

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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

The optimal value of ridge and lasso regression is as Alpha value 0.001, Lasso regression 0.89

R2 values slightly dropped if alpha value is doubled like Alpha value 0.002, Lasso regression 0.88

The most Predictive variable is GarageArea(0.88)

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?

Once model is created in Ridge and Lasso, we see that r2 scores are same for both but lasso will penalize more on the dataset and can also help in feature elimination

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?

LotArea

OverallQual

YearBuilt

BsmtFinSF1

TotalBsmtSF

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?

The model should be accurate for datasets other than the ones which were used during training. Too much importance should not be given to the outliers so that the accuracy predicted by the model is high. The outliers analysis needs to be done and only those which are relevant to the dataset need to be retained. Those outliers which it does not make sense to keep must be removed from the dataset.