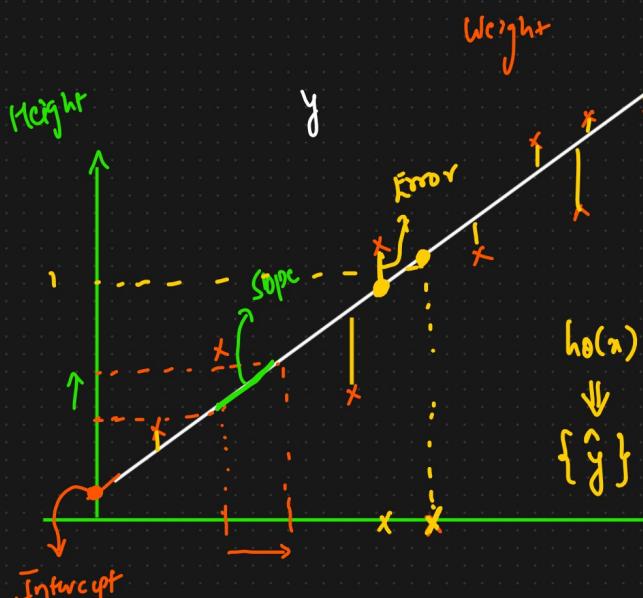
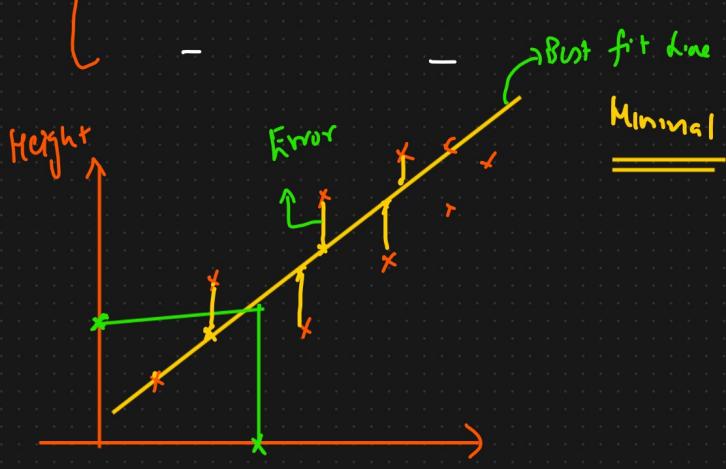
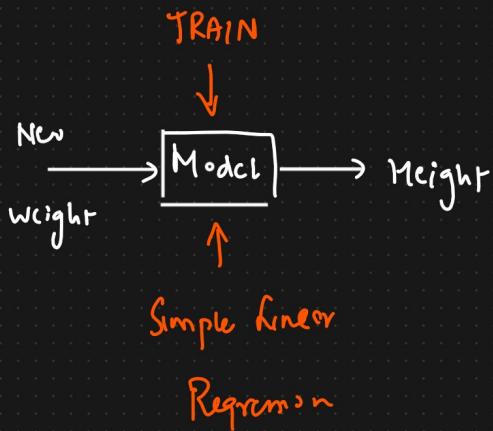


Simple Linear Regression

Supervised ML \rightarrow Regression

<u>Dataset</u>	IIP features
Weight	X
74	$y \uparrow$ O/P or dependant feature
80	Height 170cm
75	180cm
-	175.5cm



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

$$\theta_0 = \text{Intercept}$$

$$\theta_1 = \text{slope or coefficient}$$

$$\text{if } x=0 \quad \text{Error } (y - \hat{y})$$

$$h_{\theta}(x) = \theta_0$$



$$y = mx + c$$

$$y = \beta_0 + \beta_1 x$$

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

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$\dots + \theta_n x_n$$

Cost function

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right)^2 \Rightarrow \text{Mean Squared Error}$$

↑
predicted
↑
True O/P
Error

Final Aim: What we need to solve

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

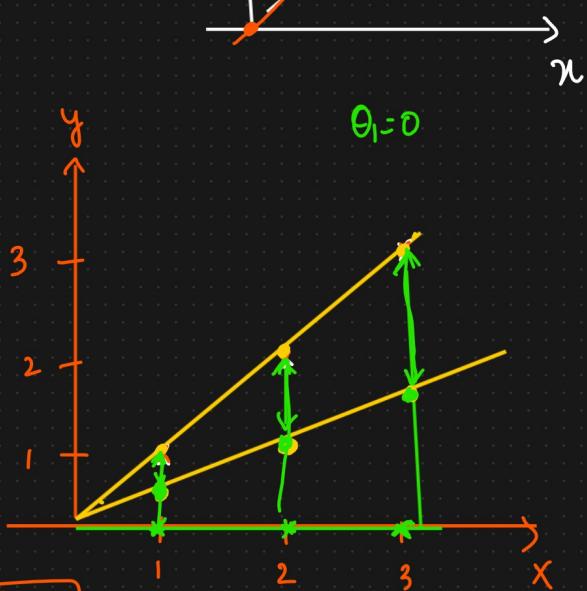
θ_0, θ_1

① $h_{\theta}(x) = \theta_0 + \theta_1 x$

$$\boxed{\theta_0 = 0}$$

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

<u>DATASET</u>	
x	y
1	1
2	2
3	3



$\theta_0 = 0$
 $h_{\theta}(x) = \theta_1 x$
 $\text{det } \theta_1 = 1 \quad \{ \text{slope} \}$

$h_{\theta}(x) = \theta_1 x$
 $\text{det } \theta_1 = 0.5$

$h_{\theta}(x) = 1 \quad x=1$

$h_{\theta}(x) = 0.5 \quad \text{if } x=1$

$h_{\theta}(x) = 2 \quad x=2$

$h_{\theta}(x) = 1 \quad \text{if } x=2$

$h_{\theta}(x) = 3 \quad x=3$

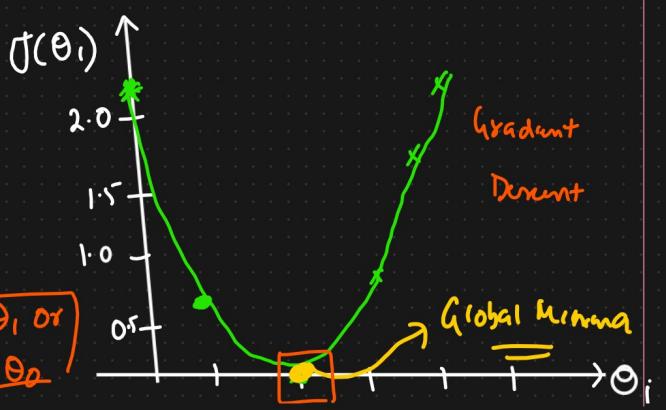
$h_{\theta}(x) = 1.5 \quad \text{if } x=3$

$$\boxed{\theta_1 = 1}$$

$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right)^2$$

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

$$\boxed{\theta_1, \text{ or } \theta_0}$$



$$J(\theta_1) = 0 \leftarrow$$

$$\underline{\underline{\theta_1}} = 0.5$$

Error has been
minimized

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

$$J(\theta_1) \approx 0.58$$

$$\underline{\underline{\theta_1}} = 0$$

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

$$J(\theta_1) \approx 2.3$$

$$\underline{\underline{=}}$$

Convergence Algorithm {Optimize the changes of
 θ_1 value}

Repeat until convergence

{

θ_1 value much
more efficiently

$$\left\{ \theta_j := \theta_j - \alpha \left[\frac{\partial J(\theta_j)}{\partial \theta_j} \right] \rightarrow -ve \right. =$$

}

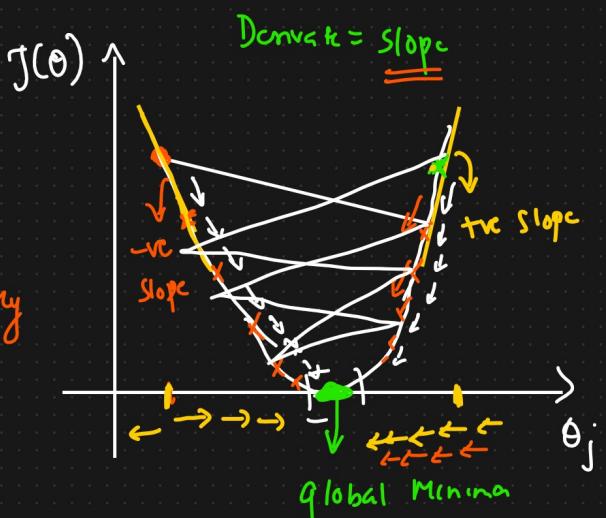
$$\theta_j = \theta_j - \alpha (+ve)$$

$$= \theta_j - (+ve)$$

α : Learning Rate

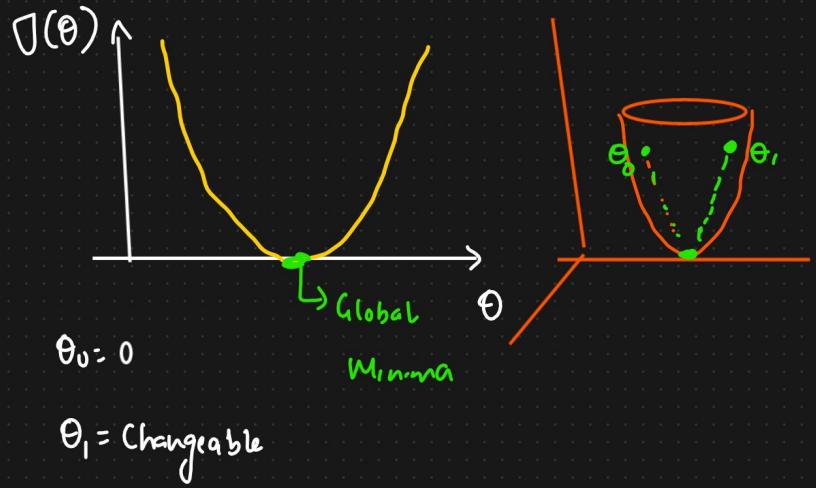
$$\left\{ \begin{array}{l} \theta_j = \theta_j - \alpha (-ve) \\ \theta_j = \theta_j + (+ve) \end{array} \right.$$

$$\boxed{\alpha = 0.001} \leftarrow$$



Final Conclusion

GRADIENT DESCENT



Convergence Algorithm

repeat until convergence

{

$$\theta_j := \theta_j - \alpha \left[\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \right]$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2$$

$$j = 0 \text{ and } 1 \quad \frac{\partial}{\partial x} (x)^2 = 2x$$

$$\frac{\partial}{\partial x} x^h = x x^{n-1} \quad \frac{\partial}{\partial x}$$

$$\rightarrow \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2$$

if

$$h_\theta(x) = \theta_0 + \boxed{\theta_1 x} \rightarrow 0$$

$$\begin{aligned} j=0 \Rightarrow \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) &= \frac{\partial}{\partial \theta_0} \frac{1}{2m} \left[\sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2 \right] \\ &= \frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) \times 1 \end{aligned}$$

$$j=1 \Rightarrow \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_1} \frac{1}{2m} \left[\sum_{i=1}^m ((\theta_0 + \theta_1 x) - y^{(i)})^2 \right]$$

$$= \frac{1}{m} \sum_{i=1}^m ((\theta_0 + \theta_1 x) - y^{(i)}) (x)$$

$$\boxed{\frac{\partial}{\partial \theta_1} [\theta_0 + \theta_1 x] \Rightarrow x =}$$

Repeat until Convergence

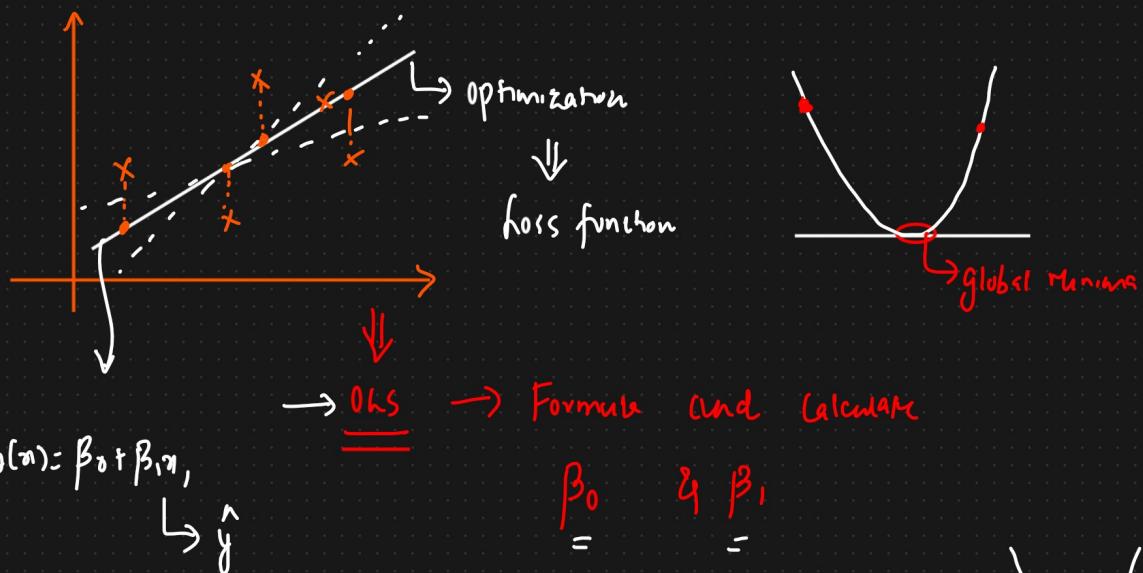
{

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x)^{(i)} - y^{(i)})$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x)^{(i)} - y^{(i)}) x^{(i)}$$

}

Linear Regression Using OLS {Ordinary Least Square}



Ordinary Least Square

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

$$\text{find } \beta_0, \quad \& \quad \beta_1$$

$$= \quad \quad \quad =$$

$$\begin{aligned} \frac{\partial S}{\partial \beta_0} (\beta_0, \beta_1) &= \frac{2}{n} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) (0 - 1 - 0) \\ &= -\frac{2}{n} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \rightarrow ① \end{aligned}$$

$$\begin{aligned} \frac{\partial S}{\partial \beta_1} (\beta_0, \beta_1) &= \frac{2}{n} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) (-x_i) = 0 \rightarrow ② \\ &= -\frac{2}{n} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) (x_i) = 0 \rightarrow ② \end{aligned}$$



$$y = x^2$$

$$\begin{aligned} \frac{\partial y}{\partial x} &= 2(x)^{2-1} \frac{\partial}{\partial x} (x) \\ &= 2x \end{aligned}$$

Eq → ①

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

↓

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

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

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

$$\beta_0 = \frac{\sum_{i=1}^n y_i - \beta_1 \sum_{i=1}^n x_i}{n}$$

Intercept
↓

$$\boxed{\beta_0 = \bar{y} - \beta_1 \bar{x}}$$

Eq 2 ÷

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

↓

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

↓

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

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

$$\text{Replace } \beta_0 = \bar{y} - \beta_1 \bar{x}$$

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

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

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

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

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

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

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

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

Coefficient \Rightarrow

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

$$\begin{array}{c}
 \text{Intercept} \\
 \Downarrow \\
 \boxed{\beta_0 = \bar{y} - \beta_1 \bar{x}}
 \end{array}
 \quad
 \begin{array}{c}
 \text{Ohs} \\
 \equiv \\
 \uparrow \\
 \text{Coefficient}
 \end{array}$$

$$\begin{array}{ccccccc}
 & & & \text{coefficient} & & & \\
 & & & \downarrow & & & \\
 x & y & (y_i - \bar{y}) / (x_i - \bar{x}) & \beta_1 & \beta_0 = (\bar{y} - \beta_1 \bar{x}) \\
 \hline
 & & & & & & \\
 & & & & & & \\
 & & & & & & \\
 & & & & & & \\
 & & & & & & \\
 & & & & & & \\
 & & & & & & \\
 & \bar{x} & \bar{y} & & & &
 \end{array}$$

Ohs ≈ linear Regression (sklearn)

Multiple Linear Regression

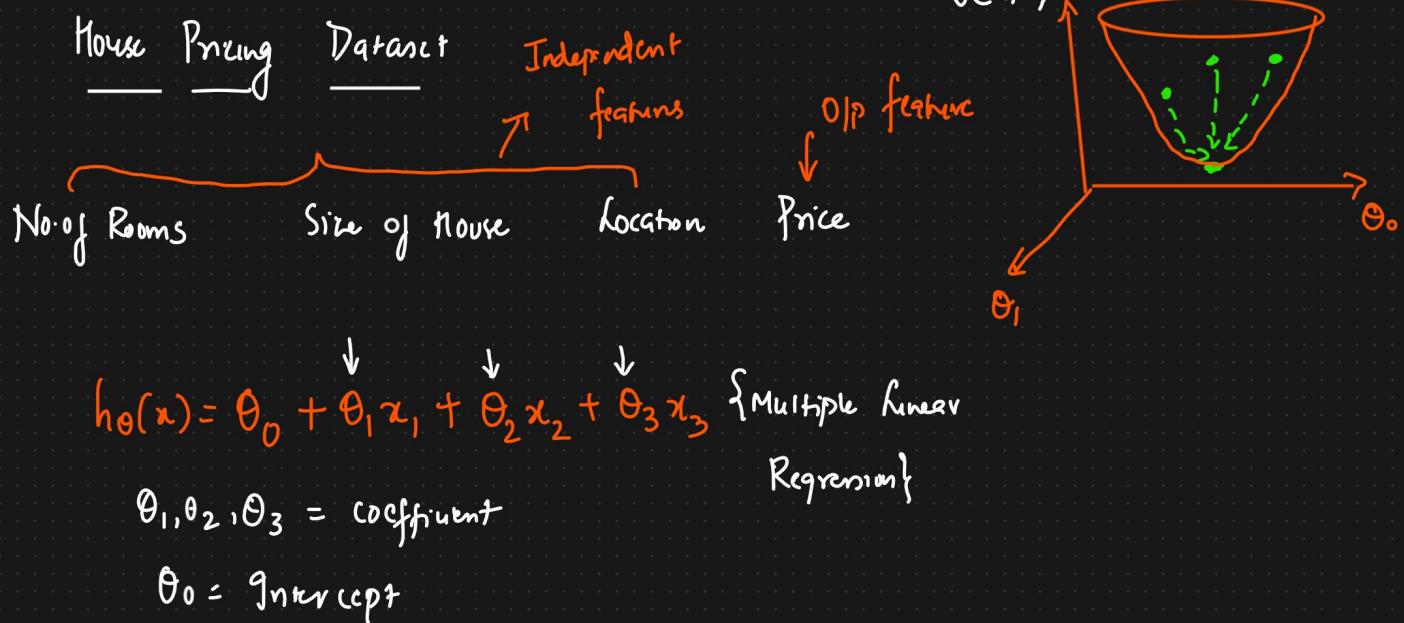
Dataset

Weight	Height
-	-
-	-

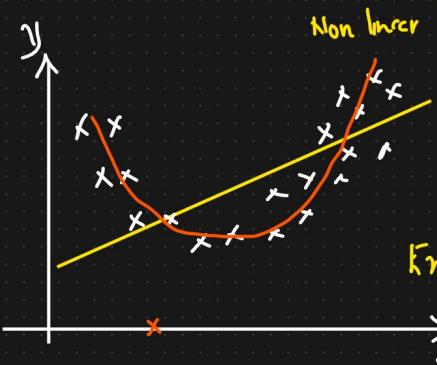
$$h_{\theta}(x) = \theta_0 + \theta_1 x^{\downarrow} \text{ I/P or Independent}$$

θ_0 = Intercept

θ_1 = Slope



Polynomial Regression



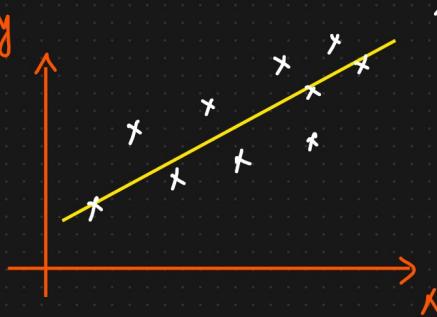
Non linear Relationship

$$h_{\theta}(x) = \beta_0 + \beta_1 x - \text{Simple Linear Regression}$$

$$h_{\theta}(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 - \text{Multiple linear Regression}$$

Error $\uparrow\uparrow$ Error $\downarrow\downarrow \rightarrow$ Polynomial Regression

Polynomial Degrees



\rightarrow Hyperplane degreec = 0

Simple Polynomial Regression {1 I/p and 1 o/p feature}

deg=0

polynomial degree=0

$$h_{\theta}(x) = \beta_0 * x^0 \Rightarrow \text{constant value}$$

polynomial degree=1

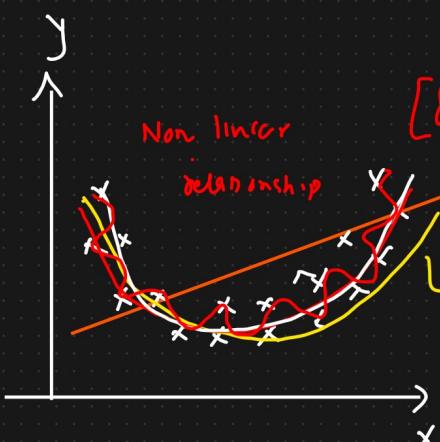
$$\boxed{h_{\theta}(x) = \beta_0 * x^0 + \beta_1 * x^1} \rightarrow \text{Simple Linear Regression}$$

polynomial degree=2

$$h_{\theta}(x) = \beta_0 * x^0 + \beta_1 * x^{(1)} + \beta_2 * x^{(2)}$$

polynomial degree=n

$$\boxed{h_{\theta}(x) = \beta_0 * x^0 + \beta_1 * x^{(1)} + \beta_2 * x^{(2)} + \beta_3 * x^{(3)} + \dots + \beta_n * x^{(n)}}$$



Non linear

relatnship

[degree \rightarrow values]

degree = 1

\hookrightarrow

\curvearrowleft

{2 independent feature}

degree = 1

$$h_{\theta}(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

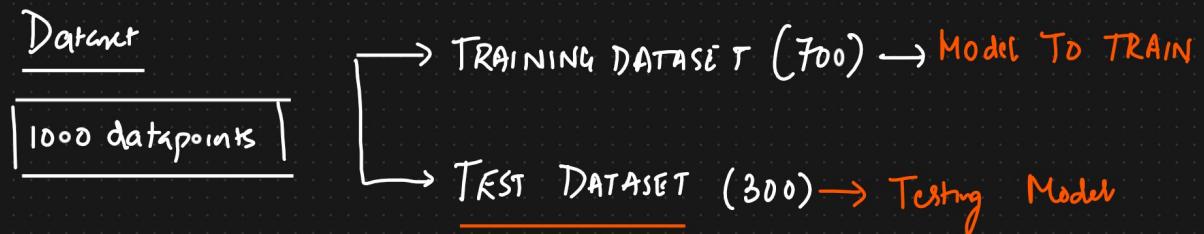
$\boxed{\text{degree} = 2}$

$$h_{\theta}(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2$$

Overfitting And Underfitting (Bias And Variance)

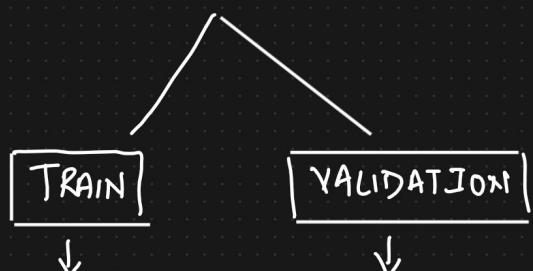
- ① Training dataset
- ② Test dataset
- ③ Validation dataset

For 30%



Size of House	No. of bedrooms	Price
-	-	-
-	-	-
-	-	-
-	-	-

TRAINING DATASET



↓
Train the model

↓
Hyperparameters

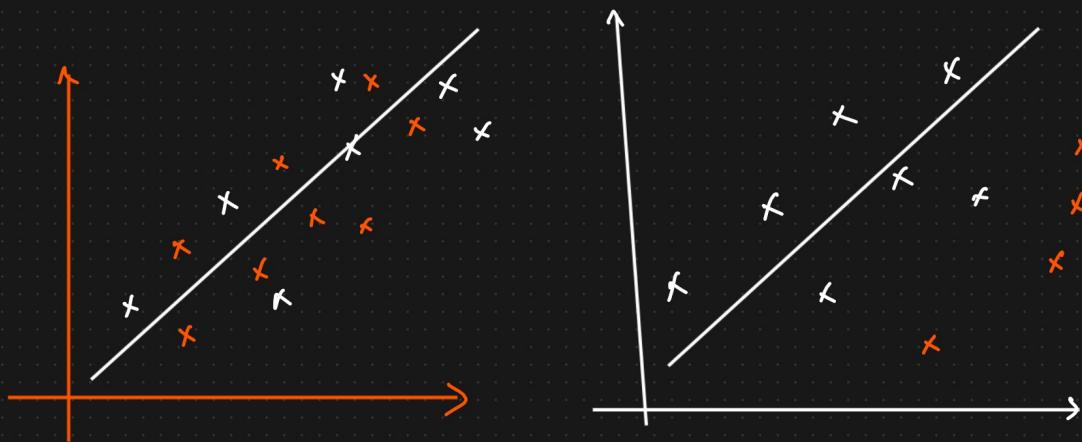
Tunning your
Model

TRAIN	(90%)	Very Good Accuracy [low Bias]	Very Good Accuracy (90%) [low Bias]
TEST	Very Good Accuracy [low Variance] (85%) ↑	Bad Accuracy (50%) ↓	[High Variance]
→ Generalized Model		Model is Overfitting	

TRAIN Model Accuracy is low [High Bias]

TEST Model Accuracy is low [High Variance]
↓

Model is Underfitting



Generalized Model

↓
Low Bias, Low Variance

Overfitting

Low Bias, High Variance

Performance Metrics Used In Linear Regression

① R squared

② Adjusted R squared

$$R^2 = 1 - \frac{SS_{\text{Res}}}{SS_{\text{TOTAL}}}$$

$$= 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

$$= 1 - \frac{\text{Small number}}{\text{Big number}}$$

$$= 1 - \text{Small number}$$

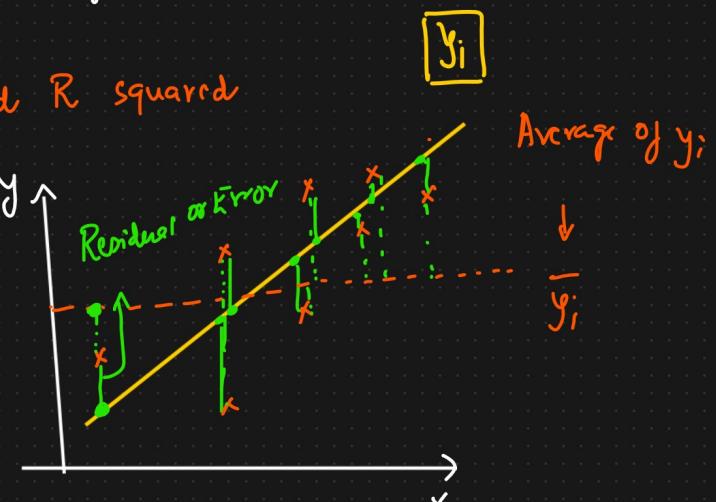
$$\approx 1$$

$$0.70 \Rightarrow 70\%$$

$$0.85 \Rightarrow 85\%$$

$$0.90 \Rightarrow 90\%$$

{ Overfitting, Underfitting }



1

↓

Accurate My
Model Is?

R squared ↑↑↑

Size of the house ↑ Price ↑

tve correlation

② Adjusted R squared

Dataset

→ (Price)

↖

Price

Gender

Size of the house No. of bedrooms location

No. of bedrooms ↑ Price ↑

tve correlation

$$R^2 = 75\% \Rightarrow 0.75$$

This is the problem of R squared

$$R^2 \text{ squared} \Rightarrow 80\% \Rightarrow 0.80$$

$$R^2 \text{ squared} \Rightarrow 85\% \Rightarrow 0.85$$

$$R^2 \text{ squared} \Rightarrow 87\% \Rightarrow 0.87$$

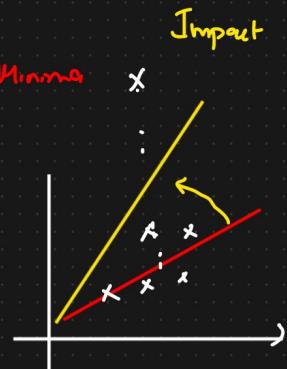
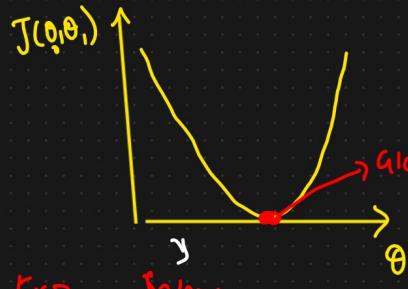
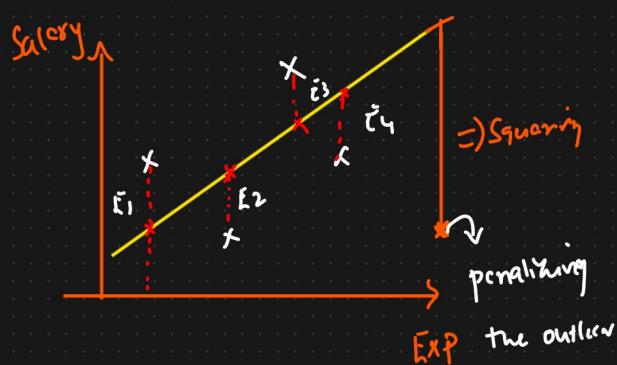
$$\text{Adjusted } R^2 \text{ squared} = 1 - \frac{(1-R^2)(N-1)}{N-p-1}$$
$$\left\{ \begin{array}{lll} p=2 & R^2 = 90\% & R^2 \text{ adjusted} = 86\% \\ p=3 & R^2 = 92\% & R^2 \text{ adjusted} = 82\% \end{array} \right.$$

N = No. of data points

p = No. of Independent features

MSE, MAE, RMSE [Cost function] → Performance Metrics

R^2 and Adjusted R^2



$NSE \uparrow \uparrow$

MSE

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \Rightarrow \text{Cost function } \downarrow \downarrow$$

Quadratic Equation

$$ax + by + c$$

$$(a-b)^2 = a^2 - 2ab + b^2$$

Advantage

- ① Differentiable ✓
- ② JT has one local and one global Minima

Disadvantage

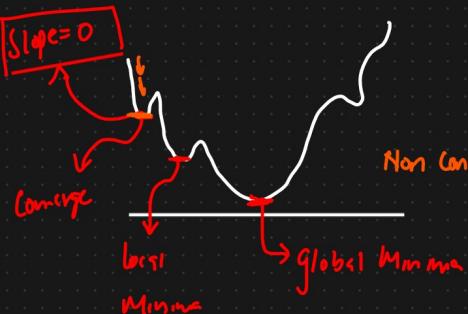
- ④ Not Robust to outliers
- ④ It is not in same unit.



③ Convex function

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

$$\text{Salary (Lakhs)} \quad \underline{y} \quad \underline{(y_i - \hat{y}_i)^2} \quad \underline{\text{MSE}}$$



Non-convex functions

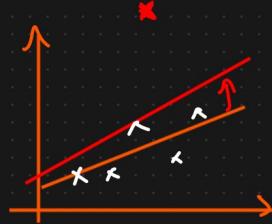
$$\text{Error} < 2.5 \Rightarrow (10\text{K})^2 \leq$$

② Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

factors

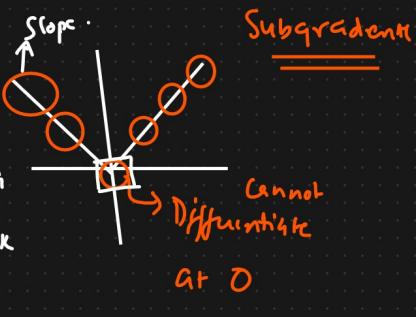
MAE↑↑



Advantage

- ① Robust to outliers ✓
- ② It will be in the same unit
- ④ Convergence usually take more time. Optimization is a complex task
- ⑤ Time consuming

Disadvantage



③ RMSE {Root Mean Square Error}

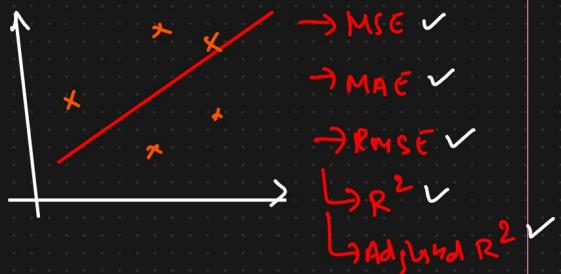
$$\text{RMSE} = \sqrt{\text{MSE}}$$

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



MAD, MSE, RMSE

Performance metric



Advantage

- ① Same unit
- ② Differentiable

Disadvantage

- ④ Not Robust to outliers

MSE vs MAE vs RMSE

R^2 vs Adjusted R^2