

What is multi collinearity ?

Multi collinearity is phenomena where 2 or more independent variables are high correlated with high others .

Why mult collinerity is bad?

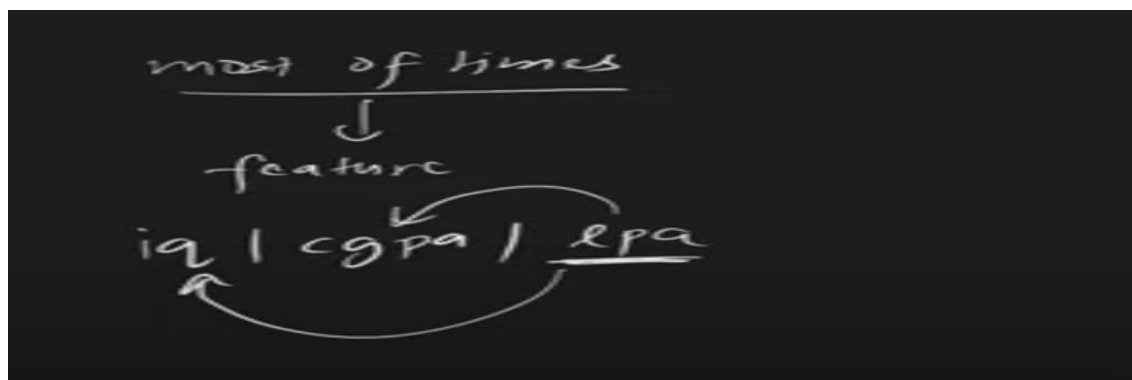
The diagram shows a regression equation  $lpa = \beta_0 + \beta_1 \times iq + \beta_2 \times cgpa$  with arrows pointing to each term. Below the equation, it says "multicollinear" and "exam". An example is given: "Skills - exactly same" with a circled 2 and arrows pointing to "iq" and "cgpa". Another "multicollinear" label is written below the example.

Is multi collinearity bad every time ?

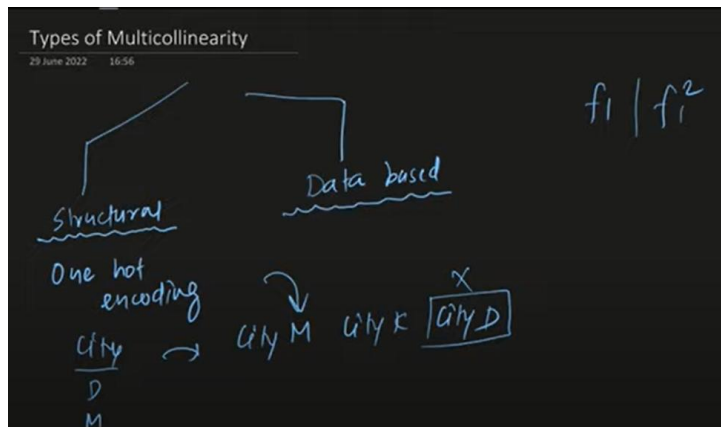
Model is used for 1. prediction 2. Inference

When we will use linear for inference then multi collinearity is bad .If we use linear model for prediction then it will not affect too much .We mostly use linear regression model for inference(feature importance) because other models gives us good results.

We want know that how much iq is contributing in LPA and how much cgpa is contributing in lpa. Whenever we want to know the importance of feature separately .



## Types of Multi collinerity



### How to Detect Multi collinerity?

1. Domain Knowledge
2. Scatter Plot of independent columns
3. Heatmap of Corr
4. VIF > 5%

### How to remove

1. Add more Data
2. Remove one of columns from dependents
3. Ridge / Lasso
4. PCA

### Does it affect all algorithms ?

There are 2 types of ML algo Parametric .and Non-Parametric.

Parametric are affected .

Paramteric Linear/KNN/Logistic/Naive Bayes Non-parametric(Tree Based)

### What is paramteric ?