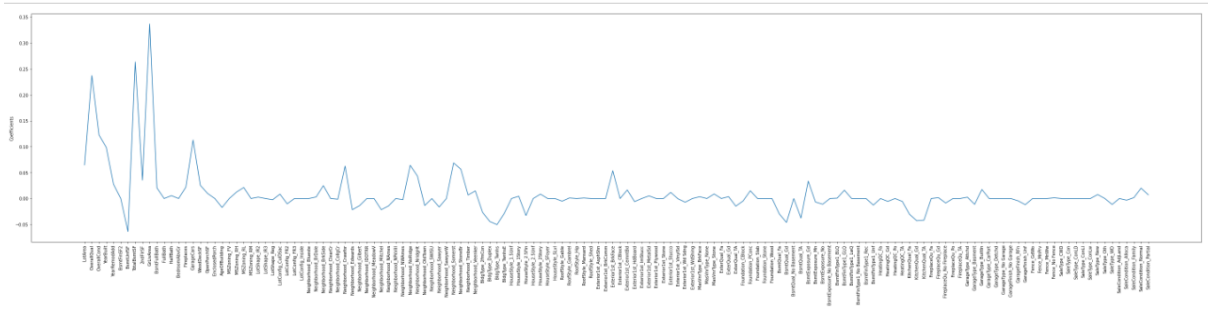
Question 1

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?

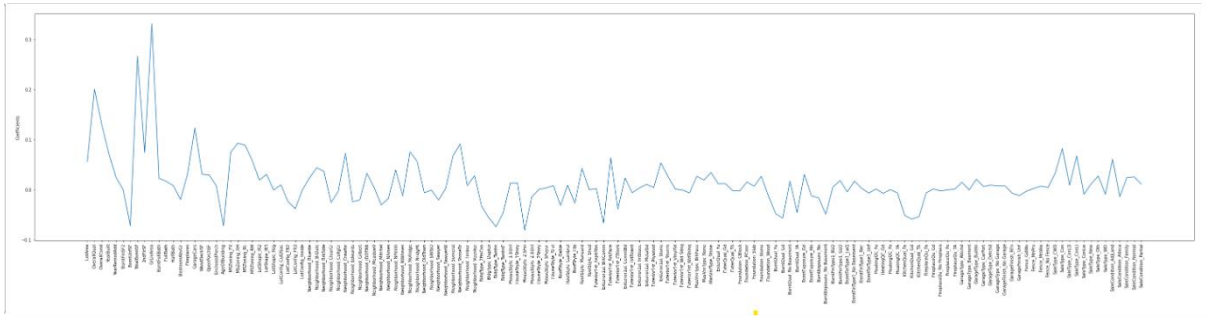
- Alpha for Lasso: 0.0001
- As per the question, if alpha for Lasso is doubled then 0.0002
- If the alpha is doubled in the Lasso then r2_score for both training set is same. Also r2_score for test is decreased when Alpha is doubled.

Below graph is the Coefficients values and its names for Lasso Regression:



- Alpha for Ridge: 0.1
- As per the question, if alpha for Ridge is doubled then 0.2
- If the alpha is doubled in the Ridge then r2_score for both training and test set is same.

Below graph is the Coefficients values and its names for Ridge Regression:



Important variables in Lasso after the change implemented is :

GrLivArea	0.336654
TotalBsmtSF	0.263436
OverallQual	0.237323
OverallCond	0.122696
GarageCars	0.112816
YearBuilt	0.098529
Neighborhood_Somerst	0.068622
LotArea	0.064826
Neighborhood_NoRidge	0.064117
Neighborhood_Crawfor	0.062592

Important variables in Ridge regression after the change implemented is :

GrLivArea 0.331268
TotalBsmtSF 0.266605
OverallQual 0.200655
OverallCond 0.130759
GarageCars 0.123107
MSZoning_RH 0.092882
Neighborhood_StoneBr 0.091234
MSZoning_RL 0.088954
SaleType_Con 0.082169
Neighborhood_NoRidge 0.075295

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?

- Alpha for Lasso: 0.0001
- Alpha for Ridge : 0.1
- Choosing between ridge and lasso completely depends on the output that we need.
- If we want to choose only the columns where the coefficients are not zero then we go for Lasso regression.
- If we have more number of predictor variables than the number of observations then ridge will the best technique

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?

Before with alpha = 0.001, the top 5 variables in lasso regression are below :

- GrLivArea
- TotalBsmtSF
- OverallQual
- OverallCond
- YearBuilt

After removing these variables, again if we build the lasso model again then the below are the 5 important predictor variables:

- Exterior1st_Stone
- GarageCars
- 2ndFlrSF
- Neighborhood_StoneBr
- Fireplaces

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?

- Model should be overfitting or under fitting.
- It should not be the case that model performs well on the training data set where as it fails to predict the test dataset
- Model should be able to generalise the unseen data
- Too complex model will have high accuracy
- To make model more generalisable and robust we need to decrease the variance which lead to bias.