Machine Learning

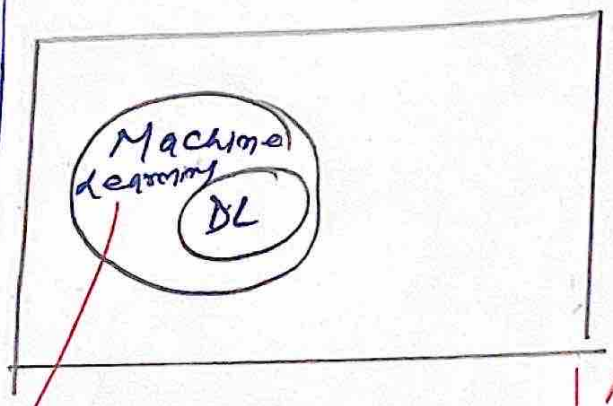
Machine Learning Day-1

Introduction

Today's agenda

1. Machine Learning introduction
2. AI vs ML vs DL vs DS
3. Simple Linear Regression (Mathematical Intitution)

Lets Start with AI vs ML vs DL vs DS



→ Machine Learning
Subset of A.I. ML
Provides us stats tools to explore, visualise, analyse & Perform Prediction & other task

Lets consider Netflix APPLICATION
↳ Recommendation System

A.I
:→ It is creating an APPLICATION where it perform all its tasks without any human intervention.

example:→ Alexa
chatbot - AI
Self-Driving
Youtube Recommendation

Q. What is Deep Learning?

⇒ 1950's the Scientist & Reasearchers think can we make Machine Learn How we we human Being Learn.

Multi-Level Nueral Network invented to Mimic the human Brain.

NLP - Technique
Computer vision

Machine Learning & Deep Learning

- Supervised
  - Regression
    1. Linear Regression
    2. Multi-Linear
    3. Polynomial
    4. SVR
    5. Decision tree
    6. Random forest
    7. xg boost
    8. Naive Bayes
    9. KNN
  - Classification
    1. Logistic Regression
    2. SVM
    3. Decision tree
    4. Random forest
    5. Naive Bayes
    6. KNN
- Unsupervised
  - Clustering Algorithum
    - → DBScan
    - → KMeans
    - → Hierarical
    - → Scoring

Supervised :→ You know the target variable

Unsupervised :→ You Domot know the output features.

Difference between Supervised & Unsupervised

Supervised

Degree	exp	Salary → continuous feature
B.E	7	50K
PHD	2	70K
—	—	65K
—	—	66K
—	—	68

} Regression
it's a Regression Statement Problem

No. of Play	No. of Study	Pass/Fail 1 0
	Indipent feature	} Dependent feature
9	1	0
7	2	0
3	5	1

} classification

Flight Price Prediction → Regression category

Algerian forest fire → classification

Air quality index → Regression

Tommorow Rain/Not Rain → classification

* Dependent features → Output Features

* Indipendent features → Input features

Unsupervised

Customer Segmentation

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

Product

* We divide group Based on Spending Score to recommend offers

This ~~Pati~~ Particular Scenerio is said Customer Segmentation

* And after doing this: Sale will increase nearly by 20%.

Simple Linear Regression

~~For~~

Main Aim :→ To create a Model

① 1 independent var

② 1 dependent var

I/P = Height

Predict O/P = weight

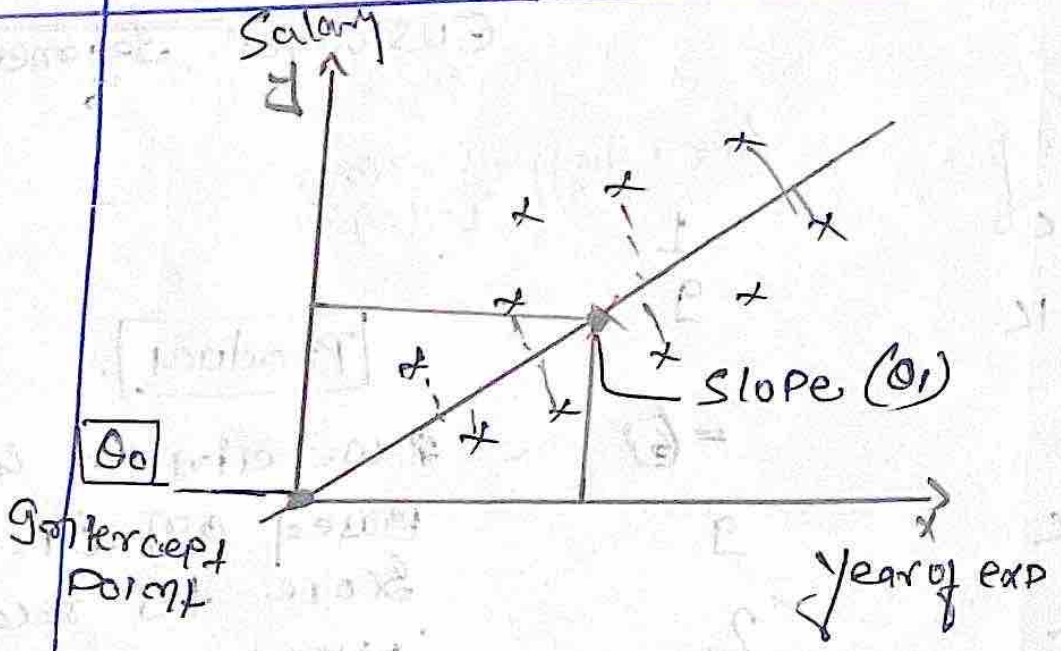
Dataset

(Independent) Year of Exp

(Dependent) Salary

Model training on Year of experience & Salary

Predict Salary on the Based of No. of Experience



Main aim of simple linear Regression is to find the best line & so, the sum of Year of exp Distance of Real Point & Predicted Point minimul.

Best fit line equation of straight line

$$y = mx + c$$

or

$$y = \beta_0 + \beta_1 x_1$$

The Difference Between Predicted Point & Real Point called Residual / error.

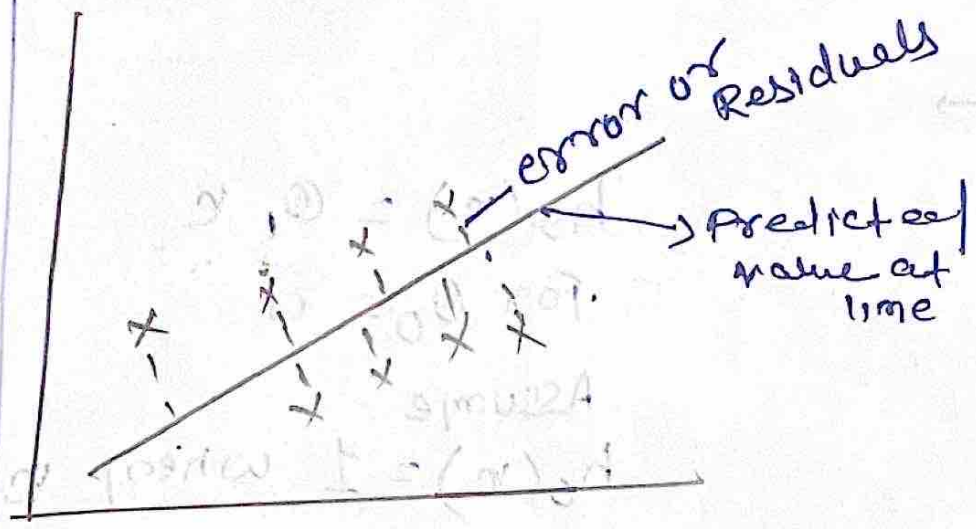
$$h_\theta(x) = \theta_0 + \theta_1 x$$

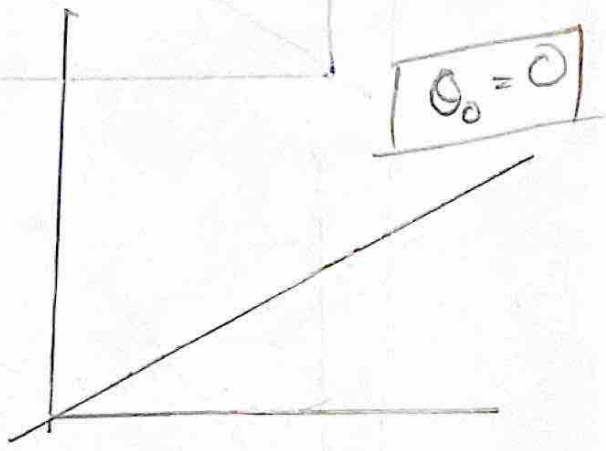
θ_1 → slope (in the unit movement in x-axis what is movement in y axis)

Intercept θ_0

Example 1 Fresher

⇒ If we want to change the line then our intercept (θ_0) & slope (θ_1) also change. To change the line for the best fit this process is called Training of the Model





For Best Prediction we have to minimize the error & it can be done by

Cost Function

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

m = No. of Data Points

$h_\theta(x)^{(i)}$: Predicted Point

$(y)^{(i)}$: Actual Data Point

This technique is called Mean Square error

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

MSE = Mean Square error

m = nuber of Data Points

y_i = Observed Point

$\hat{y}$ → Predicted Point

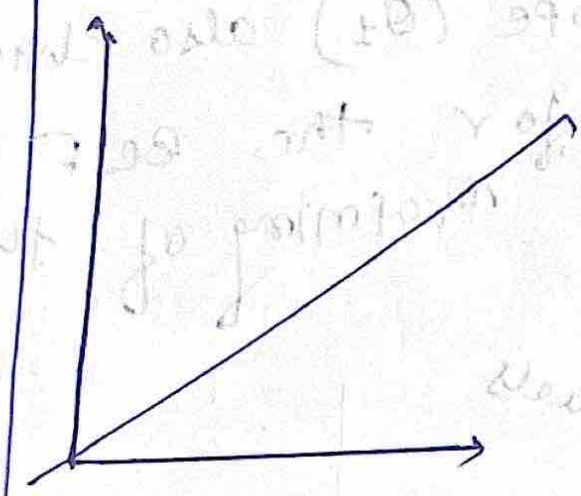
Final Aim

Minimise $J(\theta_0, \theta_1)$

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

By changing θ_0 & θ_1 value.

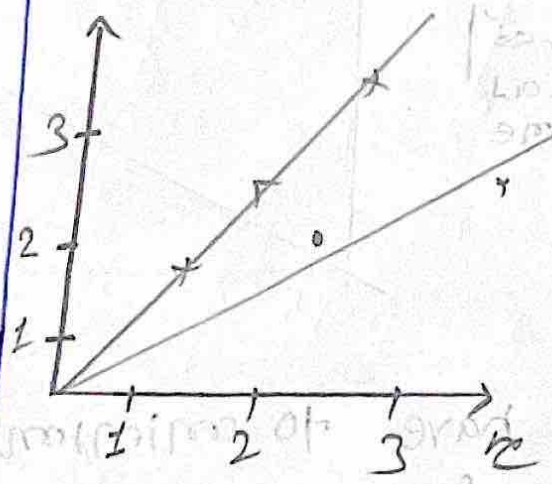
$h_\theta(x) = \theta_0 + \theta_1 x$



Let us consider $\theta_0 = 0$

then

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



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

For $\theta_0 = 0$

Assume

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

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

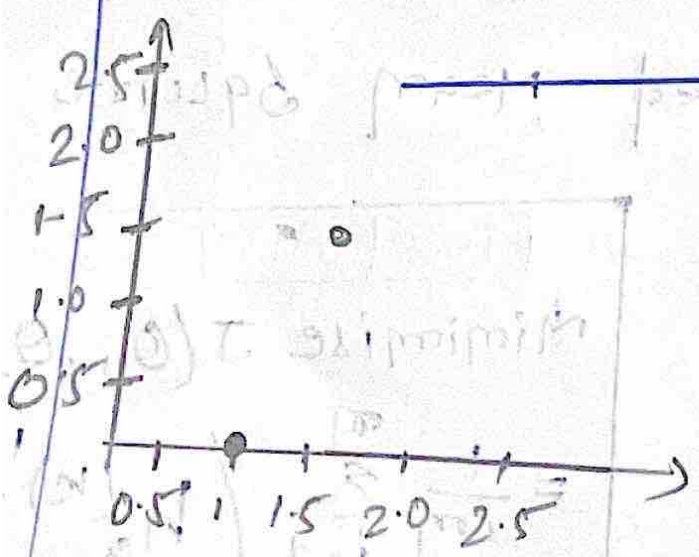
$h_\theta(x) = 3$ when $x = 3$

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

$$= \frac{1}{3} [0 + 0 + 0]^2 = 0$$

Training Dataset

x	y
1	1
2	2
3	3



Assume $\theta_1 = 0.5$

then $h_\theta(x) = 0.5 x$

$h_\theta(x) = 0.5$ where $x = 1$

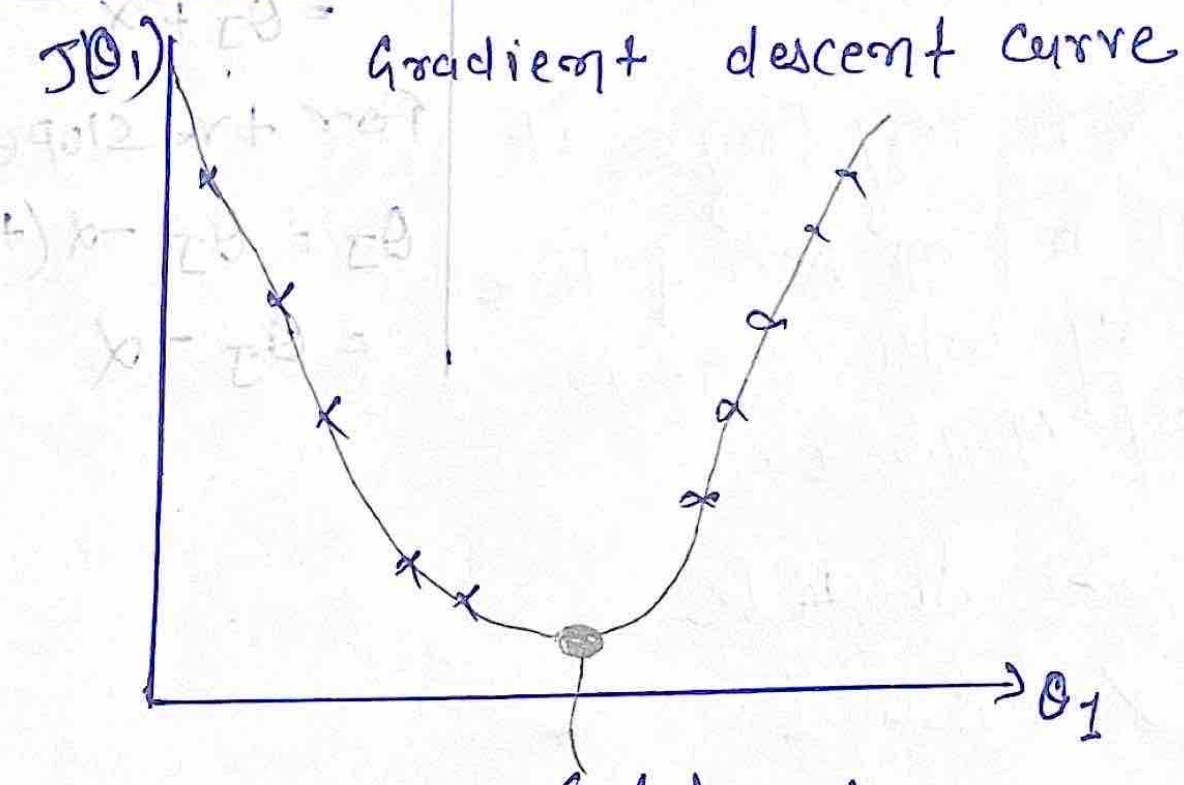
$h_\theta(x) = 1$ $x = 2$

$h_\theta(x) = 1.5$ $x = 3$

$$J(\theta_1) = 1/3 \left[0.25 + 1 + 0.25\right]$$

$$= 1.16$$

When $\theta_1 = 0$



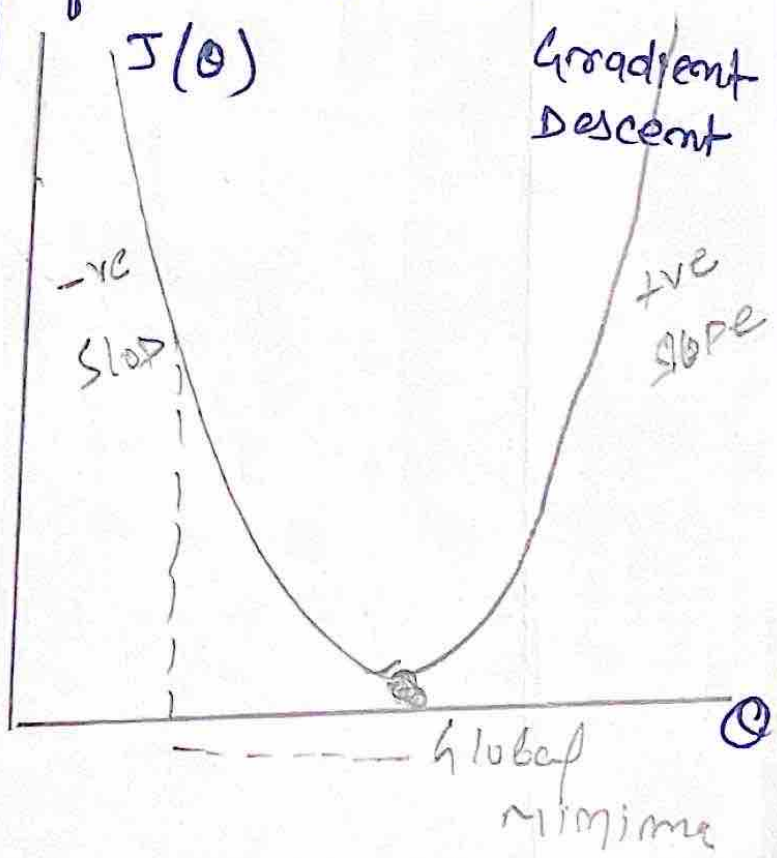
Convergence Algorithm

Repeat until Convergence

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

* Main Aim to come over Global minima because there Difference is very less

to optimize the change of value θ_1 (slope) value



α (learning rate) decides the speed of convergence

If α is small then it will take time to come at global minima.

* if α is very large it will jump here & there hardly will come to global minima.

So, value should be

$\alpha = 0.001$

At global minima,

Slope = 0

$\alpha \rightarrow$ Learning Rate

$\frac{\partial}{\partial \theta_j} J(\theta_j) \rightarrow$ Slope

For -ve slope

$\theta_J = \theta_J - \alpha(-ve)$

$= \theta_J + \alpha$

For +ve slope

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

$= \theta_J - \alpha$