

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:

Optimal Value for Ridge regression is 1

Optimal Value for Lasso regression is 0.0001

If value is increased more than optimal, they will become underfit.

OverallQual, GrLivArea can be the most important predictor

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:

As the number of significant variables are 5, which is relatively small it is better to approach with Lasso Regression.

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:

TotalBsmntSF

1stFlrSF

FullBath

YearBuilt

TotalRmsAbvGrd

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 make sure a model is robust and generalisable

1. Remove Outliers and make the model more Outliers resilient
2. Apply log transformations on the data if needed
3. Instead of relying on Mean Square error need to calculate and evaluate more on RMSE and MAD metrics
4. Simpler models are more acceptable compared to complex models