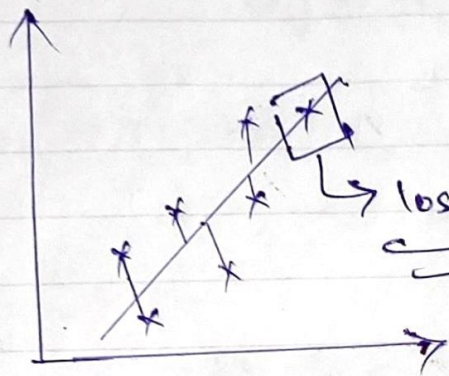


$h(x) = \theta_0 + \theta_1 x \rightarrow 1$ Independent 1 Dependent

$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots \rightarrow$ Multiple independent
Dependent

cost function (MSE) :



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

\downarrow predicted \downarrow truth point
 (with $m \rightarrow$ no. of data-points above the sum)

convergence Algorithm:

~~Every~~ ~~Obstac~~ Every Observation [loss vs cost function]

$$\text{loss function} = (h(x)^{(i)} - y^{(i)})^2 \Rightarrow (\hat{y}_0 - y^{(i)})^2$$

\downarrow predicted \downarrow Actual

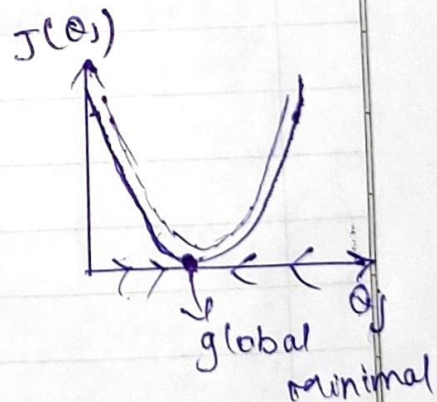
convergence Algorithm:

Repeat until convergence

{

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

}



$$\frac{\partial}{\partial Q_0} J(Q_0, Q_1) = \frac{\partial}{\partial Q_0} \left[\frac{1}{m} \sum_{i=1}^m (h_0(x)^{(i)} - y^{(i)})^2 \right]$$

J=0

$$\therefore h_0(x) = Q_0 + Q_1 x$$

$$= \frac{\partial}{\partial Q_0} \left[\frac{1}{m} \sum_{i=1}^m (Q_0 + Q_1 x)^{(i)} - y^{(i)} \right]^2$$

$$\frac{2}{2m} \sum_{i=1}^m (h_0(x) - y^{(i)})$$

$$= \frac{2}{m} \sum_{i=1}^m [(Q_0 + Q_1 x) - y^{(i)}] \times 1$$

J=1

$$= \frac{\partial}{\partial Q_1} \left[\frac{1}{m} \sum_{i=1}^m ((Q_0 + Q_1 x))^{(i)} - y^{(i)} \right]^2$$

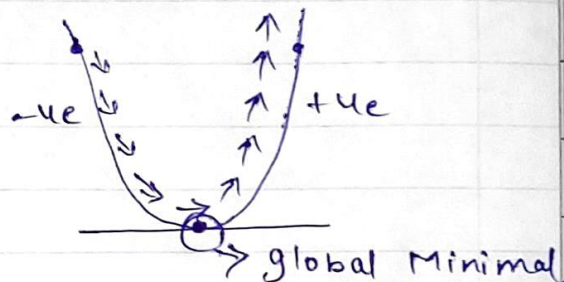
$$= \frac{2}{m} \sum_{i=1}^m \left[(Q_0 + Q_1 x^{(i)}) - y^{(i)} \right] \cdot [X]$$

Repeat until convergence

α = speed of convergence

$$\begin{cases} Q_0 := Q_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h(x)^{(i)} - y^{(i)}) \\ Q_1 := Q_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h(x)^{(i)} - y^{(i)}) x^{(i)} \end{cases}$$

$$Q_1 := Q_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h(x)^{(i)} - y^{(i)}) x^{(i)}$$



MSE { Mean Squared Error }

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

$\frac{1}{n}$ \rightarrow Quadratic Equation.

$$(a-b)^2 = a^2 - 2ab + b^2 \Rightarrow \text{Linear Algebra}$$

\downarrow

$$ax^2 + by + c = 0$$

$$\hat{y} = Q_0 + Q_1 x \Rightarrow \text{Predicted value}$$

Advantage

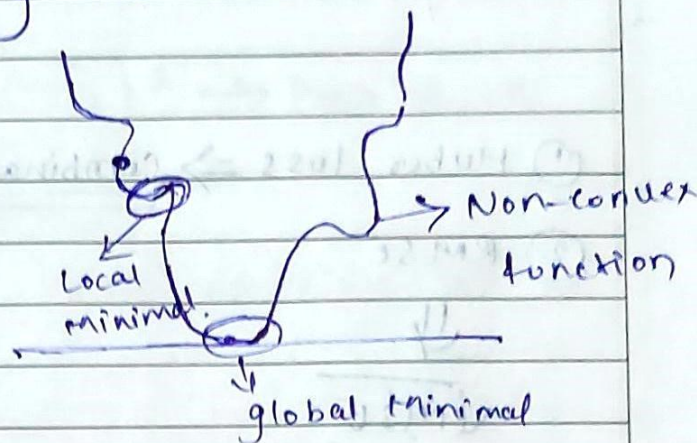
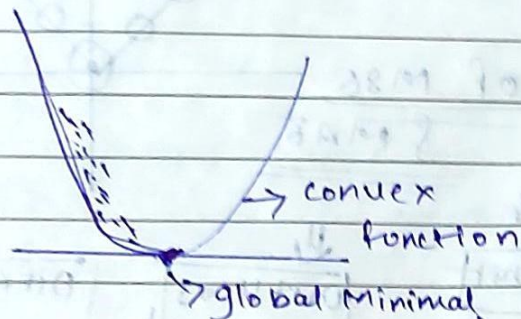
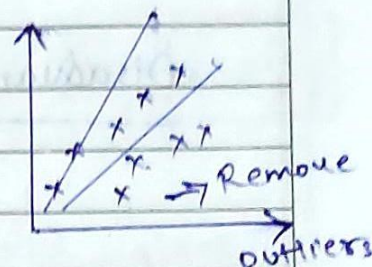
- (1) This equation is differentiable
- (2) This equation also has one global Minimal.

cost function

↑ ↑ ↑

Disadvantage

- (1) This is not robust to Outliers.
- (2) penalizing the Error changing the unit



x	y
Exp	salary
—	—
—	—
—	—

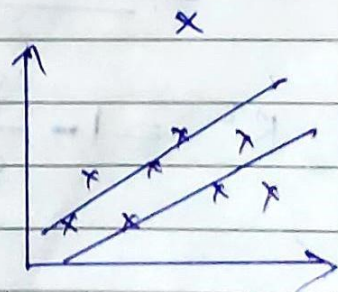
Dependent Feature

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

cost fun

(2) Mean Absolute Error (MAE)

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



MAE

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

Disadvantage ① Convergence usually takes more time.
optimization is a complex task
② Time consuming.

① Huber loss \Rightarrow combination of MSE & MAE

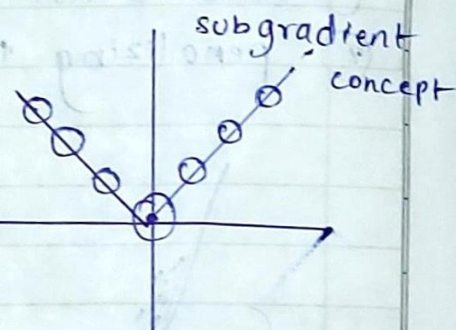
② RMSE

\Downarrow
 $\sqrt{\text{MSE}}$

\Downarrow
unit

\Downarrow
Outliers

\Downarrow
Differentiable



Loss Functions \div MAE, RMSE, Huber loss, MSE

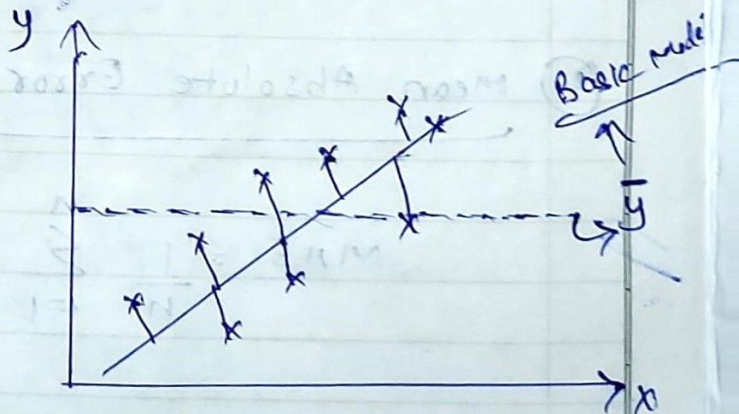
Performance Metrics.

① R squared

② Adjusted R squared

① R squared \div

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



SS_{Res} = Sum of Square Residuals

SS_{Total} = Sum of Square Average

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

\bar{y} = Average of y

$$= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \rightarrow \text{low values,}$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 \rightarrow \text{high values}$$

R^2 squared, be $-ve \Rightarrow$ Model is very bad.

$$R^2 = 1 - \left\{ \frac{\text{small number}}{\text{Bigger number}} \right\}$$

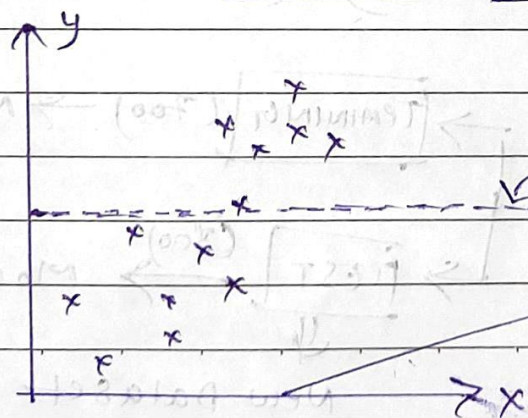
< 1

0.85%

\Downarrow

85% Accurate

performance of Model that you have created.



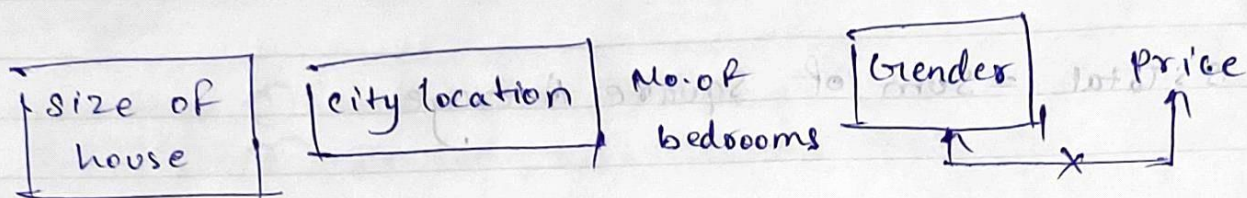
$$R^2 = -ve$$

$$R^2 = 1$$



Teacher's Signature.....

* Adjusted R Squared



Overfitting And underfitting

$R^2 = 65\%$ ✓ $Adj R^2 = 63\%$ $P=1$
 $R^2 = 75\%$ ✓ $Adj R^2 = 73\%$ $P=2$
 $R^2 = 88\%$ ✓
 $R^2 = 90\%$ } \Rightarrow No

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

N = No. of data points

P = No. of Independent feature

Overfitting And underfitting (Bias And Variance)

DATASET

1000 datapoints

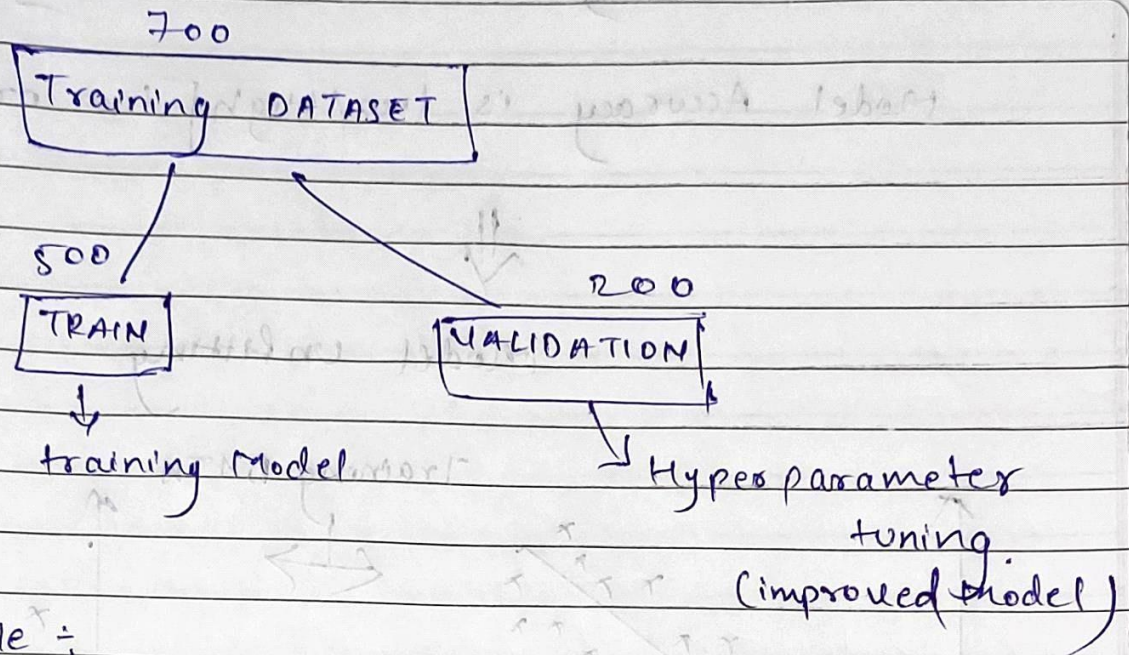
TRAINING (700)

Model training

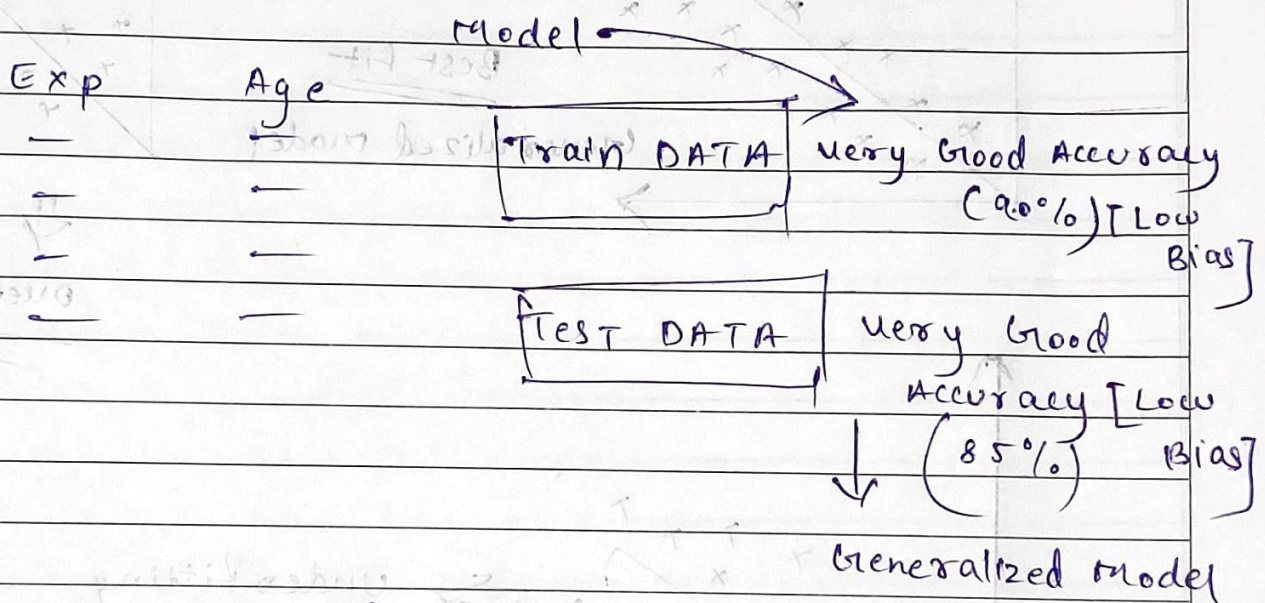
TEST (300)

Model Test

New Dataset



Example ÷



TRAIN Very Good Accuracy [90%] [Low Bias]

TEST Bad Accuracy [50%] [High variance]

⇓
overfitting

⇓
Hyper parameter tuning

Model Accuracy is low [Low Bias]

Model Accuracy is low/High [low or High variance]



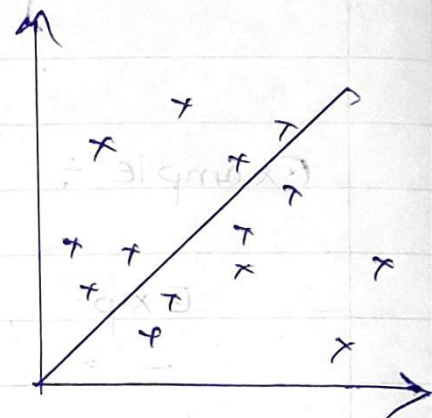
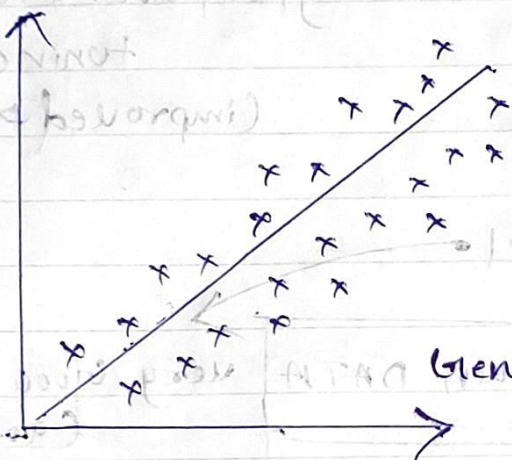
Model unfitting

Training DATA

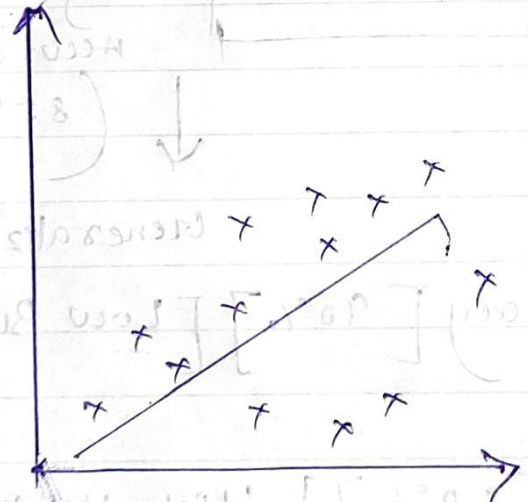


Best Fit

Generalized model



overfitting



underfitting