Simple Linear Regression

Supervised ML $\longrightarrow$ Regression

TRAIN

$\downarrow$

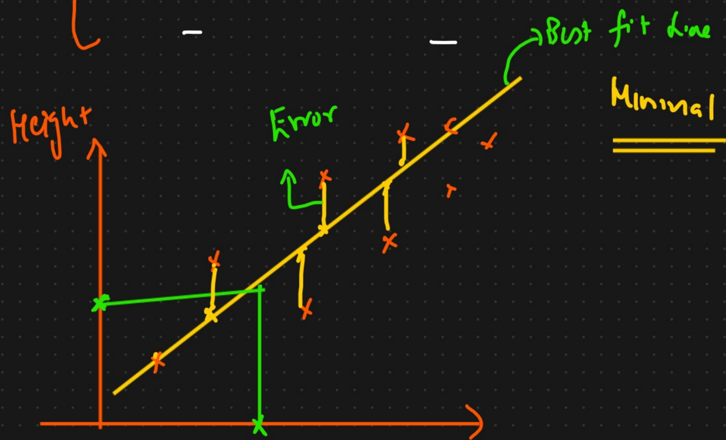
New weight $\longrightarrow$ $\boxed{\text{Model}}$ $\longrightarrow$ Height

$\uparrow$

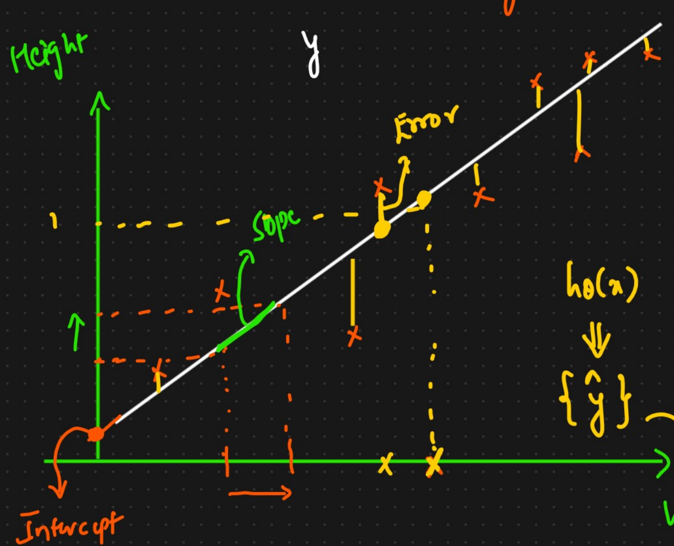
Simple Linear Regression

<u>Dataset</u> I/P features

$\uparrow$ X <u>Weight</u>	Y <u>Height</u> $\uparrow$ O/P or dependent feature
74	170cm
80	180cm
75	175.5cm
–	–



Minimal



$h_\theta(x)$

$\Downarrow$

$\{\hat{y}\}$ $\longrightarrow$ predicted point

θ_0 & θ_1 Minimal Error

Dataset

$y = mx + c$

$y = \beta_0 + \beta_1 x$

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

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

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

if $x = 0$

$h_\theta(x) = \theta_0$

Error $(y - \hat{y})$

θ_0 = Intercept

θ_1 = slope or coefficient

Cost function

predicted ⇑ True O/P ⇑

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

↳ Error

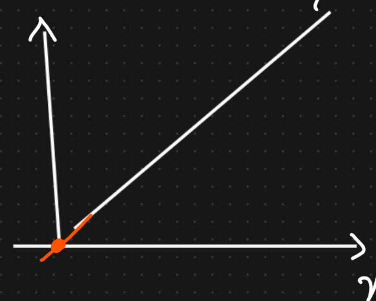
Final Aim what we need to solve

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

DATASET

x	y
1	1
2	2
3	3

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



$\boxed{\theta_0 = 0}$

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

$\theta_0 = 0$

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

Let $\theta_1 = 1$ {slope}

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

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

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

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

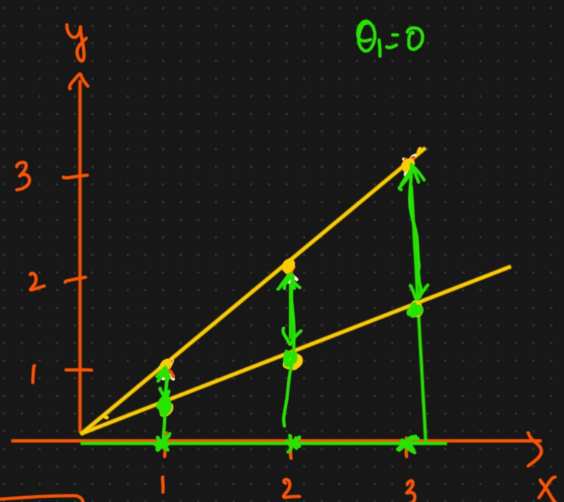
↓

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 $x = 3$

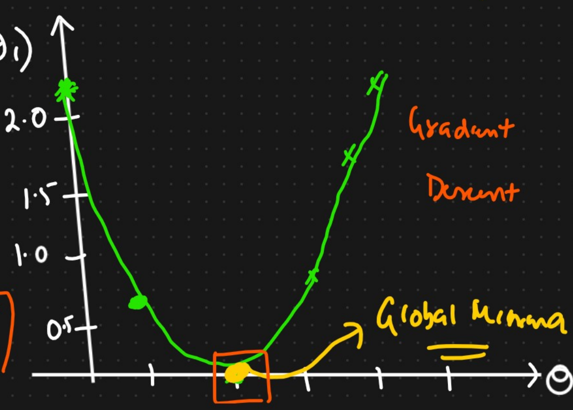


$\boxed{\theta_1 = 1}$

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

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

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



$J(\theta_1) = 0$ ←

$\theta_1 = 0.5$

Error has been minimized

0.5 1 1.5 2.0 2.5

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

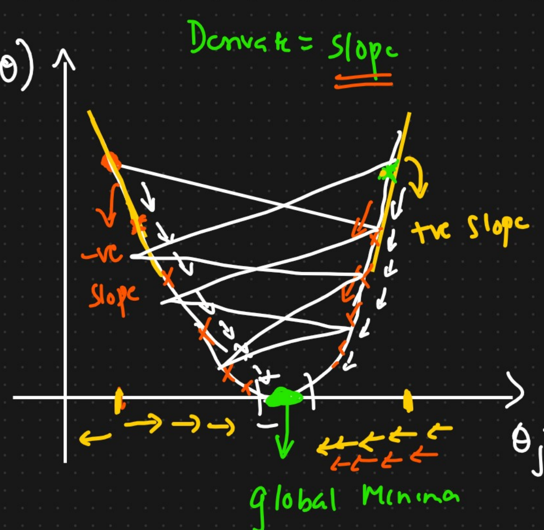
$$J(\theta_1) \approx 0.58$$

if $\theta_1 = 0$

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

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

Convergence Algorithm {Optimize the changes of θ_1 value}



Repeat until convergence

θ_1 value much more efficiently

{

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

}

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

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

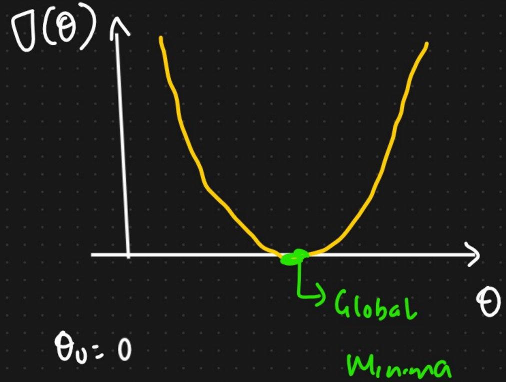
$$\left\{ \begin{aligned} \theta_j &= \theta_j - \boxed{\alpha}(-ve) \\ \theta_j &= \theta_j + (+ve) \end{aligned} \right\}$$

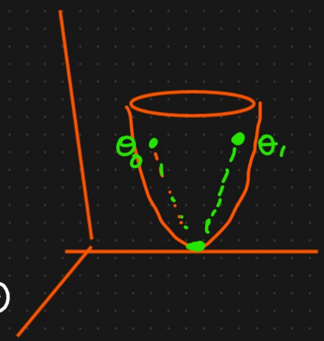
$$\boxed{\alpha = 0.001} \leftarrow$$

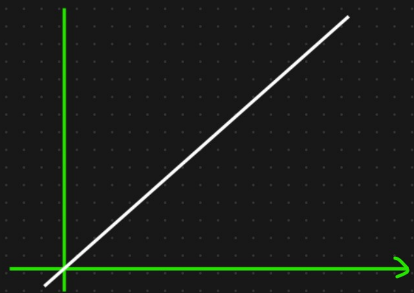
α: Learning Rate

Final Conclusion

GRADIENT DESCENT







$\theta_0 = 0$

$\theta_1 =$ Changeable

Convergence Algorithm

repeat until convergence

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

{

$$\theta_j := \theta_j - \alpha \boxed{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)}$$

}

$j = 0$ and 1

$$\frac{\partial}{\partial x} (x)^2 = 2x$$

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

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

if

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

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

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

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

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

$$\boxed{\frac{\partial}{\partial \theta_1}\left[\theta_0 + \theta_1 x\right] \Rightarrow \underline{\underline{x}}}$$

Repeat until Convergence

$$\left\{ \begin{aligned} \theta_0 &:= \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right) \\ \theta_1 &:= \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right) x^{(i)} \end{aligned} \right\}$$

Multiple Linear Regression

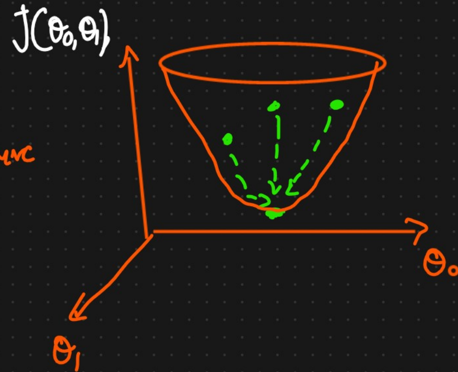
$h_\theta(x) = \theta_0 + \theta_1 x$ ← I/p or Independent

θ_0 = Intercept

θ_1 = Slope

<u>Dataset</u>

I/p	O/p feature
Weight	Height
—	—
—	—



<u>House</u> <u>Pricing</u> <u>Dataset</u>

Independent features

O/p feature

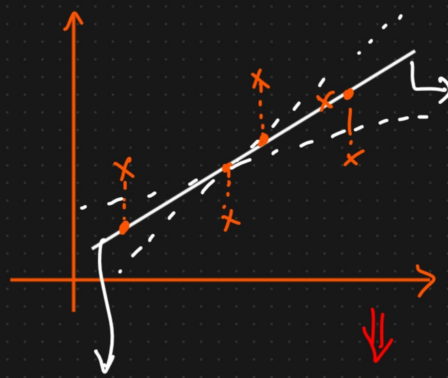
No.of Rooms   Size of House   Location   Price

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

$\theta_1, \theta_2, \theta_3$ = coefficient

θ_0 = Intercept

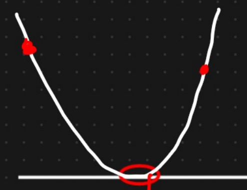
Linear Regression Using OLS {Ordinary Least Square}



↳ optimization

⇓

Loss function



⇓

→ OLS → Formula and calculate

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

↳ $\hat{y}$

β_0 & β_1

Ordinary Least Square

Error ⇑ = MSE

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

⇓

Aim β_0, & β_1

$y = x^2$

$$\frac{\partial y}{\partial x} = 2(x)^{2-1} \frac{\partial (x)}{\partial x}$$

$$= 2x$$

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

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

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

$$= -\frac{2}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_1)(x_1) = 0 \longrightarrow (2)$$

Eq → ①

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

$$\Downarrow$$

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

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

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

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

Intercept →

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

$\boxed{\text{Eq } 2 \div}$

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

$$\Downarrow$$

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

$$\Downarrow$$

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

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

$$\Downarrow$$

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

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

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

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

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

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

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

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

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

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

Intercept

$\Downarrow$

OLS

Coefficient

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

Coefficient

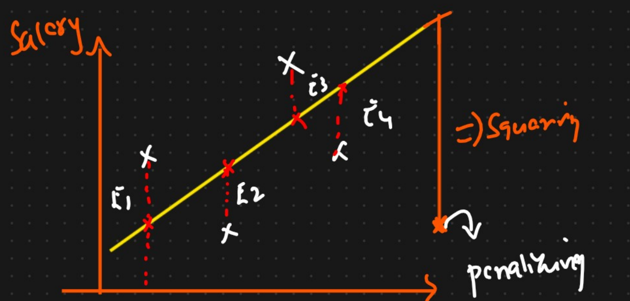
Intercept

x	y	$(y_i - \bar{y}) \div (x_i - \bar{x})$		β_1	$\beta_0 = (\bar{y} - \beta_1 \bar{x})$
–	–	–	–		
–	–	–	–		
–	–	–	–		
–	–	–	–		
$\bar{x}$	$\bar{y}$				

OLS $\approx$ Linear Regression (sklearn)

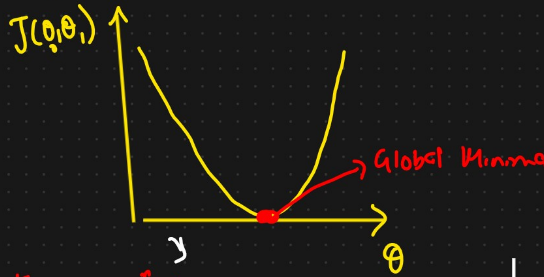
MSE, MAE, RMSE [Cost function] → Performance Metrics

R^2 and Adjusted R^2

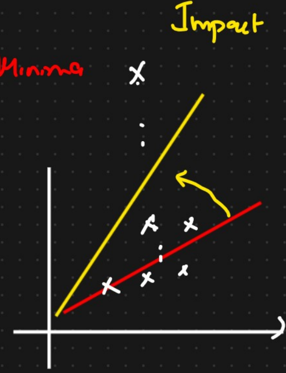


⇒ Squaring

penalizing the outlier



Exp	Salary
—	—
—	—
—	—

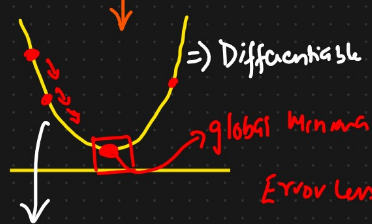


$\boxed{MSE\uparrow\uparrow}$

MSE

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

→ Quadratic Equation



Convex function

$ax+by+c$

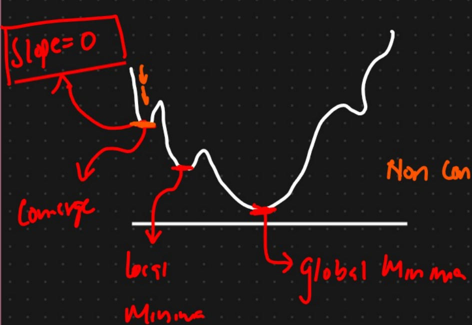
$\boxed{(a-b)^2 = a^2 - 2ab + b^2}$

Advantage

1. Differentiable ✓
2. It has one local and one global Minima
3. Converges faster

Disadvantage

1. Not Robust to outliers
2. It is not in same unit.



Non Convex functions

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

⇓ $(lakh)^2$

Salary (lakhs)

X	y	$(y_i-\hat{y}_i)^2$ (MSE)

Error c 2.5 ⇒ $(lakh)^2$ ⇐

② Mean Absolute Error (MAE)

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

lakhs

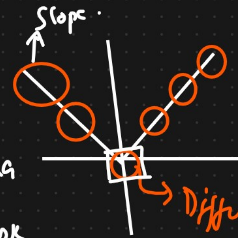
MAE↑↓

Advantage

1. Robust to outliers ✓
2. It will be in the same unit

Disadvantage

1. Convergence usually take more time. Optimization is a complex task
2. Time consuming



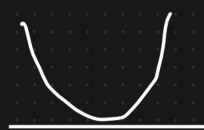
Subgradient

Cannot Differentiate at 0

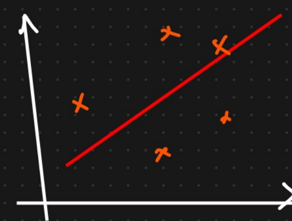
3. RMSE {Root Mean Squared Error}

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

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



$\boxed{\text{MAE, MSE, RMSE}}$ ⇑ Performance metric



→ MSE ✓
→ MAE ✓
→ RMSE ✓
↳ R^2 ✓
↳ Adjusted R^2 ✓

Advantage

1. Same unit
2. Differentiable

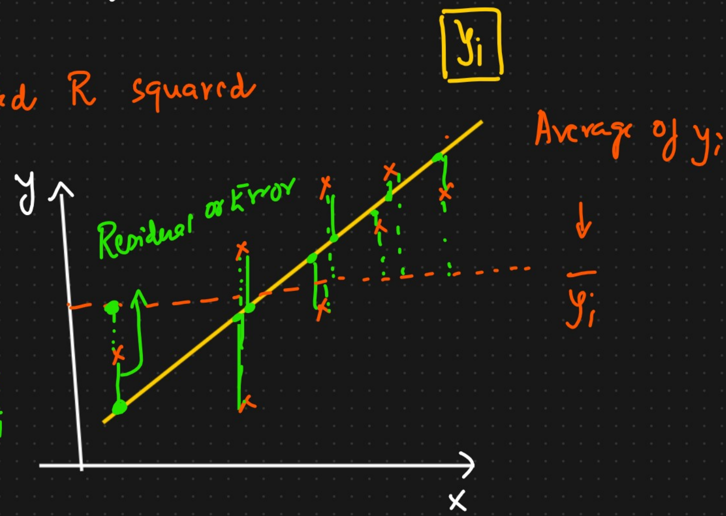
Disadvantage

1. Not Robust to outliers

$$\left.\begin{array}{l} \text{MSE vs MAE vs RMSE} \\ \\ R^2 \text{ Vs Adjusted } R^2 \end{array}\right\}.$$

Performance Metrics Used In Linear Regression

① R squared   ② Adjusted R squared



$$R\text{squared} = 1 - \frac{SS_{Res}}{SS_{TOTAL}} \qquad \hat{y}_i$$

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

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

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

$$\approx 1 \qquad \boxed{1}$$

⇓

Accurate My Model Is?

0.70 ⇒ 70%

0.85 ⇒ 85%

{Overfitting, Underfitting}   0.90 ⇒ 90%

R squared ↑↑↑

Size of the house ↑ Price ↑

+ve Correlation

② Adjusted R squared

Dataset → (Price)

No. of bedroom ↑   Price ↑

+ve correlation

↓

Gender	Size of the house	No. of bedrooms	Location	Price

R squared = 75% ⇒ 0.75

This is the problem of R Squared

R squared $\Rightarrow$ 80% $\Rightarrow$ 0.80

R squared $\Rightarrow$ 85% $\Rightarrow$ 0.85

R squared $\Rightarrow$ 87% $\Rightarrow$ 0.87

$$\text{Adjusted R squared} = 1 - \frac{(1-R^2)(N-1)}{N-P-1}$$

$$\begin{cases} P=2 & R^2 = 90\% \quad R^2 \text{ adjusted} = 86\% \\ P=3 & R^2 = 92\% \quad R^2 \text{ adjusted} = 82\% \end{cases}$$

N = No. of data points

P = No. of Independent features

Overfitting And Underfitting (Bias And Variance)

1. Training dataset
2. Test dataset
3. Validation dataset

70%-30%

Dataset

$\boxed{\text{1000 datapoints}}$ → TRAINING DATASET (700) → Model To TRAIN

→ TEST DATASET (300) → Testing Model

Size of House	No. of bedrooms	Price
—	—	—
—	—	—
—	—	—
—	—	—

TRAINING DATASET

$\boxed{\text{TRAIN}}$ → Train the model

$\boxed{\text{VALIDATION}}$ → Hyperparameter Tunning your Model

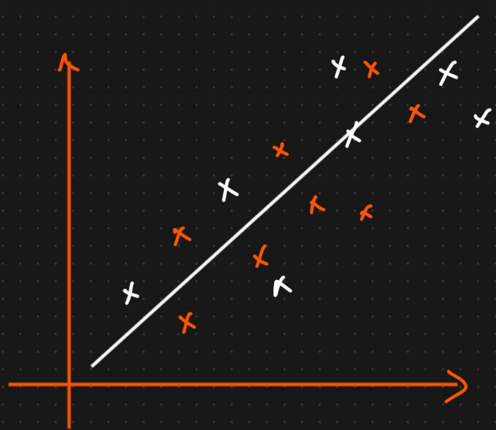
TRAIN	Very Good Accuracy (90%) [Low Bias]	Very Good Accuracy (90%) [Low Bias]
TEST	Very Good Accuracy (85%) [Low Variance] $\Uparrow$	Bad Accuracy (50%) [High Variance] $\Downarrow$
	→ Generalized Model	Model is Overfitting

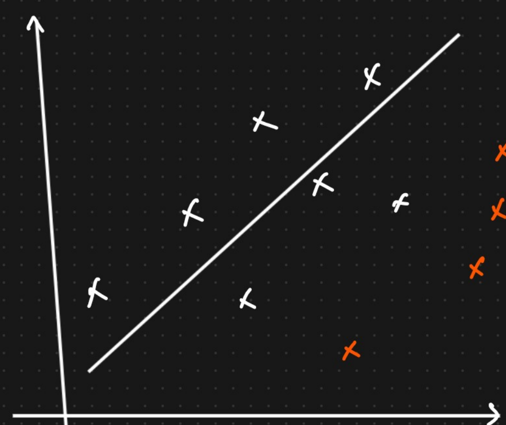
TRAIN   Model Accuracy is Low [High Bias]

TEST   Model Accuracy is Low [High Variance]

$\Downarrow$

Model is Underfitting





Generalized Model

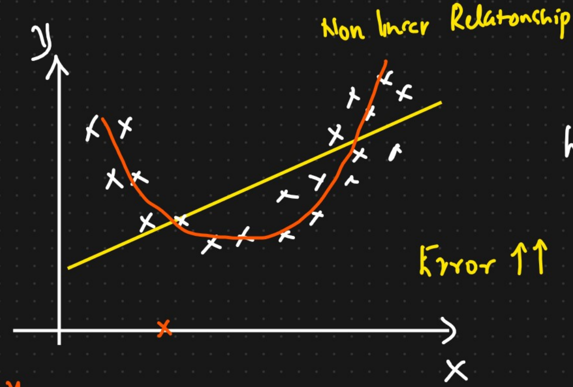
$\Downarrow$

Low Bias, Low Variance

Overfitting

Low Bias, High Variance

Polynomial Regression



$h_\theta(x) = \beta_0 + \beta_1 x$ — Simple Linear Regression

$h_\theta(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ — Multiple Linear Regression

Error ↑↑ Error ↓↓ ⟶ Polynomial Regression

Polynomial Degrees

→ Hyperplane

degree = 0

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

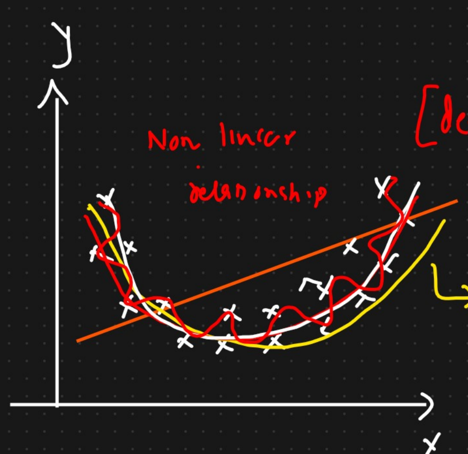
deg=0

polynomial degree=0 $h_\theta(x) = \beta_0 \times x^0$ ⇒ Constant value

polynomial degree=1 $\boxed{h_\theta(x) = \beta_0 * x^0 + \beta_1 * x^{(1)}}$ → Simple linear Regression

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

polynomial degree=n $\boxed{h_\theta(x) = \beta_0 * x^0 + \beta_1 * x^1 + \beta_2 * x^{(3)} + \beta_3 * x^{(3)} + \dots \beta_n * x^{(4)}}$



[degree → values]

degree = 15

↳ degree = 1

{2 independent feature}

degree = 1 $h_\theta(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

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