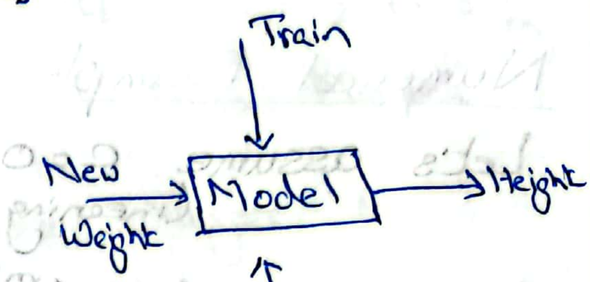
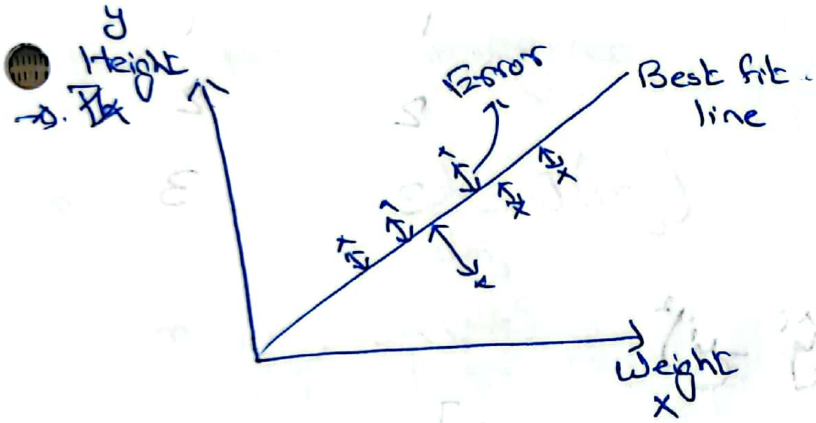
Linear Regression Model
Leature 5

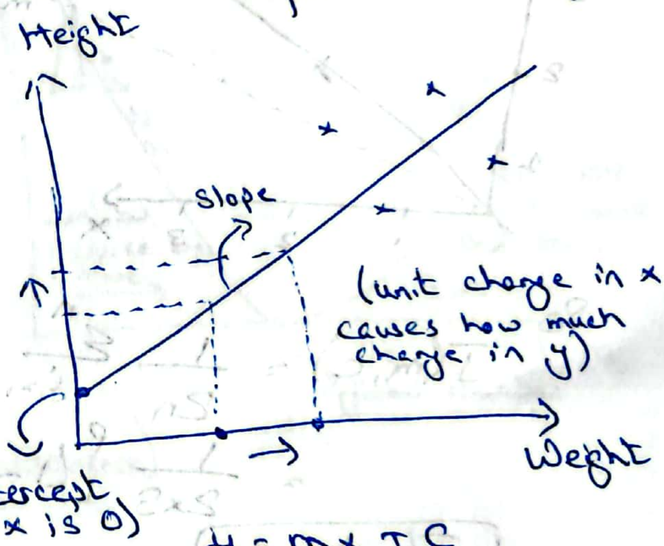
→ Important Algo to understand concepts such as cost function, gradient decent, Vectorization

→ Data set

(Independent/Input) feature Weight (x)	(Dependent/output) feature Height (y)
74	170 cm
80	180 cm
75	175.5cm







Error

Where $\hat{y}$ is predicted value

- Error $= y - \hat{y}$

$y = mx + c$ for one feature

$y = mx_1 + c$ for many features

$y = c + m_1x_1 + m_2x_2 + \dots + m_nx_n$

→ The objective is to minimize the error by rotating best fit line by changing m & c in an optimization technique to get best fit line.

→ **Cost Function**

$$J(m, c) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}^i - y^i)^2$$

↳ (Mean Squared Error)

Final object of what we need to do is

Minimize

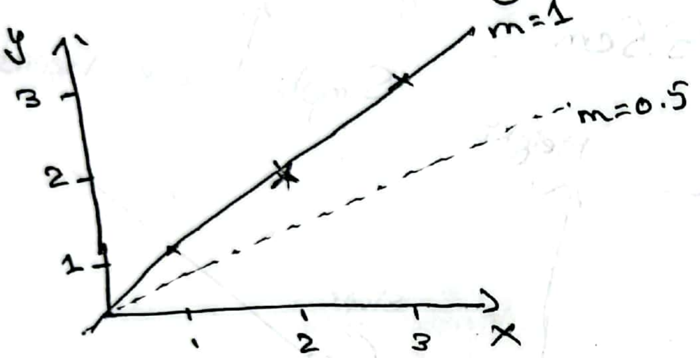
$$J(m,c) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}^i - y^i)^2$$

by changing m & c values.

Numerical Example

Let's assume $c = 0$

(meaning x is passing through origin).



x	y
1	1
2	2
3	3

So

$$J(m,c) = \frac{1}{2n} \sum_{i=1}^{n} (\hat{y}^i - y^i)^2$$

$$= \frac{1}{2\times3} \left[(1-1) + (2-2) + (3-3)\right]$$

$$\boxed{J(m) = 0}$$

Now suppose $m = 0.5$ $\quad \hat{y} = mx$

$$J(m) = \frac{1}{2\times3} \left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right] \text{ if } \begin{array}{ll} x=1 & \hat{y}=0.5 \\ x=2 & \hat{y}=1 \\ x=3 & \hat{y}=1.5 \end{array}$$

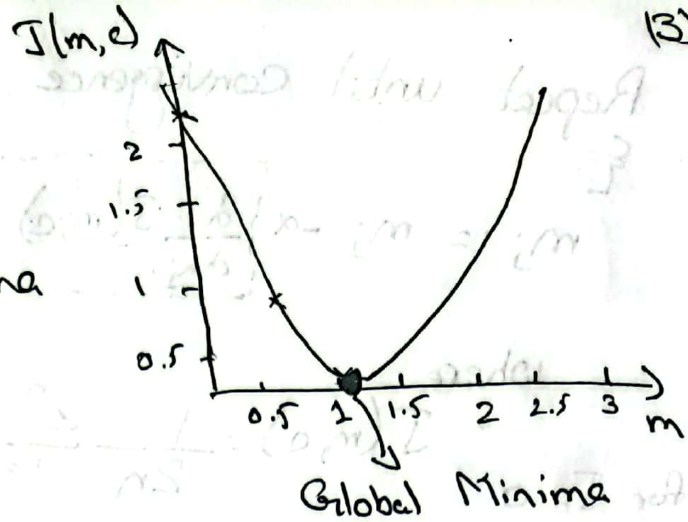
$\simeq 0.58$

Suppose $m = 0$

$$J(m) = \frac{1}{2\times3} \left[(0-1)^2 + (0-2)^2 + (0-3)^2\right]$$

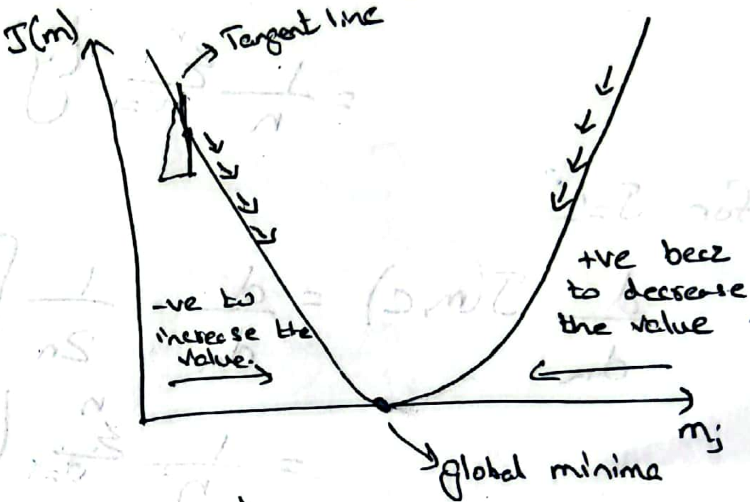
$\simeq 2.3$

This curve is known as gradient decent and is derived by changing m & c. Best fit line is on global minima



→ Convergence Agorithm

The objective is to optomize the changes in 'm' values.



Repeat until Convergence

{

$$m_j := m_j - \alpha \frac{d}{dm_j} J(m_j)$$

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

$$= m_j - (+ve)$$

$m_j - \alpha$

α = Learning rate (the rate of convergence)

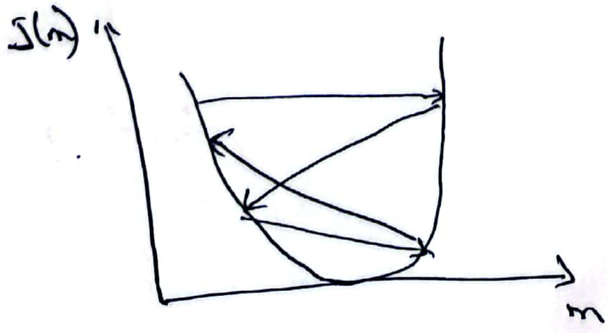
e.g. $\boxed{\alpha = 0.001}$

$$m_j := m_j - \alpha \frac{d}{dm_j} J(m_j)$$

$$m_j = m_j - a(-ve)$$

$$= m_j + \alpha$$

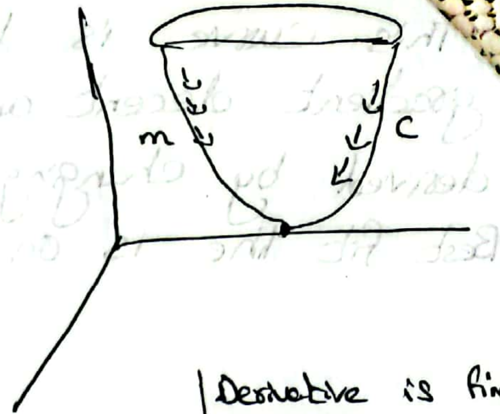
So till now we were considering that the best fit line is originating from origin, hence $c = 0$. This was for better understanding on a 2D plot.



Repeat until convergence

{

$$m_j := m_j - \alpha \boxed{\frac{d}{dm_j} J(m,c)}$$



where

$$J(m,c) = \frac{1}{2n} \sum_{i=1}^{n} (\hat{y}^i - y^i)^2$$

Derivative is finding the slope

$$\frac{d}{dx}(x)^2 = 2x$$

$$\frac{d}{dx} x^n = n x^{n-1}$$

for J= 0

$$\frac{d}{d(m,c)} J(m,c) = \frac{d}{dm_j} \cdot \frac{1}{2n} \left[\sum_{i=1}^{n} (\hat{y}^i - y^i)^2 \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} (\hat{y}^i - y^i)$$

for J=1

$$\frac{d}{dm} J(m,c) = \frac{d}{dm} \cdot \frac{1}{2n} \left[\sum_{i=1}^{n} ((c+mx) - y^i)^2 \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} ((c+mx) - y^i) x$$

Repeat until convergence

{

$$c := c - \alpha \frac{1}{n} \sum_{i=1}^{n} (\hat{y}^i - y^i)$$

$$m := m - \alpha \frac{1}{m} \sum_{i=1}^{n} (\hat{y}^i - y^i) x^i.$$

}

Confusion Matrix:-

		Predicted		
	165	No	Yes	
Actual	No	[TN] 50	[FP] 10	60
	Yes	[FN] 5	[TP] 100	105
		55	110	

→ Accuracy $= \frac{TP + TN}{Total}$

$= (100+50)/165$

$= 0.91$

→ Error $= 1 -$ accuracy

or

$\frac{FP + FN}{Total}$

$= 0.09$

Precision $= \frac{TP}{\text{Predicted Yes}}$

$= \frac{100}{110} = 0.64$

Recall $= \frac{TP}{\text{Actual Yes}}$

$= \frac{100}{105} = 0.95$

t_test

Left Tailed Test

H_0: ~~μ 60~~ $\mu \geq 60$

H_1: $\mu < 60$

α: $5\% = 0.05$

Sample size $n = 16$

Standard deviation $S = 40$

Mean $\bar{x} = 70$

$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{70-60}{40/\sqrt{16}}$

$t = 10/10 = 1$

Df(Degree of freedom) $= 16-1$

$= 15$

Right Tailed Test

H_0: $\mu \leq 400$

H_1: $\mu > 400$

α: 5%

$n = 12$

$S = 40$

$\bar{x} = 415$

$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$

$= \frac{415-400}{400/\sqrt{12}}$

$t = 1.297$

$df = 12-1 = 11$

Two tailed test

H_0: $\mu = 10$

H_1: $\mu \neq 10$

$n = 16$

$S = 20$

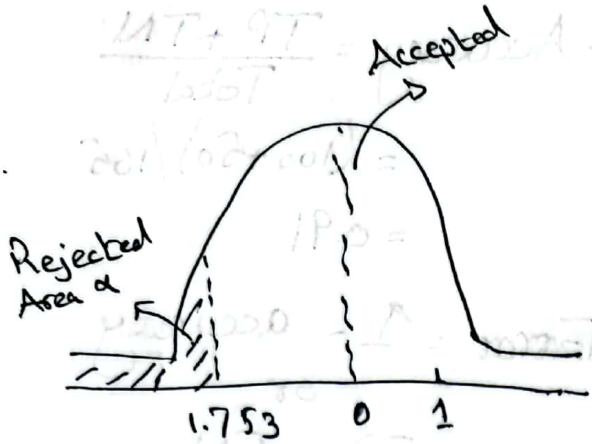
$\bar{x} = 20$

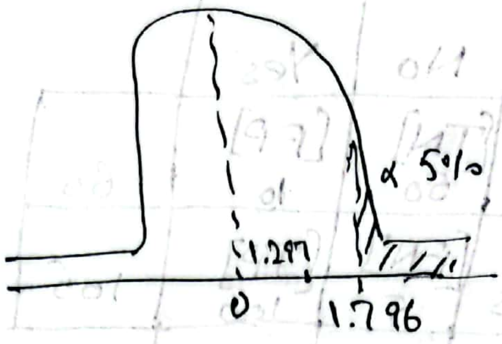
$\alpha = 5\%$

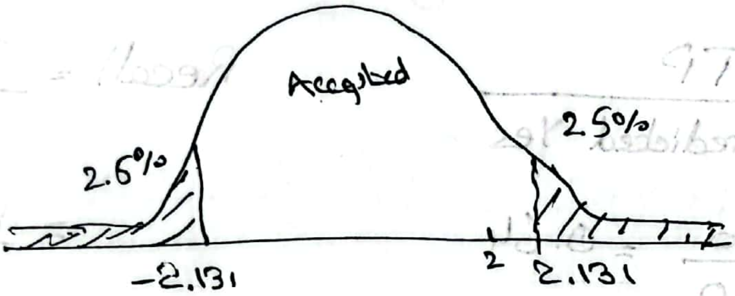
$t = \frac{20-10}{20/\sqrt{16}}$

$= 10/5$

$t = 2$


Accepted
Rejected Area α
1.753
0
1


α 5%
1.297
0
1.796


Accepted
2.6%
2.5%
-2.131
2
2.131