. If α is small then it will take time to come at Global Minimal.
- ∴ If α is very large it will Jump here and there and hardly will come at Global Minimal.

So, value should be $\alpha = 0.001$

∴ at Global Minimal, Slope = 0.

Note :— For -ve Slope :—

$$\theta_j = \theta_j - \alpha\,(-ve)$$
$$= \theta_j + \alpha$$

Note :— For +ve Slope :—

$$\theta_j = \theta_j - \alpha\,(+ve)$$
$$= \theta_j -$$