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 is 5.0 and 0.001 respectively. If we choose double double the value of alpha for both ridge and lasso. There is an increase in r2 score of test set by 0.37 and 0.03 for ridge and lasso. The most important predictor variables after the change in implemented are:

1. MSSubClass
2. KitchenAbvGr
3. 1stFlrSF
4. TotalBsmtSF

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: We will choose the optimal value of lambda for ridge and lasso regression during the assignment that we have got in GridSeachCV in 1st because there is no huge increment in mean test score if we double the alpha value. So we will go with alpha value 1.0 and 0.001 for ridge and lasso regression respectively.

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: The five most important predictor variables after excluding the five most predictor variables from previous model are:

1. LotArea
2. GarageArea
3. GrLivArea
4. LandContour_HLS
5. LowQualFinSF

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: We can make sure that a model is robust and generalisable by checking the graph of mean test and train score at particular alpha value. If there is a big difference in mean test score and mean train score then the model is actually overfitting and it

will not perform well in unseen data. If the difference is very low then we are confident that the model will perform well in unseen data.