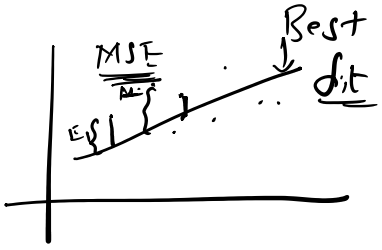
16 March 2023 20:07

Linear Regression :- Cost function of linear Regression is MSE

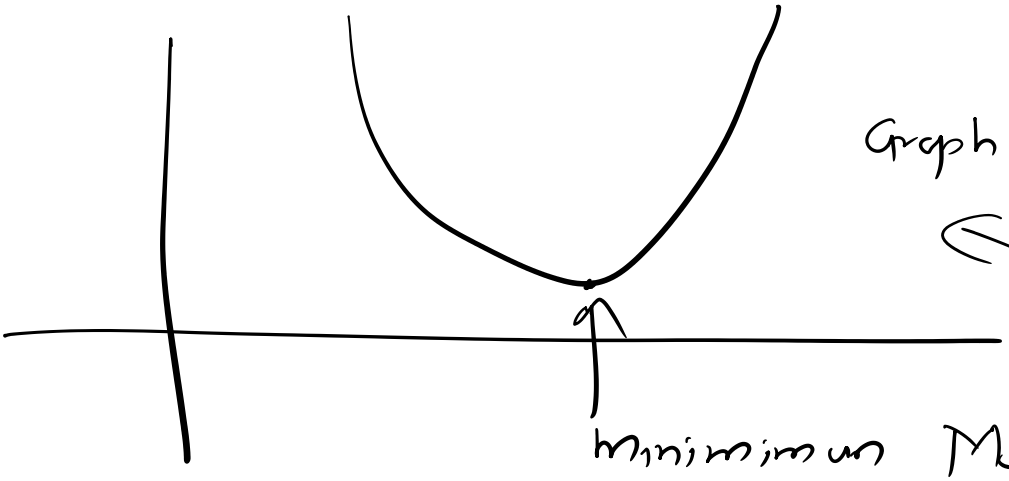


$$MSE = \underbrace{\frac{1}{n}}_{\text{constant}} \underbrace{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}_{\text{Squared function.}}$$

Goal of LR is to minimize the cost function which is the Mean squared Error.

When I say, I want to minimize the MSE, It means that I want to find the minimum value of MSE.

Graph of MSE :-



It means that MSE is minimum at this point

Q. How to reach the minimum point ?

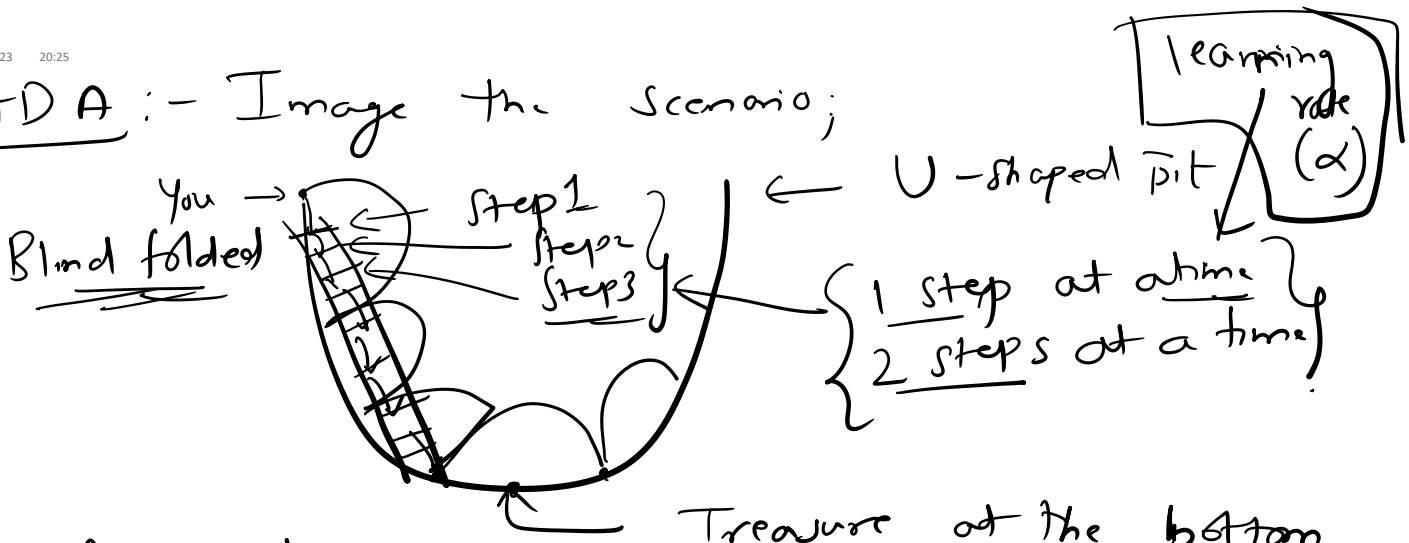
If I want to reach min point there is a step by step process.

↓

Gradient Descent Algorithm :- It is an optimization algorithm to find the minimum value of any cost function

16 March 2023 20:25

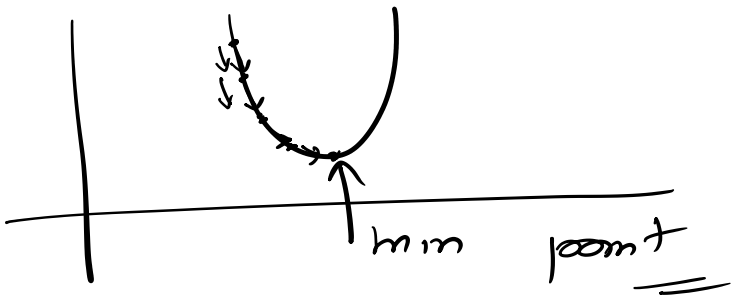
GDA :- Image the scenario;



(2) Approach :-

(1) Step by step descend using some ladder

(2) Directly jump to the bottom ×



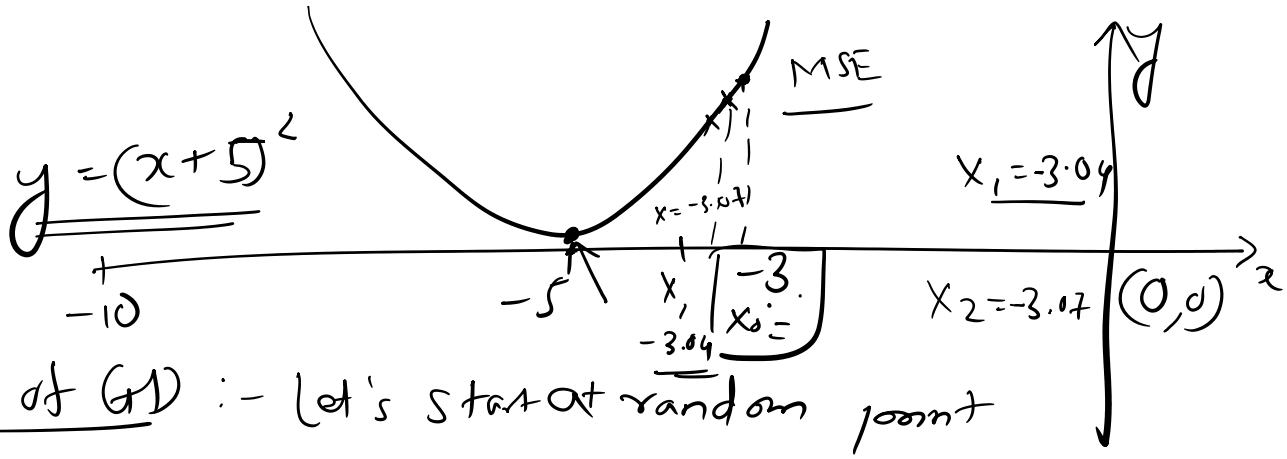
Learning rate(α) :- The α decides how fast the algorithm reaches the minimum point.

16 March 2023 20:34

eg:-

CF:-

$y = (x+5)^2$



Step 1 of GD :- Let's start at random point

$x = -3$

Step 2 of GD :- Move in the negative direction of the gradient (derivative)

$y = (x+5)^2 \Rightarrow \frac{dy}{dx} = 2(x+5)$

$y = x^2 = \frac{dy}{dx} = 2x$

Q. But how much to move? For this we have our learning rate (α) = 0.01 ← Assumption

Step 3 :- Perform 2 iterations of ~~g~~ GD, to understand how GD works.

Initilize Parameters :-

$X_0 = -3$ $\quad \frac{dy}{dx} = 2(x+5)$ $\quad \alpha = 0.01$

Iteration 1 :-

$$X_1 = X_0 - \alpha * \left(\frac{dy}{dx}\right)$$

$$X_1 = (-3) - (0.01) * (2*(-3)+5)$$

$$\boxed{X_1 = -3.04}$$

Iteration 2:- $X_1 = -3.04$ $\alpha = 0.01$ $\frac{dy}{dx} = 2(x+5)$

$$X_2 = X_1 - \alpha * \frac{dy}{dx}$$

$$= (-3.04) - (0.01) * 2(0.01+5)$$

$$\boxed{X_2 = -3.07}$$

16 March 2023 21:08

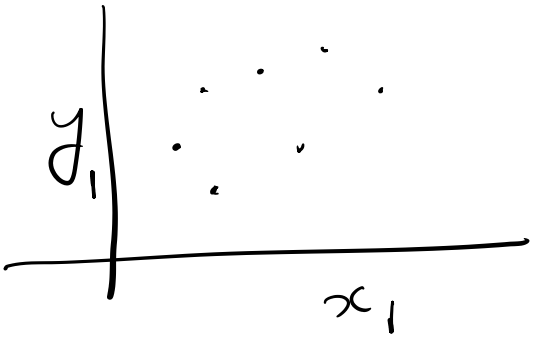
Multiple Linear Regression :-

1. Many Independent Variables
2. 1 Dependent or target varible.

The eqn of MLR :-

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots\dots$$

Q. What is important condition for Linear Regression?



The problem arises when there are multiple independent variables!

Q What is the problem.

16 March 2023 21:15

For real estate dataset,

(1) Lin Reln between house age & price is ok

(2) Lin Reln between MRT stores & price is ok

But

(1) Lin Reln between house age & MRT store is a problem because both house & MRT store are independent variable

16 March 2023 21:18

MLR :- (1) Lin Relⁿ betⁿ target & independent variables should be present

(2) Lin Relⁿ betⁿ any 2 independent variable should not be present

Multicollinearity :- Lin Relⁿ betⁿ 2 independent variables is called Multicollinearity

Q How to check Multicollinearity?

Ans :- (1) Scatter plot (2) Correlation Map