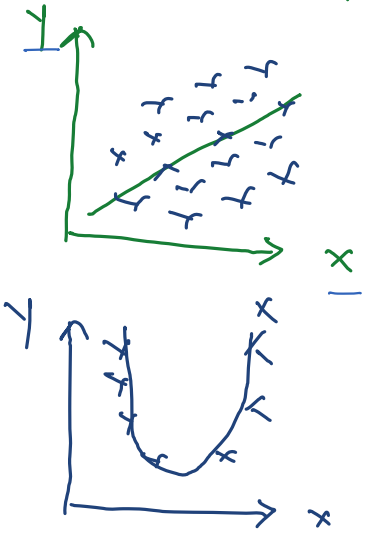
# Assumptions of Linear Regression

22 November 2022 04:13 PM

① Linear Relationship betn Dependent and Independent Variables

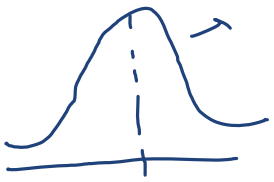



⇒ Linear Relation betn X & Y
(Linearly seperable)




2] Normally Distributed error component :

↳ Data Need to be Normally Distribution



Bell Shape curve

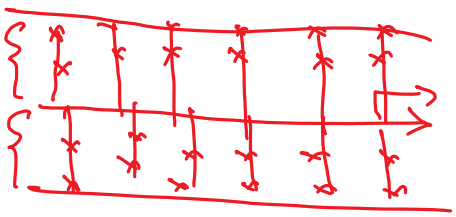
1] K-S Test
2] Shapiro Wilk Test } Analytical Test

i.e. Normality Test

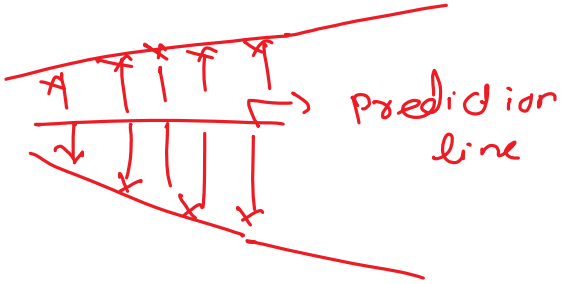
3] There should not be "Heteroscedacity" i.e.

→ Variance of residuals/error must be constant across predicted values.

Homocedacity ✓




Heteroscedacity X




4] No multicollinearity : →

4] No Multicollinearity : →

→ instability of Regression coefficient

→ two or more predictor variables are strongly correlated with other

$$\hat{y} = b_1 x_1 + b_2 x_2 + \dots + b_n x_n + a$$

($b_1$: 50, $b_2$: 50, $a$: intercept)

coefficient.

→ Let's Assume $x_1$ & $x_2$ feature having some coefficient.

* Multicollinearity Diagnosis: →

$$\hat{y} = b_1 x_1 + b_2 x_2 + \dots + b_k x_k + a$$

↳ Tolerance

$$T = 1 - R^2$$

↳ coefficient of determination

↳ VIF : variation inflation factor

$$VIF = \frac{1}{1 - R^2}$$

↳ coefficient of determination

$T < 0.1$

$VIF > 10$