Machine learning

Supervised

Regression
1) linear
2) polynomial
3) SVR
4) Decision Tree
5) Random forest
6) Xg boost
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 clustering
4) silhouette score

1) supervised => dataset => output is known
2) un supervised => dataset => output is unknown.

① supervised Learning examples

Ex: 1

Degree	Experience	Salary
BE	7	50k
PHD	2	70k
ME	1	30k
M Tech	4	40k
—	—	—

Independent features: Degree, Experience

Dependent feature: Salary.

since scalary is continuous feature
∴ It is an example of Regression.

Ex: 2 House price prediction
∵ price is continuous feature
∴ Regression.

Ex: 3. Flight fare prediction.
Since fare prediction will be a continuous feature
Therefore it is an example of regression.

Ex: 4 Predict Air quality Index.
since AQI is continuous feature
Therefore Regression example

② classification (supervised learning).

No. of playing hours	No. of study hours.	Exam Result
9	1	0
7	2	0
3	3	1
4	5	1
⋮	⋮	⋮

Pass = 1
fail = 2

I.F ⇒ No. of playing hours and No. of study hours.

D.F ⇒ Exam Result.

Since Exam Result is categorical feature.
Therefore it is an example of classification.

Ex: 2 Algerian Forrest fire dataset.
Since Dependent feature is categorical
Therefore Classification example.

③ unsupervised learning

Dataset

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

Prediction
↓
Product discount

Cluster 1
→ Earn more
→ Spend more

Cluster 2
→ Earn less
→ spend less

Cluster 3.
→ earn less
→ spend more.

Above is customer segmentation example.

① Simple Linear Regression:

It is used when we have one Independant variable and one dependant variable.

Ex: Training dataset.

Height	Weight
6	29
5	55
4	66
6	72

Aim: To create a model which takes input as height and predicts weight.

Ex:

No. of Rooms.	Price of house
2	20k
3	40k
4	60k

Aim:

model ⇒ I/P No. of rooms

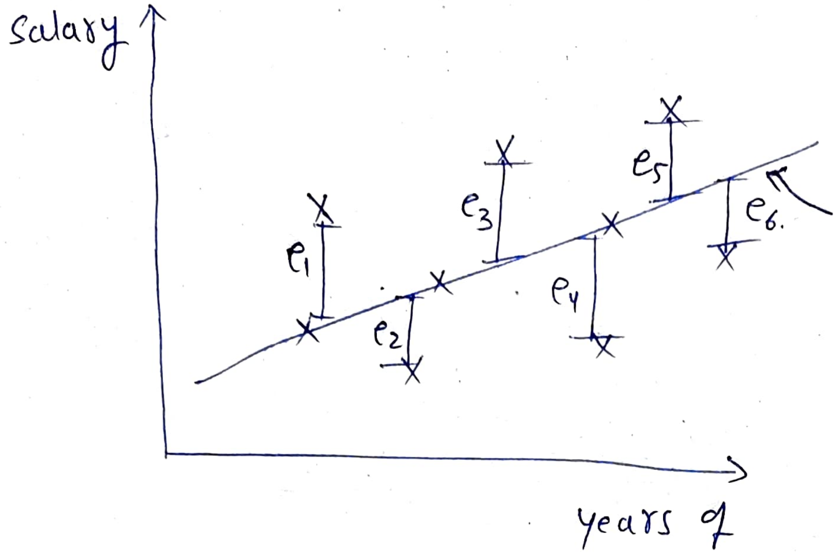
└→ predicts. price of house.

Ex:

Years of Experience	Salary
2	10k
3	40k
5	70k

Aim: model ⇒ I/P years of experience

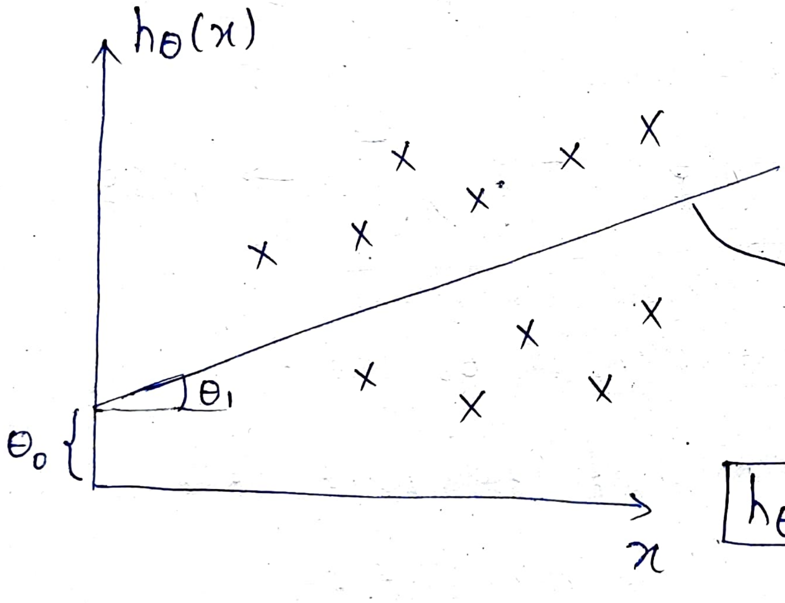
└→ predicts. salary



. Best fit line.

Σ e $\equiv$ error or Residual.

We need to find Best fit line in such a way that Σ error or Residuals is minimum.

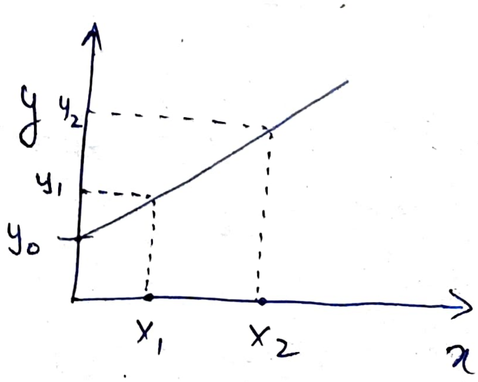


Best fit line Given by

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

where θ_0 is Intercept.
θ_1 is Slope.

To get best fit line we change θ_0 and θ_1.



$$\text{slope} = \frac{y_2 - y_1}{x_2 - x_1}$$

$$\text{Intercept} = y_0$$

Cost Function

A cost function is an important parameter that determines how well a machine learning model performs for a given dataset. It calculates the difference between the expected value and predicted value and represents it as a single real number.

Cost function is a measure of how wrong the model is in estimating the relationship b/w input and output parameter.

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

Here $J(\theta_0, \theta_1)$ = cost function at θ_0 and θ_1

m = Total number of data points in dataset.

$y^{(i)}$ = actual value of dependent feature.

$h_\theta(x)^i$ = predicted value of dependent feature.

The above cost function is called mean square Error (MSE)

The aim is to minimise this mean square error ie the cost function.

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

consider $\theta_0 = 0$

lets have a dataset for $h_\theta(x) = \theta_1 x$

X	y
1	1
2	2
3	3

=> Training dataset.

Here y => dependent featu
X => Independent feature.

X	y	$h_\theta(x)$ $\theta=0$	$h_\theta(x)$ $\theta=0.5$	$h_\theta x$ $\theta=1$	$h_\theta x$ $\theta=1.5$
1	1	0	0.5	1	1.5
2	2	0	1	2	3
3	3	0	1.5	3	4.5

at $\theta=0$

$$J(\theta_1) = \frac{1}{3}\left((0-1)^2 + (0-2)^2 + (0-3)^2\right) = 4.67$$

at $\theta=0.5$

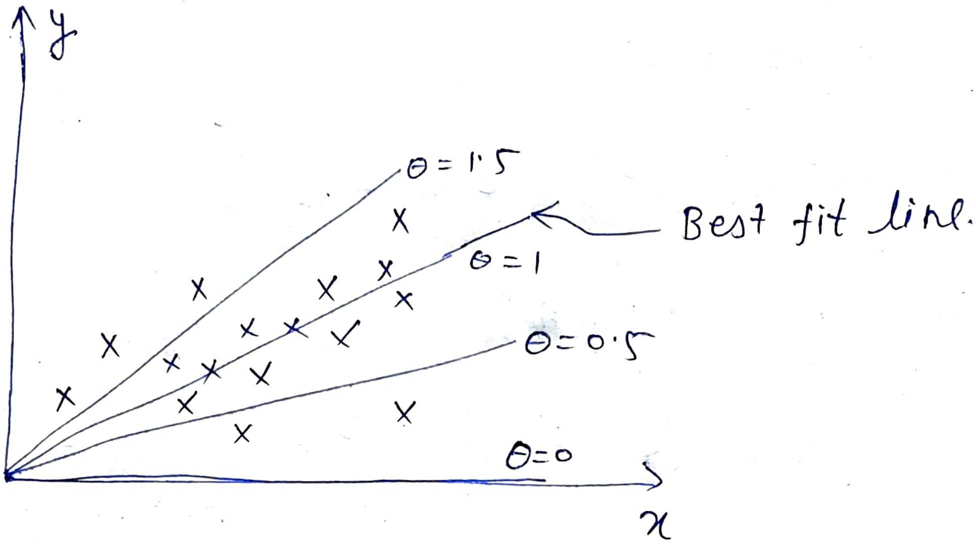
$$J(\theta_1) = \frac{1}{3}\left((0.5-1)^2 + (1-2)^2 + (1.5-3)^2 = 1.16\right.$$

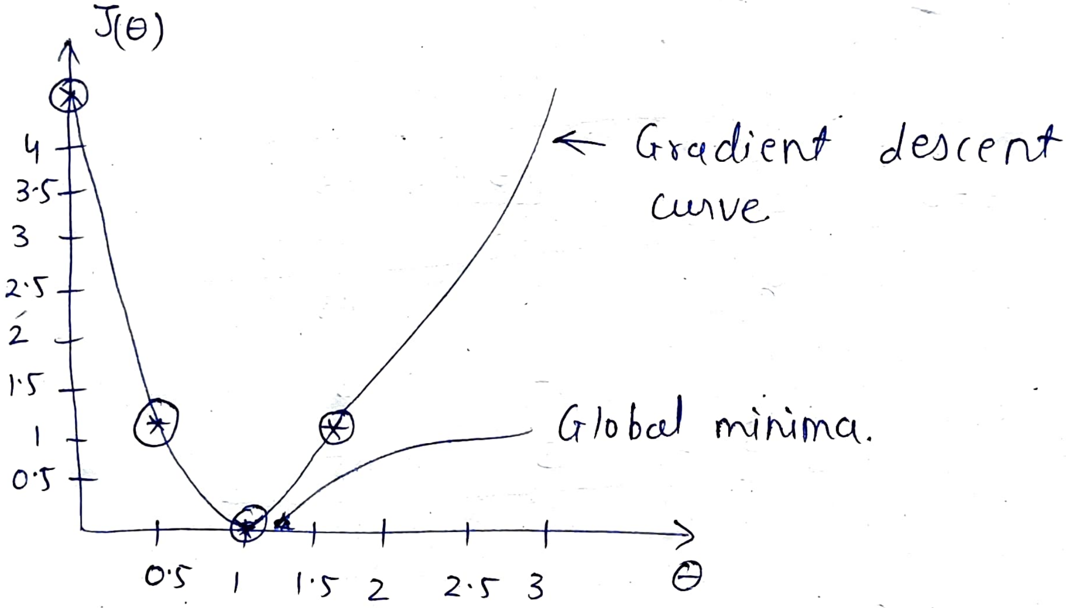
at $\theta=1$

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

at $\theta=1.5$

$$J(\theta_1) = \frac{1}{3}\left((1.5-1)^2 + (3-2)^2 + (4.5-3)^2\right) = 1.16$$





Gradient descent is an optimization algorithm, which is used for optimizing the cost function or error in the model.

Gradient descent is an iterative approch where model gradually converges towards minimum value. and if model further iterates it produces little or zero change in Loss.

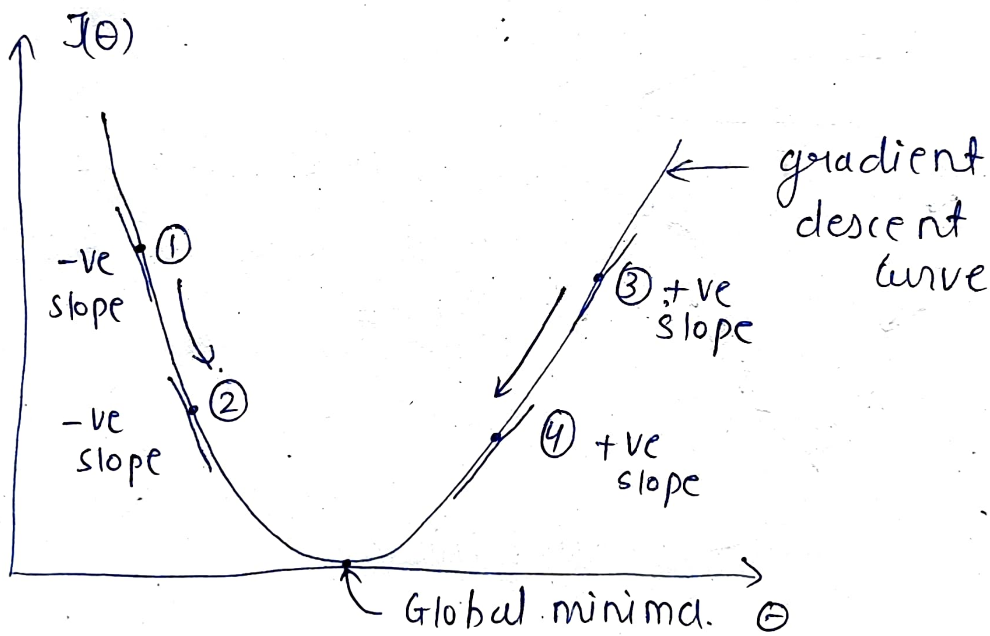
Convergence Algorithm.

Equation for gradient descent in Linear regression.

$$\boxed{\theta_j = \theta_j - \alpha \frac{d\, J(\theta_j)}{d\theta_j}}$$

Here $\alpha \equiv$ learning rate

$$\frac{d\, J(\theta_j)}{d\theta_j} = \text{Slope of } J(\theta) \text{ and } \theta \text{ curve.}$$



at point 1 slope = (-ve)

$$\theta_j = \theta_j - \alpha(-k_1)$$

$$\theta_j = \theta_j + \alpha k_1$$

at point 2 slope = -ve

$$\theta_j = \theta_j - \alpha(-k_2)$$

$$= \theta_j + \alpha k_2.$$

.... we are moving close to global minima.

at point 3 $\quad$ slope = +ve

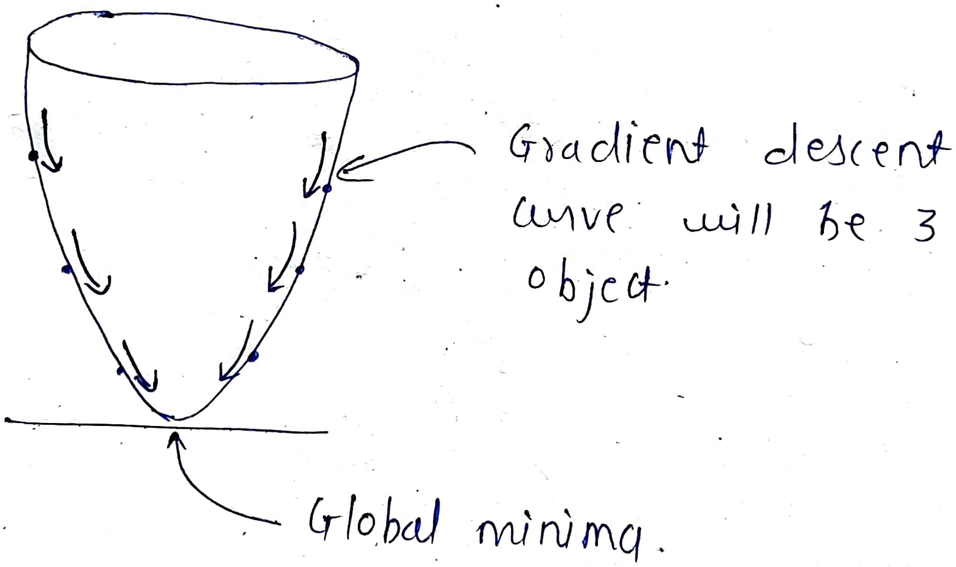
$$\theta_j = \theta_j - \alpha(+k_3)$$
$$= \theta_j - \alpha k_3$$

at point 4 $\quad$ slope = +ve

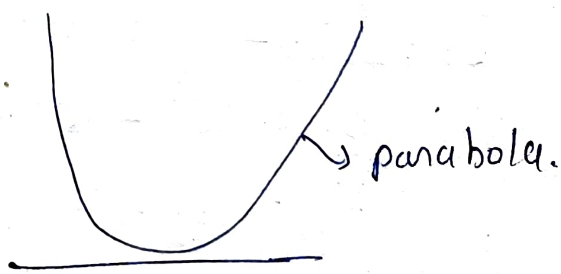
$$\theta_j = \theta_j - \alpha(+k_4)$$
$$= \theta_j - \alpha k_4$$

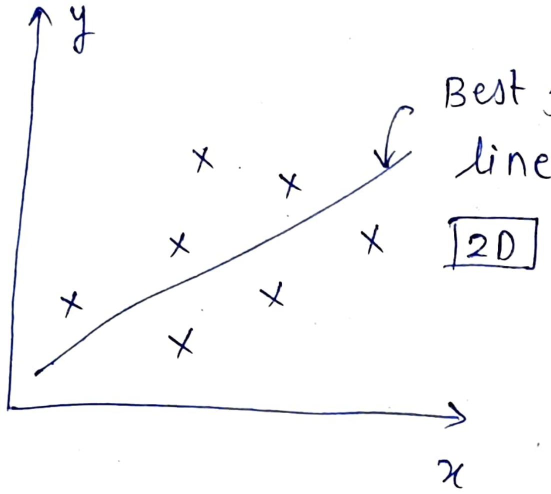
..... we are moving close to global minima.

Note: If θ_0 and θ_1 both considered.

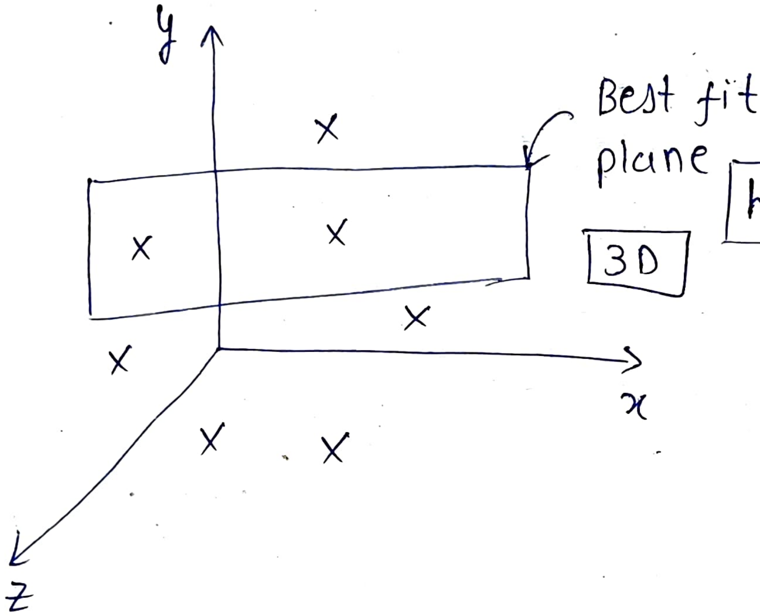


MSE ⇒ is reason we get gradient descent curve





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



$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x}$$

* If there are more than 3 Dimensions then we get hyper plane.

Ex: $$\boxed{h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x + \theta_3 x}$$

hyper plane

Ex:

$I.F_1$	IF_2	IF_3	D.F.
No. of rooms	city	Room size	House price