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:

a. For ridge regression the optimal value of alpha is 10, and for lasso regression the optimal value of alpha is 100.

b. If we choose double the value of alpha fro both ridge and lasso regression, then the following changes will happen:

I. In case of Ridge regression coefficient values increases as alpha increases, and r2_score of train data drops from .807 to 0.45.

II. In case of Lasso regression as the alpha value increases more features are removed from model and r2_score is also dropped by 1% in both test and train data.

c. The most important predictor variables after the change is implemented are:

1. *Neighborhood_NoRidge*
2. *Neighborhood_NridgHt*
3. *2ndFlrSF*
4. HouseStyle_1Story
5. 1stFlrSF

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:

After determining the optimal value of lambda for ridge and lasso regression, we should choose lasso because it gives us feature selection options.

Lasso regression removes unwanted features from model without affecting the model accuracy, it makes our model generalized.

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 most important top 5 predictors are as follows:

1. OverallQual
2. TotalBsmtSF
3. Foundation_Slab
4. GarageType_BuiltIn
5. Neighborhood_Crawfor

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:

a. To make model robust and generalisable 3 features are required:
    1. Model accuracy should be greater than 70 – 75% range.
    2. P-value of all the features should be less than 0.05
    3. VIF of all the features should be less than 5.

b. If we don't generalize the model then even if it provides accurate prediction on the trained data, but as soon as new data comes the model starts giving inaccurate predictions.