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 - Optimal value for alpha which I got for ridge regression is 0.3 and for lasso regression its 0.001. R2 value is almost the same when I doubled the value of alpha for both ridge and lasso regression.

Important predictor variables for ridge are (top 10) –

- GrLivArea
- OverallQual
- MSZoning_FV
- MSZoning_RL
- TotalBsmtSF
- MSZoning_RH
- GarageArea
- MSZoning RM
- LotArea
- GarageQual

Important predictor variables for Lasso are (top 10) –

- GrLivArea,
- OverallQual,
- TotalBsmtSF
- GarageArea
- BsmtFinSF1
- LotArea
- MSZoning_FV
- GarageQual
- MSZoning_RL
- WoodDeckSF

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 - R2 value for train and test is almost same for both Ridge and Lasso. I will prefer using Lasso as it has reduced the number of predictor variables and helped in feature selection. It helped in selecting 28 features of 30.

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 excluding the top 5 predictor variables in lasso, below are the important variables:

- GarageArea
- LotArea
- LotFrontage
- BsmtFinSF1
- Neighborhood_Crawfor

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?

We can make the model as robust as possible by ensuring model is not overfitting or underfitting and is as simple as possible. Simpler models are more robust as they are not sensitive to the specifics of the training dataset. Simpler models are more generic as well they are more widely generalisable.

Regularizations is one of the techniques used to simplify models. It controls the model complexity. In Regression model, this involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.

Regularization can lead to more accurate models as it avoids overfitting. The shrinking coefficients minimizes the bias and improves the variance of models.