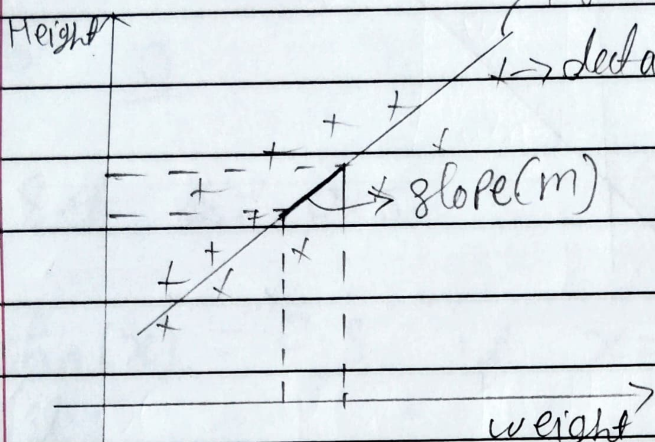
Regression Analysis

Linear regression.

fitted line / best fit line.



data points.

slope(m)

$y = mx + c$

$y = \beta_0 + \beta_1 x$

slope

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

Intercept

Cost function.

predicted O/P

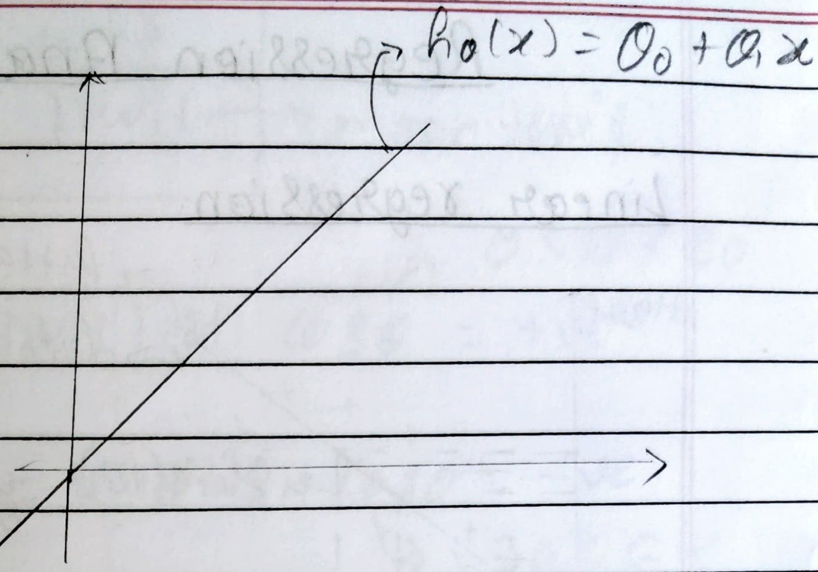
$$\boxed{J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x)_i - y_i \right)^2}$$

actual o/p

error (mean Squared error)

no of Data points you have

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



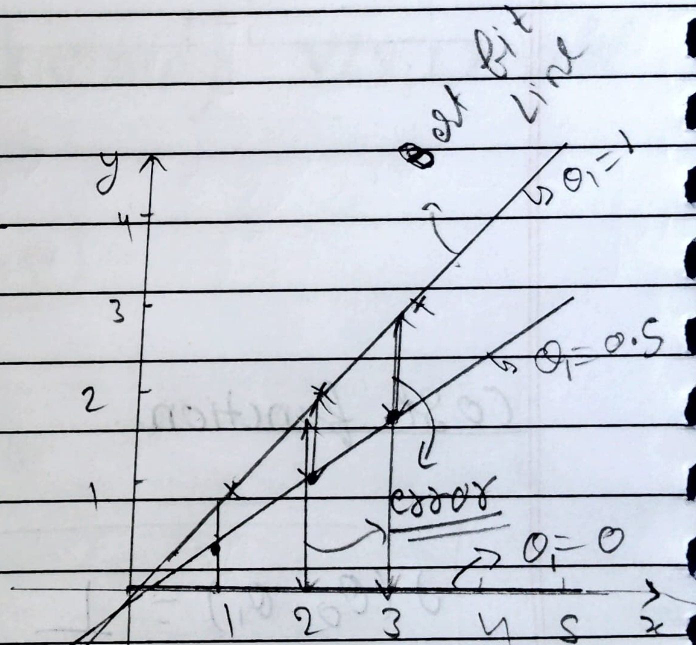
Let's take

$\boxed{\theta_0 = 0}$

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

<u>Input</u>

X	Y
1	1
2	2
3	3



let $\theta_1 = 1$

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

$x = 1, \; y = 1$
$x = 2, \; y = 2$
$x = 3, \; y = 3$

<u>cost function</u>: $J(\theta_1) = \frac{1}{2m} \sum_{i=0}^{m} \left(h_\theta(x)_i - y_i\right)^2$

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

$\Rightarrow \underline{0}$

Let $\theta_1 = 0.5$

$h_\theta(x) = 0.5$ if $x = 1$

$h_\theta(x) = 1$ if $x = 2$

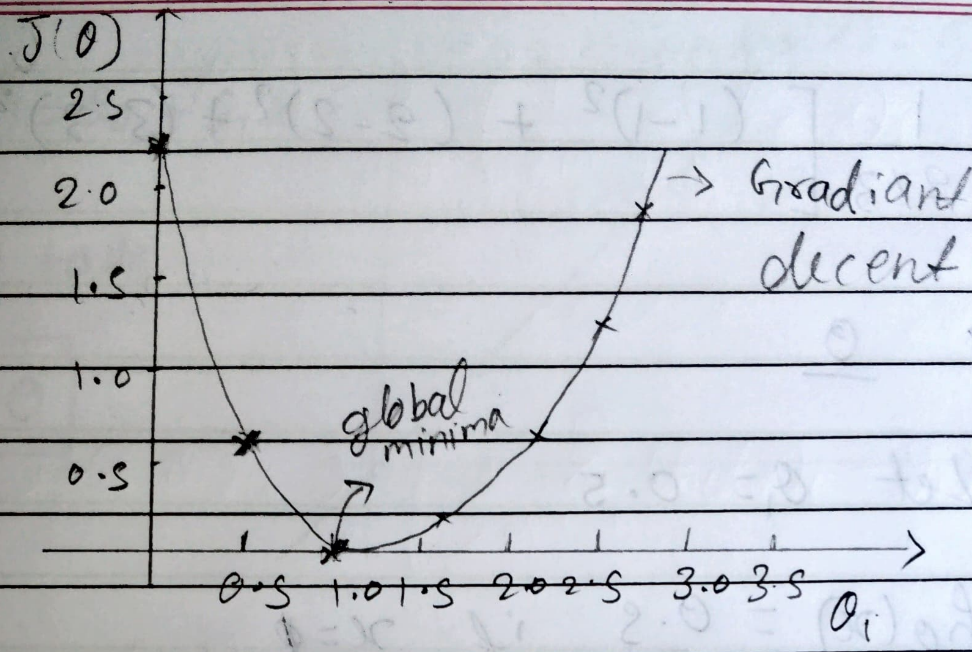
$h_\theta(x) = 1.5$ if $x = 3$

<u>cost function w.r.t $\theta_1 = 0.5$</u>;

$$J(\theta_1) = \frac{1}{2m}\sum_{i=0}^{m}(h_\theta(x)_i - y_i)^2$$

$$\Rightarrow \frac{1}{2\times3}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$$

$\Rightarrow \underline{0.58}$



Let $\theta_1 = 0$

$$J(\theta_1) = \frac{1}{6}\left[(0-1)^2 + (0-2)^2 + (0-3)^2\right]$$

$$J(\theta_1) = \underline{2.3}$$

Final Aim what we do / need to solve

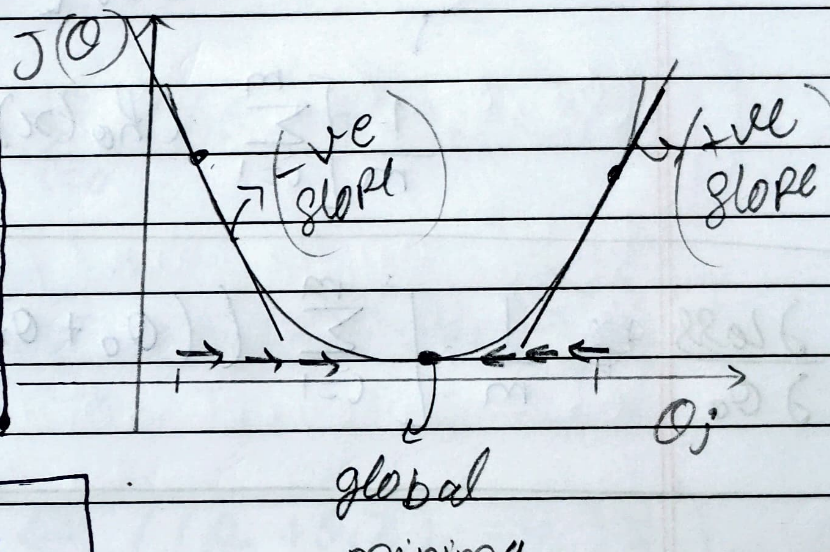
minimize $\{J(\theta_0, \theta_1) = \frac{1}{2m}\left[\sum_{i=0}^{m} (h_\theta(x)_i - y_i)^2\right]$

Convergence Algorithm {Optimization of θ

Repeat until convergence

learning rate (it should be small)

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



$$\boxed{\text{if } \frac{\partial J(\theta)}{\partial \theta_j} \text{ is -ve} \qquad \theta_j = \theta_j + \alpha \frac{\partial J(\theta)}{\partial \theta_j}}$$

$$\boxed{\text{if } \frac{\partial J(\theta)}{\partial (\theta)_j} \text{ is +ve} \qquad \theta_j = \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}}$$

So this θ_j is used to update the θ to minimize the error and try to findout the Best fit line.

$$\Rightarrow \left\{ \frac{\partial}{\partial \theta_j}\left(J(\theta_0,\theta_1)\right) = \frac{\partial}{\partial \theta_j}\left(\frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)_i - y_i\right)^2\right)\right.$$

if $J=0 \Rightarrow$

$$\frac{\partial}{\partial \theta_0} J(\theta_0,\theta_1) = \frac{\partial}{\partial \theta_0}\left[\frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)_i - y_i\right)^2\right]$$

$$\frac{1}{m}\left[\sum_{i=1}^{m}\left(h_\theta(x)_i - y_i\right)\right]$$

$$\boxed{\frac{\partial Loss}{\partial \theta_0} = \frac{1}{m}\left[\sum_{i=1}^{m}\left((\theta_0+\theta_1 x) - y_i\right)\cdot 1\right]}$$

$\because \frac{\partial(\theta_0+\theta_1 x)}{\partial \theta_0}$

$\Rightarrow 1+0=1$

if $J=1$

$$\frac{\partial}{\partial \theta_1} J(\theta_0,\theta_1) = \frac{\partial}{\partial \theta_1}\left[\frac{1}{2m}\sum_{i=0}^{m}\left(h_\theta(x)_i - y_i\right)^2\right]$$

$$\Rightarrow \frac{1}{m}\left[\sum_{i=0}^{m}\left((\theta_0+\theta_1 x) - y_i\right)\times x_i\right]$$

$\because \frac{\partial(\theta_0+\theta_1 x)}{\partial \theta_1} = x$

$$\boxed{\frac{\partial Loss}{\partial \theta_1} = \frac{1}{m}\left[\sum_{i=0}^{m}\left((\theta_0+\theta_1 x) - y_i\right) * x_i\right]}$$

⇒ Final equation for updating the θ_0 & θ_1 to finding the best fit line.

update the θ_0, θ_1 until the convergence

#

$$\theta_0 = \theta_0 - \alpha\left(\frac{1}{m}\left[\sum_{i=0}^{m}(h_\theta(x)_i - y_i)\right]\right)$$

$$\theta_0 = \theta_0 - \alpha\left(\frac{1}{m}\left[\sum_{i=0}^{m}((\theta_0 + \theta_1 x)_i - y_i)\right]\right)$$

$$\theta_1 = \theta_1 - \alpha\left(\frac{1}{m}\left[\sum_{i=0}^{m}(h_\theta(x)_i - y_i)\cdot x_i\right]\right)$$

$$\theta_1 = \theta_1 - \alpha\left(\frac{1}{m}\left[\sum_{i=0}^{m}((\theta_0 + \theta_1 x)_i - y_i)\cdot x_i\right]\right)$$

here $h_\theta(x)$ = the equation of line.

Multiple Linear Regression:

$h_\theta(x) = \theta_0 + \theta_1 x$ → Simple Linear Regression

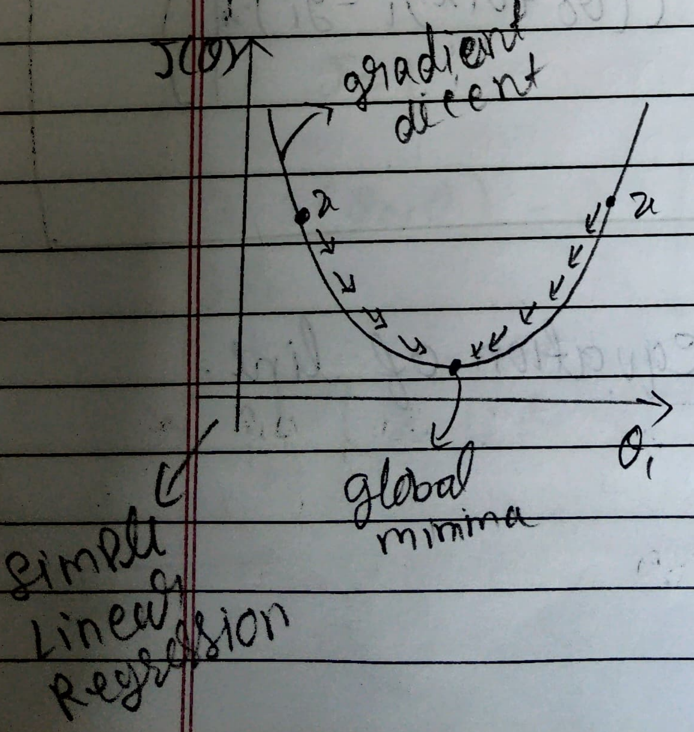
$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots \theta_n x_n$

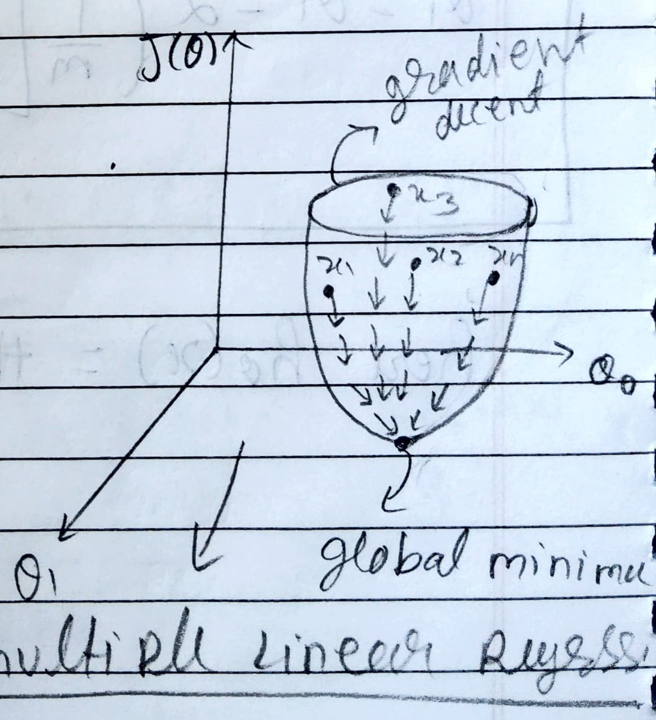
multiple Linear Regression

$(\theta_1, \theta_2, \theta_3, \cdots \theta_n)$ are the slopes / coefficients

based on the input features $(x_1, x_2, x_3, \cdots x_n)$

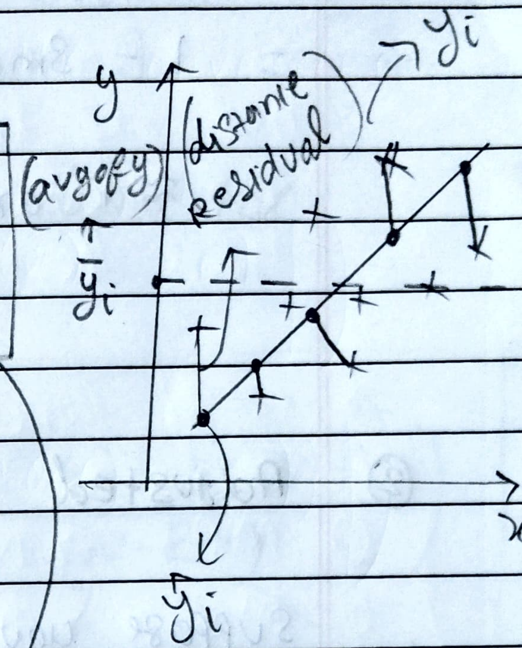
θ_0 = Intercept → and it is the same for all the input features.





Performance metrix used in Linear Regression.

(I) R-Squared.
(II) Adjusted R-Squared.



(I) $$\boxed{\text{R-Squared} = 1 - \frac{SS_{Res}}{SS_{Total}}}$$

Sum of Squared of residuals or errors
(SS_{res})

Sum of Squared of Total

predicted value

$$\boxed{\text{R-Squared} = 1 - \frac{\Sigma (y_i - \hat{y}_i)^2}{\Sigma (y_i - \bar{y}_i)^2}}$$

actual value

avg of y

$$= \left\{ 1 - \frac{\text{Small number}}{\text{Bigger number}} \right\}$$

= 1 - Small value

≈ 1 or 0.70 ⇒ 70%
0.85 ⇒ 85%
0.90 ⇒ 90%
} ⇒ accurate my model is.

② Adjusted R-squared :-

suppose you have a dataset with 2 features.

(i) size of the hous
(ii) Price
} ⇒ shows the positive co-relation.

I calculate our R-squared and get 75%.
thend i add one more features called (Location)

so based on location our R-squared value is 85%

but if we add the "~~age~~" Gender feature.

then we will calculate the R-Squared value it is 87% and we are know that there is no cor-relation between "Gender" and "price" of house prediction.)

to over come this we define the Adjusted R-Square ⇒

$$\text{Adjusted R-Square:- } 1 - \left(\frac{(1-R^2)(N-1)}{(N-P-1)} \right)$$

N ⇒ No of data points

P ⇒ No of Independent feature.

Let suppose a condition:

P = 2 $R^2 = 90\%$ R^2 Adjusted = 86%

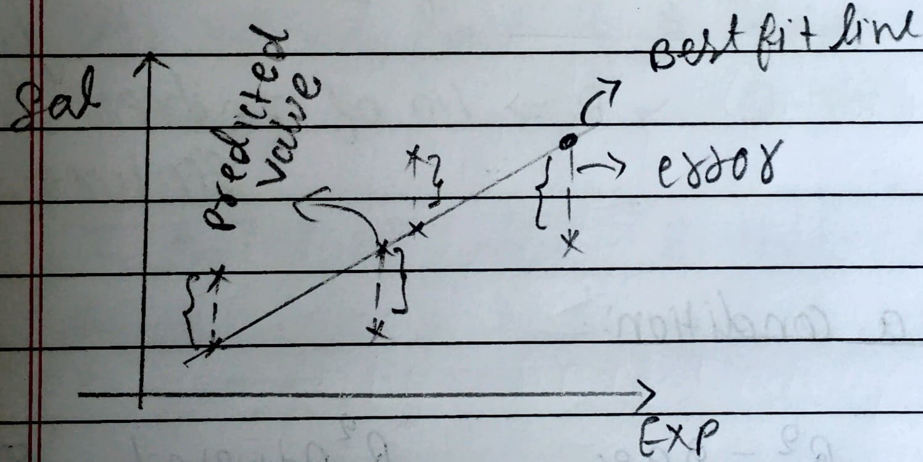
P = 3 $R^2 = 92\%$ R^2 Adjusted = 82%

in the P = 3 may be the feature is not Impacted or directly corelated to the output that's why we got the Less accuracy as compaired to the R^2 accuracy.

but if the features are dependent or corelated then it is increasing rather than decreasing.

⇒ cost functions → Performence matrixs

1. MAE → Mean Absolute Error.
2. MSE → Mean Squared Error.
3. RMSE → Root mean Squared Error.



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

Advantage:

(i) Differentiable

(ii) It has just one local and global minima

(iii) fast convergence

Disadvantage:

(i) not robust to outliers. (Penalise)

(ii) It is not longer in a same unit

(ii) Mean Absolute error:

$$\boxed{MAE = \frac{1}{n}\left(\sum_{i=1}^{n} |y_i - \hat{y}_i|\right)}$$

Advantage:

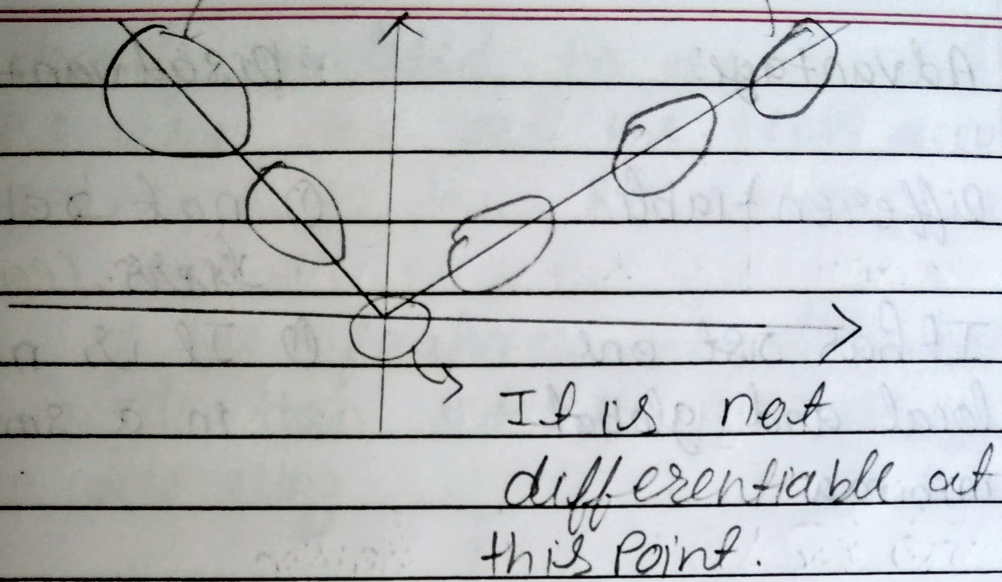
(i) Robust to outliers

(ii) It will be in same unit.

Disadvantage:

(i) Convergence take more time the optimization is complex task. because it is not differentiable at 0. it use the sub-gradient.

subgrediant



(III) RMSE :- Root mean Squared error

$$RMSE = \sqrt{MSE}$$

$$\Rightarrow \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Over fitting And Underfitting

(I) Training dataset
(II) Testing dataset
(III) Validation dataset.

1000 data points

Training Dataset (700) 70%

(300) 30% Testing Dataset

TRAIN → Train the model

Validation → Hyper Parameter Tunning your model

[Low Bias]

Train: very good acc

[Low variance]

Test: very good acc

⇓

Generalized model

[Low Bias]

very good acc (90%

Bad acc (50% -60%)

[High variance]

⇓

model is overfitting

Train: model acc is Low → [High Bias]

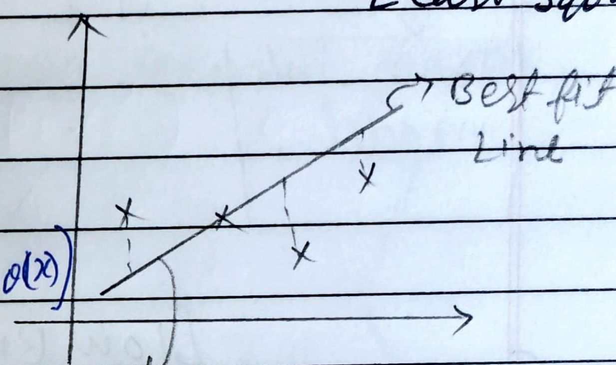
Test: model acc is Low → [High variance]

$\Downarrow$

model is underfitting.

⇒ Linear Regression using OLS {ordenary Least Square}

$$OLS = S(\beta_0, \beta_1)$$



$$S(\beta_0, \beta_1) = \frac{1}{n}\left(\sum_{i=1}^{n} (y_i - h_\theta(x)\right)$$

$$h_\theta(x) = \beta_0 + \beta_1 x_1$$

$$S(\beta_0, \beta_1) = \frac{1}{n}\left(\sum_{i=1}^{n} \underbrace{(y_i - \beta_0 - \beta_1 x_1)^2}_{\text{error}}\right)$$

finding out the β_0 & β_1

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = \frac{2}{n}\left(\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)(-1)\right)$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = \frac{-2}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \quad \text{--- (I)}$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = \frac{2}{n} \left(\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_1)(-x_1) \right) = 0$$

$$= \frac{-2}{n} \left(\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_1)(x_1) \right) = 0 \quad \text{--- (II)}$$

Taking eq (I)

$$-\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_1) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i + n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\Rightarrow \beta_0 = \underbrace{\frac{\sum_{i=1}^{n} y_i}{n}}_{\bar{y}} - \beta_1 \underbrace{\frac{\sum_{i=1}^{n} x_i}{n}}_{\bar{x}}$$

$$\boxed{\Rightarrow \beta_0 = \bar{y} - \beta_1 \bar{x}} \Rightarrow \underline{\text{Intercept}}$$

Taking eqn ②

$$\Rightarrow -\frac{2}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)(x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)(x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} (x_i) - \beta_1 \sum_{i=1}^{n} (x_i)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} (x_i y_i - \beta_0 x_i - \beta_1 x_i^2) = 0$$

$\Downarrow$

Replace with $(\beta_0 = \bar{y} - \beta_1 \bar{x})$

$$\Rightarrow \sum_{i=1}^{n} (x_i y_i - (\bar{y} - \beta_1 \bar{x}) x_i - \beta_1 x_i^2) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (x_i y_i - x_i \bar{y} + \beta_1 \bar{x} x_i - \beta_1 x_i^2) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \bar{y} + \beta_1 \bar{x}_i - \beta_1 x_i) \cdot x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} ((y_i - \bar{y}) + \beta_1 (\bar{x}_i - x_i)) = 0$$

classmate

Date

Page

$$\sum_{i=1}^{n} (y_i - \bar{y}) + \sum_{i=1}^{n} \beta_1 (\bar{x} - x_i) = 0$$

$$\sum_{i=1}^{n} \beta_1 (\bar{x} - x_i) = -\sum_{i=1}^{n} (y_i - \bar{y})$$

$$\Rightarrow \beta_1 = \frac{-\sum_{i=1}^{n} (y_i - \bar{y})}{\sum_{i=1}^{n} (\bar{x} - x_i)}$$

#

$$\boxed{\beta_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})}}$$ ⇒ equation for coefficient.

⇒ $\boxed{\beta_0 = \bar{y} - \beta_1 \bar{x}}$ ⇒ equation for Intercept

⇒ So based on the β_0 & β_1 values it is adjust the line to minimize the error and find the bestfit line.