08-OCT-2022

Introduction to Machine Learning

AI vs ML vs DL vs DS

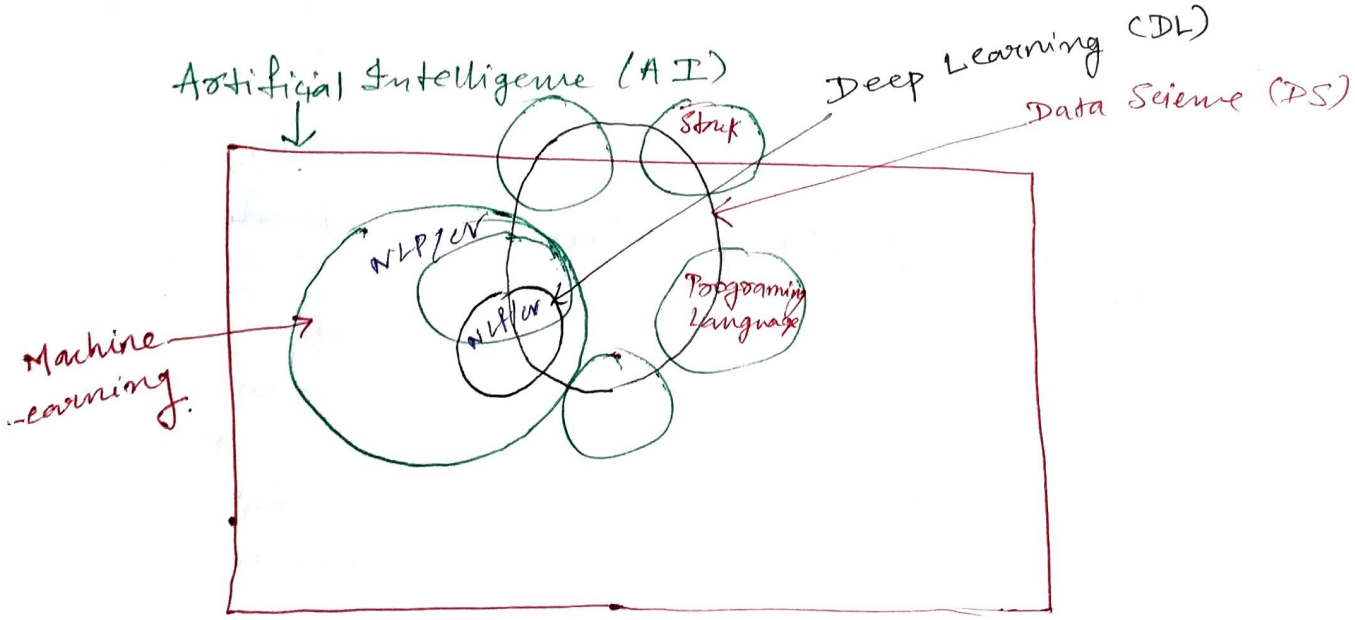
AI application

Streaming Platform

↑

Netflix → Recommendation System (AI model)

> Netflix is a software in which Recommendation system is integrated as AI application.
> At the end of the day, AI application is needed.



AI

It is defined as creating an application where it performs all its task without any human intervention.

eg: chatbox, Amazon with recommendation system, Alexa, Google Assistant, Self driving car, Youtube recommendation,

ML

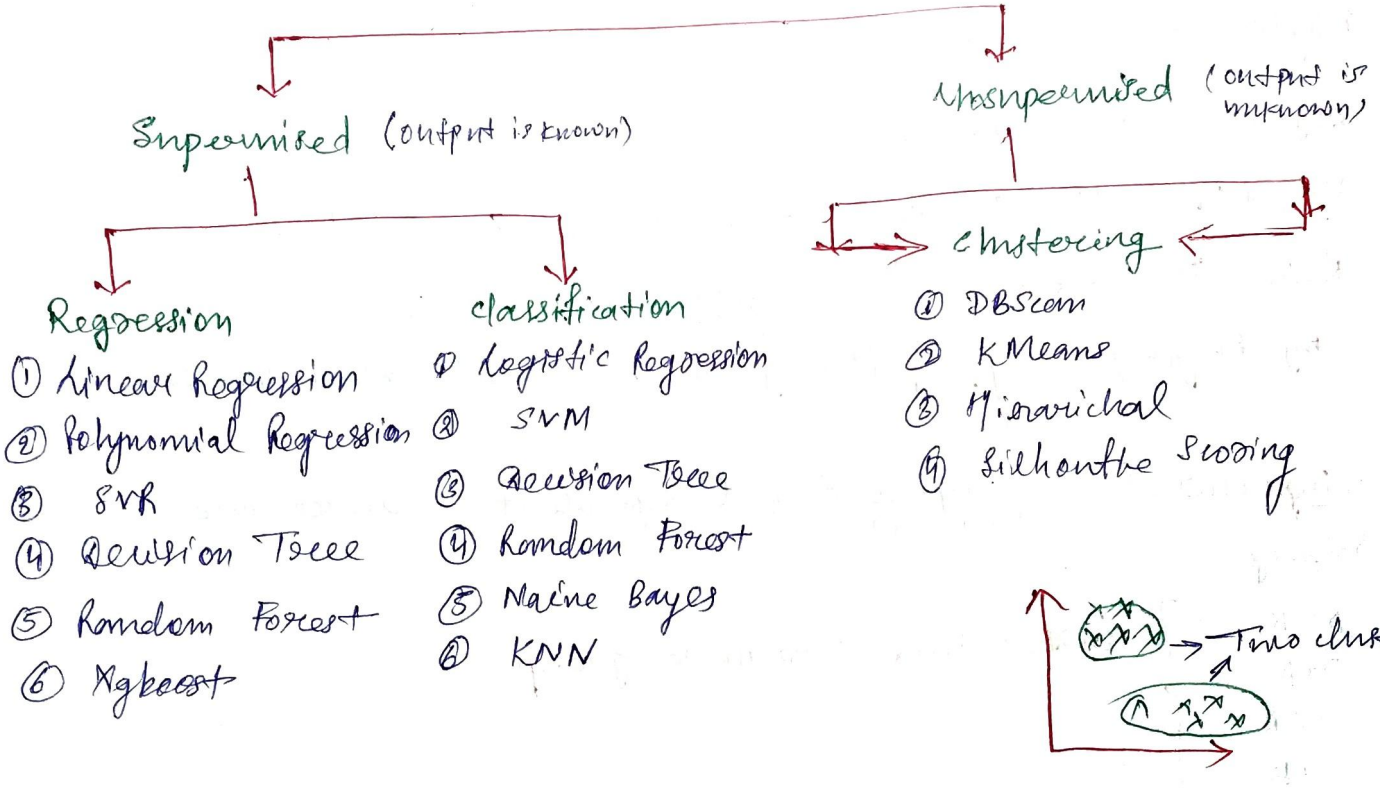
- It provides us statstools to explore, visualize, analyze and perform prediction and other task with the help of data.
- It is a subset of AI.

DL

- In 1950's, Researchers started thinking about can we make a machine learn like how we human learn. This is the intention behind DL. Initially they started with and now they use multi layer neural network to perform the task.
- It is invented to mimic the human brain.

NLP / Computer vision : these two are technique.

ML and DL



Supervised

eg Independent Features, O/P = continuous → Regression Problem

Degree	Exp	Salary → Dependent Feature
B.E	7	50K
PHD	2	70K

eg Independent Features, O/P = ~~Discrete~~ categorical → classification Problem

No. of Play hours	No. of study hours	Pass/Fail → Dependent features
9	1	0
7	2	0
3	5	1

eg Flight Price Prediction → Regression Problem

eg Algerian Fire Forest → classification Problem

eg Predict Air Quality Index → Regression Problem

eg Rain Prediction → classification Problem

eg By Buying day of the person → classification problem

Unsupervised (Independent and dependent features are known)

Salary

80K
70K
80K
90K
100K

} cluster into 2 or more group

Note: Independent features is basically input features.

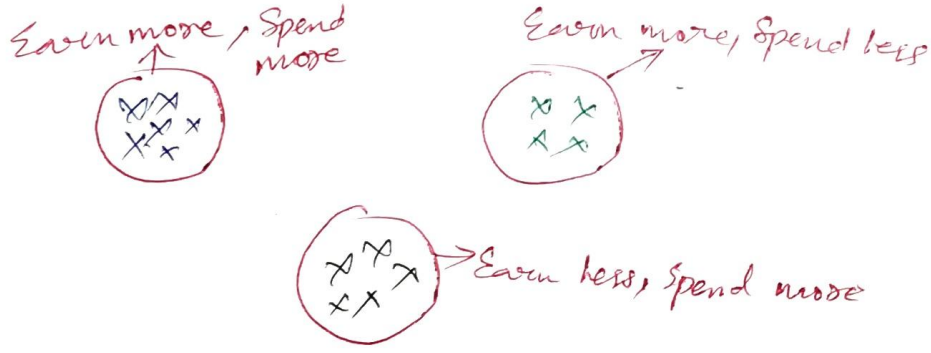
~~Simple Linear Regression~~

Customer Segmentation

Age	Salary	Spending-Score (1-10)
24	70k	1
26	100k	9
25	60K	6
21	20K	9
25	120K	2

Product
Online Promotions
No discount
10% discount
20% discount
10% discount

⇒ We can make cluster (group) based on product owner requirement



- No dependent feature.
- Sale can be increased by doing all this steps.

eg Titanic dataset (Person survived or not) ⇒ classification problem
Iris dataset (Flower belong to which category) ⇒ classification problem

Simple Linear Regression

eg [1 independent feature and 1 dependent feature]

< AIM: To create a model, which takes input as height and predict weight. >

DATASET

Height	Weight

eg AIM: Based on the no. of rooms, predict price.

Dataset

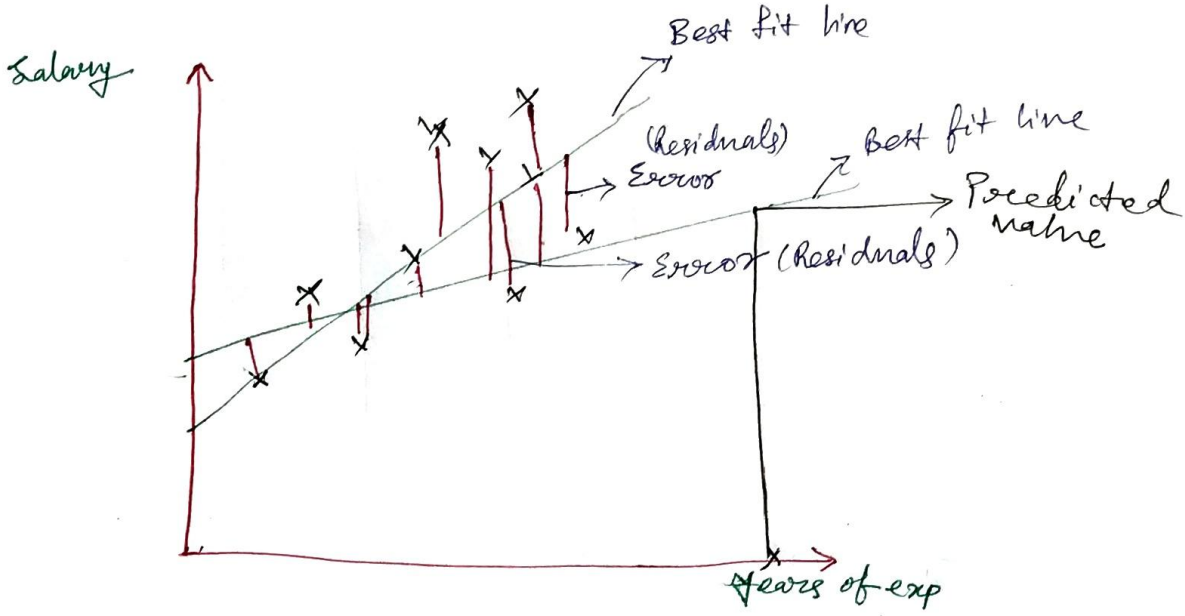
No. of Rooms (X)	Price (Y)

eg AIM: Based on the year of exp, predict salary.

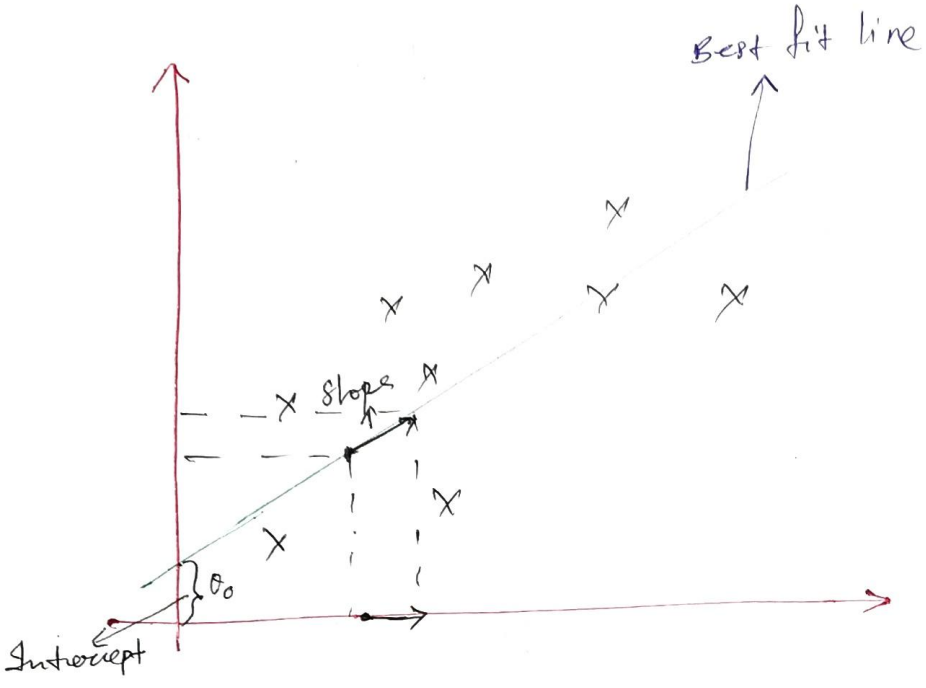
Dataset

Training dataset

Year of Exp	Salary
–	–
–	–
–	–
–	–
–	–
–	–



→ Based on the training dataset, it finds the best fit line in such a way that the sum of difference b/w real points and predicted should be minimal.



Equation of a straight line,

$y = mx + c$

$y = \beta_0 + \beta_1 x_1$

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

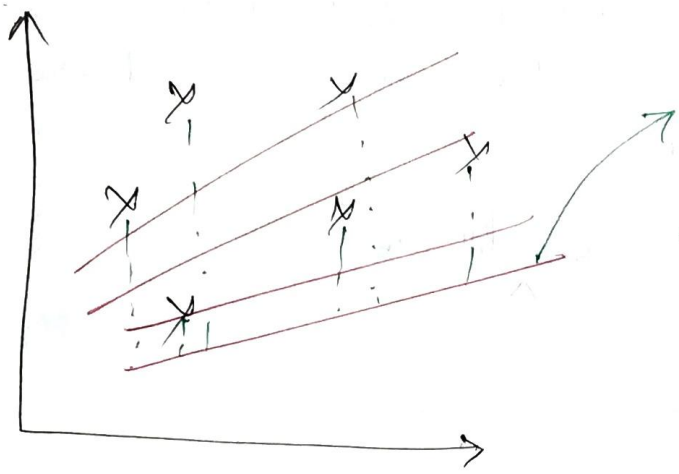
} All are same.

θ_0 = Intercept (value of y for $x = 0$)

Exp	Salary
0	3.25 lakhs

θ_1 = Slope (with the unit movement in the x-axis, what is the unit movement in the y-axis)

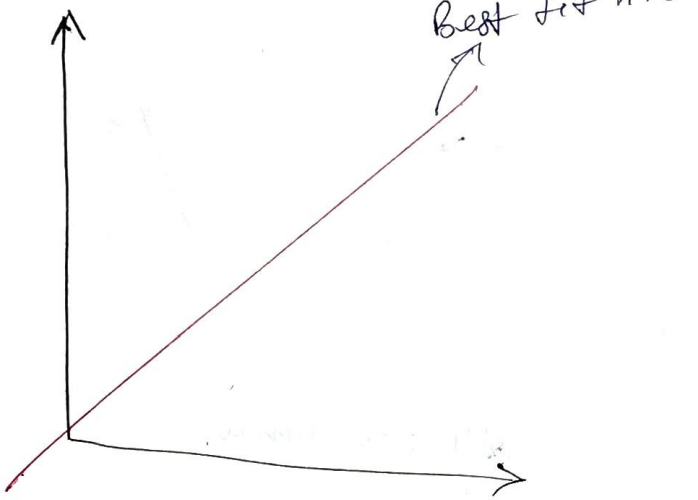
→ By changing θ_0 and θ_1, ~~bet~~ best fit line will be change.



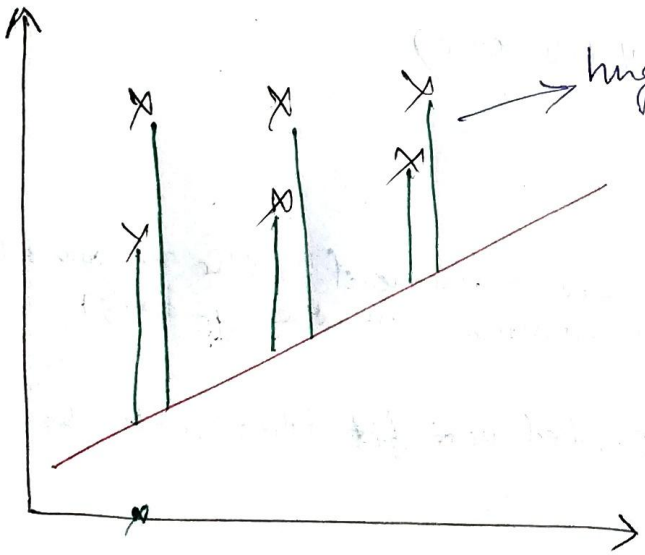
Error is big

θ_0, θ_1 → value

Training of the model change θ_0, θ_1 to get the best fit line



$\theta_0 = 0$ (Intercept)



→ huge error

Do we need to minimize the error?

→ Yes, we need to minimize the error using cost function.

Cost function

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

← Mean Squared Error ↓ one of the cost function.

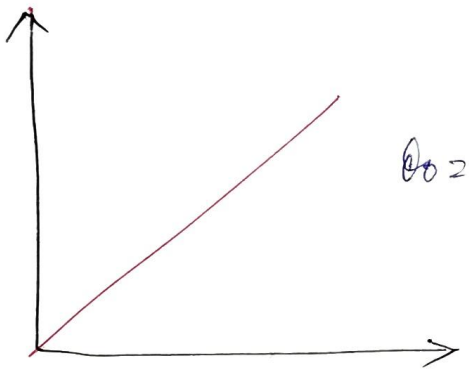
- Minimize it to get the best fit line.
- Predicted
- Actual.

- Dividing by m to get the average (mean).

Final aim in linear regression

Minimize, $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$ by changing θ_0 & θ_1.

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



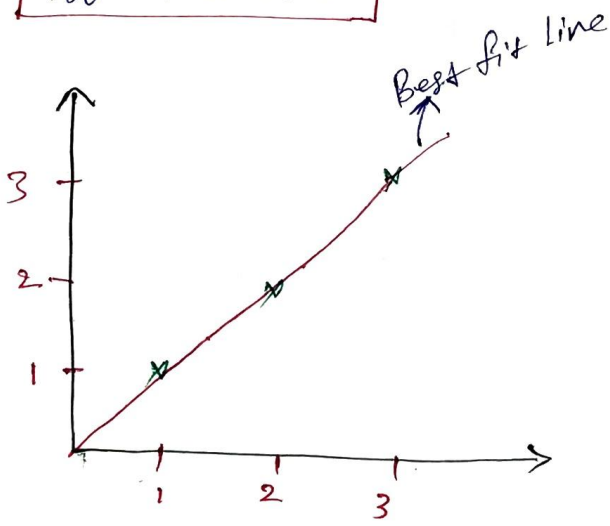
$\theta_0 = 0$

Training Dataset

x	y
1	1
2	2
3	3

Let us consider $\theta_0 = 0$

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



Let's assume $\boxed{\theta_1 = 1}$

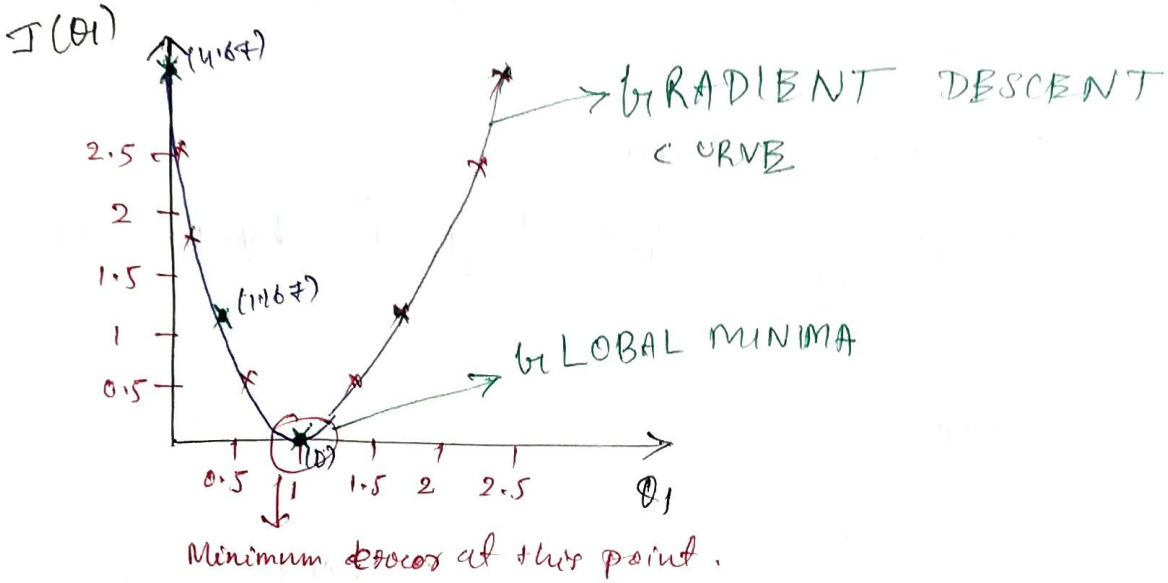
$h_\theta(x)$	x
1	1
2	2
3	3

$$J(\theta_1) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 = \frac{1}{3}\{(1-1)^2 + (2-2)^2 + (3-3)^2\}$$

$$= \frac{1}{3}\{0 + 0 + 0\}$$

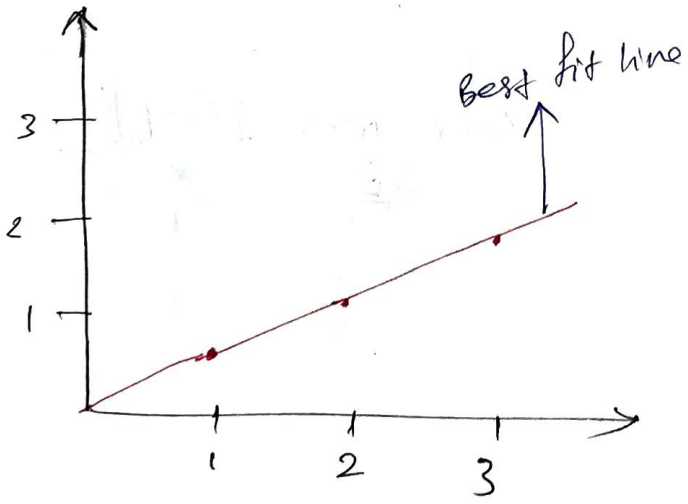
$$= 0$$

$J(\theta_1)$ VS θ_1



Let's assume - $\theta_1 = 0.5$

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



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

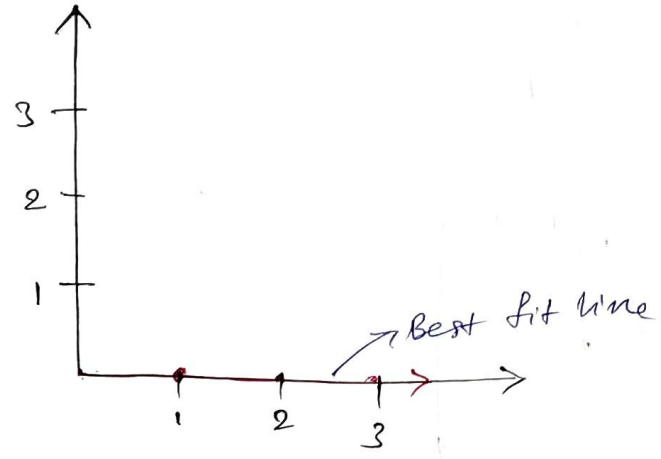
$$= \frac{1}{3} \{ (0.5-1)^2 + (1-2)^2 + (1.5-3)^2 \}$$

$$= \frac{1}{3} \{ 0.25 + 1 + 2.25 \}$$

$$= 1.167$$

Let's assume, $\boxed{\theta_1 = 0}$

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



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

$$= \frac{1}{3} \{ (0-1)^2 + (0-2)^2 + (0-3)^2 \}$$

$$= \frac{1}{3} \{ 1 + 4 + 9 \}$$

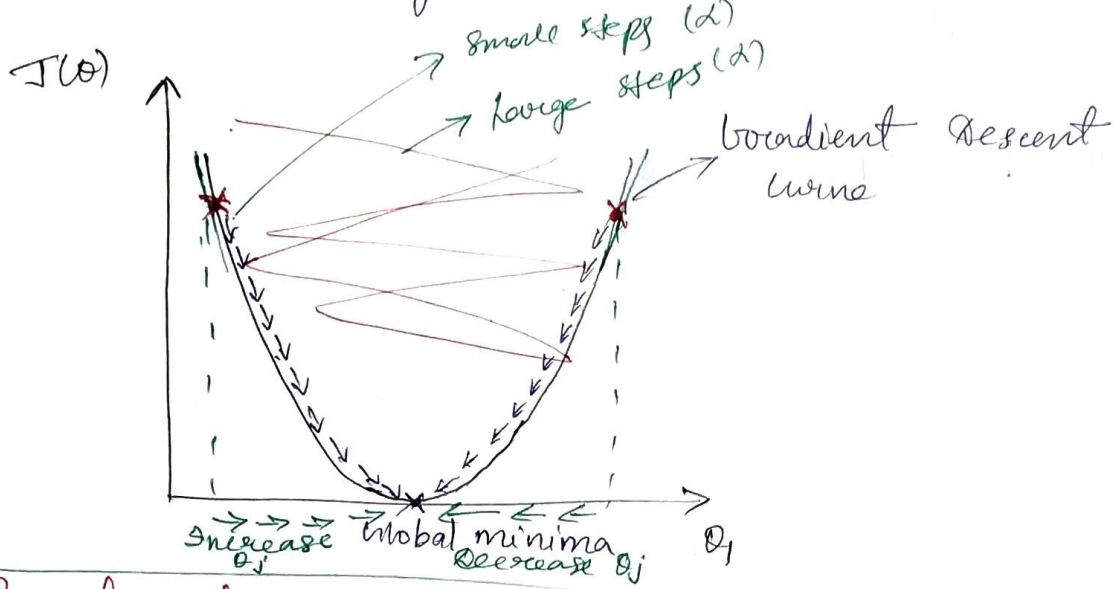
$$= ~~4.33~~ \; 4.67$$

→ Our main aim is to come near Global Minima.

→ We cannot change θ value manually, there should be some mechanism to change θ value to get the global minima. To overcome this issue we use convergence algorithm.

Convergence Algorithm

- Optimize the changes of θ₀ value.



Repeat until Convergence,

{

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

(α: Learning Rate; $\frac{\partial J(\theta_j)}{\partial \theta_j}$: slope)

}

+ve or -ve slope

Right side of the line facing downwards → '-ve' slope
Right side of the line facing upwards → '+ve' slope

'-ve' slope

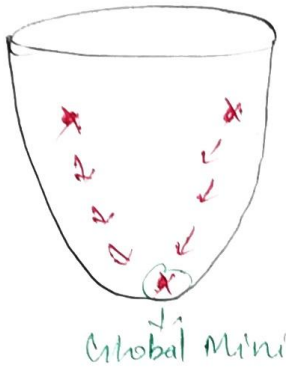
$$\theta_j = \theta_j - \alpha(-ve) = \theta_j + \alpha$$ (It means we are increasing θ_j)

'+ve' slope

$$\theta_j = \theta_j - \alpha(+ve) = \theta_j - \alpha$$ (It means we are decreasing θ_j)

* θ_j = θ_0 and θ_1

3-D diagram (if both θ_0 and θ_1 present)



α, Learning Rate

⇒ It decides the speed of the convergence.

If α is very small ⇒ It will take more time to reach global minima.

If α is very large ⇒ It will jump here and there, global minima will not be reached.

→ α should be around 0.001 for the smaller steps.