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:

Using RidgeCV and LassoCV classes, the alpha value for ridge regression was found to be 1 and for lasso regression the alpha value is 0.0001.

If the alpha value is doubled the model complexity will be reduced. Hence, tends to increase bias and decrease variance. This may lead to underfitting issue.

GrLivArea, OverallQual, GarageCars, Neighborhood_NoRidge, RoofMatl_WdShngl and so on…

Above are the most important predictors variables for lasso regression after implementation.

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one would you choose to apply to and why?

Answer:

I'll choose lasso regression with 0.0001 alpha value as its accuracy of 88.61% for training dataset and 87.88% for testing dataset with no of 108 predictor variables which makes model less complex with moderate accuracy.

Question 3

After building the model, you realized 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:

After retraining the model, following are the topmost important predictors:

Neighborhood_NridgHt

Neighborhood_StoneBr

LotArea

2ndFlrSF

OverallCond