★ Train data accuracy will always be higher than test data accuracy.

November-18:

Multi Linear Regression:

Whenever we have multiple features & one target, we will go with multi-linear Regression.

features (x) → target (y)

Age (x_1) height (x_2) weight (x_3) BMI (y)

To use MLR we need to satisfy some assumptions.

★ Assumptions of (MLR) Linear Regression:

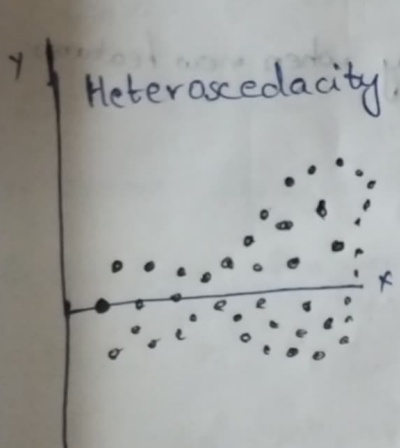
1.) ~~and~~ Linearity - The relationship b/w dependant (y) & independent (x) variables should be colinear.

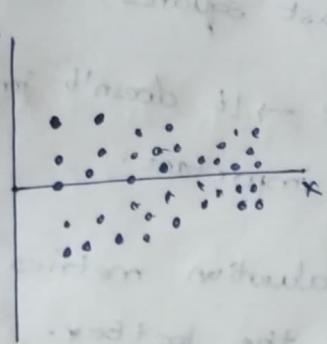
2) Independence - ~~Independ~~ Features should be independent on each other. or else they will cause multicolinearity problem.

3) Auto regression - It means the present value will be dependent on previous value. Auto regression should not be present in the data.

If continuous date is present, then they are dependant.

4) Homoscedasticity: Constant / uniform variance of residuals / error.





* Our data should follow homoscedacity.

5) Normality -

Features, residuals should be normally distributed

6) Mathematical equation:

SLR ⇒ $\hat{y} = \alpha x + \beta$.

MLR ⇒ $\hat{y} = \alpha_1 x_1 + \alpha_2 x_2 + \dots \alpha_n x_n + \beta$.

α_n, α_1 → coefficients

β → intercept

* Auto regression is also called as Auto correlation.

November-19:

To overcome assumptions:

As our cars data failed assumptions let's try to overcome them.

1. Linearity → increase the data.
2. Normality → increase the data
   Remove any outliers if present.
3. Multicollinearity → try different ways in presence of features, absence of one, absence of one. choose whichever gives more accuracy.
   - Before removing, check how much collinear they are based on it take decision.

Ols - ordinary least squares.

Adjusted R^2 score → It doesn't improve accuracy when new features that are not important.

aic, bic → evaluation metrics

↳ lower the better.