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:

In case of ridge regression the alpha value increase as error term decreases. If you look into the graph, as error value started to decrease, alpha started to remain same. When the alpha value is 0 the error is started remain same and for alpha value 2 it started to remain same.

In case of lasso regression model, kept small value which is 0.02 or o0.00 but when increase the value the alpha, model tries to penalize more and makes the most of the coefficient values to 0.

Below are the most important variables.

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:

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: Below are the important predictor variables:

- GrLivArea
- OverallQual
- OverallCond
- Total BsmtSF
- GarageArea

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:

The model should be as simple as possible as it will be more robust and generalizable but tradeoff is accuracy. As we know the simpler is the mode more the bias but less variance and more generalizable

Bias: It is the error in the model. High bais means model is not able to learn from the data and model performs poor on training and testing data.