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:

Alpha for ridge = 4

Alpha for ridge with rfe = 0.3

Alpha for lasso = 0.0004

Ridge Regression: If we double the Alpha, we are increasing the penalty term in the cost function thereby increasing the bias and decreasing the variance of the model. Ridge Regression includes all variables in the final model unlike Lasso Regression. Coefficients are decreased in magnitude. R2 values are reduced slightly

Top features are GrLivArea, 1stFlrSF, TotalBsmtSF, OverallQual_excellent, 2ndFlrSF

Lasso Regression: Like Ridge regression answer. Few coefficients are made 0 with doubling the alpha. R2 values are reduced slightly. Few of the features are eliminated as coefficients are made 0.

Top features are GrLivArea, TotalBsmtSF, OverallQual_excellent, Neighborhood_MeadowV, OverallQual _vexcellent

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: I will choose to apply Lasso regression model as it has feature elimination as well embedded in the model. R2 score of Lasso is slightly higher compared to Ridge regression model. Feature elimination is made easy in Lasso regression if not it would be laborious process

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?

Ans: Five most important predictor variables after removing five most important predictor variables in the lasso model are

Taking the absolute values of coefficients as both positive and negative needs to be considered

1stFlrSF, 2ndFlrSF, Neighborhood_MeadowV, BsmtFinSF1, Neighborhood_StoneBr

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:

- Model should not be impacted by outliers in the data
- Training/Test accuracy should be very close which makes it more generalizable and robust
- Model accuracy should be high for datasets other than the ones used for training the model
- Minimal weightage should be given to the outliers so that accuracy of the model is high. To make sure of this, outlier treatment needs to be done
- Regularization can be done while creating a model so that penalty term takes care of model complexity.