Introduction to Machine Learning

16/04/23

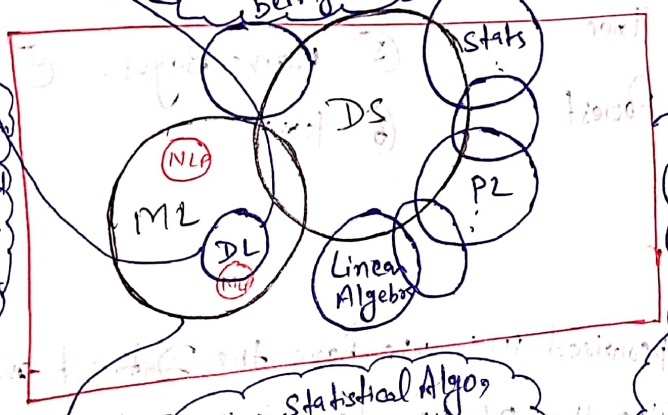
Agenda

1. Machine Learning Introduction
2. AI Vs ML Vs DL Vs DS
3. Simple Linear Regression.

1. AI Vs ML Vs DL Vs DS

Intention behind D.L
1950s Scientist started thinking can Machine learn how we human being learn.

D.L they started with Perceptron but now multi layered Neural N/w is part of D.L



Streaming Platform, S/w Product

AI Application → Netflix ← Recommend-ation System.

Seperate AI Module Integrated into it.

A.I (Artificial Intelligence)
It is creating an application where it performs all its task without human Intervention.
Eg:- Amazon.in, Google Assistant, Alexa, Chat Bot, Self driving car, YT recommendations

ML Provides stats tools to analyze, visualize & perform prediction and other task with the help of data.
M.L is subset of AI.

Statistical Algo, tools etc..

NLP:- NLP is a technique which Works with text data.
It falls in Both the catogries ML & DL..

CV (Computer Vision):- It's CV also falls in both ML & DL.
these are Some of the technique which we use to Solve Some of the problems.

DS (Data Science) :- W.r.t DS everything is involved like DL, ML, NLP, CV etc and also need to learn technology like stats, Programming lang., Linear Algebra etc.

Big Data :→ Used for collecting and storing the data efficiently. DS's completely a different field.

ML and DL

ML and DL

- Supervised
  - Regression
    1. Linear Regression
    2. Polynomial Regression
    3. SVR
    4. Decision Tree
    5. Random Forest
    6. Xgboost
    7. KNN
  - Classification
    1. Logistic Regression
    2. SVM
    3. Decision tree
    4. Random Forest
    5. Naive Bayes
    6. KNN
- Unsupervised
  - Clustering Algorithm
    1. DB Scan
    2. K-Means
    3. Hierarchical Mean Clustering
    4. Silhouette Scoring.

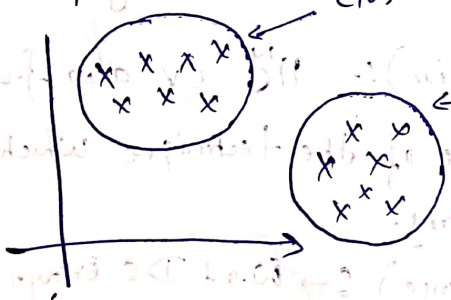
(*) In case of Supervised M.L we have the Dataset and we will be knowing the o/p or target variable.

(*) In case of unsupervised M.L; We never be knowing the o/p. only features are given like

f_1 f_2 f_3 f_4 ...

here we try to find out the similar data points in the form of cluster.

eg:-



Here we don't know the specific o/p

Difference between Supervised and Unsupervised M.L

Supervised M.L

Q: Predict the Salary based on Degree and Exp.

Predict

Degree	Exp	Salary
B.E	7	50k
PHD	2	70k
-	-	65k
-	-	66k
-	-	68k

Regression

Q Why is it a Supervised M.L?

→ Since we have our o/p feature i.e Salary to predict.

Q Salary is continious or catogrical feature.

Ans Continious feature.

Since it can be 80.5, 20.7 etc.

Hence it is a Regression Problem.

Another Example: →

Independent Feature

dependent Feature

No. of Play hours	No. of Study hrs.	Pass / Fail
9	1	0 ← Fail
7	2	0
3	5	1 ← Pass

Since it is completely dependent on other two features based on that its value will change.

Q Whether this a classification Problem or Regression Problem?

Ans Classification Problem.

Since o/p is having only two catogries: Pass (1), fail (0)

Q. "Price of the house" is Regression Problem or classification Problem?

Ans. Regression Problem.

(*) Whenever the o/p feature is a continious value then it becomes a Regression Problem and

(*) Whenever the o/p feature is a catogorical value then it becomes a classification Problem.

→ In case of Unsupervised learning we don't have any dependent feature. We consider every feature as Indepand feature and we try to group it.

Q. "Flight Price Prediction" is a Regression Problem or Classification Problem?

Ans. Regression Problem.

Q. (*) Algerian Fire Forest → classification Problem. (Fire is there / Fire is not there)

(*) AIR Quality Index → Regression Problem

(*) Tommorow Rain/Not → classification Problem.

(*) Buy day of the Person → classification Problem (Mon, Tue --- Sun)

Difference b/w dependent & Independent features.

Ans

I/P feature → Independent feature (Eg:- Degree, Exp)

O/P feature → Dependent feature (Eg:- Salary).

Unsupervised Machine Learning :→ There is no dependent feature.

Customer Segmentation Problem :→

	Age	Salary	Spending-Score (1-9)	
1.	24	70k	1	← Earns more spends less
2.	26	100k	9	← Earns more spends more
	-	-	-	
	-	-	6	← Moderate. (20% discount may buy.
19.	-	20k	9	← Earns less Spends more
20.	22	120k	2	← Very stingy Person
21.	25			

Q For a Particular Product which kind of Promotion to give which Person?

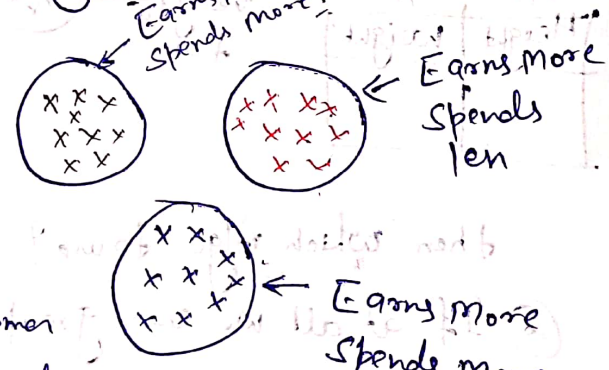
⇒ Product → 10% off → 20s! Person will buy

Product → For first Person even if you give free he won't take ← Observation.

Q. How to decide? whome to give how much offer?

So! Cluster the data.

Hence by clustering the data into different Catagories. It is easy to decide which type of Customer should be given what kind of discount.



∴ This Problem is called Customer Segmentation Problem

Unsupervised M.L :→ Here we don't have any dependent feature. He we create a Cluster groups so that we will be able to segment them together in diff. diff. groups and we will be able to apply to take the decisions.

→ By doing this there are chances of increase in sale may be 20%.. Sales 20% ↑

③ # Simple Linear Regression: →

(*) It has only 1 Independent feature and 1 dependent feature.

(*) In Case of multiple linear regression it will have many Independent feature and 1 dependent feature.

Hence most of the time while solving linear regression problem mostly, we have only 1 dependent feature.

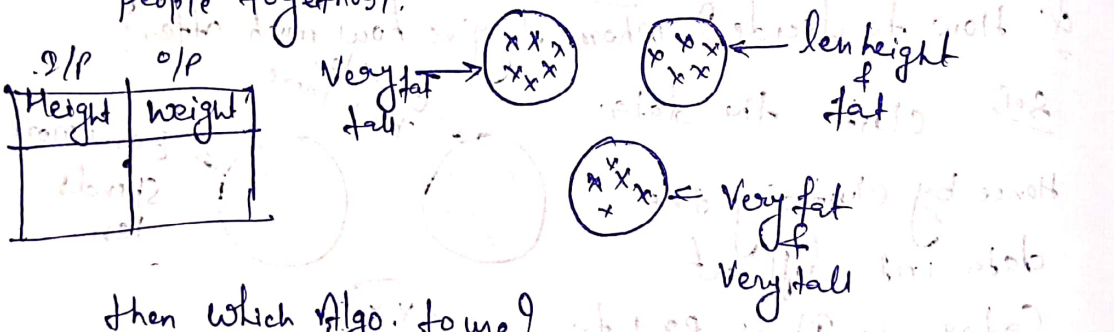
Examples of Linear Regression: →

i) Problem Statement

AIM: → To Create a Model which takes I/P as height and It should predict the weight.

Based on height and weight we need to group common people together.

I/P	O/P
Height	weight



then which Algo. to use?

(*) If at all we are grouping it then we need to apply for Clustering or unsupervised learning.

(*) But here problem is based on height we need to give the prediction for weight.

I/P → Height Predict → weight.

So here In this Case we use Regression i.e. Linear Regression.

ii) Problem Statement :- Predict the Price where No. of Rooms given

Dataset :-

X	Y ← Predict.
No. of rooms.	Price
—	—
—	—
⋮	⋮

Mostly, when No. of rooms ↑ Price ↑

No. of Rooms ↓, Price ↓ } → Example of Linear Regression.

iii) Problem Statement :- Based on 'Year of Exp' Predict the Salary.

I/P (Training Dataset) O/P — Continious.

Year of Exp	Salary ← Predict
—	—
—	—
—	—
—	—
⋮	⋮

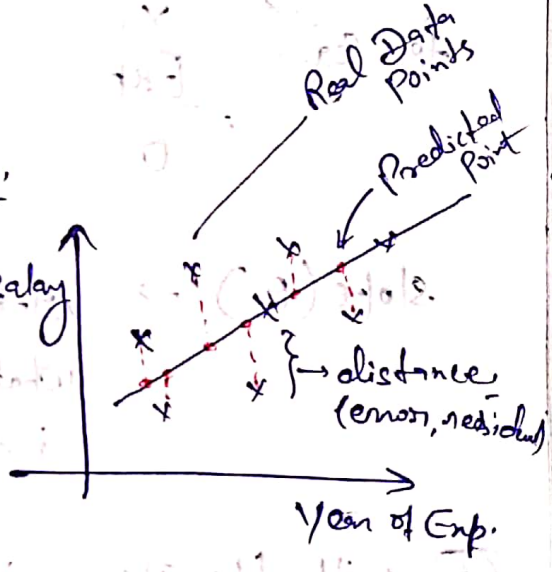
} Supervised & Linear Regression.

Model :- Year of Exp and Salary.

Predict :- Salary Based on I/P year.

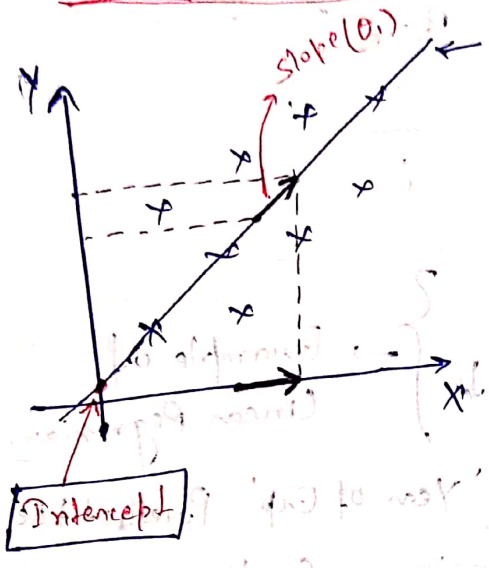
#AIM of Linear Regression :-

→ Based on training dataset it tries to find out the best fit line such that error b/w real and expected data point is minimal.



→ The difference b/w Real Point and Predicted Point is called Residuals or Errors.

Mathematical Intiduction about Linear Regression :→



Equation of straight line

$$Y = mx + c$$

or

$$Y = \beta_0 + \beta_1 x$$

or

In research Paper

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

(θ_0 ← Intercept, θ_1 ← slope)

Intercept (θ_0) → When $x = 0$ at that point the best fit line meets at y axis. that point is called as intercept.

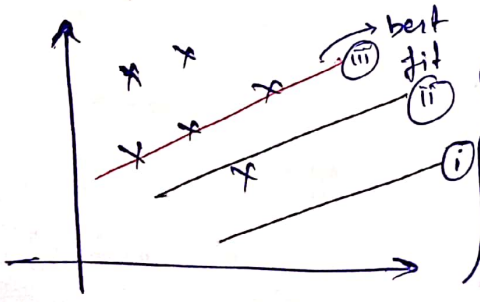
for Eg:-

X	Y
Exp	Salary
0	3.25 lakh ← The value of 'Y' when X is 0. i.e called Intercept

slope (θ_1) :→ With the unit movment in x-axis what is the unit movment y-axis's. i.e called as slope.

(*) With the change in Intercept and slope the best fit line will be moving here and there.

(*) Hence we can say that by changing the slope (θ_1) and Intercept (θ_0) value we can get the best fit line.



(i) θ_0, θ_1 → Initial Value (Initially)

(ii) θ_0, θ_1 → next Value (changed)

(iii) θ_0, θ_1 → next value (changed)

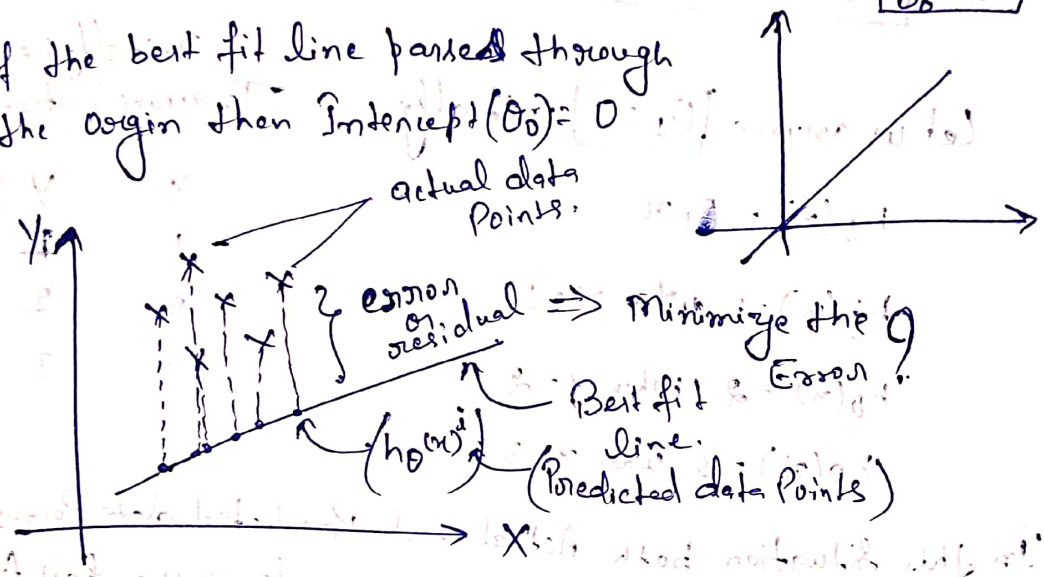
↳ TRairing of the model

Training of the Model :→

By changing the Value of θ_0 and θ_1 i.e Intercept and slope we try to get the best fit line and that process is called training of the model.

(*) If the best fit line passed through the origin then Intercept(θ_0) = 0

$\theta_0 = 0$



⇒ Minimize the Error ?

To minimize the error we have an eqn and it is called cost function.

Cost function :→ To minimize the Error.

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

- Mean Square Error
- Divided by m to find the avg
- Predicted data Point
- actual data Point

Where m = All the data Points.

(MSE)

(*) This cost function $J(\theta_0, \theta_1)$ is called Mean Square Error.

Final AIM of Linear Regression :→

To Minimize MSE

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

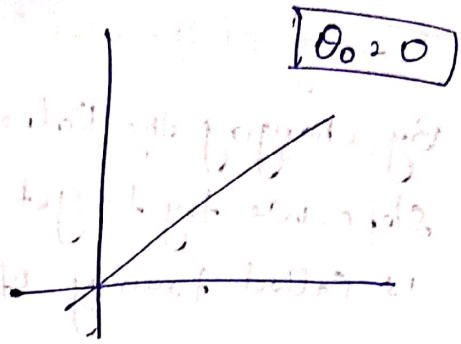
By changing θ_0 and θ_1 Value.

Let us Consider $\theta_0 = 0$

$\therefore h_\theta(x) = \theta_0 + \theta_1 x$

then,

$$\boxed{h_\theta(x) = \theta_1 x}$$



Let us assume $\boxed{\theta_1 = 1}$ initially.

Initially i.e Actual Training Dataset

X	Y
1	1
2	2
3	3

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

$h_\theta(x) = 1$ when $x = 1$

$h_\theta(x) = 2$ when $x = 2$

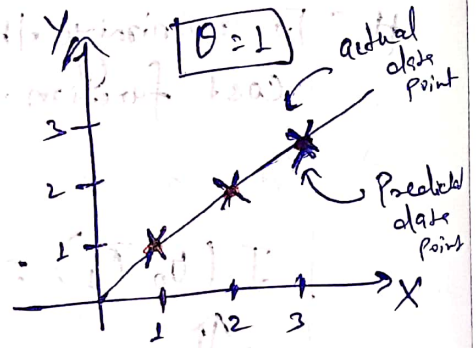
$h_\theta(x) = 3$ when $x = 3$

⇒ In this situation both Actual and Predicted data Points will die at the same point once we draw the best fit line.

Case1

In this case $\boxed{\theta_0 = 0}$, $\boxed{\theta_1 = 1}$

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



no. of Data Points i.e $\boxed{m = 3}$

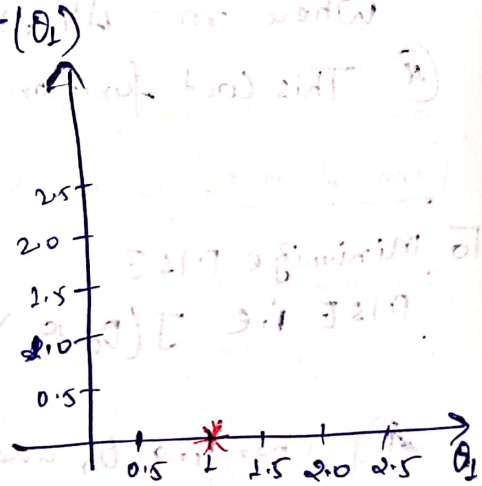
Hence,

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

$$J(\theta_1) = \boxed{0} \downarrow\downarrow\downarrow$$

when $\boxed{\theta_1 \neq 1}$, $\boxed{J(\theta_1) \neq 0}$

When $\theta_1 = 1$ then: $J(\theta_1) = 0$



Case 2

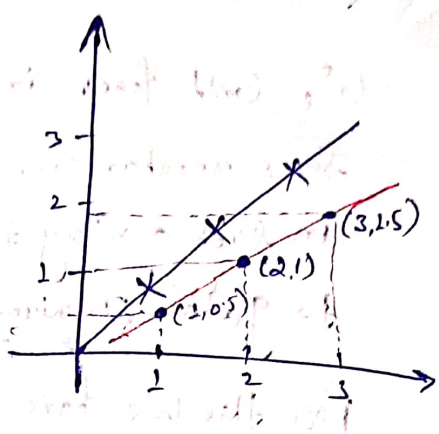
If $\theta_1 = 0.5$, $h_\theta(x) = \theta_1 x$

then,

$h_\theta(x)$	x
0.5	1
1	2
1.5	3

Predicted data Points

Actual Data Points



$$J(\theta_1) = \frac{1}{3}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$$

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

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

When $\theta_1 = 0.5$ then $J(\theta_1) = 1.16$

Case 3

If $\theta_1 = 0$, $h_\theta(x) = \theta_1 x$

$h_\theta(x)$ (Expected)	x (Actual)
0	1
0	2
0	3

$$J(\theta_1) = \frac{1}{3}\left[(0-1)^2 + (0-2)^2 + (0-3)^2\right]$$

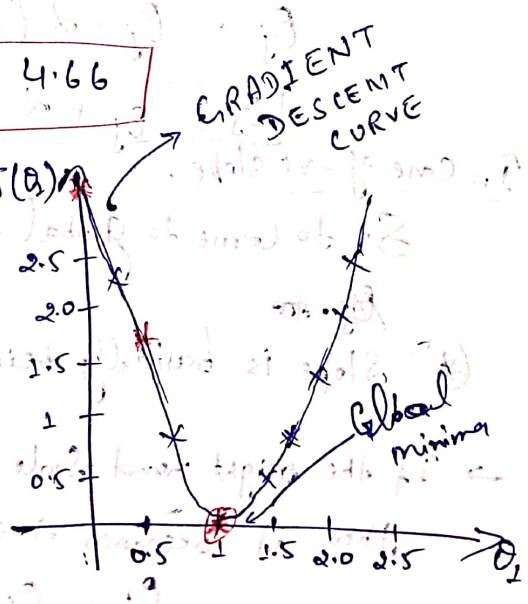
$$= \frac{1}{3}\left[1 + 4 + 9\right]$$

$$= 14/3 \approx 4.66$$

When $\theta_1 = 0$ then $J(\theta_1) = 4.66$

(*) When we keep plotting these points on we get a curve

Called: GRADIENT DECENT CURVE



(*) At global minima only minimal distance was there. Hence the goal is to reach near global minima.

Problem with this model

We can't keep initializing θ_j value there should be some mechanism. It should change the θ_j value in such a way that it should be able to reach to global minima.

For this we have sth called - Convergence Algorithm.

V.V.I

Convergence Algorithm :→ { optimizes the changes of θ_j valu

It says

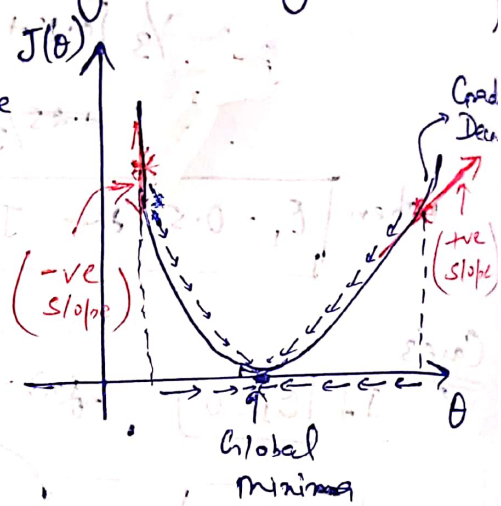
Repeat until Convergence {

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_j)$$

}

learning Rate

slope (derivative)



Q. How to know if the slope is +ve or -ve?

→ If right side of the line is facing downwards then it becomes a -ve slope.

$\theta_j \approx \theta_0, \theta_1$

$\theta_j = \theta_j - \alpha(-ve)$ ← Since slope is -ve

$\theta_j = \theta_j + \alpha$

In case of -ve slope.

So to come to global minima we need to increase the θ_j

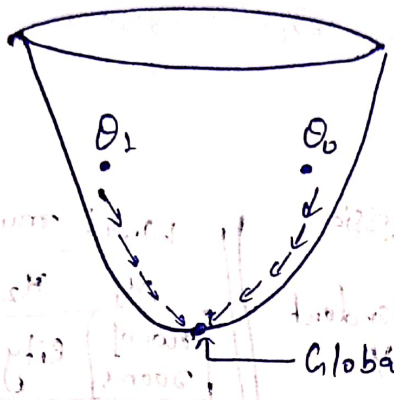
(*) Slope is basically being used to update the θ_j value

→ If the right hand side (RHS) of the is facing upwards then it becomes +ve slope ← Since slope is +ve.

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

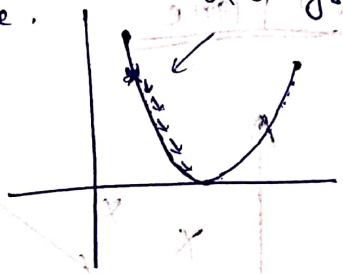
$\theta_j = \theta_j - \alpha$ ← decreasing the θ_j to Reach the global minima

3-D Diagram:-



Q What is learning rate (α)?

Learning rate decides the speed of Convergence.



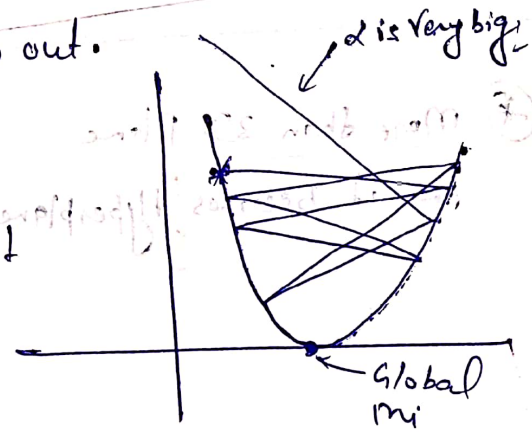
If α is very-very small:-

Suppose $\alpha = 0.0000001$

then it will take so much time to converge or we can say it will take lot of time to reach to Global minima.

If α is very-very big value

In that case it will be jumping from one corner to another corner and sometime it may go out.

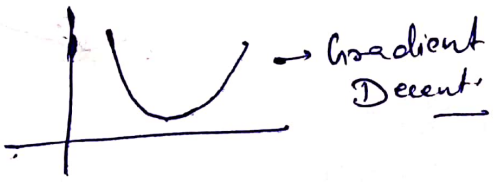


Hence we select $\alpha = 0.001$

i.e step size based on that it will converge.

(*) At Global Minima, Slope = 0

(*) Hence we can say that bez of MSE (Mean Square Error) we get Gradient Decent.



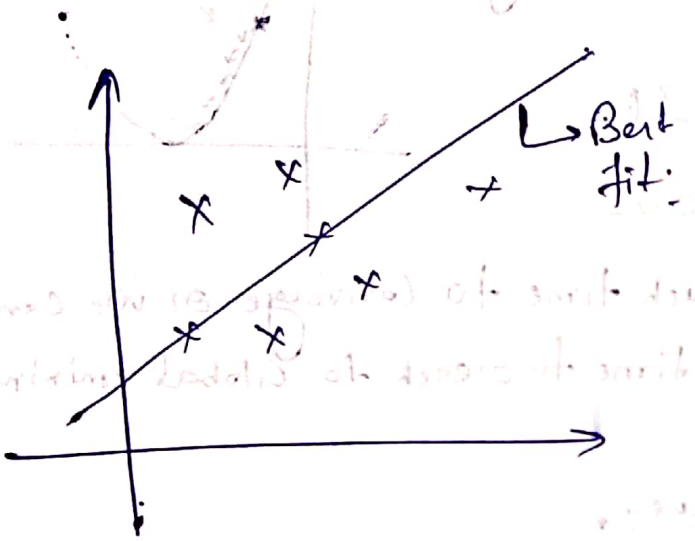
Q: What is the difference b/w Single and multiple linear regression?

Sol: W.r.t Single linear Regression

1 independent x — 1 dependent y

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

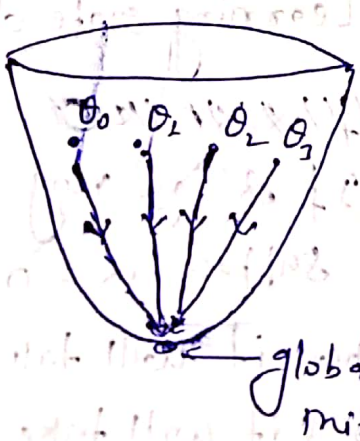
2D Plane



W.r.t multiple linear Regression

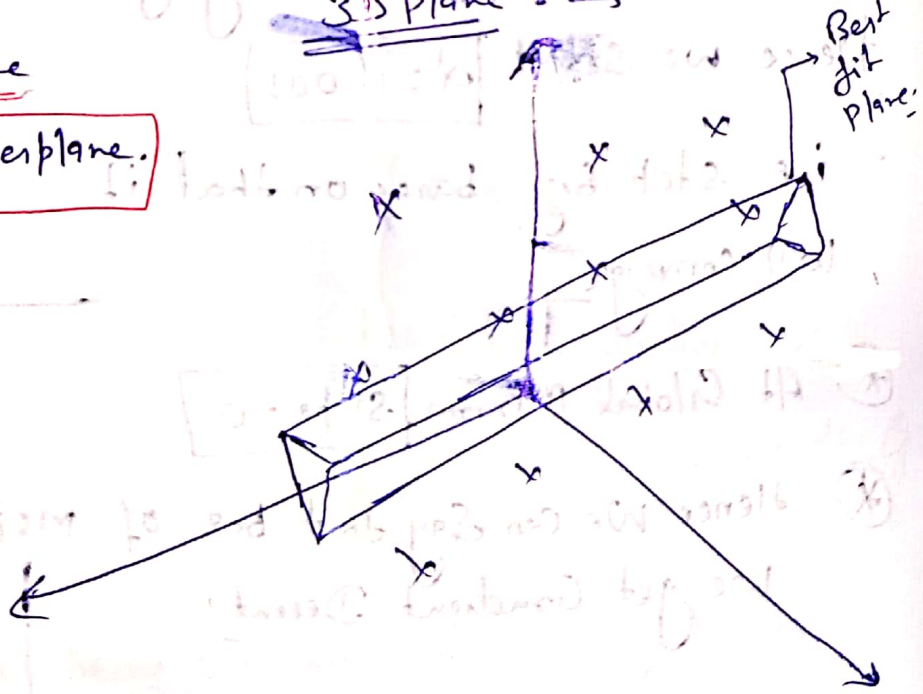
x_1	x_2	x_3	y
no. of rooms	city	Room size	Price

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$



→ 3 Independent & 1 dependent feature

3D Plane :→



(4) More than 3D Plane

then it becomes Hyperplane.