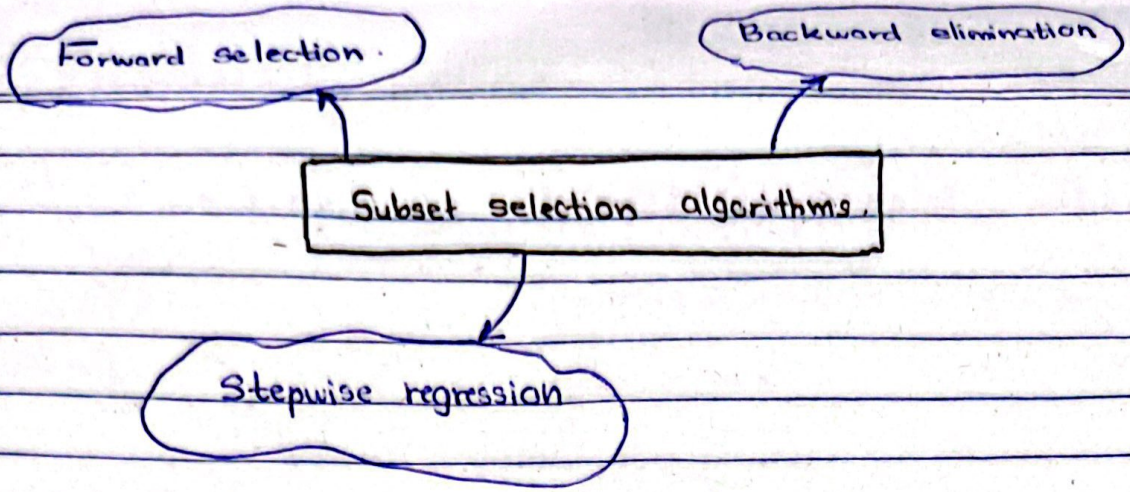


* Forward selection.

Starts with no predictors. ⟹ Adding predictors one by one.

Which predictor add next?
One that increases R^2_{adj} the most.

When / Where to stop?
When R^2_{adj} stops increasing.

* Backward elimination.

Starts with all predictors. ⟹ Removing predictors one by one.

Which predictor remove next?
One that increases R^2_{adj} the most.

When / Where to stop?
When R^2_{adj} stops increasing.

* Stepwise regression.

Starts with no predictors. ⇒ Adding and removing predictors

Which predictor add / remove next?
Based on AIC / R^2_{adj}.

When / Where to stop?
When dropped R^2_{adj} increases.

-⚹- Problems :-

1) Backward elimination methods are not feasible when $p > n$.
   - p: Number of variables / features.
   - n: Number of observations.

2) Different methods select different models.

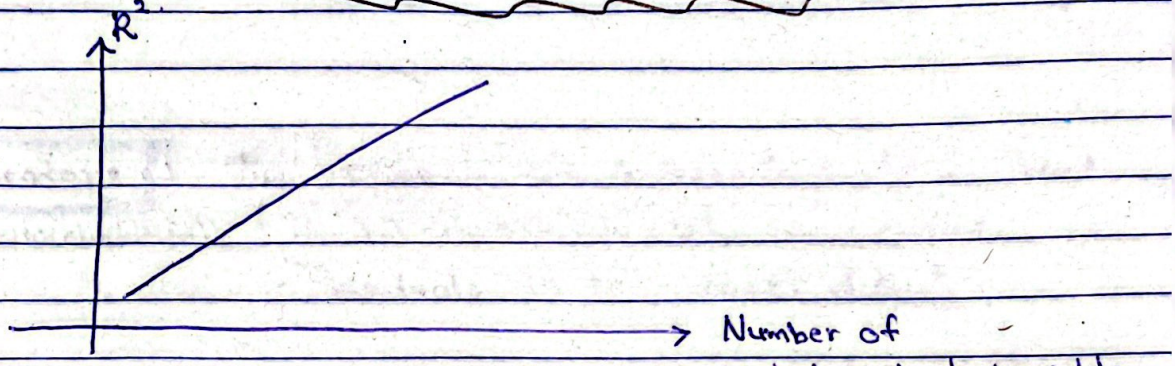
3) There is no gurantee to provide optimal solution where the model with the highest R^2_{adj}.

Accuracy measure

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

R^2 = coefficient of determination.
RSS = Residuals sum of squares.
TSS = total sum of squares.

Interpretation: The percentage of the response variable variation that is explained by a linear model.

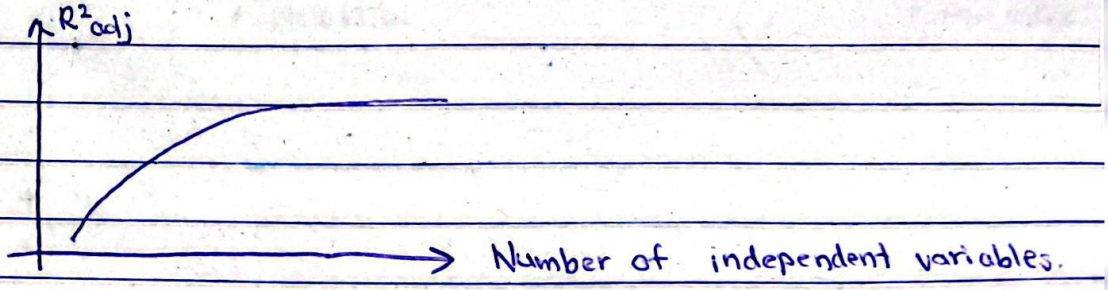


$$* R^2_{adj} = 1 - \left[\frac{(n-1)}{(n-p-1)}(1-R^2)\right]$$

n = Number of observatio
p = Number of variables.

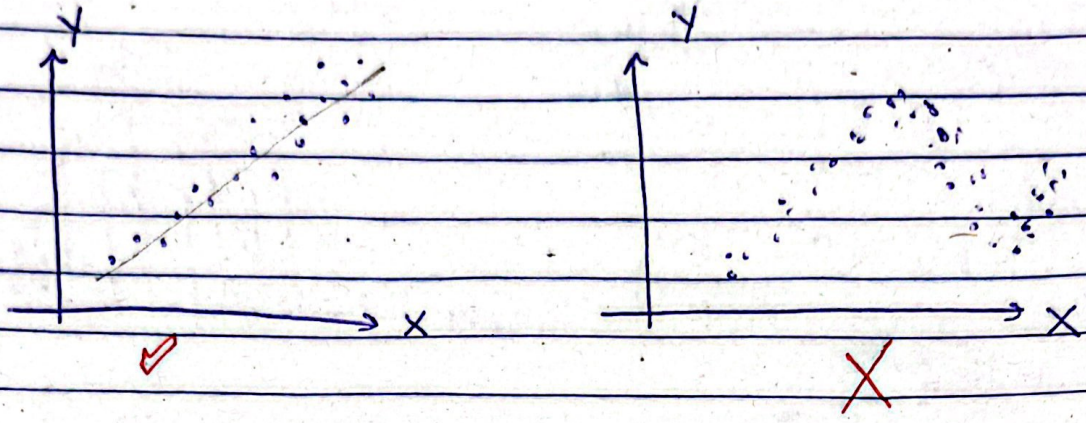
Difference

- R^2_{adj} : adjusted for the number of predictors.
- R^2_{adj} increases only if the new term improves the model.
- Always less than R^2.



* Assumptions of linear regression

① **Linear relationship** : There exists a linear relationship between the independent variable, X, and the dependent variable, Y.

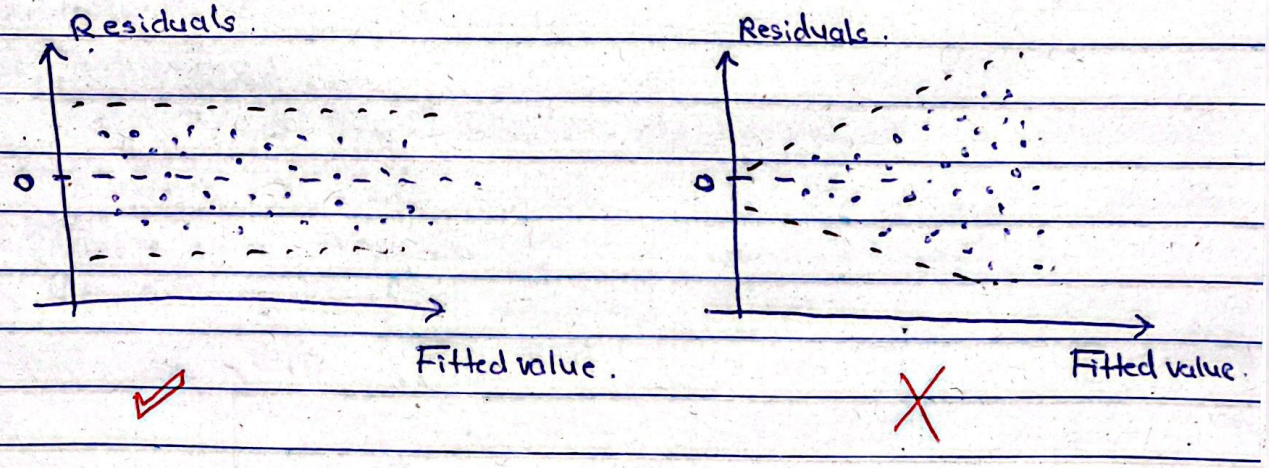


✓ ✗

② **Independence** (No Autocorrelation) : The residuals are independent. (i.e. there is no correlation b/w consecutive residuals in time series data.)

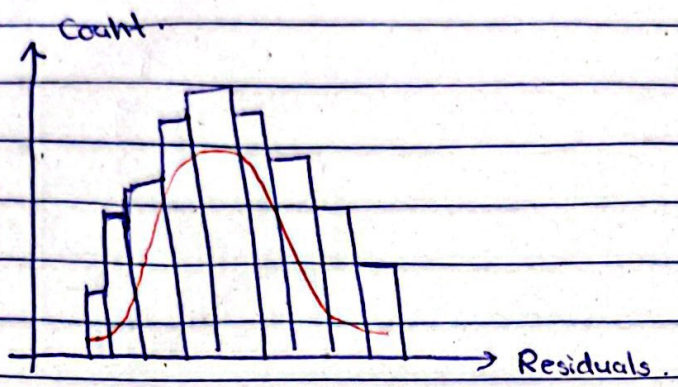
⟹ Check : Durbin-Watson test. ← Not required in this tutorial

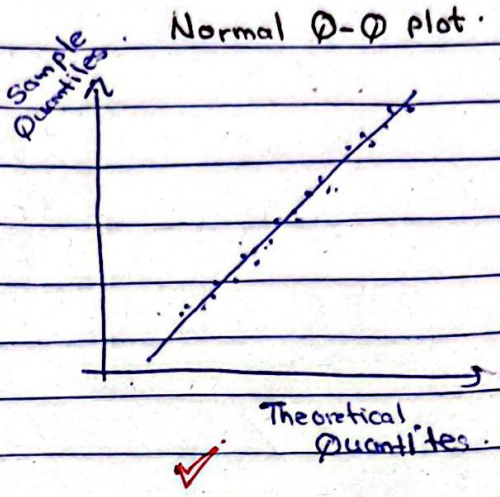
③ **Homoscedasticity** : The residuals have constant variance at every level of X.



✓ ✗

④ **Normality**: The residuals of the model are normally distributed.

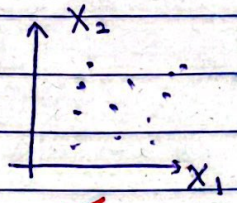




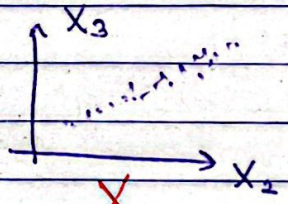
✓

⑤ **No multicolinearity**: There is no relationship between predictor variables.

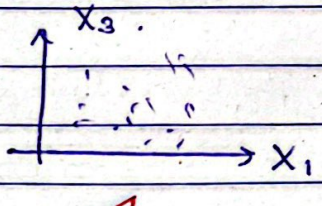
⟹ Check pairwise plots.



✓



✗



✓