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 new model is created and demonstrated in the Jupiter notebook.

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:-

The optimum lambda value in case of ridge and lasso is

as follows:- Ridge – 2

Lasso – 0.0001

The mean squared error in case of ridge and lasso

are : Ridge – 0.0018396090787924262

Lasso – 0.0018634152629407766

The mean squared error for both the models are almost

same. Since lasso helps in feature reduction.

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 :-

1. Total_sqr_footage
2. GarageArea
3. TotRmsAbvGrd
4. OverallCond
5. LotArea

We build a lasso model in the Jupiter notebook after removing these attributes from the dataset. The R2 of the new model without the top 5 predictors drops to 0.73

The mean squared error increases to 0.0028575670906482538.

The new top 5 predictors are :-

1. LotFrontage
2. Total_porch_sf
3. HouseStyle_2.5Unf
4. HouseStyle_2.5Fin
5. Neighborhood_Veenker

Question 4: How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer: -

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:-

1. Simpler models are usually more generic and are more widely applicable
2. Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
3. Simpler models are more robust.

   Complex models tend to change widely 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 for other test.

Therefore to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use.

Also making a model simple leads to bias- Variance Trade-off:

1. 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.
2. A simpler model that abstracts out some pattern followed by the data points given is unlikely to change widely even if more points are added or removed.