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:

Value near by 10^2 is optimal value for ridge and lasso as per analysis.

Doubling the value of alpha will make model insignificant to available variables.

After implementing changes Lasso gives better result in term of r-squared.

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:

Optimal value of lambda for ridge: 5

Optimal value of lambda for lasso: 100

I will choose lasso as the MSE value of lasso is lowest among both.

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:

Below variable are important as per EDA analysis:

Foundation

1stFlrSF

TotalBsmtSF

GrLivArea

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?

As RFE is applied and top 50 feature are taken for further analysis.
And Lasso also eliminated many variables and provided 24 features which impacting on prediction of price.

As MSE value for Test dataset coming lowest among all three analysis, Lasso providing best result.
Also R2 and Adjusted R2 score do not have much gap and all predictor are able to explain model.