Various ways to handle Multicollinearity except VIF.

1.Correlation Matrix

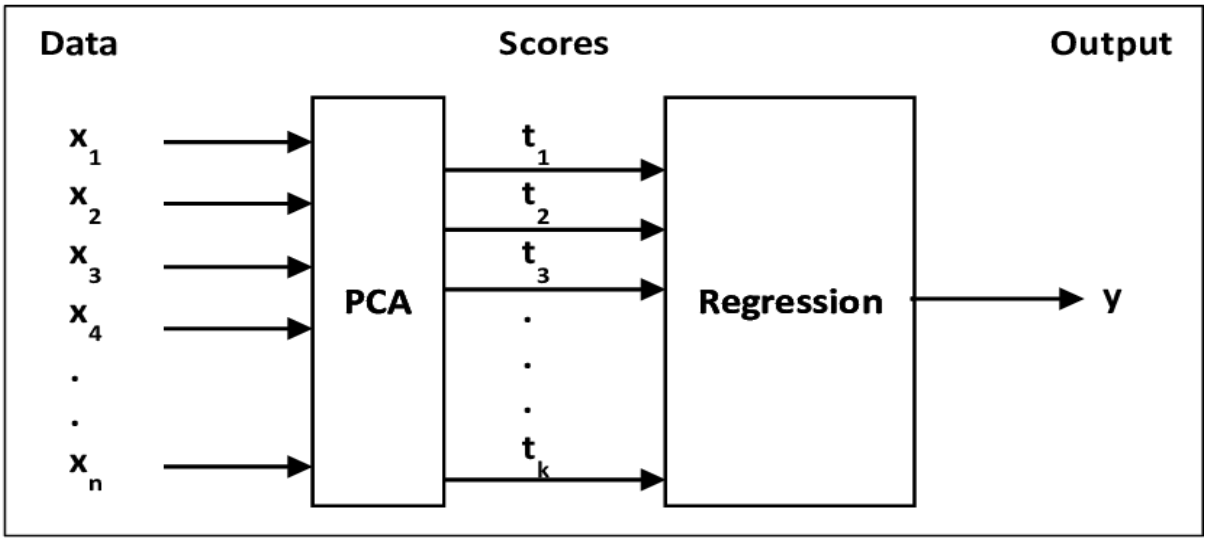
It is a key tool in exploratory data analysis, as it helps identify relationships between variables and understand their linear associations.

It computes pairwise correlation coefficients between independent variables to identify strongly correlated variables.

The correlated values close to +1 or -1 indicate strong multicollinearity, while values close to 0 indicate weak or no multicollinearity.

2.Principal Component Analysis (PCA)

Principal component analysis (PCA) is a dimensionality reduction and machine learning method used to simplify a large data set into a smaller set while still maintaining significant patterns and trends. So the idea of PCA is, reduce the number of variables of a data set, while preserving as much information as possible.



A Principal Components analysis (PCA) plot shows the similarities between groups of samples in data set.Each point on a PCA plot represents a correlation between initial variable and the first and second principal components.

Steps to be followed in PCA:

1. Standardize the range of continuous initial variables

2.Compute the covariance matrix to identify correlations

3.Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal components

4.Create a feature vector to decide which principal components to keep

5.Recast the data along the principal components axes

3.Ridge and lasso regression

- Lasso Regression is a form of regularization that seeks to minimize the magnitude of coefficients so that more relevant variables are included in the model.
- Ridge Regression works by introducing an additional term that penalizes large coefficient values. Both techniques help to reduce overfitting and improve prediction accuracy.

4.Increase Sample Size

Collecting more data can reduce the impact of multicollinearity, especially if the relationship between variables is not perfectly linear.

5.Domain Knowledge for Variable Selection

Use expert insights to prioritize and select variables that are theoretically important and relevant to the problem.