Name: Mehta Akshat Rohitkumar

Roll No:4609206

Batch:ML-C45

Advanced Regression Assignment

(Subjective Questions)

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?

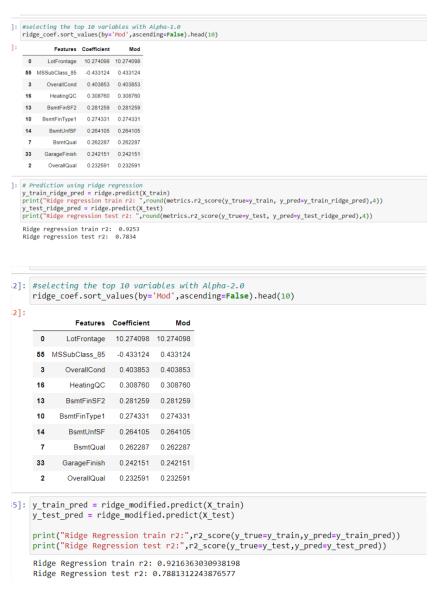
Answer 1

Optimal Values of Alpha

- ✓ Ridge Regression Model-1.0
- ✓ Lasso Regression Model-**0.0001**

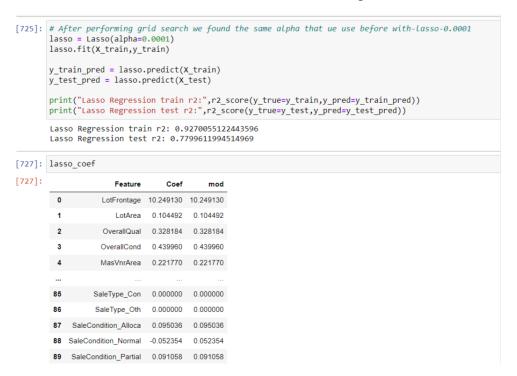
When Double value of Alpha of Ridge Regression-

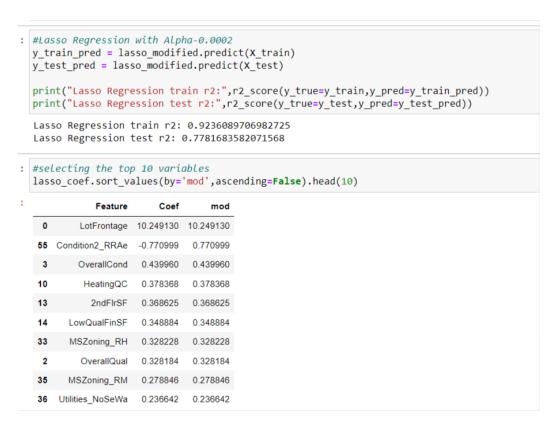
Feature Coefficients and R2 Value is Same for Both Alpha Values(From Below 2 Pics).



When Double value of Alpha of Lasso Regression-

Feature Coefficients and R2 Value is Same for Both Alpha Values(From Below 2 Pics).





• Overall since the alpha values are small, we do not see a huge change in the model after doubling the alpha.

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?

- ✓ The optimum lambda value in case of Ridge and Lasso is as follows:-
 - \circ Ridge -1
 - Lasso 0.0001
- ✓ The Mean Squared Error in case of Ridge and Lasso are:
 - o Ridge 0.0018396090787924262
 - o Lasso 0.0018634152629407766
- ✓ The Mean Squared Error of both the models are almost same.
- ✓ We will make use of Lasso Regression model because it is using less numbers of variables and giving almost the same accurate. Its more efficient model than Ridge regression model.

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?

```
#selecting the top 5 variables in lasso model with alpha=0.001
lasso_coef.sort_values(by='mod',ascending=False).head(5)

Feature Coef mod

LotFrontage 10.249130 10.249130

55 Condition2_RRAe -0.770999 0.770999

3 OverallCond 0.439960 0.439960

10 HeatingQC 0.378368 0.378368

13 2ndFirSF 0.368625 0.368625

X_train_new = X_train.drop(['LotFrontage','Condition2_RRAe','OverallCond','HeatingQC','2ndFlrSF'],axis=1)
X_test_new = X_test.drop(['LotFrontage','Condition2_RRAe','OverallCond','HeatingQC','2ndFlrSF'],axis=1)
X_test_new.head()
X_train_new.shape
```

- ✓ The five most important predictor variables in the current lasso model is:-
 - 1. LotFrontage
 - 2. Condition2 RRAe
 - 3. OverallCond
 - 4. HeatingQC
 - 5. 2ndFlrSF

✓ We build a Lasso model in the Jupiter notebook after removing these attributes from the dataset.

The new Top 5 predictors are:-

- 1.LotArea
- 2.KitchenAbvGr
- 3. Condition2_RRAn
- 4. MasVnrArea
- 5. OverallQual

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 Per, Occam's Razor— given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:-
- ✓ Simpler models are usually more 'generic' and are more widely applicable
- ✓ Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
- ✓ Simpler models are more robust.
 - Complex models tend to change wildly with changes in the training dataset.
 - Simple models have low variance, high bias and complex models have low bias, high variance
 - Simpler models make more errors in the training set. Complex models lead to overfitting — they work very well for the training samples, fail miserably when applied to other test samples

- ✓ Therefore, to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use.
- ✓ Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use.
- ✓ For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.
- ✓ Also, Making a model simple leads to Bias-Variance Trade-off:
 - A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.
 - A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed.
- ✓ Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high.
- ✓ Variance refers to the degree of changes in the model itself with respect to changes in the training data.
- ✓ Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph.

