Multicollinearity

Multicollinearity is a statistical phenomenon in which two or more predictor variables in a model are highly correlated with each other.

It can cause issues such as unstable parameter estimates and inflated standard errors, making it difficult to discern the individual effects of each predictor variable on the response variable.

Identify and remove redundant predictor variables that are highly correlated with each other using **Variance Inflation Factor (VIF)**

**Variance Inflation Factor (VIF)**: Calculate the VIF for each predictor variable. The VIF measures how much the variance of an estimated regression coefficient is increased because of multicollinearity. A VIF greater than 10 or 5 is often considered indicative of multicollinearity.

**After detecting the highly correlated variables, the another way to handle it is** combining correlated variables into composite variables

1. Identify Correlated predictors
2. Create composite variables
3. Summation: Add the values of correlated variables to create a new composite variable

Xcomposite=X1+X2

1. Weighted Average: Calculate a weighted average of the correlated variables, weights based on their importance of correlation coefficients.
2. Create Interaction terms

Xinteraction=X1\*X2

3. Evaluation: Using VIF to verify the effectiveness of the approach

4. Interpretation: Consider the interpretation of the composite variables rather than the individual predictors