Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer: Optimal value for alpha for ridge regression is : 'alpha': 0.9

Optimal value for alpha for ridge regression is : 'alpha': 0.0001

When the value of alpha for the ridge regression was doubled, the R2 score on the test data was increased by approx.. 1%.

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer: For the alpha value which was identified in Ridge and Laso Regression, I was getting approx.. the same R2 score on training and test data.

Question 3

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Answer: A model can be considered as robust and generalisable when:

1. Model also has better accuracy on test data. There shouldn't be overfitting scenario.
2. VIF score the model should be less than 5.
3. P-value of the predictor variables should be less than 05 or 0.1

Question 4

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answer

The important predictor variables are:

BsmtFullBath, HeatingQC, 2ndFlrSF, MasVnrArea, GrLivArea, OverallQual