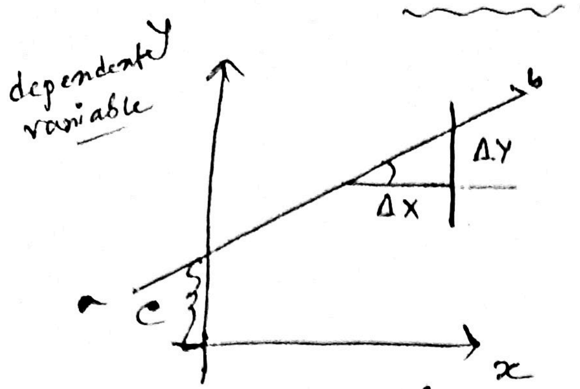
Linear Regression



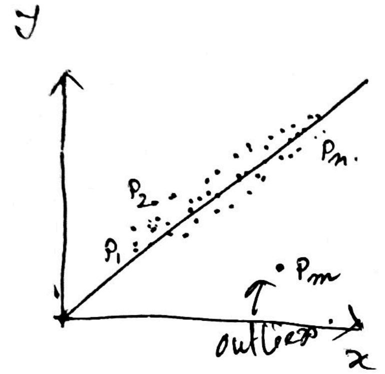
Equation of SL - ab is

$y = mx + c$

slope (m), intercept on y axis (c)

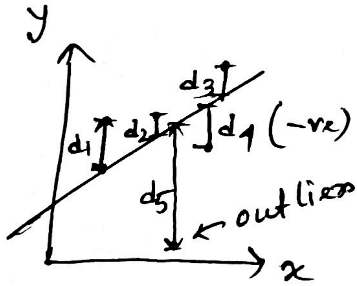
So $m = \frac{\Delta Y}{\Delta X}$ means for unit change in x what is the change in y.

c = When $x = 0$ what is the value of y.



line fitting to approximate the points. → linear regression.

So we have to find out the equation of SL that is __best fit__ for the given points $P_1\ P_2\ P_3 \cdots P_n$.



Intution of best fit is to minimize the total error distance of all the points from the fitted line.

$$\text{minimize}\left(d_1 + d_2 + d_3 \cdots d_n\right)$$

Remember if we consider d_3, d_1, d_2 be positive (+ve) d_4 should be -ve. So __Just__ adding the distance ~~wold~~ would not work. So we need to -

1. Take square of all distance,
2. To minimize outlier's effect take average dis.

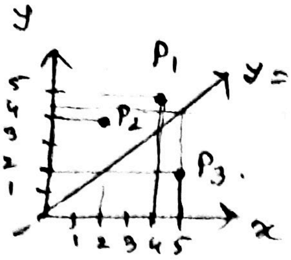
$$\text{MSE (mean square error)} = \underline{\text{minimize}} \sum_{i=1}^{n} \frac{(y - y_i)^2}{2n} \Rightarrow \text{Cost function}$$

⇓ need to minimize

Let us consider for our case in

$y = mx + c \qquad c = 0. \qquad$ so.

(1)



When $m = 1 (45^\circ)$ $y = x$.

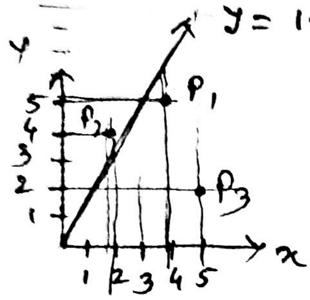
$d_1 = 1$
$d_2 = 2$
$d_3 = -2.5$

$$MSE = \frac{1}{2 \times 3} \; 1^2 + 2^2 + (-2.5)^2$$

$$= \frac{1+4+6.5}{6}$$

$$= \boxed{1.875}$$

(2)



When $m = 1.73 (60^\circ)$

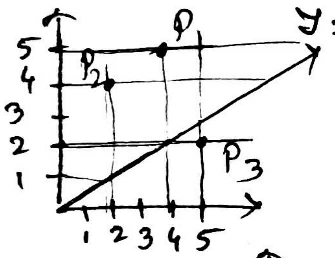
$d_1 = -1$
$d_2 = 1$
$d_3 = -7$

$$MSE = \frac{(-1)^2 + 1^2 + (-7)^2}{6}$$

$$= \frac{49+2}{6}$$

$$= \boxed{8.5}$$

(3)

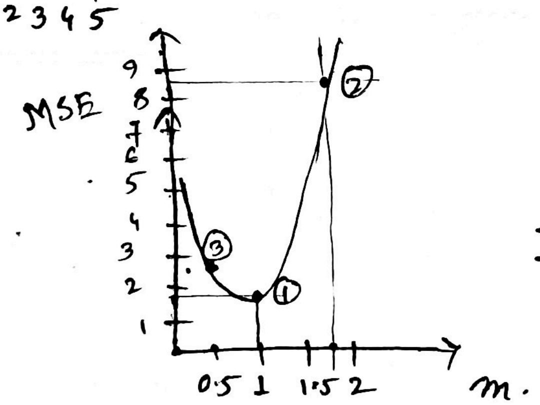


When $m = .577 (30^\circ)$

$d_1 = 2.5$
$d_2 = 3.$
$d_3 = -1$

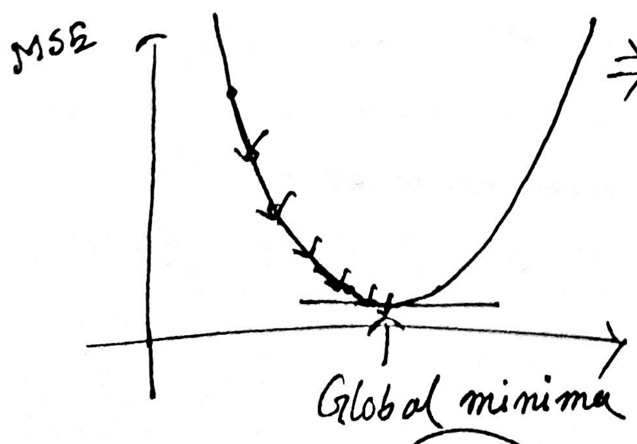
$$MSE = \frac{(2.5)^2 + 3^2 + (-1)^2}{6}$$

$$= 2.70$$



If we plot m and MSE we will get a curve like in the left hand.

Gradient Descent Algorithm:



$$\Rightarrow Y = F(\theta_1)$$

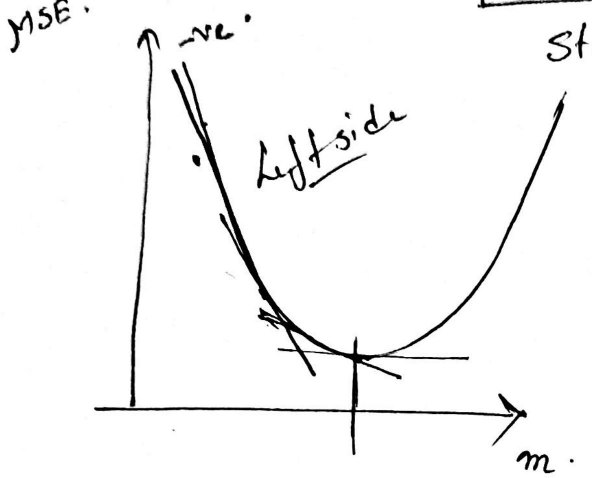
To minimize - Y

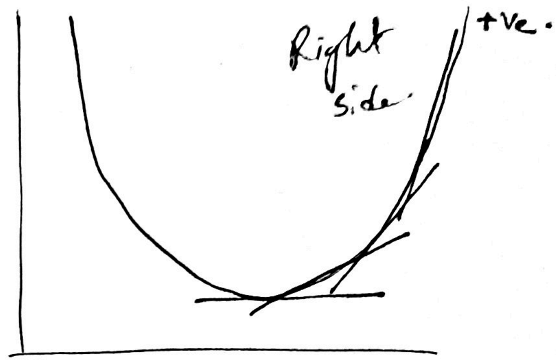
$$\frac{\delta}{\delta \theta_1} Y \Rightarrow \frac{\delta}{\delta \theta_1} F(\theta_1)$$

$$\theta_1 \Rightarrow \frac{\delta}{\delta \theta_1} F(\theta_1) = 0$$

Theory of Convergence

$$\theta_{t+1} = \theta_t - \underbrace{\frac{\delta}{\delta\theta} F(\theta)\,\alpha}_{\text{Step size}} \leftarrow$$ learning rate (very small value like 0.001)





at the left side

$$\frac{\delta}{\delta\theta} F(\theta) = -ve$$

$$\theta_{t+1} \Rightarrow \theta_t - (-ve) \Rightarrow \theta_t + (+ve)$$

$$\theta_{t+1} > \theta_t$$

On Right

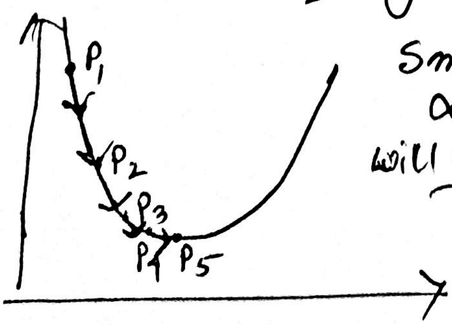
$$\frac{\delta}{\delta\theta} F(\theta) = +ve$$

$$\theta_{t+1} \Rightarrow \theta_t - (+ve)$$

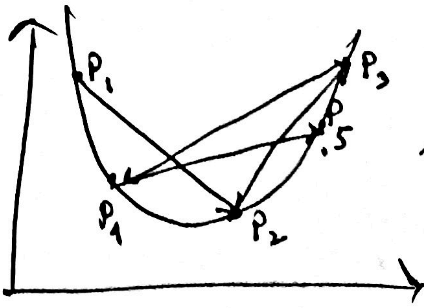
$$\theta_{t+1} < \theta_t$$

$\Rightarrow$ So the eqn. converge.

Learning Rate



Small α will converge.



Big α may not converge