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:

As my inicial value of the alpha are 0.005 and 4 as it is double the values it become 0.010 and 8 for both ridge and lasso regression

And I don’t see much r2 score difference in it.

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:

Lasso and ridge there are no much difference in r2 score and predictions lasso have advantage over ridge it will make most of variables to 0

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 variables after creating the model again (lasso) by excluding the top five features are:

1. LotFrontage

2. Totalbsmtsf
3. Grilvarea
4. MasVnrArea
5. OverallQual

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: To keep the model simple and robust by having only the significant variables and balancing the bias and variance.

By ignore the little bias, we can reduce the larger amount of variance and so keeping the model simple and robust.

More bias and less variance will cause the model to become underfit likewise more variance and less bias makes the model overfit.