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?

ANS:

1. The optimal value of alpha for ridge is: 1.0 and for Lasso is 0.001

2. The initial r2 score for train and test dataset for Lasso when alpla=0.001 is 0.9323520569674042 and 0.8612291009064933 respectively; and for Ridge when alpha = 1.0 is 0.951494613029963 and 0.8640161933454418 respectively. After doubling value of alpha, the r2 score for Lasso becomes 0.9160594727767669 and 0.8534844225263987 respectively ; and for Ridge it becomes 0.9487745092139233 and 0.8670146042044159 .

3. The most important predictor variables after the change is implemented will be: MSSubClass, BsmtFullBath, 2ndFlrSF, OverallCond, YearBuilt.

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?

Ans:

1. Comparing the 'R2 score Vs alpha' graphs of ridge and lasso regression, we can clearly see that lasso regression graph has more overlapping of train and test datapoints as compared to ridge regression.

2. Also the model parameter values for columns is better in Lasso as compared to Ridge.

3. Thus, our final prediction model would be Lasso Regression.

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?

Ans:

After removal of the 5 predictor variables from our current model, now the five most important predictor variables would be:

1. YearRemodAdd

2. GarageArea

3. BsmtFinSF2

4. SaleType_Oth

5. GarageYrBlt

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?

Ans:

For a model to be Robust and generalizable, we need to keep in mind that our model is fit to perform well on any unseen data and to accomplish this, we will have to make a model which deals with outliers and null values very efficiently without any complications and give the desired results. We also need to keep in mind that our model is simple enough and to achieve this, we can make use of the bias and variance trade-off method. i.e. if our model has more bias and less variance then we can say that it is more generalizable.