Problem Statement - II

Assignment Part - II

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?

Ridge Regression

- The Alpha in Ridge is the penalty or the tuning parameter. Thus the ridge regression is sometimes called shrinkage penalty.This terms act as the deciding parameter themselves and determine the significant value for Xi and Bi Coefficient .
- If Alpha is large - The shrinkage grows and thecoeffients tends to be zero and thus the model will be over biased and we will get higher MSE.

Lasso Regression

- In Lasso less the alpha produces the same coefficient as the linear regression and when alpha si too high all the feature will be zero value
- Doubling the alpha value will make the value of coefficient to get shrink

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?

The lambda value in case of Ridge and Lasso is

1. Ridge - 5.0
2. Lasso - 0.03

The MSE are

1. Ridge - 0.017
2. Lasso - 0.014

The MSE is slightly lower in Lasso and it helps in feature selection and reduction by making unwanted coefficients to zero.
Thus Lasso helped in selecting 80 features from 215 features but the ridge took almost 200+ features which was more complex.
Thus Lasso made our choice better and helped in predicting the values and finding the variables predicting the price of the house.

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?

After the Lasso Model building the top 5 features are

- **GrLivArea** - Above Ground Living Area Square Feet
- **OverallQual**- Overall material and finish of the house
- **Neighborhood** - Physical locations within Ames city limits - **Northridge Heights**
- **Neighborhood** - Physical locations within Ames city limits - **Northridge**
- **GarageCars** - Size of garage in car capacity

Now ignoring these features in second model prediction we got the following list of 5 predictor values

- **BsmtExposure_Gd** - Refers to walkout or garden level walls with Good Exposure
- **RoofMatl** - Roof Material with **Wood Shingles**
- **YearBuilt** - Original construction date
- **BsmtFullBath** - Basement full bathrooms
- **LotFrontage** - Front Lot Area

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?

To make sure that our model robust and generalised we need to

1. Make sure model is resistant to outliers
2. It follows and handles error metric
3. It should have a effective training and should have undergone the regressions
4. The model should show similar results in the test and training set which means that model should have gone through all the possibilities.
5. The model should be more and more simpler
6. Creation of complex models will result in new data every time as they rely on multiple parameters.
7. Complex Model may try to change the model and which causes it to be unstable and sensitive to any changes in training data.

8. Regularisation helps in maintaining the model more simpler and this adds the absolute value to the cost and square parameters of the model.
9. Bias quantifies how good the model is with test data and thus we should need low bias and variance to determine the change in model regards to training data and accuracy will be balanced with this both parameters.