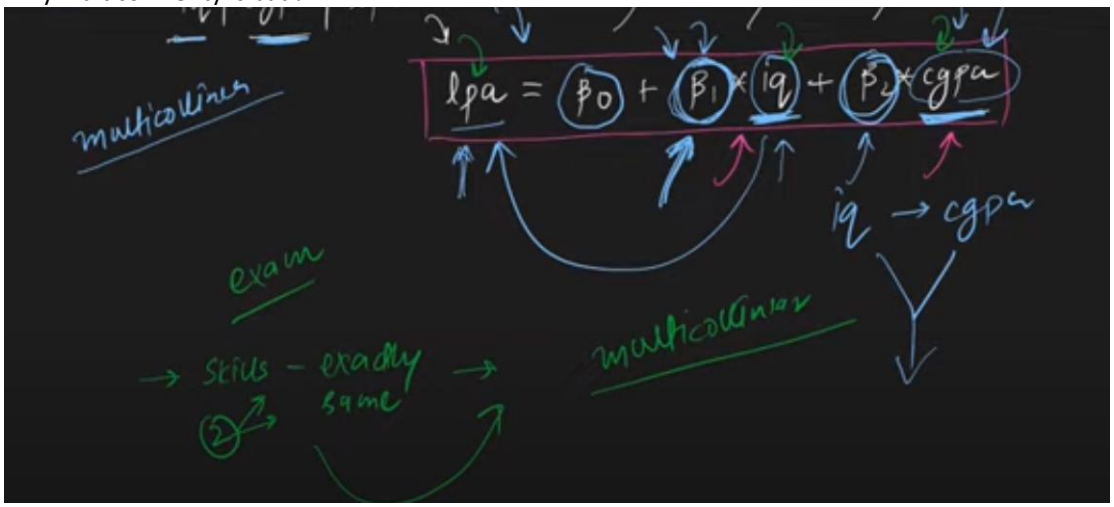
What is multi collinerity ?

Multi collinerity is phonemena where 2 or more independent variables are high correlated with high others .

Why mult collinerity is bad?

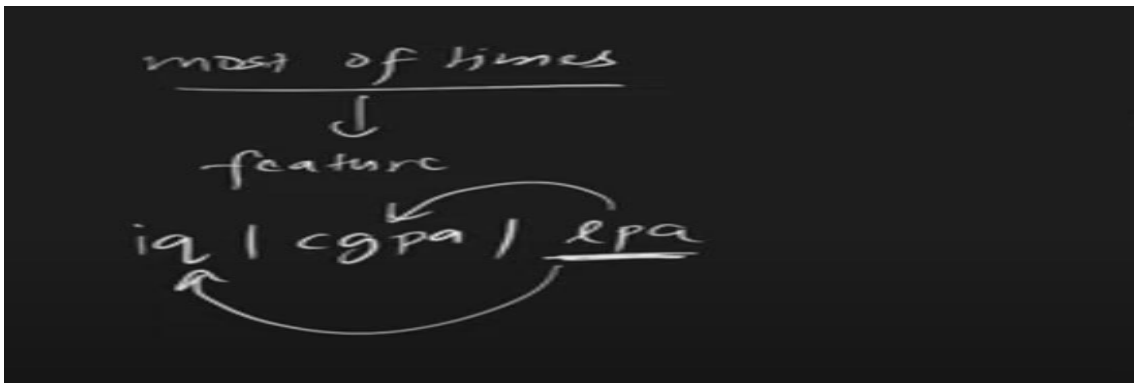



Is multi collinerity bad every time ?

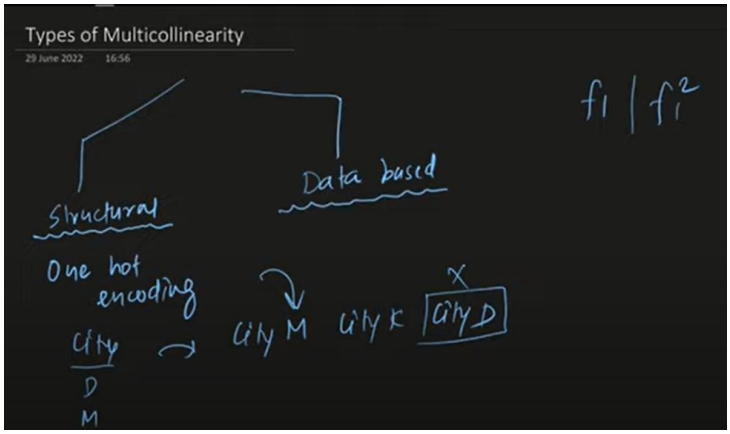
Model is used for 1. prediction  2.Inferenece

When we will use linear for inference then multi collinerity 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 ?