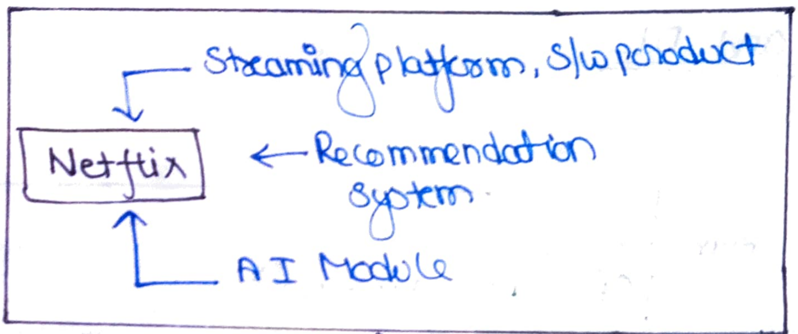
9 oct 2022

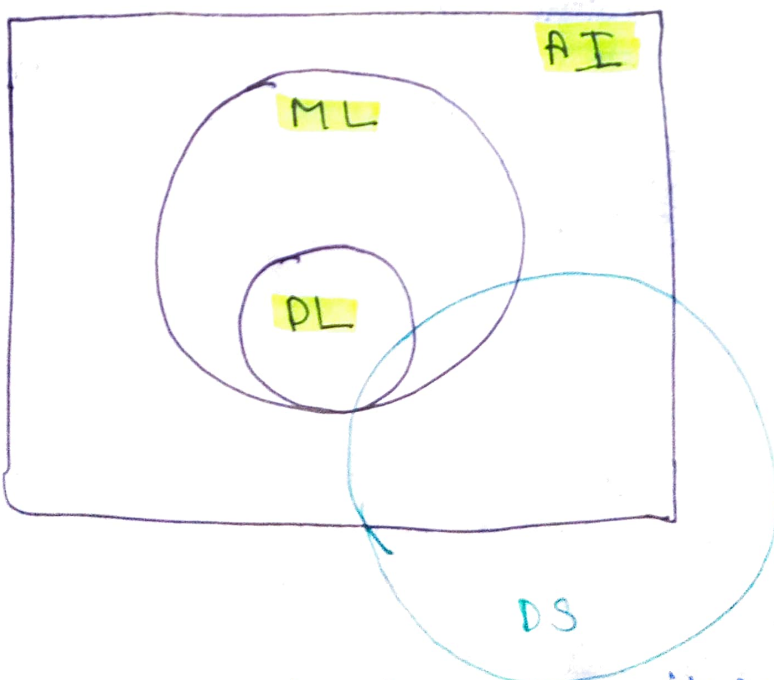
Introduction to Machine Learning

1. Machine Learning Introduction
2. AI vs ML vs DL vs DS
3. Simple Linear Regression → Mathematical ~~Intuition~~ Intiution

AI vs ML vs DL vs DS



AI Application



AI It is an application where it perform all its task without any human Intervention

→ ML provides stats tool to analyze, visualize, & perform prediction & other task with the help of data.

ML is subset of AI

→ DL created to mimic human brain.

ML and DL

Supervised
(Know target variable)

Regression
1. Linear Regression
2. polynomial
3. SVR
4. Decision Tree
5. Random forest
6. xgboost

Classification
1. logistic Regression
2. SVM
3. Decision Tree
4. Random forest
5. Naive Bayes

Unsupervised

* Clustering Algorithms
- DBScan
- Kmeans
- hierical mean clustering
- Silhouette clustering

- A Machine is said to be learning from past Experience with respect to some class of task if its performance in a given task improves with the Experience.

- Types of learning
  1. Supervised
  2. unsupervised
  3. Semi supervised
  4. Reinforcement

Supervised learning

when model is getting trained on labelled dataset (one having both input & output ~~variable~~ parameter)

eg

Degree	Exp	Salary
B.E.	7	50K
PHD	2	70K
–	–	–
–	–	–

→ continous feature

Input feature or Independent feature

output feature or Dependent feature

eg

No. of play Hours	No. of study hours	Pass/fail
		0
9	1	0
7	2	1
3	5	

} Classification Problem

~~If outfeature~~

If output feature

is continous feature → Regression

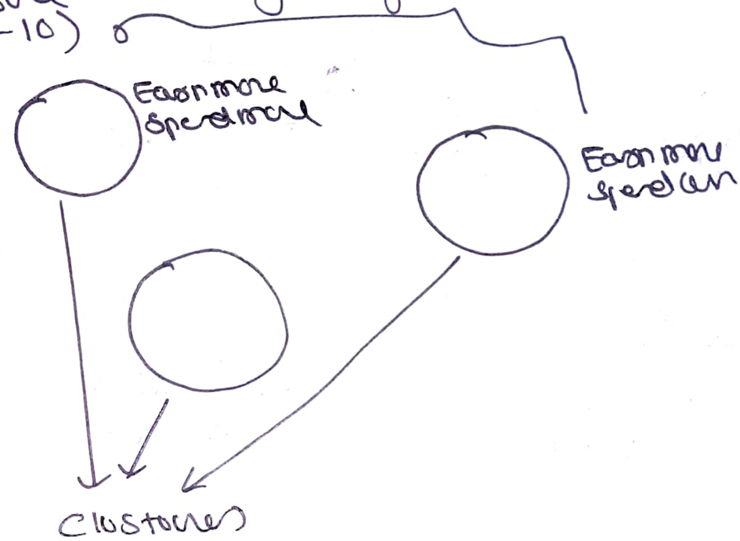
or is, categorical feature → Classification

- House price prediction → Regression
- Algerian forest fire → Classification
- Air Quality Index → Regression
- Tommorow Raincoat → classification
- Person will buy on which Day → classification

Unsupervised ML

Customer Segmentation type of Problem

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



Simple Linear Regression → Try to find the best fit line such that the summation of error should be minimal.

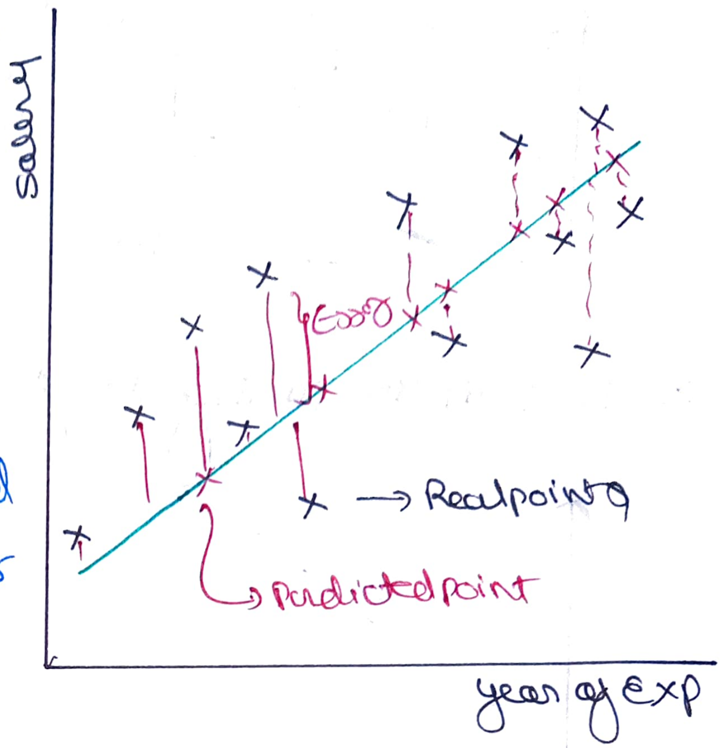
1 independent feature & 1 Dependent feature.

eg Height (x) → wt (y), no. of Rooms (x) $\xrightarrow{\text{Predict}}$ Price (Y),

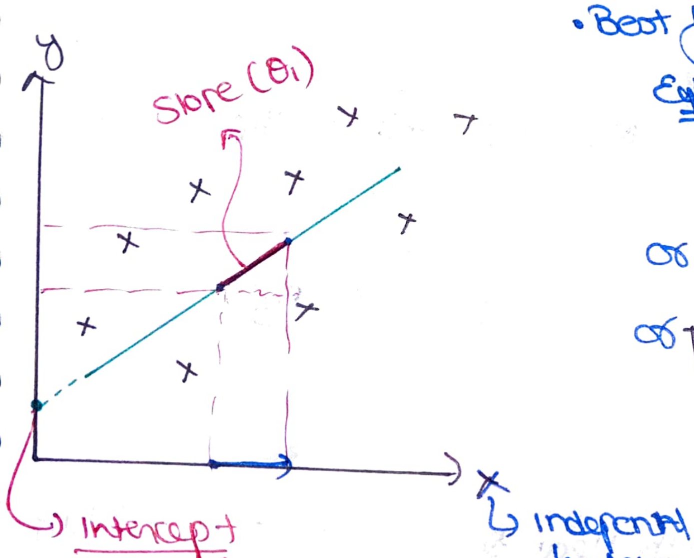
year of exp (x) $\xrightarrow{\text{Predict}}$ Salary (Y)

Eg

Years of Exp	Salary
—	—
—	—
—	—
—	—
—	—
—	—



- Diff b/w the Real & predicted point is called Residuals or Errors



- Best fit line is Eqn of a st. line

Eqn of st. Line

$$y = mx + c$$

or $y = \beta_0 + \beta_1 x_1$

or $\boxed{h_\theta(x) = \theta_0 + \theta_1 x}$

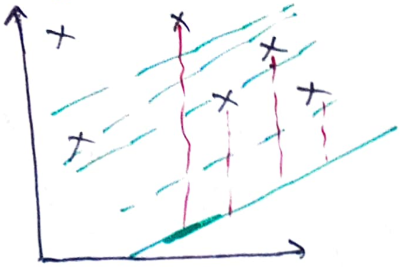
θ_0 — Intercept (when x = 0 where line meeting at y)

θ_1 — Slope
In the unit movement in the x-axis what is the movement in y axis

Exp(x)	Salary(y)
0	3.25 LPA

Even when x = 0 there is some default value of y & that of the best line

• To get the best fit line only need to change θ_0 & θ_1

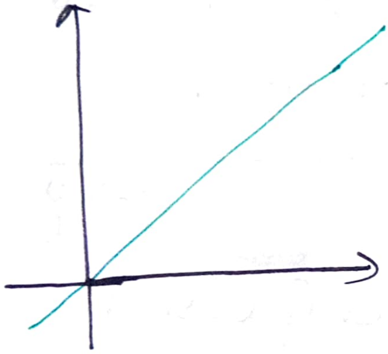


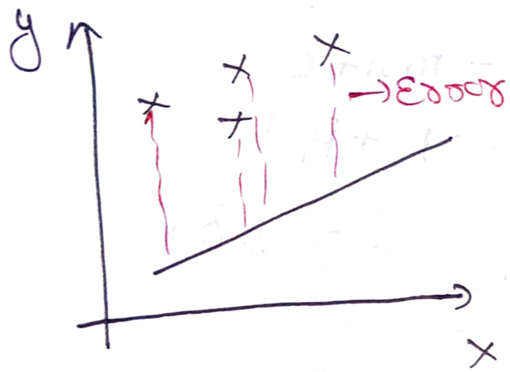
Some iterations to find best fit line $\xrightarrow{\text{by changing}}$ θ_0, θ_1

↓

It is called the Training of the Model

* best fit line passing through origin $\boxed{\theta_0 = 0}$





Need to Minimize the Error

↓ by

Cost function

Cost function (J) - It is the Mean Squared Error (MSE) b/w predicted & actual value

$$\boxed{J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\Big(\underbrace{h_\theta(x)^{(i)}}_{\text{Predicted}} - \underbrace{y^{(i)}}_{\text{Actual}}\Big)^2}$$

MSE - Mean Squared Error

$$\boxed{MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

n = no. of data points
Y_i = obs values
$\hat{Y}_i$ = predicted values

→ **Aim in Linear Regression is**

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

θ_0 = intercept θ_1 = slope

→ to Minimize by changing θ_0 & θ_1 value

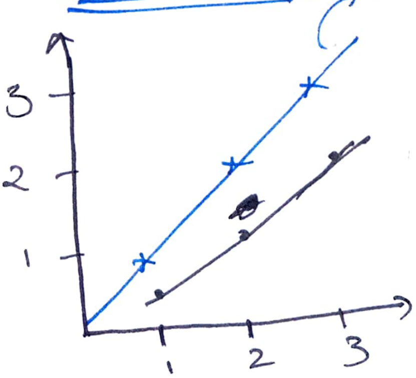
* $h_\theta(x) = \theta_0 + \theta_1 x$

let us consider $\boxed{\theta_0 = 0}$

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

let assume

Best fit line



Assumed $\boxed{\theta_1 = 1}$

$h_\theta(x) = \theta_1 x$
$h_\theta(x) = 1$ when $x = 1$
$h_\theta(x) = 2$ " $x = 2$
$h_\theta(x) = 3$ $x = 3$

Train dataset

x	y
1	1
2	2
3	3

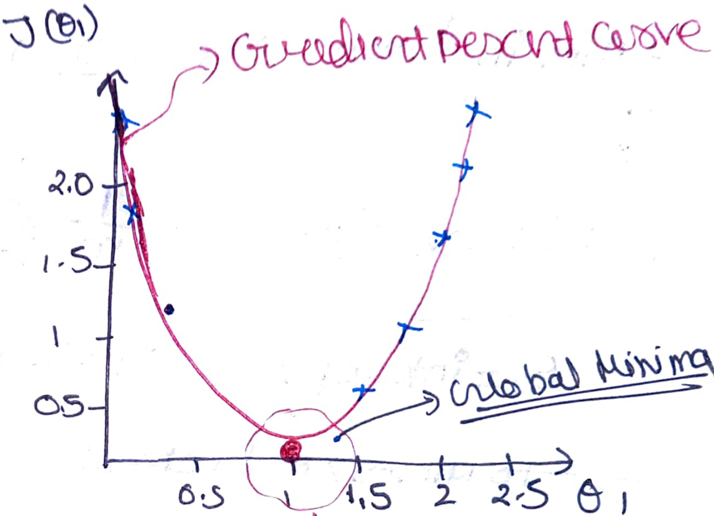
$$J(\theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^i - y^{(i)}\right)^2$$

In our assumed example ~~total~~ datapoints are

$m = 3$

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

$$= 0$$



- when $\boxed{\theta_1 = 1 \quad J(\theta_1) = 0}$

- Similarly

if $\theta_1 = 0.5$

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

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

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

$J(\theta_1) = 1.16$

when $\boxed{\theta_1 = 0.5 \quad J(\theta_1) = 1.16}$

This is our Aim to get so able to see our cost fun^n is Reduce & we get our best fit line. This point is called Global Minima

when $\theta_1 = 0$ — the best fit line along x-axis

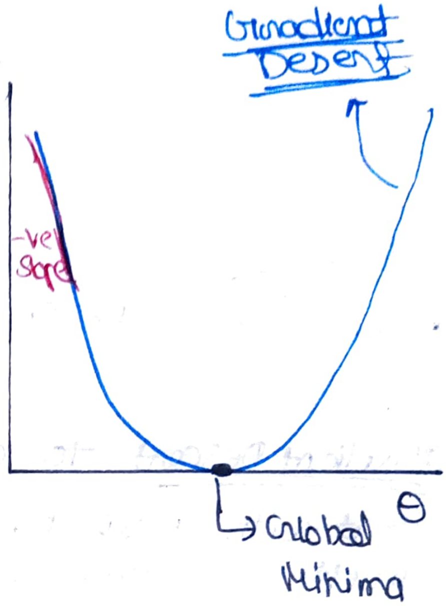
$$J(\theta_1) = \frac{1}{3}\left[(0-1)^2 + (0-2)^2 + (0-3)^2\right]$$

$J(\theta_1) = 4.66$

- There is need to have a Mechanism to change the θ value in Such a way that we can come near over Global Minima so for this Convergence Algorithm is used.

Convergence Algorithm

- optimize the changes of θ_1 value
- also work for θ_0



Repeat until Convergence

{

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

}

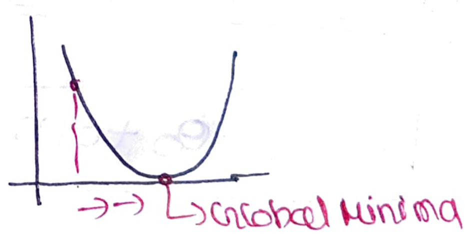
α — Learning Rate — It Decides speed of convergence

$\frac{\partial}{\partial \theta_j}(J(\theta_j))$ = Slope at particular Region

- if Righside of line facing downwards $\Rightarrow$ -ve slope & vice-versa.

- **if -ve slope**

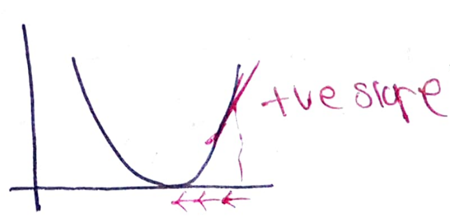
$$\theta_j = \theta_j - \alpha(-ve)$$
$$= \theta_j + \alpha$$



- In -ve slope θ_j is going to inc to get near the Global Minima

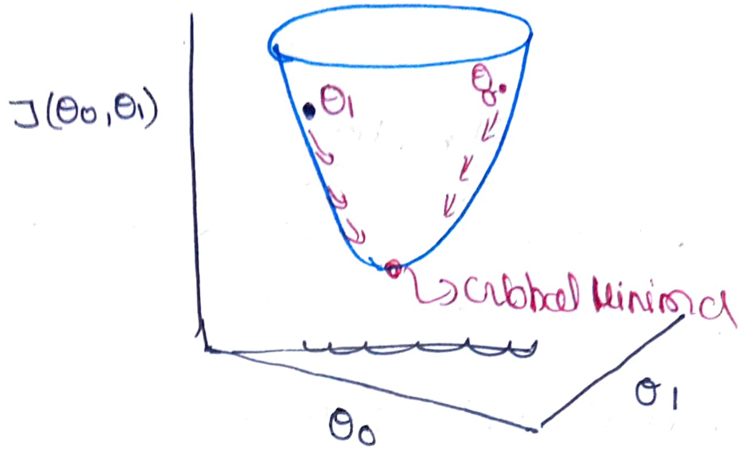
- **if +ve slop**

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



-

• only considered θ_1, if θ_0 is accounted then we get 3-D Diagram



generally

• $\alpha = 0.001$ — for smaller step size

Gradient Descent to acheive best fit line it is used. It is start with Random θ_0 & θ_1 values & then iteratively updating the values, Reaching min cost.

• MSE is the reason coz of which we are getting Gradient Descent

→ our cost function

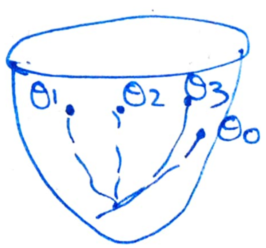
Diffence b/w simple & Multiple Linear Regresion

$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x}$$

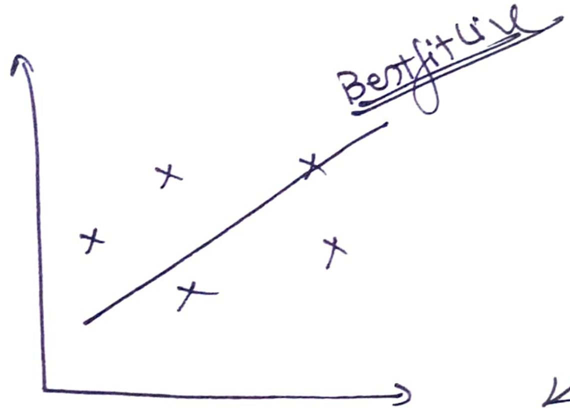
→ 1 independent feature & 1 dependent feature

no of Rooms (x_1)	city (x_2)	Room size x_3	Price

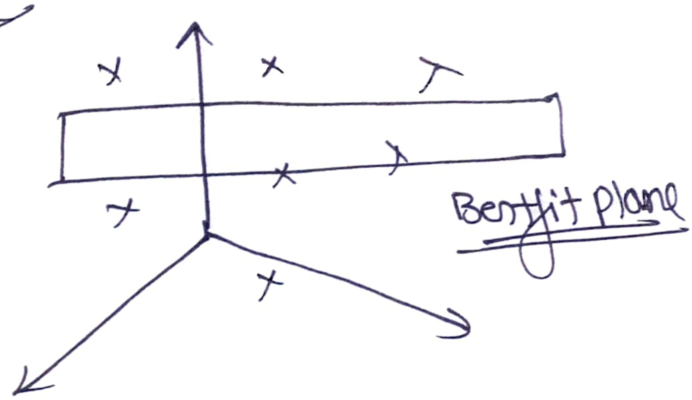
$$\boxed{h_\theta(x) = \underbrace{\theta_0}_{\text{Intercept}} + \underbrace{\theta_1 x_1}_{\text{Slope1}} + \underbrace{\theta_2 x_2}_{\text{Slope2}} + \underbrace{\theta_3 x_3}_{\text{Slop3}}}$$



•



2D



3D

• more than 3-D → hyperplane