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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

- The optimal value of lambda for Ridge Regression = 6
- The Optimal value of lambda for Lasso = 0.0001

After doubling the value of the alpha:-

- The r-square of the train and test of Ridge becomes 0.9068 and 0.8634 earlier it was 0.9176 and 0.8709
- The r-square of the train and test of Lasso becomes 0.9165 and 0.8724
- Earlier it was 0.9239 and 0.8733

The most important predictors after the change is implemented are:-

- OverallQual_9
- TotRmsAbvGrd
- GarageArea
- FullBath
- Neighborhood_Crowfar

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'll choose Lasso as the model has too many features and many features are highly correlated to each other. Lasso will do feature selection and set the coefficient of some irrelevant and redundant features to zero.

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?

Answer:-

Five most important predictor variables now are:-

- 1stFlrSF
- TotRmsAbvGrd
- 2ndFlrSF
- YearRemodAdd
- Neighborhood_NridgHt

Question 4

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

Answer:-

The model is robust when changes in the data have little impact on its performance. general, we need to reduce the variance that could lead to bias. The model should be as simple as possible, though its accuracy will decrease it will be more robust and generalizable. It can be also understood using the bias tradeoff. The simpler the model the more the bias but less variance and more generalizable. Its implication in terms of accuracy is that a robust and generalizable model will perform equally well on both training and test data