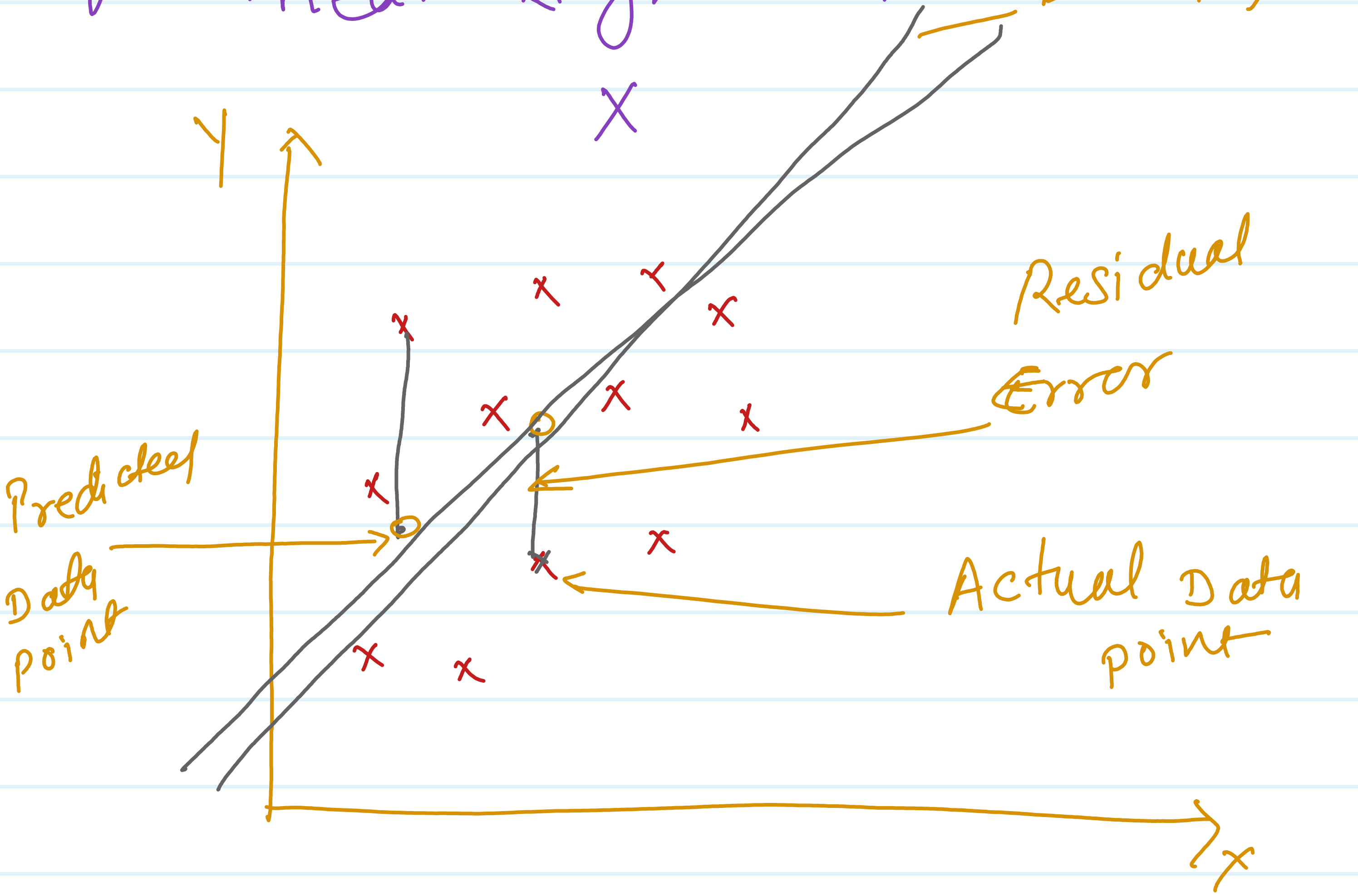
Linear Regression



y = Actual Data point

$\hat{y}$ = predicted data point

Residual Error (Distance b/w Actual and predict point)

Line eqn

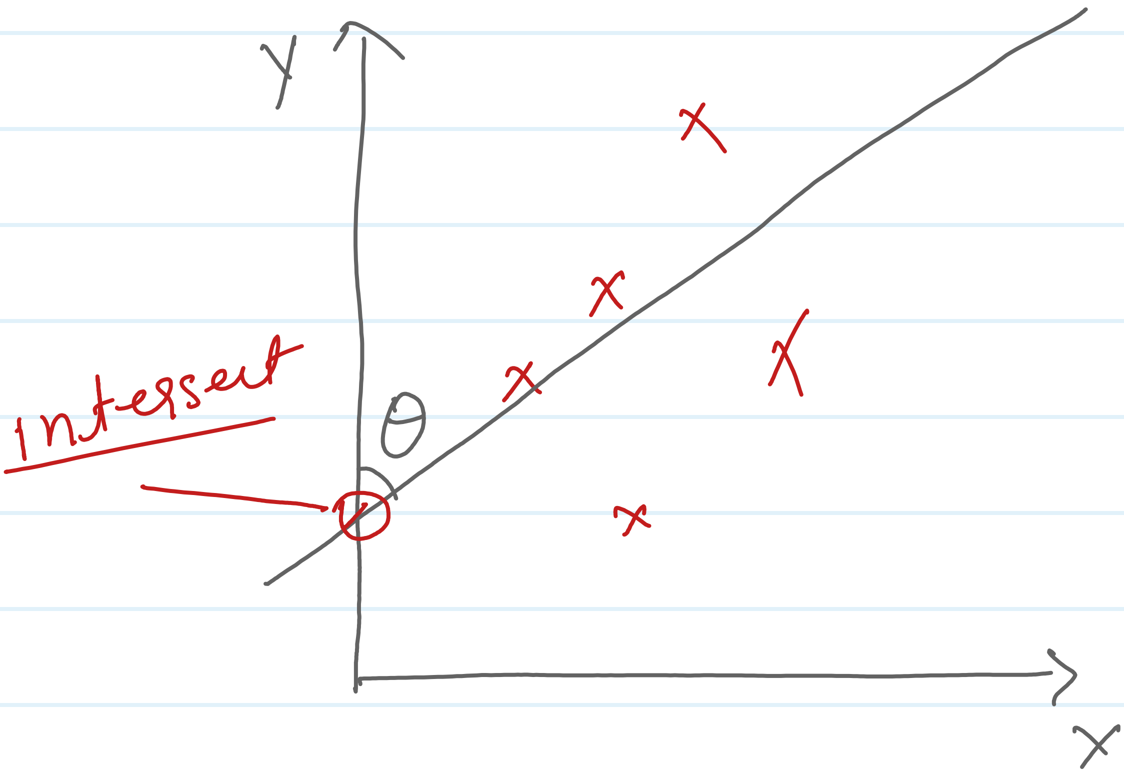
$$\hat{y} = mX + C$$

m - slop

x - Data point

c - Intersect

$\hat{y}$ = predicted point



$$\left[\; h_\theta(x) = \hat{y} \; \right]$$

$$h_\theta(x) = \theta_0 + \theta_1 x_1$$

single variable

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

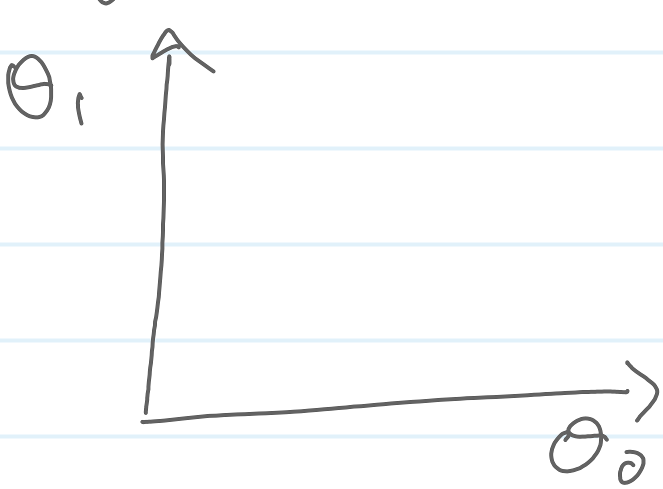
The main aim is, Reduce Error,

\# Cost function (It is also called squared error func)

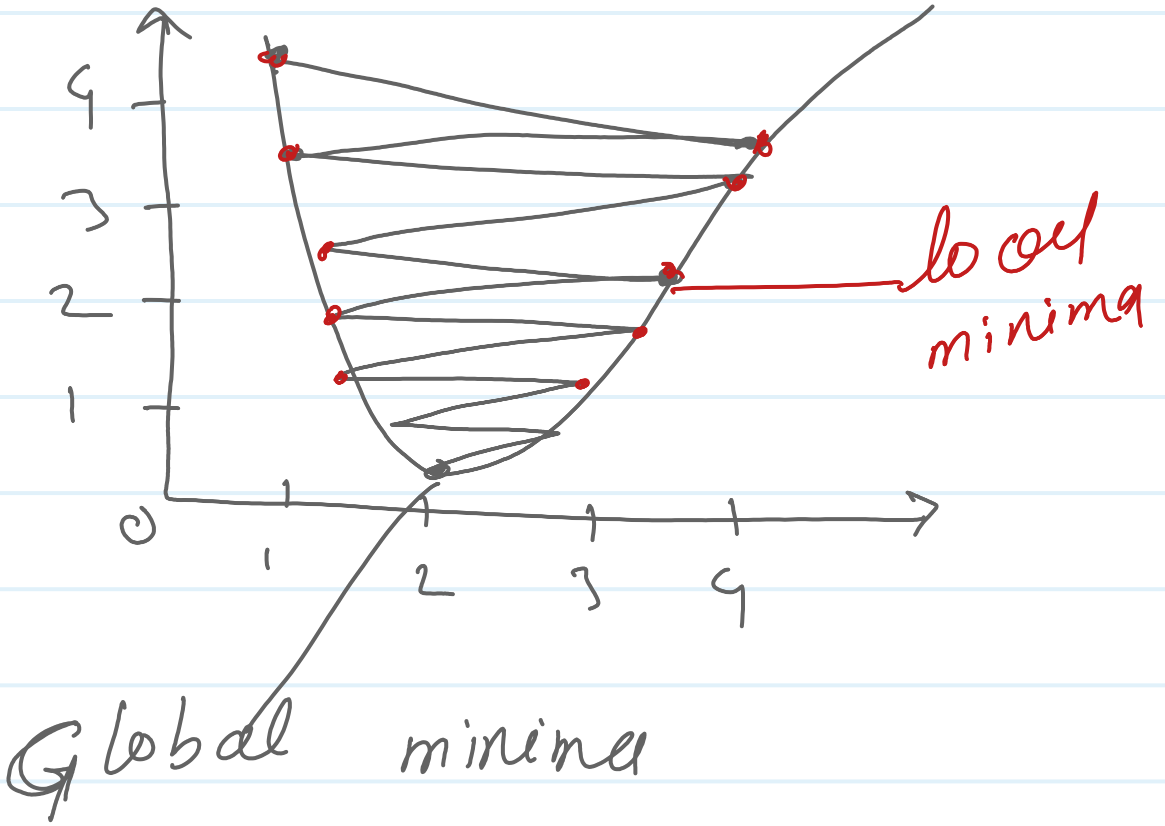
$$J(\theta_0, \theta_1)$$

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

This is cost function to mini error by changing value of θ_0, θ_1.



value change θ, x,



* Repeat convergien theoram

$$\theta_j = \theta_j - \alpha \boxed{\frac{d}{d\theta_j}\left[J(\theta_1)\right]}$$

slop funct.

α = learning rate

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

Type of cost function –

1. MSE
2. RMSE
3. MAE

① MSE (mean squared Error]

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

Adv –
① This eqn is differnciable
② This eqn also has one globale minima.

② RMSE [Root mean squared error]

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[y-(\theta_0+\theta_1 x)^2\right]}$$

Dis adv – ① This is not robust to outliers.
② It create local minima.

③ MAE (Mean Absolute Error)

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

Adv →

① Robust to outliers

② It will be in same unit

Dis Adv —

① Conversion usually take more time.

☆ To find accuracy of model there are two matrix or method.

① R^2

⑪ Adjusted R^2

* R^2 statistics

$$R^2 = 1 - \frac{RSS}{TSS}$$

R^2 = Coefficient of determination

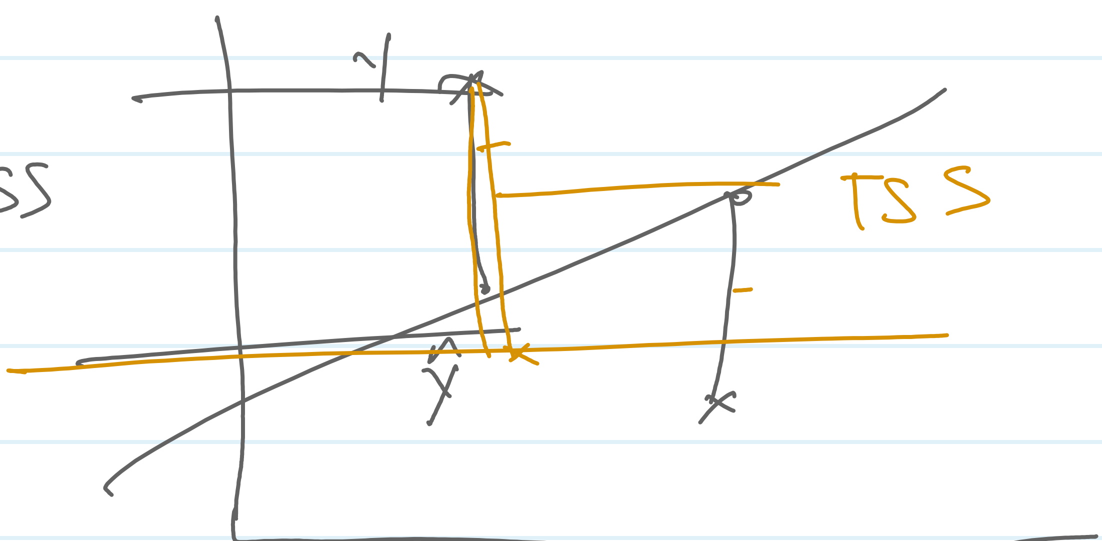
RSS = sum of square of residual

TSS = Total sum of squares

RSS - Dist. b/w y and $\hat{y}$

TSS - Dist b/w y and Avg $\bar{y}$

RSS



$$RSS = \sum (Y - \hat{Y})^2$$

$$TSS = \sum (Y - \bar{Y})^2$$

Adj. R^2

$$\text{Adj } R^2 = 1 - \frac{(1-R^2)(N-1)}{N-p-1}$$

N = no of datapoint in our dataset

p = no. of independent variable

x_1	x_2	x_3	Y

Assumption of linear Regression -

① Ind. and depn. variable has to be linearly connected.

② no multicolinerlity ($X_1 \neq X_2 \neq X_3 \neq X_n$

③ Error term must showcase constant variance.

✯ Regularization

① L-1 / Lasso Regularization -

It is used to select features in dataset while making models. Like "feature Reduce" or "Feature selection"

X_1	X_2	X_3	X_4	Y

X = feature column

Y = target column

$$L_1 = \frac{1}{m}\sum_{i=1}^{m}\left[h_\theta(x)^i - y^i\right] + \lambda\,|slop|$$

λ is hyperparameter.

② L-2 / Ridge Regularization :-

To reduce overfitting of model. in Linear Regress.

* overfitting =>

Training Accuracy - 90%

New data

Testing Accuracy - 55%

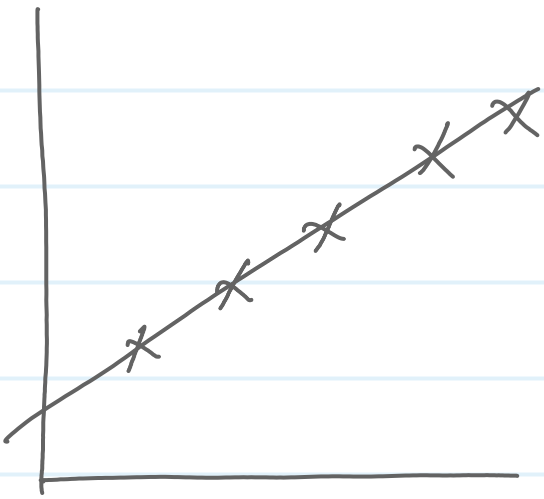
* underfitting

Training Accuracy - 60%

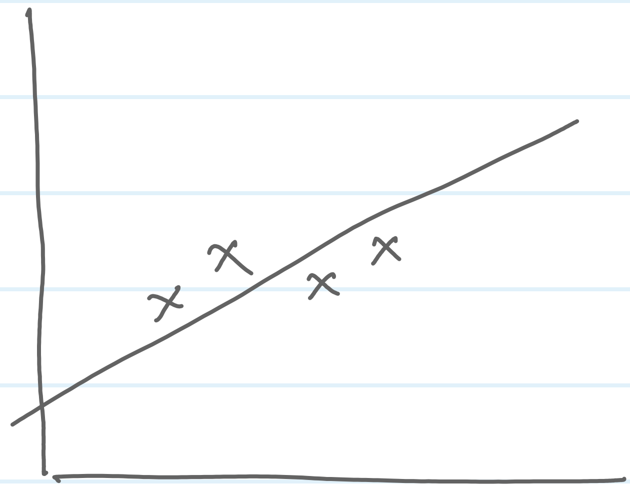
Test Accuracy - 80%

overfitting — Low bias
high variance

underfitting — high variance / low variance
High biase



overfitted



after Regularization.

$$L_2 = \frac{1}{m} \sum_{i=1}^{m} \left[h_\theta(x)^i - y^i \right] + \lambda \, (slop)^2$$

③ Elasticnet Regularization -

Combination of L-1 and L-2

$$\text{Elasticnet} = \frac{1}{m}\sum_{i=1}^{m}\underbrace{\left[h_\theta(x)^i - y^i\right]}_{\text{cost funct}} + \overbrace{\lambda\,|slop|}^{L-1} + \underbrace{\lambda(slop)^2}_{L-2}$$