**Question 1**

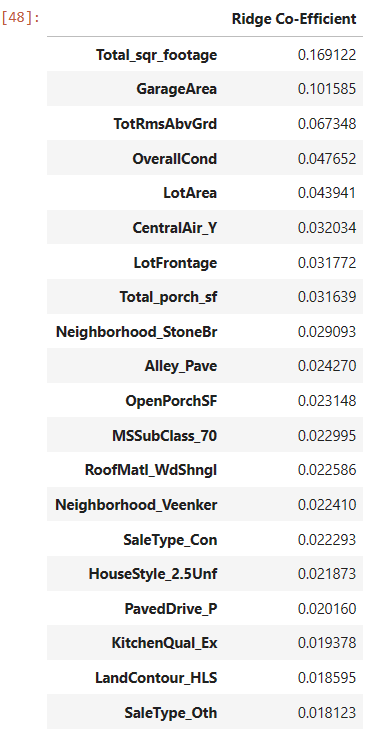
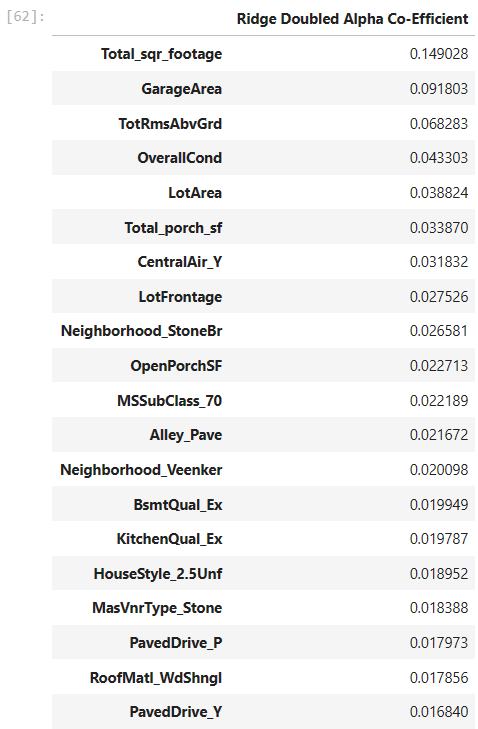
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

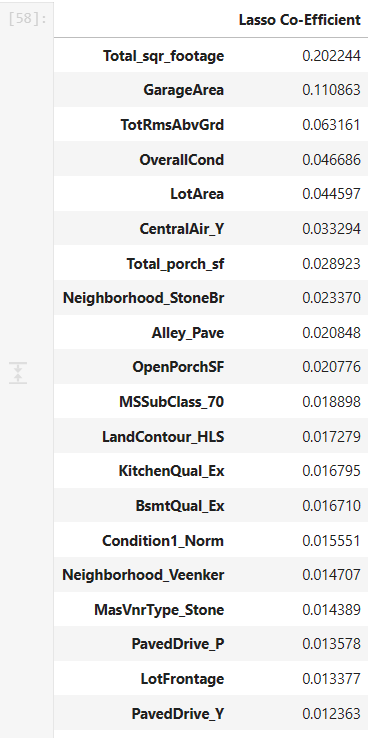
Answer –

The optimal value of alpha for Ridge is 2 and for Lasso it is 0.001. With these alphas the R2 of the model was approximately 0.83.

After doubling the alpha values in the Ridge and Lasso, the prediction accuracy remains around 0.82 but there is a

small change in the co-efficient values. The answer was prepared in the. ipnyb. Below are the changes in the co-efficient.

 A screenshot of a computer

Description automatically generated

**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:** We will choose Lasso as its giving feature selection option also. It has removed unwanted features from model without affecting the model accuracy. Which makes are model generalized and simple and accurate.

**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 five most important predictor variables in the current lasso model is:-

* Total\_sqr\_footage
* GarageArea
* TotRmsAbvGrd
* OverallCond
* LotArea

This based on Lasso model build on top of given CSV data.

The R2 of the new model without the top 5 predictors drops to .73. The Mean Squared Error increases to 0.0028575670906482538.

**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**: To make the model more robust and generalizable, make the model simple but not

simpler which will not be of any use.

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 data set.
* 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 another test.
* samples

Feature wise to make model robust and generalisable 3 features are required:

1. Model accuracy should be > 70-75%: I our case its coming 80%(Train) and 81%(Test)

which is correct.

2. P-value of all the features is < 0.05

3. VIF of all the features are < 5