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 ridge and lasso are 0.6 and 0.001 respectively

For Lasso, more coefficients are becoming zero if alpha is doubled and r2 decreases on train dataset

For ridge, coefficients are decreasing if alpha is doubled and r2 score also decreases on train data set.

Important Predictor - MasVnrArea, BsmtFinSF1, 2ndFlrSF, KitchenQual_Fa, KitchenQual_Gd

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 - I will Lasso as Lasso and Ridge both are giving same r2 score. But lasso is reducing some of the feature which makes model less complex and keep it simple.

Question 3

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 - dropping first 5 important predictor, next five are as following - MasVnrArea, BsmtFinType2_None, KitchenQual_Gd, KitchenQual_TA, KitchenQual_TA

Question 4

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 – Below are the features which makes model robust and genearalisable –

1. R2 score of model - it should be between 70%-80%.
2. There should not be less difference between r2 score of train and test data.
3. There should not be overfitting and under fitting.
4. Regularization should be used to reduce number of features.
5. P values of variables should be less than 0.05
6. VIF values for the variables should be less than 5