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?

- Optimal Values for Ridge: 8
- Optimal Values for Lasso: 50

We can see that after doubling the alpha values, the r2 score decreases on training and test data and the RMSE increases.

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?

Ridge:

- 0.9454354136238928 –– r2_score – train
- 0.909276793523196 –– r2_score– test

Lasso:

- 0.9474552454906371 r2_score – train
- 0.913308626850547 r2_score – test

We will choose Lasso because:

- It gives higher r2 score for both test and train data set.
- RMSE is less than Ridge
- It helps us identify important variables.

Question 3

After building the model, you realized 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?

After dropping the top 5 we get next five most important predictors which are:

ExterQual, BsmtQual, OverallCond, KitchenQual,MSZoning

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?

The model be such that the test accuracy is close the training accuracy (r2_score)to avoid overfitting . The model should be predict for datasets other than the ones which were used during training. Outlier should be treated correctly. We should drop columns/rows very carefully. We need to impute values where applicable so that we do not lose the important feature. The one hot encoding should be used for categorical variables and also label encoding can be used. Trying out log transformation if the values are more pronounced right tail.