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>The Optimal Value of alpha for ridge and lasso regression:

Ridge: 0.2

Lasso: 0.01

If we double the Value of alpha for Ridge and Lasso:

Ridge: 0.4

Lasso: 0.02

The value of MSE in Lasso increases from 0.019 to 0.027

The value of MSE in Ridge has little decrease from 0.0196 to 0.0191

The predictor variables are not changed after doubling the alpha Value for Ridge and Lasso. The most important Variable in Predicting House Sale Price are:

1. OverallQual: Rates the overall material and finish of the house
2. LotArea: Lot size in square feet
3. RoofMatl: Roof material
4. GrLivArea: Above grade (ground) living area square feet
5. MSZoning: Identifies the general zoning classification of the sale

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 > The Optimal Value of alpha for ridge and lasso regression:

Ridge: 0.2

Lasso: 0.01

The Ridge Regression has better scoring metric (Rsquare) on train and test dataset. However, the predictor variables we get using lasso helps in determining House Prices much more effectively.

The value of MSE in Lasso is 0.019

The value of MSE in Ridge is 0.0196

Hence, we can use Lasso Regression as Business demands to identify Predictor Variables that has a significant impact on the House Prices.

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 >If we have top 5 predictor variables missing, we consider next 5 variables with a significant impact on the Response Variable.

These Feature could be:

1. OverallCond: Rates the overall condition of the house
2. GarageCars: Size of garage in car capacity
3. Condition2: Proximity to various conditions
4. Neighborhood: Physical locations within Ames city limits
5. LandContour: Flatness of the property

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 >The concept of Occam's Razor comes handy while selecting a Generalized model Occam's Razor: If we have two models that work about as well as each other, we should choose the simpler one.

Below points should be considered for a Robust Model:

- Simpler models are more generic and widely Applicable
- Complex Models tend to change with changes in Training Dataset
- Simpler models has Low Variance, High Bias. Contrary to Complex Models with low Bias and High Variance.

- Complex Models tend to perform extremely well on Training Dataset but fail on Testing Set. However, Simpler models might make some error on Training Set but works well with the Unseen Data or Test Data.
- Bias-Variance Tradeoff should be taken into consideration based on Business Requirements

Regularization can be used for making a Simple and Robust Machine Learning Model. For Regression, Regularization involves adding an Error Term to the Cost Function.

Error term : RSS + alpha * (sum of absolute value of coefficients)