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 – When alpha is 2 the test error is minimum, however when we increase the alpha value from 0, the error term decreases. Hence decided to go with alpha as 2.

Lasso – We have kept a small value as 0.01 and when we increase the vale of alpha the model penalises and make most of the coefficients value to zero.

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?

Since Lasso does variable selection as well, we will go with Lasso Regression. And also, the model performs pretty well.

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?

- GRLIVArea
- OverallQual
- OverallCond
- TotalBSmtSF
- Garage 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?

We need to try to make the model simple. Also it depends on the Bias and Variance Trade off. Model should not be too simple to under fit the data and also not too complex to over fit the data. Based on the bias and variance trade-off, we need to select an optimal simple model which can be generalised.