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?

Answer:

Optimal Value for ridge alpha is 10 and for lasso is 100.

After Double alpha the value will be 20 and 200.

Changes will be as follows:

Lasso: More Features removed but r2 score dropped by 1%.

Ridge: Coeff value increased.

Important Predictor Variables Are:

- Neighborhood NoRidge (Northridge)
- Neighborhood Veenker (Veenker)
- Neighborhood NridgHt (Northridge Heights)

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 per my understanding Lasso is a better option as it provides feature selection option

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:

predictors 1stFlrSF, MSSubClass_90, MSSubClass_120, TotalBsmtSF, HouseStyle_1Story

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:

We can consider features as below:

1. VIF should be <5.
2. Model Accuracy >75%
3. P value should be < .05