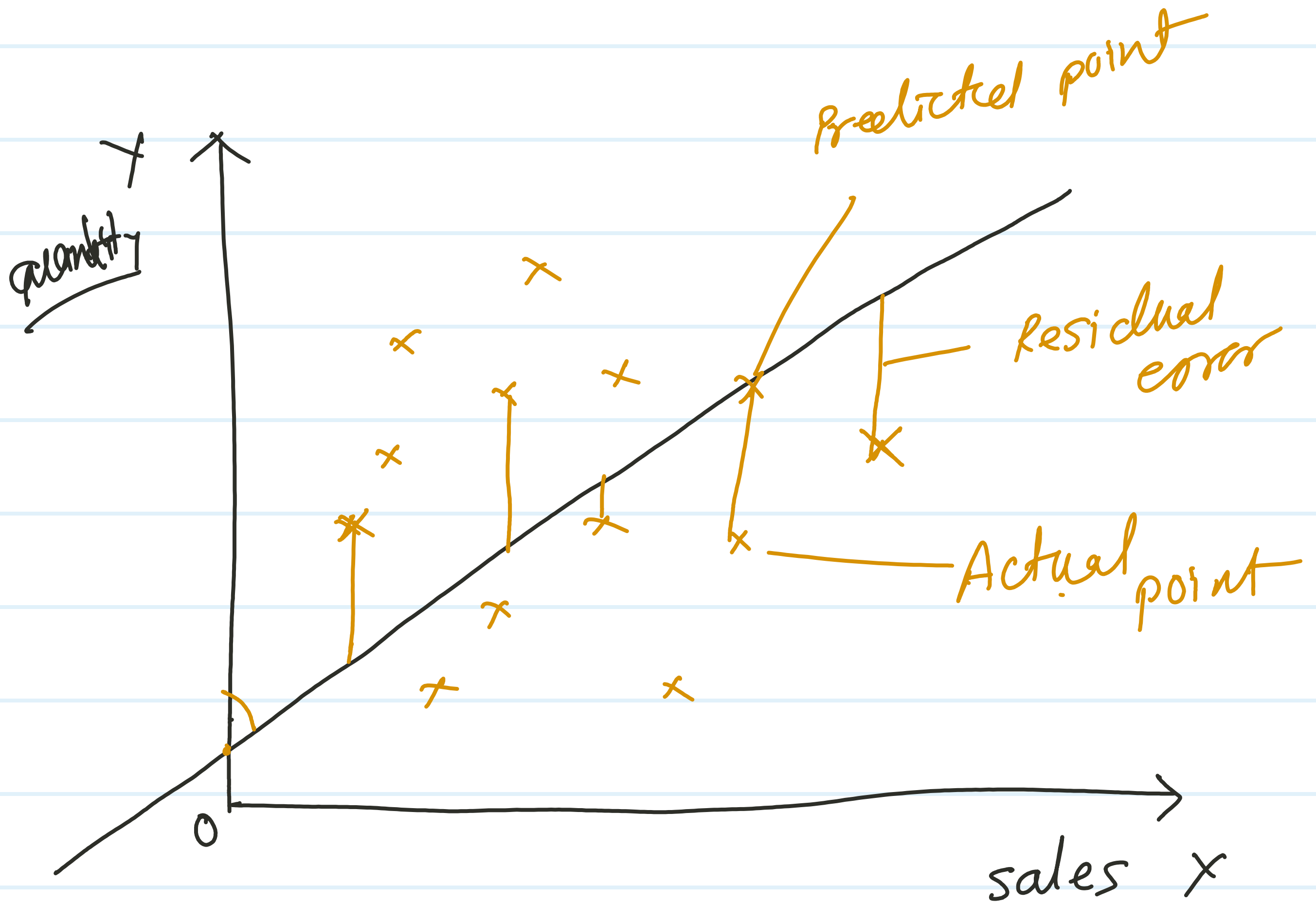
✱ Linear Regression

$$y = mx + c$$

$m =$ slop or coeff.

$x =$

$c =$ Intercept $(0, x=0)$



Residual error $(y - \hat{y})$

To find best fit line with minimal error.

$$y = mx + c$$

$$h_\theta(x) = \hat{y}$$

Single linear Regr.

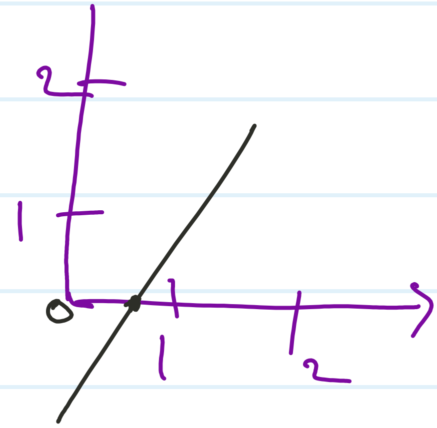
$$h_\theta(x) = \theta_0 + \theta_1 x$$

multipoint linear Regress

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

$\theta_0 = 1.5$

$\theta_1 = 0$



$\theta_0 = 0$

$\theta_1 = 0.5$

$\hat{y} = 0 + (0.5) \times 1 \Rightarrow 0.5$

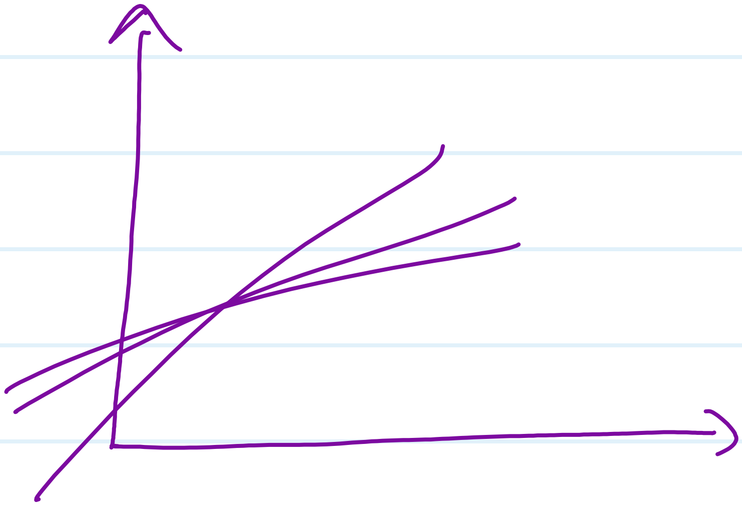
$\hat{x} = 0 + (0.5) \times 2 \Rightarrow 1$

$\hat{y} = 0 + (0.5) \times 3 \Rightarrow 1.5$

* Cost function =

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

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



θ_0

θ_1

Repeat convergen theorem

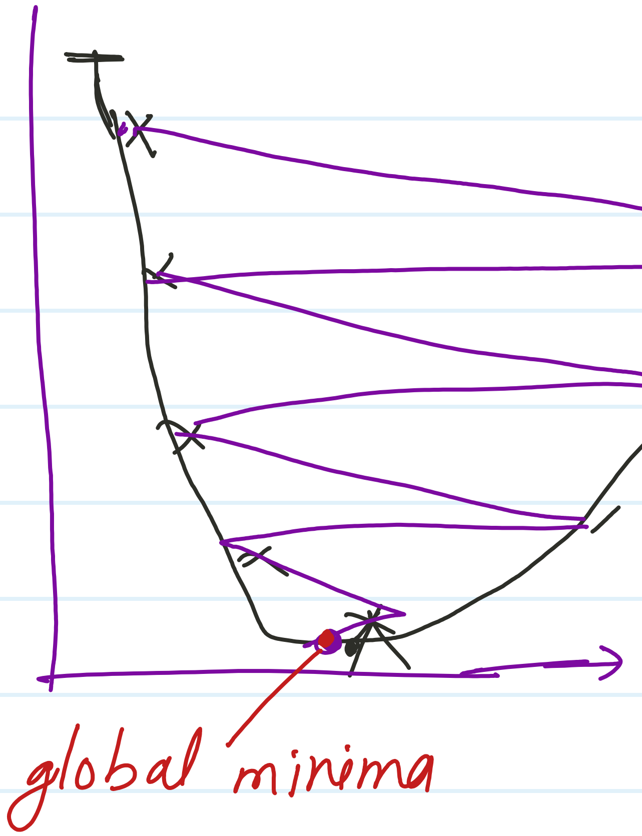
$$\theta_j = \theta_j - \alpha \boxed{\frac{d}{d\theta_j}\left(J(\theta_1)\right)}$$

α = learning rate

0.05, 0.01

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

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



gradient Decent

① MSE (mean square error)

② RMSE (Root mean square error)

③ MAE (mean Absolute error)

① MSE

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

It create global minima

② RMSE

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

It is not robust to outlier.
Create local minima

③ MAE

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

pros: Robust to outlier.

Cons. - It take usually more time
to optimization.

* performance matrix

R^2 statistics →

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

RSS = sum of squ. of residuals

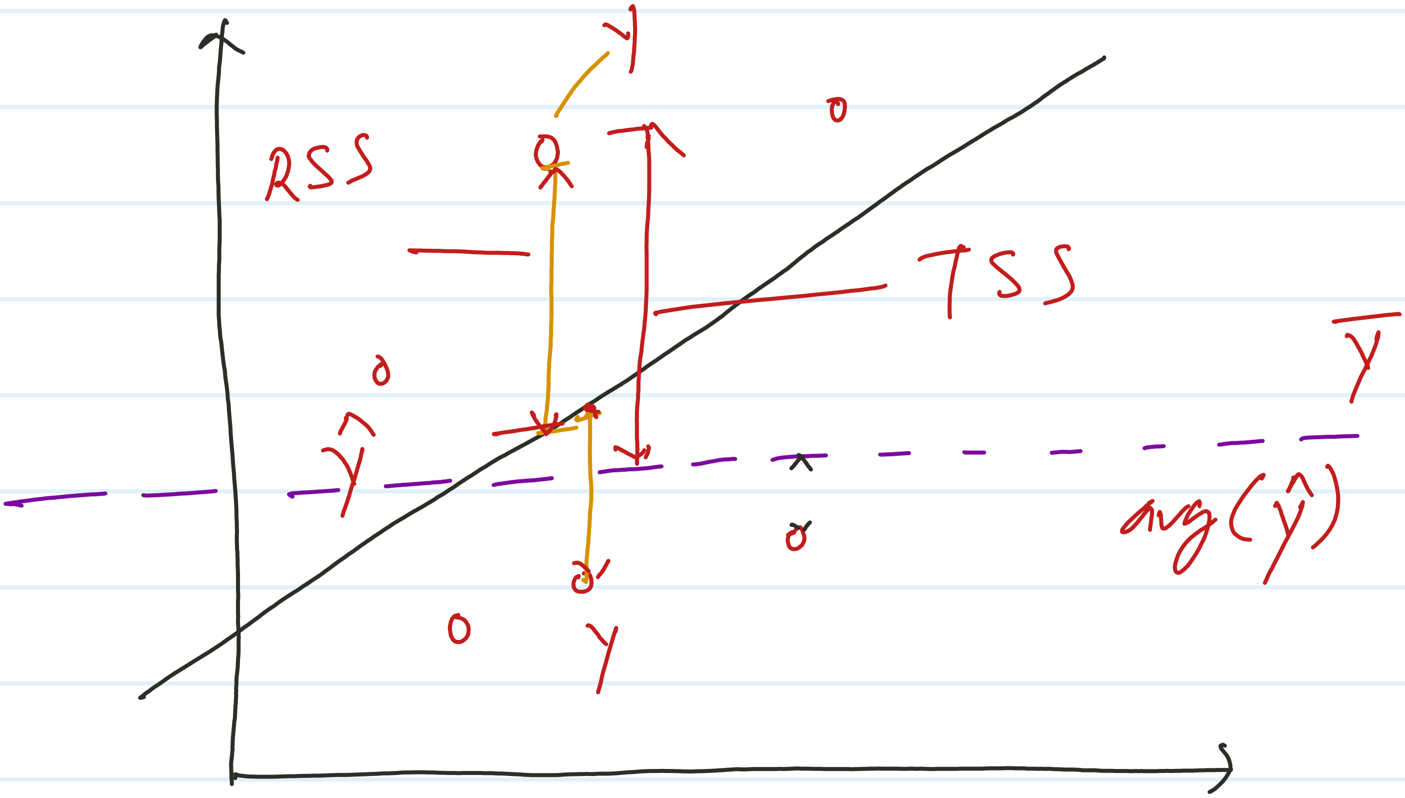
TSS = Total sum of squ.

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

TSS = Dist. b/w y and avg($\hat{y}$)

$$RSS = \Sigma (y - \hat{y})^2$$

$$TSS = \Sigma (y - \bar{y})^2$$



Adjusted R^2 - statistics

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

N = number of datapoint in dataset

P = number of independent variable

overfitting and underfitting

overfitting

=> Train — 90% high variance low biased

=> Test = 50%

underfitting

=> Train = 50% low variance high biased

=> Test = 90% | 40%

1
2
3
4
5
6
7
8 9 10

6 train

4 test

Best fit model

low variance
low biased

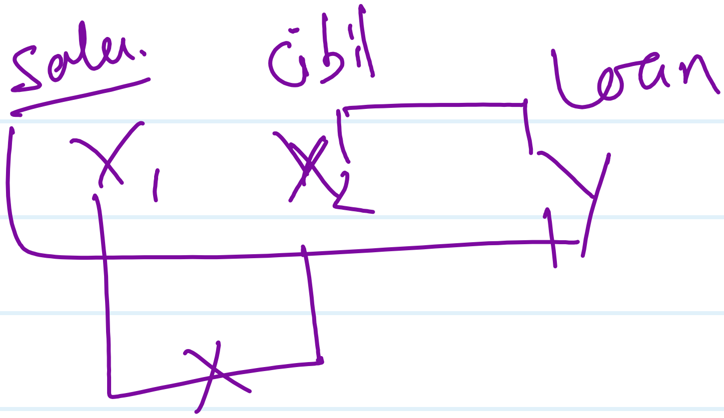
Important Assumption of LR

1. There should be linear relationship b/w dependent and independent variable.

2. Error term are not suppose to co-related.

3. Ind. variable (x) and residual error should be uncorrelated.

4. No - multicolinearity



X_1 X_2 X_3 X_4

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

cost function — MSE, RMSE, MAE

evaluation | performance metrix

R^2 or Adj R^2

Data leakage

Dataset

split → train test

train - transform

R^2