Linear regression

Regression – linear, non linear

STEPS 1 – LINEAR OR POLYNAMIAL

STEP2- GOOD FIT MODEL

STEP3 – COST FUNCTION/ERROR FUNCTION

STEP4 -GRADIENT DESCENT ALAGORITHM

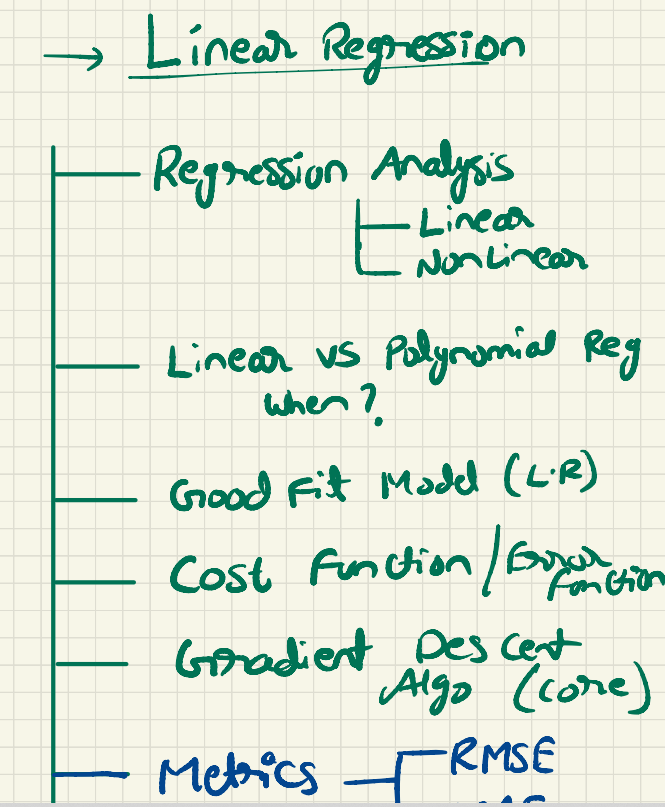
METRICS – RMSE, MAE,MPA

R2 AND ADJ R2

PVALUE

R2 and adjutants R2 is only for linear regression

It is used to evalute the performance of linear regression

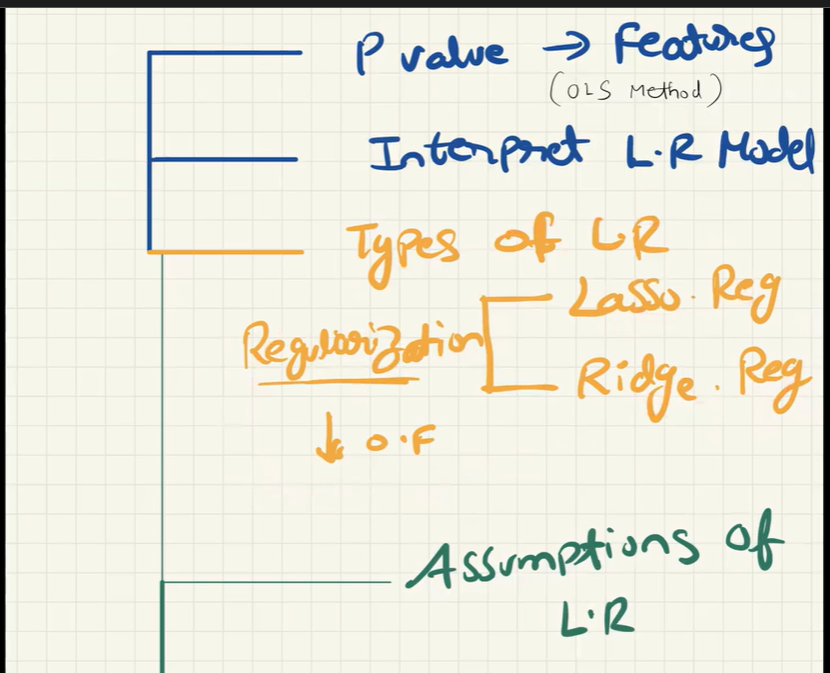


Regularization

Lasso/Ridge

Linear regression is used in two ways one is statical another one using ml

Regulation and assumption important topic

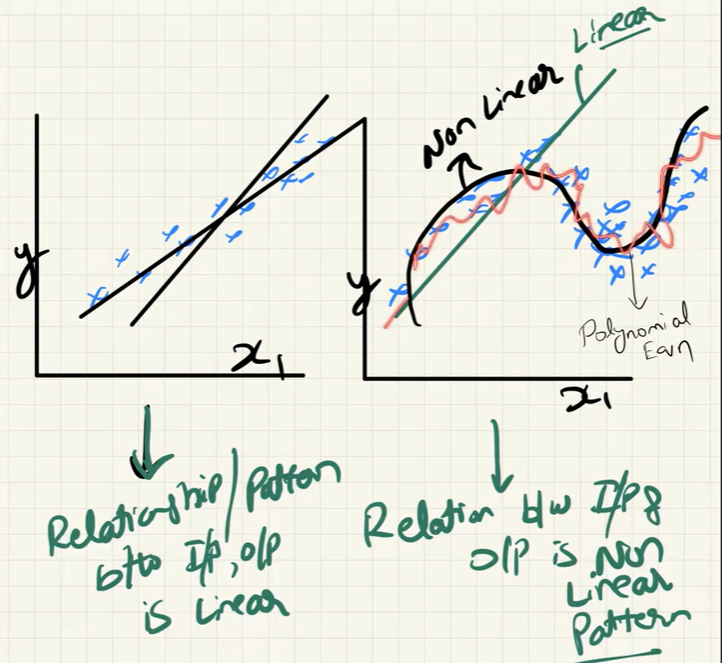


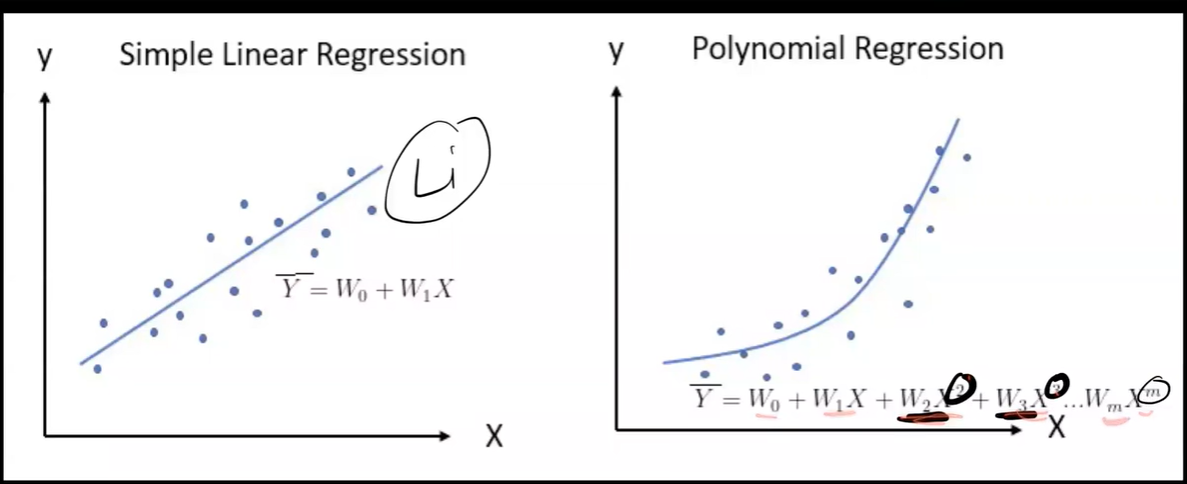
Linear regulation is used to find relationship between input and output for numerical data only

From – infinity to + infinity

Linear – staright line in graph

Non – linear – curve or polynomial curve





Suppose only one input feature like x1 then it is simple linear model

And more then one input feature it is called multiple linear model

BEST FIT MODEL

Overfit or underfit

Which means model should predict the output close to actual output

COST FUNCTION/ERROR FUNTION

That determines the difference between the actual and predicted output

Here first take error with predicted output and we compare with actual ouput using cost function and the we means the mean square error .

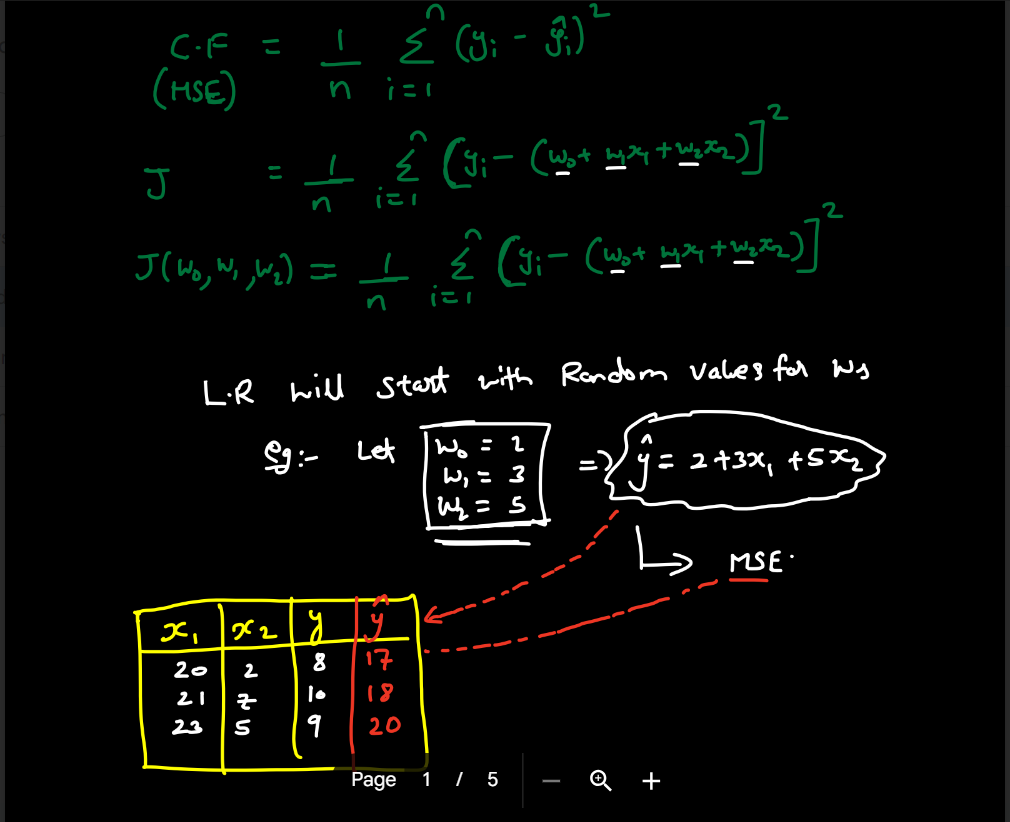
Weight is nothing but the coeffienct like w0, b0

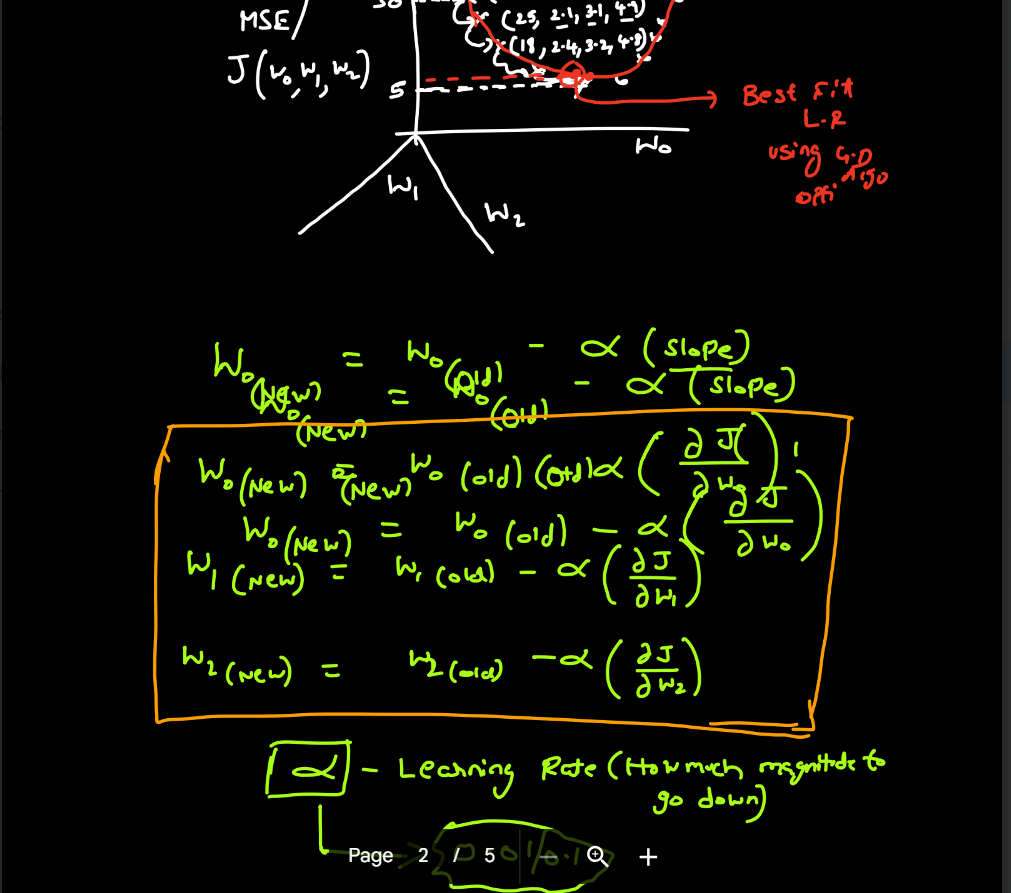


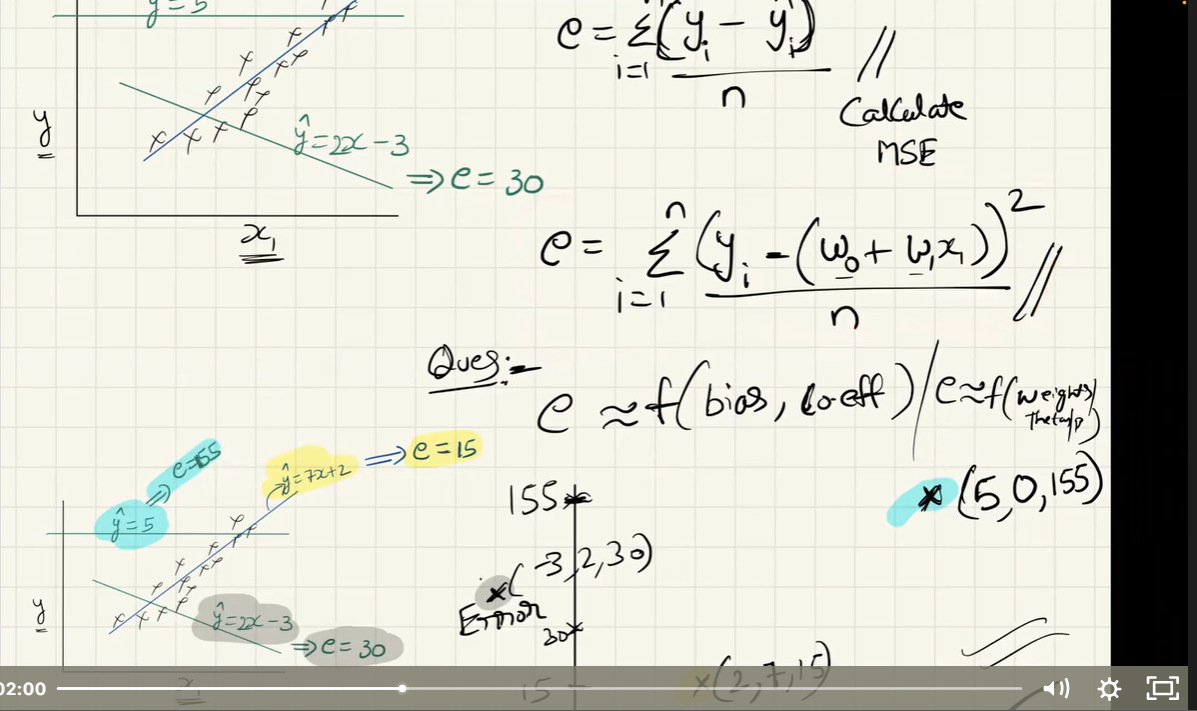
Gradient descent algorithm

It is used to determine the best fit model

It will find best slope by going down which means it reduces the error until it reaches zero





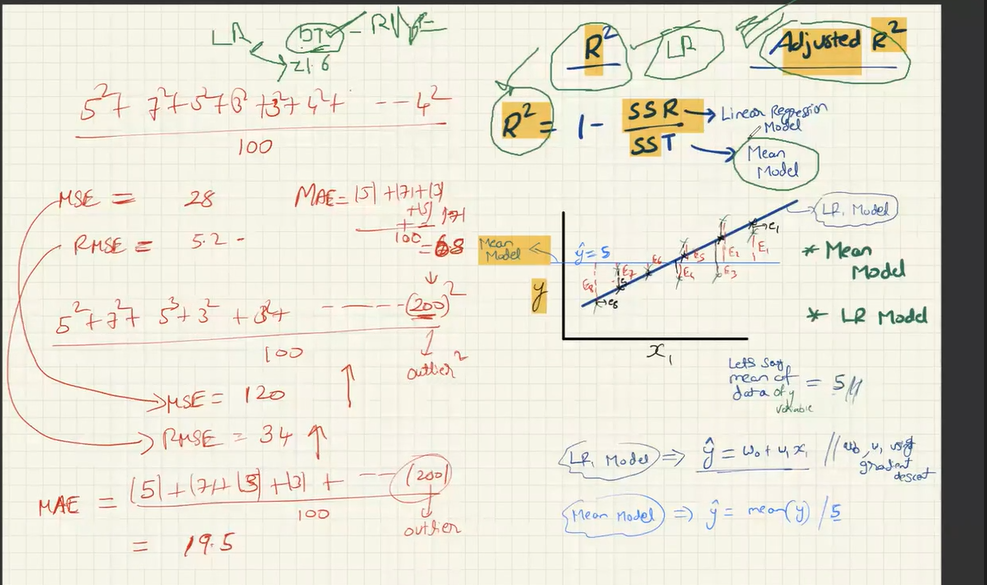


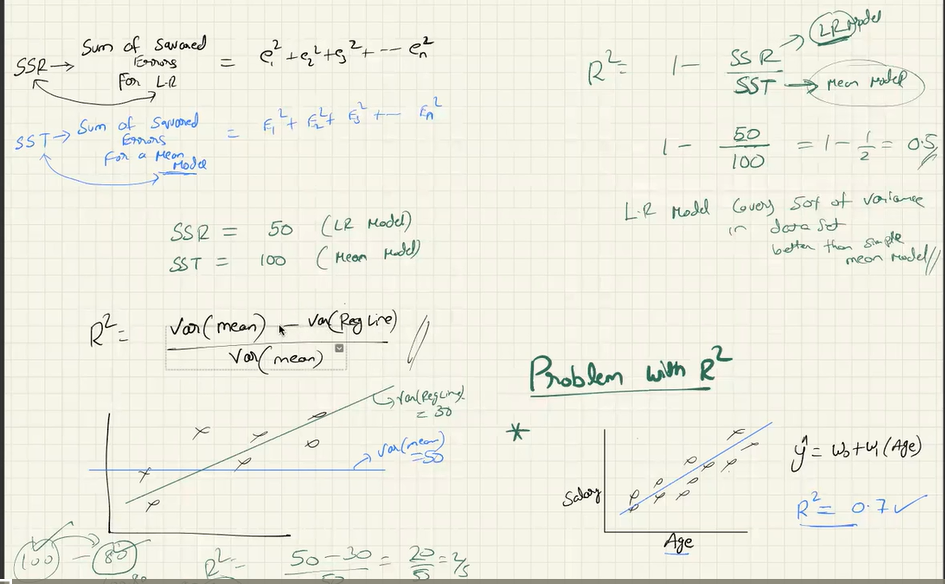
**R2**

**R2=1-linear regression/mean model**

**It is to prove LR better than the mean model**

**Based on the error difference from mean model and linear regression model we tell linear regression is best. Like suppose linear regression model error is less than mean model then it is best.**

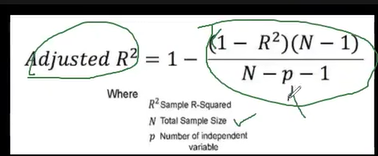


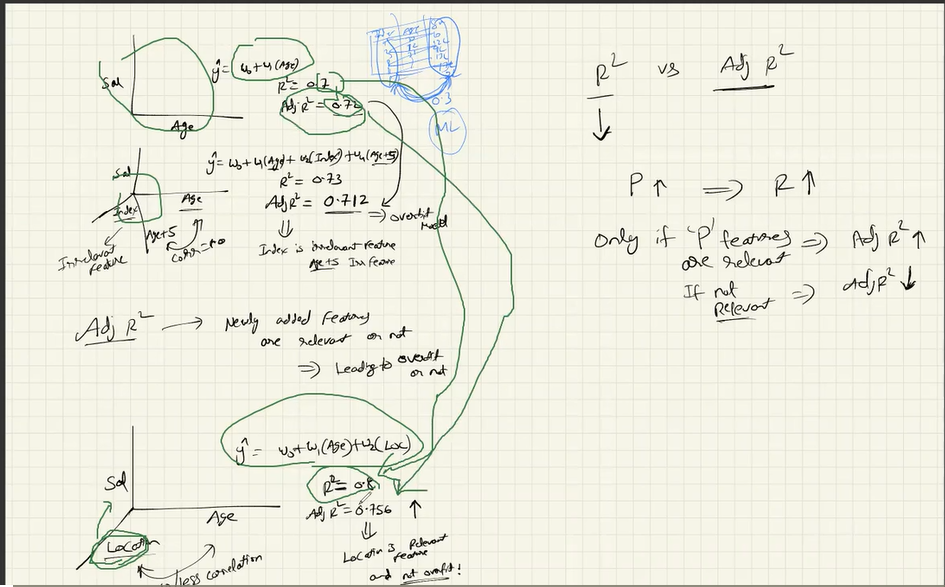


R2 cannot difference between the unimportant features which means extra features added and strongly correlated features.

**ADJUSTED RSQUARE**

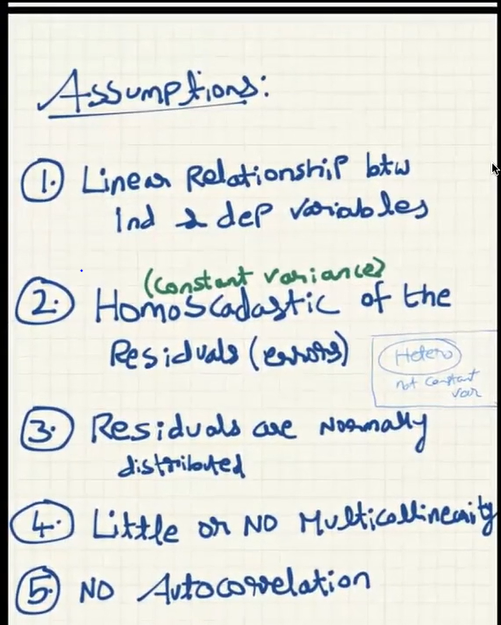
**It will define the feature which is added by r2 is important or not. By comparing with r2,adjusted r2 value will very less then say r2 is okay and not overfit.**



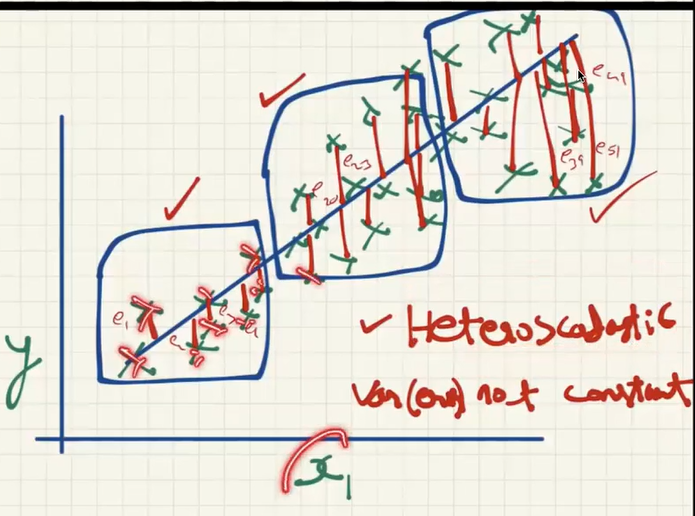


1. improve intro 2. aws why? if you know then only mention 3. improve explanation of project, 4. why 1 year as duration 5. not clear about her role, team size not clear, 6. random forest tree? 7. hyperparameter tuning in log reg?

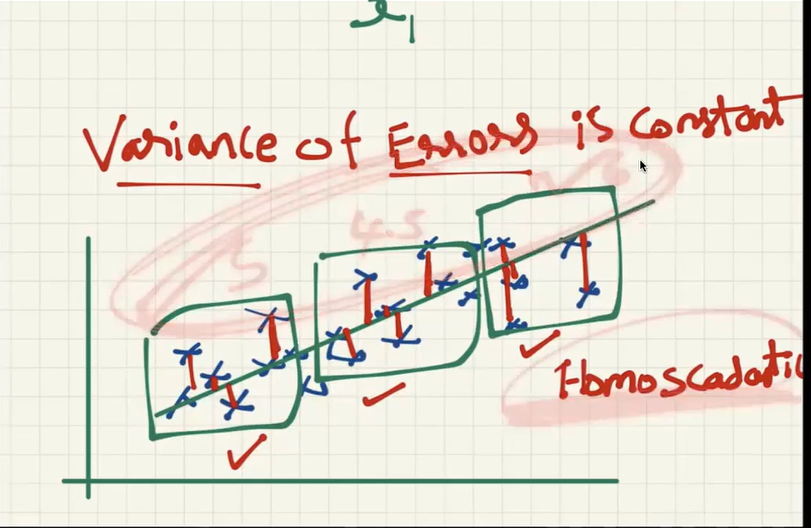
**ASSUMPTIONS**



**Homoscedastic – home means same scedastic means scatter in nature – constant variance**



**First find the difference which is error between actual and predicted output find the variance of all those error for one one sample see the difference if it is same then homescedatic orelse hetroscedatic**



**No auto correlation means output depends on provoisus output means is not possible in linear regression so for time series compantant is applied**